Visualization of water drying in porous materials by X-ray phase contrast imaging

Supplementary Materials

F. Yang*,♣, M. Griffa*, A. Bonnin†,♦ , R. Mokso†, C. Di Bella*,♣, B. Münch*, R. Kaufmann*, P. Lura*,♣

*Empa, Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, 8600, Switzerland

♣Institute for Building Materials (IfB), Swiss Federal Institute of Technology Zurich (ETHZ), Zürich, 8093, Switzerland

†Swiss Light Source, Paul Scherrer Institute, Villigen, 5232, Switzerland

♦Center for Biomedical Imaging, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, 1015, Switzerland

Key words: X-ray tomographic microscopy, X-ray phase contrast imaging, image analysis, porous materials, liquid transport, water

S1. Theoretical calculations of the X-ray complex index of refraction for water and $CaCO_3$

We report in this Section the plots of the $\frac{\delta}{\beta}$ ratio (Fig. S1) and of the absolute values of δ (Fig. S2(a)) and β (Fig. S2(b)) as a function of photon energy *E*, in the range 1 - 100 keV, for both $CaCO_3$ (solid line) and water (dotted line). The complex index of refraction is defined as $n = 1 - \delta + i \cdot \beta$.The plots were produced by using the NIST FFAST database of photon cross-sections and atomic form factors [FFAST,Chantler1995, Chantler2000].

In Figure S1 one can notice that $\frac{\delta}{\beta}$ varies between 10^2 and 10^4 supporting the idea proposed in Section I of the article that a partial or complete filling/emptying of pores at length scales above the spatial resolution should be more easily resolvable in a δ-based image than in a β-based image. Figure S2 is used to support the second main idea at the basis of using XPCI for visualization of pure water transport in porous materials. If one considers water volumes at the voxel or sub-voxel scale, moving from one voxel to another, one should expect to see a more pronounced voxel value change in an image based upon δ compared with an image based upon β, because the difference in β

between water and the porous substrate is one to two orders of magnitude larger than the same type of difference for δ.

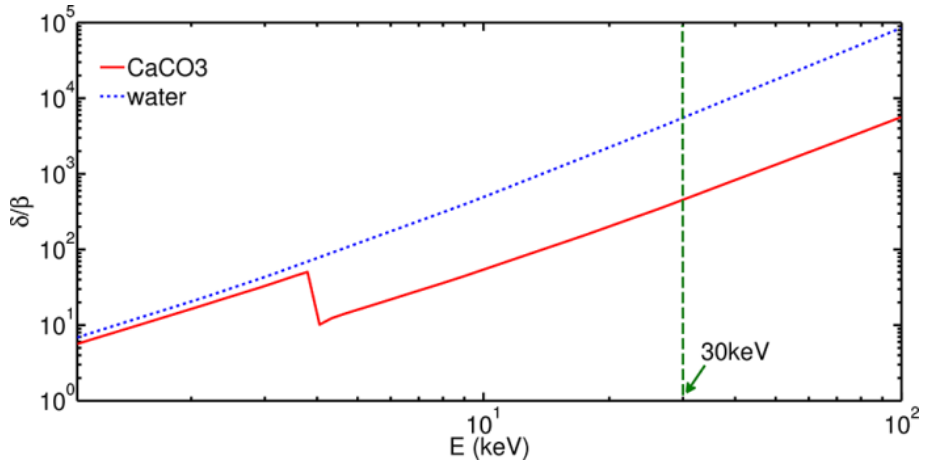

Figure S1: plot of the ratio between the decrement δ of the real part of the index of refraction and its imaginary part β as a function of the photon energy E. The solid line refers to $CaCO_3$ as representative main chemical component of the Globigerina limestone solid phase, while the dotted line refers to water. The large drop for $CaCO_3$ at about 4 keV is due to the large and sudden increase of β in correspondence of that energy (see also Fig. S2(b), solid line). Such a β increase is due to K absorption edge of Ca, occurring at 4.0381 keV according to theoretical calculations based upon the NIST FFAST database of photon cross-sections and atomic form factors [FFAST, Chantler1995, Chantler2000]. The vertical dashed line indicates the photon energy at which the X-ray tomographic microscopy measurements were performed.

