The Matter Compiler—towards atomically precise engineering and manufacture

Dung Q. Ly,[1] Leonid Paramonov,[1] Calvin Davidson,[5] Jeremy Ramsden,[2] Helen Wright,[3] Nick Holliman,[4] Jerry Hagon,[6] Malcolm Heggie[5] and Charalampos Makatsoris[1,*]

[1] *School of Engineering and Design, Brunel University London*
[2] *School of Applied Sciences, Cranfield University*
[3] *Department of Computer Science, University of Hull*
[4] *Innovative Computing Group, School of Engineering and Computing Sciences, Durham University*
[5] *Department of Chemistry, University of Sussex*
[6] *School of Chemistry, Newcastle University*

The Matter Compiler is a novel framework for a computer toolset that allows the design of arbitrary molecular structures and their assembly by means of radical synthesis pathways that resemble "pick and place". Such synthesis pathways have been demonstrated experimentally in the literature in the recent past using scanning probe microscopes (SPM). The Matter Compiler framework entails first-principles modelling and novel control approaches that together assess the feasibility of user-defined molecular structures, screen alternative options for their assembly and finally suggest synthesis by means of physical equipment such as scanning probe microscopes. Nanoscale structure design is driven by first-principles calculations as well as by any relevant physical and chemical knowledge that can be accessed by the system from chemical databases. The first main consideration is therefore a suitable modelling methodology enabling the design and assessment of user-defined molecular structures. This entails the computer representation of such structures, methods to access and use available chemical and physical knowledge and, subsequently, an overall approach to plan the synthesis of such structures. Currently, it is assumed that assembly takes place on suitable substrates molecule-by-molecule and uses first-principles calculations to determine synthesis pathways (plans) on those substrates. The second main consideration is to use those synthesis plans to determine how

* Corresponding author: Tel.: +44 1895265063. E-mail: harris.makatsoris@brunel.ac.uk

Nanotechnology Perceptions **7** (2011) 199–217
Received 4 November 2011

synthesis will take place within a physical instrument. For scanning probes, for example, this translates to a collection of set points for the instrument controller to follow that collectively comprise the trajectory of the scanning probe for each step of the process. Image processing for monitoring progress, visualization and haptic interaction are also part of the Matter Compiler—they provide insight, feedback and determination of intervention strategies in case of errors. In this paper a description of the Matter Compiler methodology and framework architecture is presented. Some computational results are also presented.

1. Introduction

The ambition to assemble molecules and materials under atomically precise control demands a big leap forward in control engineering and computer science. Is it possible to anticipate the properties and needs of a *"universal nano-assembler"*? If so, there is a need for a high level instruction language and a computer compiler that translates commands in this language into instructions for such a *"nano-assembler"* to execute. This development requires a breakthrough in understanding of chemical synthesis, which must embrace the radically new "pick and place" assembly method that is now possible in scanning probe microscopy (SPM). Such deterministic synthesis, driven by scanning probes, has been demonstrated experimentally in the recent past: The experiments involved manipulating molecules [1–5]; creating and breaking bonds at the single atom level [6]; and synthesizing molecules [3, 7] using scanning probes on specific substrates. This type of research is, to many, a definitive example of the potential of nano-engineering. However, neither a systematic methodology nor enabling software technology for the mechanistic realization of such type of synthesis of arbitrary molecules is available. Such capability would offer tremendous benefits to the design and synthesis of new materials and the devices of the future. For instance, one can envisage the precise assembly of drug molecules to engineer crystal seeding. The seed can then be allowed to grow conventionally, resulting in a user-designed crystal that has desirable physical properties such as particular dissolution rates. Another example is the engineering of surfaces by creating atomically precise patterns with specific properties and function, such as an atomic switch for ultrahigh density storage.

In this paper we attempt to anticipate developments in this area and provide a prototype framework for the engineering control and computer science aspects of computer-directed molecular assembly. We use as inputs data from SPM experiments, energy landscapes for "pick and place" reactions and the vast knowledge base of classical synthetic chemistry. Target structures are analysed into simpler molecules using an algorithmic approach that is analogous to existing methodologies such as retrosynthesis. This is then followed by calculating "pick and place" trajectories deduced from energy landscape calculations. The approach has been captured in an object-oriented software framework.

The outcome is a specification and prototype design of what we call "Matter Compiler" methodology. It addition, a detailed investigation using first-principles simulations of Si(100) as a substrate for adsorption and reactions involving organic molecules is also presented. In particular, our case study calculations relate to the manipulation of a Si(100) 2×1 reconstructed surface such that the adatoms of the Si dimers on the surface switch their position with the lower atoms in the presence of a scanning probe acting like an atomic 1-bit switch. Our investigation

demonstrates the steps of the Matter Compiler process. The results of these simulations can be regarded as the future inputs into a knowledge base as well as providing an understanding of the behaviour of this type of surface in the presence of scanning probes, modelled by the system as molecular structures that are controlled by a computer or directly by a user.

In Section 2 the overall methodology is described and also an account of the main use cases for the Matter Compiler is provided. In Section 3 the Matter Compiler framework is presented, followed by the case study description and results. Then a discussion on visualization and haptic interfaces is provided, together with a description and demonstration of an approach for scanning tunnelling microscopy image simulation before a summary of conclusions and an outline of future work are given.

2. System description and use cases

The Matter Compiler has the long-term aim of producing full instructions to direct the synthesis and assembly of arbitrary nanostructures, informed by all relevant physical and chemical knowledge. It is being developed as a novel general-purpose software system that will assist researchers in nanotechnology and surface science to determine optimal reaction pathways for the fabrication of arbitrary molecular structures. Currently design and development assumes that fabrication of such surfaces will take place on surfaces and will be driven by scanning probes. However, the core of the system (i.e., the data structures and methodology) could be applicable in different synthesis regimes and with approaches not currently available or known.

The system is being developed for the following three *use cases*:

- planning synthesis pathways for currently unknown molecules and structures or for those lacking a systematic pathway for their synthesis driven by a scanning probe;
- predicting new synthesis pathways of known structures driven by a scanning probe on a given surface;
- employing first-principles energy landscape calculations combined with suitable path-planning algorithms for determining the trajectory of a scanning probe tip required to execute the fabrication of the target nanostructure.

"Pick and place" synthesis driven by scanning probes has been demonstrated experimentally in the recent past [1–7].[1] However, those experiments were designed specifically for the particular systems under investigation and also involved a large number of failed attempts before achieving the desired outcomes.[2] The Matter Compiler aims to introduce a systematic methodology and, eventually, a software implementation that predicts manipulation or synthesis pathways that can take place experimentally. The Matter Compiler can then translate the results of those calculations into the form of a reusable set of instructions that a physical instrument can execute. This capability will enable the mechanistic and repeatable synthesis of arbitrary molecules in user-defined systems.

The main assumptions employed in designing the framework and the case study are:

- synthesis takes place within an SPM on a given surface. At its core, the framework is not

[1] The actual process is, however, an extremely laborious manually driven operation.
[2] The iconic assembly of 18 Xe atoms on a nickel surface to form the letters "IBM" took almost 24 h [1].

restricted to SPM although all first-principles calculations in the case studies are solely based on SPM;
- a single scanning probe executes synthesis. The data structures support any number of scanning probes present simultaneously although the case study in this paper only considers a single probe being present in the unit cell;
- in the case study the tip was assumed to have three degrees of freedom and is thus able to translate in any desired direction along any Cartesian axes. The data structures in the Matter Compiler framework do, however, support any number of degrees of freedom;
- bond formation and bond breaking take place only by the mechanical force exerted by the tip in chemically compatible systems;
- at present small organic molecules are considered as targets and broken down into simpler molecules or synthons using the same strategies as in retrosynthesis [8, 9].

