

# Polygranular image guided atomistic reconstruction: a parametric model of pyrocarbon nanostructure

Franck Polewczyk<sup>a,b,c</sup>, Paul Lafourcade<sup>a,b</sup>, Jean-Pierre Da Costa<sup>d,e</sup>, Gérard Vignoles<sup>f</sup>, Jean-Marc Leyssale<sup>c,\*</sup>

<sup>a</sup>CEA DAM DIF, 91297 Arpajon, France

<sup>b</sup>University of Paris-Saclay, LMCE, 91680 Bruyères-le-Châtel, France

<sup>c</sup>University of Bordeaux, CNRS, Bordeaux INP, ISM, UMR 5255, F-33400 Talence, France

<sup>d</sup>University of Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400 Talence, France

<sup>e</sup>Bordeaux Sciences Agro, F-33175 Gradignan, France

<sup>f</sup>University of Bordeaux, CNRS, CEA, Safran: LCTS (UMR5801), Pessac F-33600, France

---

## Abstract

Atomistic modelling of disordered yet textured carbons is notoriously difficult even though it can prove extremely helpful in rationalizing structure-property relationships for this class of materials. In this work we introduce a polygranular image-guided atomistic reconstruction method, which allows building models with fine-tuned values of the in-plane ( $L_a$ ) and out-of-plane ( $L_c$ ) coherence lengths, and of the orientation angle (OA). Applying a parametric study of grain size and orientation distribution, a database of 210 models is presented with parameters spanning domains characteristic of high and medium textured pyrolytic carbons: 1.5-8 nm, 2-5.5 nm and 25-110°, for  $L_c$ ,  $L_a$  and OA, respectively. A machine learning model based on a random forest regression shows that these three measurable properties can be accurately predicted from a limited set of microscopic information character-

---

\*Corresponding author

Email address: jean-marc.leyssale@u-bordeaux.fr (Jean-Marc Leyssale)

izing the distribution of local atomic environments in the models. Finally, the computed diffraction properties and high-resolution transmission electron microscopy images of a series of six models, that best match the properties of a set of well-characterized pyrocarbons, are extensively compared to experimental data, showing excellent agreement and drastically improving over former modelling studies on high textured pyrocarbons, in addition to providing the first atomistic model of a medium textured pyrocarbon.

*Keywords:* pyrolytic carbon, structure, texture, modelling, machine learning

---

## **1. Introduction**

Low temperature pyrolytic carbons, or pyrocarbons (pyCs), are dense carbon coatings deposited by chemical vapor deposition (CVD) on a surface, or infiltration (CVI) on the pore surface of a porous material [1, 2]. Conversely to the formation of soot and blacks, the lower temperatures and pressures considered during pyC preparation ( $T \sim 1000^\circ\text{C}$  and  $P \sim 1\text{-}10\text{ kPa}$ ), allow avoiding gas phase carbon nucleation and ensure the deposition of a dense and relatively homogeneous turbostratic carbon [2, 3]. Owing to their excellent mechanical and thermal properties, high melting point and low porosity, the principal application of these materials is as constituents of continuous-fiber reinforced composite materials: as matrices of C/C composite materials for thermal protection systems and solid rocket motor parts in spatial applications [4, 5] or for brakes in aeronautic or terrestrial transport [6, 7], as interphases in C/SiC materials [8, 9] for the same types of applications [10] or as interphases in SiC/SiC ceramic matrix composites for GenIV nuclear reactors or nuclear fusion reactors [11–13] . PyCs are also considered as nuclear fuel

cell coatings, in the tristructural isotropic (TRISO) particle fuel cell of the high temperature reactor (HTR) technology [14], or around individual carbon nanotubes for electron emission applications or near field microscopy probes [15]. Pyrocarbons also have attractive properties when used in Li-ion battery electrodes [16] or in supercapacitors [17]. Finally, due to their biocompatibility, pyC monoliths or coatings are also commonly used in medical implants or prostheses [18, 19].

It has been long-known that pyC can present a wide variety of microstructures, depending on the gas phase precursor and deposition/infiltration conditions [20–23]. PyC microstructures were first classified according to polarized light optical microscopy (PLOM) observations of growth features [24]. A metric, the extinction angle ( $A_e$ ) between crossed polars, was used to characterize the texture anisotropy and three classes of microstructures were observed, the high anisotropy rough laminar (RL), low anisotropy smooth laminar (SL), and isotropic (ISO) pyCs [24, 25]. The dark laminar (DL), intermediate between ISO and SL, was also identified [26].

High-resolution transmission electron microscopy (HRTEM) imaging as well as selected area electron diffraction (SAED) have improved our understanding of pyC nanostructure and texture. Bourrat *et al.* have proposed a sub-micrometer metric of the texture anisotropy, the orientation angle (OA) defined from intensity profiles along the 002 arc in SAED patterns [27]. A classification of pyC textures, based on both  $A_e$  and OA parameters, was proposed by Reznik and Hüttinger, defining the bounds for isotropic (incl. ISO), low (incl. DL), medium (incl. SL) and high (incl. RL) textured pyCs

[28]. Few other pyC types were also identified, namely the granular (G) [29] and regenerative laminar (ReL) [30] textures. Interestingly, G and ReL pyCs show similar anisotropy to the SL and RL pyCs, respectively, yet with a few different features. Especially, the ReL pyC shows a much broader Raman D band, suggesting a larger amount of structural defects, than the latter, while conversely, the G pyC presents a thinner D band than the SL pyC [29].

The structure of the different pyCs was investigated using HRTEM and X-ray diffraction (XRD). As prepared, all these materials show a turbostratic structure with an interlayer distance ( $d_{002}$ ) significantly larger than in graphite, and the absence of the diffraction peaks characterizing 3D order. Diffraction patterns are generally analyzed in terms of the widths of the 002 peak and 10 band, used to define the in-plane ( $L_a$ ) and out-of-plane ( $L_c$ ) coherence lengths, which are usually found in the 1-5 nm range for as-prepared pyCs [31–34]. Analysis of lattice fringe images from HRTEM experiments confirmed the nm-scale extent of graphene domains and stacks in pyCs [21, 30], although such methods cannot be quantitative due to image artifacts like Moiré patterns or signal superimposition effects, especially for medium or low textured pyCs. This results in apparent fringe lengths ( $L_2$ ) that are generally shorter than the coherence length  $L_a$  obtained from XRD[34]. These techniques were also applied to investigate heat-treated pyC samples and have shown that only high-textured pyCs, both RL and ReL, can graphitize when taken to temperatures above 2500°C [1, 31]. Also, pair distribution functions of the as-prepared RL and SL pyCs [33], and of the ReL pyC as-prepared and heat-treated at moderate temperatures (up to 1700 K)[34], were determined

using neutron diffraction, allowing for a more detailed and resolved description of the nanostructure. It was shown that when heat-treated at 1500° C the structure of the ReL pyC is almost identical to the one of the as-prepared RL pyC [34].

Electron energy loss spectroscopy (EELS) was used to characterize carbon hybridization, showing that the materials are essentially constituted of  $sp^2$  hybridized carbons [31, 35]. More precisely, by analyzing the relative intensities of the  $\pi^*$  and  $\sigma^*$  peaks, Vallerot et al. determined a fraction of  $sp^2$  hybridized carbon atoms of about 80 % for SL, RL and ReL pyCs, the authors recognizing that the missing 20 % cannot be accounted for by  $sp^3$  hybridized carbon, but more certainly by disordered  $sp^2$  environments[31]. These environments, and their EELS signature, have been evidenced later in the case of irradiated graphite using a combination of classical molecular dynamics and density functional theory predictions of EELS features [36].

Hydrogen content was measured using elastic recoil detection analysis (ERDA) and/or secondary ion mass spectroscopy (SIMS). Earlier investigations report values increasing from about 2 at. % for SL pyC to 4 at. % for RL pyC, the ReL pyC showing an intermediate value [30, 37]. More recent investigations using the same techniques report values of 0.7 and 1.1 at. % for the RL and SL pyCs, respectively [33], and a value of 2.5 at. % for the ReL pyC, decreasing down to a non detectable value after heat treatment at 1300°C [34].

Modeling the structure of such materials from the atomic scale up to the nanotexture scale, with a typical representative length scale of 10-100

nm, is extremely challenging. Common modeling strategies, either based on introducing defects within crystalline materials [38–40], or on structure reconstruction approaches [41] or quench molecular dynamics simulations [42] commonly used for extremely disordered or amorphous materials, obviously do not apply here. About a decade ago, an original modeling approach was developed to generate realistic atomistic representations of high-textured pyCs by combining image analysis and synthesis techniques with classical molecular dynamics simulations [43, 44]. In this so-called image-guided atomistic reconstruction (IGAR) method, grey-level statistics and spatial correlations, measured on experimental HRTEM images, through a multi-resolution framework, are extended to 3D under an orthotropy condition (*i.e.* the statistical equivalence of all directions normal to the pyC deposition axis) and imposed on a 3D image block. This 3D textured image is then used as an external potential during a liquid quench MD, bringing the atoms to sit on dark areas while creating a graphite-like network of bonds.

Following preliminary reports on the methodology [43, 44], atomistic models were produced for a series of well-characterized pyCs – including as-prepared RL and ReL pyCs, as well as a series of heat-treated ReL PyC at 1300, 1500 and 1700°C [34] – and validated against experimental HRTEM (fringe and orientation statistics) and diffraction (coherence lengths, structure factors and pair distribution functions) data. It was shown that those materials count from 96 to 98 % of  $sp^2$  hybridized C atoms, the remaining being mostly  $sp^3$ , and a few  $sp$ , atoms. These atoms are arranged as stacks of faulted graphene layers, especially, the intra-sheet structure is made

of misoriented nanometric graphenic domains connected by grain boundaries made of pentagonal and heptagonal rings, in analogy with those found in CVD graphene [45, 46]. While some edges may be saturated by hydrogen atoms, the H content being very low with respect to  $L_a$  [30], the majority of graphene edges are interconnected to neighboring domains via a disordered network of screw dislocation dipoles - the so-called “car park access ramp” arrangement [44] - where most  $sp^3$  and  $sp$  hybridized carbon atoms reside. These IGAR pyC models were used to investigate elastic properties [47] and rationalize indentation measurements by identifying the mechanisms underlying out-of-plane deformations under in-plane compressive loads, and their effects on the apparent indentation modulus [48]. IGAR pyC models were also used to investigate the behavior SiC/pyC interphases under shear [49].

Despite significant success, the applicability of the IGAR method remains limited and has not allowed yet to propose a detailed structure/property relationship for pyrocarbons. First, even though IGAR models have shown to be extremely accurate in predicting pair distribution functions (*i.e.* the real space distribution of interatomic distances), these models were shown to significantly underestimate stacking order, as manifested by the underestimation of the  $L_c$  parameter obtained from the 002 peak in Fourier space [34]. We note that higher-order peaks, like 004, were barely visible on the predicted structure factors. A more fundamental limitation is that existing experimental data on pyCs are rather scarce, which clearly limits the number of models that can be reconstructed, noting that for each model, the IGAR approach requires a high quality HRTEM image and accurate estimations of the H con-

tent and density. Furthermore, by construction, only high-textured pyCs can be modeled due to the 3D image construction strategy, which basically limits its applicability to RL and ReL pyCs. Finally, establishing a well-informed structure/property relationship would require to properly discriminate the effects of the important structural and/or textural parameters like  $L_a$ ,  $L_c$  and OA on the considered properties. However, as pointed out earlier [34, 47], these properties are often correlated in actual pyCs.

In this work, we propose a variant of the IGAR method in which textured image blocks are constructed using *a priori* defined grains. This polygranular image-guided atomistic reconstruction (PG-IGAR) method is used to build a large database of atomistic models with a parametric sampling of the  $L_a$ ,  $L_c$  and OA parameters. This database is then analysed using a machine learning (ML) model to derive correlations between these parameters, measurable using diffraction data, and the distribution of local atomic environments and of ring structures present in the models. Finally, a set of models, matching best a series of experimental data on actual pyCs, is identified and their structure and texture are compared to their experimental counterparts.

