Terahertz time-domain spectroscopy as a novel tool for crystallographic analysis in cellulose: the potentiality of being a new standard for evaluating crystallinity

Han Wang†, Satoru Tsuchikawa†, Tetsuya Inagaki†*

†Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Aichi, 464-8601, Japan.

*E-mail: inatetsu@agr.nagoya-u.ac.jp

ABSTRACT

Given that terahertz (THz) radiation responds to intermolecular forces such as hydrogen bonds, THz time-domain spectroscopy (THz-TDS) has expanded possibilities in cellulose research. In this study, THz-TDS was used to investigate the crystallinity of three types of cellulose-based materials. Microcrystalline cellulose (MCC) and wood were ball milled at different times, and pseudo-wood was a mixture of MCC and lignin of different mass fractions. All the samples showed peaks at 3.04 THz in the THz mass absorption coefficient spectra. Further, the spectra from 2.79 THz to 3.32 THz were cut out and detrended by subtraction from a baseline. The integrated intensity of the detrended spectra showed a correlation with the mass fraction of lignin of the pseudo-wood samples, and ball milling time of the MCC and wood samples. The correlation was similar with the crystallinity index calculated from X-ray powder diffraction. Moreover, the original wood sample without ball milling had an integrated intensity that was about 30% that of the original MCC sample, matching with the cellulose concentration of the wood (about 30% to 40%). We normalized the integrated intensity of 2.79 THz to 3.32 THz into 1 to 0 by a min-max algorithm and proposed a new “index” for evaluating crystallinity.

Keywords: THz-TDS, XRD, Cellulose, Crystallinity, MCC, Wood

Declarations.

Funding: JSPS KAKENHI Grant Number 16H02559
Conflicts of interest/Competing interests: None
Availability of data and material: None
**Code availability:** Matlab 2018b

**Authors' contributions:** Han Wang and Tetsuya Inagaki conceived and designed the experiments. Han Wang performed the experiments, analyzed the data and wrote the paper, Tetsuya Inagaki and Satoru Tsuchikawa gave final approval of the manuscript.
Introduction

Cellulose is a macromolecular polysaccharide composed of glucose units, where the $\beta (1 \rightarrow 4)$-D-glucose units are linked to a linear chain by hydrogen bonds (Updegraff 1969) and the cellulose chains further are linked to a sheet structure. The abundant intra- and inter-chain hydrogen bonds in cellulose make its structure relatively stable (Nishiyama et al. 2002; Langan et al. 2005). The sheets of cellulose are spontaneously and orderly packed by hydrophobic forces such as van der Waals forces, and further aggregated into larger microfibrils (Nishiyama 2009). The degree of order of cellulose crystals can be evaluated in general by using the crystallinity index (CrI), which is an important property of cellulose, and measured by X-ray powder diffraction (XRD). Cellulose is abundantly contained in the cell walls of plants and algae and plays a role in supporting cell structure. Wood is a typical natural cellulose-based material with a wide application and long history. The CrI of cellulose reflects the physical and chemical properties of wood to a certain extent. The tensile strength, Young’s modulus, density, thermal stability, and dimensional stability of wood all increase as the cellulose CrI increases (Density et al. 1975; Thygesen et al. 2005; Poletto et al. 2012), whereas the swelling and reaction speed of enzymatic conversion processes decrease (Jörgensen et al. 1950; Hall et al. 2010). This condition means that crystalline cellulose is one of the main rate-limiting barriers for the efficient transformation and utilization of plant biomass (Himmel et al. 2007). Therefore, the crystallinity of cellulose must be investigated.

In addition to XRD, the methods for investigating the crystallinity of cellulose include solid-state carbon 13 nuclear magnetic resonance (13C NMR), cross polarization/magic angle spinning (CP/MAS), infrared spectroscopy (IR), near-infrared spectroscopy (NIR), and Raman spectroscopy (Teeäär et al. 1987; Kataoka and Kondo 1998; Inagaki et al. 2010; Kim et al. 2013). Among these methods, XRD is the most widely used and considered as a standard method for evaluating crystallinity in most cases given that its pattern reflects the arrangement of atoms, and it has a rigorous theoretical calculation formula. However, the CrI obtained by XRD will be significantly different depending on the selected calculation method. The most used methods for determining the CrI from XRD are the Segal peak height method (Segal et al. 1959), amorphous subtraction method (Thygesen et al. 2005), and the peak deconvolution method (Hermans and Weidinger 1948).
minimum diffraction intensity between the (200) and (110) crystalline diffraction peaks usually appears at about 18°, where was attributed to amorphous by Segal et al., however, the diffraction intensity at this position overlaps the crystalline intensity peaks; thus, the reliability of the first two methods is questionable (French and Santiago Cintrón 2013; French 2014; Ling et al. 2019). Therefore, the peak deconvolution method was most discussed in recent studies. For the deconvolution process, the most used peak profiles are Gaussian, Lorentzian, Voigt, and pseudo-Voigt (Hindeleh and Johnson 1972; de Keijser et al. 1983; Wada et al. 1997). However, the accurate expression of amorphous curves using only such functions is difficult (del Cerro et al. 2020). Yao et al. reported an amorphous curve fitted using Fourier series modeling, which gave a result that was consistent with the measured amorphous cellulose pattern (Yao et al. 2020).