Until now we have primarily pursued the understanding, through first-principles models, of "pick and place"-type synthesis by studying the behaviour of certain systems that are popular in the research community and also have many industrial applications. One such system is Si(100). Our studies have focused on complex unit cells comprising model scanning probes, adsorbates and substrates with the aim of understanding how they would all interact in a "pick and place" scenario assuming that the process takes place within a scanning probe microscope (SPM). These studies, which involved a large number of simulation results, have been crucial for elucidating sufficient understanding of how the design of such experiments could be automated and subsequently executed automatically. Our understanding was then captured in suitable data structures, an application framework, and the overall specification of the Matter Compiler.

The Matter Compiler framework is designed as a client to external codes for density functional theory (DFT) and for atomistic and molecular dynamics calculations. The Matter Compiler coordinates simulations performed by such systems, which can relate to structural optimizations, tip–substrate interactions, transition states and, eventually, "pick and place" reactions. Analysis of targets into synthons currently relies on external databases such as Beilstein, PubChem and ChemSpider [10–12]. This analysis is of importance as it not only determines those synthetic precursors that are potentially easier to manipulate with a scanning probe but also identifies the molecular fragments that can be obtained via conventional chemical routes. Then a calculation of transition states follows, leading to an appropriate synthesis sequence (pathway). To determine reaction sequences that can be realizable experimentally, a large number of options are available. They comprise combinations of potential synthons, alternative synthetic equivalents, different substrates or even a variety of scanning probes with appropriate functionalizations. Suitable object-oriented data structures have been designed to represent these considerations, which the system handles as inputs into the synthesis schemes determined by the system.

A schematic diagram of the Matter Compiler is shown in Figure 1. At the core of the framework is a first-principles [13, 14] code for energy landscape calculations. The use of reactive force fields (and, in particular, ReaxFF [15]) to complement DFT with fast screening of potential structures and approaches to fabricate targets has also been considered. This will extend the capability of the Matter Compiler to execute simulations across multiple length and time scales in the future. It is to be noted that in contrast to existing uses of quantum mechanical

and molecular mechanics calculations, which are primarily performed for explanation or analysis of behaviour in experiments already performed, the Matter Compiler uses such calculations to design the synthesis protocols, based on which experiments will be subsequently carried out. The system then translates simulation outputs into set points for the control system of the instrument to follow while executing the experiment.

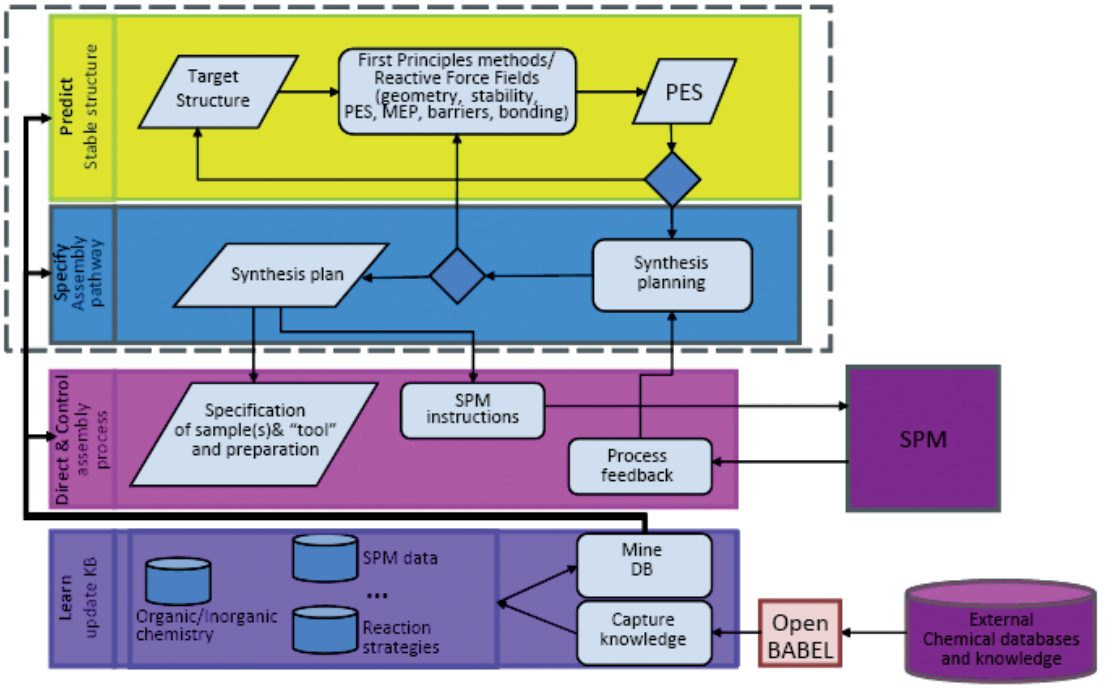

Figure 1. Matter Compiler methodology.

3. Design of the Matter Compiler

The Matter Compiler framework comprises 12 components in total. Each component contains the data structures and classes implementing specific functionality of the system. Table 1 provides the names and short descriptions of the main components of the system.

There are three main important data structures of the system. The first describes molecular structures. This represents a molecular structure or a crystal structure as a graph. This representation is widely used in chemoinformatics. It allows the use of graph traversal and matching algorithms that are used for searching molecular substructures, identifying key features such as cycles in the graph (e.g., carbon rings) and also allow efficient storage and retrieval from a database. Of particular importance is the finding and matching of substructures. With this functionality molecular fragments or moieties can be discovered within molecular graphs and tagged. Knowledge of the existence and position of such fragments within a given molecule is used as an input to the Synthesis Planner, which determines how to analyse the target structure into substructures, or synthons [8, 9], which must be sourced to be eventually assembled together to form the ultimate, desired, structure. Furthermore, the algorithm

Table 1. The main components of the Matter Compiler, reflecting the methodology shown in Figure 1.

Component name	Description
Nanostructure Designer	Comprises data structures that represent an arbitrary molecular structure including crystals by means of the atom connexion table and algorithms that traverse and manipulate those molecular graphs.
Synthesis Pathway Planner	Comprises data structures that represent synthesis pathways for a target in a hierarchical tree-type structure and the algorithms to create, traverse and manipulate those data structures. In addition it uses the interfaces to first-principles and molecular dynamics codes that plan and control atomistic and molecular simulations. This is achieved by path-planning algorithms that determine simulation snapshots, initializing them appropriately and then assembling results into a synthesis pathway plan, which is then mapped by the Transition Planner (described below) into macroscopic control instructions to the physical instrument that will execute synthesis.
Transition Planner	This component is driven by the Synthesis Pathway Planner. It is responsible for mapping results from atomistic or first-principles simulations onto macroscopic instructions that the controller of a physical instrument can execute.
Knowledge Base	This component is responsible for the capture, persistent storage and management of the synthesis protocols that have been produced by the Matter Compiler. It also contains interface classes to external databases.
Molecular Simulator	Contains interface classes that connect to external simulators such as CASTEP [13, 14] and ReaxFF [15] implemented in LAMMPS [16].
Synthesis Planner	This component is responsible for designing the optimal assembly sequence of a given molecular structure. It does this by determining the molecular fragments that require the least energy to join. It then determines suitable precursors that can be sourced and introduced into the process. The algorithm employs an approach that extends retrosynthesis but, in contrast, it can handle organic as well as inorganic structures and covalent bonds as well as noncovalent interactions. All calculations for the target structure are performed in the gas phase without the influence of other components such as substrates.
Transition Planner	This component calculates trajectories of the scanning probe at each step of the assembly process. The calculations performed at the atomic scale use a reactive force field to determine bond creation and removal under the influence of all components involved, including substrates and the scanning probe. In addition it determines the molecular conformations at which bond manipulation will likely take place. The calculations then inform higher level continuum models that describe how the scanning probe will be affected at each manipulation and assembly step. By doing so, set points for the instrument controller are determined.

determines which parts of the target structure must remain intact. For example, it may be desirable to consider benzene rings or C dimers as synthons and therefore not analyse them into finer substructures. There is indeed a good reason for doing so because the synthesis of benzene, for instance, has a high energy requirement that may be difficult to provide by means of mechanical force with a scanning probe.