## 2. Methods

### 2.1. Atomistic reconstruction

A schematic of the PG-IGAR process is presented in Fig. 1. After defining the volume of the simulation cell and the target values of the  $L_a$  and  $L_c$  parameters, a Voronoï tessellation with periodic boundary conditions (Fig. 1a) is created to accommodate  $n_G$  grains with average volume  $V_G = \pi L_a^2 L_c / 4$ ,



where, in agreement with the common definition of the structural parameters [50], we assume that  $L_a$  and  $L_c$  correspond to the diameter and height, respectively, of a cylindrical crystallite. For the sake of simplicity, prescribed values of  $L_a$  and  $L_c$  are set to be equal in this work, since grain orientations, defining the  $a$  and  $c$  directions in the grains, are defined at a later stage. Therefore, a unique parameter  $L_t = \left(\frac{4V_G}{\pi}\right)^{\frac{1}{3}}$ , where subscript  $t$  refers to the “target” value, is used to define the average grain size in textured images.

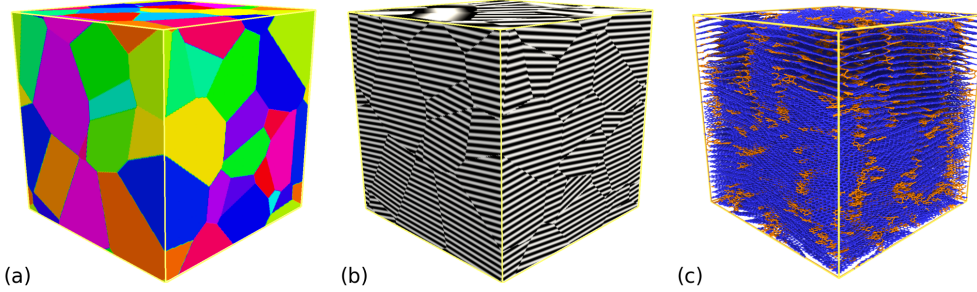


Figure 1: Principle of the PG-IGAR method. (a) 3D Voronoi tessellation with average grain size  $V_G$ . (b) Same image after decorating each grain with perfect 002 fringes and a statistical distribution of grain orientations according to  $OA_t$ . (c) 3D atomistic model after a liquid quench MD guided by the image in (b). The represented case, containing 38 grains, corresponds to a cube of 12.4 nm width with imposed  $L_t = 4$  nm and  $OA_t = 30^\circ$ . In (c) atoms in graphene-like and defect environments appear in blue and orange, respectively.

In the texture mapping step, illustrated in Fig. 1b, each grain is decorated with a 3D HRTEM-like pattern, corresponding to perfectly flat 2D fringes with a stacking periodicity given by  $d_{002}$ . Assuming that the  $z$  (vertical) axis defines the growth direction of the pyC layer, such that the  $xy$  (horizontal) plane is parallel to the fiber or substrate surface, the 002 direction in each grain can

be defined using two angles: a polar angle  $\theta$  defining the angle between the 002 direction and the z axis, and an azimuthal angle  $\varphi$ , defining the angle between the projection of the 002 direction in the xy plane and the x axis. In a purely isotropic state,  $\varphi$  would be uniformly distributed in  $[0 : 2\pi]$  and  $\cos(\theta)$  in  $[0 : 1]$ . However, in textured pyCs, a certain degree of preferential orientation is observed towards an alignment of the 002 direction parallel to the z axis. Therefore, while a uniform distribution of  $\varphi$  was adopted, the distribution of  $\cos(\theta)$  had to be biased to match prescribed  $OA_t$  values. To achieve this, for each grain, a random number  $q$  was drawn from a normal distribution centered on 0 and with standard deviation equal to  $1 - \cos(\sigma_\theta)$  where  $\sigma_\theta = OA_t/2.355$  (note that the division by 2.355 correspond to the conversion of the FWHM into standard deviation for a normal distribution). Then, the value of  $\cos(\theta)$  was obtained as  $1 - \|q\|$ . Examples of obtained distributions for various  $OA_t$  values are given in Fig. S1.

The 3D textured image is then filtered by applying a radial and directional filter in order to obtain smoother transitions at grain boundaries. Therefore, the filter performs in the Fourier domain and is defined by its band-pass transfer function:

$$H(\rho, \theta) = H_{radial}(\rho)H_{angular}(\theta) \quad (1)$$

where  $H_{radial}(\rho) = \frac{1}{\sigma_\rho\sqrt{2\pi}} \exp(-\frac{(\rho-\rho_0)^2}{2\sigma_\rho^2})$  and  $H_{angular}(\theta) = \frac{1}{\sigma_\theta\sqrt{2\pi}} \exp(-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2})$ . Parameters  $(\rho, \theta)$  are the polar frequency coordinates in the Fourier domain with  $\rho_0 = \frac{1}{d_{002}}$  corresponding to the interlayer spacing and  $\sigma_\rho = \alpha\rho_0$  is defined from  $\rho_0$  with  $\alpha \in ]0, 1[$ .  $H_{radial}(\rho)$  performs a Gaussian filtering around  $\rho = \rho_0$  reducing high and low frequencies artifacts while preserving  $d_{002}$  fringes

whatever their orientation.

The angular filter  $H_{angular}(\theta)$  aims at preserving patterns which are oriented around a reference orientation  $\theta_0$  that fits the preferential orientation in the image. The parameter  $\theta_0$  is found by looking for the largest magnitude peak in the Fourier transform modulus, yet by construction remains  $\sim 0$ .  $H_{angular}(\theta)$  also performs a Gaussian filtering centered on  $\theta_0$  with a standard deviation of  $\sigma_0$ . Typical values  $\alpha = 0.5$  and  $\sigma_0 = 50^\circ$  have proved to be a good compromise between artifact removal and structural preservation for anisotropic pyrolytic carbons. Note that angular filtering is only applied to high textured images ( $OA_t < 50$  degrees) in order to prevent the apparition of unwanted artificial structures in less anisotropic images.

In the third step, as was done in former IGAR work [34, 43, 44], carbon in the liquid state is progressively cooled down to room temperature under the influence of an image potential  $U_{Im} = \sum_{i=1}^N k_{im} I(\mathbf{r}_i)$ , where  $N$  is the number of atoms,  $I(\mathbf{r}_i)$  is the grey level of the textured image at position  $\mathbf{r}_i$  of atom  $i$  and  $k_{im}$  is the proportionality factor between grey level and energy. After the quench, the resulting atomistic model with prescribed structure and texture is relaxed at room temperature and pressure in the absence of any external potential (Fig. 1c).

All the PG-IGAR simulations reported in this work use a cubic cell of 12.4 nm width with periodic boundary conditions. 3D image templates of  $1024 \times 1024 \times 1024$  voxels are used, giving a resolution of  $\sim 0.012$  nm per voxel. Texture mapping is performed using a constant  $d_{002}$  value of 0.345 nm, a typical value for low temperature pyCs [33, 34]. The carbon atom

density, constant during the reconstructions, is set to  $d = 2.16 \text{ g/cm}^3$ , which corresponds well to experimental data for high textured pyCs [34], yet may be slightly larger than those measured for medium textured pyCs like the SL pyC[33]. This results in systems composed of 206,950 atoms for the considered volume. Finally, for the sake of simplicity, hydrogen, which can count for up to 2.5 at. % in the worst case scenario of the as prepared ReL PyC, is neglected in this work.

A total of 210 pyC models were constructed by varying two parameters in the textured image synthesis – the grain size  $L_t$  and the width of the orientation distribution  $OA_t$  – and the quenching scheme in the MD process, defined here by the total quench time  $\tau_Q$ . Note that quench rates, which vary during simulations, have been adjusted with respect to the evolution of the average potential energy with temperature, as prescribed in former work [34]. The five quench schemes used in this work are detailed in Fig. S2 and have total quench time values  $\tau_Q$  of 0.5, 1.5, 4.0, 4.4 and 8.4 ns. For comparison, the IGAR reconstructions in Ref. 34 used a quench with  $\tau_Q = 3.23 \text{ ns}$ . Regarding the structural and textural parameters, the database was constructed based on two sets of simulations. In the first one,  $OA_t$  is fixed at  $30^\circ$  and  $L_t$  varies in the in the 1-8 *nm* range by step of 0.5 *nm*. In the second set,  $OA_t$  varies from  $10^\circ$  to  $100^\circ$  by steps of  $10^\circ$ , excluding the case  $OA_t = 30^\circ$ , for three fixed values of  $L_t$ : 2, 4 and 6 nm.

The image-guided quench is performed with the same in-house code as in former studies [34, 36, 43, 44]. In these simulations, interatomic interactions are described by the second-generation reactive empirical bond order potential

(REBO2) [51]. The image potential at every atomic position is computed by trilinear interpolation using the eight closest data points in the 3D textured images (*i.e.* the eight vertices of the enclosing voxel) and applying a conversion factor  $k_{im} = 4$  eV. Newtonian equations of motions are integrated with a velocity-Verlet integrator [52] using a time step of 0.25 fs. The system temperature, imposed through the stochastic thermostat proposed by Andersen [53], is progressively decreased from an initial value of 8000 K down to 0 K, with a minimal quench rate operating around the carbon melting point (see Fig. S2 for details). The collision frequency of the thermostat is set to  $4 \times 10^{12} \text{ s}^{-1}$ .

The initial atomistic configuration, common to all reconstructions, has been prepared as follows. First, a  $24 \times 24 \times 23$  replication of the 16 atoms graphite orthorhombic unit cell was rescaled to a cubic box at the suited volume, and a number of atoms were removed to match the suited density. Then this configuration has been equilibrated for 2 ns at 8000 K, using MD in the NVT ensemble, with the LCBOPII potential [54] implemented in the STAMP code [55]. Melting was quickly observed in the simulation and the well-equilibrated liquid configuration obtained at the end of the 2 ns was selected as initial configuration for the IGAR quench simulations. The LCBOPII potential, improving over earlier empirical potential by including both medium-range and long-range (*i.e.* van der Waals) interactions, was selected because it provides a good description of carbon phases in both liquid [56] and solid [57–59] states. Recent work has also shown that it is one of the most suitable potentials to investigate elastic properties at high and ultra-high temperatures

[60], which will be the purpose of future investigations.

After the reconstruction, this code and potential are used to relax the atomistic models using MD in the NPT ensemble at 300 K and 1 atm, in the absence of image potential. These simulations also use a velocity Verlet integrator with the same time step as in the IGAR quench (0.25 fs). Temperature (both NVT and NPT) and pressure (NPT) are controlled with Nosé-Hoover style thermostat and barostat [61, 62], operated with damping constants of 0.1 and 1 ps for temperature and pressure, respectively. Note that the barostat acts independently on the three cell lengths to fix the three diagonal elements of the pressure tensor to the set value. Relaxations are run for 2 ns, amongst which the last 200 ps are used to determine equilibrium properties such as density, enthalpy as well as structural and textural properties, as described below.

## 2.2. Structural analysis

Standard, single configuration coordination analysis is performed on the reconstructed models. As in former studies [34, 36, 63], a 1.85 Å cutoff distance is used to determine the network of covalent bonds and determine bond statistics. This allows determining atoms forming bonds with two, three and four neighbors, named as C<sub>2</sub>, C<sub>3</sub> and C<sub>4</sub> atoms, respectively and generally associated to sp, sp<sup>2</sup> and sp<sup>3</sup> hybridized carbon atoms, even though some C<sub>2</sub> atoms could also correspond to sp<sup>2</sup> radicals. Then, an analysis of the ring structure is performed on the subset of C<sub>3</sub> atoms (i.e. sp<sup>2</sup>) using the “shortest path ring” algorithm proposed by Franzblau [64], yielding ring statistics. Coordination and ring structure data are finally used to identify six different

atom types depending on their local environment, as proposed in Ref. 36 and then redefined in Ref. 63. In this classification, the  $C_3$  atoms are split into four categories:  $C_3^\alpha$  corresponds to  $sp^2$  atoms involved in three hexagonal rings;  $C_3^\beta$  are  $sp^2$  atoms involved in three rings, including at least one non-hexagonal ring;  $C_3^\delta$  are  $sp^2$  atoms bonded to at least one  $C_4$  ( $sp^3$ ) atom; and  $C_3^\gamma$  are  $sp^2$  atoms bonded to at least one  $C_2$  ( $sp$ ) atom.