The terahertz (THz) band has received considerable attention; this band is located in the far-infrared region where the boundary area is between light and radio waves, with a frequency range of 0.1 THz to 10 THz, corresponding to wavelengths of 3 mm to 0.03 mm. The earliest commercial employed application of this band was THz time-domain spectroscopy (THz-TDS). The THz band responds to the intermolecular vibration and rotation of numerous biological macromolecules; thus, the THz band has been used in biology (Plusquellic et al. 2007; Markelz 2008), biomaterials (Inagaki et al. 2014a, b; Wang et al. 2019), hydrodynamics (Braly et al. 2000), and medicine studies (Xie et al. 2014). Given the response of THz radiation to phonons in the crystal lattice, THz has been used in several cases to study crystalline cellulose; in one case, the CrI of microcrystalline cellulose (MCC) was determined by THz-TDS combined with a partial least squares (PLS) model (Vieira and Pasquini 2014), and the capability to distinguish between cellulose crystalline forms of Iα and Iβ was detected (Wang et al. 2020).

This study is a follow-up of a previous research (Wang et al. 2020). We continuously explored the possibility of THz-TDS in cellulose research. THz-TDS was used to investigate the cellulose-based materials, including MCC and wood, both ball milled at different times, and pseudo-wood which was a mixture of MCC and lignin. All materials showed a peak at 3.04 THz of the mass absorption coefficient spectra, and the peak intensity decreased as the crystallinity decreased by ball milling. Further, the mass absorption coefficient spectra of 2.79 THz to 3.32 THz were cut out and
detrended by subtraction of a baseline. The integrated intensity of the detrended spectra showed a correlation with the mass fraction of lignin of the pseudo-wood, and ball milling time of MCC and wood. Different from determination of the CrI of cellulose by deconvolution of the XRD pattern, the processing of THz mass absorption coefficient spectra was relatively simple. Moreover, the integrated intensity of 2.79 THz to 3.32 THz was normalized to 0–1 by a min-max algorithm, and it can be regarded as a new “index” for evaluation of the cellulose crystallinity.

**Experimental**

**Sample preparation**

To obtain cellulose and wood powder with different CrI, the MCC powder (EMD Millipore 1.0231.0500) were ball milled with a benchtop ball mill (AV-2, Asahi Rika Factory. Ltd) at 200 rpm using ceramic spheres and jars for 0, 12, 24, 48, 72, and 144 hours. An air-dried softwood, Hinoki cypress (*Chamaecyparis obtuse*), was first crushed by a rotary crusher and sieved with 10 US standard mesh to collect wood powder with a particle size of 200 µm. The wood powder was then ball milled for 0, 6, 12, 24, 32, 48, 72, and 144 hours using the same procedure as the MCC powder.

The pseudo-wood powder was a mixture of MCC and organic-solvent lignin powder (Guangzhou Yinnovator Bio-tech Co., Ltd), where the lignin powder was extracted from Eucalyptus (*Eucalyptus*) by a high-concentration glacial acetic acid (>70%). To obtain different CrIs, we used the following mass fractions of lignin in the powder: 0%, 25%, 50%, 75%, and 100%. The 100% organic-solvent lignin powder can be considered as a totally amorphous powder.

The prepared powders with different CrIs were collected at a mass of 0.075 g by an electronic balance (±0.0001 g). All the powders were compressed into tablets with a diameter of 14 mm and a thickness of approximately 0.35 mm using a hand-pressing tableting kit (IMC-180C, Imoto Machinery Co., LTD). Three tablets were prepared for each powder. The thickness of these tablet samples was measured using a micrometer (± 0.001 mm).
XRD and THz-TDS measurement

The XRD measurements of all the tablet samples were performed with Cu-Kα radiation ($\lambda = 0.1542$ nm) using an X-ray diffractometer (Ultima IV, Rigaku) at a voltage of 40 kV and a current of 40 mA. Diffractograms were recorded from $5^\circ$ to $40^\circ$, where the scan range included the peak of the main crystal lattice of cellulose and wood. The scan speed was set to $5^\circ$ min$^{-1}$, and the sampling step was $0.05^\circ$. The background diffractogram was obtained from an empty sample holder.

