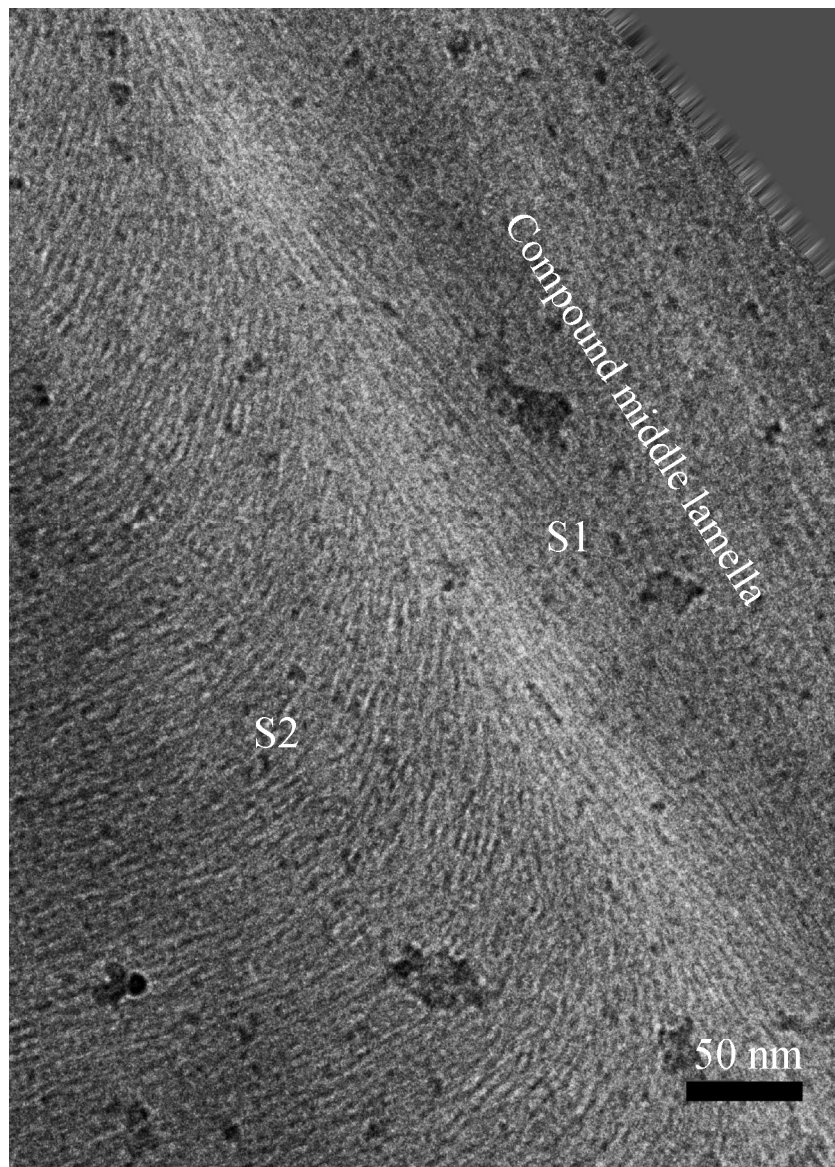Supporting Information for

Cellulose Elementary Fibrils Assemble into Helical Bundles in S_1 Layer of Spruce Tracheid Wall

Mehedi Reza, Carlo Bertinetto, Janne Ruokolainen, Tapani Vuorinen

Supplementary Figure 1. A TEM image of spruce transverse section at 0 tilt shows dif - ferent cell wall layers and location of tomography.

Description of the fitting algorithm

The algorithm used in this work to fit the geometry of the EFs is an adaptation of the one presented by Ciesielski *et al.*[1]. The selected tomographic subvolume is initially normalized and rotated so that one of the axes (here chosen to be *z*) is roughly aligned to the filamentous structural patterns that can be distinguished by visual inspection. The tomographic density is then fitted by parametric space curves of the form:

$$\mathbf{h}(t) \equiv \begin{cases} x = c_1 \sin(\omega t + \varphi) + P_x(t) \\ y = c_2 \cos(\omega t + \varphi) + P_y(t) \\ z = t \end{cases} \tag{1}$$

Where $\mathbf{h}(t)$ is an *xyz* triple that defines the point of a space curve $\mathbf{h}$ at the parameter value of *t*. The trigonometric operators allow the space curve to adopt a helical geometry with *x* and *y* amplitudes (c_1, c_2), angular frequency (ω), and phase shift (φ). The terms $P_x(t)$ and $P_y(t)$ are polynomials, which confer the curve flexibility to deviate from a straight helix. The algorithm places space curves in the regions of the subvolume with the highest relative density, by varying the parameters of equation (1) while optimizing an appropriate cost function.