The second data structure is the Synthesis Pathway Planner (Figure 2). It is a hierarchical tree-type structure specifically designed to describe reaction pathways, comprising multiple levels of parent–children nodes. Each node in this graph is potentially a synthon, while one level in the hierarchy represents a reaction step. A collection of child synthons represents the reactants with the parent synthon being the product. The data structure also supports by-products or side reactions as well as any number of alternatives.

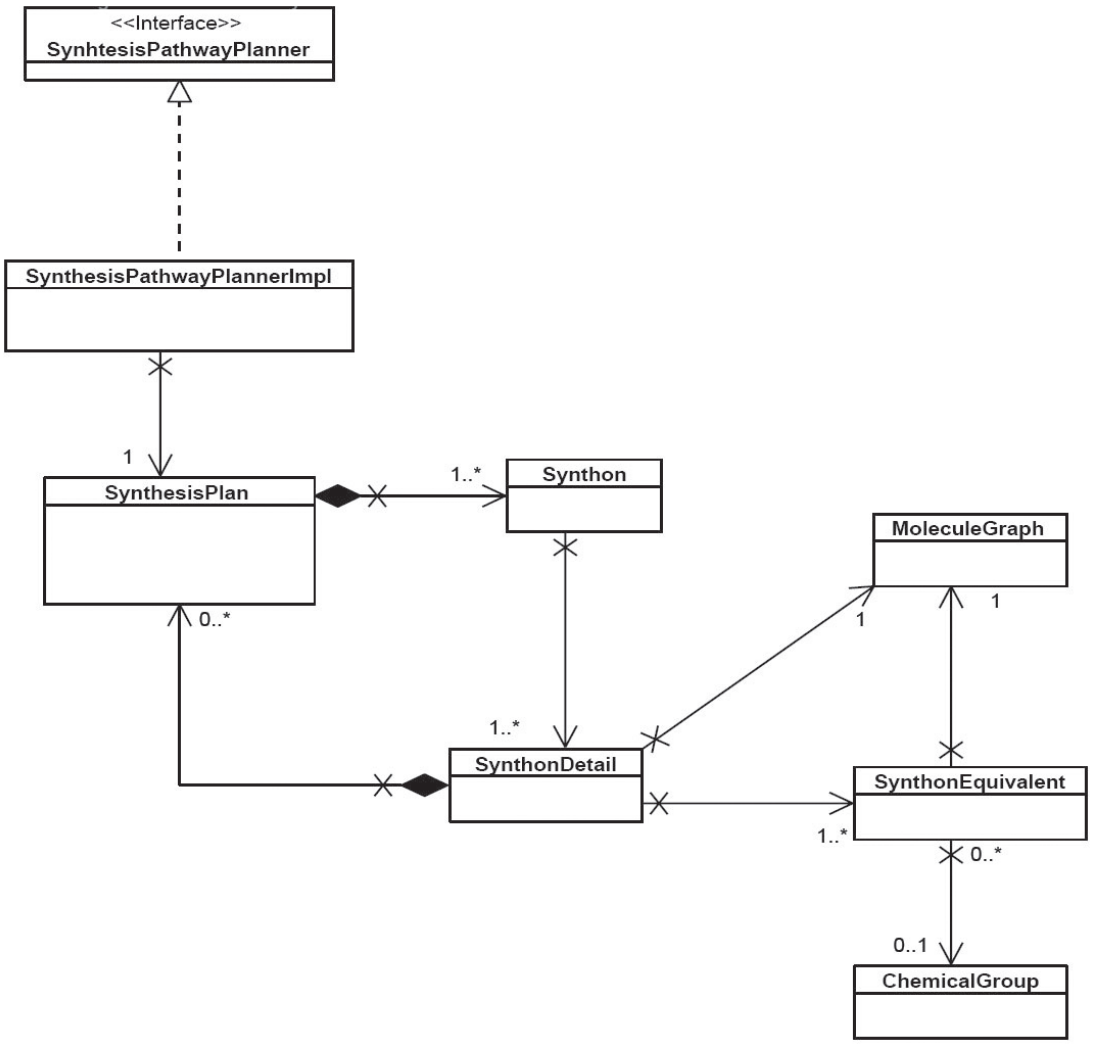

Figure 2. Synthesis Pathway Planner class diagram.

This data structure is also at the core of the Synthesis Planner, which uses it to search for and evaluate alternative options systematically by manipulating the nodes that comprise this data structure. The Synthesis Planning algorithm iteratively breaks the molecule down into synthons, subject to constraints described above, and within every iteration the algorithm creates instances of the Synthon class, shown in Figure 2, with associated details included in the SynthonDetail class. Such details are, for instance, the commercial name, sourcing information and molecular structure of the associated synthetic equivalent in case it can be sourced commercially (or the pathway by which it can be obtained). Different instances of the SynthonEquivalent class refer to alternative or substitute molecules.

Queries to the Matter Compiler knowledge base or external database are constructed with the information contained in those objects. The class also contains search information suitable for database queries within the Matter Compiler knowledge base or external databases. There are two approaches to querying molecular databases and both are considered here. One is graph similarity matching [17], the other is molecular fingerprinting [18, 19]. The latter is a faster but potentially less accurate approach (although already widely used in searching chemical databases).

After a synthesis tree has been constructed, a parent synthon or synthetic equivalent together with its children synthons or synthetic equivalents are screened for structural feasibility. This is achieved by means of a DFT calculation. Reaction trajectories are also calculated via, for example, DFT simulations. The Matter Compiler framework assumes the availability of external DFT codes such as CASTEP [13, 14], or AIMPRO [20] codes. However, the Matter Compiler coordinates simulations by initializing the external DFT code and launching the simulation. Simulations are quasistatic, meaning that the results obtained at the nth step (i.e., the optimized system) are then used as an input (or "seed" structure) in the $(n + 1)$th simulation step. The Matter Compiler thus appropriately initializes the external simulator for that next simulation step.

This process is represented and managed by the third data structure (Figure 3), called "TransitionPlan". This is also a hierarchical data structure but at each level peer nodes are connected sequentially. This structure reflects the approach described above linking the results of the n simulations in the correct sequence. Higher levels in the hierarchy provide an abstraction of levels below, aggregating length, timescale and information. For instance, the trajectory of the scanning probe, from the perspective of the SPM controller, is a "macroscopic" problem that is solved using continuum mechanics, while the interaction between the tip and the substrate or the manipulation of particular particles on the surface is a problem in the "microscopic" (or ultramicroscopic) domain best addressed by quantum mechanical calculations. Both problems, though, are linked as the microscopic domain directly affects the macroscopic domain. Therefore, this data structure has been designed to coordinate microscopic simulations (executed, for example, by quantum mechanical codes) that inform macroscopic calculations in the continuum domain and at the same time describe the trajectory of the scanning probe by means of the set of coordinates it should follow during execution of a synthesis step in the context of a particular experiment. Furthermore, each node (or assembly step) in a trajectory within the transition plan is a collection of all items in the unit cell participating in the particular manipulation or synthesis experiment such as substrate, synthon and any number of other molecular structures that are used to facilitate synthesis.