The THz transmission spectra of all the tablet samples were measured by using a Tera Prospector-Kit model (NIPPO PRECISION Co., Ltd.), and the reference signals were obtained by measurement of air before and after the sample measurement. The THz beam was horizontally polarized with the bandwidth from about 0.1 THz to 4.00 THz, and the spectral resolution was 0.02 THz, corresponding to the inverse of the temporal scan range (50 ps). The diameter of the THz beam spot on the sample was around 3 mm. Each measurement was recorded by averaging 100 scans to improve the signal-to-noise ratio. For reproducibility, all measurements were conducted thrice. To avoid the influence of water vapor on the measurement caused by THz absorption, we placed the whole THz optical system in an almost-closed acrylic box that was continuously filled with dry air to ensure the stability of humidity. All samples were placed in the box for 24 hours before measurement to balance the ambient humidity. When the samples were placed in the box, dry airflow was filled into the box until all the THz measurements were completed.

Results and Discussion

Calculation of CrI with XRD pattern

The original XRD pattern was cut out with a scattering range from $10^\circ$ to $40^\circ$ of pseudo-wood, MCC, and wood samples, as shown in Fig. 1(a), 1(b), and 1(c), respectively. Gradient colors were used to express changes in the ball milling time of samples and mass fractions of lignin. Background subtraction and baseline correction were performed on all original XRD patterns before further calculation. The background pattern was obtained as mentioned in the experimental section. The
baseline was fitted as a first-order polynomial after background subtraction; the process is shown in Fig. 2(a).

After background subtraction and baseline correction, the corrected XRD patterns can be considered as composites of an amorphous intensity curve and five main crystalline peaks for the $I_\beta$ type cellulose, a dominant type in vascular plants such as wood. The five main crystalline peaks had Miller indices of (1 1̅0), (110), (102), (200), and (004). The amorphous intensity curve was determined by two different methods (Park et al. 2010), which will be discussed in detail later.

The deconvolution of all the crystalline peaks was carried out with a curve-fitting process using a pseudo-Voigt profile, which is a linear combination of a Gaussian curve and a Lorentzian curve as shown respectively in the following equations:

$$I_G(2\theta) = I_{max} \cdot \exp\left\{-4 \ln(2) \cdot \left(\frac{2\theta - 2\theta_{max}}{\beta}\right)^2\right\}$$  \hspace{1cm} (1)

$$I_L(2\theta) = I_{max} \left\{1 + 4 \left(\frac{2\theta - 2\theta_{max}}{\beta}\right)^2\right\}^{-1}$$  \hspace{1cm} (2)

$$I_{PV}(2\theta) = \mu I_L + (\mu - 1)I_G$$  \hspace{1cm} (3)

where $I_{max}$ is the intensity of peaks, $2\theta_{max}$ is the peak position, and $\beta$ is the full width at half maximum (FWHM); $I_G(2\theta)$ and $I_L(2\theta)$ are the Gaussian and Lorentzian curves, respectively (de Keijser et al. 1983; Wada et al. 1997).

The positions ($2\theta$) of the crystalline peaks were fixed at 14.9°, 16.7°, 20.5°, 22.6°, and 34.5°, corresponding to the Miller indices of (1 1̅0), (110), (102), (200), and (004), respectively. The other parameters of FWHM, the peak intensity, and coefficient $\mu$ were all determined by the curve-fitting process.

In this study, the amorphous intensity curve was first fitted as a pseudo-Voigt profile, as shown in Fig. 2(b), the same fitting process as that of other crystalline peaks. Given that recent research showed that the maximum of the amorphous intensity curve is slightly more than 20°, therefore, the peak position ($2\theta$) of the amorphous intensity curve was fixed at 20.6°(Yao et al. 2020), where the position was very close to the (102) crystalline peak, and the influence caused by this will be discussed in detail later.
Yao et al. reported the Fourier series modeling of amorphous intensity curve, and this method was used in this study (Yao et al. 2020). The amorphous intensity curve was determined by fitting the averaged XRD pattern of three 100% organic-solvent lignin samples using a 7th-order Fourier series as shown in Fig. 2(c). The obtained Fourier series model can be considered as the basis function of the amorphous intensity curve. The XRD patterns of other different samples only showed the difference in intensity. The intensity can be adjusted by multiplying this Fourier series by a coefficient $k$ in the deconvolution of XRD patterns:

$$f(2\theta) = k \cdot \{a_0 + \sum_{i=1}^{7} a_i \cos(2\theta \cdot w) + b_i \cos(2\theta \cdot w)\}$$  \hspace{1cm} (4)

Supplementary Information provides the details of the fitting of the Fourier series of 100% organic-solvent lignin samples.

