Terahertz radiations for breast tumour recognition

Quentin Cassar

▶ **To cite this version:**

Quentin Cassar. Terahertz radiations for breast tumour recognition. Biotechnology. Université de Bordeaux, 2020. English. NNT : 2020BORD0032 . tel-03053722

HAL Id: tel-03053722
https://theses.hal.science/tel-03053722

Submitted on 11 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

université de BORDEAUX

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

DÉPARTEMENT DES SCIENCES PHYSIQUES ET DE L'INGÉNIEUR

SPÉCIALITÉ LASERS, MATIÈRE, NANOSCIENCES

présentée et soutenue publiquement par

Quentin CASSAR

le 25 Mars 2020

Terahertz Radiations for Breast Tumour Recognition

Directeur de thèse : **Patrick MOUNAIX**
Encadrant : **Jean-Paul GUILLET**

Jury

M. Zimmer Thomas,	Professeur des universités, IMS, Bordeaux	Président
M. Gallot Guilhem,	Directeur de recherche, LOB, Paris	Rapporteur
M. Garet Frédéric,	Professeur des universités, IMEP-LAHC, Chambéry	Rapporteur
M. Peretti Romain,	Chargé de recherches, IEMN, Villeneuve d'Ascq	Examinateur
M. Pfeiffer Ullrich,	Professeur des universités, IHCT, Wuppertal	Invité
M. MacGrogan Gaëtan,	Docteur en biopathologie, Institut Bergonié, Bordeaux	Invité
Mme. Tunon de Lara Christine,	Docteur en chirurgie, Institut Bergonié, Bordeaux	Invitée

Abstract (EN)-

The failure to accurately delineate breast tumor margins during breast conserving surgeries results in a 20% re-excision rate. Consequently, there is a clear need for an operating room device that can precisely define intraoperatively breast tumor margins in a simple, fast, and inexpensive manner. This manuscript reports investigations that were conducted towards the ability of terahertz radiations to recognize breast malignant lesions among freshly excised breast volumes. Preliminary works on terahertz far-field spectroscopy have highlighted the existence of a contrast between healthy fibrous tissues and breast tumors by about 8% in refractive index over a spectral window spanning from 300 GHz to 1 THz. The origin for contrast was explored. Results seem to indicate that the dynamics of quasi-free water molecules may be a key factor for demarcation. On these basis, different methods for tissue segmentation based on refractive index map were investigated. A cancer sensitivity of 80% was reported while preserving a specificity of 82%. Eventually, these pilot studies have guided the design of a BiCMOS-compatible near-field resonator-based imager operating at 560 GHz and sensitive to permittivity changes over breast tissue surface.

Keywords: Terahertz spectroscopy imaging; Biomedical engineering; Breast cancer; Data science; Inverse electromagnetic problem.

Résumé (FR)-

La faible précision avec laquelle sont délimitées les marges d'exérèse des adenocarcinomes mammaires se traduit par un recours régulier à un second acte chirurgical. Afin d'en limiter la fréquence, la communauté scientifique tente de définir les grands axes de conception d'un système intraopératif permettant la reconnaissance des lésions mammaires malignes. Ce manuscrit de thèse rapporte les investigations menées sur la capacité des ondes térahertz à fournir un contraste entre tissus mammaires sains et malins. Les premiers travaux ont montré l'existence d'un contraste sur l'indice de réfraction entre 300 GHz et 1 THz, évalué en moyenne à 8%. Ce dît contraste semble prendre origine dans la dynamique des molecules d'eau intrinsèques aux cellules cancéreuses. Différentes techniques de segmentation d'image, basées sur l'indice de refraction des zones tissulaires, ont permis de rapporter une sensibilité au cancer jusqu'à 80% tout en maintenant un taux de spécificité de l'ordre de 82%. L'ensemble de ces études a guidé la conception d'un imageur champ-proche opérant à 560 GHz, dont la réponse des différents senseurs est sensible à la permittivité en surface des tissus du sein.

Mots Clefs: Spectro-imagerie térahertz; Ingénierie biomédicale; Cancer du sein; Science des données; Problème électromagnetique inverse.

Université de Bordeaux
Laboratoire de l'Intégration du Matériau au Système (IMS)
UMR CNRS 5218, F-33400 Talence, France

Contents

Resumes iii

Contents v

List of Figures vii

List of Tables xv

Acknowledgments 1

Acronyms 3

Symbols 5

Medical Glossary 7

General Introduction 9

I Context and Key Concepts 13
- I.1 Introduction . . . 14
- I.2 Terahertz Radiations . . . 14
- I.3 Terahertz Field and Biological Samples . . . 15
- I.4 Breast Cancer . . . 39
- I.5 Conclusion . . . 43
- I.6 Bibliography . . . 44

II Terahertz Spectroscopy 53
- II.1 Introduction . . . 54
- II.2 Generation, Detection and Sensing . . . 54
- II.3 Spectroscopy: Principles . . . 62
- II.4 Water . . . 75
- II.5 Phantoms . . . 80
- II.6 Excised Breast Tissues . . . 89
- II.7 Cancerous Cells . . . 99
- II.8 Conclusion . . . 104
- II.9 Bibliography . . . 105

III Terahertz Imaging: Far-field Survey 109
- III.1 Introduction . . . 110
- III.2 Far-field Imaging: Principles . . . 110
- III.3 Self-Reference Image Inversion . . . 114
- III.4 Image Registration . . . 121
- III.5 Correlation . . . 124
- III.6 Study Cases . . . 131
- III.7 Principal Component Analysis . . . 148

III.8 Conclusion . . . 160
III.9 Bibliography . . . 161

IV Breast Cancer Imaging: Near-field Survey **165**
IV.1 Introduction . . . 166
IV.2 Beating the Diffraction Limit of Resolution . . . 166
IV.3 Towards a Silicon-Based Terahertz Near-field Imaging Matrix . . . 175
IV.4 Conclusion . . . 194
IV.5 Bibliography . . . 195

General Conclusion **199**

Author Publication List **201**

Introduction en Français **203**

List of Figures

I.1 Location of the terahertz regime within the electromagnetic spectrum. 14

I.2 Refraction and reflection occurring at a dielectric interface between two media of propagation. 16

I.4 Example of a simple Debye treatment for $\Delta\epsilon_r - \epsilon_\infty = 12$, $\epsilon_\infty = 2$ and $\tau_D = 1ps$. The inflection point of $\epsilon'_r(\omega)$ and the maximum of the dielectric loss factor $\epsilon''_r(\omega)$ are located at $\omega_C = 1.01$ rad.ps^{-1}. 20

I.5 Illustration of the Cole-Cole treatment for the Debye parameters used in Figure I.4 (Top: blue; Bottom: red) and for $\alpha = 0.25, 0.50, 0.75$ subsequently. 21

I.6 Illustration of the Cole-Davidson treatment for the Debye parameters used in Figure I.4 and for $\beta = 0.25, 0.50, 0.75$ subsequently. 22

I.7 Illustration of the Havriliak-Negami formulation for the Debye parameters used in Figure I.4 and for α and β taking the successive values: 0.25, 0.50, 0.75. 23

I.8 Illustration of a Lorentz oscillator for $\epsilon_0 = 75$, $A/(2\pi)^2 = 0.25$ THz^2, $\Gamma/(2\pi) = 0.32$ THz and a characteristic frequency at $\omega_0/(2\pi) = 500$ GHz. 24

I.9 Schematic representation of the real and imaginary part of the relative permittivity, ϵ'_r and ϵ''_r as a function of the frequency and the location of the dispersive domains α, β, δ and γ. Where α denotes KHz relaxation of ions, β describes the MHz relaxation of large biomolecules, δ depicts the hydration shell relaxation within protein surface - water complex and γ stands for bulk and free-water relaxations. Simulation performed by a tetra-Debye treatment (equation I.22 wherein N = 4) for $\epsilon_\infty = 3.33$, $\Delta\epsilon_1 = 2.49 \cdot 10^8$, $\Delta\epsilon_2 = 6.66 \cdot 10^5$, $\Delta\epsilon_3 = 900$, $\Delta\epsilon_4 = 96.67$, $\tau_1 = 10^{-2}$s, $\tau_2 = 10^{-6}$s, $\tau_3 = 5 \cdot 10^{-10}$s, $\tau_4 = 5 \cdot 10^{-11}$s. 25

I.10 Water molecules. Right: the hydrogen-bond network. Left: closer look on a hydrogen-bond and a covalent bond. 26

I.11 Structure of the α-form (left) and the β-form (right) of D-glucose. 28

I.12 Structure of the heparan sulfate polysaccharide subunit. 28

I.13 Chemical structure of the fatty acid $C_{16}H_{32}O_2$ (palmitic acid). 29

I.14 Schematic diagram of the DNA helix structure. 30

I.15 Protein synthesis in the ribosome complex. Proteins are assembled as stipulated by the encoded genetic information. The messenger-RNA (mRNA) serves as a template while the transfer-RNA (tRNA) acts as the physical link between the mRNA and the amino acid sequence of growing proteins. 31

I.16 Skin diagramm. 34

I.17 Eye anatomy. 35

I.18 Tooth structure. 36

I.19 Breast anatomy. 40

I.20 Evolution of a ductal adenocarcinoma. 40

I.21 Breast cancer surgery management: from initial partial mastectomy to second intervention, and histological examination in between. Step 1: abnormal tissue may extend beyond the visible fraction of the tumor. If not completely removed, the cancer will recur. Step 2: The surgeon removes the visible portion of the tumor. Step 3: The excised tissue is formalin-fixed and paraffin-embedded. Then, the tissue is divided into thin section of around 5µm. Hematoxylin and eosin (H&E) stain is carried out to highlight cell cytoplasm and nuclei and the extracellular matrix. Hematoxylin stains cell nuclei blue, eosin staining pink the remainder of the section. Step 4: the pathologist microscopically examines the margins and the undersurface of each section. Step 5: if margins are defined as positive, the surgeon returns to the patient to remove the remaining abnormal cells. 42

II.1 Illustrative example of pulsed terahertz generation by means of a photoconductive antenna. LCL: long carrier lifetime. SCL: short carrier lifetime. 55

II.2 Details of a free-space terahertz detection using a $\chi_{(2)}$ susceptibility crystal. 57

II.3 Reconstruction of a terahertz pulse by means of electro-optic sampling method. . . 58

II.4 Schematic details of a time-domain system for terahertz spectroscopy in transmission configuration. BS stands for beam splitter, and M depicts mirrors. 59

II.5 Schematic details of a time-domain system for terahertz spectroscopy in reflection configuration. BS stands for beam splitter, and M depicts mirrors. 59

II.6 Illustrative terahertz time-domain waveform (blue) and spectrum (red) recorded through the use of a TDS setup. 60

II.7 Photograph of the TPS Spectra 3000 (left) and the Terapulse 4000 (right). 61

II.8 Scheme of the iterative determination of $S_{sam}(\omega)$. Each k-incrementation gives access to further possible events based on the previous ones at $k-1$. Detected pulses have been marked with an eye in the node box. The incident angle was tilted for visual convenience. 63

II.9 Microscopic examination of the calibration sample. 64

II.10 Refractive index $n(f)$ and extinction coefficient $k(f)$ of each medium belonging to the calibration sample. 65

II.11 Progressive reconstruction of $S_r(t)$ (red) in comparison to the experimental signal (black) for the calibration sample made of four layers each of tens of microns. 66

II.12 Top: experimental signal of the stratified structure in reflection geometry (black) and its reconstruction via the iterative tree algorithm (red). Bottom: deviations observed from the measured signal using the $\Lambda(t)$ metric. 67

II.13 Profile of the integrated comparison metric between 22 ps and 37 ps as a function of the number of iterations. The non-monotonous nature of the dependence reflects that time features are in majority due to the combination of several physical processes. Despite additional iterations, the error does not fall to zero. That is caused by unrecovered time features for the previously mentioned reasons and because time features corresponding to $k \to \infty$ are basically out of the time window in which the comparison metric is defined. 67

II.14 Relative field at different frequencies as a function of the $k-$value and their respective relative noise level (dashed lines). The relative noise level for each frequency is defined as the ratio of the noise level of the reference with respect to the reference field carried at these frequencies. The two dotted black lines represent the maximum and the minimum relative field value over the 0.2 - 2 THz spectral range observed at each $k-$value. Due to the dispersive behavior of coating layer, the difference between the field carried at low frequencies and the field carried at high frequencies increases as k increases. 68

II.15 Top: Reconstruction of the signal provided by the reduced iterative tree algorithm (black cross) with the 33 selected contributing paths (red). The correlation is almost perfect between the two signals over the entire time window. Bottom: $\Lambda(t)$ metric. Differences are observed in regions where additional subdivided pulses have been ommitted (*e.g.* $k > 17$). 69
II.16 Representation of the main optical paths sorted by the iterative tree procedure for k-index ranging from 7 through 17. The optical paths at each k-values are sorted from left to right from the highest contribution to the signal to the lowest. Particular optical paths exhibit a symmetric one represented in dashed lines. These symmetric paths participate in an equivalent fashion to the signal. 70
II.17 $\sum_{t=0}^{t_{max}} |S^M(t) - S^S(t)|^2$ as a function of the tested thicknesses contained in the d vector. The global minimum was found to be 63.4 μm. 72
II.18 Time-domain error of the thickness global minimum using the Λ metric. 73
II.19 Optical properties $n(\omega)$ and $\kappa(\omega)$ for 63.4 μm. 73
II.20 1st order Debye (*i.e.* $n = 1$ equation I.22.) fit of the IEP extracted properties. 74
II.21 Reconstructed waveform after fitting the optical properties. 75
II.22 Refractive index and extinction coefficient of the sapphire substrate. 76
II.23 Typical time-domain waveform recorded without (black) and with (red) water film on the top side on the sapphire substrate. 77
II.24 Water optical properties extracted in reflection geometry, averaged over ten independent measurements. 77
II.25 Cole-Cole plot of water and approximated representation of characteristic relaxations domains. In black dashed lines are schematically represented the portion of the semi-circle that describes the double-Debye relaxation nature. Frequency axis decreases from 200 GHz to 1 THz from right to left. 78
II.26 Residuals at each ω with respect to both the real and complex plane for 2-Debye terms and 3-Debye terms. 79
II.27 Fit to 2-Debye terms model and the experimental $n(\omega)$ and $\kappa(\omega)$. 80
II.28 Frequency dependence of the refractive index and the extinction coefficient of each samples. 82
II.29 Fitting of the dispersion profiles of each phantoms with respect to II.33. Blue: #1 phantom; Green: #3 phantom; Yellow: #4 phantom; Purple: #5 phantom. Black dashed line represents the Λ metric. 85
II.30 Fitting of the real part of the dispersion profiles of each phantoms. Blue: #1 phantom; Green: #3 phantom; Yellow: #4 phantom; Purple: #5 phantom. Black dashed line represents the Λ metric. 87
II.31 Λ metric on the comparison between Truong *et al* model and the modified-Lorentz one for phantoms. Blue: #1 phantom; Green: #3 phantom; Yellow: #4 phantom; Purple: #5 phantom. Black dashed line represents the Λ metric. 89
II.32 I.I and K.I samples sandwiched between two sapphire substrates to prepare the tissue measurement. 91
II.33 Refractive index and extinction coefficient for malignant lesions, healthy fibrous and adipose tissues. 91
II.34 Refractive index and extinction coefficient difference between adipose and healthy fibrous as a percentage of healthy fibrous characteristics. 93
II.35 Refractive index and extinction coefficient difference between healthy fibrous and cancer as a percentage of cancer characteristics. 94
II.36 Fitting of the dispersion profiles of the mean response of cancer (solid red), fibrous (solid blue) and adipose tissue (solid green). Dashed black line represents the Λ metric between measurement and fitting. 96
II.37 Averaged distribution of model parameters of each poles for tissue types. 98
II.38 Sealed liquid cell. 100

II.39 Dispersion profiles of the real and imaginary part of the dielectric constant for each cell concentration. . . . 101
II.40 Variations in ϵ_∞, $\Delta\epsilon_1$ and $\Delta\epsilon_2$ as a function of MCF7 cell concentration. . . . 103
II.41 Contribution of $\epsilon\infty$, ϵ_1 and ϵ_2 to the dielectric constant dispersion profiles. . . . 104

III.1 Schematic representation of the information volume acquired during 2D material imaging. The third axis, t represents the time. . . . 111
III.2 Abbe's limit between 200 GHz and 1 THz in free space for $\theta_0 = \frac{\pi}{6}$. . . . 112
III.3 Resolution test-target. . . . 112
III.4 Minimum resolved spatial frequency corresponding to at least a contrast of 30% as a function of the sensing terahertz frequency. . . . 113
III.5 Smallest resolved line pairs (dashed red square) for 200, 400, 600, 800 and 1000 GHz. The smallest resolved spatial frequency at 200 GHz was assumed to be half the spatial frequency of group -2, element 1. . . . 114
III.6 Flow chart to illustrate the procedure to improve sample characterization. R_1 and R_2 depicts the lower and upper reflected pulses at dielectric interfaces. . . . 115
III.7 Schematic illustration of the plane of interest that contains the information about the tissue sample. . . . 116
III.8 Example of a typical pixel that allows the reference retrieval. . . . 116
III.9 Scaling function and shape of the db6 wavelet. . . . 118
III.10 Block diagram of the self-reference method. (i). A retrieval pixel is selected; (ii). Signal inversion is performed to calculate the reference signal; (iii). The reference signal is denoised by means of db6 wavelet. Eventually, the complete signal of the retrieval pixel arising from the interaction of the calculated reference with the two dielectric interfaces is retrieved. Note that the time-scale is relative rather than absolute. . . . 119
III.11 Representation of a pixel with the four associated pointel (0;0), (0;1), (1;1), (1;0) and the four linels joining them. . . . 120
III.12 Illustration of the contouring procedure to reduce the number of pixels treated by the inverse electromagnetic problem. (i). Raw terahertz image of a sample; (ii). Contour levels extraction; (iii). Enclosed surface associated to a selected contour level; (iv). Delineation of the region to be treated by inverse electromagnetic problem. In the example, the number of pixel to be considered was reduced from 101,440 pixels to 56,296 pixels. . . . 121
III.13 Problems associated with 2D histological images. (a) Presence of tears and holes. (b) Global shape deformation and missing information. (c) Artefacts in and around. (d) Glue deposited on sample slide. (e) Local misalignment of tissue substructures. (f) Vignette arising from difference of thickness around tissue edges. Reproduced with the courtesy of Abiodun Ogunleke **[20]**. . . . 122
III.14 Representation of the registration process. The terahertz and pathology original images are treated by a contouring algorithm to extract the edges of each image. The pathology image resolution is reduced to match the lower resolution of the terahertz image. Then, the terahertz image is reoriented with respect to the pathology image until the best correlation is achieved. These modifications are therefore applied to the refractive index map and the pathology mask. . . . 124
III.15 Representation of the refractive index thresholding. a. Schematic refractive index map. b. Refractive index thresholding for $t = 2.4$. c. Refractive index thresholding for $t = 2.1$. d. Refractive index thresholding for $t = 1.8$. . . . 125
III.16 Spreading of an object due to PSF. From left to right: punctual object, PSF and spread object image. . . . 126
III.17 Representation of the class overlapping for the modulus amplitude distribution C_1 (blue) and C_2 (orange). Green fills the overlapping region where it is not possible to clearly define pixel's belonging. Black curve represents the sum of the two distributions and therefore the observed global behavior. . . . 126

III.18 Schematic representation of Σ^1 and Σ^2 structuring elements. Left: Σ^1 structuring element that may also be named local connectivity with four connectivity. Right: Σ^2 structuring element that is not solely local as it consists of twelve connectivity pixels. 128
III.19 Diagram of the classification process from the refractive index map to both the hard and connected thresholding diagnosis. The black dashed arrow on the top of the diagram refers to the inverse electromagnetic problem (IEP) and upstream data treatment performed to derive the refractive index map. Note that it is possible to think about connected thresholding as an expansion of the hard thresholding rather than an independent process. For clarity purposes, the two strategies have been organised in two distinct arms. Black dashed circles highlights 129
III.20 Flow chart of the method followed to classify tissue section terahertz pixels. [1]: Note that the histology treatment is performed after the terahertz imaging. 131
132figure.caption.148
III.22 Receiver operating characteristic for the different classification methods, at 560 GHz. The black line represents TPR = FPR. 133
III.23 TPR – FPR as a function of the refractive index threshold for HT, CT1, CT2 and CT3, at 560 GHz. 134
III.24 NS1 tissue sample classification map at 560 GHz for HT, CT1, CT2 and CT3 and their respective first two best thresholds, at 560 GHz. 136
III.25 Overview of the results obtained via the procedure for the NS2 sample: i. photograph; ii. pathology image with four enlightened regions. A gap is shown in the middle of the section that is due to the histology chain; iii. Pathology mask derived from the pathologist diagnosis. iv. Terahertz image of the tissue sample at 560 GHz; v. Terahertz mask derived from the 560 GHz raw image; vi. Refractive index map calculated from the self-reference procedure and the inverse electromagnetic problem; vii. Raw and denoised retrieval pixel signals used to retrieve the reference signal - relative time-scale; viii. Reconstructed signal from the retrieval denoised pixel signal - relative time-scale. a, b, c and d are pathological view of the regions related in the pathology image (ii) and the raw terahertz image (iv). 137
III.26 Receiver operating characteristic for the different classification methods, at 560 GHz. The black line represents TPR = FPR. 138
III.27 TPR – FPR as a function of the refractive index threshold for HT, CT1, CT2 and CT3, at 560 GHz. 139
III.28 NS2 tissue sample classification map at 560 GHz for HT, CT1, CT2 and CT3 and their respective first two best thresholds, at 560 GHz. 141
III.29 Overview of the results obtained via the procedure for the NS3 sample: i. photograph; ii. pathology image with four enlightened regions. A gap is shown in the middle of the section that is due to the histology chain; iii. Pathology mask derived from the pathologist diagnosis. iv. Terahertz image of the tissue sample at 560 GHz; v. Terahertz mask derived from the 560 GHz raw image; vi. Refractive index map calculated from the self-reference procedure and the inverse electromagnetic problem; vii. Raw and denoised retrieval pixel signals used to retrieve the reference signal - relative time-scale; viii. Reconstructed signal from the retrieval denoised pixel signal - relative time-scale. a, b, c and d are pathological view of the regions related in the pathology image (ii) and the raw terahertz image (iv). 142
III.30 Receiver operating characteristic for the different classification methods. The black line represents TPR = FPR, at 560 GHz. 143
III.31 TPR – FPR as a function of the refractive index threshold for HT, CT1, CT2 and CT3, at 560 GHz. 144
III.32 NS3 tissue sample classification map at 560 GHz for HT, CT1, CT2 and CT3 and their respective first two best thresholds, at 560 GHz. 145
III.33 Data compression: from 2D to 1D data set. 148

III.34 Schematic representation of a data matrix to be inserted into the PCA routine. 149
III.35 Integrated loading weighted frequencies for modulus and phase data set for the first three principal components. The reader is referred to Figure III.21 for eyeball comparison. 151
III.36 Modulus data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection. 152
III.37 Phase data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection. 153
III.38 Integrated loading weighted frequencies for modulus and phase data set for the first three principal components. The reader is referred to Figure III.25 for eyeball comparison. 154
III.39 Modulus data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection. 155
III.40 Phase data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection. 156
III.41 Integrated loading weighted frequencies for modulus and phase data set for the first three principal components. The reader is referred to Figure III.29 for eyeball comparison. 157
III.42 Modulus data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection. 158
III.43 Phase data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection. 159

IV.1 Delineation between the near-field domain and the far-field regime. The near-field regime is defined as the sphere of radius $\lambda/2\pi$ whose center is the punctual electromagnetic source. Beyond the near-field sphere, interactions with matter should be regarded as far-field effects. 167
IV.2 Near-field sensing with terahertz source near the sample to be inspected. a: approach developed by Lecaque *et al* in **[5; 6]**. The sample is placed on the emitter. b: coupling between an optical fiber and a $\chi^{(2)}$ crystal as reported by **[7]** 168
IV.3 Near-field sensing with terahertz detector near the sample to be inspected. a: approach developed by **[9]**. The terahertz field induces a refractive index change in the electro-optic crystal onto which is mounted the sample. Polarization ellipticity changes in the counter-propagating probe fs-laser beam are recorded. b: alternative idea suggested by **[11]** where the electro-optic crystal is replaced by H-shaped electrodes patterned on a silicon layer. 170
IV.4 Near-field sensing with both terahertz electro-optic source and detector close to the sample to be inspected **[12]**. 170
IV.5 Near-field sensing with illuminated sub-wavelength aperture. Note that the sample may either be upstream or downstream the sub-wavelength aperture. a: the terahertz beam passes through the sample and subsequently through a sub-wavelength gold aperture. The electric component of the transmitted field is detected by a low temperature grown GaAs photoconducting antenna **[15]**. b: the incident terahertz beam is sent onto a bull's eye structure that activates surface waves. The resonant excitation of those waves allow to increase the amplitude of the electric field that is detected. Terahertz passes through a bow-tie aperture and interacts with the investigated sample **[17]**. 171

IV.6 Scheme of a S-NSOM setup. A terahertz beam is focused onto the sample surface by means of lenses. The evanescent field is probed with an oscillating scatterer that vibrates at the frequency of the field to be recorded. The measurement of the near-field suffers from the inavoidable presence of a strong far-field background. 173

IV.7 Conceptual design of a near-field device that senses the evanescent field arising from the interaction of terahertz radiations with a sample (right). The device is aimed to consist of an illumination source, a near-field surface sensor and a high-responsivity terahertz power detector, with all mounted on an unique chip. Thanks to subwavelength resolution, features whose size is smaller than the probing wavelength can be resolved. [a]: the smallest size of features than can be resolved depends on the near-field device performances. In comparison, far-field probing (left) does not allow to resolve sample features smaller than $\lambda/2$. 175

IV.8 Representation of different split-ring resonator geometries. a: circular SRR; b: double circle SRR; c: complementary spiral SRR; d: square SRR; e: U-shaped SRR; f: dual-square SRR. 176

IV.9 Schematic representation of a split-ring resonator probe. The resonator behaves like a LC-component whose characteristic frequency is about to be shifted in the presence of a sample, at the neighborhood of the resonator gap. 176

IV.10 Block-diagram of the measurement operation. The resonance shift is interpreted as amplitude transmission changes from opaque to lucent at the terahertz power detector output. 177

IV.11 Sensor scheme. a: 3/4 view of the complete sensor made of a transmission line, and cross-bridged dual-square SRR; b: top view of the resonator with L = 19 µm, W = 4.5 µm, and S = 3 µm. 178

IV.12 S_{21} parameter profile between 400 GHz and 600 GHz highlighting the presence of a nominal resonance frequency ($\frac{\omega_0}{2\pi}$) at 540 GHz. The sensor illumination ($\frac{\omega_{osc}}{2\pi}$) is proceeded at 560 GHz. 179

IV.13 Profile of S_{21} as a function of the PEC distance from the surface. Left: Highlight of the induced frequency shift due to the PEC relative altitude. Right: Bi-exponential behavior of S_{21} as a function of the PEC relative distance at 560 GHz, with α_1 = -6.273, α_2 = 0.099, β_1 = 4.082 and β_2 = -1.189. 180

IV.14 Profile of S_{21} as a function of different ϵ'. Left: Highlight of the induced frequency shift due to the progressive increase of ϵ' . Right: Bi-exponential behavior of S_{21} as a function of the ϵ' relative permittivity at 560 GHz, with α_1 = -6.675, α_2 = -0.488, β_1 = -5.755 and β_2 = -0.038. 181

IV.15 Profile of S_{21} as a function of different ϵ''. Left: Highlight of LC-circuit response due to the progressive increase of ϵ'' . Right: Bi-exponential behavior of S_{21} as a function of the $\tan\delta$ relative loss angle at 560 GHz, with α_1 = -9.677, α_2 = -0.029, β_1 = 0.041 and β_2 = -10.55. 181

IV.16 Slope as defined in equation IV.7 as a function of $tan\delta$ with α_1 = 0.106, α_2 = -1.761, β_1 = 0.255 and β_2 = -0.371. 182

IV.17 Profile of S_{21} as a function of ϵ' and ϵ''. Left: Highlight of LC-circuit response due to changes in ϵ' and ϵ'' . Right: Map of the S_{21} value at 560 GHz for $\epsilon' \in [1;6]$ and $\tan\delta \in [0;0.5]$. 182

IV.18 Profile of S_{21} for water methanol, ethanol and propanol. Left: Highlights of the LC-circuit response to polar liquids. The inset depicts the values of ϵ' and ϵ'' for each sample at 560 GHz as derived from the parameters in Table IV.1. Right: Behavior of S_{21} at 560 GHz as a function of the considered liquids. Simulated differences in S_{21} between the different samples are highlighted. 183

IV.19 Profile of S_{21} for water, BCC, normal and adipose tissue. Left: Highlights of the LC-circuit response to biological tissues. The inset depicts the values of ϵ' and ϵ'' for each tissue types and water at 560 GHz as derived from the parameters in Table IV.3. Right: Behavior of S_{21} at 560 GHz as a function of the considered tissues. Simulated differences in S_{21} between the different samples are highlighted. 185
IV.20 Micrograph of the near-field sensing array. The sensing strip length is 3.2 mm long. The black region contains additional circuits that are not associated to the near-field sensing. ASIC: application-specified integrated circuits. ADC: analog-to-digital converted. Refgen: reference generator. BP: bandpass. 186
IV.21 Block diagram of the system on a chip with the required external components. The chip contains 2 rows of 64 pixels, each divided into 16 subarrays of 4 entities illuminated by a common oscillator and an on-chip lock-in amplifier read-out. Pixel selection (row,col)) is achieved via an application-specific integrated circuit (ASIC). $adlc_{clk}$: analog-to-digital converter clock. FADC: flash analog-to-digital converter. GND: ground. LIref: lock-in reference. OSC: oscillator. TP: test point. UART: universal asynchronous receiver/transmitter. VDD: digital supply voltage. VCC: analog supply voltage. 187
IV.22 Photograph of the packaged imaging module. a: 3/4 view of the module with 1-euro coin for scale. b: top view of the module, highlighting the sensing line, the spacer and the wire bonds encapsulated in epoxy resin. c: zoom on the system on-chip (x20). 188
IV.23 Surface profile analysis of a small fraction of the sensing strip. Strip lines are elevated by about 12µm. Such a topology makes it difficult to accurately sense stiff material as a perfect control of the altitude and the flatness with respect to the sensing line are required. 190
IV.24 Droplet deposited onto the top of the sensing line. 190
IV.25 Response of the near-field array pixels as a function of the deposited liquids to that of free space. a: global matrix response to water. b: global matrix response to methanol. c: global matrix response to ethanol. d: global matrix response to propanol. Logarithmic scale in arbitrary units. The reader is invited to note that the scale is logarithmic rather than linear to highlight matrix response difference between liquids. 191
IV.26 Pixel response to water (blue), methanol (red), ethanol (green) and propanol (yellow). The mean level of propanol is slightly lower than that of ethanol. The mean level of the matrix response to methanol is higher than for other liquids. Water stands in the middle as a result of the impact of its lossy part of permittivity. 192
IV.27 Biometric fingerprint data comparison [**35**]. a: ink-and-paper fingerprint. b: 560 GHz near-field image of the equivalent finger region. Points 1, 2 and 3 depict the presence of ridge splitting that are well retrieved from ink-and-paper fingerprint to terahertz image. 193
IV.28 Terahertz near-field imaging of a deparaffinized breast section [**35**]. a: micrograph of the deparaffinized imaged breast section; b: height profile of the sample; c: terahertz near-field image at 560 GHz with the section in contact with the array. 194

List of Tables

I.1 Dielectric properties of some alcohols and water at 70 GHz. [**28**] 19

I.2 Comparison of the dielectric Debye parameter for water molecules found in the literature.
[a] Parameter fixed during the Levenberg–Marquardt fitting procedure [**41; 47**]. [b] Includes intermolecular libration amplitude. 27

I.3 Relative mass (%) of water, lipid, protein, carbohydrate for various body tissues [**66**].
[a]: Approximations are particularly schematic for breast composition since component content varies among individuals and tissue excision location [**67**]. In addition, significant variations in composition have been observed for adipose tissue and mammary glands as reported by [**66**]. 33

I.4 Cancer classification system. Cancers are sorted by histological appearance. Depending on the histological type, they are classified as ductal or lobular carcinoma. Carcinoma *in situ* denotes cancers which have not spread out the initial tissue compartment, while invasive carcinoma have spread in surrounding tissues. The grade states for the differentiation level of cells. Tissues exhibiting cells least like normal ones are by and large high grade carcinomas. 41

II.1 Properties of some ultrafast photoconductive substrates for pulsed terahertz emission. 56

II.2 Characteristics of the TPS Spectra 3000 and the Terapulse 4000. 61

II.3 Fitting Debye parameters for the four coatings of the calibration sample. ϵ_∞ stands for the dielectric constant at high frequencies, $\Delta\epsilon$ is the permittivity difference between the static dielectric constant and the relaxation process permittivity, occurring over the time τ. d stands for the thicknesses extracted from optical measurement. . 65

II.4 Number of computed paths for each k-incrementation and the number among them outgoing the structure being thus detected. 68

II.5 Fitting Debye parameters for the unknown sample. 74

II.6 Fitting parameters for 2-Debye terms and 3-Debye terms. 79

II.7 Fraction of protein, oil and water in each phantom in percent-by-weight. 81

II.8 Fitting parameters of each phantom with respect to equation II.33. 84

II.9 Fitting parameters of each phantom for (II.36a). 87

II.10 Parameters of the model proposed by Truong *et al.* aR^2 is the adjusted R-square value denoting the square of the correlation between the response values and the simulated ones. When equal to unity, the adjusted R-square value denotes a perfect correlation while being equal to 0 it denotes a null correlation. BCC: basal cell carcinoma. 88

II.11 List of the fifteen freshly excised tissues analyzed in reflection spectroscopy. Samples having the same letter reference code belong to the same patient and have been excised during the same procedure. The diagnosis column refers to the pathology from which is suffering the patient. ILC: invasive lobular carcinoma; IDC: invasive ductal carcinoma; MBC: metaplastic breast cancer; BPT: breast phyllodes tumor; BR: breast reduction. 90

II.12 Fitting parameters for the mean complex dielectric function of each tissue types for (II.33). . . . 95
II.13 Comparison of ϵ_∞. . . . 97
II.14 MCF7 cell concentrations. . . . 100
II.15 Debye parameter for the best fitted relaxation strengths. . . . 102

III.1 Thickness and spatial frequency of each element of the USAF-1951 test-chart. . . . 113
III.2 Compliance decision rule between the terahertz image and the pathology one. . . . 130
III.3 Area under the curve for each classification method. . . . 133
III.4 First two best refractive index threshold characteristics considering the trade-off between TPR and FPR. . . . 134
III.5 Area under the curve for each classification. . . . 138
III.6 First two best refractive index threshold characteristics considering the trade-off between TPR and FPR. . . . 139
III.7 Area under the curve for each classification. . . . 143
III.8 First two best refractive index threshold characteristics considering the trade-off between TPR and FPR. . . . 144
III.9 Contribution to variance (%) for the first six principal components based on the modulus and the phase of each pixel between 200 GHz and 1 THz. . . . 150
III.10 Contribution to variance (%) for the first six principal components based on the modulus and the phase of each pixel between 200 GHz and 1 THz. . . . 154
III.11 Contribution to variance (%) for the first six principal components based on the modulus and the phase of each pixel between 200 GHz and 1 THz. . . . 157

IV.1 Debye parameter of water, methanol, ethanol and propanol as used to simulate their respective dispersion profiles. . . . 183
IV.2 Simulated sensor response to water, methanol, ethanol and propanol. Values are given with respect to that of free space. $P_{0'}/P_0$ denotes the gain in transmitted power for a particular sample with respect to that of free space. . . . 184
IV.3 Parameters of the model proposed by Truong *et al.* aR^2 is the adjusted R-square value denoting the square of the correlation between the response values and the simulated ones. When equal to unity, the adjusted R-square value denotes a perfect correlation while being equal to 0 it denotes a null correlation. . . . 185
IV.4 Simulated sensor response to basal cell carcinoma (BCC), normal fibrous and healthy adipose. Values are given with respect to those of free space. $P_{0'}/P_0$ denotes the gain in transmitted power for a particular sample with respect to that of free space. . . . 185
IV.5 Simulated sensor response to water, normal fibrous and healthy adipose. Values are given with respect to those of **basal cell carcinoma** (BCC). $P_{0'}/P_{BCC}$ denotes the change in transmitted power for a particular sample with respect to that of BCC. . . . 186
IV.6 Permittivity lossless and lossy part for water, methanol, ethanol and propanol at 560 GHz from [**36**] and this work. . . . 190

Acknowledgments

Je souhaiterais avant tout remercier Dr. Patrick Mounaix, mon directeur de thèse, pour m'avoir laissé me tromper. C'est ainsi -je crois- que j'ai pu tant apprendre et ne jamais cesser de questionner.

Mes sincères remerciements aux Dr. Jean-Paul Guillet, Dr. Thomas Zimmer et Dr. Gaëtan MacGrogan, qui m'ont tous trois permis de progresser dans chacun de leur domaine d'expertise respectif. Au même titre, j'aimerais remercier M. Frédéric Fauquet, pour sa gentillesse et son enthousiasme permanent à l'égard de mon travail.

Mes pensées vont également au Dr. Christine Tunon de Lara, engagée au sein de ce projet en tant qu'experte en chirurgie du sein.

Mein Dank gilt auch dem gesamten Forschungsteam in Wuppertal, unseren Partnern in diesem Projekt und insbesondere Prof. Ullrich Pfeiffer und Philipp Hillger für ihre Erläuterungen und Hilfe bei der Nutzung der Nahfeldmatrix.