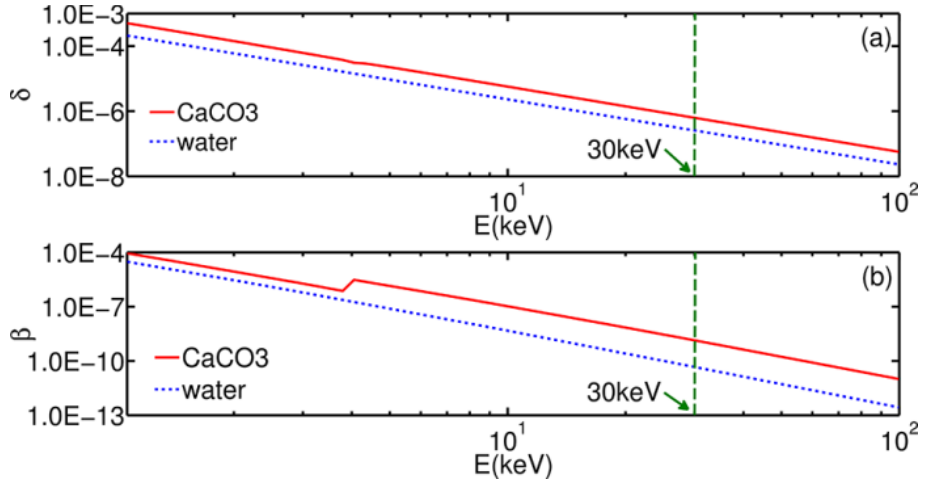

Figure S2: individual plots of the decrement δ of the real part (inset (a)) and the imaginary part β (inset (b) of the index of refraction n for the same photon energy E as in Fig. S1. The solid line still refers to $CaCO_3$ while the dotted line to water. The vertical dashed line indicates the photon energy at which the X-ray tomographic microscopy measurements were performed.

S2. Segmentation of pore voxels

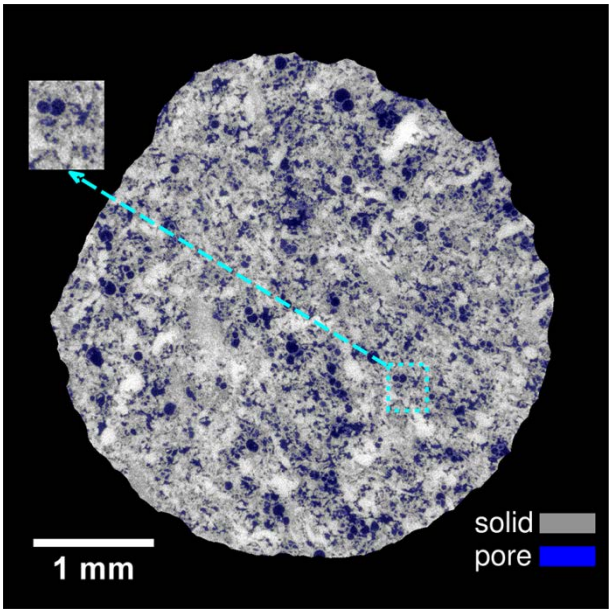

Figure S3: overlapping of the 1st slice of the *AS2* 3D image (grey tones) and of the corresponding slice of the *PORE* mask (semi-transparent blue color), the latter obtained by *k*-means clustering of the *AS1* and *AS2* 3D images with a 1D feature space based on the voxel value only. The inset shows a zoom into a smaller ROI, for providing an example of the segmentation results quality.

S3. Voxel value distributions for the XACI and XPCI 3D images

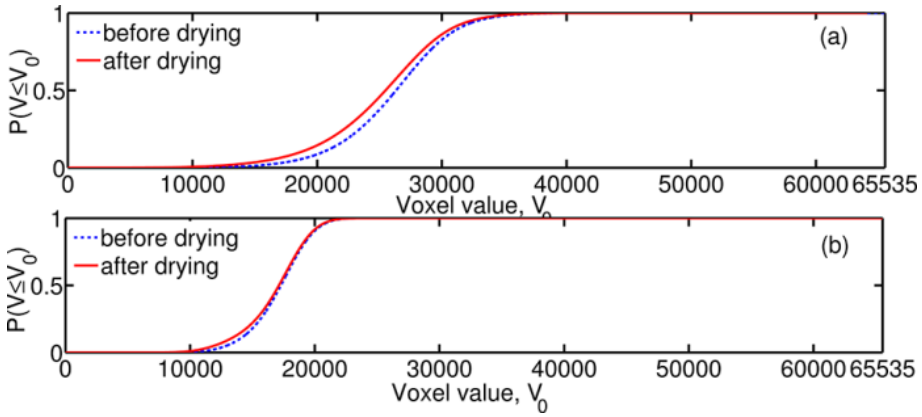

Figure S4: voxel value, V_0, cumulative distribution functions (CDFs) for the X-ray phase ((a)) and attenuation ((b)) contrast datasets (3D images), before (dashed curve) and after (solid curve) drying. $P(V \leq V_0)$ means the probability for the event $V \leq V_0$, where V indicates generically the voxel value as a random variable. For each

3D image, the statistical ensemble is made of all the voxels belonging to the sample region, i.e., not belonging to the region surrounding the sample (3D binary image *R*).