The deconvolution results were evaluated with a coefficient of determination $R^2$:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_i - Y_i')^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$  \hspace{1cm} (5)

where $X_i$ is the intensity after background subtraction and baseline correction, $Y_i$ is the fitted intensity, and $\bar{X}$ is the average of intensities. Table 1 summarizes the calculated $R^2$ of the pseudo-wood fitted by two different amorphous intensity curves, and Supplementary Information summarizes the $R^2$ values of other samples. As shown in Table 1, the curves fitted by both methods showed good $R^2$, which indicates that the curve-fitting of the two methods was credible.

**Table 1** Calculated $R^2$ values of the pseudo-wood fitted by two different amorphous intensity curves

| Amorphous curve   | Volume fraction of lignin |
|-------------------|---------------------------|
| pseudo-Voigt      | 0% | 25% | 50% | 75% | 100% |
| $R^2$             | 0.992 | 0.993 | 0.996 | 0.993 | 0.909 |
|                  | 0.990 | 0.987 | 0.987 | 0.989 | 0.960 |
| Fourier series    | 0.982 | 0.996 | 0.996 | 0.994 | 0.898 |
|                  | 0.991 | 0.993 | 0.997 | 0.994 | 0.910 |
| $R^2$             | 0.988 | 0.986 | 0.988 | 0.991 | 0.962 |
|                  | 0.979 | 0.996 | 0.996 | 0.995 | 0.902 |

Once the peaks were deconvoluted by the curve-fitting process, the CrI was determined by the following equation:

$$CI = \frac{S_{Cr}}{S_{Cr} + S_{Am}}$$  \hspace{1cm} (6)
where $S_{Cr}$ and $S_{Am}$ are the sum of the integrated intensity of five crystalline peaks and the integrated intensity of amorphous intensity curve, respectively.

As shown in Fig. 2(b) and 2(c), the amorphous intensity curve fitted with a pseudo-Voigt profile and a Fourier series, respectively, gave different deconvolution results. The most evident intensity of the (102) peak in Fig. 2(b) was almost invisible, since the peak position of (102) peak was fixed at 20.5° and the amorphous peak was fixed at 20.6°, the contribution of the (102) peak was almost systematically ignored in the fitting process. On the other hand, the (102) peak can be clearly observed in Fig. 2(c). Thus, the CrI calculated with a pseudo-Voigt profile amorphous intensity curve was smaller than that calculated with a Fourier series amorphous intensity curve on average.

Fig. 3 shows the CrI calculated from the different methods. In the figure, the amorphous curve for (a) pseudo-wood, (b) MCC, and (c) wood was fitted by pseudo-Voigt profile, whereas that for (d) pseudo-wood, (e) MCC, and (f) wood was fitted by Fourier series. As shown in Fig. 3, two different fittings gave a similar crystallinity, whereas for the pseudo-wood, the crystallinity decreased from about 85% to about 5% with the increase in mass fraction of lignin. For MCC and wood, the crystallinity decreased from about 90% to 60% and 80% to 40% with the increase in ball-milling time, respectively.

As shown in Fig. 3(a) and 3(d), for the pseudo-wood, the samples with a 25% mass fraction of organic-solvent lignin showed a similar crystallinity with the samples without organic-solvent lignin (100% MCC samples). A non-linear correlation was observed between the CrI and mass fraction of lignin, where the CrI decreased sharply when the mass fraction of lignin was over 50%. Regardless of the fitting method used to obtain the curve, the sample that had 100% organic-solvent lignin, which was supposed to have no crystalline region, still showed a crystallinity of about 5%. The crystallinity for MMC and wood showed a similar downtrend. In general, the concentration of cellulose in different species of wood is about 30% to 40% (PETTERSEN 1984). However, in this study, for most of the ball-milling times, the CrI of wood was about 10% lower than that of MCC.