The mass enclosed within a chosen distance *L* on the *x* and *y* dimensions from the curve $\mathbf{h}(t)$ is given by:

$$M_{\mathbf{h}} = \int_{\mathbf{h}_x-L}^{\mathbf{h}_x+L} \int_{\mathbf{h}_y-L}^{\mathbf{h}_y+L} \int_{t_0}^{t_f} \rho\left(\mathbf{h}(t)\right) dx\,dy\,dt \tag{2}$$

where t_0 and t_f are the lower and upper bounds of the parametric variable domain, $\mathbf{h}_x$ and $\mathbf{h}_y$ are the *x* and *y* components of the ordered triplet generated by $\mathbf{h}(t)$ and $\rho(\mathbf{h}(t))$ the density of the voxel at the location $\mathbf{h}(t)$. For a discrete tomographic volume, this expression may be approximated as:

$$M_{\mathbf{h}} \approx \sum_{i=\mathbf{h}_x-L}^{\mathbf{h}_x+L} \sum_{j=\mathbf{h}_y-L}^{\mathbf{h}_y+L} \sum_{k=t_0}^{t_f} \rho(x_i, y_j, t_k) \cdot L^3 \tag{3}$$

In order to further attract the space curves to the densest regions of the tomogram, the L^3 term was neglected from equation (3) and the tomographic density was weighted by the inverse square of its distance from the space curve, as:

$$C \approx \sum_{i=\mathbf{h}_x-L}^{\mathbf{h}_x+L} \sum_{j=\mathbf{h}_y-L}^{\mathbf{h}_y+L} \sum_{k=t_0}^{t_f} \frac{\rho(x_i, y_j, t_k)}{(\mathbf{h}_x - x_i)^2 + (\mathbf{h}_y - y_j)^2} \tag{4}$$

The cost function used to optimize the parameters of $\mathbf{h}(t)$ consisted thus of the term C, sometimes with the addition of a case-dependent penalty term that avoided solutions known to be unrealistic (see below).

The optimization was carried out employing two different methods in sequence: i) Particle Swarm Optimization (PSO)[2] and ii) Nelder-Mead simplex method[3]. The former is one of the so-called "global optimization" methods, which performs a wide exploration in the variable space to find (or at least attempt at finding) the global optimum of the cost function. Its result depends very little on the starting value and is therefore more suitable for the initial search. However, since it tends to explore with such freedom that it can easily yield extreme and unphysical solutions, its search space is limited to the phase shift φ and the constant terms in P_x and P_y (which determine the location of the helix for $t = 0$), with the remaining parameters manually set to reasonable values. For the S_1 layer containing fibrous structures, these values were: $c_1 = c_2 = 4.5$ nm, $\omega = 1.12$ deg nm^{-1} and all polynomial terms set to zero. For the subvolumes containing bundles, they were: $c_1 = c_2 = 13.4$ nm, $\omega = 1.12$ deg nm^{-1} and all polynomial terms set to zero.

The result from PSO was used as starting point for the second part of the optimization using the simplex method. The latter is more efficient at converging to a nearby local minimum and was thus employed to optimize all the parameters of the space curve. The cost function's penalty terms mentioned above were added in this phase: for bundles we penalized excessively high ω (which could generate a single curly path fitting the whole bundle).

Once an optimal space curve was found, the tomographic density within 6 nm of the curve was removed from the dataset, a new curve was added and the whole process was repeated, until a designated fraction of the density was removed or a user-defined maximum number of curves were placed. This procedure is illustrated also in the Supporting Videos V1 and V2.