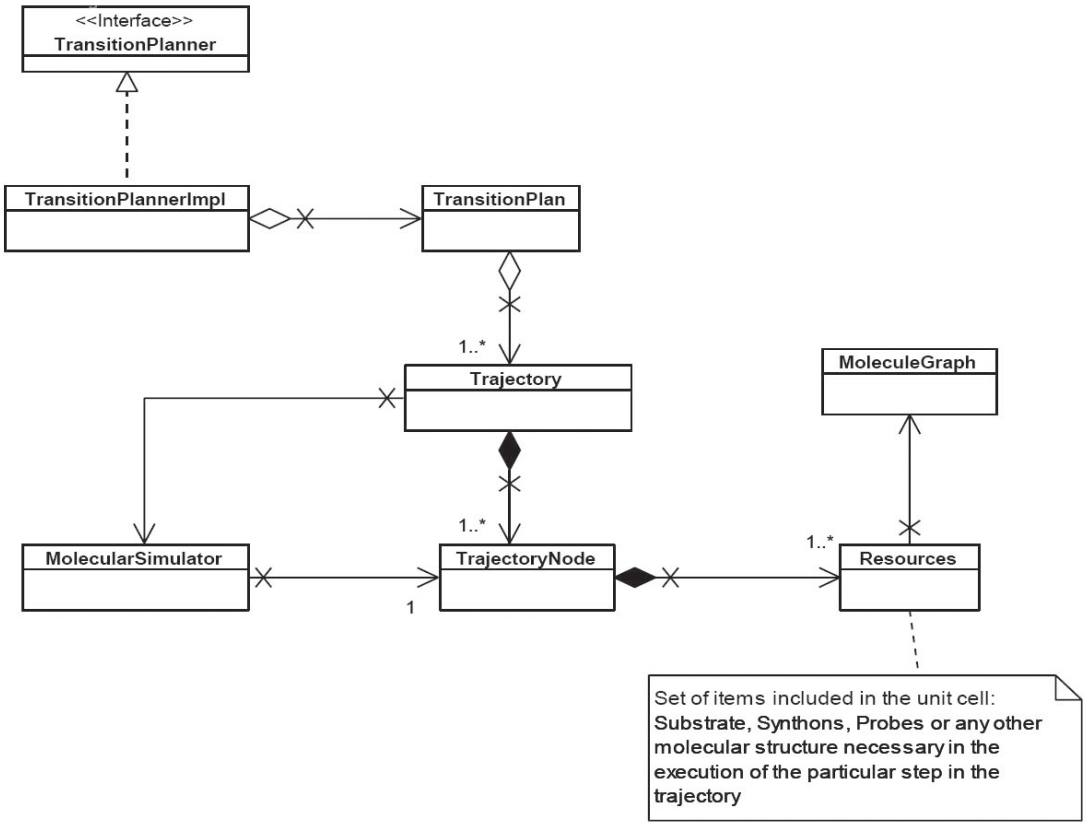

Figure 3. Transition Planner class diagram.

It is noteworthy that the framework is independent of particular experimental equipment (such as an SPM). Hence, although synthesis using SPM is assumed for now, the framework is equally applicable to future approaches and equipment not currently available.

4. Case study: Adsorption of a single synthon of the (4-mercaptophenyl)-phenylacetylene molecule on a Si(100) substrate

4.1. Analysis and optimization of the substrate

Extensive studies [21–25] provide an understanding of the behaviour of a Si(100) surface in the presence of a scanning probe. Building on such understanding, one can then design manipulation or synthesis experiments on this substrate. Si(100) is popular in industrial and academic research due to its stability [26, 27]. The behaviour of Si(100) 2×1 is shown in Figure 4 for lateral tip displacements. Reconstructions of Si(001) when displacing the scanning probe vertically above particular positions are shown in Figures 5(a) and 5(b) [23, 24].

The system in the case study comprised a Si(001)-p(2×1) substrate and a 4-atom Si tip as described in refs 22 and 24. The aim was to adsorb at least one synthon of the (4-mercaptophenyl)-phenylacetylene (MPPA) molecule [28]. The reasons for selecting this molecule were its simple

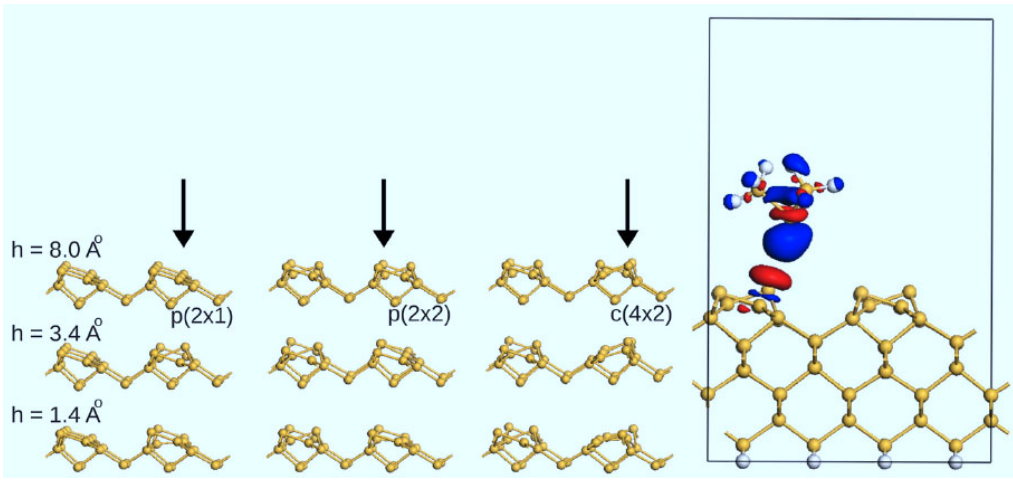

Figure 4. Manipulation of Si(100) 2 × 1 surface as a scanning probe approaches the surface at the positions indicated by the arrow. (a) Reconstructions of Si(001) under the influence of the scanning probe at the distances indicated between the probe's apex atom and the surface's adatoms; (b) Charge differences between the tip and Si(001) surface. Figure 4(a) reproduced from ref. 22 with permission and Figure 4(b) reproduced from ref. 23 with permission.

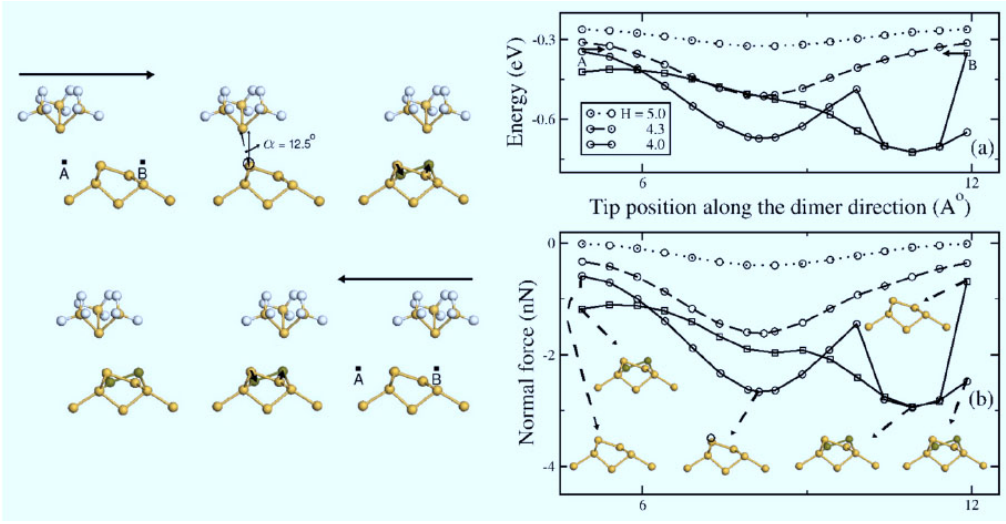

Figure 5. (a) Total energy as a function of tip position over a scanned direction for different tip–surface distances. For a tip–surface distance (h) of about 4.3 Å or more, the behaviour of the energy curve is almost the same for the tip scans starting from A → B or B → A, so we just show the case with the tip scans from A → B. For tip–surface distance h = 4 Å we show energy curves for both directions of the scan, starting from A → B (○), and from B → A (□). (b) The same as (a) but this is for the normal force. Different dimer orientations when the tip scans from A → B and B → A are also shown. The new dimer after the flipping has occurred is shown in a darker colour; a small dashed circle represents the new position of the adatom when the tip scans above it. Hydrogen atoms are represented by the white spheres (reproduced from ref. 22 with permission of the Institute of Physics).

structure, containing two C-rings, one triple C bond and one S atom, and its conducting properties. A chain of such molecules has been used in ref. 28 to build a molecular wire. Beilstein [10], PubChem [11] and ChemSpider [12] have been explored to elucidate conventional synthesis pathways and synthons that can provide initial guesses for assembly pathway planning. The last two chemical data bases provide an XML-based computer interface to execute automated searches but contain no reaction data. Beilstein does contain reaction data but does not offer such an interface so searches can either be manual or using software made available from partner companies to the publishers (such as Elsevier in the case of Beilstein or the Royal Society of Chemistry for ChemSpider). For our prototype case study system, Beilstein was selected as access to reaction data was necessary.