For the same type of samples, the changes in the CrI calculated from the XRD patterns was easy to observe and evaluate. However, when the cellulose concentration of wood was about 30% to 40%, the original wood samples still showed a relatively similar CrI compared with that of the original
MCC samples. Thus, accurate comparison of the differences in CrIs between various types of samples using the same standard can be a challenge. Table 2 summarizes the detailed crystallinity of the pseudo-wood sample calculated from the two curve-fittings and the crystallinity calculated from the THz mass absorption coefficient spectrum, which will be discussed in detail in the next section. Supplementary Information summarizes the crystallinities of other samples.

Fig.1 Original XRD patterns of (a) pseudo-wood with a decreased mass fraction of organic-solvent lignin, (b) MCC, and (c) wood samples.
Fig. 2 Curve-fitting process of a pseudo-wood with 50% lignin. (a) Background subtraction and baseline correction, (b) deconvolution of peaks by using a pseudo-Voigt profile only, and (c) the deconvolution of peaks where the crystalline peaks fitted by using a pseudo-Voigt profile; the amorphous curve was fitted by a 7th-order Fourier series model.
Fig. 3 CrI calculated from two different curve-fitting processes of the amorphous curve, where (a) – (c) with circle markers were fitted by a pseudo-Voigt profile, and (d) – (e) with diamond markers were fitted by a 7th-order Fourier series.

**Evaluation of CrI by THz mass absorption coefficient spectra**

The measured THz time-domain signal was Fourier transformed into the frequency domain. Given that the samples used in this study were all “optically thick samples,” where the backward and forward reflections in the sample (Fabry–Pérot effect) can be ignored (Duvillaret et al. 1996), further to correct the possible inhomogeneity of hand-made samples, the mass absorption coefficient $\alpha$ that also considers the influence of density was calculated using the following equations:

$$\alpha = \frac{1}{\rho L} \ln\left(\frac{R_r}{R_s}\right)$$

where $\rho$ is the sample density, $L$ is the sample thickness, $R_r$, $R_s$ are the phase and amplitude in the frequency domain of the reference and measured samples, respectively, and (Kore and Pawar 2014).

Fig. 4 (a), 4(b), and 4(c) showed the original THz mass absorption coefficient spectra from 0.2 THz to 4 THz of the pseudo-wood, MCC, and wood, respectively. The same gradient colors as that in Fig. 1 were used. The calculated mass absorption coefficient spectra were all first corrected by baseline fluctuations with a standard normal variate algorithm (SNV) and then smoothed by the
application of a Savitzky–Golay filter with a second-order polynomial and fifteen smoothing points, as shown in Fig. 4(d) – 4(f). The pseudo-wood and MCC samples showed absorption peaks at around 2.1 THz, and these peaks gradually decreased with the increase in mass fraction of lignin or ball-milling time, whereas at this frequency position, no evident absorption peaks of wood samples were observed. Previous research discovered that the absorption peaks at 2.1 THz correlated with crystalline structures, where the I_β- dominant and I_α-rich types showed absorption peaks at 2.11 and 2.38 THz, respectively. Furthermore, the intensity of the peaks here was correlated with the fraction of I_α. The larger the fractions of I_α, the stronger the absorption at 2.38 THz. On the contrary, the smaller the fraction of I_α, the stronger the absorption at 2.11 THz (Wang et al. 2020). Similar to the case of wood samples, the absorption peaks of pseudo-wood samples with the mass fraction over 50% of organic-solvent lignin were difficult to observe at around 2.1 THz, whereas the cellulose concentration of wood samples was about 30%. Therefore, no evident absorption peak, such as that of MCC samples at around 2.1 THz can be observed. This phenomenon might have been caused by the low cellulose concentration (about 30%) of wood. The thickness of cellulose samples used in the previous research was about 0.5 mm to 1 mm (Wang et al. 2020), which was thicker than that of the samples in this study. Given that the absorption around 3 THz is correlated with the CrI of cellulose (Vieira and Pasquini 2014), the samples were made thin in this study to improve the signal-to-noise ratio to investigate further the changing absorption peaks around 3 THz. However, this condition may lead to difficult observation of the small absorption peaks around 2 THz.