Calculation of Nearest-Neighbor distances

For each space curve fitted to a subvolume, the average distance between its position and that of the other curves was approximated by equation (5), and the minimum of these was reported as the nearest-neighbor spacing. For a space curve of the form given by equation (1), the average Euclidian distance average between the ith and j^{th} space curves is approximated over the parametric interval $[t_0, t_f]$ as:

$$\langle D_{i,j} \rangle \approx \frac{\mathbf{1}}{n} \sum_{t_0}^{t_f} \sqrt{(\mathbf{h}_i(t) - \mathbf{h}_j(t))^2} \tag{5}$$

Where $\langle D_{i,j} \rangle$ is the average Euclidian distance, $\mathbf{h}_i(t)$ and $\mathbf{h}_j(t)$ are the *xyz* triples of the ith and j^{th} space curves, respectively, at the parametric value of t, and n is the number of discrete values in the interval $[t_0, t_f]$. Large deviations in the axial direction of the space curve from the parametric axis can result in an overestimate of the spacing calculated by equation (12), thus the nearest-neighbor spacing values represent an approximation of the upper limit of the distance between the curves.

References

1. Ciesielski, P. N.; Matthews, J. F.; Tucker, M. P.; Beckham, G. T.; Crowley, M. F.; Himmel, M. E.; Donohoe, B. S. *ACS Nano* **2013**, 7, 8011-8019.

2. Kennedy, J.; Eberhart, R. Particle Swarm Optimization. *Proceedings of IEEE International Conference on Neural Networks* 1**995**, IV, 1942–1948.

3. Lagarias, J.C.; Reeds, J. A.; Wright, M. H.; Wright, P. E. *SIAM Journal of Optimization* **1998**, 9, 112-147.
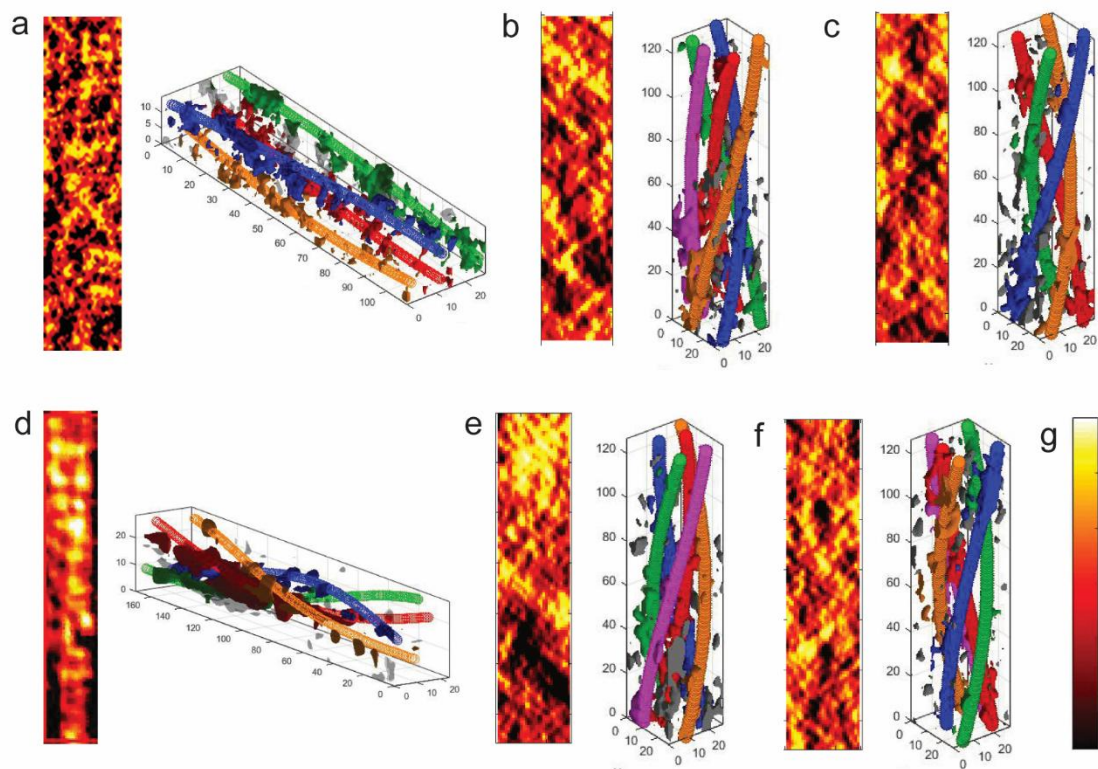
Supplementary Figure 2. Tomographic slices through the subvolumes and the resultant fitted space curves show fibrous structure (a) and helical bundles (b-f); plot units are nm. Color bar in *g* represent tomographic density for all tomographic slices.