MD trajectories of the relaxed models were also analyzed to characterize structural properties. First, the atomic pair distribution function is computed for each model:

$$g(r) = \frac{1}{4\pi r^2 \rho N} \sum_{i=1}^N \sum_{j \neq i}^N \delta(r - r_{ij}) \quad (2)$$

where  $N$  is the number of atoms,  $\rho$  the total number density ( $N/V$ ) and where  $\delta(r - r_{ij}) = 1$  when carbon atoms  $i$  and  $j$  are distant by  $r$  and 0 otherwise. The total reduced pair distribution function  $G(r) = 4\pi r \rho [g(r) - 1]$  can be Fourier transformed to produce the structure factor:

$$S(Q) = 1 + \frac{1}{Q} \int_0^{r_{max}} G(r) \sin(Qr) dr \quad (3)$$

where  $r_{max}$  is set to 6 nm (*i.e.*  $\sim$  half the simulation cell width).

The structural parameters  $L_c$ ,  $L_a$  and  $d_{002}$  are extracted from  $S(Q)$  according to the following procedure. First, structure factors are converted from  $Q$  space to  $2\theta$  space according to the relation  $Q = \frac{4\pi \sin(\theta)}{\lambda}$  where  $\lambda = 1.542 \text{ \AA}$  is taken as the X-Ray Cu  $K_\alpha$  wavelength. As in former works [33, 34], the 002 peak is fitted with a pseudo-Voigt function and the  $d_{002}$  and  $L_c$  parameters are obtained from the peak position and width at half maximum, using Bragg's

law and Scherrer's relation, respectively. The  $2\theta$  domain of the 10 band and 004 peak is adjusted with the sum of a skewed Gaussian (10) and a Pearson VII function (004).  $L_a$  is then obtained from the width at half maximum of the fitted 10 band using Warren's relation:  $L_a = 1.77\lambda/(\beta\cos\theta)$ , where  $\beta$  is the width of the peak and  $\theta$  its position.

In the last section of this paper, these data are compared to experimental data, including structural parameters obtained from XRD, and  $G(r)$  and  $S(Q)$  data obtained from Neutron diffraction [33, 34]. While the computed data can be directly compared to XRD data,  $S(Q)$  data derived from neutron diffraction are subject to an important experimental broadening. This artifact needs to be accounted for in the computed  $S(Q)$  and  $G(r)$  for quantitative comparison[34]. The neutron-corrected structure factor  $S^N(Q)$  is thus obtained as :

$$S^N(Q) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^{Q_{max}} S(Q) \exp - \left( \frac{(Q - q)^2}{2\sigma^2} \right) dq \quad (4)$$

where  $\sigma = (\sigma_{instru}^2 - \sigma_{rmax}^2)^{\frac{1}{2}}$  is the broadening parameter accounting for the instrumental broadening  $\sigma_{instru}$  and subtracting the artificial broadening  $\sigma_{rmax}$  due to the finite  $r$  range in the computed  $S(Q)$ . The first term  $\sigma_{instru}$  is given by a 6<sup>th</sup> order polynomial:  $\sigma_{instru} = -13.33 \times 10^{-9} Q^6 + 788.8 \times 10^{-9} Q^5 - 10.4 \times 10^{-6} Q^4 - 24.31 \times 10^{-6} Q^3 + 8.756 \times 10^{-3} Q^2 - 62.51 \times 10^{-3} Q + 239.8 \times 10^{-3}$ . Evaluations of  $S(Q)$  using various  $r_{max}$  values have shown that  $\sigma_{rmax} = \frac{3}{r_{max}}$  provides a good estimation of the truncation broadening. Finally, the neutron-corrected pair distribution function  $G^N(r)$  is given by the inverse



Fourier transform of  $S^N(Q)$ :

$$G^N(r) = \frac{2}{\pi} \int_0^{Q_{max}} Q [S^N(Q) - 1] \sin(Qr) dQ \quad (5)$$

where  $Q_{max}$  is set to  $23.5 \text{ \AA}^{-1}$  [34].

### 2.3. Textural analysis

HRTEM image simulation is performed on every reconstructed model with two main purposes: i) determine the orientation angle (OA) corresponding to the models, and ii) determine quantitative HRTEM image descriptors that can be compared to experimental data. All the image simulations are performed with Dr Probe [65]. Standard microscope parameters are used [33, 34]: 300 kV voltage, 1.2 and 1.5 mm spherical and chromatic aberration, respectively, -58 nm Scherzer defocus and 7 mm defocus.

A fast Fourier transform (FFT) is first performed on the simulated HRTEM image to obtain its spectrum. Then, 40 circular intensity profiles with splines are performed on a ring of thickness  $r$ . The limits of the ring  $r_{min}$  and  $r_{max}$  are given by the extremums of the 002 arcs present on the pattern. These 40 circular profiles are then subsequently averaged into a "master" intensity profile curve, over which OA is determined as the full width at half maximum (FWHM) of this profile, using a gaussian fit (see Fig. S3).

Simulated HRTEM images are also analyzed in terms of fringe lengths, tortuosities and orientations. As in former works, an iso-grey level curve tracking algorithm [66] allows retrieving lattice fringes in HRTEM images and computing statistics on their length  $L_2$  and tortuosity  $\tau$ , the latter be-

ing defined as the ratio of their curvilinear length  $L_2$  over their end-to-end Euclidean distance. Orientation statistics are drawn following the procedure described in [67]. Local orientation is defined at any pixel of an image as the direction of the iso-grey level curve (the direction  $a$  in the case of graphite single crystal). Local orientations are computed using convolutional operators based on differential geometry [68] and regularized using the standard structure tensor technique, which provides at every pixel  $u$  both an orientation estimate  $\theta_u$  and its confidence index  $\eta_u$ . The latter are finally processed to produce the Rotation Invariant Mean Orientation Difference  $RIMOD(r, \Psi)$ , *i.e.* the average orientation deviance between two pixels located at a given distance  $r$  and relative angle  $\Psi$  from each other:

$$RIMOD(r, \Psi) = \frac{1}{\sum_u \eta_u \eta_{u+\delta u}} \sum_u \eta_u \eta_{u+\delta u} |\theta_{u+\delta u} - \theta_u| \quad (6)$$

with  $\delta u = (r \cos(\theta_u + \Psi), r \sin(\theta_u + \Psi))$ . In such a local coordinate system,  $\Psi = 0^\circ$  and  $90^\circ$  respectively correspond to the  $a$  and  $c$  directions.  $RIMOD$  can differentiate between longitudinal (*i.e.*, along the fringes,  $\Psi = 0^\circ$ ) and transverse (*i.e.*, across the fringes,  $\Psi = 90^\circ$ ) coherence losses. It has also shown to be a good indicator as how coherence vanishes with distance, *e.g.* [47, 67]. An interesting feature is the value of the plateau  $\beta_{MOD}$  reached by  $RIMOD(r, \Psi)$  when  $r \rightarrow \infty$ ,  $\beta_{MOD}$  being interpreted as the long-term orientation deviance within the HRTEM image.

#### 2.4. Machine Learning model

A random forest regressor (RFR) [69] is a non-parametric machine learning (ML) algorithm used for predicting a specific property given a set of input features commonly called descriptors. An RFR generates a large number of decision trees, each trained on a random subset of descriptors, and combines their predictions to produce a final output. This combination of individual predictions from multiple trees helps reducing overfitting and improves the accuracy of the model. In other words, given a set  $\mathbf{x}$  of descriptors, the final predicted properties  $\mathbf{y}$  are averaged over the individual intermediate predictions  $\mathbf{y}_i$ :

$$\mathbf{y}(\mathbf{x}) = \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbf{y}_i(\mathbf{x}) \quad (7)$$

with  $N_t$  the total number of decision trees. In the present case, the structural/textural properties ( $L_c$ ,  $L_a$ , OA) of reconstructed pyrocarbons are predicted while statistics on local atomic environments ( $C_3^{\alpha,\beta,\gamma,\delta}$ ,  $C_2$ ,  $C_4$ ) and ring structures ( $R_5$ ,  $R_6$ ,  $R_7$ ) play the role of input features.

The entire dataset was split into two (training 70% and testing 30%) sub-datasets in order to assess the predictive capability of the model and two metrics were used for quantifying the error of the prediction, e.g. the Root Mean Squared Error (RMSE) and the coefficient of determination ( $R^2$ ), widely used together with ML techniques. When dealing with small databases, RFR has the advantage of possessing very few parameters such as the number of decision trees  $N_t$ , their depth  $D_t$  and their minimum number of samples  $N_{\text{samples}}^{\text{leaf}}$ . These parameters were set to 100, 15 and 5 respectively, and the RFR implementation of the `sklearn` Python library [70] was used.

### 3. Results

#### 3.1. Testing the PG-IGAR method

Although grain sizes and orientations are prescribed during the textured image construction, there is no guarantee that a one-to-one correspondence exists between the structural and textural parameters computed from the reconstructed models and these simulation inputs. We discuss this point in this section by considering how input parameters and quench time actually impact the computed  $L_a$ ,  $L_c$  and OA parameters.

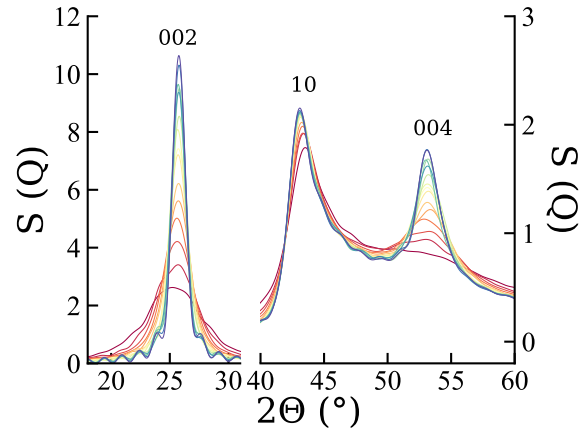


Figure 2: Structure factors  $S(Q)$  of the PG-IGAR models reconstructed using  $OA_t = 30^\circ$ ,  $\tau_Q = 4.0$  ns and grain sizes ranging from 1 (dark-red) to 8 (dark-blue) nm, by steps of 0.5 nm.

Fig. 2 shows the evolution of the computed  $S(Q)$  with increasing grain size at fixed OA ( $30^\circ$ ) and  $\tau_Q$  (4.0 ns). We observe a significant increase of the 002 and 004 peaks with increasing grain size from 1 to 8 nm, indicating an increase in the “out-of-plane” organization. However, although some increasing trend is also observed for the 10 band, the latter is much more

moderate than for 002 and 004, indicating a more limited increase in the “in-plane” order. Oscillations also appear for the models with the largest grain sizes. This corresponds to an expected truncation effect for models in which stacking order exceeds the distance cutoff  $r_{max}$  (6 nm) used in the Fourier transform of  $G(r)$ .

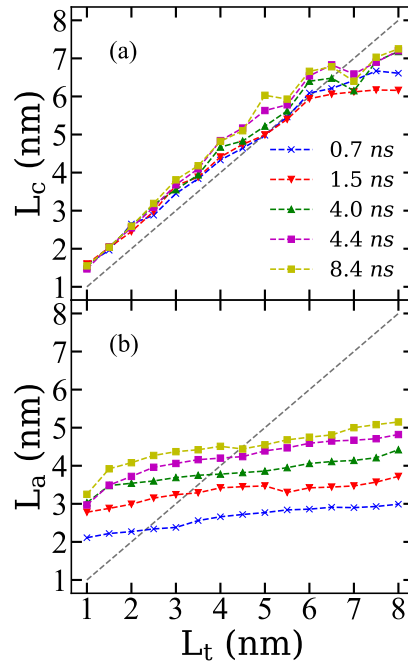


Figure 3: Evolution of (a)  $L_c$  and (b)  $L_a$  with imposed grain size ( $L_t$ ) obtained at various quench rates at fixed  $OA_t = 30^\circ$ .