4.2. Synthesis pathway planning for the MPPA molecule

Using the Beilstein database and interface software from Molecular Networks [10], the MPPA molecule was analysed into synthons in the gas phase. In addition, a number of alternative reaction paths for conventional synthesis in solution were also extracted from the database. The hits from the Beilstein database provided clues (or starting points) on how synthesis planning could take place by the Matter Compiler, analysing the structure into synthons that would eventually result in minimum energy trajectories in a "pick and place" scheme, molecule-by-molecule. The selected retrosynthetic tree for MPPA is shown in Figure 6. For step 1, eleven hits were found in Beilstein, representing alternative published reactions. An equal number of alternative reactions were found for step 2. One could analyse further synthons (e.g., those containing rings) but the synthesis tree shown in Figure 6 is the simplest to simulate in DFT and, potentially, a viable one to execute with an SPM.

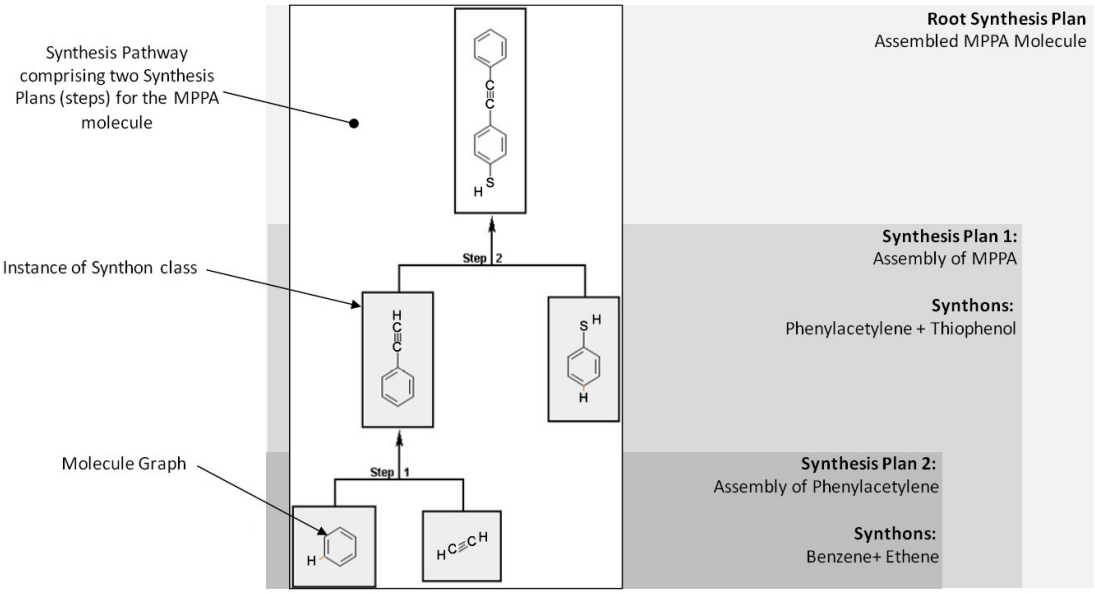

Figure 6. Retrosynthetic tree of MPPA in the gas phase and its mapping onto the Synthesis Pathway Planner construct.

4.3. Design of tooling and adsorption calculations

The prototype synthesis selected involved the adsorption of a single carbon ring on the Si(001)-p(2 × 1) (see Figure 7). The tip was functionalized with a tool comprising a diadamantane molecule. First-principles calculations were carried out to determine whether and how such a tool can hold, with its two bridging central atoms, a C-dimer or small organic molecules such as a benzene ring. This tool acts like a donor source that can be positioned in specified locations on the surface and deposit the molecule with which it is charged. Extensive theoretical analysis by means of first-principles calculations done in the past has shown that the two bridging central atoms can be replaced by two other atoms of a different element in the same group (IV—C, Si, Ge, Sn, Pb) [28]). The choice of atom determines whether the molecule can be deposited on the surface. Our calculations with several alternative configurations involving different bridging atoms revealed that Si and Ge are the best options, due to a weaker bond with benzene allowing deposition to take place (Figure 7).

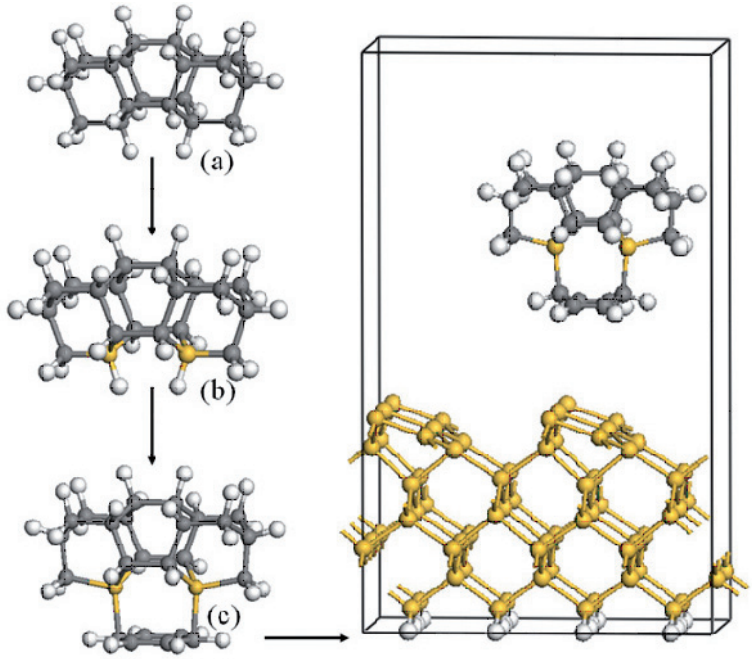

Figure 7. Diadamantane tool design for manipulation of benzene, the basic synthon of MPPA, and its adsorption onto Si(100). The white spheres are the H atoms, the dark grey spheres are the C atoms, while the light shaded (yellow online) spheres are the Si atoms.

5. The Matter Compiler interface

Having demonstrated a single prototype synthesis with our Matter Compiler framework, our future objective is to create a complete system. We anticipate this will require image processing for monitoring progress, visualization and haptic interaction to provide insight, and an intervention strategy in case of errors.

5.1. Visualization and haptic interaction

Within SPM there is already a body of research exploring the potential of visualization and haptic investigation. For example, Rubio-Sierra et al. [29] have used a force–feedback joystick (FFJ) to interface to an atomic force microscope (AFM); Fok et al. [30, 31], in their work on nanoindentation, also use a FFJ to help position the AFM tip, but furthermore introduce simulation to estimate the indentation of the sample according to the applied force; Varol et al. [32] have used a Phantom Omni® [33] to interface to a virtual environment (VE) for planning and testing manipulation strategies; Gao and Lécuyer [34] also report a VE but for the manipulation of nanotubes; and Iwata et al. [35] report controlling an AFM directly, also with an Omni®.

The current state-of-the-art for haptic interaction with SPM thus comprises a number of reports detailing contact with objects or indentation of the substrate; however, the scanning probe trajectories of the Matter Compiler's "Transition Planner" component (see Section 3) necessarily include the volume *above* the substrate and sample, too. Indeed, for pick-and-place interactions understanding tip behaviour in noncontact mode is arguably more important than when near to the surface. Whereas haptic technology applied in SPM has thus far concentrated largely on rendering objects, to create visual representations of the complex data variables involved in noncontact mode, the Matter Compiler will be more concerned with rendering fields.