Je souhaiterais également remercier le Dr. Guilhem Gallot et le Dr. Frédéric Garet pour avoir acceptés de participer au processus d'évaluation de mon mémoire de thèse en tant que rapporteurs, ainsi que le Dr. Romain Peretti pour le temps accordé à mon travail en tant qu'examinateur de ces travaux.

Je souhaite également remercier M. Samuel Caravera, technicien supérieur à l'institut Bergonié, centre régional de lutte contre le cancer, pour son enthousiasme et sa patience à l'égard de mes demandes concernant les échantillons excisés, même les plus farfelues. C'est également l'ensemble des équipes à Bergonié qu'il me faut remercier, je pense à Marie-Claude, Reynlad, Fabien, Valérie et Stéphane.

Il me faut adresser mes sincères remerciements à Mme. Florence Poulletier de Gannes, pour sa patience et son aide concernant la culture des cellules MCF7.

Mes pensées vont à Amel Al-Ibadi, Joyce Bou-Sleiman et Corinna Koch-Dandolo, avec qui j'ai pu, de près ou de loin travailler sur diverses thématiques.

Хочу выразить искреннюю благодарность сотрудникам Международного института «Фотоники и оптоинформатики» Университета ИТМО, которые работают в области исследования терагерцового поля. За их гостеприимство, отзывчивость и профессионализм. Особую благодарность выражаю Ольге Алексеевне Смолянской, с которой мне удалось сотрудничать во время написания диссертации.

日本・大阪大学の斗内政吉教授のTHz研究チームとは、乳癌のTHz近接場イメージングに関して短いながらも充実した共同研究を行わせていただき、ここに厚く御礼を申し上げます。特に本共同研究の中心メンバーである岡田航介氏に深く感謝申し上げるとともに、彼の博士号取得の成功を願います

Je souhaiterais également remercier Simone, Fabienne, Nicolas, Nathalie, Laëtitia et l'ensemble des équipes administratives du laboratoire IMS.

Sur des plans plus amicaux, c'est l'ensemble des doctorants et amis du laboratoire IMS qu'il me faut maintenant remercier, je pense à Marine, Florent Alb., Matthieu, Mathieu, Adrien, JB, Djeber, Soumaya, Florent Abd., Jean-Marie, Thomas, Isabelle, Margaux, Yuanci, Marco, Olivia...

Mais aussi, mes amis d'enfance, Tom, Andréas, Léa, Matthieu, Sofiane, Sébastien, Mathilde, Baptiste, Jean-Melchior, Alexandre, Florent, Marie-Camille, Ludovic, Rémi, Manu...

Je souhaiterais aussi remercier les équipes du planétarium de Lyon, qui m'ont données l'opportunité d'enseigner dès mon plus jeune âge: Adrien, Walter, Pierre, Simon, Philippe, Matthieu, Serge,

Alex...

Mes amis du handball, pour les soirées "handiablées": Aurore, Laurence, Antoine, Franck, Maud, Véronique, Vincent, JB, Adrian...

Mes pensées à mes élèves, à l'IUT GEII de Bordeaux qui n'ont eu de cesse de me surprendre, jusqu'à scander mon nom dans les tramways bordelais.

Mon colocataire et ami, Romain Matthieu, pour tous ces moments de détente ô combien forts nécessaires durant cette épreuve.

Mes pensées vont à mes parents, pour leurs amours inconditionnelles et leur support sans mesure, sans lesquels, rien de tout cela n'aurait jamais été réalisable. Mais également, mon frère Benjamin et ma soeur Auriane, pour notre merveilleuse enfance, pleine de complicité et d'amour. Et si l'amour se veut, sur ces lignes maître, il me faut remercier mes grands-parents qui n'ont eu de cesse d'alimenter ce ruisseau d'amour et de bienveillance à mon égard, de ma naissance jusqu'à aujourd'hui.

最后的最后, 我所有的思念, 爱怜和温柔都涌向明明。在读博的三年里,你给予了我无与伦比的陪伴,我们一起克服了所有的困难。我们紧握双手，把挡在我们面前的每一个障碍都推倒，对彼此没一丝迟疑对方。我希望并且相信, 从这场考验开始, 我们的余生紧密地连接在一起。明明，我爱你。

Acronyms

A

- **ADC**: Analog-to-Digital Converter;
- **A-NSOM**: Aperture-Near-field Scanning Optical Microscopy;
- **ASIC**: Application Specified Integrated Circuits;
- **ATR**: Attenuated Total Reflection;
- **AUC**: Area Under the Curve;

B

- **BCC**: Basal Cell Carcinoma;
- **BCS**: Breast Conserving Surgery;
- **BP**: Band-Pass;
- **BS**: Beam Splitter;

C

- **CT1**: 1^{st}-order Connected Thresholding;
- **CT1**: 2^{nd}-order Connected Thresholding;
- **CT1**: 3^{rd}-order Connected Thresholding;

D

- **dB**: Decibels;
- **DFT**: Direct Fourier transform;
- **DNA**: Deoxyribonucleic Acid;

E

- **EMF**: Electromotive Force;

F

- **FADC**: Flash Analog-to-Digital Converter;
- **FDTD**: Finite Difference Time Domain;
- **FFT**: Fast Fourier Transform;
- **FIT**: Finite Integration Technique;
- **FPGA**: Field-Programmable Gate Array;
- **FPR**: False-Positive Rate;
- **FT**: Fourier Transform;

G

- **GHz**: Giga-Hertz;
- **GND**: Ground;

H

- **HT**: Hard Thresholding;

I

- **IEP**: Inverse Electromagnetic Problem;
- **ITA**: Iterative Tree Algorithm;

K

- **KHz**: Kilo-Hertz

L

- **LC**: Inductance-Capacitance;
- **LT-GaAs**: Low-Temperature grown Gallium Arsenide;

M

- **MCF7**: Michigan Cancer Foundation - 7;
- **MDCK**: Madin-Darby Canine Kidney;
- **MHz**: Mega-Hertz

N

- **NA**: Numerical Aperture;
- **NIPALS**: Nonlinear Iterative Partial Least Squares;
- **NMR**: Nuclear Magnetic Resonance;
- **NSOM**: Near-field Scanning Optical Microscopy;

O

- **OSC**: Oscillator;

P

- **PC**: Principal Component;
- **PCA**: Principal Component Analysis;
- **PCA**: Photoconductive Antenna;
- **PEC**: Perfect Electric Conductor;
- **PSF**: Point Spread Function;

R

- **RD-SOS**: Radiation-Damaged Silicon-on-Sapphire;
- **RNA**: Ribonucleic Acid;
- **RoI**: Region of Interest;
- **ROC**: Receiver Operating Characteristics;

S

- **S-NSOM**: Scattering-Near-field Sensing Optical Microscopy;
- **SRR**: Split-Ring Resonator;
- **SVD**: Singular Value Decomposition;

T

- **TDS**: Time-Domain Spectroscopy;
- **THz**: Tera-Hertz;
- **TP**: Test-Point;
- **TPR**: True-Positive Rate;

U

- **UART**: Universal Asynchronous Receiver/Transmitter;
- **USAF-1951**: United-States Air Force - 1951;

V

- **VCC**: Analog Supply Voltage;
- **VDD**: Digital Supply Voltage;

Symbols

Chapter I.

E: Photon energy;
h: Planck constant;
ν: Frequency;
ω: Angular frequency;
t: Transmission coefficient;
r: Reflection coefficient;
$\hat{\boldsymbol{n}}$: Complex refractive index;
n: Real refractive index;
κ: Extinction coefficient;
α: Absorption coefficient;
δ: Penetration depth;
$\vec{\boldsymbol{k}}$: Wave vector;
$\vec{\mathbf{E}}$: Electric field vector;
$\vec{\mathbf{D}}$: Dielectric field vector;
$\hat{\boldsymbol{\epsilon}}_{\boldsymbol{r}}$: Complex rel. permittivity;
$\boldsymbol{\epsilon_0}$: Free-space permittivity;
$\boldsymbol{\epsilon}'_{\boldsymbol{r}}$: Real rel. permittivity;
$\boldsymbol{\epsilon}''_{\boldsymbol{r}}$: Imaginary rel. permittivity;
$\boldsymbol{\epsilon_d}$: Dipole losses;
σ: Conductivity losses;
$\boldsymbol{\epsilon_\infty}$: ∞-frequency permittivity;
Δϵ: Permittivity step;
τ: Characteristic time;
Γ: Damping factor;
$\mathbf{A_S}$: Intermolecular stretching;
$\mathbf{A_L}$: Intermolecular libration;

Chapter II.

$\mathbf{E_{THz}}$: Scalar THz field;
$\boldsymbol{q}$: Elementary charge;
ν: Charge velocity / frequency;
$\boldsymbol{n}$: Charge density;
$\boldsymbol{t}$: Time;
J: Transient current;
$\vec{\mathbf{E}}$: Electric field vector;
$\vec{\mathbf{P}}$: Polarization vector;
$\boldsymbol{\epsilon_0}$: Free-space permittivity;
$\boldsymbol{\chi}^{(n)}$: n_{th} order non-linearity;
ω: Angular frequency;
$\boldsymbol{f_s}$: Sampling frequency;
$\boldsymbol{f_{max}}$: Maximum frequency;
$\boldsymbol{\Delta x}$: Delay-line displacement;
$\boldsymbol{c}$: Light velocity in free-space;
$\mathbf{S_{THz}}$: THz signal;
$\mathbf{S}'_{\mathbf{THz}}$: Recorded THz signal;
$\mathbf{R_{PCA}}$: PCA response;
$\mathscr{F}$: Fourier operator;
$\boldsymbol{c_n}$: n_{th} Fourier coefficients;
$\boldsymbol{\sigma_s}$: Modulus variance;
$\boldsymbol{\sigma_e}$: Emitter noise;
$\boldsymbol{\sigma_d}$: Shot noise;
$\boldsymbol{\sigma_i}$: Other sources of noise;
ρ: Dielectric properties;
T: Transfer function;
$\mathbf{S}_{\boldsymbol{sam}}$: THz sample signal;
$\mathbf{S}_{\boldsymbol{ref}}$: THz reference signal;
$\boldsymbol{\phi_{sam}}$: THz sample phase;
$\boldsymbol{\phi_{ref}}$: THz reference phase;
$\mathbf{H}_{\boldsymbol{t}}$: Transmission function;
$\mathbf{H}_{\boldsymbol{r}}$: Reflection function;
M: Propagation matrix;
$\boldsymbol{i \longrightarrow j}$: i; j dielectric interface;
$\boldsymbol{t}$: Transmission coefficient;
$\boldsymbol{r}$: Reflection coefficient;
$\boldsymbol{k}$: Pulse internal subdivision;
Λ: Discrepancy metric;
χ: Objective function;
δM: Amplitude penalty;
δϕ: Phase penalty;
ζ: Fibonacci like sequence / regularization parameter;
$\boldsymbol{|Res|}$: Fitting residuals;
$\boldsymbol{\epsilon_\infty}$: ∞-frequency permittivity;
Δϵ: Permittivity step;
τ: Characteristic time;
$\boldsymbol{d}$: Layer thickness;
$\boldsymbol{\epsilon}'_{\boldsymbol{r}}$: Real rel. permittivity;
$\boldsymbol{\epsilon}''_{\boldsymbol{r}}$: Imaginary rel. permittivity;
$\boldsymbol{\Phi}''$: Problem Hessian function;
σ: Conductivity losses;
$\boldsymbol{\alpha_0}$: Fitting term;
$\boldsymbol{\alpha_1}$: Fitting term;
$\boldsymbol{\beta_0}$: Fitting term;
$\boldsymbol{\beta_1}$: Fitting term;

Chapter III.

$\boldsymbol{d}$: Focus diameter / thickness;

$\boldsymbol{\lambda}$: Wavelength;

NA: Numerical aperture;

$\boldsymbol{n}$: Real refractive index;

$\boldsymbol{\theta}$: Convergence angle;

$\mathbf{E_{RP}}$: Retrieval pixel waveform;

$\mathbf{E}_{\boldsymbol{ref}}$: Reference waveform;

$\boldsymbol{t}$: Transmission coefficient;

$\boldsymbol{r}$: Reflection coefficient;

$\boldsymbol{\omega}$: Angular frequency;

$\hat{\boldsymbol{n}}$: Complex refractive index;

$\boldsymbol{c}$: Light velocity in free-space;

X: Reduction of Snell's law;

$\mathbf{N_X}$: Number of pixel in X-axis;

$\mathbf{N_Y}$: Number of pixel in Y-axis;

$\mathscr{R}$: Real number;

$\boldsymbol{s}$: Scalar field;

$\boldsymbol{c}$: Scalar field contour;

$\mathbf{S}_{\boldsymbol{x,y}}$: Scalar field coordinates;

$\boldsymbol{\chi}$: Objective function;

$\boldsymbol{\zeta}$: Regularization parameter;

$\mathbf{I}_{\boldsymbol{h}}$: Image height;

$\mathbf{I}_{\boldsymbol{w}}$: Image width;

$\boldsymbol{\delta_\Sigma}$: Morphological dilation;

$\boldsymbol{\Sigma}$: Structuring element;

$\boldsymbol{\epsilon}$: Structuring element component;

$\mathbf{P}'$: Transposed projection matrix;

$\boldsymbol{s_n}$: n_{th} score;

$\boldsymbol{l_n}$: n_{th} loading;

T: Projection plane;

R: Residual matrix;

$\mathbf{O}_{\boldsymbol{n}}$: n_{th} observable;

Chapter IV.

$\boldsymbol{\eta}$: Wave impedance;

$\boldsymbol{\mu}$: Magnetic permeability;

$\boldsymbol{\mu_0}$: Magnetic permeability of free-space;

$\boldsymbol{\epsilon}$: Electric permittivity;

$\boldsymbol{\epsilon_0}$: Electric permittivity of free-space;

$\boldsymbol{j}$: Imaginary number;

$\vec{\mathbf{E}}$: Electric field vector;

$\vec{\mathbf{H}}$: Magnetic field vector;

$\boldsymbol{\lambda}$: Wavelength;

$\boldsymbol{\chi}^{(n)}$: n_{th} order non-linearity;

$\vec{\mathbf{S}}$: Poynting vector;

$\boldsymbol{z}$: Distance from aperture;

$\boldsymbol{d}$: Aperture diameter;

$\boldsymbol{\omega_0}$: Resonance frequency;

L: Inductance;

C: Capacitance;

$\mathbf{V_0}$: Sample-free detector response;

$\mathbf{V_0'}$: Sample-loaded detector response;

$\mathbf{P}_{\boldsymbol{osc}}$: Oscillator supplied power;

$\mathbf{R_V}$: Detector voltage responsivity;

$\boldsymbol{\Delta T_{0,0'}}$: Transmitted power difference with and without sample;

$\boldsymbol{\omega_{osc}}$: Oscillator angular frequency;

$\mathbf{S_{21}}$: Transmission parameter between port #1 and port #2;

$\boldsymbol{\alpha_1}$: Fitting term;

$\boldsymbol{\alpha_2}$: Fitting term;

$\boldsymbol{\beta_1}$: Fitting term;

$\boldsymbol{\beta_2}$: Fitting term;

$\boldsymbol{\epsilon_r'}$: Real rel. permittivity;

$\boldsymbol{\epsilon_r''}$: Imaginary rel. permittivity;

tan: Loss angle;

$\boldsymbol{\Delta\hat{\epsilon}^{min}}$: Minimum detectable permittivity change;

$\mathbf{V}_{\boldsymbol{n,int}}$: Total noise integrated over the sensor;

$\mathbf{P}_{\boldsymbol{t}}$: Transmitted power;

$\boldsymbol{\epsilon_\infty}$: ∞-frequency permittivity;

$\boldsymbol{\Delta\epsilon}$: Permittivity step;

$\boldsymbol{\tau}$: Characteristic time;

$\boldsymbol{\sigma}$: Conductivity losses;

DR: Dynamic range;

Medical Glossary

A

- **Adenocarcinoma**: Specific type of malignant tumor that arises from the neoplasia of epithelial tissues.
- **Aneuploidy**: Presence of an abnormal number of chromosomes in a cell.
- **Angiogenesis**: Formation of new blood vessels.
- **Apoptosis**: Programmed cell death process in which redundant or flawed cells destroy themselves.

C

- **Carcinogenesis**: Process by which a healthy cell becomes cancerous.
- **Centromeres**: Region where the chromatids of a chromosome meet.
- **Cirrhosis**: Cirrhotic diseases of the liver that progressively destroys the liver's ability to aid in digestion and desintoxification.

D

- **Deoxyribonucleic acid**: Molecule that contains the genome of living organisms.
- **Differentiation**: Process through which a cell changes from on cell type to another.
- **Ductal**: Refers to breast ducts.
- **Duct**: Tube that carries breast milk from the glandular tissue to the nipple.
- **Dysplastic**: Abnormal changes in cells of a tissue. Cells are not cancerous but they may progress to cancer.

E

- **Epithelial**: Cellular structure lining, protecting and enclosing the organs.

F

- **Fibroblast**: Cell that helps the formation of collagen and elastic fiber of connective tissue.

G

- **Gene**: Sequence of nucleotides that encodes the synthesis of a proteins and RNA.
- **Genome**: Referring to the inherited ensemble of chromosomes and genes.
- **Glandes**: Group of cells that produces specific substances.
- **Glandular**: Refers to glandes.
- **Glioblastoma**: Most frequent and most aggressive brain tumor.
- **Grade**: Measure of the cell appearance in tumors and other neoplasms.

H

- **Haploidy**: Presence of a single complete set of chromosome in a cell.

K

- **Keranocyte**: Cell of the epidermis that produces a protein called keratin and forms a soft protective sheet for the body.

L

- **Leukemia** Group of blood cancers that commonly starts in the bone marrow.
- **Liposome**: Spherical vesicle exhibiting at least one lipid layer.
- **Lobule**: Gland that produces breast milk.
- **Lobular**: Refers to lobules.

- **Lumpectomy**: Surgical removal of a discrete portion of the breast.
- **Lymph**: Fluid that flows through the lymphatic network.
- **Lymphatic**: Organ system that is a part of the circulatory and immune system.
- **Lymphocyte**: Type of white blood cell that can recognize foreign substances in the body.

M

- **Mastectomy**: Surgical removal of the breast.
- **Metastasis**: Spread of cancer cells from primary site to another part of the body.
- **Mitosis**: Part of the cell cycle during which replicated chromosomes are divided into two new nuclei.
- **Mucosa**: Tissue that lines the tube-like structures of the body.

N

- **Necrosis**: Premature death of living cells.
- **Neoplasia**: Abnormal growth of tissues either benign or cancerous.
- **Nucleotides**: DNA and RNA building block molecules.

O

- **Orthotopic**: Referring to an event that occurs in the normal or usual place in the body.

P

- **Pleomorphism**: Ability of an organism to modify its intrinsic functions in response to environmental conditions.
- **Polyp**: Noncancerous growth that protudes from mucous membranes.

R

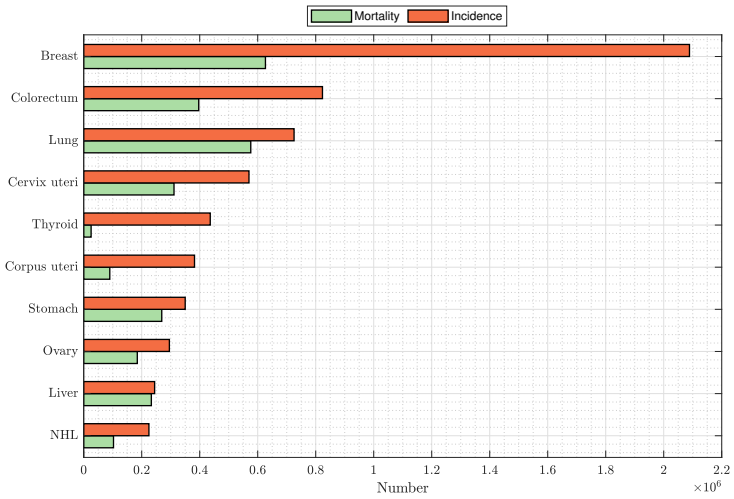
- **Ribonucleic acid**: Molecules that takes part in coding, decoding, regulation and expression of genes.
- **Ribosome**: Cytoplasmic organite where takes place protein synthesis.

S

- **Squamous**: Referring to a flat, scaly epithelial cell.

General Introduction

Breast cancer is the most frequently diagnosed cancer in women. Worldwide, nearly one in four of new cancer cases in women are identified as breast carcinoma. The chart below summarizes the GLOBOCAN survey released by the International Agency for Research on Cancer (IARC) in 2018.



Estimated number of cancer incident cases and deaths in women, worldwide, in 2018. Source: Global Cancer Observatory, 2018.

Breast cancer is not an unique disease, but rather gathers a constellation of distinct subtypes associated with diverse clinical outcomes. Most common tumor initiation and progression occur in cells that line the ducts and the terminal ductal lobular units, that supply the ducts with milk. Despite considerable advances towards diagnosing and curing breast cancer, several crucial clinical and scientific issues have not been entirely resolved. These problems are related to prevention, to diagnosis specificity and sensitivity, to cancer prediction, to the type and to the length of treatment and, to therapeutic resistance. Addressing and resolving all these problems remain drastically challenging as breast cancer is not a single disease but is highly heterogeneous at both the clinical level and the cellular one.

Breast cancer management depends on the disease characteristics established by a set of indicators: (i). the tumor stage; (ii). the tumor grade; (iii) the tumor immunophenotype. Breast conserving surgery, followed by radiation therapy, is nowadays the standard therapy to promote healing for early stage breast cancer. Such a local therapy is often preferred since it is safe and provides an overall equivalent survival rate as mastectomy. In 2005, the international consensus conference defined breast conserving surgery as "[...] complete removal of the breast issue with a concentric margin of surrounding healthy tissue performed in a cosmetically acceptable manner (lumpectomy) usually followed by radiation therapy". It is, however, worth mentioning that women undergoing breast conserving surgery to treat cancer carry slightly greater risk of breast cancer local recurrence, for life. The success of breast conserving procedures is therefore dictated by the accuracy with which are delineated the concentric margins of excised breast volumes during the surgery. Although there is no generic definition of ideal margins, it is common to consider

a 10-mm wide edge, if feasible, followed by postoperative radiation management to eradicate the eventual remaining malignant cells. The cleanliness of edges are assessed during biopsy examination for which, excised tissues are successively fixed into formalin solution, paraffin embedded, sliced in micrometric section and immersed into different alcohol and biological stain baths. Usually, hematoxylin and eosin stains are used as hematoxylin stains cell nuclei blue and eosin stains both the cytoplasm and the extracellular matrix pink. The stain draws the global layout of a tissue structure so that a pathologist can easily state on margin cleanliness. Overall, two extreme cases towards edge definition can be distinguished: (i). positive margins - malignant cells are located at the edge of the excised volume; (ii). negative margins - absence of tumor cells at the edge or the distance of abnormal cells from the edge is at least more than 1-mm. Currently, following histopathologic inspection, up to 20% of excised volumes are reported to exhibit positive margins. Reasons behind tumor edge delineation failure are often due to random specimen for biopsy, discontinuous tumor spread from the original disease site or inappropriate excision during surgery.

With the aim to address the issue of margin positiveness, this work reports investigations on a specific window of the electromagnetic spectrum, the terahertz band, towards its potential for demarcation between healthy and malignant breast tissues. Terahertz radiations have been reported to be particularly sensitive to water content. Indeed, water exhibits relaxation mechanisms in the terahertz band at approximately few tens of gigahertz and 900 GHz, respectively[1]. At the same time, water is the predominant fluid among living organisms. Unifying concepts around carcinogenesis suggested that cell hydration could be on the front lines. All of these reports strongly advert that terahertz waves may help surgeons to accurately define breast tumor margins intraoperatively, based on both terahertz radiations and tumor specificities.

Terahertz waves cover photon energies that are orders of magnitude smaller (0.4-40 meV) than the visual spectrum. Furthermore the low energy carried by such waves indicates that they do not have an ionizing effect and are overall considered to be weakly hazardous for cells, organs and organisms.

This manuscript attempts to answer questions fundamental to the terahertz band's ability to manage demarcation between abnormal and normal breast tissues. The title of this thesis work should be interrogative rather than simply affirmative: *Terahertz radiations for breast tumour recognition ?* Throughout this manuscript different strategies and methodologies are subsequently considered, described, applied and discussed in trying to respond to such a question.

The first chapter *Context and Key Concepts*, establishes the physics that is proving to be useful to study the interaction of electromagnetic waves in the sub-millimeter range with biological entities, whereupon interaction mechanisms from a sub-cellular point of view are discussed. In this regard, the most promising candidates for interaction are reported and a related overview of the research field is provided. The discussion then turns to the super cellular level responsiveness to terahertz stimuli, where the main findings and biomedical application are listed, as exhaustively as possible. Ultimately, the chapter ends on a section dedicated to the specific case of breast carcinoma, from breast anatomy review to cancer treatment, through carcinogenesis.

The second chapter *Terahertz Spectroscopy* begins by introducing the different concept and techniques to generate and to detect terahertz waves. Subsequently, the experimental set-up employed to conduct the investigations on biological samples is presented. Such a system allows to measure the fraction of electromagnetic energy that is either transmitted through or reflected by the sample under inspection. Therefore, important concepts to extract the intrinsic dielectric properties of the different types of tissue found in breast volume are discussed. Some specific contributions to the field are reported and the concept of inverse electromagnetic problem is described as it is omnipresent throughout this work. These tools are then used to extract the dielectric profile of water and the associated relaxations are addressed. As the actual breast environment is complex and definitely not homogeneous, preliminary studies are conducted on tissue mimicking breast phantoms. These samples are aimed to provide generic information on the theoretical behavior of real tissues, by mimicking the actual chemical composition. Specific trends

[1] Note that these values may slightly vary from one study to another.

are reported that tend to suggest that terahertz is globally sensitive to the fraction of protein, fat and water found in tissues. In an attempt to provide a generic formulation to describe the overall dielectric profiles, the grounds of a mathematical model are discussed. Then, based on the characteristics exhibited by tissue mimicking breast phantoms, a campaign of measurements on freshly excised breast tissues conducted at Bergonié Institute (Bordeaux, FRANCE) is reported. Global dielectric profiles of adipose, fibrous and malignant tissues are reported. Clear evidence of demarcation between adipose tissue and other types are reported. Subtler changes in optical constants are additionally underlined between healthy fibrous and malignant tissues. Ultimately, to understand the source of contrast between tissue types, investigations on different concentration of cancer cells in solution are reported and some possible explanations towards tissue contrast are postulated.

The third chapter *Terahertz Imaging: Far-field Survey* is dedicated to the acquisition of terahertz images of freshly excised breast tissues in far-field. First and foremost, the performances of the imaging set-up are identified and sorted. Then, a self-reference image inversion procedure is described that aims to improve sample characterization. From raw images, the described process and inverse electromagnetic problem, the refractive index map of different tissues are extracted. The refractive index derived over the sample surface are then aimed to be employed to segment within the terahertz image. Image registration and correlation with pathology images are detailed. Thereafter different classification strategies are investigated and their relative performances in terms of sensitivity and specificity are reported. Compliance map between pathology images and terahertz classifications are provided. Last, to complete the field of inquiry, principal component analysis, that is an unsupervised data analysis, is investigated as a potential additional tool towards segmentation. Reported results on principal component analysis are in good agreements with refractive index classification.

The last chapter *Breast Cancer Imaging: Near-field Survey* reports the development of a terahertz near-field solid-state imaging matrix working at a frequency that is identified in preceding chapters as promising towards tissue differentiation. The chapter begins by pointing out the diffraction limit of resolution that severely affect far-field imaging systems. The state-of-the-art of the near-field research domain is discussed, whereupon the conceptual working principle of a resonator-based permittivity-sensitive sub-wavelength near-field imager is described. Such a device is aimed to break the diffraction limit of resolution and to provide an image resolution that is overall at the scale of the typical size of eukaryote cells. A battery of simulations are reported that support the idea that such a device could effectively break the diffraction limit while remaining sensitive to tissue dielectric behavior. The task is however reported to be challenging as only subtle changes in response of such an imager between healthy and cancer are anticipated. The fabrication process and the overall architecture of the solid-state array are described and its intrinsic imaging performances are provided. Preliminary studies with the sub-wavelength near-field imaging sensor are reported and an image of a deparaffinized breast section concludes this manuscript.

The work reported in this manuscript takes part in collaborative projects: the *Far and Near-sense projects*, respectively funded by Région Nouvelle-Aquitaine and the German Research Foundation as a part of the Priority Program ESSENCE (SPP 1857). Three entities were part in these joint projects and have to be acknowledged. Studies on interaction mechanisms between terahertz radiations and biological tissues were conducted at the University of Bordeaux (FRANCE) under the supervision of Dr. Dir. Rrch. Patrick Mounaix (DR). The development of the near-field imager was addressed by the University of Wuppertal (GERMANY) under the guidance of Prof. Dr. rer. nat. Ullrich Pfeiffer. Finally, the medical background for the study was assured by the department of pathology at Bergonié Institute (Bordeaux, FRANCE) under the watch of Dr. Gaëtan MacGrogan.

Chapter I

Context and Key Concepts

They didn't know it was impossible, so they did it.

Mark Twain

Sommaire

I.1 Introduction . . . 14
I.2 Terahertz Radiations . . . 14
I.2.1 Physical Features . . . 14
I.2.2 Safety Issues . . . 15
I.3 Terahertz Field and Biological Samples . . . 15
I.3.1 Physical Quantities in Electromagnetism . . . 15
I.3.2 Interaction Mechanisms . . . 19
I.3.3 Candidates for Interaction: Sub-cellular Level . . . 25
I.3.4 Super Cellular Level Responsiveness . . . 32
I.4 Breast Cancer . . . 39
I.4.1 Breast Anatomy & Malignant Illness . . . 39
I.4.2 Treatments . . . 40
I.4.3 Terahertz Waves for Breast Cancer Detection . . . 41
I.5 Conclusion . . . 43
I.6 Bibliography . . . 44

I.1 Introduction

The first chapter *Context and Key Concepts* prepares the ground for the study of the interaction between terahertz radiations and breast cancer tissues, to which this thesis is dedicated.

A brief statement on the physical characteristics of sub-millimeter waves is established, whereupon safety surveys on their use are reported.

Due to those specific properties, these radiations are expected to be a potential tool for advances in biology. Consequently, the subsequent section reviews the physical concepts in electromagnetism useful to study the interaction between external terahertz fields and biological objects. In light of these considerations, past and ongoing scientific efforts on the application of terahertz waves for biology are discussed.

In order to address the specific case of breast cancer, preliminary knowledge on breast structure and oncology are required. To this aim, a succinct review is provided and the prospective sources of tissue distinction in the terahertz regime are raised.

I.2 Terahertz Radiations

Sub-millimeter waves are known since the end of the XIXth century thanks to Samuel Langley and the invention of the bolometer **[1]**. However, it was not until the early-*1900s* that such radiations will begin to be employed with the introduction of effective, compact, powerful and broadband sources and detectors. Mainly used at the beginning for plasma diagnostics **[2]**, molecular spectroscopy **[3]** and astrophysics **[4; 5]**, it took a decade before their apparition in the field of biology and medicine. To understand the current motivations for the implementation of this particular region of the electromagnetic spectrum in various research fields, an introduction to its peculiar characteristics is primarily given.

I.2.1 Physical Features

Terahertz regime, also called far-infrared, is between the microwave and the infrared domain. It is commonly considered that frequencies spanning from 100 GHz to 10 THz belong to the terahertz spectral range (Figure I.1.).

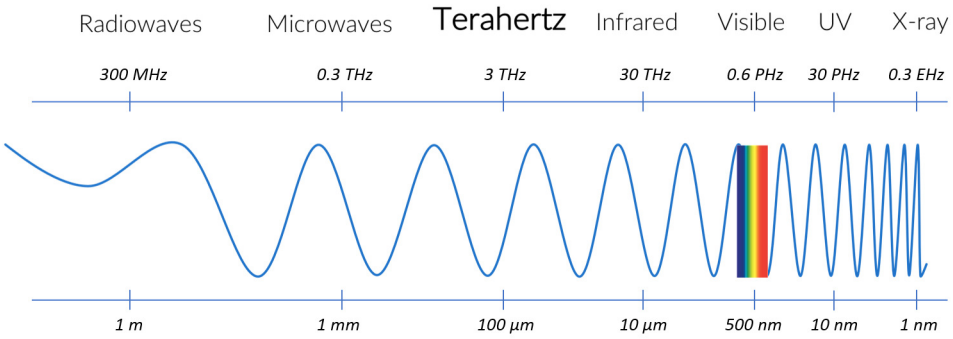


Figure I.1: Location of the terahertz regime within the electromagnetic spectrum.

Still, in some studies, these boundaries are extended to a lower limit in the gigahertz range down to 60 GHz **[6]**, or further within the far-infrared domain up to 30 THz **[7]**. A terahertz emission possesses a single photon energy of 0.41-41 meV, a wavelength ranging from 30 µm to 3 mm, a wavenumber of 3.34 to 334 cm^{-1} and an equivalent transient time of 0.1-10 ps. In view of the foregoing, the terahertz regime makes the link between energies used for electronics and the associated classical interpretation of light-matter interaction and to higher energies for which quantum mechanics is employed to describe interactions. Due to the energy of terahertz waves, the reaction following the irradiation of a material is generally in the form of collective rotations of dipoles.

This feature reflects in an increase of absorption for polar materials, which nevertheless remains globally marginal for non-polar ones. Hence, terahertz spectroscopy is able to provide absorption spectra for the identification of substances and for the characterization of their chemical composition. For instance, applications have emerged in the scope of pharmacology **[8]** and of heritage conservation **[9]** for the characterization of drugs and historical pigments. On the opposite, while non-polar and non-conductive materials, *i.e.* dielectrics, do not exhibit specific spectral features, their sufficient transparency to terahertz rays makes possible in-depth (compared to infra-red) and non-destructive (compared to X-rays) control, with an adequate spatial resolution (compared to millimeter-waves).

I.2.2 Safety Issues

The first question to be raised when dealing with radiation sources is the one of the safety. *Are these radiations deleterious for cells, tissues, organs and individuals ?* However, assessing it requires a distinction between the energy and the power of the source but also the exposure time to those radiations.

The Planck's energy–frequency relation states that the frequency ν is proportional to the energy E carried by a photon:

$$\mathrm{E} = h\nu, \tag{I.1}$$

where $h \approx 6.626 \times 10^{-34}\mathrm{J}.s^{-1}$ and denotes the electromagnetic quantum. The energy limit carried by terahertz photons is 41.4 meV at 10 THz, far from the first ionization level of carbon (11.26 eV), the main constituent of living organisms. Single terahertz photons are thus assumed to be non-ionizing and harmless for biology samples from an energy point of view.

To evaluate the impact of source power and exposure time, Wilmink *et al* **[10]**, employed, in 2011, a 2.45 THz monochromatic radiation with an average surface power of 84.8 $\mathrm{mW.cm}^{-2}$ to irradiate dermal fibroblasts. Fibroblasts were exposed for 0 to 80 minutes. They did not report variations in comparison to the control line, emphasizing that THz radiations do not appear to damage intracellular proteins or cell DNA. Another study conducted on lymphocytes exposed to 120 GHz, 1mW for 20 minutes did not report appreciable changes at a cellular and molecular level **[11]**. However, contrary to the previous study, another research **[12]** observed gene instabilities in lymphocytes exposed to an equivalent dose (0.45 $\mathrm{J.cm}^{-2}$) following an exposure time of 2 hours. They reported the induction of a significant number of cases and disturbances in the replication process of chromosome centromeres. Additional studies have also reported alterations in gene expressions **[13]**, in blood antioxidant activity **[14]** , in liposome permeability **[15]**, in seed germination **[16]** and in yeast colony growth rate **[17]**. However, it has not been demonstrated that terahertz radiation causes harmful effect to keratinocytes, epithelial cells, nucleobases, desoxyribonucleic acid or tissue matrix **[18–20]**. The effect of terahertz radiations on biology samples is still under discussion but are, overall suggested, less hazardous than smaller wavelength waves by using mild-power sources.

I.3 Terahertz Field and Biological Samples

Nowadays, scientific efforts are not solely restricted to safety issues but have expanded to dermal control of skin injuries **[21]**, hydration level monitoring in stratum corneum **[22]**, lumpectomy margin checks **[23–25]** or early screening of diabetic syndrome **[26]**.

I.3.1 Physical Quantities in Electromagnetism

The interpretation of recorded data arising from the interaction between terahertz light and biosystems usually deals with the analysis of a set of parameters denoting for instance the light velocity changes from a material to another, the wave attenuation or the dipole sensitivity of the probed

structure to the propagating electromagnetic field. For this reason, the mathematics behind interactions between media are, hereafter, described.

Light Transport Theory

Electromagnetic wave propagations are disturbed when encountering a change of medium. This disruption reflects in a portion of light which is reflected back while the rest is transmitted (also called refracted) through the subsequent medium (Figure I.2).

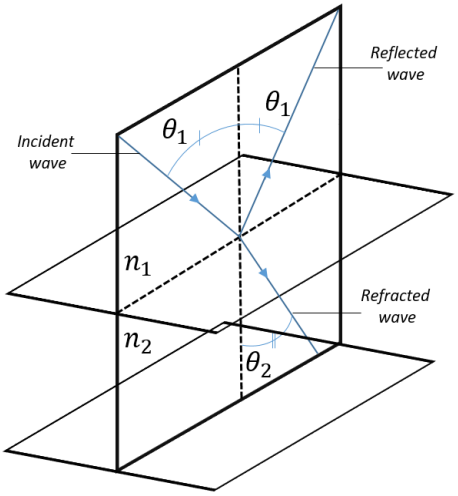


Figure I.2: Refraction and reflection occurring at a dielectric interface between two media of propagation.