Computed values of  $L_c$  and  $L_a$  at fixed  $OA_t = 30^\circ$  and various quench rates are presented as a function of  $L_t$  in Figure 3. We observe that the computed  $L_c$  value is always close to the imposed grain size  $L_t$ , regardless of the quench rate (Fig. 3a), showing that  $L_c$  is easily parameterized in the PG-IGAR method. Although, looking more closely, computed  $L_c$  are slightly

larger than  $L_t$ , typically by  $\sim 0.5$  nm, for low values of  $L_t$ , and then slightly lower, by 0.5 to 2 nm and with some dependence on quench rate, at  $L_t = 8$  nm.

The correlation between computed  $L_a$  and  $L_t$  is less clear, as shown in Fig. 3b. First, although  $L_a$  does increase with  $L_t$ , we observe that  $L_a$  is significantly larger than  $L_t$  at short  $L_t$ , and lower at large  $L_t$ . Furthermore, we observe that  $L_a$  strongly increases with quench time  $\tau_Q$  in the investigated range. These observations can be easily rationalized by considering that increasing  $\tau_Q$  leads to a thermodynamic stabilization of the graphenic domains by increasing the fraction of hexagonal ring, and decreasing the fraction of non-hexagonal rings, in the models [46]. Therefore, we can assume that with the considered  $\tau_Q$ , the system does not have enough time to build graphene domains as large as 8 nm and thus possess intragranular boundaries and point defects. To support this statement, we add that in Ref. 46 hexagonal domains no larger than 3-5 nm were obtained using a similar quench scheme for a perfectly flat monolayer. Conversely, for low  $L_t$ , hexagonal domains can extend through neighboring grains provided that the latter have a similar orientation, which is often the case at  $OA_t = 30^\circ$ .

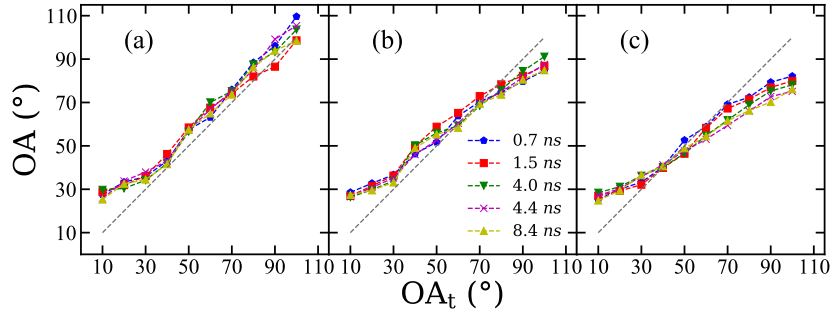


Figure 4: Evolution of the computed OA as a function of the target value  $OA_t$  for different quench times and  $L_t = 2$  (a), 4 (b) and 6 (c) nm.

Figure 4 presents a comparison of the computed orientation angle (OA) and the imposed orientation angle ( $OA_t$ ) for various grain sizes and quench times. The PG-IGAR method is largely successful in imposing OA and there is no significant dependence of the results on quench time. However, a noticeable deviation is always present at low  $OA_t$ , typically below 30 degrees, where the computed OA values surpass the corresponding  $OA_t$  values. To investigate the source of this behavior, the OA values corresponding to 2D projection images of textured image blocks constructed with various  $OA_t$  were computed and a similar behavior was observed (Fig. S4). This suggests that the behavior originates from the construction of the HRTEM signal, which is 2D in nature and is not related to the atomic aspect of the reconstruction. Additionally, while OA and  $OA_t$  show similar values for small grains at large  $OA_t$ , computed OAs tend to be lower than  $OA_t$ s for large grains for  $OA_t > 70^\circ$ . We show in Fig. S4 that this deviation occurs during the (post-reconstruction) relaxation of the models.

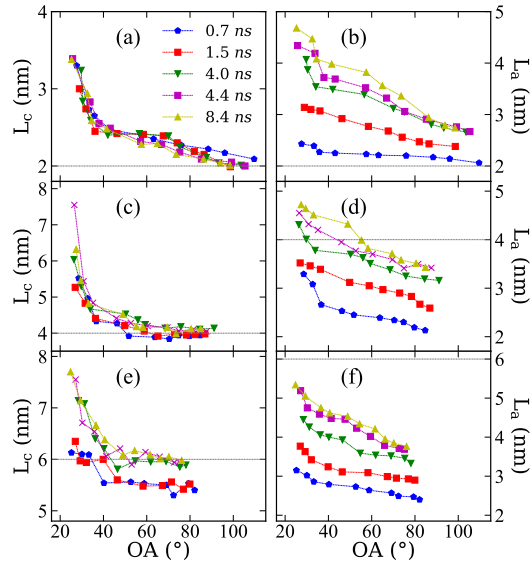


Figure 5: Evolution of computed coherence lengths  $L_c$  (a,c,e) and  $L_a$  (b,d,f) as a function of the computed OA for target grain sizes  $L_t = 2$  (a,b), 4 (c,d) and 6 (e,f) nm, and different quench times  $\tau_Q$ .  $L_t$  values are indicated with grey solid lines for clarity.

Fig. 5 shows the effect of OA on the obtained values of the structural parameters for a few values of the target grain size, namely 2, 4 and 6 nm. The dependence of both  $L_c$  and  $L_a$  on OA is obvious. Whatever the quench rate, both  $L_c$  and  $L_a$  decrease with increasing OA for a given target grain size. In the case of  $L_c$ , computed values are always larger than  $L_t$  at low OA, and seems to converge towards it at large OA. The only exception to this behavior is observed for the largest grain size and shorter  $\tau_Q$  (Fig. 5), which clearly relates to a convergence issue (quench time effect) of the reconstructed model. These observations indicate that stacking coherence easily extends to neighboring grains in high textured (low OA) materials. As mentioned above, the convergence of  $L_a$  with quench time is generally not



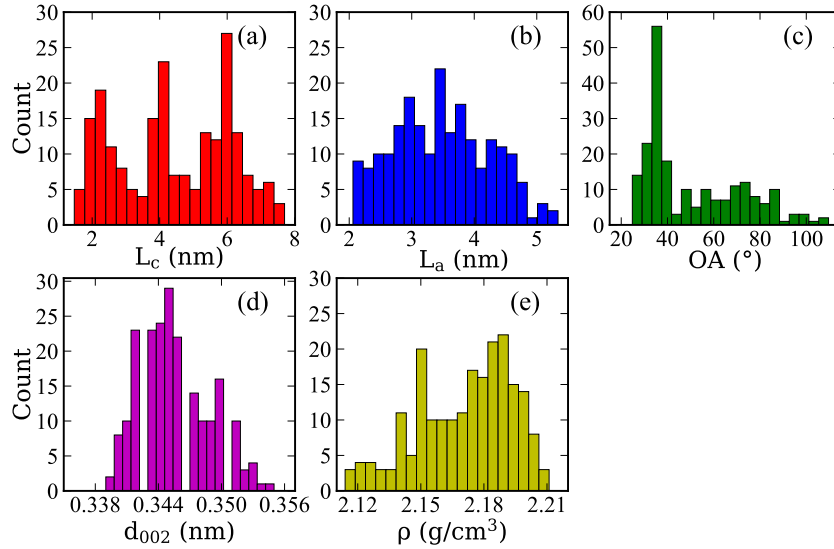


Figure 6: Distribution of (a)  $L_c$ , (b)  $L_a$ , (c) OA, (d)  $d_{002}$  and (e)  $\rho$ , in the database.

achieved in our reconstructions, especially for large grain sizes. Yet, looking at the case  $L_t = 2$  nm (Fig. 5d), for which data seem converged with respect to  $\tau_Q$  for  $\tau_Q \geq 4$  ns, we see that  $L_a$  behaves as  $L_c$ , namely that it is significantly larger than  $L_t$  for low OA and should converge towards  $L_t$  as OA increases towards isotropy. No clear conclusion can be inferred from the data at  $L_t = 4$  nm (Fig. 5e) and 6 nm (Fig. 5f), because of the magnitude of quench time effects.

### 3.2. Presentation of the database

Fig. 6 shows the distributions of some important measurable properties within the database, namely  $L_c$ ,  $L_a$ , OA,  $d_{002}$  and  $\rho$ . As shown in Fig. 6a, computed  $L_c$  span values ranging from  $\sim 2$  to 7.5 nm, with distinct peaks in the distribution at 2, 4 and 6 nm, corresponding well to the distribution of input  $L_t$  in the reconstructions. As discussed above,  $L_a$  values are less directly

related to  $L_t$  and the distribution is slightly narrower, 2 to 5 nm, and uniform, with a peak around 3.5 nm (Fig. 6b). As for  $L_c$ , the distribution of OA (Fig. 6c) showing a broad range of values from 20 to 110°, with a strong peak around 30°, also well reflects the distribution of  $OA_t$ .

Interlayer spacing  $d_{002}$  ranges from 0.340 to 0.354 nm with a maximum distribution at 0.345 nm (Fig. 6d) while density ranges from 2.12 to 2.21 g/cm<sup>3</sup>, with a maximum distribution at 2.19 g/cm<sup>3</sup> (Fig. 6e). It is interesting to note that the covered domains for  $L_c$ ,  $L_a$  and OA are considerably larger than the ones covered by the previously reported IGAR models[34], especially for  $L_c$  and OA. Also, the  $d_{002}$  values of the PG-IGAR models is well centered around the target (experimental) value whereas IGAR models were significantly overestimating this quantity (i.e. values in the 0.353 to 0.366 nm).

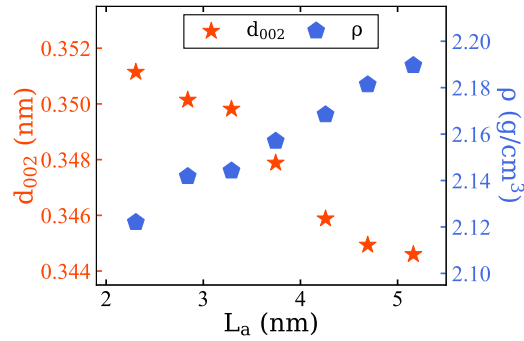


Figure 7: Evolution with  $L_a$  of the average interlayer spacing  $d_{002}$  and density  $\rho$  restricted to models having OA in the 20-40 ° range. Data have been block-averaged by subsets of data having a 0.5 nm  $L_a$  spread.

Fig. 7 shows the values of  $d_{002}$  and  $\rho$  as a function of  $L_a$  for the subset of models having OA values in the 20-40° range. While all the models were reconstructed with the same input values, namely  $d_{002} = 0.345$  nm and  $\rho =$

2.16 g/cm<sup>3</sup>, we see that both parameters show a significant dependency on  $L_a$  after relaxation. As expected, with increasing  $L_a$ ,  $d_{002}$  decreases and  $\rho$  increases towards the values expected for graphite, even though remaining in the common ranges for turbostratic carbons.

Snapshots of some selected models are shown in Fig. 8 to better illustrate the variability in the structure and texture within the database. Models shown in Fig. 8a-c have similar OA (35-39°) and  $L_a$  (2.3-3.0 nm) but show an increasing value of  $L_c$  from 2.6 to 6.6 nm. Fig. 8d-f have similar  $L_c$  (6.1-7.7 nm) and OA (25-28°) but increasing  $L_a$ , from 3.1 to 5.4 nm. Finally, models in Fig. 8g-i have close values of  $L_c$  (2.1-2.5 nm) and  $L_a$  (2.7-2.9 nm) but increasing OA from 35 to 98°. It is obvious that models with small  $L_a$  and  $L_c$ , and large OA, like the model in Fig. 8i have more defect than those with large  $L_a$  and  $L_c$  and small OA as in Fig. 8f. We also note that models in Fig. 8d-f have been obtained with the same textured image, yet with different quench times, leading to very different  $L_a$  values.