Each 3D image, of each type (phase or attenuation contrast), was originally voxel-wise subtracted by the average voxel value inside *R*, in order to remove the effect of different offsets due to temporal fluctuations in the time-lapse measurements. Given an image type, the minimum of the minima, $V_{min}^{1,2}$, and maximum of the maxima, $V_{max}^{1,2}$, voxel values was then used as common dynamic range for both images. That voxel value range was then linearly rescaled to the integer interval [0;65535] (16 bit unsigned integer encoding). With such choice of dynamic range for each image, of each type, the two time-lapse images of a given type can be compared with each other without any bias and the comparison can be performed similarly independently of the image type.

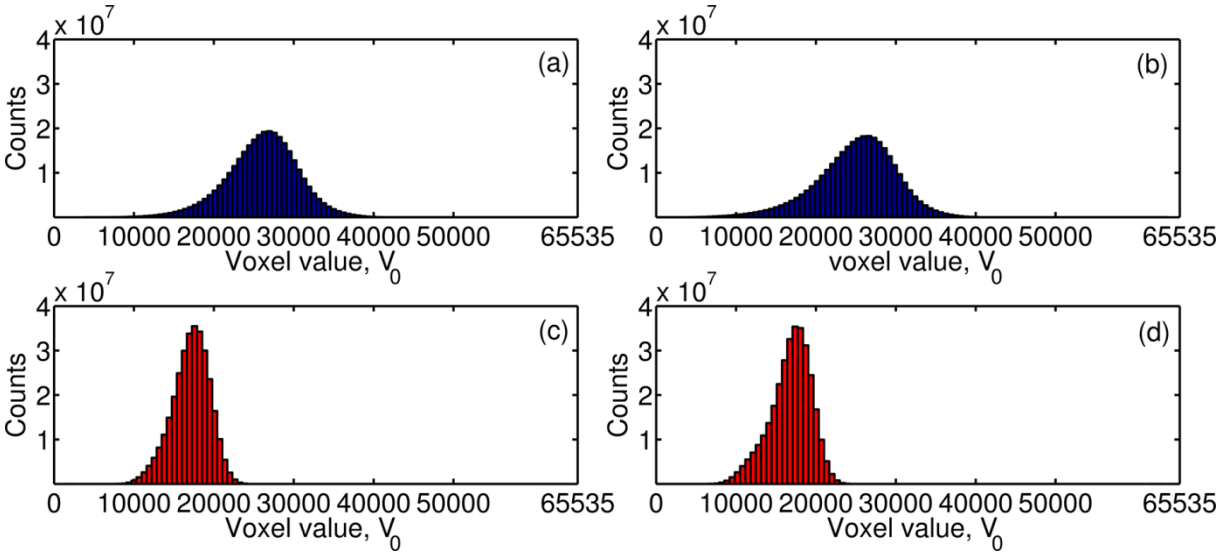

Figure S5: voxel value histograms the CDFs of Fig. S4 are derived from. (a) and (b): XPCI datasets, before and after drying, respectively. (c) and (d): XACI datasets, before and after drying, respectively.