These aforementioned portions of light either reflected or transmitted at the interface between two media are illustrated by the Fresnel's coefficients:

$$t = \frac{2\hat{n}_1 cos\theta_1}{\hat{n}_1 cos\theta_1 + \hat{n}_2 cos\theta_2}, \tag{I.2}$$

$$r = \frac{\hat{n}_1 cos\theta_1 - \hat{n}_2 cos\theta_2}{\hat{n}_1 cos\theta_1 + \hat{n}_2 cos\theta_2}, \tag{I.3}$$

where t and r are the Fresnel's coefficients in transmission and reflection, ($\hat{n}_1$,θ_1) and ($\hat{n}_2$,θ_2) are the refractive index and the characteristic angle couple of incidence and refraction, respectively. For instance, let's now consider an incident radiation whose electric field satisfies the Helmholtz's equations and is of the form:

$$\vec{\mathrm{E}}(\vec{r}, t) = \mathrm{E}_0 cos(\omega t - \vec{k} \cdot \vec{r})\vec{u}_r, \tag{I.4}$$

where $\mathrm{E}_0 = |\vec{\mathrm{E}}_0| = \mathrm{E}(0,0)$ is the amplitude and phase of the electric field at the origin at $t_0 = 0$, $\omega = 2\pi f$ is the wave pulsation, $\vec{r}$[1] is the spatial vector, $\vec{k}$ is the wave vector denoting the propagation direction and $\vec{u_r}$ is a unit vector. One can recognize the formulation of a monochromatic plane wave. We can then express the transmitted (I.5) and reflected (I.6) part of the electric fields as a function of the incident field:

$$|\vec{\mathrm{E}}^t(\vec{r}, t)| = \mathrm{E}^t(r,t) = t \times \mathrm{E}(r,t) = \frac{2\hat{n}_1 cos\theta_1}{\hat{n}_1 cos\theta_1 + \hat{n}_2 cos\theta_2} \times \mathrm{E}(r,t), \tag{I.5}$$

$$|\vec{\mathrm{E}}^r(\vec{r}, t)| = \mathrm{E}^r(r,t) = r \times \mathrm{E}(r,t) = \frac{\hat{n}_1 cos\theta_1 - \hat{n}_2 cos\theta_2}{\hat{n}_1 cos\theta_1 + \hat{n}_2 cos\theta_2} \times \mathrm{E}(r,t). \tag{I.6}$$

The $\hat{n}$ parameter, referred to the complex refractive index in the following, is of great importance when dealing with light-matter interaction. In fact, the complex refractive index denotes:

[1] Note: $||\vec{r}|| = \sqrt{x^2 + y^2 + z^2}$

first, the relative velocity of light $n = \frac{c}{v}$ in a medium, wherein c is the light velocity in vacuum and v the light phase velocity in the medium; Second the radiation extinction κ. The complex refractive index is of the form:

$$\hat{n} = n - i\kappa. \tag{I.7}$$

Most of materials are dispersive, which means that the real refractive index and the extinction coefficient are frequency dependent. Thus one can rewrite equation (I.7):

$$\hat{n}(\omega) = n(\omega) - i\kappa(\omega). \tag{I.8}$$

From the extinction coefficient κ, one can express the absorption coefficient α which addresses in a similar fashion the losses within a material:

$$\alpha(\omega) = \frac{2\omega\kappa(\omega)}{c}, \tag{I.9}$$

from which we can intuitively derive the penetration depth δ of an incident radiation:

$$\delta(\omega) = \alpha(\omega)^{-1}. \tag{I.10}$$

These representations come from the optics field. However, as we mentioned before, terahertz regime lies between the optics and the electronics fields. Thus, we will see in the following that these two quantities, namely n and κ are linked to electronics parameters.

Permittivity

The permittivity $\epsilon(\omega)$ of a material can be seen as its sensitivity in regards of an external electric field $\vec{E}$ in which the material is emerged. Alternatively, it is a measure of the motion of charge carriers due to the presence of an external electric field. To illustrate this statement, one can draw the following scheme:

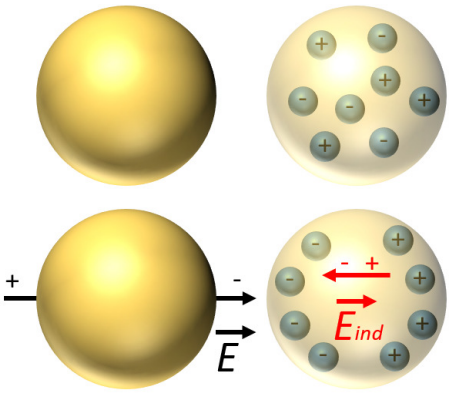


Figure I.3: Representation of charge carrier motions induced by an external electric field $\vec{E}$ inside a charged sphere. One unique field line is drawn for visual convenience.

The application of an external electric force through the charged sphere in Figure I.3. induces the motion of charged elements with respect to the electric field lines. The new charge distribution within the sphere triggers an *induced* electric field $\vec{E}_{ind}$ whose field lines are in opposite direction to $\vec{E}$. The permittivity is then an intrinsic parameter denoting the weight of an external field on an internal electric redistribution:

$$\vec{D}(\omega) = \epsilon(\omega)\vec{E}(\omega), \tag{I.11}$$

where $\overrightarrow{D}(\omega)$ is the induced electric displacement field. In general, the permittivity of a material is given relative to that of free space. Introducing then ϵ_0 the vacuum permittivity and ϵ_r the relative permittivity, equation (I.11) becomes:

$$\overrightarrow{D}(\omega) = \epsilon_0 \epsilon_r(\omega) \overrightarrow{E}(\omega). \tag{I.12}$$

Let's note that the permittivity ϵ in (I.11) and the relative permittivity ϵ_r in (I.12) are not necessary scalars. For anisotropic materials, the permittivity is a third rank tensor, *i.e.* a matrix.

In most cases, the permittivity of materials is complex. The real part of the permittivity $\epsilon'_r(\omega)$ is related to the dynamical response of a system to an external field and the imaginary part $\epsilon''_r(\omega)$ denotes the losses within the lossy material:

$$\hat{\epsilon}(\omega) = \epsilon_0(\epsilon'_r(\omega) - i\epsilon''_r(\omega)). \tag{I.13}$$

We can however distinguish different sources for losses:

- Losses due to the transformation of the absorbed electric energy into thermal energy through polar relaxation: *dipole losses*.
- Losses due to the conduction currents: *conductivity losses*.

The imaginary part of the permittivity is then a measure of these two sources of losses:

$$\epsilon''_r(\omega) = \epsilon_d(\omega) + \frac{\sigma(\omega)}{\omega}, \tag{I.14}$$

here $\epsilon_d(\omega)$ stands for the dipole losses and $\sigma(\omega)$ for conductivity losses. It must be highlighted that ϵ'_r and ϵ''_r are not independent. Insofar as science accepts the principle of causality, and inasmuch as the complex permittivity is a linear complex function, one can obtain the relation between the real and the imaginary part of the complex permittivity via the Kramers–Kronig relations **[27]**:

$$\epsilon'_r(\omega) = 1 + \frac{2}{\pi} \int_0^\infty \frac{\omega' \epsilon''_r(\omega')}{\omega'^2 - \omega^2} d\omega', \tag{I.15}$$

$$\epsilon''_r(\omega) = -\frac{2\omega}{\pi} \int_0^\infty \frac{\epsilon'_r(\omega') - 1}{\omega'^2 - \omega^2} d\omega'. \tag{I.16}$$

Beyond the mathematical complexity of those relations, the calculation of the imaginary part from the real part (or vice versa) requires the integration of (I.15) or (I.16) from the 0 frequency to the infinity. Alternatively said, one has to characterize over the entire electromagnetic spectrum either the imaginary or the real part of the permittivity to extract the other one. From a theoretical point of view, $\omega = 0$ is verified for $\lim_{t \to +\infty} \frac{1}{t} = 0$. That means that the wave has only a single crest and dip and is, obviously, not observable. On the other hand, $\omega = \infty$ is equivalent of $\lambda = 0$ which defies the definition of wave itself. From an experimental point of view, there is no existing systems that may provide a full characterization. However, one can still extract one parameter from an other through Kramers-Kronig transformations by extrapolation or by making assumptions on the behavior of the real permittivity (or the imaginary one) outside the experimental spectral window. This complexity is nevertheless overcame for terahertz radiations. In fact, one of the most used system in terahertz spectroscopy is the time-domain spectroscopy set-up based on coherent detection and thus is able to provide both the real and imaginary part of the permittivity without the use of the Kramers-Kronig relations. This asset will be discussed in greater details in *Chapter II, II.2 Generation, Detection and Sensing*.

We have seen that when dealing with a lossy material, the permittivity takes complex values. In such a case, one can show by solving Maxwell's equations, that for non-magnetic materials:

$$\hat{n}(\omega) = \sqrt{\hat{\epsilon}_r(\omega)}, \tag{I.17}$$

which in turn leads to:

$$n(\omega)^2 + 2in(\omega)\kappa(\omega) + i^2\kappa(\omega)^2 = \epsilon_r'(\omega) - i\epsilon_r''(\omega). \tag{I.18}$$

Through identification along the real and imaginary planes, we can define the relations between the electronics parameters and the optics ones:

$$\begin{cases} \epsilon_r'(\omega) = n(\omega)^2 - \kappa(\omega)^2 & n(\omega) = \sqrt{\frac{|\epsilon_r(\omega)| + \epsilon_r'(\omega)}{2}} \\ \epsilon_r''(\omega) = 2n(\omega)\kappa(\omega) & k(\omega) = \sqrt{\frac{|\epsilon_r(\omega)| - \epsilon_r'(\omega)}{2}}. \end{cases} \tag{I.19}$$

Equation (I.9) prescribed that the absorption $\alpha(\omega)$ of a radiation is proportional to the imaginary part of the complex refractive index $\kappa(\omega)$ and thus, the higher $\kappa(\omega)$ is, the smaller $\delta(\omega)$ is. Considering the way (n, κ) and $(\epsilon_r', \epsilon_r'')$ are related to each others (eq. I.19), one can draw a similar conclusion for the relative dielectric loss factor $\epsilon_r''(\omega)$. For instance, let us consider the dielectric properties of different polar liquids at 70 GHz and at room temperature:

Table I.1: Dielectric properties of some alcohols and water at 70 GHz. **[28]**

Liquid	ϵ_r'	ϵ_r''	κ	$\delta(mm)$
Propanol	3.00	0.80	0.19	3.60
Ethanol	3.47	0.98	0.26	2.62
Methanol	4.63	2.36	0.54	1.26
Water	9.93	17.56	2.26	0.30

As we can see, the penetration depth for each of those compounds are inversely proportional to both the dielectric and optics loss factors. Furthermore, losses related to water are by far the highest in the set described above and consequently the penetration depth is the smallest. Water is the polar liquid having the highest extinction. One can roughly estimate that the amplitude of a 70 GHz monochromatic wave would be attenuated by about 95% after a penetration of 3δ (0.9 mm). We can emphasize that water absorption is -at the edge of the terahertz domain- particularly strong. At the same time, water is one of the main constituent of organisms and bio-like objects in general. Intuitively, one can thus anticipate that water is about to play a crucial role in the interaction between terahertz fields and biological systems.

I.3.2 Interaction Mechanisms

The previous section established a qualitative framework for the understanding of the main physical quantities involved in the interaction between terahertz radiations and biosystems. Now that the mathematics behind the quantification of interactions are settled, one can assess there physical and chemical forms.

Relaxation & Distribution

When an external stimulation is applied to biological objects, the molecular dipoles tend to align with respect to the field lines and against the surrounding friction forces. The directional change of the polarization vector across the irradiated structure is not instantaneous and is subject to time lag with respect to the alternating field. The kinetic of the charge redistribution is frequency dependent but may additionally be affected by the viscosity, the temperature and the molecule mobility. When illuminating a sample with a given energy range, one can analyze the frequency dependence of the relative permittivity: the dielectric spectra. The greatest variation among the dielectric spectra arises when the frequency of the external field matches a characteristic frequency of a system and is called a permittivity step $\Delta\epsilon_r$. The relaxation time associated to the characteristic angular frequency of a system ω_C is defined as:

$$\tau_D = \omega_C^{-1}. \tag{I.20}$$

Such a behavior is modeled by the simple Debye model in which **[29]**:

$$\hat{\epsilon}_r(\omega) = \epsilon_\infty + \frac{\Delta\epsilon_r}{1 + i\omega\tau_D}, \tag{I.21}$$

where ϵ_∞ is the permittivity at infinite frequency (in a relative meaning). Upstream and downstream of $\Delta\epsilon_r$, the permittivity is roughly constant. When the pulsation ω_E of an incident radiation is to small to stimulate the resonant peak located at ω_C, the permittivity takes the value $\Delta\epsilon_r + \epsilon_\infty$. In fact, charge carrier alignment is proceeded in time if $\omega_E << \omega_C$. On the other hand, if $\omega_E >> \omega_C$, dipoles cannot reorient as fast as is the external stimulus and thus, no redirection polarization is observed and $\epsilon_r(\omega) \approx \epsilon_\infty$. In the analysis of dielectric relaxations, it is common to draw the Argand diagram in which the imaginary part of the permittivity ϵ_r'' is plotted as a function of the real part ϵ_r'. The dielectric function describes a semi-circle which links ϵ_∞ to $\Delta\epsilon_r + \epsilon_\infty$. The Debye treatment assumes, in a sense, that all molecules belong to the same environment. Additionally, to accurately extract the dielectric parameters from the Debye equation, one has to ensure that the characteristic frequency ω_C at which, $\epsilon_r'(\omega)$ observes an inflection point and $\epsilon_r''(\omega)$ has its maximum value (Figure I.4.), belongs properly to the sensing range. In other words, if the higher limit of the probing frequency is lower than ω_C, the accuracy of the parameter extraction is anticipated to be rough **[28]**.

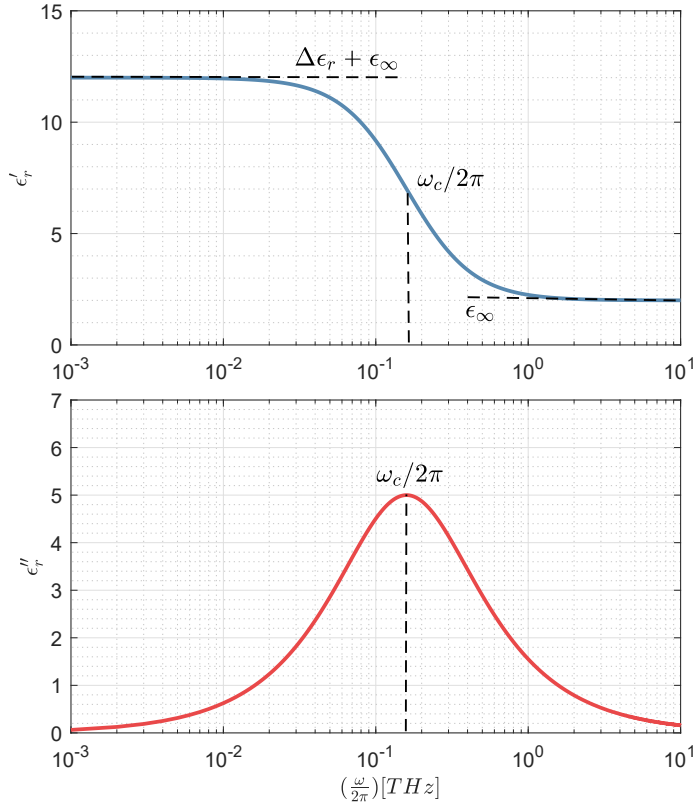


Figure I.4: Example of a simple Debye treatment for $\Delta\epsilon_r - \epsilon_\infty = 12$, $\epsilon_\infty = 2$ and $\tau_D = 1ps$. The inflection point of $\epsilon_r'(\omega)$ and the maximum of the dielectric loss factor $\epsilon_r''(\omega)$ are located at $\omega_C = 1.01$ rad.ps^{-1}.

The Debye model is a simplistic view of the process. Let's note the vast majority of molecules does not follow the description established by (I.21). One can in fact encounter polyrelaxations within material dielectric spectrum. Their description can be achieved mathematically by a superposition of Debye functions with different relaxation times by introducing the concept of relaxation time distribution, leading to a new formulation **[28]**:

$$\hat{\epsilon}_r(\omega) = \epsilon_\infty + \sum_{n=1}^{N} \frac{\Delta\epsilon_r^n}{1 + i\omega\tau_n}, \tag{I.22}$$

from which we can sort the real and the imaginary part of the permittivity:

$$\epsilon'_r(\omega) = \epsilon_\infty + \sum_{n=1}^{N} \frac{\Delta\epsilon_r^n}{1 + \omega^2\tau_n^2}, \tag{I.23}$$

$$\epsilon''_r(\omega) = -\sum_{n=1}^{N} \frac{\Delta\epsilon_r^n}{1 + \omega^2\tau_n^2}\omega\tau_n. \tag{I.24}$$

In the following, for simplicity, $\epsilon''_r(\omega)$ will consistently refers to $|\epsilon''_r(\omega)|$.

Other models have been suggested yielding more complicated formulations. Among them, one can mention the empirical Cole-Cole model **[30; 31]**, developed on the basis of polyrelaxation distributions but assessing the relaxation time width over the frequency band:

$$\hat{\epsilon}_r(\omega) = \epsilon_\infty + \sum_{n=1}^{N} \frac{\Delta\epsilon_r^n}{1 + (i\omega\tau_n)^{\alpha_n}}, \tag{I.25}$$

where $0 \leqslant \alpha \leqslant 1$ and denotes the aforementioned width of relaxation time. The α-term reflects the permittivity step damping, spreading incrementally the energy dissipation over the frequency range as it decreases (Figure I.5). Let's note that $\alpha = 0$ depicts the absence of relaxation features and $\alpha = 1$ reduces to the Debye model.

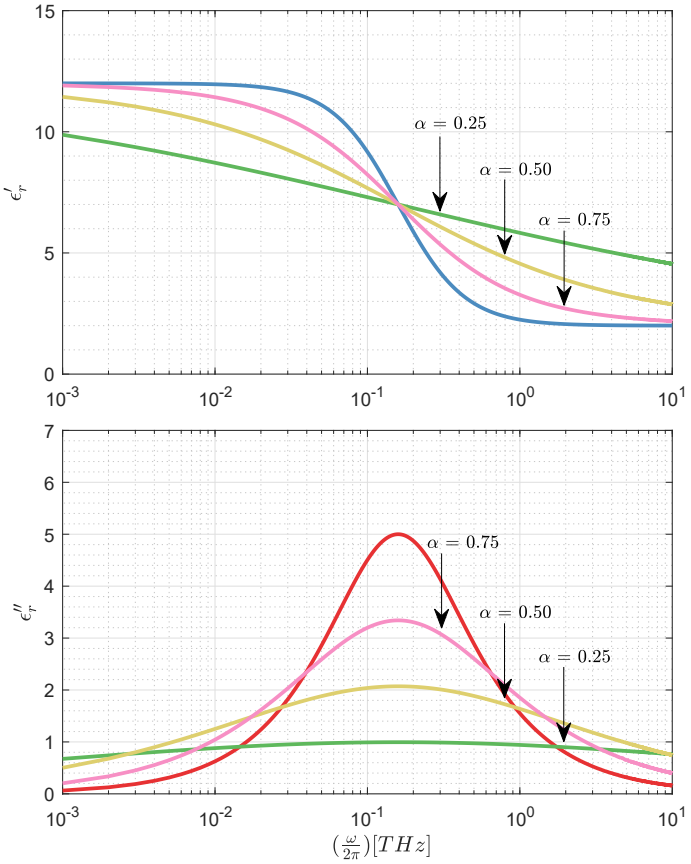


Figure I.5: Illustration of the Cole-Cole treatment for the Debye parameters used in Figure I.4 (Top: blue; Bottom: red) and for $\alpha = 0.25, 0.50, 0.75$ subsequently.

Another experimental approach is the Davidson-Cole model **[32; 33]**:

$$\hat{\epsilon}_r(\omega) = \epsilon_\infty + \sum_{n=1}^{N} \frac{\Delta\epsilon_r^n}{(1 + i\omega\tau_n)]^{\beta_n}}, \tag{I.26}$$

where the empirical parameter β is positive and no greater than unity. Similarly to the Cole-Cole model, the Davidson-Cole treatment turns into the Debye model if assuming $\beta = 1$. The impact of the β parameter on relaxations is shown in Figure I.6.

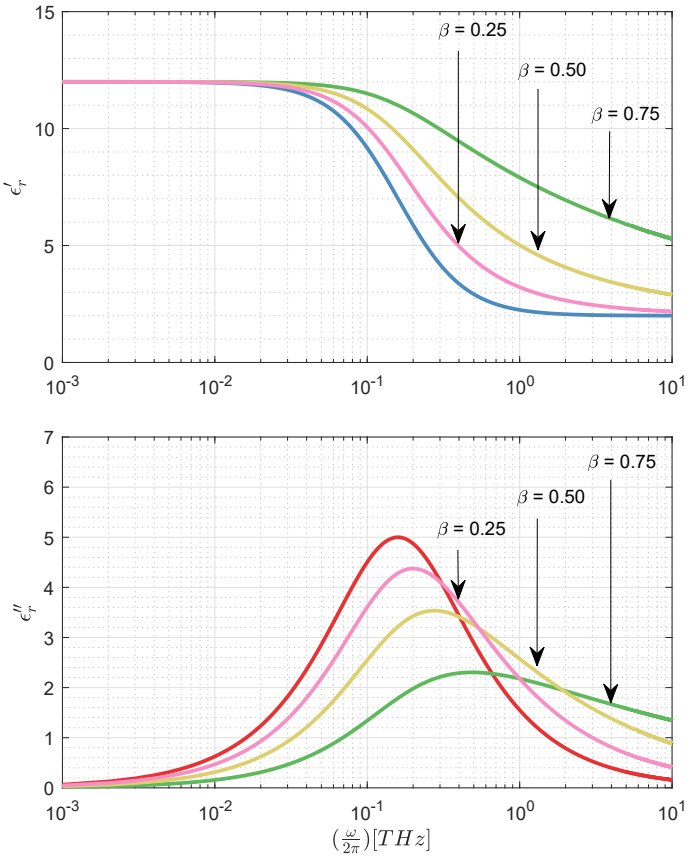


Figure I.6: Illustration of the Cole-Davidson treatment for the Debye parameters used in Figure I.4 and for $\beta = 0.25, 0.50, 0.75$ subsequently.

Ultimately, a formal generalization of both the Cole-Cole and Davidson-Cole models is given by the Havriliak-Negami representation **[34]** whose formulation is:

$$\hat{\epsilon}_r(\omega) = \epsilon_\infty + \sum_{n=1}^{N} \frac{\Delta\epsilon_r^n}{[1 + (i\omega\tau_n)^{\alpha_n}]^{\beta_n}}, \tag{I.27}$$

here, α_n and β_n are necessary parameters for the treatment of the aforementioned model. The Havriliak-Negami expressions reduces to the Cole-Cole equation for $\beta_n = 1$ and reduces to the Cole-Davidson formulation for $\alpha_n = 1$. A complete description of the relaxation behavior under different α and β is given in Figure I.7, for an unique relaxation.

Heretofore, only models describing relaxations have been assessed. However, when stimulated with external electromagnetic fields, other processes than those reported may occur. Permittivity steps can, in fact, be due to molecular vibration. Complementary to the previous models, one can consider the Lorentz model to describe molecular motions induced by a force field:

$$\hat{\epsilon}_r(\omega) = \epsilon_0 + \sum_{n=1}^{N} \frac{A^n \omega_{0n}^2}{(\omega_{0n}^2 - \omega^2) - i\Gamma_n\omega}, \tag{I.28}$$

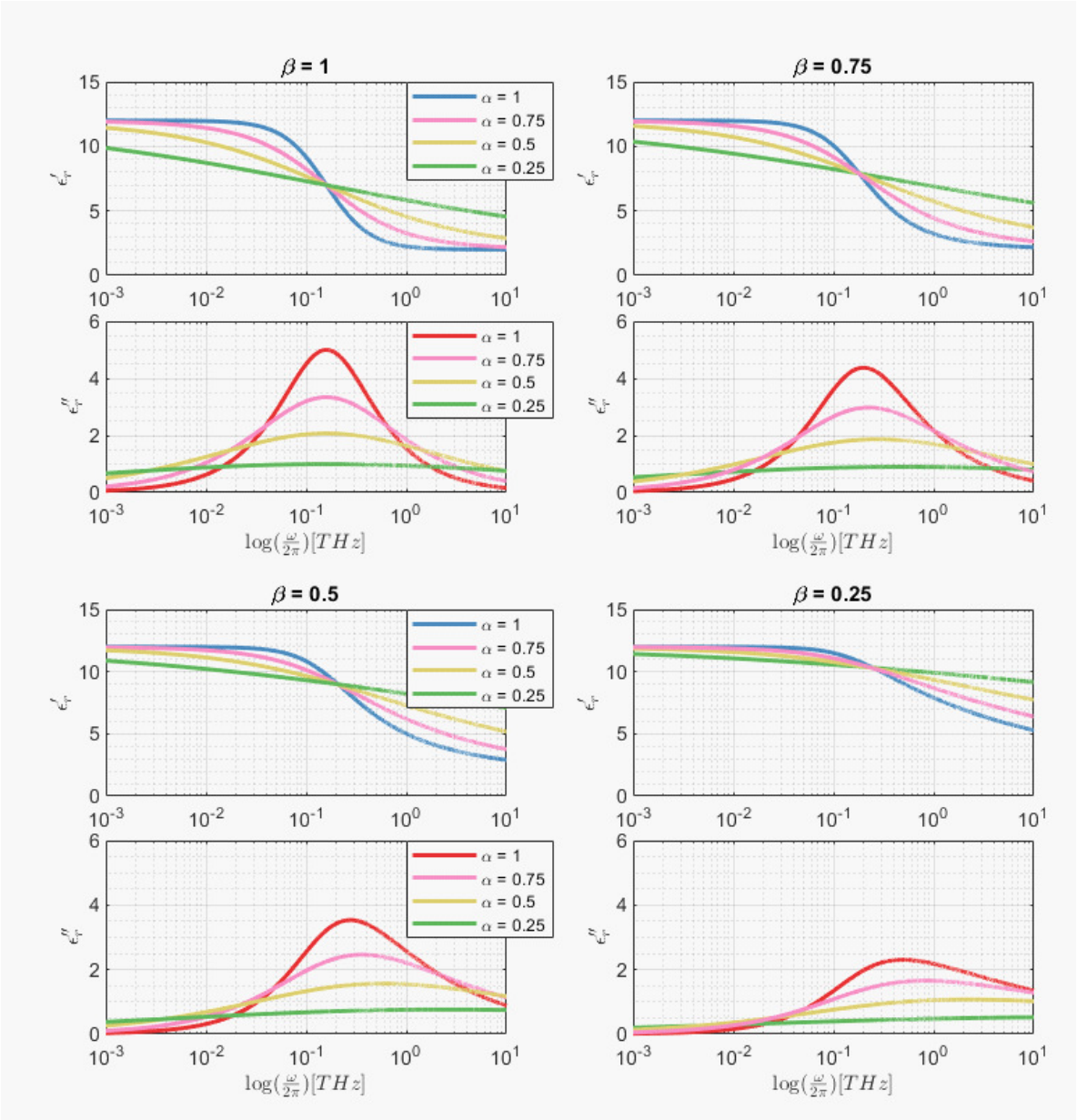


Figure I.7: Illustration of the Havriliak-Negami formulation for the Debye parameters used in Figure I.4 and for α and β taking the successive values: 0.25, 0.50, 0.75.

where, ϵ_0 is the low-frequency permittivity, ω_{0n} the resonant frequency of the n_{th} process, A_n the corresponding oscillation amplitude and Γ_n an associated damping factor (Figure I.8).

These models are usually used to fit and characterize the dielectric spectrum recorded from dielectroscopy measurements with a Levenberg–Marquardt algorithm **[35; 36]** or equivalently the damped least-squares method. The need for relaxation distribution to describe system responses is however not instinctive. Furthermore these parameters have to be adapted to the physical systems under inspection, that is, with a physical sense. The difficulty is even more harder when dealing with biological systems. The complexity of biological systems made of a lot of different molecules is expected to blur resonant peak(s) of individual compound. For instance, biological tissues are made of an extracellular matrix and an ensemble of cells assuming various functions. The extracellular matrix, holding up the biochemical support for cells, is mainly made of collagen, enzymes and glycans linked to amino acid side-chains through covalent bonds. Cells are in turn made of a phospholipid shell enclosing proteins, DNA, RNA, proteinogenic amino acids, antibiotics, inorganic ions (Na^+, K^+, Ca^{2+}, and so on), salts and unsurprisingly water, accounting for at

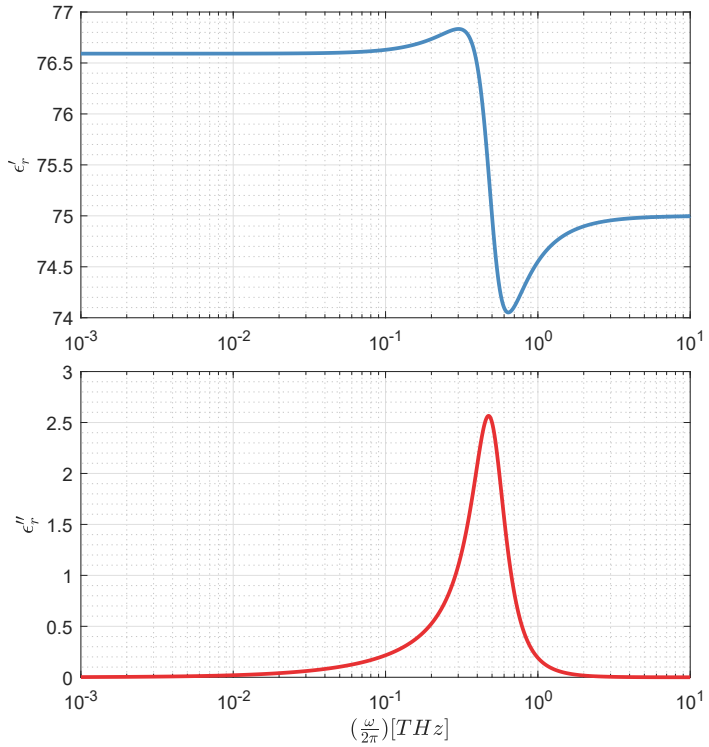


Figure I.8: Illustration of a Lorentz oscillator for ϵ_0 = 75, $A/(2\pi)^2$ = 0.25 THz^2, $\Gamma/(2\pi)$ = 0.32 THz and a characteristic frequency at $\omega_0/(2\pi)$ = 500 GHz.

least 70% of the total cell mass **[37]**.

Interaction Bands

The dielectroscopy of biological objects is expected to be highly dispersive and inhomogeneous, to exhibit spatial dependence and superimposition of dielectric domains owing to the complexity of their structure. One usually distinguish four different spectral bands among the electromagnetic spectrum. These bands, namely α, β, δ and γ, arise from relaxation or dispersion processes occurring within the biological object due to external stimuli **[38; 39]**. The representation of these four relaxation domains is shown in Figure I.9. The α and β bands are located far in the microwave regime (in the KHz for α and in the MHz for β) and denote relaxation of ions and large biomolecules. The δ and γ relaxations are however taking place at higher frequencies, between about 1 and 5 GHz for δ while above 100 GHz for γ processes. Let's note that the γ-band starts from where we have initially defined the low limit of the terahertz regime. The γ dispersion is assumed to be responsible for the relaxation processes within small and mobile polar molecules. In fact, at these frequencies, the energies (eq. I.1) are yet too low to impact the electronic cloud, and are absorbed in the form of resonant molecular vibrations. Mainly, these relaxations are attributed to bulk water and hydration water molecules. Water is the leading role in all living organisms **[39]**. At the microscopic level of tissues, it governs the formation of cell phospholipid membrane, it regulates concentration gradient in and out cells for osmosis to occur or its participates in enzymatic reaction for DNA molecule synthesizing. In a nutshell, water is essential for tissues, is the main constituent and thus is expected to dictated the interaction between terahertz radiations and organic structures by ruling the global dielectric response.

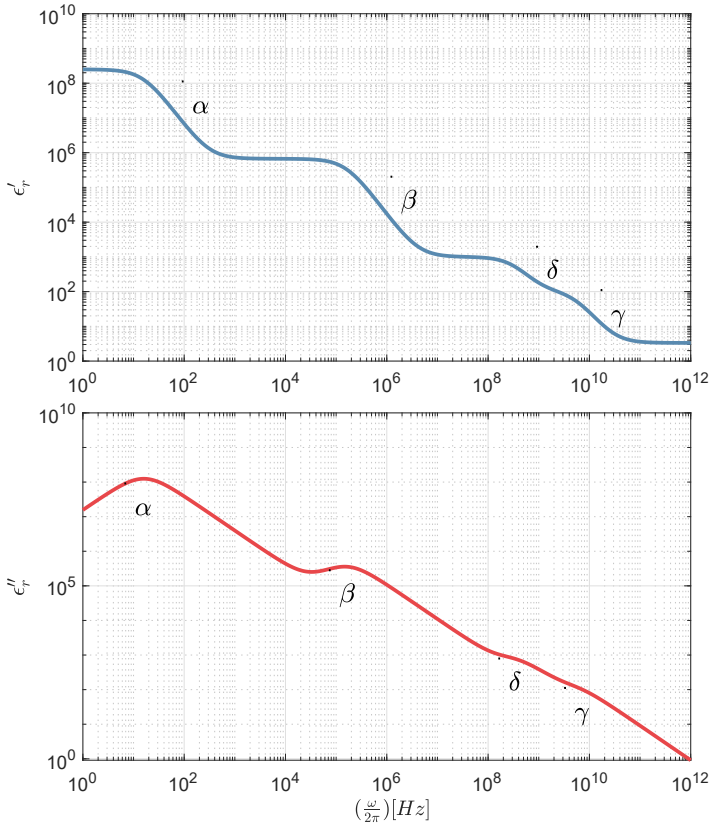


Figure I.9: Schematic representation of the real and imaginary part of the relative permittivity, ϵ'_r and ϵ''_r as a function of the frequency and the location of the dispersive domains α, β, δ and γ. Where α denotes KHz relaxation of ions, β describes the MHz relaxation of large biomolecules, δ depicts the hydration shell relaxation within protein surface - water complex and γ stands for bulk and free-water relaxations. Simulation performed by a tetra-Debye treatment (equation I.22 wherein N = 4) for $\epsilon_\infty = 3.33$, $\Delta\epsilon_1 = 2.49 \cdot 10^8$, $\Delta\epsilon_2 = 6.66 \cdot 10^5$, $\Delta\epsilon_3 = 900$, $\Delta\epsilon_4 = 96.67$, $\tau_1 = 10^{-2}$s, $\tau_2 = 10^{-6}$s, $\tau_3 = 5 \cdot 10^{-10}$s, $\tau_4 = 5 \cdot 10^{-11}$s.

I.3.3 Candidates for Interaction: Sub-cellular Level

The informativeness of terahertz radiations when applied on biosystems depends on the specific dielectric responses of the various biomolecules in presence. Since the early stage of bio-terahertz investigations, it was certain that water could be the primordial feature. However, water is not the only object that exhibits specific features in the terahertz band. There is actually plenty of biological entities that may additionally interact with terahertz radiations. In regards to the incredible wealth of biological compound interoperating with each other among bio-organisms, the analysis of those biostructures requires preliminary investigations on these individual elements.

Water

By assuming 70% of the total mass of cells, water is the key factor which in a lot of biological systems dictates the dielectric response. Water is a tiny bent molecules whose chemical formulation is H_2O **[40]**. Water molecules are electrically neutral but polar. The center of the slightly negative pole is located at the oxygen atom and the positive one between the two hydrogen atoms, at 0.8 Å from each, approximately. The hydrogen atom is a simple structure made of one proton and an electron located in the first orbital level *1s*. Oxygen atom is formed of a nucleus consisting of eight protons and neutrons, surrounded by the electronic cloud made of eight electrons. On forming the water molecule, each hydrogen atom shares an electron with the oxygen atom to form a so-called covalent bond, leaving the positively-charged hydrogen nucleus *alone*. On the

contrary, the oxygen side of the molecule is consequently partially negatively charged. Individual water molecules are then similar to dipoles in the sense that they exhibit a negative and a positive pole. This specificity allows water molecules to bind with each others through hydrogen-bonding in which, the negatively charged side of the molecule binds with the positively charged hydrogen nucleus of another one. These hydrogen-bonds give the particularity to water to have a lower density in the solid state than in the liquid one. It additionally explains the high boiling point of liquid water due to the high energy required to break those bonds. The highest water density is reached at 277.15 K. The mean distance between two oxygen atoms in bulk water is 2.82 Å and between hydrogen-bonded hydrogen and oxygen is 1.88 Å. Water molecules that are hydrogen-bonded are represented in Figure I.10.

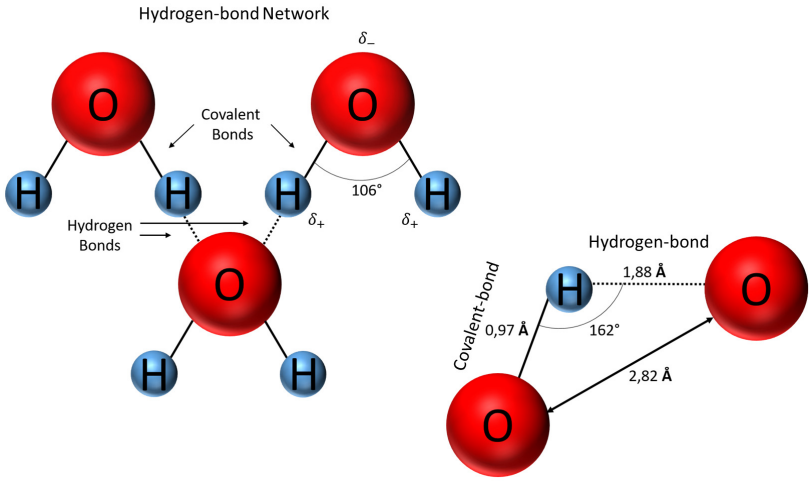


Figure I.10: Water molecules. Right: the hydrogen-bond network. Left: closer look on a hydrogen-bond and a covalent bond.