Although technically these two goals appear very different, in fact they each rely on a key characteristic of any haptic device, namely the specification of its input and output degrees of freedom (DoF). A device with a 3DoF input accepts the x, y, z coordinates of the stylus in whatever position the user moves it to, and outputs 3DoF of force; that is, a force vector comprising x, y and z components. When applied to "feel" an object, the event-processing loop continuously monitors the modelled object boundary for penetration by the stylus and outputs a restoring force, often based on a simple spring model. Different properties can be simulated: a step function (i.e., a strong repulsion acting at very small penetrations) gives the impression of a hard object; whereas returning forces that are proportional to the penetration depth simulates something that appears to give way (moelleux).

During contact mode AFM the topography signal is usually used to model the substrate and sample, whilst forces associated with indentation and manipulation are commonly either approximated using the topography signal and a spring model, or obtained directly from the cantilever deflexion. If the haptic rendering is accompanied by visual rendering, there can be a compelling impression that the object is present in its entirety, even though haptically it is investigated only as a sequence of points presented one at a time.

Contrastingly, in noncontact mode the concern is to visualize variables occupying the volume above the sample. These involve a mixture of raw experimental data, derived quantities and simulated data. A number of reports are available that visualize such variables by traditional means. For example, Schirmeisen et al. [36] show constant-y colour plots of frequency shift, energy dissipation and potential energy, and line graphs of force in the z (i.e., vertical) direction. Ternes et al. [37], in their investigations of the interplay of force components, draw constant-z colour plots and line graphs of potential energy, vertical stiffness, conductance, and vertical and lateral force magnitudes. Ruschmeier et al. [38] also introduce arrow plots to represent lateral forces on the tip, which change with z position. If we also wish to compare with first-principles simulations such as those arising from the "Molecular Simulator" component (Section 3), then

the challenge to understand them all is greater still. The difficulty, well known in volume visualization, is how to represent several variables throughout a space, possibly in combination but without occlusion. Isosurfaces (3D contours), volume rendering (semitransparent clouds) and streamlines all help but, eventually and inevitably, the trees obscure the forest.

Haptic devices provide a way round this problem but their potential is often overlooked due to their more usual application to "feeling" virtual objects. Recall that a haptic stylus is specified in terms of its output DoFs (force) and input DoFs (position) and, moreover, that rendering is done point-by-point. Scrapping vector arrows and streamlines in favour of feeling forces directly is indeed one way to reduce the visual load, but just as important is the ability to visualize *selectively,* according to where the user has placed the stylus at any moment. We can envisage a scenario where the representation of the substrate appears permanently in the display but an isosurface of some interesting scalar above it is drawn only in the immediate region of the stylus. This leaves the space between the (partial) isosurface and the user clear to see what is going on. Of course, the computational challenge of doing this is very much greater than that presented by traditional visualization since the implementation has to keep up with the user's movements in real time. However, modern texture-rendering methods are a match for the task and the approach has already been used with some success to visualize volume vector fields interactively [39]. Figure 8 shows how the display might look.

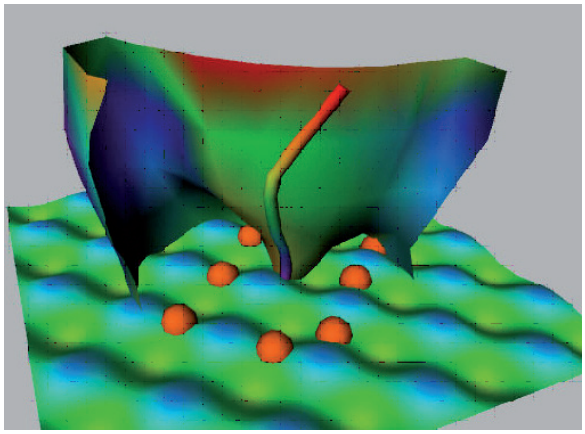

Figure 8. An isosurface of potential energy (PE) is coloured by energy dissipation at the corresponding location in the volume above a substrate and adsorbate. The haptic stylus is moved along a trajectory of constant PE (tubular structure coloured by frequency shift) and renders the force felt by the tip. Portions of the PE isosurface in front of the stylus are culled to improve visibility.

5.2. Scanning probe microscopy image simulation and visualization

We turn finally to the question of the simulation of scanning tunneling microscopy (STM) images, where the imaging is effectively that of the wave functions at the Fermi level. Such images can be simulated and interpreted using first-principles methods such as DFT, as implemented in the AIMPRO code. This is illustrated here in the case of a highly compact and flat substrate, namely a graphite surface. There are many recorded [40, 41] and some simulated [42] images available for this surface. The conjugated system offers opportunities for π–π stacking interactions with polycyclic aromatic hydrocarbons. In addition, it is relatively easy to introduce point defects with, for example, electron beam damage at energies of about 80 keV, which could provide convenient anchoring sites for covalently bonding synthons prior to reaction.

There is a variety of such defects based on the vacancy, or the bond rotation known as the Stone–Wales defect.

The method chosen for simulating STM images of pristine and defective graphitic surfaces is the following. A supercell was constructed by repeating the standard 4-atom cell 8 times along the basal lattice vectors. The c-axis lattice vector was increased to 2.5 nm in order to create a system of surfaces with sufficient vacuum between them to preclude interaction. The resulting 256 atom bilayer graphene supercell should broadly represent the features of a graphite surface, and also be of interest in its own right. Defects were introduced into this cell (Figure 9). The vacancy-related defects included: the mono-vacancy, whose ideal D_{3h} structure undergoes a Jahn–Teller distortion, forming a 9–5 ring structure with the third atom protruding from the surface by differing amounts depending on the level of theory [43, 44], and the first and third neighbour di-vacancies, forming 5–8–5 and 5–9–9–5 reconstructions, respectively. The central bond of the bridging carbon chain in the third neighbour defect has a character approaching that of a triple bond and stands proud of the surface. This rearrangement is the Stone–Wales defect, otherwise known as 5–7–7–5 bond rotation. The supercells for these were otherwise identical to the pristine supercell. DFT runs were conducted using the AIMPRO code [20] with the LDA-PW92 functional [45] and a basis set of four radial Gaussians per atom. Each of these radial functions is multiplied by a series of angular functions from s- to p- or d-symmetries (in order of increasing exponent, the upper limits of this series for the four exponents were pdpp). The Brillouin zone sampling was of the $2 \times 2 \times 1$ Monkhorst–Pack [46] type.

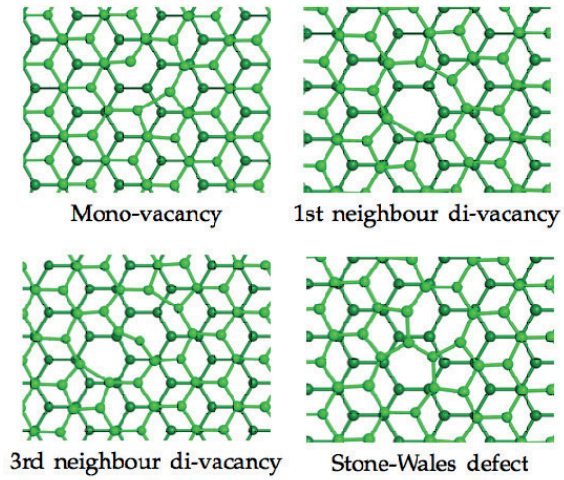

Figure 9. Graphitic surface defects.

The scanning tunnelling micrographs were then produced using the AIMview code [47], invoking the Tersoff–Hamann approximation [48], in which the tunnelling current at a given tip height is assumed to depend on the squares of the wave functions at that height for substrate energy levels or bands at or near the Fermi level. Biasing the sample with a positive potential in relation to the tip effectively lowers the Fermi level of the sample under examination. Constant height scanning tunnelling images arise from the parts of the Brillouin zone where an electron band intersects the lowered Fermi level. Plotting a cross-section of the corresponding squares of the appropriate wave functions, as shown in Figure 10(d), then yields the simulated image.