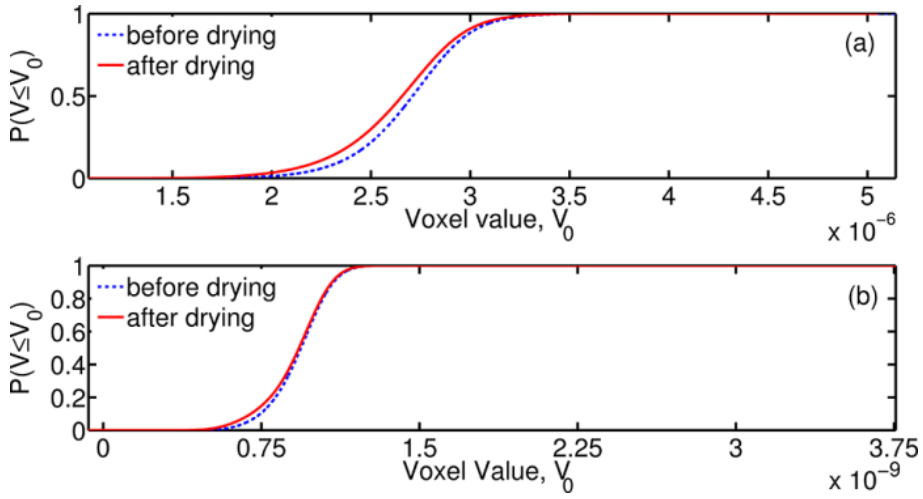

Figure S6: voxel value cumulative distribution functions (CDFs) for the two time-lapse phase (a) and attenuation (b) contrast 3D images, with the dashed line indicating the image before the drying while the solid line after drying. The only difference compared with the CDFs in Fig. S4 consists in the fact that the images before the remapping to the integer value interval [0;65535] were used.

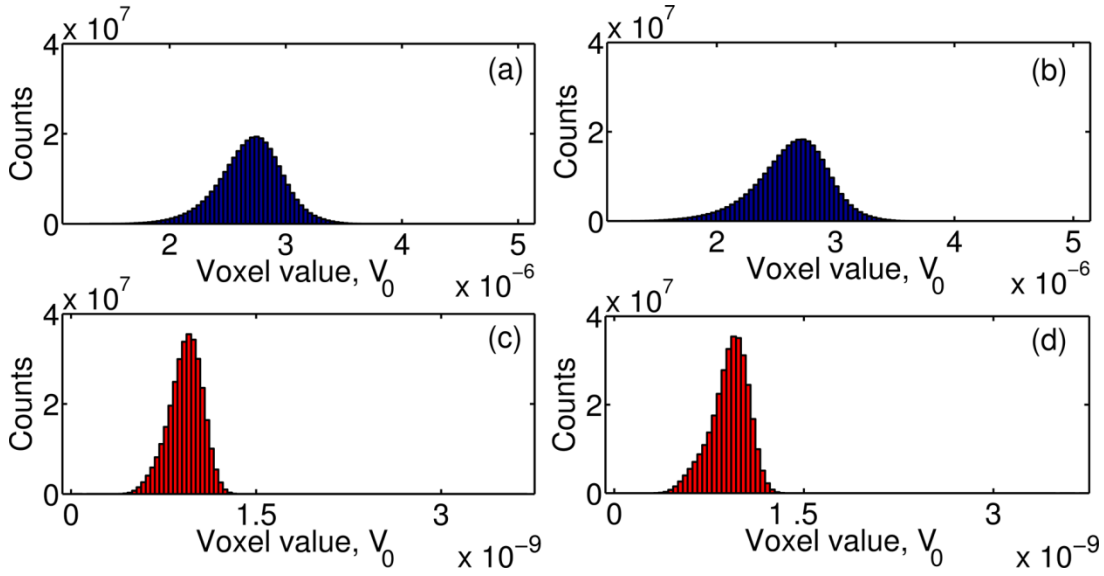

Figure S7: voxel value histograms the CDFs of Fig. S6 are derived from. (a) and (b): XPCI datasets, before and after drying, respectively. (c) and (d): XACI datasets, before and after drying, respectively.

S4. Best fit of the voxel value histogram for the *PSDIFF* dataset and voxels belonging to the *PORE* mask. 2- *vs* 3-Gaussian mixture model.

Figure S8 shows the best fit of the voxel value histogram for the time-differential XPCI dataset *PSDIFF*, considering only voxels belonging to the segmented pore regions, i.e., to the *PORE* 3D binary image, using a 2-Gaussian mixture model instead of a 3-Gaussian one as reported within the article and shown in its Fig. 4. The best fit was obtained by using the same expectation maximization (EM) algorithm as used for the results in Fig. 4.

It is immediately clear that the 2-Gaussian mixture best fit model performs worse than the 3-Gaussian one, just by qualitatively comparison of Fig. 4 within the article and Fig. S8. The 2-Gaussian model is not able, for example, to determine the plateau at about voxel value = 20000 due to the overlapping of two Gaussians.