Due to the importance of water molecules in biological object organisation, studies on the dielectric response of water in the terahertz regime started in the mid-*1990s*. Barthel *et al* **[41]** reported in 1996, preliminary investigations on the dielectric spectrum of water in the microwave region, close to the terahertz limit, up to 89 GHz. They revealed the necessity to consider two main relaxation processes for bulk water. One process exhibited a long relaxation time by about $\tau_D = 8.32ps$ and a fast relaxation whose characteristic time was about $\tau_2 = 1.02ps$. In 1997, Kindt and Schmuttenmaer **[7]** published an article in which they characterized the dielectric behavior of water from 60 GHz to 1.50 THz. They reported a slow relaxation time $\tau_D = 8.24 \pm 0.40$ ps and the other, a fast process, $\tau_2 = 0.18$ ps. One can notice the discrepancy between the two reported values for τ_2. This difference is not an inherent shortcoming in one of the studies, but reflects the restricted spectral range employed by the first study. In fact, the fast relaxation process τ_2 was reported to be equal to 1.02 ps which corresponds to a characteristic frequency of $f = \frac{1}{2\pi\tau_2} = 156$ GHz. However the frequency window used, in this pioneer study, was up to 89 GHz. In other words, as underlined in *Relaxation & Distribution*, the external field frequency ω_E was too low in comparison to the characteristic frequency ω_C of the system to accurately probe the relaxation process. Further studies **[42–45]** have confirmed the existence of these two relaxation processes. The slow relaxation τ_D whose characteristic frequency is located at ~ 19.3 GHz is commonly attributed to the breaking and reformation of hydrogen bonds within the hydrogen-bonded water molecule population. Put more simply, it consists of the cooperative reorganization among bulk water. On the other hand, the fast relaxation, occurring around 884 GHz denotes in a way *fast* water molecules, *i.e.* free or quasi-free water. These water molecules are not or poorly involved in hydrogen bonding and therefore exhibit a faster reorientation with respect to the external field lines. In 2008, Yada *et al* **[44]** have suggested the contribution of other processes occurring at higher frequencies by including in their study the results of a previous experiment from Asfar *et al* **[46]** in the far-infrared regime. They found that the total dielectric step $\Delta\epsilon_r$ could be expressed as:

$$\Delta\epsilon_r = \Delta\epsilon_{r1} + \Delta\epsilon_{r1} + \frac{A_S}{\omega_S^2} + \frac{A_L}{\omega_L^2}, \tag{I.29}$$

where A_S and A_L are intermolecular stretching vibration and intermolecular libration[2] occurring respectively at 5 and 15 THz. The following table summarizes the water properties found in the main studies.

Table I.2: Comparison of the dielectric Debye parameter for water molecules found in the literature. [a] Parameter fixed during the Levenberg–Marquardt fitting procedure **[41; 47]**. [b] Includes intermolecular libration amplitude.

Study	Spectral Band	$\Delta\epsilon_{r1}$	$\Delta\epsilon_{r2}$	$\tau_{D[ps]}$	$\tau_{2[ps]}$	$A_{\nu[THz^2]}$	ϵ_∞
Barthel *et al* **[41]**	(0.1-89) GHz	72.18	1.69	8.32	1.02	/	4.69
Kindt *et al* **[7]**	(0.06-1.5) THz	73.43[a]	1.45	8.24	0.18	/	3.48
Ronne *et al* **[43]**	(0.1-2.0) THz	72.1[a]	1.9	8.5	0.17	/	3.3
Sato *et al* **[42]**	(0.1-89) GHz	72.2	2.14	8.32	0.39	/	3.96
Yada *et al* **[44]**	(0.2-3.5) THz	74.9[a]	1.63	9.43	0.25	$31.5\times2\pi^2$	2.5
Moller *et al* **[45]**	$(10^{-4}-2.5)THz$	72.3	2.12	8.34	0.36	$28.4\times2\pi^2$	2.68[b]

To summarize, when investigating below 3 THz, one can assume that only two relaxations in the form stipulated by Debye are necessary to properly fit the data. One slow relaxation process whose characteristic frequency is located at ~ 20 GHz, denoting configurational changes within the hydrogen-bond network and one fast relaxation located at ~ 890 GHz, involving free and quasi-free water molecules rotating or translating with respect to the external stimulation. Nonetheless, if dealing with higher frequencies, one should consider intermolecular vibration stretching (~ 5 THz) and libration (~ 15 THz).

Carbohydrates

The name carbohydrate refers to molecules or group of molecules whose chemical structure is of the form $C_XH_{2X}O_X$ and commonly known as sugars **[48]**. D-glucose ($C_6H_{12}O_6$) is the only active form of glucose found in cells. D-glucose as a significant importance as it serves as the first source for cell energy, as the fundamental brick for mitosis and as a metabolic intermediate for cell-cell communication. Sugars having at least five carbon atoms can cyclize. With its six respective atoms of carbon (C = 6), D-glucose is then often cyclic. Let's however underline that its linear form exists but is naturally found in a proportion of ~ 1 for 100 among aqueous solutions. D-glucose is found in two alternative forms (anomer): the α-D-glucose and the β-D-glucose which only differ in the orientation of -H and -OH groups on carbon 1 (Figure I.11.). Nonetheless, despite their quite similar structure, the chemical association of several α-D-glucose gives arise to a polymer starch. By contrary, β-D-glucose molecules bond with each other to form a polymer cellulose. Both α-D-glucose and β-D-glucose are expected to exhibit specific features within the terahertz band thanks to their cyclic nature, presumably allowing different intermolecular vibration and phonon modes.

In an article printed in 2006, Liu & Zhang **[49]** have sensed polycristalline monohydrated-D-glucose ($C_6H_{12}O_6 \bullet H_2O$) and anhydrous-D-glucose. They reported three main absorption peaks for the monohydrated-version located at 1.82, 1.98 and 2.46 THz. On the other hand, four different absorption peaks have been observed for anhydrous-D-glucose for 1.29, 1.44, 2.10 and 2.6 THz. As α-D-glucose, β-D-glucose and amorphous-D-glucose (linear) are operating at different levels and for different functions, Fischer *et al* **[50]** conducted a study to evaluate the terahertz regime upon discrimination between anomers and the amorphous state. They have reported that the first resonant peaks observed for the α-form of D-glucose were also observed among the β-D-glucose spectra but were however slightly positively shifted along the frequency axis. Authors have emphasized that clear differences were nevertheless observed in the range between 50 cm^{-1}

[2] Motion of an object with a fixed orientation in which it repetitively rotates slightly back and forth.

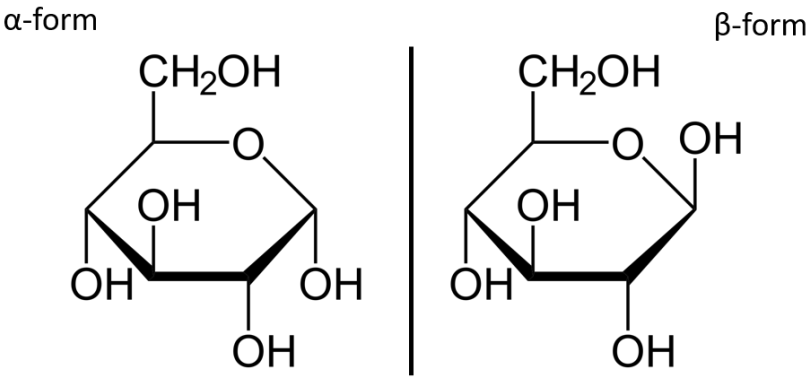


Figure I.11: Structure of the α-form (left) and the β-form (right) of D-glucose.

and 75 cm^{-1}. These measurements were conducted at room temperature and at 13 K but were, intriguingly, still comparable.

Another sugar found in histology is the linear anionic heparan sulfate polysaccharide **[51; 52]** (Figure I.12.) found in the extracellular matrix protein or at the surface of cells by bonding covalently to proteins to form proteoglycans **[53]**. It assumes various biological actions which lie at the heart of harmonizing tissue development. For instance, blood coagulation or formation of new blood vessels from pre-existing ones. It has been however shown that heparan sulfate polysaccharide was involved in tumor metastasis and could act as a cellular receptor for numerous viruses **[54]**.

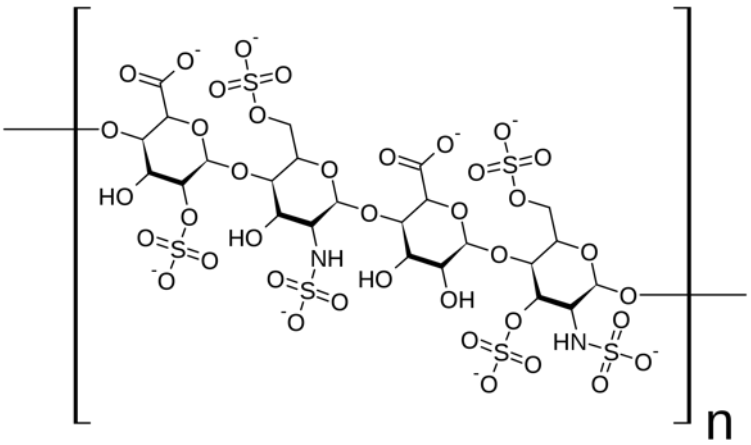


Figure I.12: Structure of the heparan sulfate polysaccharide subunit.

In **[55]**, authors have investigated the dielectric spectra of porcine mucosal heparin which mimics the heparan sulfate polysaccharide behavior from 1 to 13 THz. They inspected the six cationic form of heparin (Na^+, K^+,Ca^{2+}, Mg^{2+}, Cu^{2+} and Zn^{2+}). The overall forms exhibited a comparable spectra. However they have underlined differences in the location of the first peak. In addition they have observed a broad dip in the spectra whose centre was not similarly located for each cation form. Moreover, the intensity difference between the peak and the dip observed in absorption spectra was greater for the Na^+, K^+ and Mg^{2+} cation forms than the others. Holder and his co-authors concluded that the relatively high absorption of terahertz waves by heparin at room temperature, was caused by vacancies in fundamental modes and thus, a certain amount among these modes could be occupied by the absorption of the external field when sensing.

Lipids

Lipids refer to a heterogeneous group of biomolecules that includes fat, oil and steroids. The simplest lipid form is fatty acid that generally consists of 16-or-18 carbon atoms forming a hydrocarbon chain and ends at one side by a carboxyl group (COO^-) (Figure I.13.).

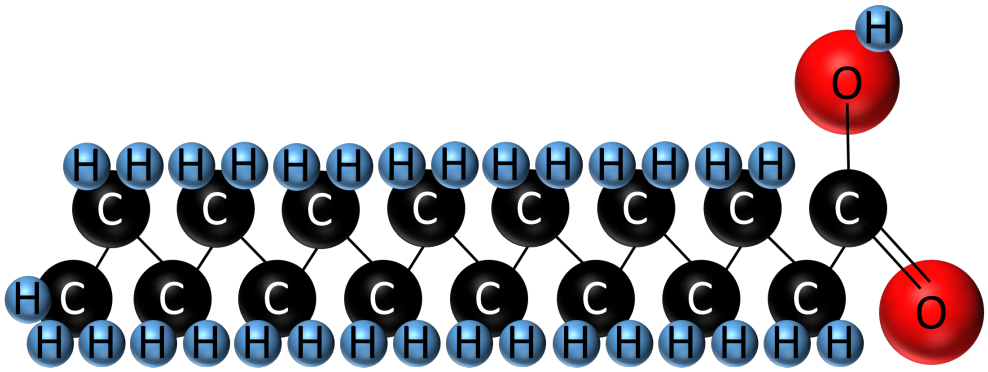


Figure I.13: Chemical structure of the fatty acid $C_{16}H_{32}O_2$ (palmitic acid).

When carbon atoms contained in the hydrocarbon chain are not bonded with the maximum number of hydrogen atoms, they are referred to as unsaturated fatty acids. On the contrary if carbon atoms are bonded to the maximum number of hydrogen atoms, they are referred to as saturated fatty acids. Lipids have major role in organisms. The most noteworthy point is their participation in the formation of the selectively permeable plasma membrane that consists of two saturated fatty acid layers and phosphate group heads on each side. Fatty acid tails are by nature non-polar elements as they only consist of non-polar C-H bonds. Consequently they cannot interact with water molecules. While tails are hydrophobic, phosphate head groups are hydrophilic. When emerged in aqueous solutions, the hydrophilic heads orient themselves to be on the outside, in contact with polar molecules and hydrophobic tails arrange to be on the inside. Such a molecular configuration is referred to as amphipathic arrangement. Sometimes, unsaturated hydrocarbon chains are inserted as tails in the phospholipid bilayer. This induces an increase of the permeability of the membrane to ions and other small molecules. Another asset of lipids is that they are a significant source of energy storage. Fatty acids are stored in the form of three fatty acids bonded to a glycerol molecule, commonly called fat. Fat is accumulated into the cell cytoplasm. When energy is required to yield biological reactions, they break down. The main difference with the energy provided by carbohydrates aforementioned is that, on breaking down, fat releases more energy per unit of volume. As an example, carbohydrates only allow energy to be accumulated in more than twice the body weight that would be needed to store the equivalent quantity of energy in fat.

Hishida and Tanaka **[56]** conducted experiments on the response of phospholipid solutions (water solvent) to an ultrafast stimulus ($\tau \sim 1.10^{-13}$s). Authors reported that both the real and imaginary part of the dielectric function were decreasing as the concentration of phospholipids was increasing. This observation yields to the distinction between bulk water from hydration water. While bulk water refers to molecules whose neighborhood is only populated by other hydrogen-bonded water molecules, hydration water consists of water molecules in close interaction with other biological compounds, *e.g.* carbohydrates, lipids, proteins and so on. Consequently, by increasing the portion of phospholipid into the bulk water solvent, they indeed progressively and concomitantly turned previously bulk-like water molecules into hydration water. The slow relaxation process of water, $\tau_D \sim 8$ ps, referring to the structural reorientation of the bulk water network, was not longer exhibited for high phospholipid concentrations. Authors concluded that on interacting with phosphate head groups, hydration water considerably loses its dynamics inducing the loss of the slow relaxation. The hydration shell length was estimated by about 10 Å

from the lipid surface. Similarly to this study, Reid and co-authors **[57]** reported analogous dielectric decrease features when increasing the lipid concentration. However, for a comparable maximum concentration (25% for Hishida and 20% for Reid), the second study did not report on the global loss of the main relaxation component.

Nucleic Acids

Desoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are the informational molecules of cells (Figure I.14). Both DNA and RNA are made of a chain of monomeric purine (adenine, guanine) and monomeric pyrimidine (cytosine, uracil, thymine) bases attached to phosphorylated sugars. Nucleotides are connected with each other through a covalent bond network. The phosphate group, to which are linked each bases, is joined to the carbohydrate group of the next one and so on. While DNA is made of two hydrogen-bonded strands, each carrying the transposed information of the other, RNA is a one strand molecule which carries the information outside the nucleus and serves as a template for protein synthesis. In regard to the complexity of the DNA structure, candidate sites for interaction with terahertz radiations are numerous. From the foregoing, terahertz waves are sensitive to hydrogen-bonds -via which, DNA strands are joined-. Furthermore, the complexity of the DNA structure may lead to interaction with aminyl radicals (NH_2), with phosphorylated sugars or may additionally involve longitudinal phonon modes. Despite this list -which was not fully representative of the DNA complexity-, theoretical works conducted toward the description of DNA within the terahertz regime have foreseen the existence of specific relaxations. Duong and Zakrzewska **[58]** have calculated the normal modes for the three canonical conformations of DNA known as the A, B and Z forms and considering the two possible nucleobase sequences. They reported a wide quantity of modes corresponding to the deformation of the double-helix between 0.3 THz and 7.50 THz. Both the A-cytosine/guanine and A-thymine/adenine exhibited the highest number of vibrational modes at 1.8 THz (by about 13 distinct vibrational modes), whereas the two other allomorphs shown the largest number of vibrational modes around 600 GHz, 3 THz and 900 GHz, 2.4 THz for B and Z conformation, respectively, without discrepancies regarding the two alternating base sequence. They additionally shown that B-DNA was the form subjected to the largest vibrational amplitudes, whereas A-DNA had the lowest, Z-DNA being the intermediate case. Other studies reported the theoretical capacity of terahertz to interact with fractional propeller twist modes or hydrogen-bond breathing modes **[59; 60]**. Heretofore, it was only theoretical investigations. To compensate for the lack of experimental data, Brucherseifer *et al* **[61]** studied the response of hybridized-DNA (double stranded) and denatured-DNA (single stranded) within the 0.2-2.5 THz window. They have shown that due to denaturing, the refractive index of single stranded DNA was globally decreased by 0.1 over the frequency range. This was the first observation of demarcation between binding-state-dependent properties of polynucleotides in the terahertz frequency range.

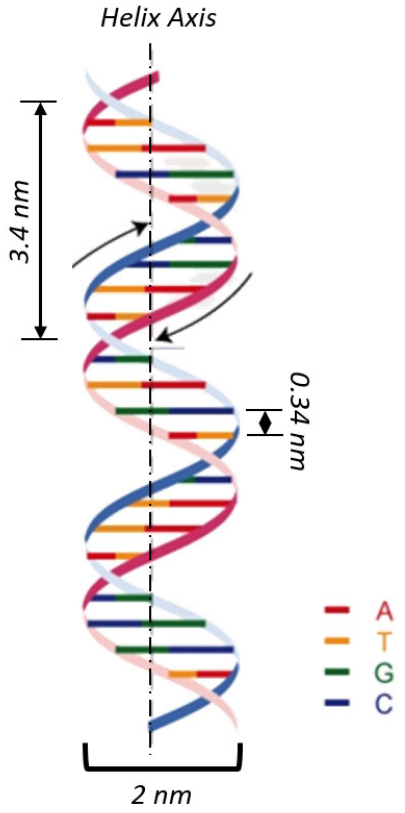


Figure I.14: Schematic diagram of the DNA helix structure.

Proteins

In the previous section we mentioned that nucleic acids carry the genome information. Proteins are the direct executors of the tasks stipulated by that information. Proteins are macromolecules synthesized in the cell ribosome (Figure I.15). They consist of chains of various amino acids whose combination is expected to drastically affect the spectrum of the molecules. Proteins are found by thousand inside cells, at their plasma membrane surface and even within the extracellular matrix (collagen, fibronectin and laminin). Proteins exhibit particularly small dipole moments. Thus dipole reorientation with regard to an external terahertz field is not expected to be of vast contribution to the spectral characteristics. One can anticipate that the most probable origins of features within the terahertz regime on probing proteins are intramolecular collective vibrations and side chain rotations. First studies on proteins have been conducted on bovine serum albumin and collagen pellets **[62]**. Authors confirmed that the dissipation of the incident energy among samples was consistent with the Beer's law behavior. In an interesting way, the bovine serum albumin absorbance increased almost linearly with frequency while for collagen, the absorbance increased more rapidly with increasing frequency. Markelz and colleagues concluded that sample broadband absorption between 0.06 and 2 THz was caused by a large density of low frequency collective modes. Nonetheless, *in vivo* proteins behave in a quite different fashion than the one that may be observed as lyophilized structure. It is in fact known that *in vivo*, proteins interact with water molecules. This interaction gives arise to a hydration shell, in which the surrounding network of water molecules drives fluctuations among the protein surface groups. The hydration sphere taking place around proteins is affected out to at least 1 - 1.5 nm from surface proteins. Since the terahertz regime covers the band of collective dynamics in proteins but is as well sensitive to the rotational and translational motion of water molecules, more recent studies aimed to probe not just the self-assembling proteins but also the hydration sphere dynamics **[63]**. To describe the hydration water as a part of the protein global dynamics, Kim *et al* **[64]** studied ubiquitin, a protein which serves as a marker for other protein degradation. Considering first hydration water behaving like ubiquitin-free molecules, they observed a stronger absorption for ubiquitin buffered solved solution than the buffer itself. They highlighted that the increase in absorption was actually caused by water dynamics alteration among the hydration shell up to 18 Å from the protein surface. The interaction induced a shift of absorbance from sub-THz to higher-frequency modes. Correcting the hydration water - protein surface group coupling, they finally observed that the protein and its hydration water absorbed less than the buffer solution. Similar work were conducted by Xiu and colleagues **[65]**, leading to an analogous conclusion for bovine serum albumin.

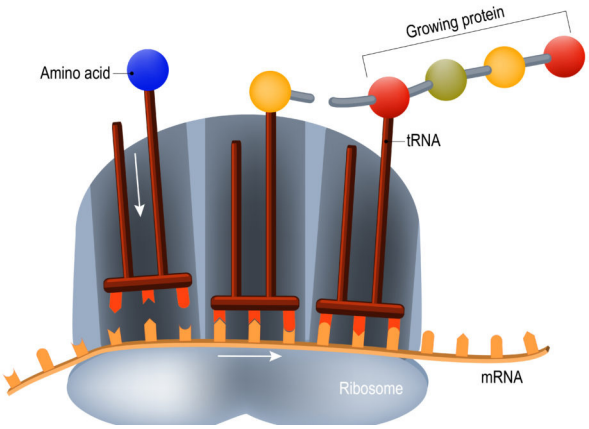


Figure I.15: Protein synthesis in the ribosome complex. Proteins are assembled as stipulated by the encoded genetic information. The messenger-RNA (mRNA) serves as a template while the transfer-RNA (tRNA) acts as the physical link between the mRNA and the amino acid sequence of growing proteins.

I.3.4 Super Cellular Level Responsiveness

Sensing life sub-units is not an end on itself. The purpose behind these investigations is evidently to understand the interaction mechanisms at a more complex stage, where sub-units join with each other to build tissues and organic fluids. Tissues and fluids are the sub-units of organs and thus are of primary importance when dealing with injuries, diseases and disorders. The biological state of tissues generally reflect the organ condition and therefore are often key factors to detect, to characterize and to monitor eventual organism disfunctions. Interest in medical applications necessitates an understanding of the fundamental interaction of terahertz radiations with biological tissues.

Heretofore, we looked at sub-cellular units from the perspective that their behavior with respect to an external terahertz field would affect the global response at a multi-cellular level. While it is true that dielectric spectra at super-cellular levels would in fact depict the complexity of biological organisation, that same complexity may concomitantly blur tissue sub-unit features. Thereupon, one can ask oneself, if at a super-cellular level the dielectric response would denote appreciable changes depending on tissue components. As has been endlessly pointed out, terahertz waves are significantly sensitive to water molecules. Water responsiveness to terahertz stimuli may mask out the global information from inspection either by strongly dissipating the incident sensing energy or by blurring slighter interactions with alternate sub-units. Although the presence of water limits the depth of tissue that may be sensed, the degree of hydration and the water content may reveal a disease stage. In addition, one can intuitively predict that the dielectric response is impacted by the fraction of aforementioned organism components (*i.e.* carbohydrates, proteins, lipids and nucleic acids for cells) regarding the fraction of water found in tissues. Table I.3. describes approximately the part of water, lipid, protein and carbohydrates found in human organism.

The question that now one should raise is in which biomedical context the use of terahertz waves may have a positive impact. Based on terahertz radiation properties, the research toward biomedicine is split along two axis; first, on prognosis near the body surface, considering the limitation on the depth penetration imposed by water; Second, on *in vivo* excised tissues which may, in the future, be attainable to terahertz irradiation thanks to eventual technology advances.

Table I.3: Relative mass (%) of water, lipid, protein, carbohydrate for various body tissues **[66]**.
[a]: Approximations are particularly schematic for breast composition since component content varies among individuals and tissue excision location **[67]**. In addition, significant variations in composition have been observed for adipose tissue and mammary glands as reported by **[66]**.

Tissue	Relative Mass (%)				
	Water	Lipid	Protein	Carbohydrate	Residuals
Adipose tissue[a]	21.2	74.1	4.4	-	0.3
Adrenal gland	58.1	26.0	15.5	-	0.5
Aorta	72.1	1.8	25.0	-	1.4
Blood—erythrocytes	64.0	0.5	34.7	0.1	1.1
Blood—plasma	91.9	0.7	6.6	0.1	1.0
Blood—whole	79.0	0.6	19.6	0.1	0.9
Brain—cerebrospinal fluid	99.0	-	-	0.1	0.9
Brain—grey matter	82.6	5.3	10.0	1.0	1.5
Brain—white matter	68.5	18.1	11.5	1.0	1.4
Breast[a] **[67]**	34	60	6	-	-
Connective tissue	60.4	1.0	36.6	0.9	1.3
Eye lens	64.1	2.0	33.6	-	0.4
Heart	75.9	6.2	17.1	-	0.9
Heart—blood filled	77.7	2.8	18.6	0.1	0.9
Intestine (small)	80.6	5.9	13.0	-	0.7
Kidney	76.6	4.8	17.7	-	0.9
Liver	74.5	4.6	17.6	2.2	1.2
Lung—parenchyma	80.6	1.3	17.1	-	1.1
Lung—blood filled	79.8	1.0	18.2	-	1.0
Lymph	91.9	0.7	6.6	0.1	1.1
Mammary gland[a]	51.4	30.9	17.4	-	0.4
Muscle	74.1	4.2	19.8	1.0	1.0
Oesophagus	74.1	4.2	19.8	1.0	1.0
Ovary	82.8	2.3	14.0	-	1.0
Pancreas	73.3	12.8	13.1	-	1.2
Prostate	83.3	1.2	15.0	-	1.1
Skeleton—cartilage	72.0	-	11.0	11.0	3.0
Skeleton—cortical bone	12.2	-	24.6	5.2	58.0
Skeleton—red marrow	39.7	39.7	20.0	-	0.3
Skeleton—yellow marrow	15.3	80.4	4.0	-	0.5
Skin	65.3	9.4	24.6	-	0.7
Spleen	78.7	1.8	18.6	-	1.3
Stomach	76.3	6.2	17.0	-	0.8
Testis	82.7	4.5	12.0	-	1.1
Thyroid	78.4	4.4	14.0	2.5	1.1
Trachea	74.1	3.8	18.9	2	1.2

Surface Diagnosis

In the field of medical diagnostic, water plays a central role since its absorption drastically limits the penetration length of terahertz radiations. Consequently, in the first instance, main studies have focused on skin examinations. The skin is the body's largest organ which prevents from homeostasis and whose water content is by about 65% per mass unit. The skin consists of two primary layers: the epidermis and the dermis. The epidermis is the outermost strata of the organ and thus serves as a barrier to external organisms, radiations, temperature changes and molecules (Figure I.16). In turn, the epidermis is made of several sub-layers each specialized to accom-

plish a dedicated function, among which the stratum corneum, the outermost layer. The stratum corneum is made in average of 20 layers of flattened cells without nuclei. The stratum corneum takes part in various physiological tasks among which hydration regulation. It is commonly admitted that hydration level in stratum corneum reflects the post-injury response. In that sense, Cole *et al* **[68]** investigated in the early-2000s the human skin and especially the stratum corneum hydration with terahertz pulses whose frequencies were ranging from 100 GHz to 3 THz. They performed various measurement to record the reflected pulses from different locations of the hand and arm. To ensure that the skin was constantly at the same position with respect to the emitter, the monitored location was placed in contact with a quartz window serving as a reference plane. They observed over time that the peak-to-peak amplitude was decreasing in an exponential fashion. The hydration rate was expected to be related to this observation and authors concluded that the decay was due to skin occlusion by the window, hindering water transport through the stratum corneum layer. The hydration level of the stratum corneum is expected to differ among the various skin tissues. Echchgadda and co-authors **[69]** have reported different absorption coefficients and refractive indexes for ventral forearm, dorsal forearm and palm, assumed to be due to varying hydration levels. Unsurprisingly, even though the dielectric properties have been reported to be different for each skin tissue, the highest variations in the refractive index were observed between 100 and 400 GHz. This spectral window is not random since it is where lies the right-hand side of the relaxation of the bulk-water dielectric-active mode at 19 GHz. The trend to be sensitive to hydration level is of clinical relevance when dealing with burn injuries. In the early post-injury period, physicians sort burns in three distinct categories; first, second and third degree. While the first group often involves superficial injuries, the second and third degree require medical care in the form of tissue grafting, surgery and careful monitoring. In some cases, second degree cases may turn in third degree due to the lack of epithelium formation sites inducing dermis desiccation. In order to evaluate injury triage by terahertz technique and thus to determine an effective therapy for patient suffering from burns, Arbab *et al* **[21]** have reported an *in vivo* study on second and third degree burns, after 72 hours, based on the rat animal model. Findings revealed that the dielectric response of the different degrees was not only consistent with the presumed water content but that scattering sites within the organ layers were also begotten response variations. These preliminary findings are relevant for future diagnostic techniques aiming to discriminate between second degree burns that will naturally heal from those requiring skin grafting and surgery procedures.

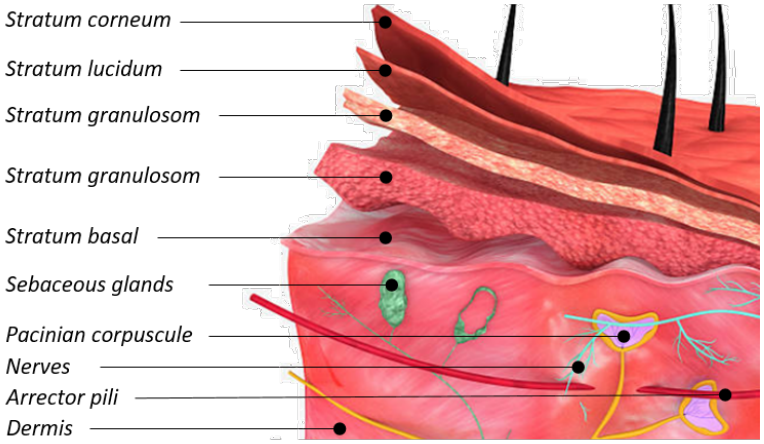


Figure I.16: Skin diagramm.

Another biomedical application on surface prognosis involves the eye cornea (Figure I.17). The cornea is the outermost layer of the eye and is often the location of disabling diseases that can impair vision. The cornea thickness varies from the outskirt (0.8 mm) to the center (0.5 mm). The cornea is the eye first refractive element and represents 66% of the total of ocular diopters, the rest of them being found in the eye lens. The hydration level of the cornea is by about 85% and the

consistency of such a hydration level is of paramount importance to preserve both the focusing and the transmission properties. Variation of the water content among the cornea is often a sign of injury or disease process which will, with time, cause eye disfunction or blindness. Currently, the corneal hydration is calculated from an acoustic measurement of the central corneal thickness and an empirical law extracted from a large set of patients. However, such a procedure suffers from the variety of changes that may occur upon patient physiology, age, healthiness and environmental factors. As reported by Bennett *et al* **[70]**, terahertz radiations would take advantage of a direct interaction between the probing beam and the cornea water. They successfully observed the evaporation of water from an *ex vivo* porcine cornea, reporting terahertz-images whose contrast are assumed to be due to the hydration level. The study was performed between 100 GHz and 1.2 THz and authors emphasized that the contrast mechanism was almost lost at upper frequencies. Additional studies have been published suggesting terahertz as a potential technique to overcome the existing clinical issues for corneal hydration measurement **[49; 71]**. However, while overcoming non-direct and invasive issues reported for other assessment techniques, terahertz suffers from the eye blinking effect which makes difficult the acquisition of accurate data. To surmount blinking, one may keep the eye right open but therefore, induces forced corneal desiccation. So far, no clinical implementations have been realized.

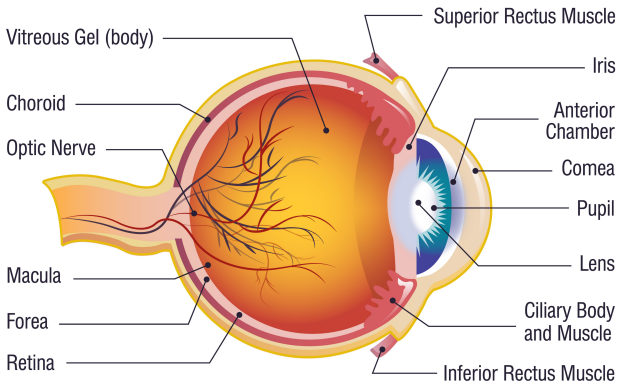


Figure I.17: Eye anatomy.

Teeth are also under investigations. Teeth consist of three distinct layers among which two are made of hard tissues; first, the enamel whose structure contains 95% of inorganic calcium and phosphate ions. Approximately 2% of the enamel is made of specific proteins known as enamelin, water making up the remainder of the enamel (Figure I.18). When a tooth erupts from the gums, the epithelial cells are broken down leaving the enamel without any living cells in its composition. The lack of living cells removes enamel's ability to regenerate or repair itself. Second, the dentin which is the major constituent of teeth. Similarly to the enamel, the dentin is mainly made of inorganic compounds which count for 70% of its structure. However the dentin is made of a larger amount of organic material representing 20% of the total weight (mainly collagen) and 10% of water. Although both the enamel and the dentin are not vascularized, the dentin is on its part, innervarted. Last, the pulp which is the neurovascular bundle central of teeth and is filled with a wide quantity of different cells having specific functions. Consequently, the water content within the pulp is by far greater than the two other layers. One could have mentioned the cementum layer, lying between the dentin and the pulp. However this specific layer takes place under the gums and thus is not of primary interest for terahertz inspections since gums are particularly wet structures. First investigations attempted to evaluate the ability of terahertz radiations to differentiate between the teeth layers. Usually, due to the dielectric sensitivity of terahertz radiations, the discrimination task focuses on the detection of different pulses exhibiting optical delay correlated to the intrinsic complex refractive indices and tooth layer thicknesses. To this aim, Berry *et al* **[72]** performed transmission measurements with a useful bandwidth of 0.5-2.5 THz on histology slides of enamel and dentin in the coronal direction. The thickness of sections were known,

and thus the only remaining unknown parameter was the complex refractive index. They reported interesting differences between both the mean refractive index and the mean absorption coefficient of each tissue. By means of 44 measurements on enamel and 72 on dentin, they extracted a mean refractive index and a mean absorption coefficient of $(3.06 \pm 0.09; 62 \pm 7 cm^{-1})$ and $(2.57 \pm 0.05; 70 \pm 7 cm^{-1})$ for enamel and dentin, respectively. Authors reported similar values to those by Ciesla *et al* **[73]** and by Hoshi *et al* **[74]**. Extracted complex refractive indices are consistent with the presumed larger absorption coefficient with greater hydration level. Whilst no specific spectral features have been reported, the variation in the refractive index and absorption are assumed to be enough to discriminate between tooth strata. To validate this assumption, Crawley *et al* **[75]** reported a non-destructive imaging method based on terahertz technology to provide accurate 3-D images of teeth structures. They successfully identified the enamel dentine junction in 91% of cases. However, these studies, have all been performed on *ex vivo* samples. The implementation for *in vivo* analysis are quite more complex, since probing inside the mouth may alter the sensing radiation due the high level of humidity. One way to overcome the natural wet within the mouth would require the development of a in-contact probe as patented by **[76]**.

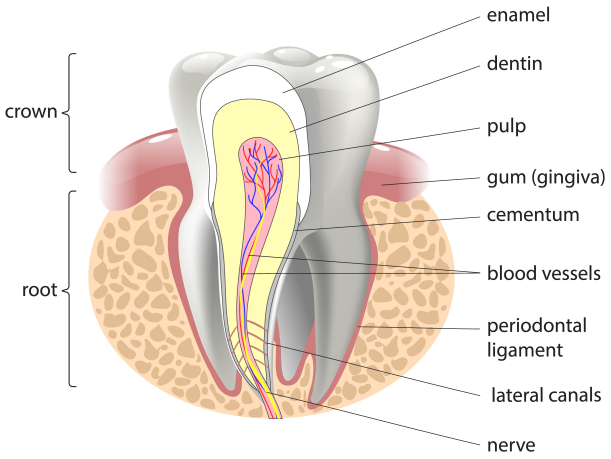


Figure I.18: Tooth structure.

Surveys for Oncology

Cancer is a general term employed to characterize the pleomorphism of cell growth diseases. It is a genome affliction involving the alteration of the genetic code of cells, including sexual haploid ones. Therefore, altered genes can be transmitted to descendants. Despite the plurality of neoplasia -alsmost all tissue systems can be affected-, in 2000 two cancer researchers named Hanahan and Weinberg published *The Hallmarks of Cancer* **[77]** in the peer-reviewed Cell journal in which, they argue that the complexity of cancer features could in fact be reduced to six hallmarks: (i) self-sufficiency in growth signals; (ii) insensitivity to antigrowth signals; (iii) evading apoptosis; (iv) limitless replicative potential; (v) sustained angiogenesis; (vi) tissue invasion and metastasis. Additionally, in 2006 McIntyre proposed an unifying concept on the carcinogenesis whose primary factor is suggested to be cell hydration **[78]**. Alternatively said, cancer cells exhibit an abnormally high water content in comparison to healthy ones.

In view of all the above considerations on cancer features, on tissue sub-unit candidates for interaction and on radiation properties, terahertz may offer some possibilities for non-invasive sensing of biological tissues for cancer recognition and demarcation. Nonetheless, let's remind that in-depth probing is not feasible since water strongly absorb terahertz radiations. Studies on the evaluation of terahertz waves regarding cancer recognition and discrimination are thus performed *ex vivo* on excised tissues. Depending on the capacity of the scientific community to pro-

vide and to prove the clinical relevance of terahertz waves for oncology, further technical developments will have to be implemented, *e.g.* endoscopes, probes, in-contact sensors and miniaturized spectroscopes, to assist oncologists and surgeons know-how in the fight against cancer.

Considering the numerous cancers that affect various organs and that have been reported in the medical literature, we propose to review some of them that have been assessed by terahertz radiations for cancer detection.