The distributions of the different atom types in the database are shown in Figure 9. Overall, the most encountered atom type is  $C_3^\alpha$ , corresponding to graphene-like environments, which shows a skewed distribution in the 40-88 % range, with a peak at 80 %. The second most encountered atom type is  $C_3^\beta$ , which corresponds to threefold ( $sp^2$ ) atoms involved in non-hexagonal rings, counting from 10-45 %, then follow threefold atoms bonded to fourfold,  $C_3^\gamma$ , from 0.5 to 9 %, and threefold atoms bonded to twofold atoms,  $C_3^\delta$ , from 0.5 to 2.8 %. Fourfold ( $sp^3$ ),  $C_4$ , and twofold,  $C_2$ , atoms are also present, yet in low amounts: 0.1-2.5 % and 0.2-2 %, for  $C_4$  and  $C_2$ , respectively.

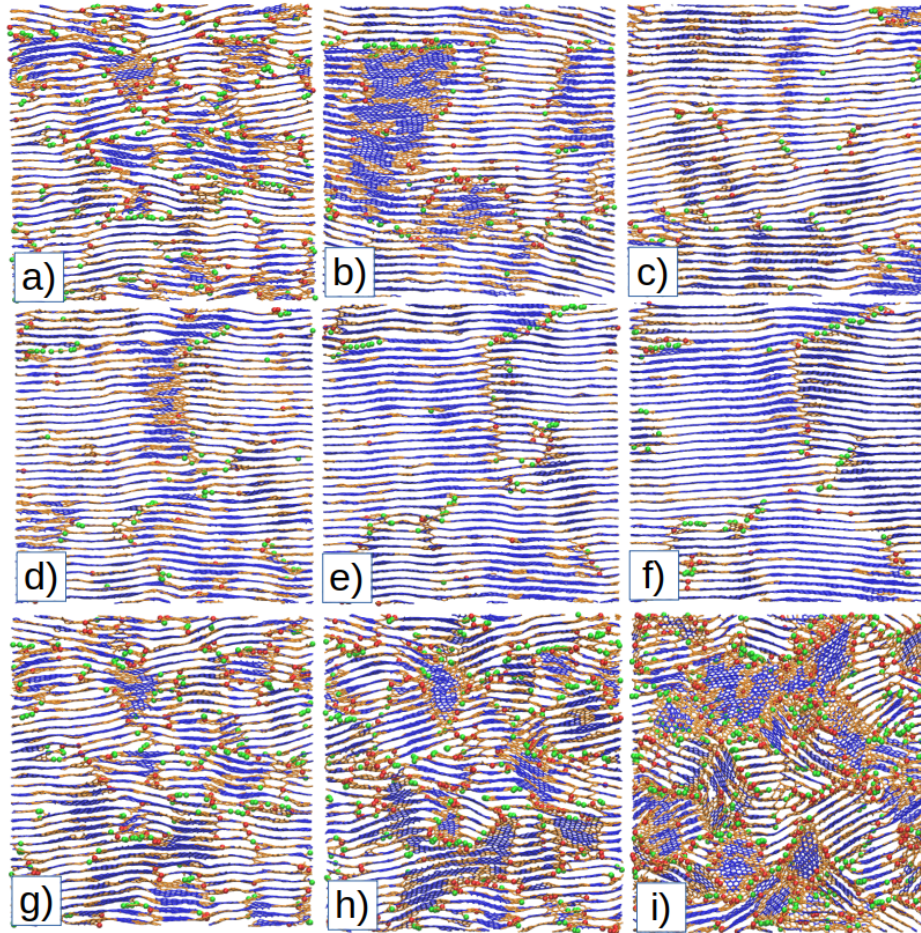


Figure 8: Snapshots of one nm thick cross-sections of PG-IGAR models of increasing (a-c)  $L_c$ , (d-f)  $L_a$  and (g-i)  $OA$ , the other parameters remaining  $\sim$  constant. Blue: bonds between  $C_6$  (graphene-like) atoms; orange: other (defect) bonds; red spheres:  $C_4$  atoms; green spheres:  $C_2$  atoms. Detailed  $\{L_c, L_a, OA\}$  triplets read (a):  $\{2.6, 2.3, 36\}$ , (b):  $\{5.0, 2.7, 35\}$ , (c):  $\{6.6, 3.0, 39\}$ , (d):  $\{6.1, 3.1, 25\}$ , (e):  $\{7.1, 4.4, 28\}$ , (f)  $\{7.7, 5.4, 25\}$ , (g):  $\{2.5, 2.9, 35\}$ , (h):  $\{2.4, 2.8, 58\}$  and (i):  $\{2.1, 2.7, 98\}$ .

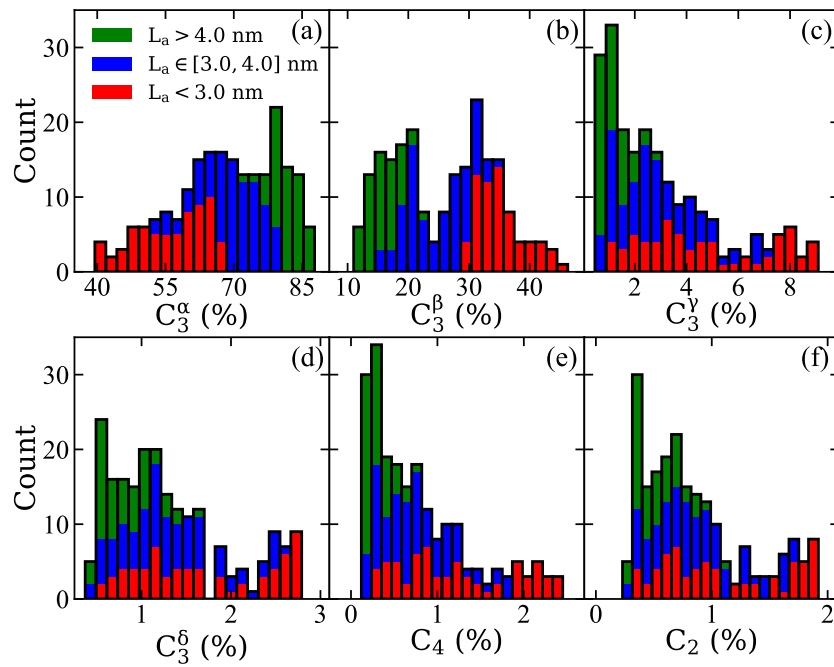


Figure 9: Distribution of the different atom types within the database sorted from (a) to (f) by relative weight. (a):  $C_3^\alpha$ , (b):  $C_3^\beta$ , (c):  $C_3^\gamma$ , (d):  $C_3^\delta$ , (e):  $C_4$ , (f):  $C_2$ . Contributions of models with small (red), medium (blue) and large (green)  $L_a$  are stacked from bottom to top, respectively.

As shown in Fig. 9, there is a marked correlation between the atom type distributions and the  $L_a$  parameter. Especially we observe that the percentage of  $C_3^\alpha$  atoms increases with  $L_a$  while the percentage of  $C_3^\beta$  atoms decreases, which is consistent with the fact that  $C_3^\alpha$  atoms form the coherent domains (or grains) that contribute to  $L_a$ , and  $C_3^\beta$  their grain boundaries and point defects (Stone-Wales, etc.). We also observe that the percentages of the four other atom types decrease with increasing  $L_a$ , indicating that fourfold and twofold atoms are mostly present at the edges of graphene domains, as free edges ( $C_2$ ) or interlayer cross-links ( $C_4$ ), as can be visually perceived on the snapshots in Fig. 8 and discussed in former investigations [34, 44].

Atom type distributions similar to Fig. 9 yet colored according to  $L_c$  and OA are given in Fig. S5 and S6, respectively. The correlation between atom types and  $L_c$  is very similar to the one with  $L_a$  as these parameters are quite correlated in the database. Regarding OA, we observe a strong correlation with the amounts of  $C_3^\beta$ ,  $C_3^\gamma$ ,  $C_3^\delta$ ,  $C_4$  and  $C_2$ , all increasing with increasing OA. This indicates that domain boundaries, which are sharper at large OA, contain a higher density of defect atom types than at low OA (see also Fig. 8g-i).

The distributions of hexagonal ( $R_6$ ), pentagonal ( $R_5$ ) and heptagonal ( $R_7$ ) rings are shown in Fig. 10. As expected the ring content is dominated by hexagons (82-96 %), which, following the distribution of  $C_3^\alpha$ , increases with increasing  $L_a$ . Pentagons and heptagons are found in even amounts over the database (i.e. from 2 to 8 %), which is expected in grain boundaries or planar point defects [45, 46].

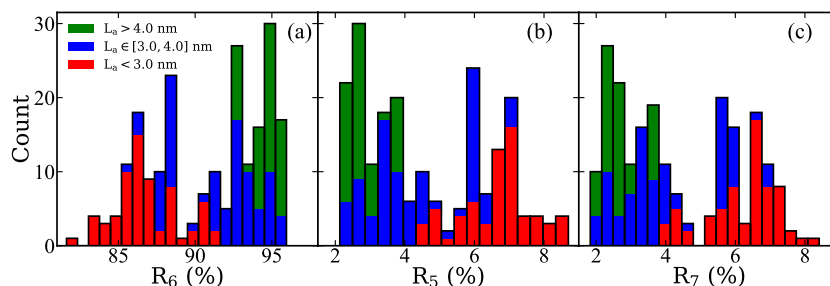


Figure 10: Distribution of (a) hexagons, (b) pentagons and (c) heptagons in the database. Contributions of models with small (red), medium (blue) and large (green)  $L_a$  are stacked from bottom to top, respectively.

### 3.3. Machine learning model

Here, we aim at connecting the structural/textural parameters  $L_a$ ,  $L_c$ , OA to the distributions of local atomic environments and ring statistics using the RFR model applied to the database detailed above. In the first place, the entire set of descriptor is used to predict  $L_a$ ,  $L_c$  and OA and the corresponding correlation graphs are presented in Figure 11. Excellent agreement is obtained

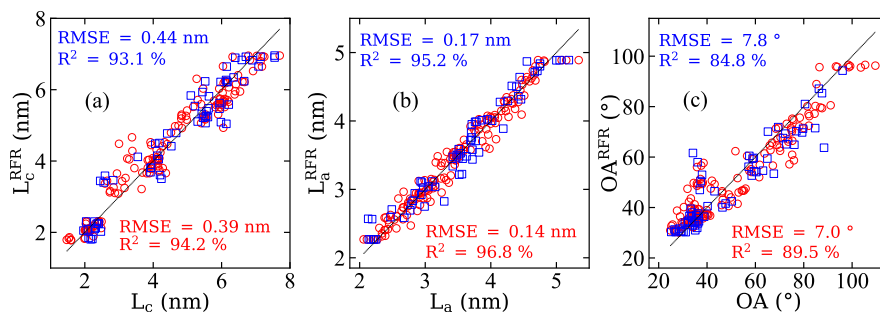


Figure 11: RFR predictions of (a)  $L_c$ , (b)  $L_a$  and (c) OA. Red and blue squares correspond to the predictions in the training and test sets, respectively. The diagonal black lines indicate the equality between actual and predicted features, as a guide to the eye.

for both structural and textural parameters and the RMSE/ $R^2$  scores for both training and testing datasets are reported on the graphs. A common practice

while using RFR consists in investigating the sensitivity of the prediction to the descriptors. In other words, one can find the features that are the most correlated to the predicted properties in order to reduce the features space, even though in the present case its dimension is rather small. Such strategy was applied to the database and the prediction results are displayed in Figure 12. It is found that  $L_a$  can be predicted from the fractions of  $C_3^\alpha$  and  $C_3^\beta$

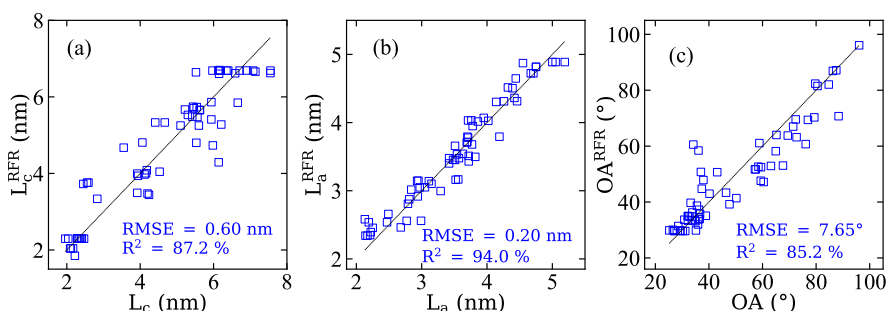


Figure 12: RFR predictions of (a)  $L_c$ , (b)  $L_a$  and (c) OA using reduced feature sets. Only the validation set is displayed. Considered features are (a)  $C_2$  and  $C_3^\delta$ , (b)  $C_3^\alpha$  and  $C_3^\beta$ , (c)  $C_3^\gamma$ ,  $C_3^\delta$ ,  $C_4$  and  $C_2$ .

atoms while considering only  $C_2$  and  $C_3^\delta$  atoms allows for a good reproduction of  $L_c$ . However, prediction of the textural parameter OA, requires more input features, namely  $C_3^\gamma$ ,  $C_3^\delta$ ,  $C_4$  and  $C_2$  atoms. An almost equivalent accuracy is obtained for the predictions of  $L_a$  and OA in comparison with the initial model. On the other hand, the reduced features model leads to a slightly larger error for the prediction of  $L_c$ , yet remaining acceptable. It is noted that some values of  $R^2$  are actually higher for the optimized model, meaning that some features were acting as noise and bias in the original RFR model.