As shown in Fig. 4(a) – 4(f), all types of samples expressed a relatively strong absorption at 3.04 THz without a peak shift. The mass absorption coefficient spectra of wood showed a relatively small intensity compared with other samples. To further study this absorption peak, we cut out the mass absorption coefficient spectra from the range of 2.79 THz to 3.32 THz and then subtracted the baseline that was fitted as a first-order polynomial. Fig. 4(g), 4(h), and 4(i) show the baseline corrected mass absorption coefficient spectra of the pseudo-wood, MCC, and wood, respectively. The integrated intensities of all the detrended mass absorption coefficient spectra were calculated, and their correlations with the mass fraction of lignin and ball-milling time are plotted in Fig. 5(a), 5(b), and 5(c).
For the pseudo-wood samples, as calculated from the XRD patterns, the integrated intensity of the mass absorption coefficient spectra from 2.79 THz to 3.32 THz showed a linear correlation with the mass fractions of lignin. Given that the mass fraction of lignin was adjusted to increase linearly (+25% each time) during the sample preparation process, the correlation obtained from THz mass absorption coefficient spectra is more reasonable than that obtained from XRD patterns. The correlations between the ball-milling time with the integrated intensity of the MCC and wood samples showed a similar trend with that of the crystallinity obtained from the XRD patterns. However, the values of 144 h ball-milled MCC samples decreased by about 50% compared with that of the original MCC samples, whereas the CrI of the 144 hours ball-milled MCC samples calculated from XRD patterns decreased by about 25%. Furthermore, the original wood samples without ball milling showed relatively small values compared with the original MCC samples. By contrast, the CrI of the original wood samples calculated from the XRD pattern showed no evident difference with that of original MCC samples. The integrated intensities of the original wood samples were about 30% those of the original MCC samples, thus matching with the approximate concentration of cellulose in wood (about 30%). Table 2 summarizes the detailed integrated intensities of the pseudo-wood samples and the CrI calculated by two curve-fittings from the XRD patterns.

The integrated intensities of all samples were normalized to the range of 0 to 1 by the min-max algorithm, and the results are presented in Fig. 4(d), 4(e), and 4(f). After this process, the normalized value can be considered as an “index” that can be used to evaluate crystallinity. Different from the CrI, which is a relative value calculated from XRD patterns, an assumption was easily reached from the above results, that is, the THz mass absorption coefficient spectra may reveal the absolute mass fractions of crystalline cellulose in all the samples used in this study. Given that the peaks of the mass absorption coefficient in the THz region did not shift with the changes the type of samples or ball-milling time and mass fraction of lignin, only the intensity changed. Moreover, the THz spectra contain rich physical and chemical information, such as the phonon frequency of crystal lattices, which provides a possibility for the comparison of the mass fraction of crystalline cellulose between different cellulose-based materials. However, the energy corresponding to 3 THz is 0.012 eV, which
is relatively small compared with the stabilization energy of hydrogen bonding in cellulose I\(_\beta\), for example, the intrasheet O6-H…O3 H-bonds with the stabilization energy of about 0.87 eV (Parthasarathi et al. 2011). Without further research based on the theoretical proof level such as the quantum mechanical calculation using density-functional theory (DFT), the assignments of absorption peaks at 3.04 THz are still uncertain.

Fig. 4 Original THz mass absorption coefficient spectra of (a) pseudo-wood, (b) MCC, and (c) wood samples. (d), (e), and (f) SNV and smoothed mass absorption coefficient spectra. (g), (h), and (i) Detrended mass absorption coefficient spectra obtained by subtraction of the baseline at the frequency range of 2.79 THz to 3.32 THz.
Fig. 5 Integrated intensity calculated from the detrended mass absorption coefficient spectra from 2.79 THz to 3.32 THz of (a) pseudo-wood, (b) MCC, and (c) wood samples. (d), (e), and (f) Normalized integrated intensity by min-max algorithm.

**Table 2** CrI of the pseudo-wood calculated from XRD patterns and the integrated intensity obtained from 2.79 THz to 3.32 THz mass absorption coefficient spectra

| Calculation method          | Volume fraction of lignin | 0%           | 25%           | 50%           | 75%           | 100%          |
|-----------------------------|---------------------------|--------------|--------------|--------------|--------------|--------------|
| XRD with pseudo-Voigt Am    |                           | 84.270       | 77.024       | 73.808       | 47.351       | 8.393        |
| CrI (%)                     |                           | 82.361       | 77.163       | 70.362       | 58.641       | 8.586        |
|                             |                           | 80.704       | 78.268       | 73.242       | 52.810       | 10.794       |
| XRD with Fourier series Am  |                           | 90.870       | 85.692       | 76.463       | 50.256       | 4.966        |
| CrI (%)                     |                           | 86.272       | 81.236       | 72.402       | 58.293       | 3.296        |
|                             |                           | 81.576       | 86.667       | 82.702       | 56.254       | 7.075        |
| Detrended THz mass absorption coefficient integrated intensity |                           | 0.171        | 0.132        | 0.095        | 0.051        | 0.011        |
|                             |                           | 0.142        | 0.133        | 0.093        | 0.048        | 0.013        |
|                             |                           | 0.178        | 0.139        | 0.083        | 0.041        | 0.013        |
| Min-maxed integrated intensity |                           | 0.957        | 0.734        | 0.518        | 0.261        | 0.027        |
|                             |                           | 0.789        | 0.740        | 0.506        | 0.246        | 0.039        |
|                             |                           | 1.000        | 0.774        | 0.448        | 0.206        | 0.044        |
Conclusions