Blood cancer generally involves the disfunction of one of the three blood cells: the white blood cells fighting against infection; the blood vessels carrying the oxygen through the organism; the platelets ensuring blood clotting when injured. The most common cancer is the leukemia in which abnormal blood cells replicate. In general the diagnostic reveals a high content of white blood cells and low amount of red blood cells and platelets. The concentration of red blood cells may thus be an indicator of leukemia. Researchers have then focused on the characterization of the blood components from the blood plasma through red blood cells. Jeong *et al* reported a linear dependence of the absorption coefficient as a function of the red blood cells concentration in the blood plasma **[79]**. Reid and colleagues also reported different refractive indices for blood plasma and blood cells, 2.05 and 2.00 at 500 GHz, respectively **[80]**.

Brain cancer, often in the form of glioblastoma, is of great challenge since neoplastic regions do not exhibit clear margins for demarcations. Oh *et al* inspected the terahertz response in reflection configuration of freshly excised brain tumors, from the orthotopic glioma rat model **[81]**. They reported higher water content in stroma tissue, caused by angiogenesis and body fluid presence around necrotic debris. However, authors emphasized that water content was not the only source for contrast. Comparing the hematoxylin-eosin histology slides with terahertz images, they reported that tissue regions whose reflectivity was relatively high were in fact corresponding to high cell density area. Other groups have also reported possible abnormal tissue demarcation from healthy matrix and differences between grey and white brain matter **[82; 83]**.

Breast cancer will be addressed in further details in the subsequent section and chapters, being the central investigation of the present work.

Colorectal cancer appears within the colon or the rectum and usually arises from an abnormal growth of the mucous membrane called a polyp that may become, over time, cancerous. The disease is the second most common cause of cancer-related death. Survival rates are greater if dysplastic lesions can be treated or removed rapidly. Reid *et al* have conducted a statistical study on freshly excised colonic tissue: healthy, dysplastic and malignant **[84]**. They used a multinomial logistic regression analysis to sort the samples as a function of their histology. Authors reported a sensitivity of 82% and a specificity of 77% when comparing parameters from healthy and malignant tissues. On the other hand, when comparing dysplasia and healthy tissue, authors returned a sensitivity of 89% and a specificity of 71%. No comparison between dysplatic and cancerous tissues was however reported. Eadie *et al* reused the data from **[84]** and enhanced the true-positive rate up to 92% by training a neuron network working with principal component analysis reduced dataset **[85]**. However, authors emphasized that a simple decision tree based on water and air reference parameters achieved surprisingly similar results, with a sensitivity and specificity of 90% and 88%, respectively.

Gastric cancer: the first investigations on the potential of terahertz radiations to distinguish between benign and malignant region were performed by Hou *et al* in 2014. They performed a battery of measurements on human dehydrated normal and cancerous gastric tissues **[86]**. Since tissues have been desiccated, authors underlined that any observed source of contrast would shed light on additional contrast-contributing factors other than water content. Thanks to the dehydration, authors were able to perform transmission measurements. They revealed spectral features arising between 200-500 GHz and 1-1.5 THz for cancerous tissues. In addition they projected the principal components with the most significant differences, with respect to each others. They reported that the best overall results were attained when using the first four significant principal components as the input data set for the normal versus tumor classification task. The complexity of probing fresh tissues have been conducted a year later by another team **[87]**. Due to the

presence of water they conducted their experiment in reflection. Eight freshly excised early gastric cancer samples were inspected. They reported higher reflection intensities for abnormal regions than normal mucosa. While authors were able to distinguish early gastric cancer from normal tissues (for 6 cases over 8), they shed light on specific limitations regarding the terahertz characterization of signet ring cell carcinoma which were not well correlated to pathologically mapped images. Consequently, although reported work on dehydrated samples have suggested that terahertz may be used for illness identification, the most recent paper shades these previous results, mentioning the need for further technical developments or numerical methods to enhance the true-positive discrimination rates for early gastric cancer and other digestive cancer organs.

Liver cancer also known as hepatic is often caused by cirrhosis of various etiologies. Like other internal organs, liver can not be directly probed using terahertz radiations. First analysis of liver cancer was thus performed by Enatsu *et al* on paraffin-embedded samples **[88]**. It was found that the refractive index and the extinction factor were less inhomogeneous for healthy tissue than malignant ones. The non-homogeneity among dielectric properties of cancerous tissues was explained by the non-uniformity of the density of necrotic tissues. Additional study on the origin of contrast were conducted by a research group in Hong-Kong **[89]**. Freshly harvested healthy and cirrhotic rat liver samples in reflection geometry with a hand-held terahertz probe **[90]** were sensed. Authors reported that the absorption coefficient of cirrhotic tissues was higher than the absorption of benign tissues. In an interesting fashion, the absorption was still higher after formalin fixing. This observation led the authors to conclude that the contrast agent was not only the water content among cirrhotic samples since formalin removes water from tissues. By doing so, formalin alter the structural organisation of the tissue. In a nutshell, terahertz radiations were not only sensitive to water content but also to structural changes among liver biological structure.

Lung cancer was investigated with terahertz radiations by Brun *et al* **[91]**. They obtained medically relevant segmented images by means of terahertz spectroscopy-imaging and clustering algorithm working with principal component analysis. They linked the contrast mechanism among samples to cellular-level differences. More than discriminating between benign and malignant regions, they additionally reported demarcation within cancer, determining cancer sub-types. However, although the aforementioned study revealed promising features, the lack of complementary studies on fresh lung cancer samples remains an ongoing issue.

Oral cancer is in 90% of cases a squamous cell carcinoma or a mucoepidermoid carcinoma **[92]**. Oral cancer is particularly tricky to detect in its early stage due to its asymptotic nature and carries a high mortality rate. As for the majority of cancer, detection in the early development would reduce both the mortality and morbidity rates. Inside the oral carcinoma, a rapid angiogenesis takes place leading to an increase of the global water content. While being a possible mechanism for contrast, water content additionally limit the penetration depth of the incident radiation. Sim and colleagues investigated the response of oral cancer tissues at room and at frozen temperature to circumvent the high hydration level found in oral tissues **[93]**. Authors underlined that based on pathology analysis, freezing and thawing processes did not alter the tissues in a significantly fashion. It was found that samples imaged in a frozen state exhibited higher contrast between normal and abnormal mucosa than images at room temperature. Furthermore, frozen state inspection allowed to detect cancer hidden deeper than 1.2 mm from the tissue surface. These promising results for cytological examination are yet to be validated by supplemental studies.

Skin cancer is probably one of the most relevant clinical case to be studied through terahertz irradiation as it is the outermost organ of the body. Terahertz waves have been reported to be serious candidates for skin cancer detection and demarcation as suggested in a first study by Woodward *et al* in 2002 **[94]**. Differences between healthy and malignant tissue refractive indices have been mentioned as well as differences regarding the dissipated energy portion. According to Yu *et al* **[95]**, the predominant source of contrast is the higher hydration level in carcinoma as reported by **[96]**. However, as shaded by Wallace *et al*, the contrast in skin cancer in the terahertz regime is not only caused by water content but additionally arises from cell density which turns into scatter-

ing sites when the aggregation size matches the incident wavelength **[97]**. Furthermore, increases in actin, amyloid, keratin, collagen within basal cell carcinoma are also contributing factor for contrast mechanisms.

Physiological changes associated with carcinogenesis commonly lead to an increase of water, the presence of some specific proteins, a modification of the vasculature, a decrease of lipids and an increase in cell density near the original cancerous site. Pre-clinical studies on the evaluation of terahertz capacities to discriminate between benign, dysplastic and malignant tissues for different organs have reported various contrast-contributing instruments. Although the biochemical mechanism behind such a contrast has yet to be fully understood, the most promising avenue seems to arise from the inherent water content discrepancy between normal and abnormal tissues. While it is true that additional sources may affect the dielectric spectrum, water has been acknowledged to be a regular source of contrast. Heretofore, results have been promising but worldwide efforts are still required to test, to validate and to establish the most suitable experimental techniques, to determine the most accurate numerical methods, and to formulate the most solid interaction models, all this with the aim of providing the most efficient tools for patient recovery.

I.4 Breast Cancer

Cancer has various etiologies, takes various forms and may occur in almost all the organism. Because of that, the present thesis work solely focuses on the specific case of breast cancer which develops in one out of every nine women. Before embarking upon an investigation, it is necessary to collect and to collate all available relevant information about the site of the disease. Thus, it is appropriate to provide a succinct description of the breast anatomy and the associated cancers. All of this will draw the framework of the present work and will highlight the main goals of this thesis.

I.4.1 Breast Anatomy & Malignant Illness

In women's anatomy **[98]**, the breast is the tissue overlaying the pectoral muscle (Figure I.19). The breast mainly consists of fat, that determines its volume. The breast is also made of ligaments and connective tissues that support the overall structure. Breast tissue also contains nerves, lymph nodes, lymph vessels and a vascular network. Below the areola, at the nipple level, opens a dozen of lactiferous ducts. On the rearward part of the nipple, those ducts divide in narrower ducts until reaching a terminal ductal-lobular unit. Terminal ductal-lobular units are epithelial structures that produce milk during lactation. Since the work of Wellings and Jensen in 1975, it was recognized that these structures are also the primary source of the largest majority of breast cancer precursors and carcinogenesis **[99]**.

Breast cancers are malignant tumors that occur in the breast. Different types of breast cancers have been reported depending on cells from which cancer develops **[100–102]**. The most frequent tumors are adenocarcinomas that arise from the epithelial cells that line the ducts and the terminal ductal lobular units. The generic evolution of a ductal adenocarcinoma is depicted in Figure I.20. Adenocarcinomas are a group of malignant tumors which clusters about 20 different types of breast cancers. Adenocarcinomas are sorted in two main categories. First, *in situ*: abnormal cells are found only in the ducts or the lobules without spreading in neighbor tissues; Second, *invasive*: abnormal cells are not confined to the initial tissue site but have started to infiltrate the fibrous or the fatty tissue surrounding the ducts and the lobules. Furthermore, depending on its origin, adenocarcinoma may be referred to as ductal or lobular. At the time of surgery, different tumor characteristics are assessed such as tumor stage, tumor grade and tumor immunophenotype. The grade aims to compare the current cell differentiation state with normal cells. In general, the higher the differentiation, the lower the grade. Breast staging is, for its part, based on the **TNM** system referring to the size of the tumor (**T**), the spreading within lymph nodes (**N**) and whether

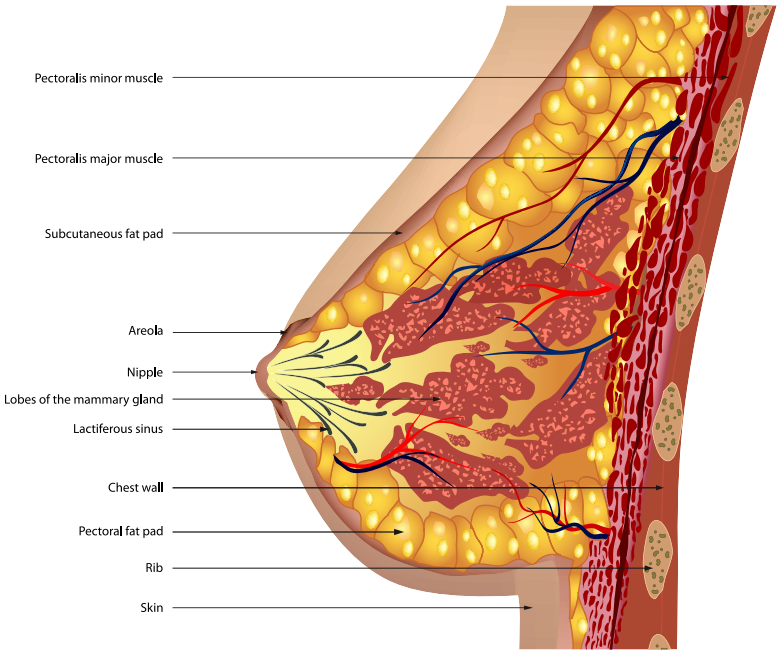


Figure I.19: Breast anatomy.

or not the tumor has metastasized (**M**). While stage 0 denotes ductal or lobular adenocarcinoma *in situ*, stages 1-3 stand for illness that has spread in the breast or the lymphatic network and stage 4 refers to cancer which has spread beyond breast tissue and local lymph nodes. Table I.4 aims to summarize the classification of breast cancers from the most favorable description to the worse.

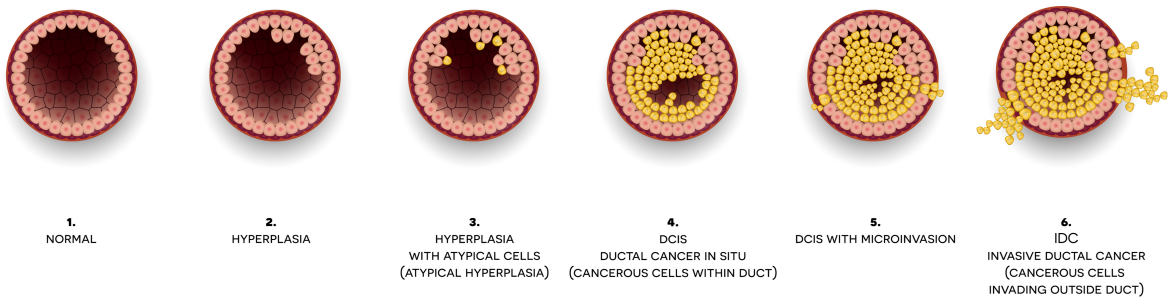


Figure I.20: Evolution of a ductal adenocarcinoma.

Now that the outline of breast cancer is established, we will address the different treatments and we will see that it is at this specific level that terahertz radiations are about to play a particular role.

I.4.2 Treatments

Various breast cancer managements exist in order to promote healing. Treatment in patient depends on the physical and biological characteristics of the disease established by the tumor stage, the tumor grade and the tumor immunophenotype. Two types of therapies can be distinguished; local and systemic treatments. Local therapies aim to prevent from the development and the spreading of early-stage cancer, while systemic therapies aim to treat micro or macro metastatic disease. The present thesis work takes place in the first category, notably in breast conserving surgery (BCS) **[103; 104]**, where a lack of intraoperative accuracy has been reported.

Table I.4: Cancer classification system. Cancers are sorted by histological appearance. Depending on the histological type, they are classified as ductal or lobular carcinoma. Carcinoma *in situ* denotes cancers which have not spread out the initial tissue compartment, while invasive carcinoma have spread in surrounding tissues. The grade states for the differentiation level of cells. Tissues exhibiting cells least like normal ones are by and large high grade carcinomas.

Prognosis	Histology	Grade
Favorable	*In situ*	Well differentiated
↓		Moderately differentiated
Worse	Invasive	Poorly differentiated

Local Breast Cancer Surgery

Surgery is often the primary approach when facing breast cancer. Depending on the diagnosis established by the pathologist, surgery management may be addressed in different ways: first, lumpectomy during which, a discrete fraction of breast tissue surrounding the tumor is removed. In such a procedure, the quantity of tissue removed is limited and is thus considered as a BCS; Second, mastectomy which is carried out when total breast removal is required. In contrast to BCS, mastectomy is not a breast conserving surgery since the entire breast has been withdrawn. Let's note that other specific procedures may exist, but are carried out in more rare cases.

Breast conserving surgery, also called tylectomy, segmental mastectomy or as aforementioned lumpectomy, is a surgery procedure during which an exeresis of an early-stage breast tumor is carried out. The removal is done along with the boarder of healthy tissue, also called surgical margins. Those margins are the normal tissue that directly surrounds the tumor. Following the removal procedure, a pathologist analyzes the excision and particularly the margins to spot the eventual presence of remaining cancer cells. The pathologist then states on the completeness of the surgery by distinguishing positive from negative margins. Positive margins reflects the presence of cancer cells on the boarders of the excision while negative or clear margins denote the absence of cancer cells **[105]**. Currently, the intraoperative precision with which are delineated tumor margins remains weak and periodically leads to a second surgery to assess the entire breast cancer removal. In fact, following histopathologic examination of excised tissues, up to 20% are shown to have tumor at the margins **[106; 107]**. A second surgical procedure is then required to minimize the recurrence risk but results in increased morbidity rate. The entire procedure, from surgery to histological analysis is depicted in Figure I.21. Consequently there is a clear need for an operating surgery room device which could accurately define breast tumor margins during breast conserving surgeries [90]. At present, there is no technology available to do this. In view of the foregoing sections, the use of terahertz radiations may be indicated to provide additional information on the cleanliness of surgical margins. Their sensitivity to water content and other aforementioned contrast-contributing factors may in fact supply a direct examination of the margins without going through the histological analysis chain. The investigation on the validity of such a statement is the central work of this thesis work. Before going on, primary and previous works on the question have to be acknowledged.

I.4.3 Terahertz Waves for Breast Cancer Detection

Since physicians continuously look for an intra-operative technique that can assess the margins of excised tissue to accurately remove the tumor with adequate negative margins, Fitzgerald *et al* **[108]** started to investigate the ability of terahertz spectroscopy-imaging to map the breast tumor margins. They analyzed several freshly excised breast tissues and used time-domain available parameters to provide contrast among images. Maximum and minimum values of the terahertz electric field recorded at each tissue locations were particularly studied. Using those parameters, they reported good correlations between histology maps of the surgical sites and the terahertz images. In addition, they extracted the refractive index and the absorption coefficient for both normal and

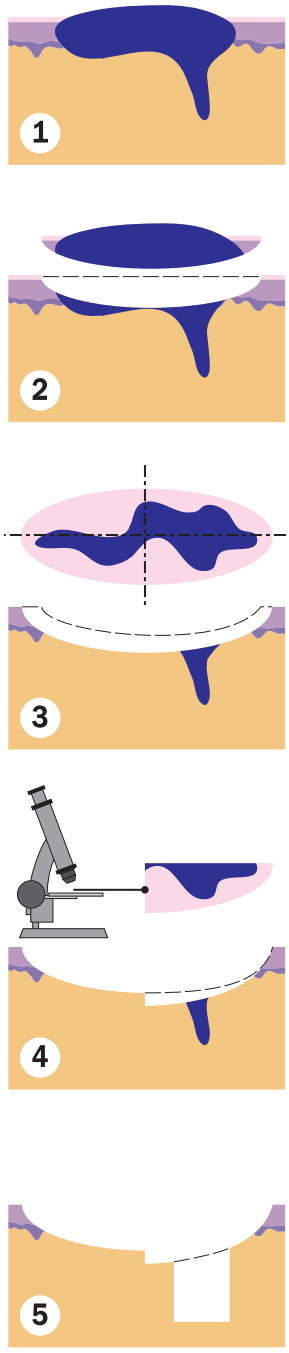


Figure I.21: Breast cancer surgery management: from initial partial mastectomy to second intervention, and histological examination in between. Step 1: abnormal tissue may extend beyond the visible fraction of the tumor. If not completely removed, the cancer will recur. Step 2: The surgeon removes the visible portion of the tumor. Step 3: The excised tissue is formalin-fixed and paraffin-embedded. Then, the tissue is divided into thin section of around 5µm. Hematoxylin and eosin (H&E) stain is carried out to highlight cell cytoplasm and nuclei and the extracellular matrix. Hematoxylin stains cell nuclei blue, eosin staining pink the remainder of the section. Step 4: the pathologist microscopically examines the margins and the undersurface of each section. Step 5: if margins are defined as positive, the surgeon returns to the patient to remove the remaining abnormal cells.

abnormal tissues. They reported higher real and imaginary refractive index for cancer tissue in comparison to those extracted from healthy sites. Such kind of observations are consistent with what has been aforementioned concerning the sensitivity of terahertz waves to water content and structural changes, such as cell and protein density increase and decreased lipid density. This pre-

liminary work prepared the ground for breast tumor margin cleanliness assessment. Since that first study, further developments and advances have been reported. Throughout the manuscript, the literature will be discussed and compared with both the results and advances achieved during the thesis.

I.5 Conclusion

Advances toward applications for life-science and medicine have been numerous during the last decades. The field has effectively attracted a lot of interest for two main reasons: first, the development of new sources have bridged the terahertz gap between the electronic and the optic domain of the electromagnetic spectrum; Second, the presence of the γ-dispersion domain, attributed to the relaxation of mobile and polar small molecules, that starts at the lower limit of the terahertz regime. That is to say, a wide quantity of biological processes occur at the pico and sub-picosecond time window.

Consequently, various questions have been raised concerning the interaction of biological objects with terahertz radiations. Rationally, when probing Life with electromagnetic radiations, one wants first to ensure that sensing would be achieved without damaging objects under inspection. We reported that even though terahertz radiations may in fact have biological effects for long and powerful exposure, powerless and short exposure should be limited to thermal effects.

We have seen that investigations have been conducted both at the sub-cellular and at the super-cellular level. Studies at the first level highlighted specific interactions between life sub-units and terahertz radiations, notably water molecule relaxations. Processes arising from terahertz-water interactions are both a predominant source of information and a robust shield against biological monitoring. In fact, water exhibits two specific relaxations within the terahertz band, leading to particular features among the dielectrograph. However, even though the presence of such processes are valuable assets, the associated energy dissipation is high and might compromise the access to the biological information beyond water. Moreover, we have seen that water is used to interact with other life sub-units, such as proteins, forming hydration shells within which, water molecules loose their dynamics and modify the natural behavior of the joined structure regarding the stimuli.

The sensitivity of terahertz waves to water initiated research at the super-cellular level, *i.e.* on tissues. Although tissues are globally made of various fractions of proteins, carbohydrates and lipids, water content is expected to dictate the dielectric response. As it has been reported by cancer researchers, water content in malignant tissues is greater than normal ones. This features has driven studies for oncology toward tissue malignancy estimation.

Breast cancer detection by means of terahertz radiations aims, ultimately, to develop an intra-operative system that provides accurately, non-invasively and quickly the surgical margin positiveness and negativeness.

The manuscript will now take the reader step by step through the investigations led during this thesis, from far-field spectroscopy-imaging of biological objects and freshly excised breast tumor to the use of a near-field silicon-based terahertz subwavelength imager for intra-operative breast tumour margin identification developed within the scope of the Farsense and Nearsense projects. These projects involve three institutions, namely the University of Wuppertal, GERMANY, the Bergonié Institute, FRANCE, and the University of Bordeaux, FRANCE.

I.6 Bibliography

[1] Samuel P Langley. Xviii. the bolometer and radiant energy. *American Academy of Arts and Sciences, Boston. Proceedings (1846-1906)*, 16:342, 1881. 14

[2] SP Jamison, Jingling Shen, DR Jones, RC Issac, B Ersfeld, D Clark, and DA Jaroszynski. Plasma characterization with terahertz time–domain measurements. *Journal of applied physics*, 93(7):4334–4336, 2003. 14

[3] M Walther, P Plochocka, B Fischer, H Helm, and P Uhd Jepsen. Collective vibrational modes in biological molecules investigated by terahertz time-domain spectroscopy. *Biopolymers: Original Research on Biomolecules*, 67(4-5):310–313, 2002. 14

[4] David T Leisawitz, William C Danchi, Michael J DiPirro, Lee D Feinberg, Daniel Y Gezari, Mike Hagopian, William D Langer, John C Mather, Samuel Harvey Moseley Jr, Michael Shao, et al. Scientific motivation and technology requirements for the spirit and specs far-infrared/submillimeter space interferometers. In *UV, Optical, and IR Space Telescopes and Instruments*, volume 4013, pages 36–46. International Society for Optics and Photonics, 2000. 14

[5] George H Rieke, MW Werner, Rodger I Thompson, EE Becklin, WF Hoffmann, JR Houck, FJ Low, WA Stein, and FC Witteborn. Infrared astronomy after iras. *Science*, 231(4740):807–814, 1986. 14

[6] David F Plusquellic, Karen Siegrist, Edwin J Heilweil, and Okan Esenturk. Applications of terahertz spectroscopy in biosystems. *ChemPhysChem*, 8(17):2412–2431, 2007. 14

[7] JT Kindt and CA Schmuttenmaer. Far-infrared dielectric properties of polar liquids probed by femtosecond terahertz pulse spectroscopy. *The Journal of Physical Chemistry*, 100(24):10373–10379, 1996. 14, 26, 27

[8] J Axel Zeitler, Philip F Taday, David A Newnham, Michael Pepper, Keith C Gordon, and Thomas Rades. Terahertz pulsed spectroscopy and imaging in the pharmaceutical setting-a review. *Journal of Pharmacy and Pharmacology*, 59(2):209–223, 2007. 15

[9] J Bianca Jackson, John Bowen, Gillian Walker, Julien Labaune, Gerard Mourou, Michel Menu, and Kaori Fukunaga. A survey of terahertz applications in cultural heritage conservation science. *IEEE Transactions on Terahertz Science and Technology*, 1(1):220–231, 2011. 15

[10] Gerald J Wilmink, Benjamin D Rivest, Caleb C Roth, Bennett L Ibey, Jason A Payne, Luisiana X Cundin, Jessica E Grundt, Xomalin Peralta, Dustin G Mixon, and William P Roach. In vitro investigation of the biological effects associated with human dermal fibroblasts exposed to 2.52 thz radiation. *Lasers in Surgery and Medicine*, 43(2):152–163, 2011. 15

[11] MR Scarfi, M Romano, R Di Pietro, O Zeni, A Doria, GP Gallerano, E Giovenale, G Messina, A Lai, G Campurra, et al. Thz exposure of whole blood for the study of biological effects on human lymphocytes. *Journal of Biological Physics*, 29(2-3):171–176, 2003. 15

[12] Avital Korenstein-Ilan, Alexander Barbul, Pini Hasin, Alon Eliran, Avraham Gover, and Rafi Korenstein. Terahertz radiation increases genomic instability in human lymphocytes. *Radiation research*, 170(2):224–234, 2008. 15

[13] Boian S Alexandrov, Kim Ø Rasmussen, Alan R Bishop, Anny Usheva, Ludmil B Alexandrov, Shou Chong, Yossi Dagon, Layla G Booshehri, Charles H Mielke, M Lisa Phipps, et al. Non-thermal effects of terahertz radiation on gene expression in mouse stem cells. *Biomedical optics express*, 2(9):2679–2689, 2011. 15

[14] VF Kirichuk and AA Tsymbal. Effects of terahertz irradiation at nitric oxide frequencies on intensity of lipoperoxidation and antioxidant properties of the blood under stress conditions. *Bulletin of experimental biology and medicine*, 148(2):200–203, 2009. 15

[15] Alfonsina Ramundo-Orlando, Gian Piero Gallerano, Pasquale Stano, Andrea Doria, Emilio Giovenale, Giovanni Messina, Mauro Cappelli, Marco D'Arienzo, and Ivan Spassovsky. Permeability changes induced by 130 ghz pulsed radiation on cationic liposomes loaded with carbonic anhydrase. *Bioelectromagnetics: Journal of the Bioelectromagnetics Society, The Society for Physical Regulation in Biology and Medicine, The European Bioelectromagnetics Association*, 28(8):587–598, 2007. 15

[16] VI Fedorov and GF Bakharev. Influence of thz radiation on early phase of seed germinating and yield of wheat. In *ICONO 2010: International Conference on Coherent and Nonlinear Optics*, volume 7993, page 799327. International Society for Optics and Photonics, 2011. 15

[17] S Hadjiloucas, MS Chahal, and JW Bowen. Preliminary results on the non-thermal effects of 200–350 ghz radiation on the growth rate of s. cerevisiae cells in microcolonies. *Physics in Medicine & Biology*, 47(21):3831, 2002. 15

[18] Henning Hintzsche, Christian Jastrow, Bernd Heinen, Kai Baaske, Thomas Kleine-Ostmann, Michael Schwerdtfeger, Mohammed Khaled Shakfa, Uwe Kärst, Martin Koch, Thorsten Schrader, et al. Terahertz radiation at 0.380 thz and 2.520 thz does not lead to dna damage in skin cells in vitro. *Radiation research*, 179(1):38–45, 2012. 15

[19] Alfonsina Ramundo Orlando and Gian Piero Gallerano. Terahertz radiation effects and biological applications. *Journal of Infrared, Millimeter, and Terahertz Waves*, 30(12):1308–1318, 2009.

[20] GP Gallerano, E Grosse, R Korenstein, M Dressel, W Mantele, MR Scarfi, AC Cefalas, P Taday, RH Clothier, and P Jepsen. Thz-bridge: an european project for the study of the interaction of terahertz radiation with biological systems. In *Infrared and Millimeter Waves, Conference Digest of the 2004 Joint 29th International Conference on 2004 and 12th International Conference on Terahertz Electronics, 2004.*, pages 817–818. IEEE, 2004. 15

[21] M Hassan Arbab, Dale P Winebrenner, Trevor C Dickey, Antao Chen, Matthew B Klein, and Pierre D Mourad. Terahertz spectroscopy for the assessment of burn injuries in vivo. *Journal of biomedical optics*, 18(7):077004, 2013. 15, 34

[22] Daniela I Ramos-Soto, Abhishek K Singh, Edgar Saucedo-Casas, Enrique Castro-Camus, and Mariana Alfaro-Gomez. Visualization of moisturizer effects in stratum corneum in vitro using thz spectroscopic imaging. *Applied Optics*, 58(24):6581–6585, 2019. 15

[23] Philip C Ashworth, Emma Pickwell-MacPherson, Elena Provenzano, Sarah E Pinder, Anand D Purushotham, Michael Pepper, and Vincent P Wallace. Terahertz pulsed spectroscopy of freshly excised human breast cancer. *Optics express*, 17(15):12444–12454, 2009. 15

[24] Tyler C Bowman, Magda El-Shenawee, and Lucas K Campbell. Terahertz imaging of excised breast tumor tissue on paraffin sections. *IEEE Transactions on Antennas and Propagation*, 63(5):2088–2097, 2015.

[25] Quentin Cassar, Amel Al-Ibadi, Laven Mavarani, Philipp Hillger, Janusz Grzyb, Gaëtan MacGrogan, Thomas Zimmer, Ullrich R Pfeiffer, Jean-Paul Guillet, and Patrick Mounaix. Pilot study of freshly excised breast tissue response in the 300–600 ghz range. *Biomedical optics express*, 9(7):2930–2942, 2018. 15

[26] GG Hernandez-Cardoso, SC Rojas-Landeros, M Alfaro-Gomez, AI Hernandez-Serrano, I Salas-Gutierrez, E Lemus-Bedolla, AR Castillo-Guzman, HL Lopez-Lemus, and E Castro-Camus. Terahertz imaging for early screening of diabetic foot syndrome: A proof of concept. *Scientific Reports*, 7:42124, 2017. 15

[27] Arthur Robert Von Hippel. Dielectrics and waves. 1954. 18

[28] Eias M Bernstein. *Bioelectrochemistry research developments*. Nova Publishers, 2008. xv, 19, 20

[29] Andrew K Jonscher. Dielectric relaxation in solids. *Journal of Physics D: Applied Physics*, 32(14):R57, 1999. 20

[30] Kenneth S Cole and Robert H Cole. Dispersion and absorption in dielectrics i. alternating current characteristics. *The Journal of chemical physics*, 9(4):341–351, 1941. 21

[31] Kenneth S Cole and Robert H Cole. Dispersion and absorption in dielectrics ii. direct current characteristics. *The Journal of Chemical Physics*, 10(2):98–105, 1942. 21

[32] D-W_ Davidson. Dielectric relaxation in liquids: I. the representation of relaxation behavior. *Canadian Journal of Chemistry*, 39(3):571–594, 1961. 21

[33] TP Iglesias, G Vilao, and João Carlos R Reis. An approach to the interpretation of cole–davidson and cole–cole dielectric functions. *Journal of Applied Physics*, 122(7):074102, 2017. 21

[34] S Havriliak and S Negami. A complex plane representation of dielectric and mechanical relaxation processes in some polymers. *Polymer*, 8:161–210, 1967. 22

[35] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978. 23

[36] Constantino Grosse. A program for the fitting of debye, cole–cole, cole–davidson, and havriliak–negami dispersions to dielectric data. *Journal of colloid and interface science*, 419:102–106, 2014. 23

[37] Michael H Ross and Wojciech Pawlina. *Histology*. Lippincott Williams & Wilkins, 2006. 24

[38] Herman P Schwan. Electrical properties of tissue and cell suspensions. In *Advances in biological and medical physics*, volume 5, pages 147–209. Elsevier, 1957. 24

[39] S Gabriel. The dielectric properties of biological tissue parametric models for the dielectric spectrum of tissues. *Phys. Med. Biol.*, 41:2251–2269, 1996. 24

[40] David Eisenberg and Walter Kauzmann. *The structure and properties of water*. Oxford University Press on Demand, 2005. 25

[41] Josef Barthel, K Bachhuber, Richard Buchner, and H Hetzenauer. Dielectric spectra of some common solvents in the microwave region. water and lower alcohols. *Chemical physics letters*, 165(4):369–373, 1990. xv, 26, 27

[42] Takaaki Sato and Richard Buchner. Dielectric relaxation processes in ethanol/water mixtures. *The Journal of Physical Chemistry A*, 108(23):5007–5015, 2004. 26, 27

[43] Cecilie Ronne, Lars Thrane, Per-Olof Åstrand, Anders Wallqvist, Kurt V Mikkelsen, and So/ren R Keiding. Investigation of the temperature dependence of dielectric relaxation in liquid water by thz reflection spectroscopy and molecular dynamics simulation. *The Journal of chemical physics*, 107(14):5319–5331, 1997. 27

[44] Hiroyuki Yada, Masaya Nagai, and Koichiro Tanaka. Origin of the fast relaxation component of water and heavy water revealed by terahertz time-domain attenuated total reflection spectroscopy. *Chemical Physics Letters*, 464(4-6):166–170, 2008. 26, 27

[45] Uffe Møller, David G Cooke, Koichiro Tanaka, and Peter Uhd Jepsen. Terahertz reflection spectroscopy of debye relaxation in polar liquids. *JOSA B*, 26(9):A113–A125, 2009. 26, 27

[46] MN Afsar and JB Hasted. Measurements of the optical constants of liquid h 2 o and d 2 o between 6 and 450 cm- 1. *JOSA*, 67(7):902–904, 1977. 26

[47] Udo Kaatze. Complex permittivity of water as a function of frequency and temperature. *Journal of Chemical and Engineering Data*, 34(4):371–374, 1989. xv, 27

[48] Abraham J Domb, Joseph Kost, and David Wiseman. *Handbook of biodegradable polymers*. CRC press, 1998. 27

[49] Hai-Bo Liu and X-C Zhang. Dehydration kinetics of d-glucose monohydrate studied using thz time-domain spectroscopy. *Chemical physics letters*, 429(1-3):229–233, 2006. 27, 35

[50] Bernd M Fischer, Hanspeter Helm, and Peter Uhd Jepsen. Chemical recognition with broadband thz spectroscopy. *Proceedings of the IEEE*, 95(8):1592–1604, 2007. 27

[51] Guilherme F Medeiros, Aline Mendes, Ricardo AB Castro, Elaine C Baú, Helena B Nader, and Carl P Dietrich. Distribution of sulfated glycosaminoglycans in the animal kingdom: widespread occurrence of heparin-like compounds in invertebrates. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1475(3):287–294, 2000. 28

[52] Merton Bernfield, Martin Götte, Pyong Woo Park, Ofer Reizes, Marilyn L Fitzgerald, John Lincecum, and Masahiro Zako. Functions of cell surface heparan sulfate proteoglycans. *Annual review of biochemistry*, 68(1):729–777, 1999. 28

[53] Renato V Iozzo. Matrix proteoglycans: from molecular design to cellular function, 1998. 28

[54] Louay K Hallak, Dorothe Spillmann, Peter L Collins, and Mark E Peeples. Glycosaminoglycan sulfation requirements for respiratory syncytial virus infection. *Journal of virology*, 74(22):10508–10513, 2000. 28

[55] Gareth M Holder, Andrew Bowfield, Mark Surman, Michael Suepfle, David Moss, Carole Tucker, Timothy R Rudd, David G Fernig, Edwin A Yates, and Peter Weightman. Fundamental differences in model cell-surface polysaccharides revealed by complementary optical and spectroscopic techniques. *Soft Matter*, 8(24):6521–6527, 2012. 28

[56] M Hishida and K Tanaka. Long-range hydration effect of lipid membrane studied by terahertz time-domain spectroscopy. *Physical review letters*, 106(15):158102, 2011. 29

[57] Caroline B Reid, Emma Pickwell-MacPherson, Jan G Laufer, Adam P Gibson, Jeremy C Hebden, and Vincent P Wallace. Accuracy and resolution of thz reflection spectroscopy for medical imaging. *Physics in Medicine & Biology*, 55(16):4825, 2010. 30

[58] Tap Ha Duong and Krystyna Zakrzewska. Calculation and analysis of low frequency normal modes for dna. *Journal of computational chemistry*, 18(6):796–811, 1997. 30

[59] VK Saxena, BH Dorfman, and LL Van Zandt. Identifying and interpreting spectral features of dissolved poly (da)-poly (dt) dna polymer in the high-microwave range. *Physical Review A*, 43(8):4510, 1991. 30

[60] Y Feng and EW Prohofsky. Vibrational fluctuations of hydrogen bonds in a dna double helix with nonuniform base pairs. *Biophysical journal*, 57(3):547–553, 1990. 30

[61] M Brucherseifer, M Nagel, P Haring Bolivar, H Kurz, A Bosserhoff, and R Büttner. Label-free probing of the binding state of dna by time-domain terahertz sensing. *Applied Physics Letters*, 77(24):4049–4051, 2000. 30

[62] AG Markelz, A Roitberg, and Edwin J Heilweil. Pulsed terahertz spectroscopy of dna, bovine serum albumin and collagen between 0.1 and 2.0 thz. *Chemical Physics Letters*, 320(1-2):42–48, 2000. 31

[63] David M Leitner, Martina Havenith, and Martin Gruebele. Biomolecule large-amplitude motion and solvation dynamics: modelling and probes from thz to x-rays. *International Reviews in Physical Chemistry*, 25(4):553–582, 2006. 31