A more quantitative comparison between the results of the two best fits was obtained by calculating for each best fit 3 different goodness-of-fit metrics:

1. $NRMSE = 1 - \|\vec{X}_{fit} - \vec{X}_{ref}\| / \|\vec{X}_{fit} - \langle\vec{X}_{ref}\rangle\|$ (normalized root-mean squared error)
2. $NMSE = 1 - \|\vec{X}_{fit} - \vec{X}_{ref}\|^2 / \|\vec{X}_{fit} - \langle\vec{X}_{ref}\rangle\|^2$ (normalized mean squared error)
3. $R = C(\vec{X}_{fit}, \vec{X}_{ref}) / (\|\vec{X}_{fit}\| \times \|\vec{X}_{ref}\|)$ (linear correlation coefficient).

Each $\vec{X}_i$, i = *fit* or *ref*, is a column vector containing the histogram values for each bin. $\vec{X}_{fit}$ is the vector from the best fit, $\vec{X}_{ref}$ the actual data from the *PSDIFF* dataset. $\|\vec{X}\|$ indicates the Euclidean norm ("magnitude") of a vector $\vec{X}$. $\langle\vec{X}_i\rangle$ indicates the average value of the components of $\vec{X}_i$. $\vec{X}_i - \langle\vec{X}_i\rangle$ should be interpreted as an element-wise operation, thus still leading to a column vector. $C(\vec{X}_{fit}, \vec{X}_{ref})$ indicates the covariance between the two vectors, defined as $C(\vec{X}_{fit}, \vec{X}_{ref}) \equiv$

$\langle(\vec{X}_{fit} - \langle\vec{X}_{fit}\rangle) \cdot (\vec{X}_{ref} - \langle\vec{X}_{ref}\rangle)\rangle$. · indicates the scalar product between two vectors while × the product between two scalar variables.

NRMSE and NMSE vary from negative infinity (poor fitting) to 1 (perfect fitting). The closer *R* is to 1, the higher the correlation between the original data and the fitting results.

NRMSE and NMSE value were calculated with the *MATLAB*^{TM} built-in function goodnessOfFit(…)., while *R* was calculated with the *MATLAB*^{TM} built-in function corrcoef(…).

Table SI reports the values of each goodness-of-fit metric for each best fit, confirming that the 3-Gaussian mixture best fit performed better than the 2-Gaussian one.

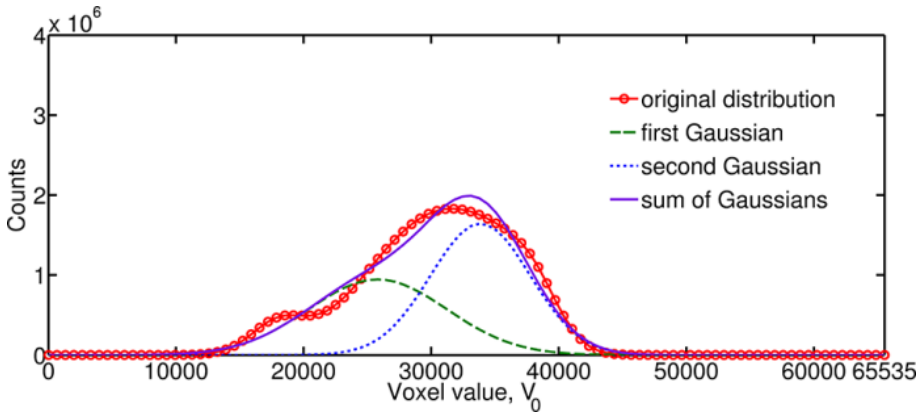

Figure S8: voxel value histogram (open circles) from the *PSDIFF* (time-differential XPCI) dataset considering only voxels belonging to the *PORE* 3D binary image, i.e., segmented pore regions. 2-Gaussian mixture best fit using the expectation maximization (EM) method. Each Gaussian function of the mixture (dash and dot-dashed lines) is plotted along with their sum (solid line).

Table SI. Values for three different goodness-of-fit metrics applied to the best fit of the voxel value histogram of the *PSDIFF* dataset, considering only voxels belonging to the pore region, and using two best fit models, a 2-Gaussian mixture and a 3-Gaussian one.

Goodness-of-fit metric type	**Value**	
	3-Gaussian mixture	**2-Gaussian mixture**
NRMSE	0.9591	0.8690
NMSE	0.9983	0.9828
R	0.9992	0.9914

References

[FFAST] http://www.nist.gov/pml/data/ffast/index.cfm

[Chantler1995] C.T. Chantler, J. Phys. Chem. Ref. Data 24, 71-643 (1995)

[Chantler2000] C.T. Chantler, J. Phys. Chem. Ref. Data 29, 597-1048 (2000)