In this study, pseudo-wood, MCC, and wood samples were adjusted to different cellulose CrI by changing the mass fraction of lignin of pseudo-wood samples and the ball-milling time of MCC and wood samples. The measured XRD patterns of all samples were deconvoluted by two different curve-fitting processes, in which one used a pseudo-Voigt profile, and the other used a 7th-order Fourier series to represent the amorphous intensity curve. Both methods gave good fitting results, and the calculated CrIs were close. All samples showed peaks at 3.04 THz of the THz mass absorption coefficient spectra. The THz mass absorption coefficient spectra were pretreated by SNV to correct baseline fluctuations, smoothed by the application of a Savitzky–Golay filter with fifteen points, and then detrended by subtraction of the baseline from 2.79 THz to 3.32 THz. The integrated intensities showed a similar changing trend with the CrIs obtained from XRD patterns. However, evident differences were observed. For the pseudo-wood, the correlation between the CrI and mass fraction of lignin were non-linear, whereas the integrated intensity of detrended THz mass absorption coefficient spectra decreased linearly with the increase in the mass fraction of lignin. The original wood sample without ball milling had an integrated intensity value that was about 30% of that in the original MCC sample, matching the cellulose concentration of wood. By contrast, the CrI of the original wood sample calculated from XRD patterns showed a similar value to that of the original MCC sample. Thus, based on the above results, THz mass absorption coefficient spectra may be used to evaluate the mass fraction of crystalline cellulose. Furthermore, after normalization of the integrated intensity by a min-max algorithm to the range of 0 to 1, the original MCC sample without lignin and the 100% lignin sample can be considered as a pair of standard samples corresponding to 1 and 0, respectively. All the cellulose crystallinity can be evaluated in this range. However, the assignments of absorption peaks of cellulose at 3.04 THz are still uncertain. Numerous optional functions that rely on the experience of analysts, such as pseudo-Voigt profile and Fourier series used in this study, are available for the deconvolution of XRD patterns. Meanwhile, the processing of THz mass absorption coefficient spectra is simple, and all cellulose-based materials
showed peaks at 3.04 THz without a peak shift. The measurement of THz signal is quick, simple, and safe and requires no sample pretreatment. Thus, THz-TDS has the potentiality of becoming a new standard for the evaluation of cellulose crystallinity.

Acknowledgments The authors would like to express our grateful appreciation to Dr. Liu and Dr. Wang for the supplement of the organic-solvent lignin samples. This work was supported by JSPS KAKENHI Grant Number 16H02559. And the first author also thanks for the Mitsutani scholarship for living support.

Reference

Braly LB, Liu K, Brown MG, et al. (2000) Terahertz laser spectroscopy of the water dimer intermolecular vibrations. II. (H2O)2. J Chem Phys 112:10314–10326.

de Keijser T, Mittemeijer EJ, Rozendaal HCF (1983) The determination of crystallite-size and lattice-strain parameters in conjunction with the profile-refinement method for the determination of crystal structures. J Appl Crystallogr 16:309–316.

del Cerro DR, Koso T V., Kakko T, et al. (2020) Crystallinity reduction and enhancement in the chemical reactivity of cellulose by non-dissolving pre-treatment with tetrabutylphosphonium acetate. Cellulose 27:5545–5562.

Density ANDC, Kellogg RM, Sastr CBR, Ivellwood RW (1975) Relationships Between Cell-Wall Composition. 7:170–177

Duvillaret L, Garet F, Coutaz JL (1996) A reliable method for extraction of material parameters in terahertz time-domain spectroscopy. IEEE J Sel Top Quantum Electron 2:739–745.

French AD (2014) Idealized powder diffraction patterns for cellulose polymorphs. Cellulose 21:885–896.

French AD, Santiago Cintrón M (2013) Cellulose polymorphy, crystallite size, and the Segal Crystallinity Index. Cellulose 20:583–588.

Hall M, Bansal P, Lee JH, et al. (2010) Cellulose crystallinity - A key predictor of the enzymatic hydrolysis rate. FEBS J 277:1571–1582.
Hermans PH, Weidinger A (1948) Quantitative x-ray investigations on the crystallinity of cellulose fibers. A background analysis. J Appl Phys 19:491–506.