[64] Seung Joong Kim, Benjamin Born, Martina Havenith, and Martin Gruebele. Real-time detection of protein–water dynamics upon protein folding by terahertz absorption spectroscopy. *Angewandte Chemie International Edition*, 47(34):6486–6489, 2008. 31

[65] Jing Xu, Kevin W Plaxco, and S James Allen. Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy. *protein Science*, 15(5):1175–1181, 2006. 31

[66] HQ Woodard and DR White. The composition of body tissues. *The British journal of radiology*, 59(708):1209–1218, 1986. xv, 33

[67] Etsuko Ohmae, Nobuko Yoshizawa, Kenji Yoshimoto, Maho Hayashi, Hiroko Wada, Tetsuya Mimura, Yuko Asano, Hiroyuki Ogura, Yutaka Yamashita, Harumi Sakahara, et al. Comparison of lipid and water contents by time-domain diffuse optical spectroscopy and dual-energy computed tomography in breast cancer patients. *Applied Sciences*, 9(7):1482, 2019. xv, 33

[68] Bryan E Cole, Ruth M Woodward, David A Crawley, Vincent P Wallace, Donald D Arnone, and Michael Pepper. Terahertz imaging and spectroscopy of human skin in vivo. In *Commercial and Biomedical Applications of Ultrashort Pulse Lasers; Laser Plasma Generation and Diagnostics*, volume 4276, pages 1–10. International Society for Optics and Photonics, 2001. 34

[69] Ibtissam Echchgadda, Jessica A Grundt, Melissa Tarango, Bennett L Ibey, Thomas D Tongue, Min Liang, Hao Xin, and Gerald J Wilmink. Using a portable terahertz spectrometer to measure the optical properties of in vivo human skin. *Journal of Biomedical Optics*, 18(12):120503, 2013. 34

[70] David B Bennett, Zachary D Taylor, Priyamvada Tewari, Sijun Sung, Ashkan Maccabi, Rahul S Singh, Martin O Culjat, Warren S Grundfest, Jean-Pierre Hubschman, and Elliott R Brown. Assessment of corneal hydration sensing in the terahertz band: in vivo results at 100 ghz. *Journal of biomedical optics*, 17(9):097008, 2012. 35

[71] David B Bennett, Zachary D Taylor, Priyamvada Tewari, Rahul S Singh, Martin O Culjat, Warren S Grundfest, Daniel J Sassoon, R Duncan Johnson, Jean-Pierre Hubschman, and Elliott Brown. Terahertz sensing in corneal tissues. *Journal of biomedical optics*, 16(5):057003, 2011. 35

[72] Elizabeth Berry, Anthony J Fitzgerald, Nickolay N Zinov'ev, Gillian Claire Walker, Shervanthi Homer-Vanniasinkam, Caroline D Sudworth, Robert E Miles, J Martyn Chamberlain, and Michael A Smith. Optical properties of tissue measured using terahertz-pulsed imaging. In *Medical Imaging 2003: Physics of Medical Imaging*, volume 5030, pages 459–470. International Society for Optics and Photonics, 2003. 35

[73] Craig M Ciesla, Donald D Arnone, Alessandra Corchia, David A Crawley, Christopher Longbottom, Edmund H Linfield, and Michael Pepper. Biomedical applications of terahertz pulse imaging. In *Commercial and Biomedical Applications of Ultrafast Lasers II*, volume 3934, pages 73–81. International Society for Optics and Photonics, 2000. 36

[74] Naoya Hoshi, Yoshio Nikawa, Keiji Kawai, and Shigeyuki Ebisu. Application of microwaves and millimeter waves for the characterization of teeth for dental diagnosis and treatment. *IEEE transactions on microwave theory and techniques*, 46(6):834–838, 1998. 36

[75] David Crawley, Christopher Longbottom, Vincent P Wallace, Bryan Cole, Don Arnone, and Michael Pepper. Three-dimensional terahertz pulse imaging of dental tissue. *Journal of Biomedical Optics*, 8(2):303–307, 2003. 36

[76] Bryan Edward Cole, Vincent Patrick Wallace, Michael John Withers, John Baker, and Brian Robertson. Scanning terahertz probe, April 14 2015. US Patent 9,006,660. 36

[77] Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *cell*, 100(1):57–70, 2000. 36

[78] GI McIntyre. Cell hydration as the primary factor in carcinogenesis: a unifying concept. *Medical hypotheses*, 66(3):518–526, 2006. 36

[79] Kiyoung Jeong, Yong-Min Huh, Sang-Hoon Kim, Yeonji Park, Joo-Hiuk Son, Seung Jae Oh, and Jin-Suck Suh. Characterization of blood using terahertz waves. *Journal of biomedical optics*, 18(10):107008, 2013. 37

[80] Caroline B Reid, George Reese, Adam P Gibson, and Vincent P Wallace. Terahertz time-domain spectroscopy of human blood. *IEEE journal of biomedical and health informatics*, 17(4):774–778, 2013. 37

[81] Seung Jae Oh, Sang-Hoon Kim, Young Bin Ji, Kiyoung Jeong, Yeonji Park, Jaemoon Yang, Dong Woo Park, Sam Kyu Noh, Seok-Gu Kang, Yong-Min Huh, et al. Study of freshly excised brain tissues using terahertz imaging. *Biomedical optics express*, 5(8):2837–2842, 2014. 37

[82] Kun Meng, Tu-nan Chen, Tao Chen, Li-guo Zhu, Qiao Liu, Zhao Li, Fei Li, Sen-cheng Zhong, Ze-ren Li, Hua Feng, et al. Terahertz pulsed spectroscopy of paraffin-embedded brain glioma. *Journal of biomedical optics*, 19(7):077001, 2014. 37

[83] Sayuri Yamaguchi, Yasuko Fukushi, Oichi Kubota, Takeaki Itsuji, Toshihiko Ouchi, and Seiji Yamamoto. Brain tumor imaging of rat fresh tissue using terahertz spectroscopy. *Scientific reports*, 6:30124, 2016. 37

[84] Caroline B Reid, Anthony Fitzgerald, George Reese, Robert Goldin, Paris Tekkis, PS O'Kelly, Emma Pickwell-MacPherson, Adam P Gibson, and Vincent P Wallace. Terahertz pulsed imaging of freshly excised human colonic tissues. *Physics in Medicine & Biology*, 56(14):4333, 2011. 37

[85] Leila H Eadie, Caroline B Reid, Anthony J Fitzgerald, and Vincent P Wallace. Optimizing multi-dimensional terahertz imaging analysis for colon cancer diagnosis. *Expert Systems with Applications*, 40(6):2043–2050, 2013. 37

[86] Dibo Hou, Xian Li, Jinhui Cai, Yehao Ma, Xusheng Kang, Pingjie Huang, and Guangxin Zhang. Terahertz spectroscopic investigation of human gastric normal and tumor tissues. *Physics in Medicine & Biology*, 59(18):5423, 2014. 37

[87] Young Bin Ji, Chan Hyuk Park, Hyunki Kim, Sang-Hoon Kim, Gyu Min Lee, Sam Kyu Noh, Tae-In Jeon, Joo-Hiuk Son, Yong-Min Huh, Seungjoo Haam, et al. Feasibility of terahertz reflectometry for discrimination of human early gastric cancers. *Biomedical optics express*, 6(4):1398–1406, 2015. 37

[88] Tokujiro Enatsu, Hideaki Kitahara, Keisuke Takano, Takeshi Nagashima, Masahiko Tani, Masanori Hangyo, Yasuhiro Miura, and Takashi Sawai. Terahertz spectroscopic imaging of paraffin-embedded liver cancer samples. In *2007 Joint 32nd International Conference on Infrared and Millimeter Waves and the 15th International Conference on Terahertz Electronics*, pages 557–558. IEEE, 2007. 38

[89] Stanley Sy, Shengyang Huang, Yi-Xiang J Wang, Jun Yu, Anil T Ahuja, Yuan-ting Zhang, and Emma Pickwell-MacPherson. Terahertz spectroscopy of liver cirrhosis: investigating the origin of contrast. *Physics in Medicine & Biology*, 55(24):7587, 2010. 38

[90] Maarten R Grootendorst, Anthony J Fitzgerald, Susan G Brouwer De Koning, Aida Santaolalla, Alessia Portieri, Mieke Van Hemelrijck, Matthew R Young, Julie Owen, Massi Cariati, Michael Pepper, et al. Use of a handheld terahertz pulsed imaging device to differentiate benign and malignant breast tissue. *Biomedical Optics Express*, 8(6):2932–2945, 2017. 38, 41

[91] Marc-Aurèle Brun, F Formanek, A Yasuda, M Sekine, N Ando, and Y Eishii. Terahertz imaging applied to cancer diagnosis. *Physics in Medicine & Biology*, 55(16):4615, 2010. 38

[92] Saman Warnakulasuriya. Global epidemiology of oral and oropharyngeal cancer. *Oral oncology*, 45(4-5):309–316, 2009. 38

[93] Yookyeong Carolyn Sim, Jae Yeon Park, Kang-Min Ahn, Chansik Park, and Joo-Hiuk Son. Terahertz imaging of excised oral cancer at frozen temperature. *Biomedical Optics Express*, 4(8):1413–1421, 2013. 38

[94] Ruth M Woodward, Bryan E Cole, Vincent P Wallace, Richard J Pye, Donald D Arnone, Edmund H Linfield, and Michael Pepper. Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue. *Physics in Medicine & Biology*, 47(21):3853, 2002. 38

[95] Calvin Yu, Shuting Fan, Yiwen Sun, and Emma Pickwell-MacPherson. The potential of terahertz imaging for cancer diagnosis: A review of investigations to date. *Quantitative imaging in medicine and surgery*, 2(1):33, 2012. 38

[96] KFA Ross and RE Gordon. Water in malignant tissue, measured by cell refractometry and nuclear magnetic resonance. *Journal of microscopy*, 128(1):7–21, 1982. 38

[97] Vincent P Wallace, Anthony J Fitzgerald, Emma Pickwell, Richard J Pye, Philip F Taday, Niamh Flanagan, and Thomas Ha. Terahertz pulsed spectroscopy of human basal cell carcinoma. *Applied spectroscopy*, 60(10):1127–1133, 2006. 39

[98] Giovanni Bistoni and Jian Farhadi. Anatomy and physiology of the breast. *Plastic and reconstructive surgery: Approaches and techniques*, pages 477–485, 2015. 39

[99] SR Wellings, Hanne M Jensen, and RG Marcum. An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. *Journal of the National Cancer Institute*, 55(2):231–273, 1975. 39

[100] Muhammad Al-Hajj, Max S Wicha, Adalberto Benito-Hernandez, Sean J Morrison, and Michael F Clarke. Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences*, 100(7):3983–3988, 2003. 39

[101] Bernard W Stewart, Paul Kleihues, et al. World cancer report. 2003.

[102] Kirby I Bland, Edward M Copeland, V Suzanne Klimberg, and William J Gradishar. *The breast: comprehensive management of benign and malignant diseases*. Elsevier Inc, 2017. 39

[103] Umberto Veronesi, Natale Cascinelli, Luigi Mariani, Marco Greco, Roberto Saccozzi, Alberto Luini, Marisel Aguilar, and Ettore Marubini. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *New England Journal of Medicine*, 347(16):1227–1232, 2002. 40

[104] J Michael Dixon. *Breast Surgery: Companion to Specialist Surgical Practice*. Elsevier Health Sciences, 2013. 40

[105] Hani A Alturkistani, Faris M Tashkandi, and Zuhair M Mohammedsaleh. Histological stains: a literature review and case study. *Global journal of health science*, 8(3):72, 2016. 41

[106] Jeffrey Landercasper, Eric Whitacre, Amy C Degnim, and Mohammed Al-Hamadani. Reasons for re-excision after lumpectomy for breast cancer: insight from the american society of breast surgeons mastery sm database. *Annals of surgical oncology*, 21(10):3185–3191, 2014. 41

[107] R Jeevan, DA Cromwell, M Trivella, G Lawrence, O Kearins, J Pereira, C Sheppard, CM Caddy, and JHP Van Der Meulen. Reoperation rates after breast conserving surgery for breast cancer among women in england: retrospective study of hospital episode statistics. *Bmj*, 345:e4505, 2012. 41

[108] Anthony J Fitzgerald, Vincent P Wallace, Mercedes Jimenez-Linan, Lynda Bobrow, Richard J Pye, Anand D Purushotham, and Donald D Arnone. Terahertz pulsed imaging of human breast tumors. *Radiology*, 239(2):533–540, 2006. 41

Chapter II

Terahertz Spectroscopy

Every portrait that is painted with feeling is a portrait of the artist, not of the sitter.

Oscar Wilde

Sommaire

II.1 Introduction . . . 54
II.2 Generation, Detection and Sensing . . . 54
II.2.1 Photoconductive Antenna . . . 54
II.2.2 Optical Rectification and Pockels Effect . . . 56
II.2.3 Terahertz Setup . . . 57
II.2.4 Noise Sources in Terahertz-TDS Systems . . . 60
II.2.5 Teraview (Ltd, Cambridge UK) Systems . . . 61
II.3 Spectroscopy: Principles . . . 62
II.3.1 Transfer Function . . . 62
II.3.2 Iterative Determination of $S_{sam}(\omega)$ . . . 62
II.3.3 Parametrization of $S_{sam}(\omega)$ . . . 66
II.3.4 Optical Property Extraction . . . 69
II.3.5 Optical Property Fitting . . . 74
II.4 Water . . . 75
II.4.1 Extracted Intrinsic Properties . . . 76
II.4.2 Relaxation Mechanisms . . . 76
II.5 Phantoms . . . 80
II.5.1 Extracted Intrinsic Properties . . . 81
II.5.2 Relaxation Mechanisms . . . 83
II.6 Excised Breast Tissues . . . 89
II.6.1 Tissue Registration & Preparation . . . 89
II.6.2 Extracted Intrinsic Properties . . . 90
II.6.3 Relaxation Mechanisms . . . 95
II.7 Cancerous Cells . . . 99
II.7.1 Concentration-Dependent Properties . . . 100
II.7.2 Influence on Relaxation Mechanisms . . . 102
II.8 Conclusion . . . 104
II.9 Bibliography . . . 105

II.1 Introduction

The second chapter *Terahertz Spectroscopy* aims to describe the experimental, the mathematical and the numerical background for research towards breast cancer terahertz spectroscopy.

The first section is a brief description on how to generate and to detect broadband terahertz pulses and how to implement related devices in experimental setups.

Sensing with those devices allows to measure the portion of the incident field that is either transmitted or reflected upon interaction with a sample. Both the transmitted and reflected signals reflect the intrinsic properties of the sensed material. Some of the mathematical and numerical tools developed during this thesis in order to qualify and to quantify these properties are described, alongside with illustrations of their effectiveness on real samples.

Thereafter, these procedures are applied to the field of biology, including successively biological water, phantoms mimicking breast tissue, a pre-clinical study on freshly excised breast tissues and finally breast cancer cells in solution.

II.2 Generation, Detection and Sensing

The terahertz regime is an attracting region of the electromagnetic spectrum since it lies where thermal **[1]** and ultrahertzian **[2]** waves meet. In addition to that, several biological processes occur at the sub-picosecond time scale such as some vibrational phonon modes of DNA macromolecules, dipole relaxations and molecule stretching. Although being of interest since the *1920s* with the "paste radiator" of Glagolewa-Arkadiewa **[3]**, the terahertz portion remained for a while the least tapped fraction of the electromagnetic spectrum. Nowadays, almost a century since prior researches, one can either use optoelectronics or optics for terahertz broadband generation. While the first employs a photoconductive antenna, the second is based on the optical rectification effect.

II.2.1 Photoconductive Antenna

Photoconductive antenna, sometimes called semiconductor photoswitch, have been widely utilized for the generation of broadband frequency terahertz pulses. Such devices are the ground of many terahertz spectroscopy and imaging setups for which, auspicious applications in academia and industries have been reported **[4]**. A photoconductive antenna consists of a photoconductive layer onto which is patterned a discontinuous current biased metallic antenna **[5]**. The idea behind emitting terahertz radiations from such a structure, is to optically pump with an infrared laser whose energy is greater than the energy band gap of the semiconductor. The incident pump laser propagates into the photoconductive layer and as it is absorbed, starts to generate electrons and holes. The generation of the carriers is carried out at a rate proportional to the time-width of the incident radiation. Those photocarriers are accelerated in the electric force field of the biased antenna. The acceleration of the photogenerated charges produces a transient current that drives the antenna and is eventually emitted as terahertz pulse radiation. Once the pump radiation is fully absorbed, mobile photocarriers recombine progressively, depending on the carrier lifetime. The longer the lifetime of photocarriers is, the longer they will drive into the antenna to contribute to the overall transient photocurrent. Consequently semiconductor with long carrier lifetime will generate limited frequency bandwidth pulses. Terahertz generation has then to be realized from sub-picosecond carrier lifetime semiconductor. The emitted terahertz radiation depends on the transient current density $\mathrm{J}(t)$ whose form can be written as **[6]**:

$$\mathrm{J}(t) = (\mathrm{E} * \left[q \times n\mu\right])(t), \tag{II.1}$$

with E being the incident field on the active layer of the semiconductor in V/m, the elementary charge q in A.s, the charge density n in m^{-3} and μ the charge mobility in $m^2/\mathrm{V}s$. Note that $*$ is the convolution operator. The equation II.2 describes the rapid motion of the carriers caused by an

incident electric field. Consequently, such a motion leads to a rapid time-varying current density. The transient current is the origin of the sub-millimeter radiation E_{THz}:

$$E_{THz}(t) \propto \frac{\partial J(t)}{\partial t}. \tag{II.2}$$

Details of how much radiations are generated depends on several parameters of the generation setup. However, further considerations are beyond the scope of this work. The working principle of the photoconductive antenna is illustrated in Figure II.1.

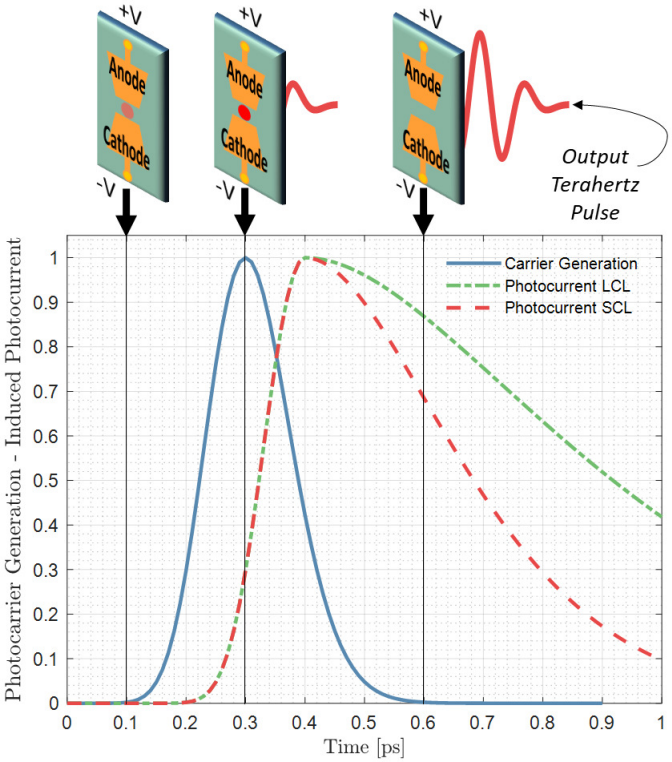


Figure II.1: Illustrative example of pulsed terahertz generation by means of a photoconductive antenna. LCL: long carrier lifetime. SCL: short carrier lifetime.

As previously said, the current driving the antenna, and thus the wealth of the pulse frequency bandwidth depend on both the lifetime of photogenerated carriers and their mobility. Globally, the carrier density and their associated velocity is controlled by the introduction of structural imperfections in perfect crystals. Research for ultrafast photoconductive materials started in the *mid-1970s*, when Auston reported the use of high-resistivity silicon, know at that time from the far infrared field **[7]**. In Table II.1 are reported some materials matching with the short carrier lifetime and the carrier high mobility requirement **[8]**. Among them, the most frequently used photoconductive materials are the RD-SOS **[9]** and the LT-GaAs **[10]**. The RD-SOS is a silicon substrate deposited on sapphire (SOS: silicon on sapphire) which is damaged by radiation exposure (RD: radiation-damaged). The radiation aims to implant argon, silicon or oxygen ions acting as defect recombination centers that ultimately shorten the carrier lifetime. The carrier lifetime of the RD-SOS photoconductive material lessens in an exponential fashion as a function of ion implantation.

An excellent trade-off between carrier lifetime and mobility has been reported since the advent of LT-GaAs material (low-temperature grown GaAs crystal). The characteristic of the LT-GaAs are directly dictated by its growth condition during the molecular beam epitaxy and the following annealing. The arsenic atoms in excess are incorporated during the growth of the crystal leading to the formation of defects in the overall structure. These imperfections compose nonradiative recombination centers that therefore, diminish the mean carrier lifetime in the active layer. The

lower the temperature under which the GaAs crystal grows is, the higher the number of recombination sites is. Although induced defects are numerous at low temperature growing process, the carrier mobility remains relatively high, contributing to preserve the short lifetime of photocarriers.

Table II.1: Properties of some ultrafast photoconductive substrates for pulsed terahertz emission.

Photoconductive Substrate	Energy Band Gap (eV, room-T°)	Carrier Lifetime (ps)	Mobility ($cm^2.(V.s)^{-1}$)
II Ge	0.66	0.6	100
RD-SOS	1.10	0.6	30
Amorphous SI	1.10	0.8-20	1
II InP	1.34	2-4	200
SI-InP	1.34	50-100	1,000
LT-GaAs	1.43	0.3	150-200
Cr:doped SI-GaAs	1.43	50-100	1000
LT-$In_{0.52}Al_{0.48}As$	1.45	0.4	5
MOCVD CdTe	1.49	0.5	180

The pulsed terahertz signal detection is achieved in an analogous way than its emission. Indeed, since the photocarrier lifetime is much shorter than the terahertz pulse (for short carrier lifetime photoconductive materials like LT-GaAs), the photoconductive antenna behaves like a sampler that monitors the incident terahertz waveform. The photoconductive antenna detector is pumped by delayed-fs-laser pulses that generate the photocarriers among the active layer. The terahertz field concomitantly stimulates the photocarriers by acting as a dynamic bias field in the detector. The recorded photocurrent is therefore proportional to the external terahertz stimuli. One can retrieve the terahertz waveform by measuring the photocurrent versus the photoconductive sampling between subsequent incident terahertz fields and the sampling optical pulses.

As already pointed out in *Chapter I, I.3.1 Physical Quantities in Electromagnetism*, a Fourier analysis of the time-domain terahertz signal shape makes possible to retrieve both the phase and the amplitude spectra.

One may have pointed out that the photoconductive materials listed in Table II.1 exhibit particularly high refractive indices in the terahertz band. Thus, due to the refractive index difference from the photoconductive material to that of free space[1], the broad majority of the generated signal and of the aimed detected one will be almost totally reflected back at the dielectric inferface. To overcome such an issue, the terahertz generation and detection sites are positioned at the focal point of a terahertz-transparent, terahertz non-dispersive and non-birefringent lens. In order to fulfill all these requirements, a high-resistivity semi-spherical silicon lens is commonly integrated.

Although such a terahertz generation-detection process is the one that was extensively used to accomplished this thesis work, an alternative way to proceed has to be acknowledged.

II.2.2 Optical Rectification and Pockels Effect

Another way to generate terahertz pulses is to focus on a second order non-linear $\chi^{(2)}$ material, visible or near-infrared light. This will generate the possible difference-frequency train. The corresponding non-linear polarization $P(\omega)$ generated by the incident electromagnetic pulse $E(\omega)$ can be expressed as **[11]**:

$$P(\omega) = \epsilon_0 \chi^{(2)}(\omega_1 - \omega_2) E(\omega_1) E(\omega_2)^*, \tag{II.3}$$

where $\omega_1 - \omega_2 = \omega$. As the optical rectification is nothing else than all the possible-frequency generations, it usually offers broader spectrum but provides lower power in comparison to pho-

[1] Often the propagation medium.

toconductive antenna. When using an oscillating laser, the emitted terahertz field may then be expressed as:

$$E_{THz}(t) \propto \frac{\partial^2 P(t)}{\partial t^2}. \tag{II.4}$$

We distinguish three $\chi^{(2)}$ susceptibility material categories that are extensively used to achieve terahertz generation: first, some III-V and II-IV semiconductors **[12]**; Second, some ferroelectrics like $LiNbO_3$ and $LiTaO_3$ **[13]**; Third, non linear organic crystals **[14; 15]**.

The coherent detection is performed by the Pockels effect which is a non-linear process. Alternatively said, it corresponds to $\omega_2 = 0$ in equation II.3. The terahertz signal to be detected changes the birefringence of the non-linear crystal which consequently turns the initially linear polarization of the optical probe beam into ellipsoid polarization **[16]**. The optical probe beam passes then through a quarter plate which converts linearly polarized light into circularly polarized one and converts ellipsoid polarization in "dilated hoop" polarization. Then, a Wallstom prism separates the two polarization components of the probing beam. The field difference between the two components is measured by a balanced photo-detector and is proportional to the terahertz field strength. The detection principle using non-linear crystal is illustrated in Figure II.2.

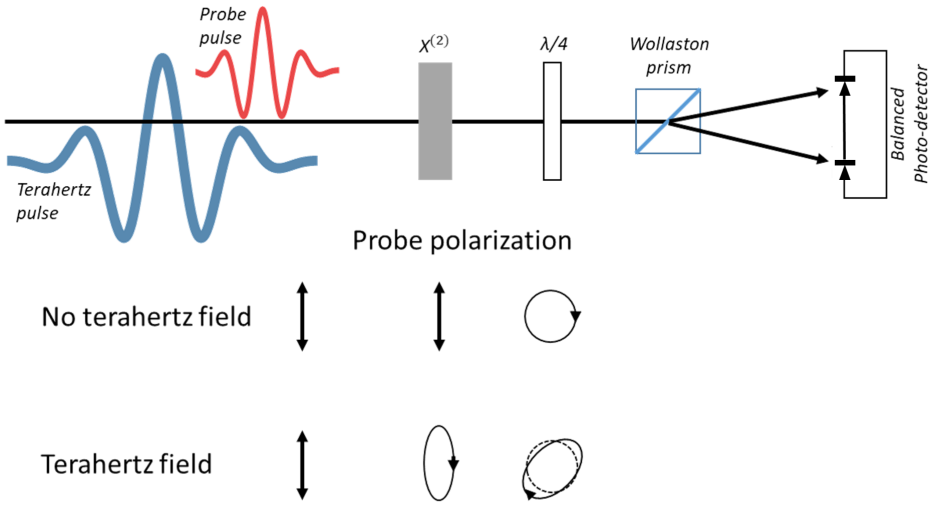


Figure II.2: Details of a free-space terahertz detection using a $\chi_{(2)}$ susceptibility crystal.

There is other ways to perform both generation and detection of pulsed terahertz radiations. One can for instance mention emission from quantum cascade structures **[17]**, from coherent longitudinal optical phonon **[18]**, or even from cold plasma oscillations **[19]**. In addition, other techniques generate continuous terahertz waves. However, these methods, although they deserve to be acknowledged, are beyond the scope of this work.

II.2.3 Terahertz Setup

Here, investigations were performed with terahertz spectroscopy systems that employ photoconductive antennas for pulse generation and detection. As terahertz pulse frequency band spans over few hundreds of gigahertz to few terahertz, it follows that their respective duration in the time-domain is particularly brief ($\leqslant 10^{-12}$s). Naturally, if one wants to measure such signals, a broadband monitoring system is required. In other words, it requires the use of a monitoring system ables to sample the signal many times over less than a picosecond to satisfy the Shannon sampling-theorem which states **[20]**:

$$f_s = 2 f_{max}, \tag{II.5}$$

where f_s is the sampling frequency and f_{max} is the maximum frequency carried by the terahertz pulse. However, such broadband systems are not available. Moreover, photocurrent induced by terahertz fields within the detecting photoconductive antenna are of picoampers to few nanoampers. Thus, it remains unrealistic to sample such signal with classical sampling method.

When the impulse response of the system exhibits a periodicity, one can employ electro-optic sampling or photoconductive sampling where the terahertz signal is sampled over different and subsequent periods (Figure II.3).

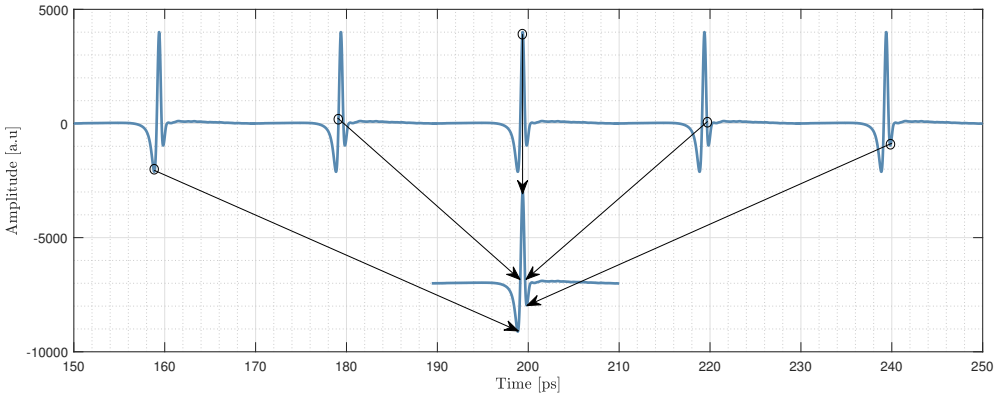


Figure II.3: Reconstruction of a terahertz pulse by means of electro-optic sampling method.

Consequently, the use of a specific component to delay precisely the sampled period is compulsory. Such device is known as a delay line. Few thousands of measurements can then be recorded over a long time scale where their respective temporal position Δt are correlated to the delay line displacement Δx through:

$$\Delta t = 2\frac{\Delta x}{c}, \tag{II.6}$$

wherein c is the light velocity in free space.

Obviously, such an approach implies the synchronicity of the signals to be measured and the recording spectrum analyzer. To this aim, the same fs-pulse triggers the terahertz photoconductive antenna (PCA) emitter and detector. The fs-pulse is divided by a beam splitter (BS) into a probe pulse (for terahertz pulse emission) and a pump pulse (for terahertz field detection) respectively. The probe pulse is focused onto the continuous external biased photoconductive antenna which serves as terahertz emitter. The pump pulse goes within the delay line before reaching the photoconductive antenna detector whose induced photocurrent is recorded by a spectrum analyzer. The acquisition time then depends on the time required by the delay line to move along a distance corresponding to the desired time window. Additionally, an equivalent period is usually measured several times to maximize the signal-to-noise ratio, increasing slightly the time of the acquisition. A schematic diagram for such systems is provided in Figure II.4.

Let's underline that heretofore, the reported setup is working in transmission geometry. In other words, it aims to record the terahertz field amplitude and phase transmitted through a given medium. Such a configuration is suitable as it provides a complete description of the in-depth intrinsic dielectric properties of objects. However, for some specific object -scatterer and absorptive- this kind of setup is not possible as the terahertz field would not reach the detector. Hence, a reflection geometry is preferred where the recorded signal depicts the reflected energy at the interface between the measurement atmosphere and the object under inspection. Although the setup is almost equivalent to the one illustrated in Figure II.4, the reflection configuration is schematically represented hereafter, in Figure II.5.

In both configuration, namely transmission and reflection, the terahertz signal $S_{THz}(t, \Delta x)$ delivered by the detector depends on the temporal evolution of the detector response $R_{PCA}(t)$ and

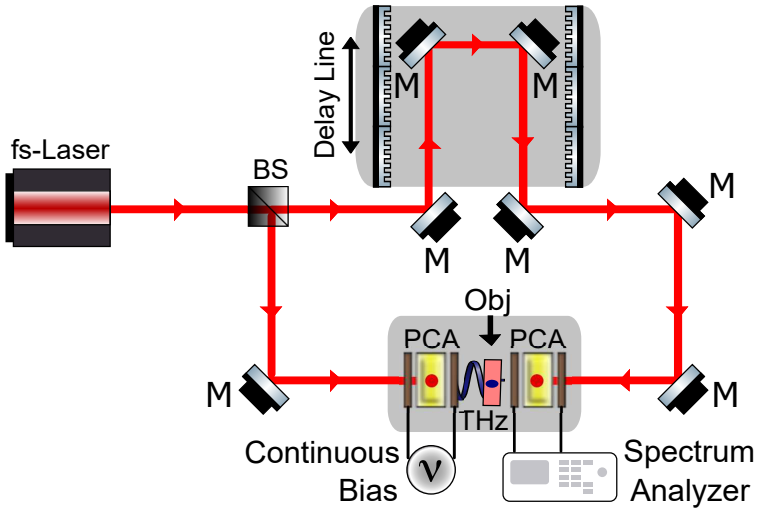


Figure II.4: Schematic details of a time-domain system for terahertz spectroscopy in transmission configuration. BS stands for beam splitter, and M depicts mirrors.

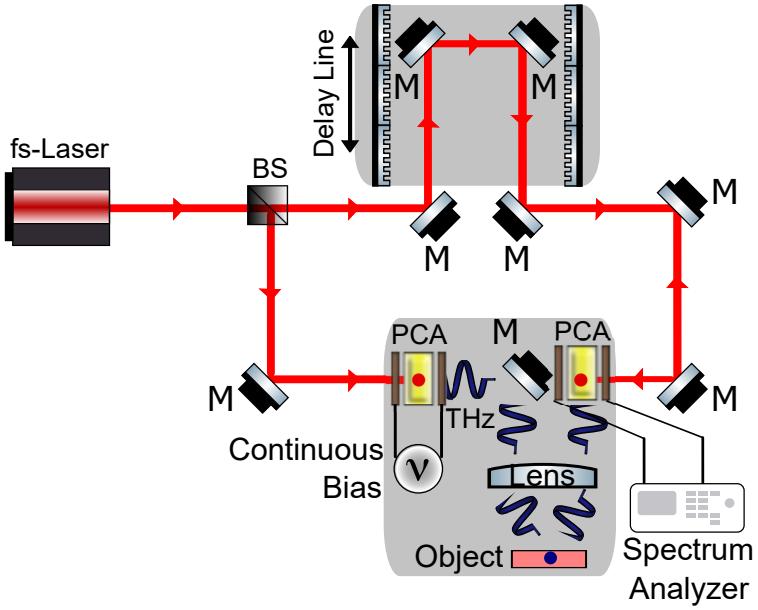


Figure II.5: Schematic details of a time-domain system for terahertz spectroscopy in reflection configuration. BS stands for beam splitter, and M depicts mirrors.

the incident terahertz field $E_{THz}(t, \Delta x)$ whose detection is delayed depending on the delay line position Δx **[21]**:

$$S_{THz}(t, \Delta x) = R_{PCA}(t) * E_{THz}(t, \Delta x), \tag{II.7}$$

where R_{PCA} is the impulse response of the photoconductive antenna. The recorded terahertz signal $S'_{THz}(\Delta x)$ is the sum of the terahertz signal delivered by the detector:

$$S'_{THz}(\Delta x) = \int_{-\infty}^{+\infty} S_{THz}(t, \Delta x)\, dt, \tag{II.8}$$

and with equation II.6:

$$S'_{THz}(\Delta t) = \frac{2}{c} S'_{THz}(\Delta x). \tag{II.9}$$

From the temporal response one can calculate the equivalent representation in the Fourier

domain. Such a transformation is famously known as the Fourier transform (FT) whose operator $\mathscr{F}$ applied on a real-valued temporal function f gives **[22]**:

$$\mathscr{F}(f) : \nu \longrightarrow \hat{f}(\nu) = \int_{-\infty}^{+\infty} f(t) e^{-2i\pi\nu t} dt. \tag{II.10}$$

The Fourier transform describes the frequency spectrum of the periodic temporal signal. However, such an operation is only feasible analytically[2]. Indeed, numerically, only the Fourier series can be calculated. Fourier series aim to describe a T-periodic -let's insist on this- function as a sum of harmonically related sinusoidal sub-functions **[23]**:

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n(f) e^{2i\pi\frac{n}{T}t}, \tag{II.11}$$

where $c_n(f)$ are the Fourier coefficients of harmonic-rank n:

$$c_n = |c_n| e^{i \arg(c_n)}, \tag{II.12}$$

where the harmonic rank n is a cosinusoidal function. However these definitions led to the development of specific algorithms, named fast Fourier transform (FFT) to calculate numerically an approximation of the Fourier transform of a temporal function. Throughout this thesis, Fourier transform of temporal signal are calculated by means of the FFT functions integrated in MATLAB® and Noctylio® softwares.

The frequency resolution is dictated by the experimental system and is equal to the inverse of the temporal window over which was performed the measurement. As an example, for a temporal window of 50 ps, the induced frequency resolution is about 20 GHz. The typical waveform and spectrum recorded by the experimental technique previously mentioned, called time-domain spectroscopy (TDS) and in association with the considered time-frequency transformation are given in Figure II.6.

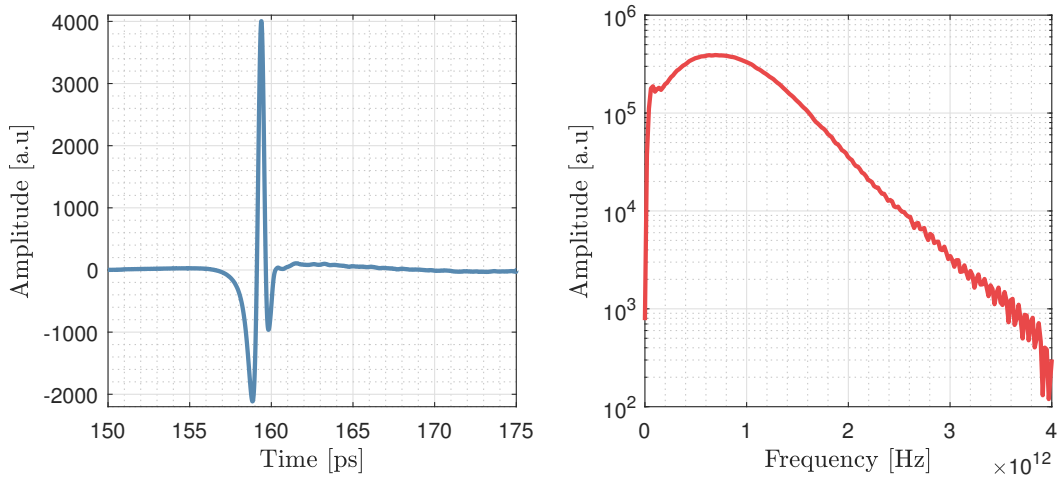


Figure II.6: Illustrative terahertz time-domain waveform (blue) and spectrum (red) recorded through the use of a TDS setup.