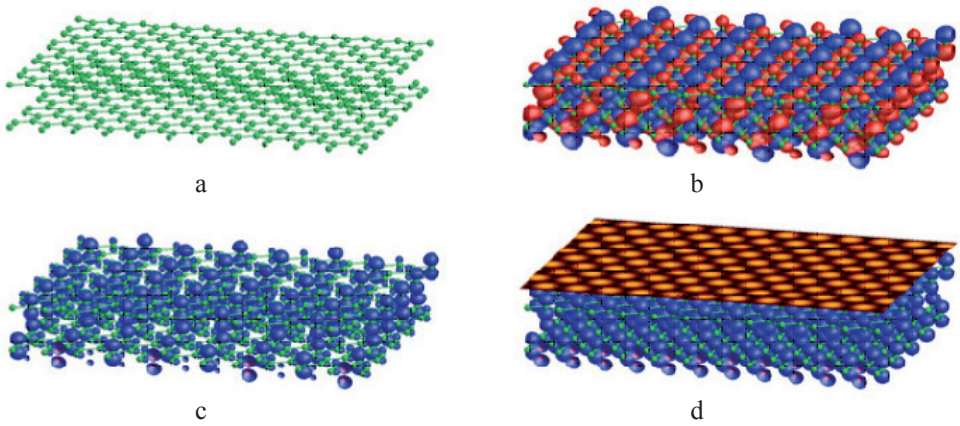

Figure 10. The general method of producing a simulated scanning tunnelling micrograph. (a) Select the area of interest. (b) Select the required wavefunction. (c) Plot the square of the modulus of the wavefunction. (d) Produce a colour-mapped cross-section at the required height.

Simulated micrographs produced by this method compare favourably with experimental images and other simulations [49]. Figure 11 shows simulated STM images of the pristine surface with tip heights of 0.80 Å and 1.60 Å, together with an experimental image showing the effect of a tip retraction of 1.00 Å during the scan. The bottom of this image corresponds well to Figure 11(a) and the top to Figure 11(b).

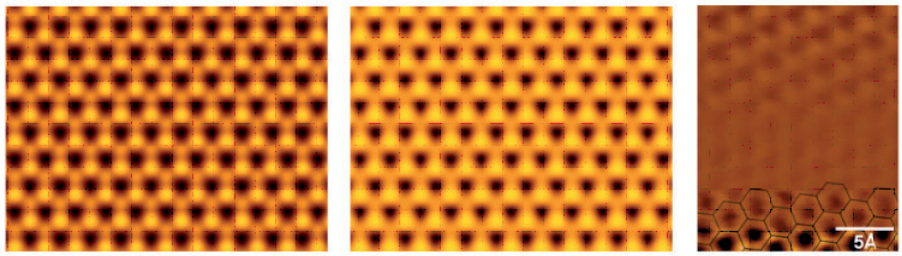

Figure 11. (a) and (b): simulated STM images of the pristine surface with a tip height of 0.80 Å and 1.60 Å, respectively. (c): is an experimental image showing the effect of a tip retraction of 1.00 Å during the scan. The lower portion of this image corresponds well to (a) and the upper to (b). Reproduced from ref. 49 (Figure 11c) with permission from the American Physical Society.

Similar simulations of the defective surfaces were produced and these are shown in Figure 12. All the simulations show increased electron density in the vicinity of the defects, which may prove useful as anchorage points for synthons during machine-phase chemistry. The 1st neighbour di-vacancy simulation appears to show evidence of a highly directional disruption to the π system perpendicular to the axis of the 5–8–5 reconstruction. In the simulation of the 3rd neighbour di-vacancy, the protruding carbon pair, being much closer to the sampling plane and having enhanced density from near-triple bond character, significantly dominates the simulation. The rotated bond in the Stone–Wales defect is clearly visible and the electron density is concentrated around the periphery of the 5–7–7–5 structure.

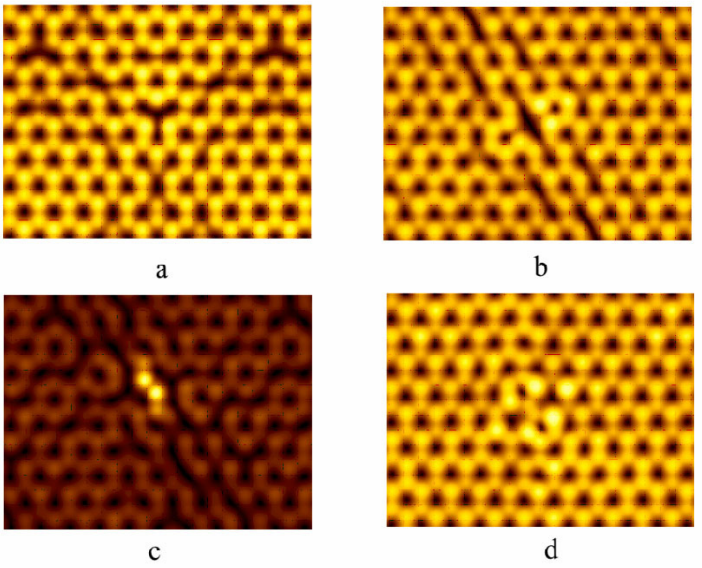

Figure 12. Simulated STM images of the defective graphitic surface. All simulations are at a tip height of 0.80 Å. (a) monovacancy; (b) 1st neighbour di-vacancy; (c) 3rd di-vacancy; (d) Stone–Wales defect.

6. Summary and conclusions

An application framework that fully describes how automated deterministic assembly of target arbitrary nanostructures can take place has been presented. This is a novel computer framework that has the potential to automate molecular assembly using a scanning probe microscope. Automation is fundamental to any future commercial realization of molecular assembly and this project, although a pilot, is a first important step towards that. Furthermore, the case study simulations contribute to the understanding of interactions that take place in an SPM for silicon substrates and allowed us to explore a potential pathway for the deposition of a benzene ring onto such a substrate. The framework has, however, been designed in anticipation of the development of new experimental protocols as well as equipment that can be deployed for molecular assembly, accommodating such developments with its fundamental data structures. Advanced visualization techniques to meet the complex needs of the Matter Compiler have been proposed; they will enable checking and error correction of the generated SPM trajectories.

There is much scope to take this research further. We are in particular now pursuing the use of reactive force fields for fast screening of potential synthons and the determination of alternative synthetic pathways; the integration of "virtual" AFM [14] codes and the development of "virtual scanning tunnelling microscopy (STM)" codes to test *in silico* synthesis before execution in a real experimental setting; the porting of algorithms to graphical processing units (GPUs) to achieve at least an order of magnitude speed up-in the calculations; experimental validation of the Matter Compiler; and development of software suitable for deployment on the internet "cloud", accessible for use by anyone in the academic community.

Acknowledgments

This work has been supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under contracts EP/F009801/1, EP/F009917/1, EP/F009933/1, EP/F010036/1 and EP/F010265/1. The authors would like to thank the EPSRC for the financial support. The

authors would also like to thank the School of Engineering and Design at Brunel University London for purchasing a computing cluster to support research in this area. Simulations were carried out on this cluster (comprising 64 processors).