Himmel ME, Ding S, Johnson DK, Adney WS (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science (80- ) 315 (5813):804–808.

Hindeleh AM, Johnson DJ (1972) Crystallinity and crystallite size measurement in cellulose fibres: 1. Ramie and Fortisan. Polymer (Guildf) 13:423–430.

Inagaki T, Ahmed B, Hartley ID, et al. (2014a) Simultaneous prediction of density and moisture content of wood by terahertz time domain spectroscopy. J Infrared, Millimeter, Terahertz Waves 35:949–961.

Inagaki T, Hartley ID, Tsuchikawa S, Reid M (2014b) Prediction of oven-dry density of wood by time-domain terahertz spectroscopy. Holzforschung 68:61–68.

Inagaki T, Siesler HW, Mitsui K, Tsuchikawa S (2010) Difference of the crystal structure of cellulose in wood after hydrothermal and aging degradation: A NIR spectroscopy and XRD study. Biomacromolecules 11:2300–2305.

Jörgensen L, Kiessling R, Lindstedt G, Kinell P-O (1950) Swelling and Heterogeneous Hydrolysis of Cotton Linters and Wood Pulp Fibers Related to Their Fine Structure. Acta Chem. Scand. 4:185–199.

Kataoka Y, Kondo T (1998) FT-IR microscopic analysis of changing cellulose crystalline structure during wood cell wall formation. Macromolecules 31:760–764.

Kim SH, Lee CM, Kafle K (2013) Characterization of crystalline cellulose in biomass: Basic principles, applications, and limitations of XRD, NMR, IR, Raman, and SFG. Korean J Chem Eng 30:2127–2141.

Kore PS, Pawar PP (2014) Measurements of mass attenuation coefficient, effective atomic number and electron density of some amino acids. Radiat Phys Chem 98:86–91.

Langan P, Sukumar N, Nishiyama Y, Chanzy H (2005) Synchrotron X-ray structures of cellulose Iβ and regenerated cellulose II at ambient temperature and 100 K. Cellulose 12:551–562.

Ling Z, Wang T, Makarem M, et al. (2019) Effects of ball milling on the structure of cotton cellulose. Cellulose 26:305–328.
Nishiyama Y (2009) Structure and properties of the cellulose microfibril. J Wood Sci 55:241–249.

Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 124:9074–9082.

Parthasarathi R, Bellesia G, Chundawat SPS, et al. (2011) Insights into hydrogen bonding and stacking interactions in cellulose. J Phys Chem A 115:14191–14202.

PetterSEN RC (1984) The Chemical Composition of Wood. In: The Chemistry of Solid Wood. pp 57–126

Poletto M, Zattera AJ, Forte MMC, Santana RMC (2012) Thermal decomposition of wood: Influence of wood components and cellulose crystallite size. Bioresour Technol 109:148–153.

Segal L, Creely JJ, Martin AE, Conrad CM. An Empirical Method for Estimating the Degree of Crystallinity of Native Cellulose Using the X-Ray Diffractometer. Textile Research Journal.

Teeäär R, Serimaa R, Paakkarl T (1987) Crystallinity of cellulose, as determined by CP/MAS NMR and XRD methods. Polym Bull 17:231–237.

Thygesen A, Oddershede J, Lilholt H, et al. (2005) On the determination of crystallinity and cellulose content in plant fibres. Cellulose 12:563–576.

Updegraff DM (1969) Semimicro determination of cellulose in biological materials. Anal Biochem 32:420–424.

Vieira FS, Pasquini C (2014) Determination of cellulose crystallinity by terahertz-time domain spectroscopy. Anal Chem 86:3780–3786.

Wada M, Okano T, Sugiyama J (1997) Synchrotron-radiated X-ray and neutron diffraction study of native cellulose. Cellulose 4:221–232.

Wang H, Horikawa Y, Tsuchikawa S, Inagaki T (2020) Terahertz time-domain spectroscopy as a novel tool for crystallographic analysis in cellulose. Cellulose 27:9767–9777.
Wang H, Inagaki T, Hartley ID, et al. (2019) Determination of Dielectric Function of Water in THz Region in Wood Cell Wall Result in an Accurate Prediction of Moisture Content. J Infrared, Millimeter, Terahertz Waves 40:673–687.

Xie L, Yao Y, Ying Y (2014) The application of terahertz spectroscopy to protein detection: A review. Appl Spectrosc Rev 49:448–461.

Yao W, Weng Y, Catchmark JM (2020) Improved cellulose X-ray diffraction analysis using Fourier series modeling. Cellulose 27:5563–5579.