### 3.4. Comparison with actual pyCs

We compare in table 1 the properties of six selected PG-IGAR models to the actual experimental properties measured for six pyC samples – including SL, RL and ReL pyCs as prepared and the ReL pyC sample after heat treatment at 1300 (ReL<sub>1300</sub>), 1500 (ReL<sub>1500</sub>) and 1700°C (ReL<sub>1700</sub>) – from which the properties have been reported in former work[33, 34]. The models were selected among the 210 models database as those minimizing the following error metric:

$$d_{err} = \frac{1}{3} \left( \frac{|L_c - L_c^e|}{L_c^e} + \frac{|L_a - L_a^e|}{L_a^e} + \frac{|OA - OA^e|}{OA^e} \right) \quad (8)$$

where  $L_c$ ,  $L_a$  and  $OA$  are the calculated structural and textural parameters and  $L_c^e$ ,  $L_a^e$  and  $OA^e$  the corresponding experimental data. Note that  $d_{002}$  and  $\rho$  data, given in table 1 and showing little variations for most models, are not considered in the model selection procedure.

Table 1: Comparison of experimental and computed structural, textural and density data.  $L_c$ ,  $L_a$  and  $d_{002}$  are given in nm,  $OA$  in degrees and  $\rho$  in  $g/cm^3$ . The error metric  $d_{err}$  defined in Eq. 8 is given in %. Experimental data are taken from Weisbecker et al. [33] for the SL pyC and Farbos et al. [34] for RL and ReL pyCs (including heat treated samples). Estimated error bars on the experimental data are of 0.1 nm, 0.3 nm and 5°, 0.001 nm and 0.02  $g/cm^3$  for  $L_c$ ,  $L_a$ ,  $OA$ ,  $d_{002}$  and  $\rho$ , respectively [34].

pyC	Experiments					Models					$d_{err}$
	$L_c$	$L_a$	$OA$	$d_{002}$	$\rho$	$L_c$	$L_a$	$OA$	$d_{002}$	$\rho$	
SL	2.6	3.3	68	0.345	1.93	2.3	3.3	67	0.342	2.17	4.9
RL	5.2	4.6	22	0.345	2.12	5.5	4.6	29	0.346	2.18	13.3
ReL	2.9	2.6	40	0.346	2.11	2.9	2.3	36	0.347	2.13	7.0
ReL <sub>1300</sub>	3.8	3.1	43	0.347	2.16	3.9	3.3	37	0.349	2.15	7.7
ReL <sub>1500</sub>	6.3	4.1	30	0.345	2.18	6.1	4.1	32	0.346	2.18	3.5
ReL <sub>1700</sub>	13.3	6.2	27	0.343	2.18	7.6	5.2	27	0.345	2.19	20.0

The computed error functions  $d_{err}$  ranges from 3.5 to 20 %, and most values of the considered parameters ( $L_c$ ,  $L_a$  and OA) are actually within the error bars from the experimental data. A noticeable exception is the case of the ReL<sub>1700</sub> pyC, whose large value of  $d_{err}$  (20 %) arises from a strong underestimation of  $L_c$  (7.6 against 13.3 nm). We note that in this case, the experimental value is larger than the considered simulation cell (12.4 nm wide) and therefore cannot be captured in such a model. Computed values of the interlayer distance,  $d_{002}$  are also very close to experimental data, for all models, and the evolution of this quantity with heat treatment for the ReL pyC is also in good agreement with experiments. The evolution with heat treatment of the density of the ReL pyC is also well accounted for by the models. However, density is overestimated for the RL and SL pyC models, certainly because of the use of a rather large density in the PG-IGAR quench simulation (2.16 g/cm<sup>3</sup>). Also, the experimental density of the SL pyC, 1.93 g/cm<sup>3</sup>, suggests the presence of some closed porosity in this material, which would require a special treatment to be accounted for in the model.

The simulated neutron weighted structure factors ( $S^N(Q)$ ) and pair distribution functions ( $rG^N(r)$ ) of the six selected models are compared to the corresponding experimental data in Fig. 13a and 13b, respectively. Overall, we observe a good agreement between the models and experiments for both  $S^N(Q)$  and  $rG^N(r)$ . Especially, the evolution of peak intensities between the different materials is well reproduced in the computed  $S^N(Q)$ , as is the development of long-range correlations in  $rG^N(r)$ . While real-space ( $G^N(r)$ ) data were already relatively-well reproduced by former IGAR high-textured pyC

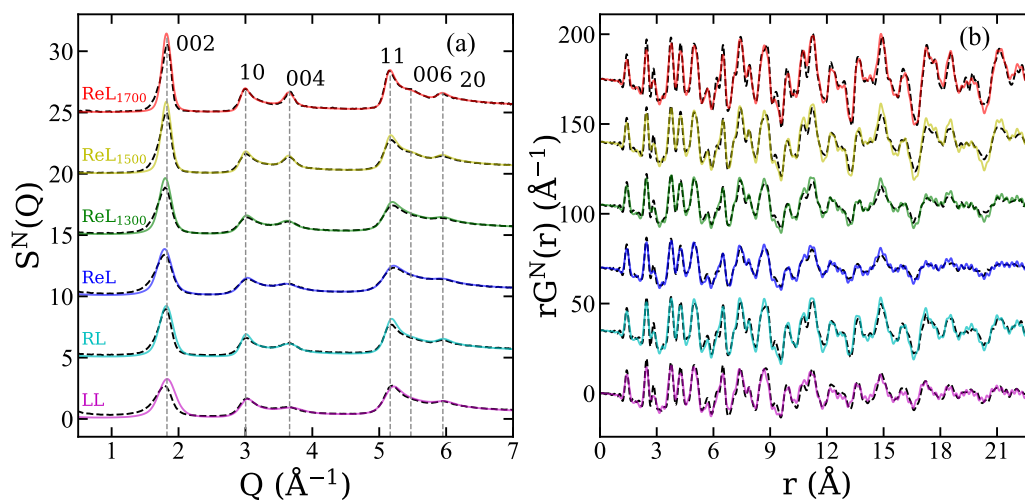


Figure 13: Comparison of experimental (dashed black lines) and calculated (straight colored lines) neutron weighted (a) structure factors  $S^N(Q)$  and (b) reduced pair distribution functions  $rG^N(r)$ . Characteristic turbostratic carbon 3D peaks and 2D bands are indicated with grey vertical dashed lines in (a).

models[34], the PG-IGAR models clearly improve over the latter regarding the reproduction of stacking order in  $S^N(Q)$ . As can be seen in Fig. 13a, the locations and intensities of the 002, 004, as well as the emergence of the 006 peak for the 1700°C heat treated ReL pyC, are extremely well reproduced by the PG-IGAR models, while these features were considerably underestimated in the IGAR models (a comparison of the experimental, IGAR and PG-IGAR  $S^N(Q)$  and  $rG^N(r)$  is given in Fig. S7 for the RL and ReL pyCs).

Table 2: HRTEM image properties of actual pyC samples and their corresponding PG-IGAR models.

pyC	Experiments			Models		
	$L_2$ (nm)	$\tau$	$\beta_{MOD}$ ( $^\circ$ )	$L_2$ (nm)	$\tau$	$\beta_{MOD}$ ( $^\circ$ )
SL	0.33	1.0088	39.3	0.50	1.0044	16.5
RL	0.72	1.0033	8.6	1.67	1.0034	6.9
ReL	0.61	1.0032	10.1	0.92	1.0040	9.3
ReL <sub>1300</sub>	0.63	1.0030	10.6	1.15	1.0039	8.7
ReL <sub>1500</sub>	0.80	1.0029	8.2	1.47	1.0037	7.5
ReL <sub>1700</sub>	1.10	1.0027	7.3	1.87	1.0035	6.8

Some real space analyses of simulated and experimental HRTEM images, including the average fringe length  $L_2$ , tortuosity  $\tau$  and the plateau value  $\beta_{MOD}$  of the rotationally invariant mean orientation difference (RIMOD) are given in table 2. RIMOD plots as a function of distance and local orientation are given in Fig. S8 in the supporting information. HRTEM images and their orientation images from both experiments and simulation are also given for the six materials in Fig. S9. Although quantitative differences exist between the fringe properties extracted from simulated and experimental images, with simulated  $L_2$  and  $\tau$  values being slightly larger than experimental ones, as was already observed in Ref. 34 with the IGAR Models, some qualitative agreement can be observed. In both sets of data, the SL pyC presents the shortest and more tortuous fringes while the ReL<sub>1700</sub> pyC has the longest and less tortuous fringes. Amongst the ReL pyC and its heat treated samples, we observe that, overall,  $L_2$  increases and  $\tau$  decreases with increasing heat treatment temperature. In both data sets (i.e. experimental and simulation

data), the RL pyC shows similar values to those observed for the ReL pyC heat treated at  $\sim 1500^\circ\text{C}$ .

Regarding  $\beta_{MOD}$  values, characterizing the average disorientation between coherent orientation domains in the HTREM images, we observe a good agreement, both qualitative and quantitative, between experimental and simulation data. All high-textured pyCs have  $\beta_{MOD}$  values in the  $6\text{-}11^\circ$  range, with data from simulated images being  $\sim$  one degree lower than those from experimental images. The medium textured pyC (SL) has considerably larger values,  $39.3^\circ$  and  $16.5^\circ$  for the experimental and simulation data, respectively. Note that the reported experimental value for the SL pyC was taken as the value of RIMOD at 6.2 nm, which is the largest distance that can be considered for the simulated images (which are periodic with a width of 12.4 nm). Yet, at this distance RIMOD has not yet reached a plateau in the SL pyC image (see Fig. S8), which indicates that orientations are correlated up to a much larger distance in this material, despite  $L_a$  and  $L_c$  values of  $\sim 3$  nm.

Finally, even though no experimental data exists, to our knowledge, regarding the energetics of pyCs, the calculated enthalpies of the six pyC models, given in table S1, indicate an increase in stability following the order  $\text{SL} < \text{ReL} < \text{ReL}_{1300} < \text{ReL}_{1500} < \text{RL} < \text{ReL}_{1700}$ .

#### 4. Conclusion

In this report we have introduced a new atomistic reconstruction technique, the PG-IGAR method, that uses texture mapping on a Voronoï tessellation to drive a liquid quench towards the formation of dense nanotextured

carbon models. Unlike its predecessor, the IGAR method, which was requiring a complex treatment of experimental HRTEM images, the PG-IGAR method is considerably simpler, as it requires very limited inputs, namely target values of grain size ( $L_t$ ), distribution of orientations ( $OA_t$ ), interlayer spacing ( $d_{002}$ ) and density ( $\rho$ ). We have shown in a first step that some parameters like the  $L_c$  and OA parameters, measurable from X-ray and electron diffraction experiments, are easily controlled in the reconstruction. Conversely, the  $L_a$  parameter, also measurable using XRD, is slightly more difficult to fine-tune, as it depends on the quench time  $\tau_Q$  and can be strongly coupled to OA.