II.2.4 Noise Sources in Terahertz-TDS Systems

One can distinguish two main configurational origins of noise in terahertz TDS setups: first, signal dependent noises and second, *a priori* signal independent noises. The first category comprises the noise arising from the emitting photoconductive antenna $\sigma_e(\omega)$ and the intrinsic shot noise of the detecting one $\sigma_d(\omega)$, proportional to the recorded induced photocurrent. The second category

[2] If the function is integrable

gathers all the other contributing sources taking place in the detector $\sigma_i(\omega)$ like the thermal noise, the Johnson noise, the laser noise or the noise arising from the signal amplification **[21]**. The variance of the modulus of any measured signal can therefore be expressed as the sum of these contributions **[24; 25]**:

$$\sigma_s^2(\omega) = \rho^2(\omega)\sigma_e^2(\omega) + \sigma_d^2(\omega) + \sigma_i^2(\omega), \tag{II.13}$$

where $\rho(\omega)$ depicts the specific dielectric properties of the sensed medium. Let's note that $\sigma_e^2(\omega)$ is multiplied by $\rho^2(\omega)$ instead of simply $\rho(\omega)$ since $\sigma_e(\omega)$ is transmitted to the detector.

An additional source of noise, which does not belong to configurational noises, is the presence of atmospheric water vapor **[26]**. Indeed, as reported in *Chapter I. Context and Key Concepts*, water molecules strongly absorb terahertz radiations. Hence, the terahertz beam path and the sample chamber have to be purged with gaseous nitrogen, almost totally transparent to terahertz fields.

II.2.5 Teraview (Ltd, Cambridge UK) Systems

Heretofore, theoretical considerations have been raised toward sensing with terahertz waves. Teraview (Ltd, Cambridge UK) has developed several commercially available systems. Among them, both the TPS Spectra 3000 and Terapulse 4000 (Figure II.7) have been used to conduct the work behind this dissertation.

Figure II.7: Photograph of the TPS Spectra 3000 (left) and the Terapulse 4000 (right).

Hereafter, in Table II.2, are listed their respective characteristics:

Table II.2: Characteristics of the TPS Spectra 3000 and the Terapulse 4000.

Characteristic	TPS Spectra 3000	Terapulse 4000
Spectral Range [THz]	0.06-4	0.06-6
Spectral Resolution [GHz]	7.5	7.5
fs-Laser	Ti:Sapphire	Ti:Sapphire
Emission-Detection	LT-GaAs PCA	LT-GaAs PCA
Portability	No	Yes

Both setups can either work in transmission or in reflection geometries whose configurations are similar to those depicted in Figure II.4 and Figure II.5 respectively. Additionally, motorized stages have been implemented in each systems to perform imaging. However, let's note that this section is focused on spectroscopy and imaging will be treated in the next chapter.

II.3 Spectroscopy: Principles

Behind spectroscopy lies the idea that materials, due to their intrinsic dielectric properties, to their specific molecular arrangement and to their possible scattering behavior, should exhibit specific features when illuminated with an external electromagnetic force field. Hence, the analysis of these features should theoretically allows one to characterize the properties behind such a response. Usually, the extraction of these (often) frequency-dependent properties can be performed either analytically or numerically. Those extractions are performed via the determination of a so-called transfer function $T(\omega)$ which denotes the disturbance to which is submitted a terahertz radiation when propagating through or when reflected by an object.

II.3.1 Transfer Function

The transfer function $T(\omega)$ is a dimensionless quantity that describes the field of the beam path disruption in the presence of specific dielectric materials, morphology and topology. The transfer function $T(\omega)$ is expressed as:

$$T(\omega) = \frac{S_{sam}(\omega)}{S_{ref}(\omega)}, \tag{II.14}$$

where $S_{sam}(\omega)$ and $S_{ref}(\omega)$ are the terahertz signal recorded with and without the sample respectively. Hence, $S_{ref}(\omega)$ is usually referred to as the reference signal. The transfer function can be alternatively expressed to highlight its complex nature:

$$T(\omega) = \frac{|S_{sam}(\omega)|e^{-i\phi_{sam}}}{|S_{ref}(\omega)|e^{-i\phi_{ref}}}, \tag{II.15}$$

where $|S_{sam}(\omega)|$ and $|S_{ref}(\omega)|$ are the modulus of the sample and reference signals and ϕ_{sam} and ϕ_{ref} their respective phase. As already pointed out in the previous chapter, one has concomitantly access to both the amplitude and the phase which allows to refrain from the use of Kramers-Kronig relations. Let's additionally note that the transfer function expresses in a similar way for both regimes, namely transmission and reflection. However, their respective frequency dependence are accordingly modified as $S_{sam}(\omega)$ corresponds in each case to different physical processes. Effectively, the terahertz field either transmitted or reflected by an object to be characterized depends on its thickness, on the corresponding Fresnel coefficients at dielectric interfaces and on the incident field. During this thesis, a mathematical model was developed to generalize the interaction for an infinite number of subsequent medium and consequently, to determine the analytical expression of $S_{sam}(\omega)$. Moreover, as interactions upon dielectric interfaces involve various possible optical paths among structures, the theoretical model was specifically designed to monitor the optical paths that contribute the most to the object response. Such an information was yet to be provided as it fully describes the physical origin of the recorded signal and allows to derive a parametric transfer function. To do so, an iterative approach was considered and is, hereafter reported.

II.3.2 Iterative Determination of $S_{sam}(\omega)$

The cornerstone of the model is to describe the signal propagation by an iterative approach, following the scheme of a tree structure **[27]**. To this aim, let us consider a plane wave incoming with normal incidence onto an object which satisfies: first, homogeneous, flat and parallel boundaries; Second, $\chi^{(1)}$ electromagnetic response; Third, absence of surface charges. Note that the mathematical background to be described remains valid for any wave incident angle via the insert of corresponding angular components. Pulse propagation depends on both the complex refractive index and the medium thickness. When a propagating pulse reaches a dielectric boundary it encounters permittivity change. Permittivity discontinuity at dielectric interfaces leads to the division of the incoming pulse into a reflected part and a transmitted one. The portions of reflected

and transmitted field at dielectric interfaces are given by Fresnel coefficients (equation I.2 and I.3). Thus, the pulse propagation from a i medium to a j medium, also referred as $i \rightarrow j$ propagation, can be set as a matrix $\mathrm{M}_{i,j}$ whose expression is **[28; 29]**:

$$\mathrm{M}_{i,j}(\omega) = t_{i,j}(\omega)^{-1} \begin{pmatrix} e^{\frac{i\omega\hat{n}_i(\omega)d_i}{c}} & r_{i,j}(\omega) \times e^{\frac{i\omega\hat{n}_i(\omega)d_i}{c}} \\ r_{i,j}(\omega) \times e^{\frac{-i\omega\hat{n}_i(\omega)d_i}{c}} & e^{\frac{-i\omega\hat{n}_i(\omega)d_i}{c}} \end{pmatrix}, \tag{II.16}$$

where ω is the pulsation, $\hat{n}_i$ and d_i are the i-medium complex refractive index and thickness, respectively, and c refers to the speed of light in vacuum. Terms $t_{i,j}$ and $r_{i,j}$ denote transmission and reflection Fresnel coefficients. The propagation functions H_t and H_r in transmission and reflection at the $i \rightarrow j$ dielectric interface can therefore be derived from equation II.16:

$$\mathrm{H}_{t_{i,j}}(\omega) = 1/\mathrm{M}^{11}_{i,j}(\omega), \tag{II.17a}$$

$$\mathrm{H}_{r_{i,j}}(\omega) = \mathrm{M}^{21}_{i,j}(\omega)\, t_{i,j}(\omega), \tag{II.17b}$$

where $\mathrm{M}^{11}_{i,j}$ and $\mathrm{M}^{21}_{i,j}$ are the a_{11} and a_{21} terms of the $\mathrm{M}_{i,j}(m \times n)$ matrix depicted in equation II.16, respectively. Each of the propagation function $\mathrm{H}_{t_{i,j}}$ and $\mathrm{H}_{r_{i,j}}$, corresponds to the portion of the light transferred or reflected at the interface between the i and the j medium following a propagation in the i medium.

These aforementioned propagation functions are then subsequently used to design the structure of the tree describing interactions within the whole object. The root node represents the transmitted field at the interface between the air and the first medium, $\mathrm{H}_{t_{0,1}}$. The access to child nodes is done through the incrementation of a mathematical object k that denotes the successive pulse subdivisions that occur upon interaction with successive dielectric interfaces. Thus, the k-index illustrates the depth of the tree structure. As each physical event, namely transmission or reflection, occurring at interfaces can be seen as the consequence of previous events at neighboring interfaces, the successive child nodes depict further practicable events. A schematic representation of the iterative determination of $\mathrm{S}_{sam}(\omega)$ is given Figure II.8.

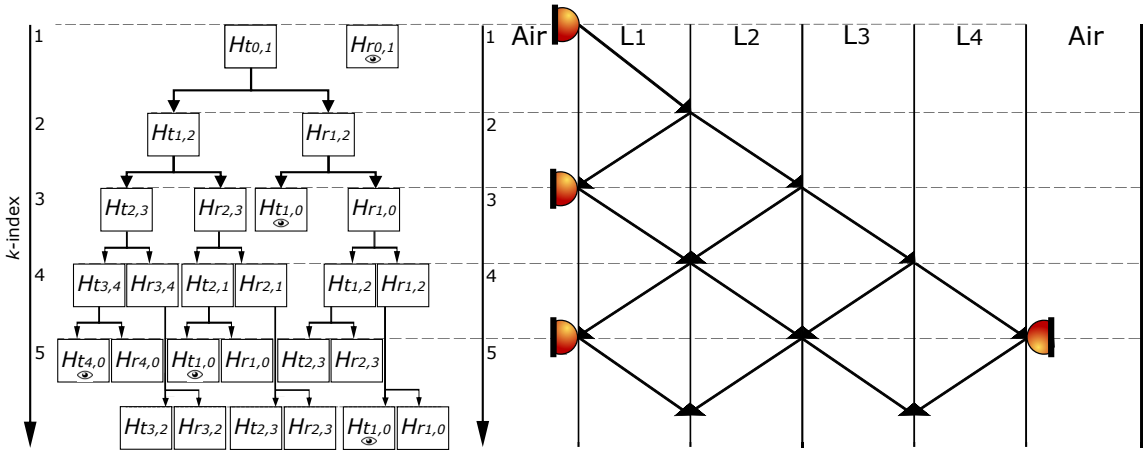


Figure II.8: Scheme of the iterative determination of $\mathrm{S}_{sam}(\omega)$. Each k-incrementation gives access to further possible events based on the previous ones at $k-1$. Detected pulses have been marked with an eye in the node box. The incident angle was tilted for visual convenience.

One has then the following equivalent mathematical description of the scheme:

$$\mathrm{S}^k_{r_{i,i+1}} = \mathrm{H}_{r_{i,i+1}}(\mathrm{S}^{k-1}_{r_{i,i-1}} + \mathrm{S}^{k-1}_{t_{i-1,i}}), \tag{II.18a}$$

$$\mathrm{S}^k_{r_{i,i-1}} = \mathrm{H}_{r_{i,i-1}}(\mathrm{S}^{k-1}_{r_{i,i+1}} + \mathrm{S}^{k-1}_{t_{i+1,i}}), \tag{II.18b}$$

$$\mathrm{S}^k_{t_{i,i+1}} = \mathrm{H}_{t_{i,i+1}}(\mathrm{S}^{k-1}_{r_{i,i-1}} + \mathrm{S}^{k-1}_{t_{i-1,i}}), \tag{II.18c}$$

$$\mathrm{S}^k_{t_{i,i-1}} = \mathrm{H}_{t_{i,i-1}}(\mathrm{S}^{k-1}_{r_{i,i+1}} + \mathrm{S}^{k-1}_{t_{i+1,i}}), \tag{II.18d}$$

where S_r^k and S_t^k are the reflected and transmitted fields, in frequency domain of the k_{th} iteration at each dielectric interfaces. Note that the set of equations II.18 is ω-dependent but the dependence has been deliberately removed for visual convenience. Hence, one can derive from the four previous equations $S_t(\omega)$ and $S_r(\omega)$:

$$S_t(\omega) = \sum_{k=\gamma+1}^{n} S_{t_{\gamma,0}}^k(\omega), \tag{II.19a}$$

$$S_r(\omega) = S_{r_{air,1}}^1(\omega) + \sum_{k=3}^{n} S_{t_{1,0}}^k(\omega), \tag{II.19b}$$

where γ stands for the total number of different dielectric medium that makes the stack. Let's however underline that $S_{t_{\gamma,air}}^k(\omega) = 0$ for $k < \gamma + 1$ and $S_{t_{1,0}}^k(\omega) = 0$ for $k < 3$. Additionally, the parity of the number of subsequent medium γ impacts the existence of $S_{t_{\gamma,air}}^k(\omega)$ and $S_{t_{1,0}}^k(\omega)$ terms. As an example, for $\gamma = 4$, $S_{t_{1,0}}^k(\omega) \neq 0$ for $k \in 2\mathbb{N}+1$ and $k \geqslant 3$. The procedure was programmed in the interpreted MATLAB® language. Ultimately, one can derive the expression of the transfer function $T_t(\omega)$ and $T_r(\omega)$ in transmission and reflection respectively:

$$T_t(\omega) = \frac{\sum_{k=\gamma+1}^{n} S_{t_{\gamma,0}}^k(\omega)}{S_{ref}(\omega)}, \tag{II.20a}$$

$$T_r(\omega) = \frac{S_{r_{0,1}}^1(\omega) + \sum_{k=3}^{n} S_{t_{1,0}}^k(\omega)}{S_{ref}(\omega)}, \tag{II.20b}$$

The indexation of each physical events occurring at singular dielectric interfaces after a given number of internal subdivision allows one to retrieve the most contributing optical routes for any recorded signal. Hence, it is possible to derive a parametric global transfer function whose input parameters only consist in the necessary transfer function combinations. To illustrate such a statement and to validate the mathematical background previously established, a calibration material made of four dispersive layers (medium) of tens of microns deposited on a metal plate was selected. The procedure is validated if and only if: first, it reconstructs adequately the experimental signal; Second, it sorts the most participating paths; Last, it derives a parametric transfer function that reduces the number of required computations to reconstruct the signal. A microscopic view of the calibration sample is provided Figure II.9.

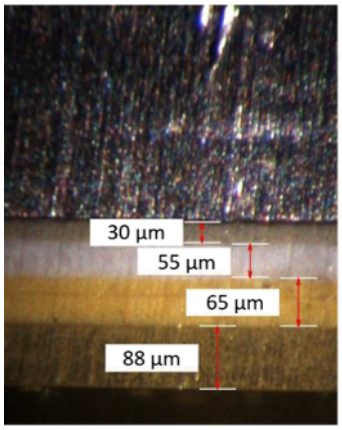


Figure II.9: Microscopic examination of the calibration sample.

Each layer dielectric properties and thickness were known as furnished by the manufacturer. The Debye parameters of each layer are given in Table II.3. From these respective parameters one

can derive the frequency dependent refractive index $n(f)$ and extinction coefficient $k(f)$, shown in Figure II.10.

Table II.3: Fitting Debye parameters for the four coatings of the calibration sample. ϵ_∞ stands for the dielectric constant at high frequencies, $\Delta\epsilon$ is the permittivity difference between the static dielectric constant and the relaxation process permittivity, occurring over the time τ. d stands for the thicknesses extracted from optical measurement.

Layer	ϵ_∞	$\Delta\epsilon$	τ [fs]	d [µm]
L_1	2.48	0.39	70.82	30 ±1
L_2	4.26	0.92	83.73	55 ±1
L_3	3.23	0.71	72.07	65 ±1
L_4	3.72	0.85	83.04	88 ±1

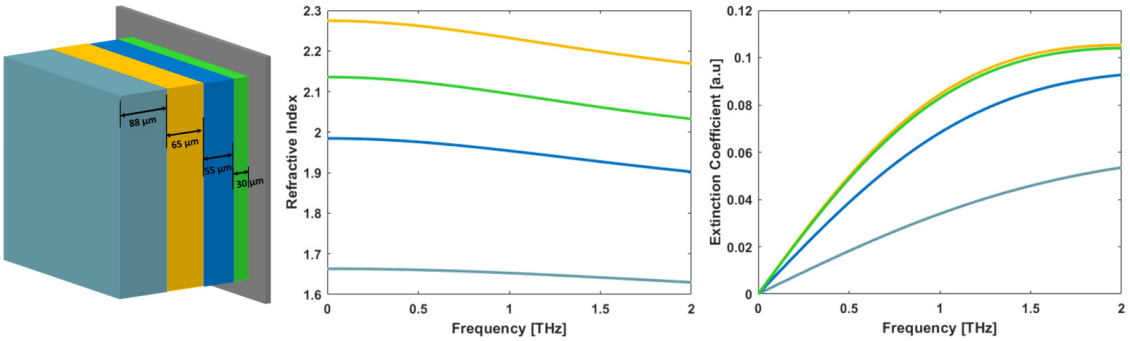


Figure II.10: Refractive index $n(f)$ and extinction coefficient $k(f)$ of each medium belonging to the calibration sample.

To test the theoretical model, a reference signal is recorded from a mirror and then a signal is measured in reflection from a 4-layers sample containing the four coating materials listed previously. Numerical nodes on layer interfaces are filled with the extracted and fitted[3] properties shown in Figure II.10. Then, the reference signal $S_{ref}(\omega)$ recorded from a reflector is numerically injected into the algorithm to compare the numerical reconstruction and the experimental signal. At each pulse division, the $k-$parameter is incremented automatically and descendant pulses are computed. The depth of the tree was limited to $k = 29$, based on the hypothesis that beyond this specific k-value, the recovered signal amplitudes are under the noise level. In Figure II.11 is shown the progressive reconstruction of $S_r(t)$. For clarity purposes, only odd $k-$iterations are displayed as even $k-$value iterations do not satisfy photo-detection conditions in reflection configuration (respectively to Figure II.8).

The metric $\Lambda(t)$ used to establish the discrepancy between the measured signal and the iterative tree algorithm (ITA) reconstruction (Figure II.12) within the time window lying between 22 and 37[4] ps denotes the amplitude time-domain quadratic difference between the two signals:

$$\Lambda(t) = |S_r^M(t) - S_r^S(t)|^2, \tag{II.21}$$

where $S_r^M(t)$ and $S_r^S(t)$ are the measured and calculated amplitudes, respectively. The variations observed are assumed to be caused by a combination of limitations arising from the inaccuracy of the Debye parameters and the possible FFT numerical roundoff errors. However, the quadratic difference between the experimental and the reconstructed signals is globally low, exhibiting a maximum value for residues by about 0.2. Additionally, $\int_{22ps}^{37ps} \Lambda(t) = f(k)$ is provided Figure II.13. It is hence reasonable to confirm the first criterion toward the validation of the procedure. In order

[3] Fitting procedure is discussed later in the manuscript.

[4] The signal recorded outside this specific time window being mostly white Gaussian noise not related to the stratified structure.

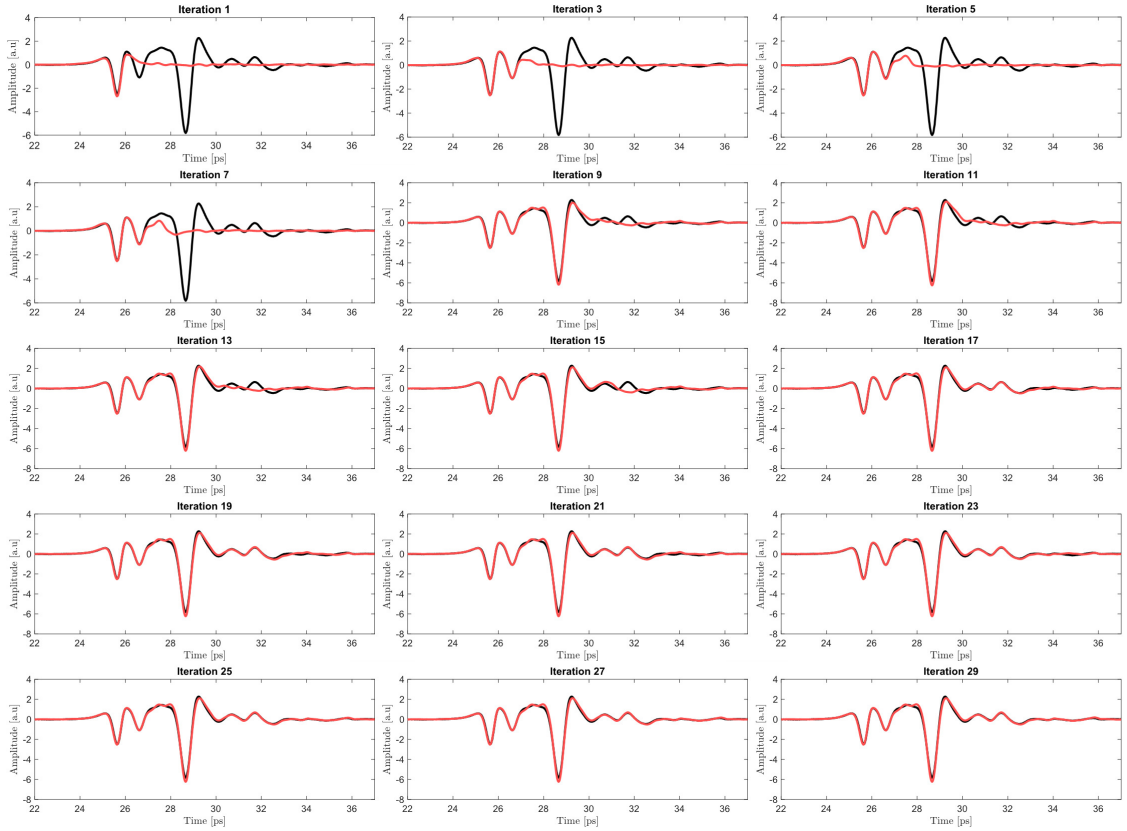


Figure II.11: Progressive reconstruction of $S_r(t)$ (red) in comparison to the experimental signal (black) for the calibration sample made of four layers each of tens of microns.

to assess the appropriateness of the assumption about the usefulness of iterating beyond $k = 29$, the frequency-dependent relative field was investigated for each k-value that satisfies $S^k_{t_{1,air}}(\omega) \neq 0$. Beyond $k = 17$, the contribution to the signal is globally low. Indeed, a large amount of information carried by the highest frequencies falls below their relative noise level. Additionally, the dispersive behavior of the subsequent medium enhances the difference between the relative field carried at low frequencies and the one carried by higher ones as k increases (Figure II.14). The previously proposed assumption seems thus to be reasonable.

II.3.3 Parametrization of $S_{sam}(\omega)$

The next step is to confirm the ability of the procedure to successfully retrieve the most contributing routes and to derive a parametric transfer function, reducing consequently the number of paths to be computed. Before deriving such a function, let's examine the room for improvement toward minimizing the number of calculated routes. In the specific case of the four layers calibration material, the total number of optical paths computed after 29 internal subdivisions is 13,153,171. Among them, 1,195,750 paths have been theoretically detected by the detector in reflection configuration. The number of detected optical paths follows a modified Fibonacci sequence $\zeta(k)$ whose expression was found to be:

$$\zeta(k) = \zeta(k-2) + 3^{(\frac{k-\gamma+1}{2})}, \tag{II.22}$$

where $k > 3$ and the number of successive medium γ is still four in this case. The value of $\zeta(k)$ for $k < 3$ is intuitive, as there is only one optical path that satisfies $S^3_{t_{1,0}}(\omega) \neq 0$, hence $\zeta(3) = 1$ and $\zeta(1) = 1$ which corresponds to $S^1_{r_{0,1}}(\omega)$. In Table II.4 are reported the number of computed optical paths and the number, among them, that are detected by the *numerical* photoconductive antenna, *i.e.* $S^k_{t_{1,0}}(\omega) \neq 0$.

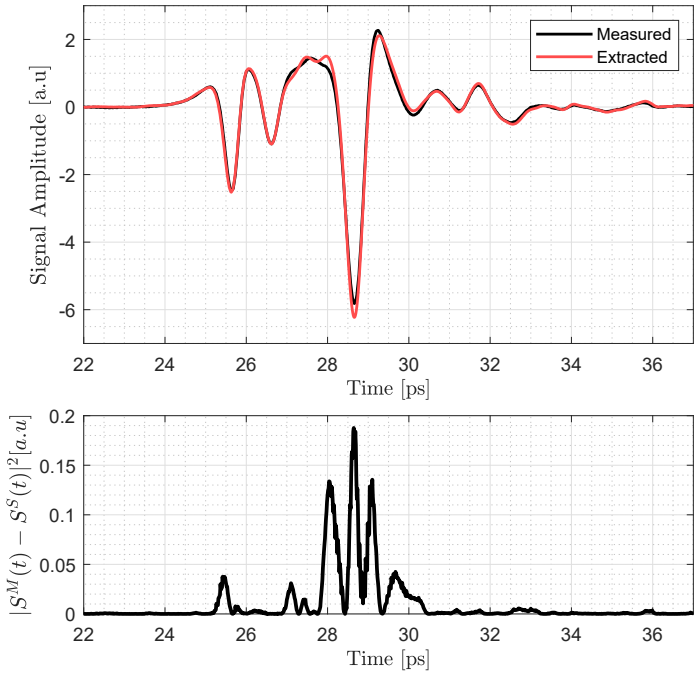


Figure II.12: Top: experimental signal of the stratified structure in reflection geometry (black) and its reconstruction via the iterative tree algorithm (red). Bottom: deviations observed from the measured signal using the $\Lambda(t)$ metric.

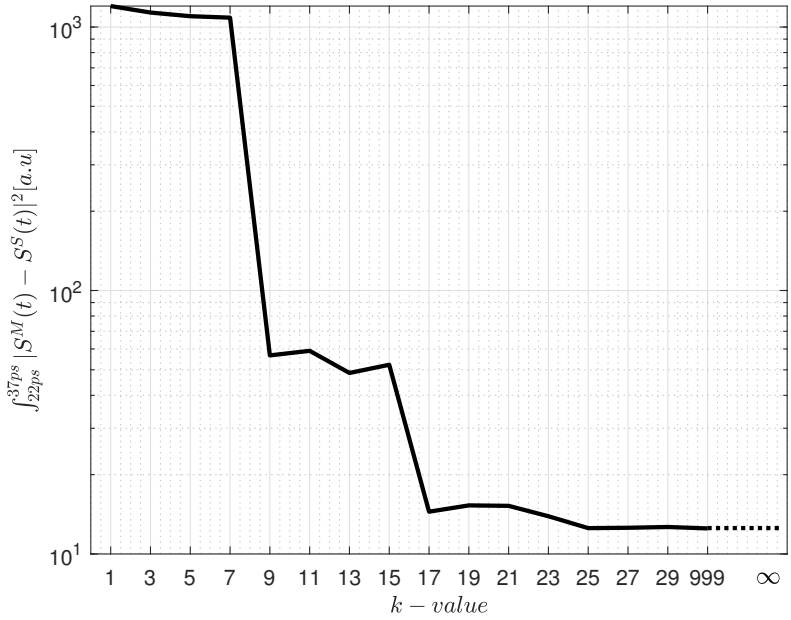


Figure II.13: Profile of the integrated comparison metric between 22 ps and 37 ps as a function of the number of iterations. The non-monotonous nature of the dependence reflects that time features are in majority due to the combination of several physical processes. Despite additional iterations, the error does not fall to zero. That is caused by unrecovered time features for the previously mentioned reasons and because time features corresponding to $k \to \infty$ are basically out of the time window in which the comparison metric is defined.

Heretofore, 1,195,750 different optical paths have been computed to reconstruct the experimental signal depicted in Figure II.12. Reasonably, one can state that among all of them, only a

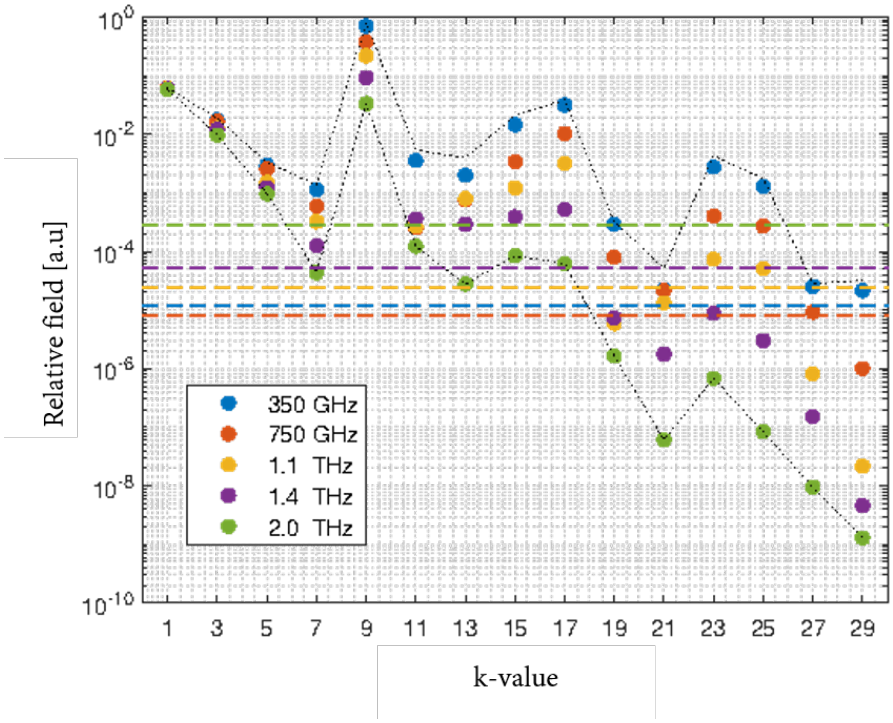


Figure II.14: Relative field at different frequencies as a function of the $k-$value and their respective relative noise level (dashed lines). The relative noise level for each frequency is defined as the ratio of the noise level of the reference with respect to the reference field carried at these frequencies. The two dotted black lines represent the maximum and the minimum relative field value over the 0.2 - 2 THz spectral range observed at each $k-$value. Due to the dispersive behavior of coating layer, the difference between the field carried at low frequencies and the field carried at high frequencies increases as k increases.

Table II.4: Number of computed paths for each k-incrementation and the number among them outgoing the structure being thus detected.

$k-$value	Total Paths	Detected Paths	$k-$value	Total Paths	Detected Paths
1	2	1	16	4,374	-
2	2	-	17	7,655	1,094
3	4	1	18	13,122	-
4	6	-	19	22,964	3,281
5	11	2	20	39,366	-
6	18	-	21	68,891	9,842
7	32	5	22	118,098	-
8	54	-	23	206,672	29,525
9	95	14	24	354,294	-
10	162	-	25	620,015	88,574
11	284	41	26	1,062,882	-
12	486	-	27	1,860,064	265,721
13	851	122	28	3,188,646	-
14	1,458	-	29	5,580,131	797,162
15	2,552	365	Total	13,153,171	1,195,750

few are actually participating to the signal. Thanks to the mathematical design of the algorithm, each optical route at each k-value has been collected and sorted as a function of their relative field contribution. The most predominant frequency contributions have been located until $k = 17$ (Figure II.13), thus only optical paths for which the number of internal subdivision is $k \leqslant 17$ were

investigated. The specific field distribution among these $k-$value is structure dependent and may drastically change from one to another dielectric assortment. At each considered $k-$value, the numerical procedure retrieves the predominant optical paths whose sum provides a quadratic difference between the reconstruction considering all optical routes and the reduced one of less than 0.001. Among the whole population of detected paths for $k \in [1;17]$, only 33 optical paths have been automatically saved corresponding roughly to 2% of the detected pulses computed previously. From the small number of retained optical paths, an optimized transfer function for radiation propagation is derived. The theoretical signal is computed back and is compared to the initial signal provided by the reconstruction that includes all optical paths. The waveform computed using these 33 optical paths is given in Figure II.15.

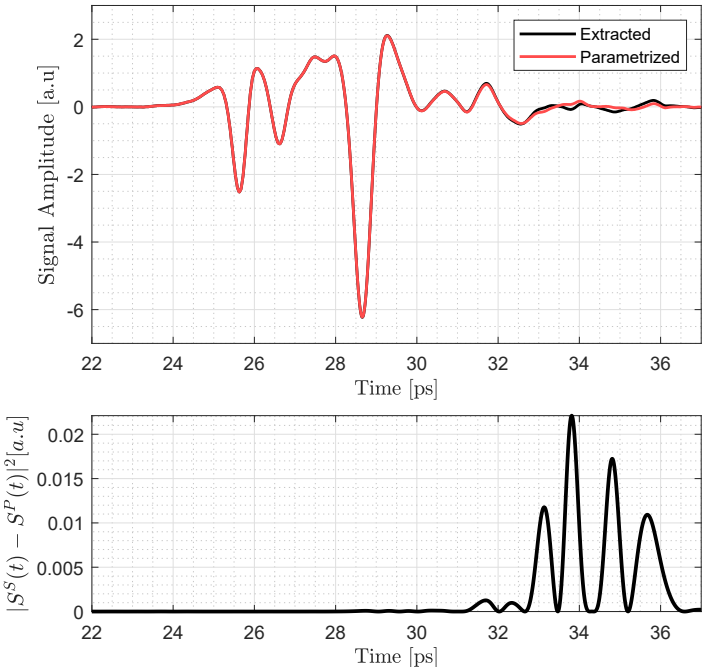


Figure II.15: Top: Reconstruction of the signal provided by the reduced iterative tree algorithm (black cross) with the 33 selected contributing paths (red). The correlation is almost perfect between the two signals over the entire time window. Bottom: $\Lambda(t)$ metric. Differences are observed in regions where additional subdivided pulses have been ommitted (*e.g.* $k > 17$).

The deviations between the two shapes are mainly observed after 32 ps from where, the $k > 17$ contributions have been ommitted. Retained optical routes for $k > 7$ are reported in Figure II.16. The parametric transfer function derived following the optical path analysis almost perfectly retrieved the initial computed signal. The assumption stating that among the wholeness of optical paths only a few were responsible for the time-domain waveform is hence valid. Indeed, considering propagation routes submitted to 17 physical processes (transmission or reflection) before being detected at the most, a total of 1645 paths were recorded (*i.e.* $\sum_{k=1}^{k=17} \zeta(k)$). However, the path analysis enlightened that only 2% of them were significantly contributing. The aforementioned procedure has therefore successfully passed through all the validation stages. Let's however underline that the validity limit of the new transfer function is sample dependent and remains genuine for small variations of the dielectric properties and for few microns concerning the thickness of individual medium.

II.3.4 Optical Property Extraction

Knowing both the analytical expression of the transfer function T(ω) reported through the couple of equation II.20 and the experimental expression, one can extract the dielectric properties of the

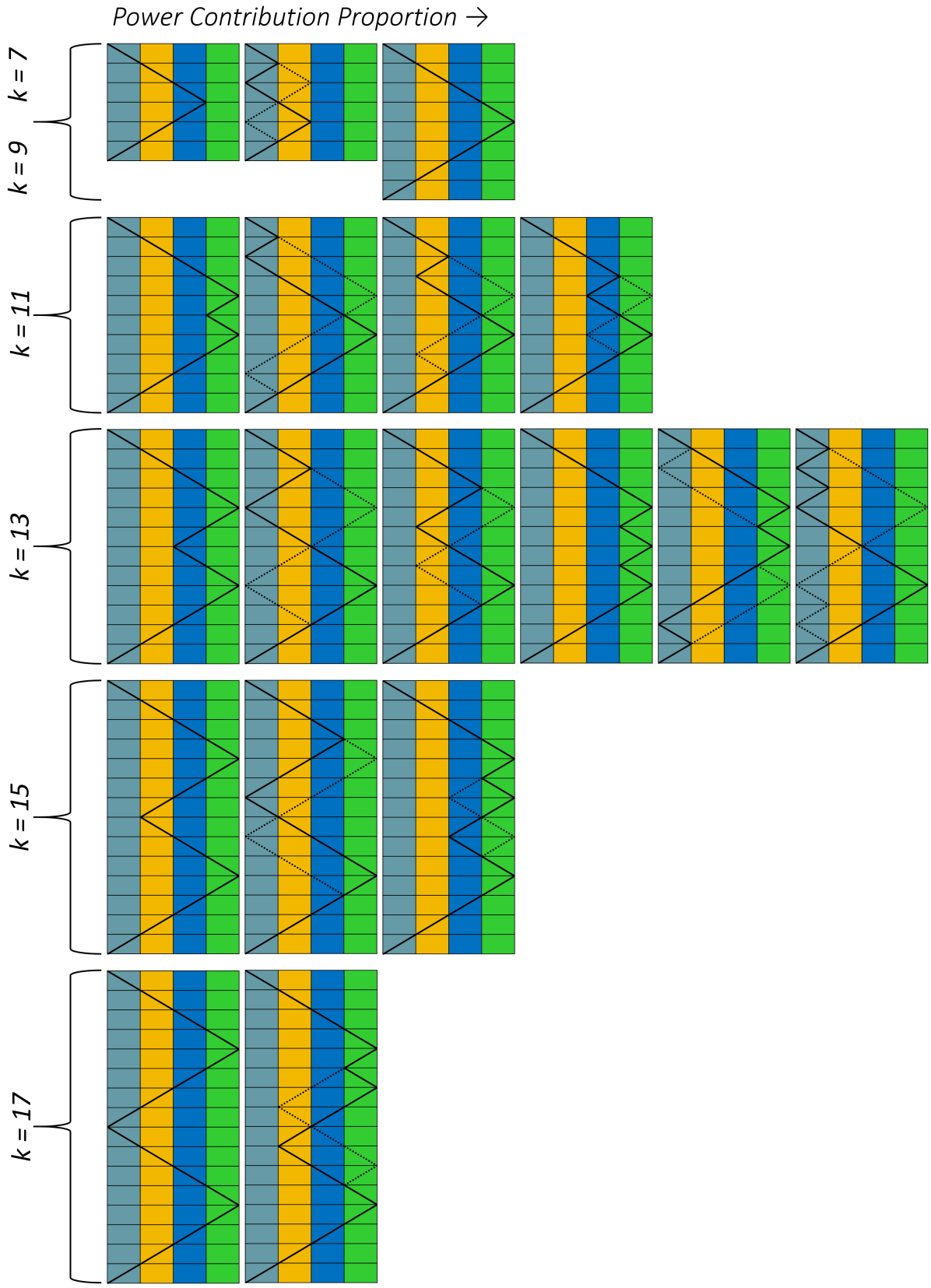


Figure II.16: Representation of the main optical paths sorted by the iterative tree procedure for k-index ranging from 7 through 17. The optical paths at each k-values are sorted from left to right from the highest contribution to the signal to the lowest. Particular optical paths exhibit a symmetric one represented in dashed lines. These symmetric paths participate in an equivalent fashion to the signal.

sample under inspection. Although these properties are *a priori* unknown, the observed waveform or spectrum are directly dictated by them. One can distinguish three properties that are going to impact the complex transfer function. Among them, two are intrinsic properties, namely the refractive index $n(\omega)$ and the extinction factor $\kappa(\omega)$ and an extrinsic parameter, the sample thickness d. Although being obvious, let's underline that d does not exhibit frequency dependence. Additionally, in view of equations II.20a and II.20b, these properties cannot be expressed analytically from the complex transfer function.