References

1. Eigler, D.M.; Schweizer, E.K. Positioning single atoms with a scanning tunnelling microscope. *Nature* (Lond.) **344**, 524 (1990).
2. Bartels, L.; Meyer, G.; Rieder, K.-H. Basic steps of lateral manipulation of single ftoms and diatomic clusters with a scanning tunneling microscope tip. *Phys. Rev. Lett.* **79**, 697 (1997).
3. Lee, H.; Ho, W. Single-bond formation and characterization with a scanning tunneling microscope. *Science* **286**, 1719 (1999).
4. Jung, T. et al. Controlled room-temperature positioning of individual molecules: Molecular flexure and motion. *Science* **271**, 181 (1996).
5. Moriarty, P. et al. Translation, rotation and removal of C60 on Si (100)-2x1 using anisotropic molecular manipulation. *Surface Science* **407**, 27 (1998).
6. Oyabu, N. et al. Mechanical vertical manipulation of selected single atoms by soft nanoindentation using near contact atomic force microscopy. *Phys. Rev. Lett.* **90**, 176102 (2003).
7. Hla, S.W. et al. Inducing all steps of a chemical reaction with the scanning tunneling microscope tip: towards single molecule engineering. *Phys. Rev. Lett.* **85**, 2777 (2000).
8. Corey, E.J.; Cheng, X.-M. *The Logic of Chemical Synthesis*. New York: Wiley (1995).
9. Corey, E.J. The logic of chemical synthesis: multistep synthesis of complex carbogenic molecules (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **30**, 455 (1991).
10. Crossfire *Beilstein*, http://xfire.mimas.ac.uk/crossfire/
11. *Pubchem* database, http://pubchem.ncbi.nlm.nih.gov/
12. *Chemspider*, http://www.chemspider.com/
13. Payne, M.C. et al. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **64**, 1045 (1992).
14. Clark, S.J.; Segall, M.D.; Pickard, C.J.; Hasnip, P.J.; Probert, M.J.; Refson, K.; Payne, M.C. First principles methods using CASTEP. *Z. Kristallographie* **220**, 567 (2005).
15. van Duin, A.C.T.; Dasgupta, S.; Lorant, F.; Goddard, W.A. ReaxFF: a reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396 (2001).
16. Plimpton, S.J. Fast parallel algorithms for short–range molecular dynamics. *J. Comp. Phys.* **117**, 1 (1995).
17. Ullmann, J.R. An algorithm for subgraph isomorphism. *JACM* **23**, 31 (1976).
18. Gasteiger, J.; Engel, T. (eds), *Chemoinformatics*. Wiley–VCH (2003).
19. Willett, P.; Barnard, J.M.; Downs, G.M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **38**, 983 (1998).
20. Briddon, P.R.; Jones, R. LDA calculations using a basis of Gaussian orbitals. *Phys. Stat. Solidi B* **217**, 131 (2000).
21. Ly, D.Q.; Makatsoris, C. First principles studies of the phase transition from Si(100)-p(2x1) to p(2x2). Submitted to *Phys. Rev. B* (2012).
22. Ly, D.Q.; Paramonov, L.; Makatsoris, C., First principles studies of an Si tip on an Si(100)2 × 1 reconstructed surface. *J. Phys. Condens. Matter* **21**, 185006 (2009).
23. Ly, D.Q.; Makatsoris, C. *Ab-initio* studies of phase manipulation from Si(001)-p(2x1) to Si(001)-p(2x2) using a silicon AFM tip (abstract and poster). UK Scanning Probe Microscopy Meeting, National Physical Laboratory (2009).
24. Molecular Networks GmbH, http://www.molecular-networks.com/
25. Object Management Group, http://www.uml.org/

26. Dabrowski, J.; Mussig, H.J. *Silicon Surfaces and Formation of Interfaces*. Singapore: World Scientific (2000).
27. Keeling, D.L; Humphry, M.J.; Moriarty, P.; Beton, P.H. Attractive mode manipulation of covalently bound molecules. *Chem. Phys. Lett.* **366**, 300 (2002).
28. Freitas, R.A.; Merkle, R.C. A minimal toolset for positional diamond mechanosynthesis. *J. Computational Theor. Nanosci.* **5**, 760 (2008).
29. Rubio-Sierra, F.J.; Stark, R.W.; Thalhammer, S.; Heckl, W.M. Force-feedback joystick as a low-cost haptic interface for an atomic-force-microscopy nanomanipulator. *Appl. Phys. A* **76**, 903 (2003).
30. Fok, L.M.; Liu, Y.H.; Li, W.J. Modelling of haptic sensing of nanolithography with an atomic force microscope. Proc. IEEE International Conference on Robotics and Automation (ICRA), pp. 2446–2451 (April 2005).
31. Fok, L.M.; Liu, Y.H.; Li, W.J. Modelling of nanomanipulation with an integrated teleoperated system. Proc. IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 83–88 (July 2005).
32. Varol, A.; Gunev, I.; Basdogan, C. A virtual reality toolkit for path planning and manipulation at nano-scale. Proc. IEEE Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS'06), pp. 485–489 (March 2006).
33. SensAble Technologies, Inc., Phantom Omni® haptic device, http://www.sensable.com/haptic-phantom-omni.htm
34. Gao, Z.; Lécuyer, A. A VR simulator for training and prototyping of telemanipulation of nanotubes. Proc. ACM Symposium on Virtual Reality Software and Technology (VRST 08), pp. 101–104 (October 2008).
35. Iwata, F.; Ohara, K.; Ishizu, Y.; Sasaki, A.; Aoyama, H.; Ushiki, T. Nanometer-scale manipulation and ultrasonic cutting using an atomic force microscope controlled by a haptic device as a human interface. *Jap. J. Appl. Phys.* **47**, 6181 (2008).
36. Schirmeisen, A.; Weiner, D.; Fuchs, H. Single-atom contact mechanics: From atomic scale energy barrier to mechanical relaxation hysteresis. *Phys. Rev. Lett.* **97**, 136101 (2006).
37. Ternes, M.; Lutz, C.P.; Hirjibehedin, C.F.; Giessibl, F.J.; Heinrich, A.J. The force needed to move an atom on a surface. *Science* **319**, 1066 (2008).
38. Ruschmeier, K.; Schirmeisen, A.; Hoffmann, R. Site-specific force-vector field studies of KBr(001) by atomic force microscopy. *Nanotechnology* **20**, 264013 (2009).
39. Emont, G. *Haptic Enhancement to Flow Visualization*. MSc thesis, University of Hull (2006).
40. Kobayashi, Y.; Fukui, K.; Enoki, T.; Kusakabe, K. Observation of zigzag and armchair edges of graphite using scanning tunneling microscopy and spectroscopy. *Phys. Rev. B* **71**, 193406 (2005).
41. Wang, Y.; Ye, Y.; Wu, K. Simultaneous observation of the triangular and honeycomb structures on highly oriented pyrolytic graphite at room temperature: An STM study. *Sur. Sci.* **600**, 729 (2006).
42. Latil, S.; Meunier, V.; Henrard, L. Massless fermions in multilayer graphitic systems with misoriented layers: *Ab initio* calculations and experimental fingerprints. *Phys. Rev. B* **76**, 201402 (2007).
43. El-Barbary, A.; Telling, R.; Ewels, C.; Heggie, M. Structure and energetics of the vacancy in graphite. *Phys. Rev. B* **68**, 144107 (2003).
44. Heggie, M.I.; Latham, C.D. The Atomic-, Nano-, and Mesoscale Origins of Graphite's Response to Energetic Particles. In: *Computational Nanoscience* (E. Bichoutskaia, ed.). London: RSC Publishing (2011).
45. Perdew, J.P.; Wang, Y. Accurate and simple analytic representation of the electron–gas correlation energy. *Phys. Rev. B* **45**, 13244 (1992).
46. Monkhorst, H.J.; Pack, J.D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188 (1976).
47. Hagon, J. *AIMview Molecular Visualisation Package (ver. 1.1)*. Newcastle University (2011).
48. Tersoff, J.; Hamann, D.R. Theory of the scanning tunneling microscope. *Phys. Rev. B* **31**, 805 (1985).
49. Cisternas, E.; Stavale, F.; Flores, M.; Achete, C.A.; Vargas, P. First-principles calculation and scanning tunneling microscopy study of highly oriented pyrolytic graphite (0001). *Phys. Rev. B* **79**, 205431 (2009).