Applying this method by varying independently three reconstruction parameters  $L_t$ ,  $OA_t$  and  $\tau_Q$ , we have built a large database of atomistic models with values of the  $L_c$ ,  $L_a$  and OA parameters covering the domains characteristic of low temperature (*i.e.*  $\leq 1700^\circ\text{C}$ ) high and medium textured pyrocarbons. These allow us identifying some interesting correlation between measurable parameters like the fact that  $d_{002}$  and  $\rho$  respectively decreases and increases with increasing  $L_a$  in high textured pyCs.

The models were then analyzed in terms of the different local atomic environments which were also correlated to the measurable (diffraction) parameters. It was shown that the proportions of atoms in graphene-like environments ( $C_3^\alpha$ ) increase with  $L_a$  while those of the other atom types, corresponding to defects, decrease. Atoms in non-hexagonal  $sp^2$  rings ( $C_3^\beta$ ) were attributed to in-plane grain boundaries and point defects while  $sp$ , or  $sp^2$  radical, ( $C_2$ ) and  $sp^3$  ( $C_4$ ) atoms were attributed to graphene edges and extended cross-links (screw dislocations), respectively.

A machine learning engine was proposed to predict structural and textural parameters from the distribution of such local atomic environments and it was shown that  $L_c$  can be accurately predicted from the proportions of  $C_2$  and  $C_3^\delta$  where the latter are  $sp^2$  atoms bonded to  $C_2$ ,  $L_a$  from the proportions of  $C_3^\alpha$  and  $C_3^\beta$ , and OA from the proportions of  $C_3^\gamma$  (i.e.  $sp^2$  atoms bonded to  $C_4$ ),  $C_3^\delta$ ,  $C_4$  and  $C_2$  atoms.

Then the six models that best match the  $L_c$ ,  $L_a$  and OA values measured for an experimental database of six pyCs were extensively compared to experimental data. It was shown that the models give an excellent reproduction of all the properties of the five high textured pyCs, including  $L_c$ ,  $L_a$ , OA,  $d_{002}$  and  $\rho$  values that were found in better agreement than those obtained with the former IGAR models. This is especially the case of the  $L_c$  parameter that was significantly underestimated with the IGAR models. Neutron diffraction structure factors and pair distribution were found in almost perfect agreement with experimental data, including again the out-of plane diffraction peaks position and width for all models.

Also, the PG-IGAR has allowed us to produce, for the first time, a model of the smooth laminar (medium-textured) pyC, with the same accuracy as the one discussed above for most properties. The only exception is the density of the reconstructed SL pyC model,  $\sim 12\%$  too large. This indicates that this material comprises some, probably closed, micropores, that should be explicitly included in the reconstruction. Another future refinement should be the inclusion of hydrogen (0 - 2.5 at. %), which could be achieved as in former works with the IGAR method [34]. Eventually, a quantitative comparison of

experimental and simulated HRTEM images was presented for the six materials. The agreement between experiments and simulations on such properties was found as good as the one obtained with the IGAR method, in which HRTEM images properties were used as inputs, validating the superiority of the PG-IGAR method.

Compared to the earlier IGAR method, PG-IGAR indeed appears more straightforward, as discussed above, more versatile, as it can be applied to any level of texture, and more accurate, as out-of-plane diffraction properties are accurately captured. Another clear advantage is that it allows constructing models of any prescribed nanotexture, existing or not, which can be a clear advantage in building structure (or texture)-property relationships for this class of materials.

### **CRedit authorship contribution statement**

**Franck Polewczyk:** Methodology, Software, Formal analysis, Investigation, Visualization, Data curation, Writing – original draft. **Paul Lafourcade:** Conceptualization, Methodology, Software, Formal analysis, Resources, Data curation, Writing – original draft. **Jean-Pierre Da Costa:** Methodology, Software, Formal analysis, Validation, Writing – review & editing. **Gérard Vignoles:** Conceptualization, Formal analysis, Resources, Validation, Writing – review & editing. **Jean-Marc Leysale:** Conceptualization, Methodology, Software, Formal analysis, Resources, Data curation, Validation, Writing – original draft.



## **Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## **Acknowledgments**

JPDC, GV and JML are grateful to the Late Patrick Weisbecker for co-initiating this research. Nicolas Pineau and Christophe Denoual are also gratefully acknowledged for useful discussions about this work. Financial support from the French Ministry of Defence - Defence Innovation Agency is gratefully acknowledged by FP. Sandra Bouiller is also acknowledged by FP for her help in data transfer. Atomistic simulations were performed using the computational resources of the TGCC-CEA.

## **References**

- [1] A. Oberlin, Pyrocarbons, *Carbon* 40 (2002) 7–24. doi:[https://doi.org/10.1016/S0008-6223\(01\)00138-5](https://doi.org/10.1016/S0008-6223(01)00138-5).
- [2] P. Delhaès, Chemical vapor deposition and infiltration processes of carbon materials, *Carbon* 40 (5) (2002) 641–657. doi:[https://doi.org/10.1016/S0008-6223\(01\)00195-6](https://doi.org/10.1016/S0008-6223(01)00195-6).
- [3] G. L. Vignoles, F. Langlais, C. Descamps, A. Mouchon, H. Le Poche, N. Reuge, N. Bertrand, CVD and CVI of pyrocarbon from various precursors, *Surf. Coat. Technol.* 188-189 (2004) 241–249, proceedings of

the 31st International Conference on Metallurgical Coatings and Thin Films. doi:<https://doi.org/10.1016/j.surfcoat.2004.08.036>.

- [4] J. Choury, Carbon-carbon materials for nozzles of solid propellant rocket motors, in: Proceedings of the 12<sup>th</sup> Propulsion Conference, 1976. doi: 10.2514/6.1976-609.  
URL <https://arc.aiaa.org/doi/abs/10.2514/6.1976-609>
- [5] G. L. Vignoles, J. Lachaud, Y. Aspa, Environmental effects: Ablation of C/C materials – surface dynamics and effective reactivity, in: N. P. Bansal, J. Lamon (Eds.), Ceramic Matrix Composites: Materials, Modeling and Technology, Wiley & American Ceramic Society, 2014, Ch. 12, pp. 353–384. doi:10.1002/9781118832998.ch12.
- [6] S. Awasthi, J. L. Wood, Carbon/Carbon Composite Materials for Aircraft Brakes, John Wiley & Sons, Ltd, 1988, Ch. 4, pp. 553–559. doi:<https://doi.org/10.1002/9780470310496.ch4>.
- [7] H. Hatta, R. Weiß, P. David, Carbon/carbons and their industrial applications, in: N. P. Bansal, J. Lamon (Eds.), Ceramic Matrix Composites : Materials, Modeling and Technology, John Wiley & Sons, Ltd, 2014, Ch. 5, pp. 85–146. doi:10.1002/9781118832998.ch5.
- [8] H. Duan, Z. Zhang, L. Li, W. Li, Effect of pyrocarbon interphase texture and thickness on tensile damage and fracture in T-700™ carbon fiber-reinforced silicon carbide minicomposites, J. Am. Ceram. Soc. 105 (3) (2022) 2171–2181. doi:10.1111/jace.18193.

- [9] J. Xu, L. Guo, H. Wang, K. Li, T. Wang, W. Li, Mechanical property and toughening mechanism of 2.5D C/C-SiC composites with high textured pyrocarbon interface, *Ceram. Int.* 47 (20) (2021) 29183–29190. doi: 10.1016/j.ceramint.2021.07.081.
- [10] B. Heidenreich, C/SiC and C/C-SiC composites, in: N. P. Bansal, J. Lamon (Eds.), *Ceramic Matrix Composites: Materials, Modeling and Technology*, John Wiley & Sons, Ltd, 2014, Ch. 6, pp. 147–216. doi:10.1002/9781118832998.ch6.
- [11] R. R. Naslain, The design of the fibre-matrix interfacial zone in ceramic matrix composites, *Compos. Part A* 29 (9) (1998) 1145–1155. doi: [https://doi.org/10.1016/S1359-835X\(97\)00128-0](https://doi.org/10.1016/S1359-835X(97)00128-0).
- [12] R. Naslain, Design, preparation and properties of non-oxide CMCs for application in engines and nuclear reactors: an overview, *Compos. Sci. Technol.* 64 (2004) 155–170. doi:[https://doi.org/10.1016/S0266-3538\(03\)00230-6](https://doi.org/10.1016/S0266-3538(03)00230-6).
- [13] E. Buet, J. Braun, C. Sauder, Influence of texture and thickness of pyrocarbon coatings as interphase on the mechanical behavior of specific 2.5D SiC/SiC composites reinforced with Hi-Nicalon S fibers, *Coatings* 12 (5) (2022). doi:10.3390/coatings12050573.
- [14] A. Koster, H. Matzner, D. Nicholsi, PBMR design for the future, *Nucl. Eng. Des.* 222 (2) (2003) 231–245, hTR-2002 1st international topical meeting on High Temperature reactor technology. doi:[https://doi.org/10.1016/S0029-5493\(03\)00029-3](https://doi.org/10.1016/S0029-5493(03)00029-3).

- [15] G. Paredes, R. Wang, P. Puech, G. Seine, J.-M. Leyssale, R. Arenal, A. Masseboeuf, F. Piazza, M. Monthieux, Texture, nanotexture, and structure of carbon nanotube-supported carbon cones, *ACS Nano* 16 (6) (2022) 9287–9296. doi:10.1021/acsnano.2c01825.
- [16] Y. Ohzawa, M. Mitani, J. Li, T. Nakajima, Structures and electrochemical properties of pyrolytic carbon films infiltrated from gas phase into electro-conductive substrates derived from wood, *Mater. Sci. Eng. B: Solid State Mater. Adv. Technol.* 113 (1) (2004) 91–98. doi:10.1016/j.mseb.2004.07.006.
- [17] N. He, O. Yildiz, Q. Pan, J. Zhu, X. Zhang, P. Bradford, W. Gao, Pyrolytic-carbon coating in carbon nanotube foams for better performance in supercapacitors, *J. Power Sources* 343 (2017) 492–501. doi:10.1016/j.jpowsour.2017.01.091.
- [18] J. Bokros, Carbon biomedical devices, *Carbon* 15 (6) (1977) 353–371. doi:https://doi.org/10.1016/0008-6223(77)90324-4.
- [19] R. H. Dauskardt, R. O. Ritchie, *Pyrolytic Carbon Coatings*, World Scientific, Singapore, 1993, Ch. 14, pp. 261–279. doi:10.1142/9789814317351\_0014.
- [20] G. Dong, K. Hüttinger, Consideration of reaction mechanisms leading to pyrolytic carbon of different textures, *Carbon* 40 (14) (2002) 2515–2528. doi:https://doi.org/10.1016/S0008-6223(02)00174-4.
- [21] V. De Pauw, B. Reznik, S. Kalhöfer, D. Gerthsen, Z. J. Hu, K. J.

- Hüttinger, Texture and nanostructure of pyrocarbon layers deposited on planar substrates in a hot-wall reactor, *Carbon* 41 (1) (2003) 71–7.
- [22] H. Le Poche, X. Bourrat, M.-A. Dourges, G. L. Vignoles, F. Langlais, Influence of the gas-phase maturation on the CVD/CVI process and the micro-texture of laminar pyrocarbon from propane, in: *Proceedings of high-temperature ceramic matrix composites*, Vol. 5, Springer, Seattle (WA, USA), 2005, pp. 293–300.
- [23] G. Vignoles, R. Pailler, F. Teyssandier, The control of interphases in carbon and ceramic matrix composites, Vol. 33 of *Ceramics Engineering and Science Proceedings*, John Wiley & Sons, Ltd, New York, 2012, pp. 11–23, ISBN: 978-1-118-53020-7. doi:10.1002/9781118217528.ch2.
- [24] M. Lieberman, H. Pierson, Effect of gas phase conditions on resultant matrix pyrocarbons in carbon/carbon composites, *Carbon* 12 (3) (1974) 233–241.
- [25] R. J. Diefendorf, W. E. Tokarsky, The relationships of structure to properties in graphite fibers, US Air Force report, AFML-TR-72-133 (1972).
- [26] P. Dupel, X. Bourrat, R. Pailler, Structure of pyrocarbon infiltrated by pulse-CVI, *Carbon* 33 (9) (1995) 1193–1204. doi:[https://doi.org/10.1016/0008-6223\(95\)00029-D](https://doi.org/10.1016/0008-6223(95)00029-D).
- [27] X. Bourrat, B. Trouvat, G. Limousin, G. Vignoles, F. Doux, Pyrocarbon anisotropy as measured by electron diffraction and polarized light, *J. Mater. Res.* 15 (1) (2000) 92–101.

- [28] B. Reznik, K. Hüttinger, On the terminology for pyrolytic carbon, *Carbon* 40 (4) (2002) 621–624.
- [29] X. Bourrat, F. Langlais, G. Chollon, G. L. Vignoles, Low temperature pyrocarbons: A review, *J. Braz. Chem. Soc.* 17 (6) (2006) 1090–5.
- [30] X. Bourrat, A. Fillion, R. Naslain, G. Chollon, M. Brendlé, Regenerative laminar pyrocarbon, *Carbon* 40 (15) (2002) 2391–45.
- [31] J.-M. Vallerot, X. Bourrat, A. Mouchon, G. Chollon, Quantitative structural and textural assessment of laminar pyrocarbons through raman spectroscopy, electron diffraction and few other techniques, *Carbon* 44 (9) (2006) 1833–1844. doi:<https://doi.org/10.1016/j.carbon.2005.12.029>.
- [32] E. López-Honorato, P. Meadows, P. Xiao, Fluidized bed chemical vapor deposition of pyrolytic carbon – I. Effect of deposition conditions on microstructure, *Carbon* 47 (2) (2009) 396–410. doi:<https://doi.org/10.1016/j.carbon.2008.10.023>.
- [33] P. Weisbecker, J.-M. Leyssale, H. E. Fischer, V. Honkimäki, M. Lalanne, G. L. Vignoles, Microstructure of pyrocarbons from pair distribution function analysis using neutron diffraction, *Carbon* 50 (2012) 1563–1573. doi:<https://doi.org/10.1016/j.carbon.2011.11.035>.
- [34] B. Farbos, P. Weisbecker, H. Fischer, J.-P. Da Costa, M. Lalanne, G. Chollon, C. Germain, G. Vignoles, J.-M. Leyssale, Nanoscale structure and texture of highly anisotropic pyrocarbons revisited with trans-

- mission electron microscopy, image processing, neutron diffraction and atomistic modeling, *Carbon* 80 (2014) 472–489. doi:<https://doi.org/10.1016/j.carbon.2014.08.087>.
- [35] B. Reznik, D. Gerthsen, K. Hüttinger, Micro- and nanostructure of the carbon matrix of infiltrated carbon fiber felts, *Carbon* 39 (2) (2001) 215–229. doi:[https://doi.org/10.1016/S0008-6223\(00\)00116-0](https://doi.org/10.1016/S0008-6223(00)00116-0).
- [36] B. Farbos, H. Freeman, T. Hardcastle, J.-P. Da Costa, R. Brydson, A. J. Scott, P. Weisbecker, C. Germain, G. L. Vignoles, J.-M. Leysale, A time-dependent atomistic reconstruction of severe irradiation damage and associated property changes in nuclear graphite, *Carbon* 120 (2017) 111–120. doi:<https://doi.org/10.1016/j.carbon.2017.05.009>.
- [37] J. Lavenac, F. Langlais, X. Bourrat, R. Naslain, Deposition process of laminar pyrocarbon from propane, *J. Phys. IV France* 11 (2001) Pr3–1013–Pr3–1021. doi:[10.1051/jp4:20013127](https://doi.org/10.1051/jp4:20013127).
- [38] M. Heggie, I. Suarez-Martinez, C. Davidson, G. Haffenden, Buckle, ruck and tuck: A proposed new model for the response of graphite to neutron irradiation, *J. Nucl. Mater.* 413 (3) (2011) 150–155. doi:<https://doi.org/10.1016/j.jnucmat.2011.04.015>.
- [39] B. März, K. Jolley, T. J. Marrow, Z. Zhou, M. Heggie, R. Smith, H. Wu, Mesoscopic structure features in synthetic graphite, *Mater. Des.* 142 (2018) 268–278. doi:<https://doi.org/10.1016/j.matdes.2018.01.038>.

- [40] M. Chen, B. Wu, L. Zhou, Y. Zhu, H. Wu, Micromechanical properties of pyrolytic carbon with interlayer crosslink, *Carbon* 159 (2020) 549–560. doi:<https://doi.org/10.1016/j.carbon.2019.12.096>.
- [41] S. K. Jain, R. J.-M. Pellenq, J. P. Pikunic, K. E. Gubbins, Molecular modeling of porous carbons using the hybrid reverse monte carlo method, *Langmuir* 22 (24) (2006) 9942–9948. doi:10.1021/la053402z.
- [42] J. Palmer, A. Llobet, S.-H. Yeon, J. Fischer, Y. Shi, Y. Gogotsi, K. Gubbins, Modeling the structural evolution of carbide-derived carbons using quenched molecular dynamics, *Carbon* 48 (4) (2010) 1116–1123. doi:<https://doi.org/10.1016/j.carbon.2009.11.033>.
- [43] J.-M. Leyssale, J.-P. Da Costa, C. Germain, P. Weisbecker, G. L. Vignoles, An image-guided atomistic reconstruction of pyrolytic carbons, *Appl. Phys. Lett.* 95 (2009) 231912. doi:10.1063/1.3272949.
- [44] J.-M. Leyssale, J.-P. Da Costa, C. Germain, P. Weisbecker, G. L. Vignoles, Structural features of pyrocarbon atomistic models constructed from transmission electron microscopy images, *Carbon* 50 (12) (2012) 4388–4400.
- [45] P. Y. Huang, C. S. Ruiz-Vargas, A. M. Van Der Zande, W. S. Whitney, M. P. Levendorf, J. W. Kevek, S. Garg, J. S. Alden, C. J. Hustedt, Y. Zhu, et al., Grains and grain boundaries in single-layer graphene atomic patchwork quilts, *Nature* 469 (7330) (2011) 389–392.
- [46] A. Gamboa, B. Farbos, P. Aurel, G. L. Vignoles, J.-M. Leyssale, Mech-



- anism of strength reduction along the graphenization pathway, *Sci. Adv.* 1 (10) (2015) e1501009.
- [47] B. Farbos, J.-P. Da Costa, G. Vignoles, J.-M. Leyssale, Nanoscale elasticity of highly anisotropic pyrocarbons, *Carbon* 94 (2015) 285–294. doi:<https://doi.org/10.1016/j.carbon.2015.06.060>.
- [48] J.-M. Leyssale, G. Couégnat, S. Jouannigot, G. L. Vignoles, Mechanisms of elastic softening in highly anisotropic carbons under in-plane compression/indentation, *Carbon* 197 (2022) 425–434. doi:<https://doi.org/10.1016/j.carbon.2022.06.063>.
- [49] E. Jin, S. Du, H. He, The evolution of mechanical and structural properties at the fiber/matrix interphase of SiC/SiC composites, *Comput. Mater. Sci.* 104 (2015) 84–91. doi:<https://doi.org/10.1016/j.commatsci.2015.03.045>.
- [50] P. Puech, A. Dabrowska, N. Ratel-Ramond, G. L. Vignoles, M. Monthieux, New insight on carbonisation and graphitisation mechanisms as obtained from a bottom-up analytical approach of X-ray diffraction patterns, *Carbon* 147 (2019) 602–611. doi:<https://doi.org/10.1016/j.carbon.2019.03.013>.
- [51] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, S. B. Sinnott, A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons, *J. Phys.: Condens. Matter.* 14 (4) (2002) 783–802.

- [52] W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson, A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters, *J. Chem. Phys.* 76 (1) (1982) 637–649. doi:10.1063/1.442716.
- [53] H. C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature, *J. Chem. Phys.* 72 (1980) 2384.
- [54] J. H. Los, L. M. Ghiringhelli, E. J. Meijer, A. Fasolino, Improved long-range reactive bond-order potential for carbon. i. construction, *Phys. Rev. B* 72 (2005) 214102.
- [55] L. Soulard, Molecular Dynamics Study of the Micro-spallation, *Eur. Phys. J. D* 50 (3) (2008).
- [56] L. M. Ghiringhelli, J. H. Los, A. Fasolino, E. J. Meijer, Improved long-range reactive bond-order potential for carbon. II. Molecular simulation of liquid carbon, *Phys. Rev. B* 72 (2005) 214103. doi:10.1103/PhysRevB.72.214103.
- [57] E. Bourasseau, N. Pineau, J.-B. Maillet, V. Dubois, Calculation of the shock temperature of moderately porous graphites, *Carbon* 103 (2016) 464–472. doi:https://doi.org/10.1016/j.carbon.2016.03.048.
- [58] G. Chevrot, E. Bourasseau, N. Pineau, J.-B. Maillet, Molecular dynamics simulations of nanocarbons at high pressure and temperature, *Car-*

- bon 47 (15) (2009) 3392–3402. doi:<https://doi.org/10.1016/j.carbon.2009.06.061>.
- [59] R. Perriot, Y. Lin, V. V. Zhakhovsky, N. Pineau, J. H. Los, J.-B. Maillet, L. Soulard, C. T. White, I. I. Oleynik, Shock compression of diamond: Molecular dynamics simulations using different interatomic potentials, *AIP Conf. Proc.* 1426 (1) (2012) 1175–1178. doi:[10.1063/1.3686489](https://doi.org/10.1063/1.3686489).  
URL <https://aip.scitation.org/doi/abs/10.1063/1.3686489>
- [60] F. Polewczyk, J.-M. Leyssale, P. Lafourcade, Temperature-dependent elasticity of single crystalline graphite, *Comput. Mater. Sci.* 220 (2023) 112045. doi:[10.1016/j.commatsci.2023.112045](https://doi.org/10.1016/j.commatsci.2023.112045).
- [61] S. Nosé, A unified formulation of the constant temperature molecular dynamics methods, *J. Chem. Phys.* 81 (1) (1984) 511–519. doi:[10.1063/1.447334](https://doi.org/10.1063/1.447334).
- [62] W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A* 31 (3) (1985) 1695–1697. doi:[10.1103/PhysRevA.31.1695](https://doi.org/10.1103/PhysRevA.31.1695).
- [63] F. Vuković, J.-M. Leyssale, P. Aurel, N. A. Marks, Evolution of threshold displacement energy in irradiated graphite, *Phys. Rev. Appl.* 10 (2018) 064040. doi:[10.1103/PhysRevApplied.10.064040](https://doi.org/10.1103/PhysRevApplied.10.064040).
- [64] D. Franzblau, Computation of ring statistics for network models of solids, *Phys. Rev. B* 44 (1991) 4925–4930. doi:[10.1103/PhysRevB.44.4925](https://doi.org/10.1103/PhysRevB.44.4925).

- [65] J. Barthel, Dr. Probe: A software for high-resolution STEM image simulation, *Ultramicroscopy* 193 (2018) 1–11. doi:<https://doi.org/10.1016/j.ultramic.2018.06.003>.
- [66] J.-P. Da Costa, C. Germain, P. Baylou, Level curve tracking algorithm for textural feature extraction, in: *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Vol. 3, IEEE, 2000, pp. 909–912.
- [67] J.-P. Da Costa, P. Weisbecker, B. Farbos, J.-M. Leysale, G. Vignoles, C. Germain, Investigating carbon materials nanostructure using image orientation statistics, *Carbon* 84 (2015) 160–173.
- [68] F. L. Pouliquen, J.-P. D. Costa, C. Germain, P. Baylou, A new adaptive framework for unbiased orientation estimation, *Pattern Recognition* 38 (11) (2005) 2032–2046.
- [69] L. Breiman, Random forests, *Machine Learning* 45 (2001) 5–32.
- [70] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, Édouard Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* 12 (85) (2011) 2825–2830.