In order to extract $n(\omega)$, $\kappa(\omega)$ and d, one has therefore to treat the underlying inverse electromagnetic problem (IEP). To do so, one generates a volume whose $a^{m\times n\times l}$ terms represent the whole possible combinations between n, κ and d. These different combinations are therefore successively injected into the analytical expression of the complex transfer function (II.20). For each conceivable values of the sample thickness d, the couple (n,κ) for which the difference with the experimental spectrum is the lowest at each frequency is kept. The statement can be formulated mathematically as:

$$\underset{n(\omega),\kappa(\omega),d}{min} \quad \mathrm{T}^{\mathrm{M}}(\omega)-\mathrm{T}^{\mathrm{S}}(\omega), \qquad \text{subject to } \begin{cases} n=(n_1;...;n_i)\in\Re_+\setminus\{0;1\} \\ \kappa=(\kappa_1;...;\kappa_i)\in\Re_+ \\ d=(d_1;...;d_i)\in\Re_+ \end{cases} \tag{II.23}$$

where, $\mathrm{T}^{\mathrm{M}}(\omega)$ and $\mathrm{T}^{\mathrm{S}}(\omega)$ are the measured and simulated complex transfer functions. However, such a problem suffers from the intrinsic definition of the complex transfer function. In fact, it has to be noticed that the complex transfer function is a mod 2π function which consequently introduces an infinity of solution instead of an unique one as expected. Searching for zeros of $\mathrm{T}^{\mathrm{M}}(\omega)-\mathrm{T}^{\mathrm{S}}(\omega)$ would therefore require to start an optimization with guess parameters close to the actual physical properties of the sensed sample. When probing totally unknown samples such an approach is hence obviously not recommended. One has therefore to treat the inverse electromagnetic problem with alternative functions to be minimized. Such functions are referred to as objective functions $\chi(\omega)$ in the following. As the oscillating behavior of the complex transfer function was the source of an infinity of solutions, one has then to express the objective function to be minimized ($\in\Re_+$) as a function of monotonous and plane discrepancy functions. As already pointed out in equation II.15, the complex transfer function $\mathrm{T}(\omega)$ can be decomposed into two entities, namely its modulus $|\mathrm{T}(\omega)|$ and its argument $\angle\mathrm{T}(\omega)$. Both $|\mathrm{T}(\omega)|$ and $\angle\mathrm{T}(\omega)$ are smooth functions when plotting $n(\omega)$ versus $\kappa(\omega)$ **[30]**. Thus, it seems convenient to determine the problem objective function as a function of the modulus and the argument such as **[31]**:

$$\chi(\omega)\propto|\mathrm{T}^{\mathrm{M}}(\omega)-\mathrm{T}^{\mathrm{S}}(\omega)|+(\angle\mathrm{T}^{\mathrm{M}}(\omega)-\angle\mathrm{T}^{\mathrm{S}}(\omega)). \tag{II.24}$$

The symbol "$\angle$" refers to the phase. During this work, the objective function to be minimized to extract the dielectric properties was generally defined as **[32; 33]**:

$$\chi(\omega)=\delta\mathrm{M}^{\mathrm{T}}(\omega)\delta\mathrm{M}(\omega)+\zeta\delta\Phi^{\mathrm{T}}(\omega)\delta\Phi(\omega), \tag{II.25}$$

where $\delta\mathrm{M}(\omega)$ and $\delta\Phi(\omega)$ are the amplitude and phase penalty respectively, and ζ is a regularization parameter weighting phase penalty regarding amplitude. Let's note that the introduction of the regularization factor ζ is not compulsory and can simply be set to unity. However, in some specific cases where one of the two entities is not reliable due to experimental configurations, the regularization parameter ζ may be set to different values. That may be, for example, the case of reflection geometry where the information on the experimental phase is often lowly reliable. The $\delta\mathrm{M}(\omega)$ and $\delta\Phi(\omega)$ matrices are respectively defined as:

$$\begin{cases} \delta\mathrm{M}(\omega)=ln(|\mathrm{T}^{\mathrm{M}}(\omega)|\times|\mathrm{T}^{\mathrm{S}}(\omega)|)^{-1}), \\ \delta\Phi(\omega)=\angle\mathrm{T}^{\mathrm{M}}(\omega)-\angle\mathrm{T}^{\mathrm{S}}(\omega). \end{cases} \tag{II.26}$$

The natural logarithmic ratio instead of a standard quadratic difference for the measured and the simulated amplitude is preferred as it is more penalizing. Let's additionally remark that as far as

$\delta M(\omega)$ and $\delta\Phi(\omega)$ are *a priori* symmetric matrices, one has the following equality:

$$\begin{cases} \delta M^2(\omega) = \delta M(\omega)\delta M^T(\omega), \\ \delta\Phi^2(\omega) = \delta\Phi(\omega)\delta\Phi^T(\omega). \end{cases} \tag{II.27}$$

To illustrate the extraction procedure, the frequency dependent optical properties as well as the thickness of a perfectly unknown sample are hereafter extracted. The only information known about the sample is that it is an amorphous thin layer of less than 100 μm deposited on a metal plate. Therefore, due to the presence of the reflector, the extraction has to be proceeded in reflection geometry. However, due to possible mismatch between the position of the reference and the one of the sample, a specificity of reflection configuration, the experimental information on the phase $\Phi^S(\omega)$ is not trustworthy. In order to prevent from extraction inconsistency, the regularization factor ζ was deliberately set to 0. The 4D problem (along n, κ, d and ω) to be solved is defined as follows:

$$\underset{n(\omega),\kappa(\omega),d}{min} \quad \chi(\omega), \quad \text{subject to} \begin{cases} \zeta = 0 \\ n = (1.6;...;1.8), \ \Delta n = 0.01 \\ \kappa = (0;...;0.08), \ \Delta\kappa = 0.001 \\ d = (10;...;99)\mu m, \ \Delta d = 0.1\mu m \end{cases} \tag{II.28}$$

Numerically, the algorithm tests successively each thickness contained in the d vector and finds the couple $n(\omega), k(\omega)$ minimizing $\chi(\omega)$. Solutions for each d component are then inputted back into a direct propagation problem. Thereafter, the thickness minimizing $\sum_{t=0}^{t_{max}} |S^M(t) - S^S(t)|^2$ is selected by the algorithm (Figure II.17). During the extraction procedure, the algorithm found 63.4 μm to be the physical thickness of the unknown sample used to illustrate the process. Let's underline that it should be considered the extent of the plateau in which this solution is found as the error bar ($\pm 5\mu m$). Now, it should be checked that the extracted waveform adequately matches the experimental one. To this aim, both these signals are compared in Figure II.18.

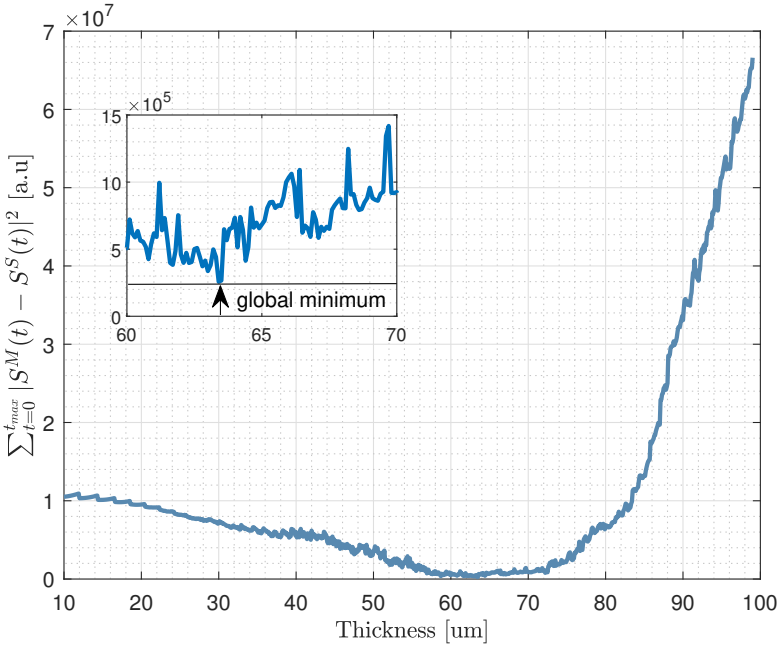


Figure II.17: $\sum_{t=0}^{t_{max}} |S^M(t) - S^S(t)|^2$ as a function of the tested thicknesses contained in the d vector. The global minimum was found to be 63.4 μm.

It is quite obvious that the experimental waveform is well retrieved. One can however argue that some oscillations appear on both sides of the main peak. To understand precisely the origin of these features, one may have a look on the optical properties that have been extracted during the

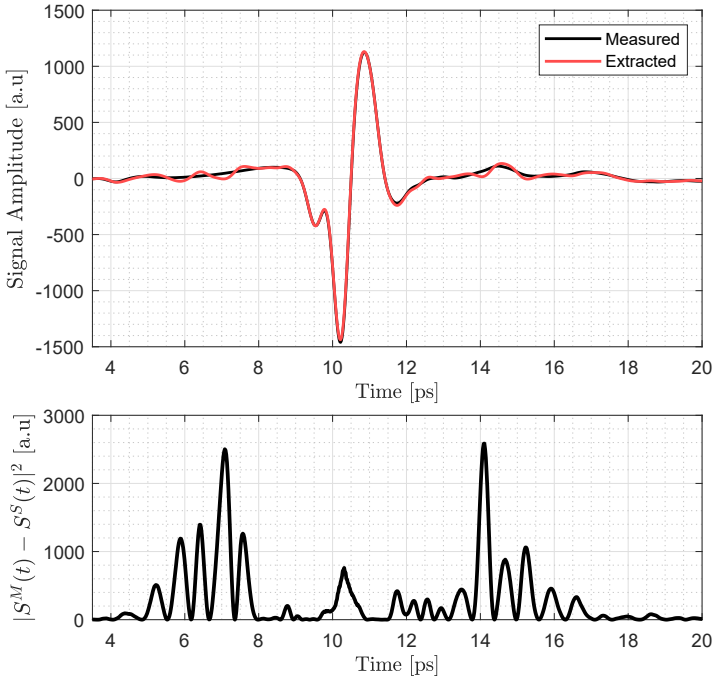


Figure II.18: Time-domain error of the thickness global minimum using the Λ metric.

objective function minimization (Figure II.19). Clearly, both these functions are not monotonous which can be due to slight mismatch between the physical thickness and the extracted one. However, disturbances are mainly observed at low and high frequencies where the signal may suffer from the presence of noise. The use of a Nelder-Mead like algorithm **[34]** that will fit the extracted $n(\omega)$ and $\kappa(\omega)$ with respect to physical models that have been described in *Chapter I, Section III.2 Interaction Mechanisms*, is then required.

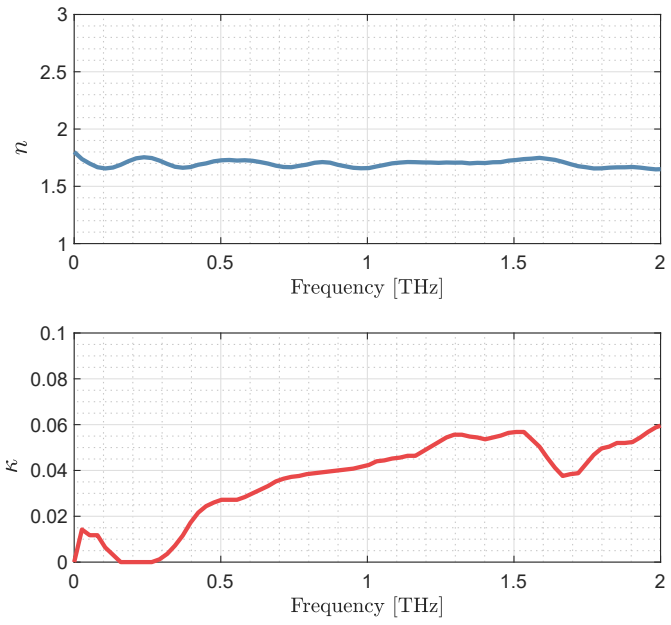


Figure II.19: Optical properties $n(\omega)$ and $\kappa(\omega)$ for 63.4 μm.

II.3.5 Optical Property Fitting

In order to refine the waveform reconstruction, one can fit the extracted optical properties with respect to different models that describe the complex permittivity as a function of the incident electric field. As far as the sample under test is amorphous, it is probable that no particular relaxation occurs excepted the one due to generalized thermal effect. Consequently, a reasonable choice would be to fit both the refractive index and the extinction coefficient with a simple Debye model. To do so, the real and imaginary part of the complex permittivity are calculated with respect to equations I.19, in Chapter I. The algorithm has to fit both the real and the imaginary part. It is therefore essential that the procedure determines the goodness of fit from the distances on the complex plane between the extracted and the calculated points rather than just their real or their imaginary parts **[35]**. To this aim, we make the assumption that residues are equal to:

$$|Res| = \sqrt{\sum_{\omega=0}^{\omega_{max}} \partial\epsilon_r'(\omega)^2 + \sum_{\omega=0}^{\omega_{max}} \partial\epsilon_r''(\omega)^2}, \tag{II.29}$$

where $\partial\epsilon_r'(\omega)$ and $\partial\epsilon_r''(\omega)$ are real and imaginary residuals in the complex plane, respectively. The search for the minimum is then performed by the *fminsearch(·)* function provided by the MATLAB® software. Results toward the fitting procedure are reported in Figure II.20.

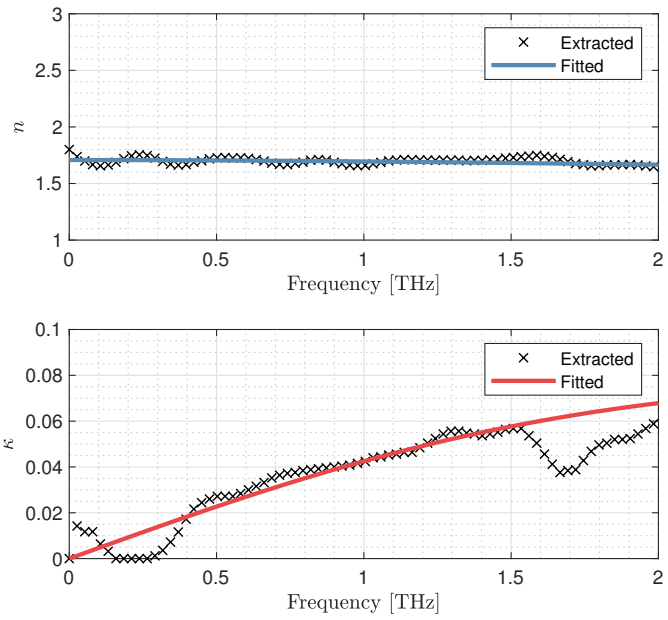


Figure II.20: 1st order Debye (*i.e.* $n = 1$ equation I.22.) fit of the IEP extracted properties.

The corresponding Debye parameters that were found out to be the most suitable to describe the dielectric behavior of the unknown sample are listed hereafter, in Table II.5.

Table II.5: Fitting Debye parameters for the unknown sample.

Parameter	ϵ_∞	$\Delta\epsilon$	τ [fs]	d [µm]
Value	2.42	0.50	50.46	63.40

One may now input back the fitted properties into a direct propagation problem. Both the measured and the extracted-fitted signals are plotted and compared using the $\Lambda(t)$ metric in Figure II.21.

Clearly, oscillations have been removed through the introduction of the monotonous fitted properties. It has to be underlined that the error values may appear quite huge. However, their

amplitude are only due to the recorded amplitude values and the comparison metric $\Lambda(t)$ depicting the square of the difference. Indeed, if considering the minimum value of the recorded signal (around 10 ps) being roughly -1500 a.u, the associated error is about 1000 a.u. Nonetheless, it only corresponds to a difference of approximately 2%, since errors depend on the square of the difference between the two signals.

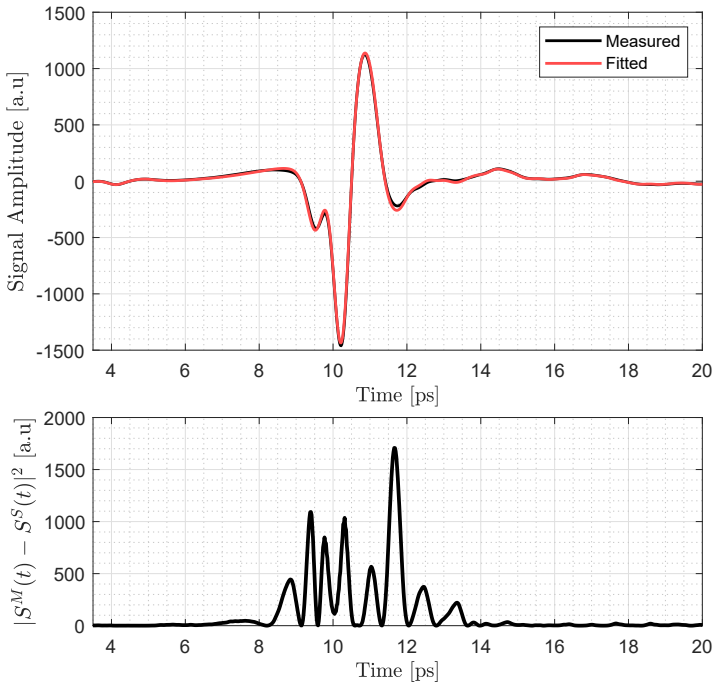


Figure II.21: Reconstructed waveform after fitting the optical properties.

Heretofore, the experimental, the mathematical and the numerical methodologies that have been developed during this work in order to define the complex transfer function of a sample and to extract the related dielectric properties were described. On these basis, one can now focus on their application for biology and further more for breast cancer investigations.

II.4 Water

Among all the fluids found in the organism, water is by far the most abundant. Water is *de facto* the highest contributor to the overall dielectric response of tissues. Consequently, one has first to investigate the liquid water behavior when stimulated with an external terahertz field. To this aim, water was inspected in reflection configuration to prevent from its significant absorption. Obviously, a support is required to drop off a thin film of water onto it. A 2 mm C-cut sapphire was therefore selected. The choice of such a material is not random and deserves further considerations. Indeed, the support has to satisfy few main requirements: first, it should not express a birefringent behavior; Second, it should be lowly absorbent in such a way that the portion of the incident field reaching the water film side is maximized; Last, the refractive index difference between the support and the water film has to be *significant* to maximize the reflection Fresnel coefficient at the dielectric interface. While the first criterion is satisfied by the intrinsic properties of the C-cut sapphire the two others require preliminary investigations toward C-cut sapphire dielectric properties. The time-domain waveform of a pulse passing through the substrate was then recorded and the respective properties derived with respect to aforementioned numerical procedures are shown in Figure II.22.

From the optical profiles extracted in Figure II.22, one can consider that the sapphire substrate is sufficiently transparent to conduct further studies. Additionally, the refractive index of water is

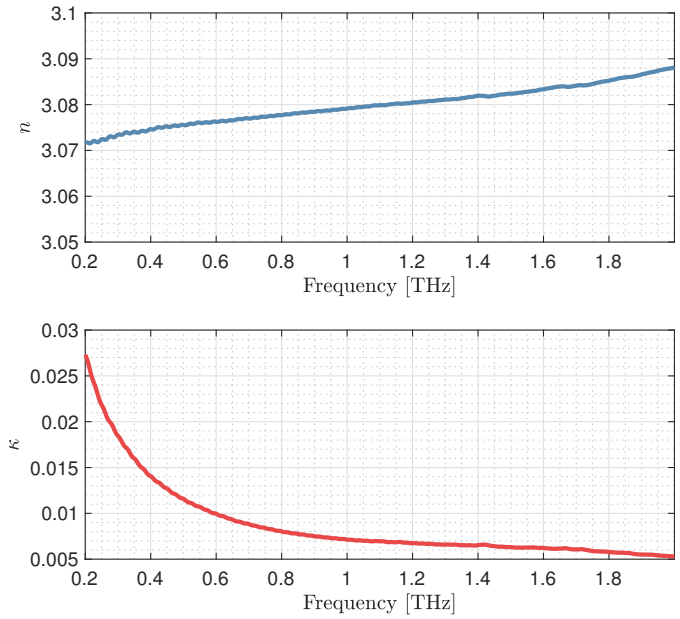


Figure II.22: Refractive index and extinction coefficient of the sapphire substrate.

expected to range from 3 to 2 between 200 GHz and 2 THz thus validating the last criterion[5].

II.4.1 Extracted Intrinsic Properties

Thirteen measurements on water films were performed successively. The water film response was averaged a hundred times for each acquisition. Between each measurement, the water film was removed and replaced by another one. All measurements were performed under controlled nitrogen atmosphere and at room temperature. The typical time-domain waveform recorded is shown in Figure II.23. The sign of each recorded peak denoting reflection at dielectric interfaces depends on the sign of the corresponding Fresnel coefficient. When a pulse is reflected at $i \rightarrow j$, with $n_i < n_j$, $R_{ij} < 0$, the corresponding interface peak is π-shifted. That is for example the case of the first peak arising from the reflection between free space and the sapphire support shown in Figure II.23. The water film information is contained in the second recorded peak, as the first one does not depend on the dielectric properties of the sample put onto the substrate. However, additional peaks may appear every $\Delta t = \frac{2dn_{sapphire}}{c}$ multiple, corresponding to k-value $\geqslant 5$ in equation II.19b and sometimes called Fabry-Pérot terms. By looking at the sign of the second peak, one can appreciate the validation of water dielectric property expectations, *i.e.* $\hat{n}_{water} < \hat{n}_{sapphire}$.

From the knowledge on the characteristics of the sapphire substrate, namely its complex refractive index and its thickness, one can extract the optical properties of water. The extracted properties are given in Figure II.24.

II.4.2 Relaxation Mechanisms

One can now fit the corresponding properties. Although throughout the literature water properties have been fitted with two-Debye terms it seems interesting to prove such a requirement. To do so one derives the real and the imaginary part of the permittivity from the extracted $n(\omega)$ and $\kappa(\omega)$. Therefore, it is relevant to explore the behavior of the imaginary part of the permittivity $\epsilon''(\omega)$ as a function of the real part $\epsilon'(\omega)$, also called Cole-Cole diagram. Indeed, in such a representation, single characteristic frequency (one-Debye term) displays the shape of half a circle. Debye overlapping resonances appear in circular loops. If it is elliptical, it denotes the spread of

[5] Expectation based on water Debye parameters found in literature.

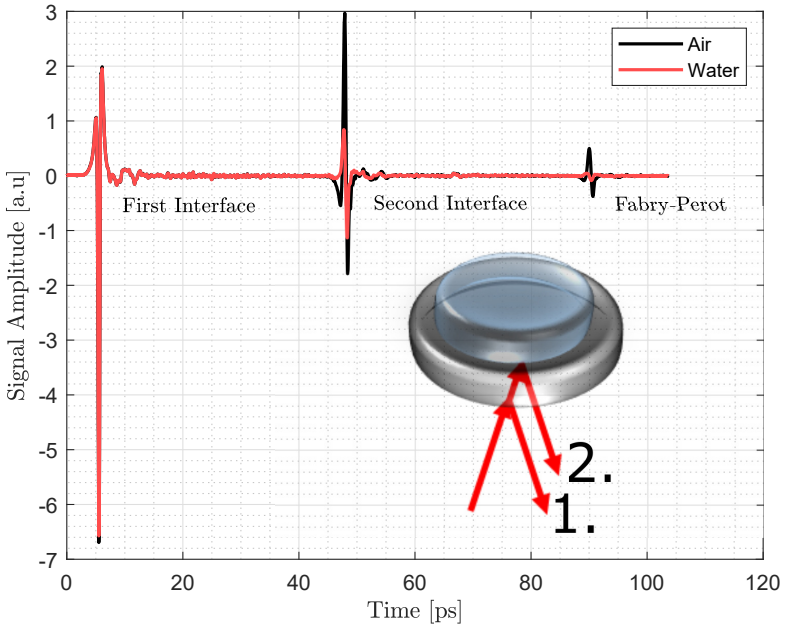


Figure II.23: Typical time-domain waveform recorded without (black) and with (red) water film on the top side on the sapphire substrate.

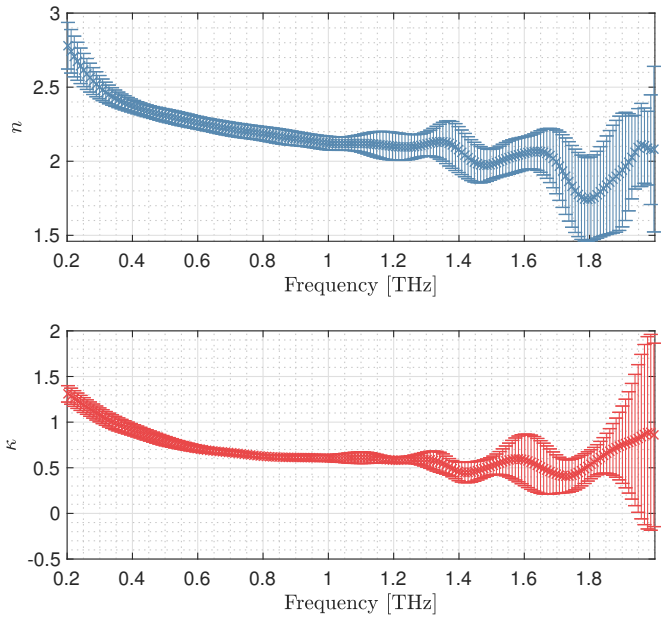


Figure II.24: Water optical properties extracted in reflection geometry, averaged over ten independent measurements.

the characteristic frequencies and thus a Cole-Cole behavior. If on the contrary the plot displays a non-symmetric tendency, it is of Cole-Davidson or Havriliak-Negami nature. The $\epsilon''(\epsilon')$ representation is plotted in Figure II.25. As beyond 1 THz the presence of noise seems to impact the data, only the information between 200 GHz and 1 THz is taken into account.

At low frequencies one observes the left-hand side of a semi-circle. Obviously, due to the lack of low frequencies, only a given portion is observable. Such an observation is of paramount importance. In fact, it tells us that the TDS system that is used to probe the dielectric behavior of water does not provide a sufficient band at low frequencies since the first characteristic relaxation seems

to arise below its monitoring frequency range. Consequently further data processing, mainly fitting processes, will have to take into account such a lack by fixing low frequencies parameters. In a nutshell, the fit of water dielectric properties extracted from TDS systems require to fix the characteristic relaxation time of the first process.

At higher frequencies, rather than having the semi-circular shape close, the slope of the dielectric permittivity tends to smooth, indicating that an additional polarization mechanism comes into play: a second characteristic frequency. One can even observe a "slight bounce" between the two relaxation domains that may indicate a third mechanism. However, we will see that it arises from experimental artifacts and will be rejected for fitting purposes. Extracted parameters are reported Table II.6.

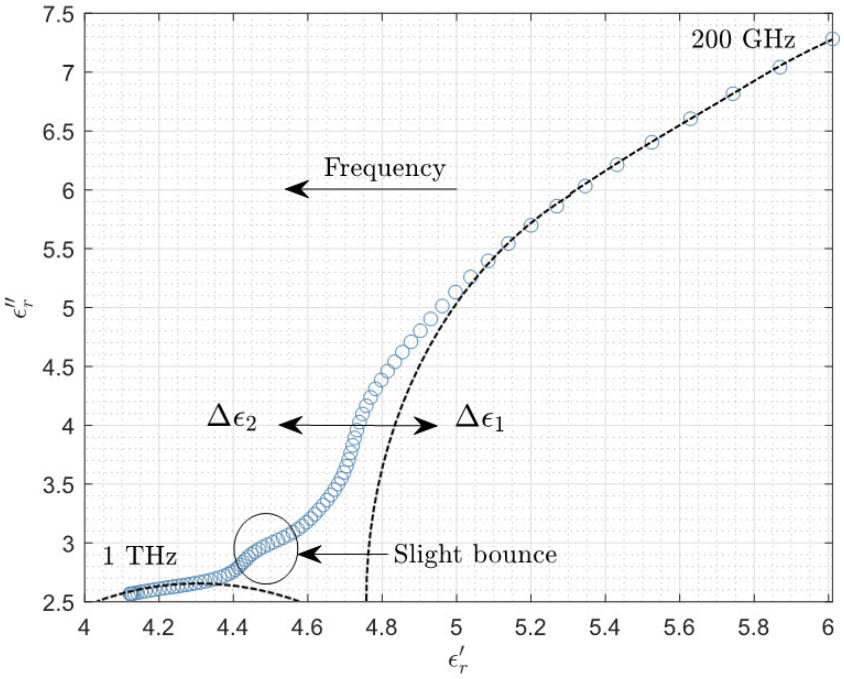


Figure II.25: Cole-Cole plot of water and approximated representation of characteristic relaxations domains. In black dashed lines are schematically represented the portion of the semi-circle that describes the double-Debye relaxation nature. Frequency axis decreases from 200 GHz to 1 THz from right to left.

On operating water data fitting, one can start by assessing the small quantity of data from low frequencies. As pointed out previously, the lack of information arising from sub-terahertz frequencies forces to fix the characteristic time of the first relaxation process τ_D from the microwave domain. To do so, we are pretty much spoilt for choice as reported in Table I.2,*Chapter I, I.3. Candidates for Interaction: Sub-cellular Level.* The relaxation time for the first process was selected from **[36]** since they provided the most recent study over a broad frequency band. Hence, τ_D is fixed and takes the value of 8.34 ps during the fitting procedure. In order to reject the spectroscopic aspect of the slight bounce, the fitting procedure was realized both for three-Debye terms and two-Debye terms, while keeping τ_D fixed.

Apart from the residual value, parameters extracted for the three-Debye terms are not consistent. Indeed, one can see that the non-linear solver returns null values for both the second and the third permittivity step amplitude. In other words, this means that the dielectric behavior of water could be described by an unique relaxation which is inconsistent with the previous observations thanks to the Cole-Cole representation. Last, to ensure that the norm of residuals is not over weighted by specific residuals at given ω in the complex plane, ω-dependent residuals have been investigated and are plotted in Figure II.26

Clearly, residuals from the 3-Debye terms are higher than the ones from the 2-Debye terms for real and imaginary residues. Smallest real residuals belong to low frequencies while smallest

Table II.6: Fitting parameters for 2-Debye terms and 3-Debye terms.

	2-Debye terms	3-Debye terms
$\Delta\epsilon_1$	73.19	85.44
$\tau_{D}[ps]$	8.34	8.34
$\Delta\epsilon_{r2}$	2.24	0
$\tau_2[ps]$	0.10	48.17
$\Delta\epsilon_{r3}$	-	0
$\tau_3[ps]$	-	29.75
ϵ_∞	2.66	4.57
|Res|	0.65	3.77

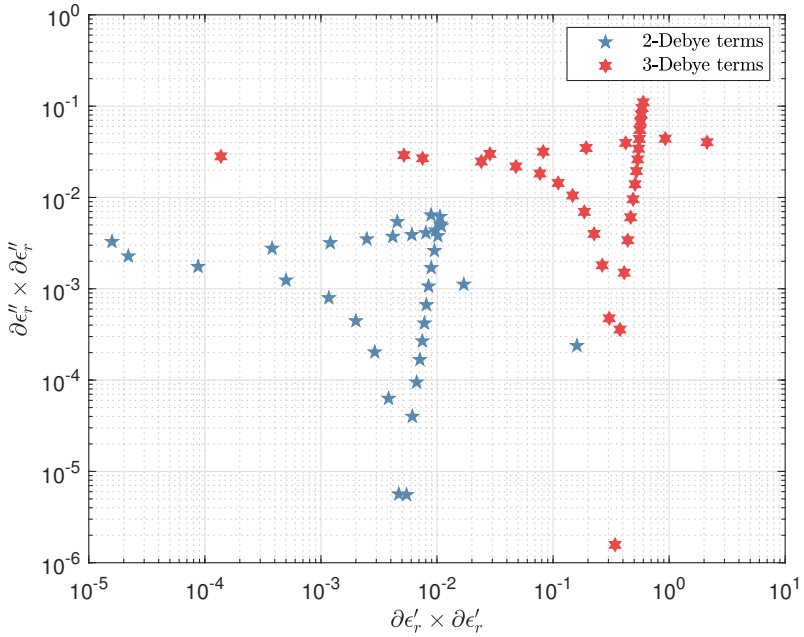


Figure II.26: Residuals at each ω with respect to both the real and complex plane for 2-Debye terms and 3-Debye terms.

imaginary ones are found at high frequencies.

In view of all these examinations, it is reasonable to reject the presence of a third relaxation in the examined frequency band. On the other hand, the presence of two overlapping relaxations is therefore validated and is in line with the literature. While the majority of the fitting parameters are in the range of reported ones, one may point out the discrepancy observed on τ_2. Although we still can state that throughout the literature the reported values for the second relaxation time are varying over a large range spanning from 0.17 to 1.02 ps, such a difference finds likely a physical explanation.

First, previous investigations were performed in transmission configuration. In other words, the terahertz beam was passing through the entire sample, carrying thus a greater amount of information about the sample than the one carried in reflection geometry. In fact, in reflection geometry, it is quite obvious that the information reported through the back propagation is only due to a surface interaction that is large by about the beam diameter.

Second the small value of τ_2 is probably arising from the spectral band over which the objective function minimization is performed. In fact, considering the fitting available spectral band ($[10^{-4} - 2.5]$ THz) in **[36]**, one can adequately state that here, the available band is quite reduced. Nonetheless, although the available spectral band seems to be of crucial importance when fitting the dielectric properties, one has still to understand why. The most reasonable hypothesis is that

due to the limited band, the second relaxation is not fully monitored by the experimental system or blurred due to low SNR. That is, only a portion of the relaxation is available and consequently the characteristic frequency is not well-defined. Hence, the solver tries to compensate the lack of information at higher frequencies. The discrepancy in τ_2 does not reflect an inherent shortcoming in the previous work, but rather its limited spectral range. It is worth mentioning that a similar value for τ_2, than the one reported here, was found in **[37]** for a spectral band ranging from 200 GHz to 1 THz and in reflection geometry.

On the physical sense of each of the characteristic frequency, the first one is assumed to correspond to the reorientation of the hydrogen-bond network while the second one to the relaxation of free and quasi-free water molecules. Further considerations were addressed previously in *Chapter I, I.3. Candidates for Interaction: Sub-cellular Level.*

On plotting the numerical solution with respect to the experimental $n(\omega)$ and $\kappa(\omega)$, one can appreciate the excellent fit shown in Figure II.27.

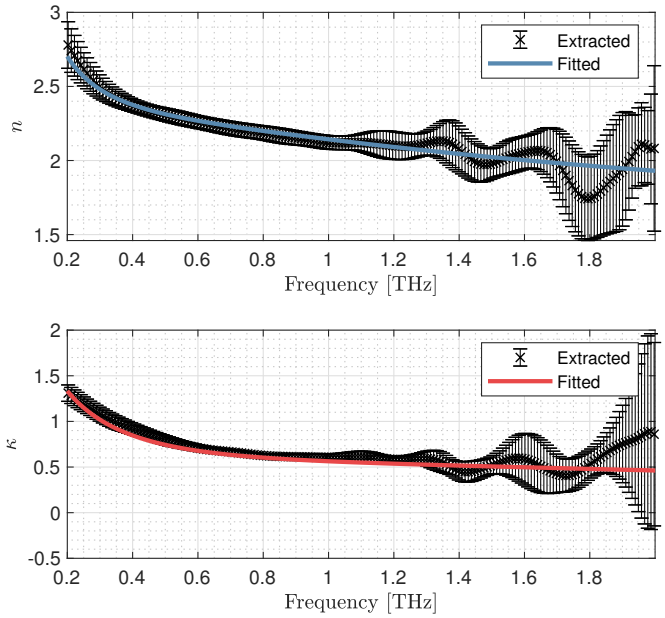


Figure II.27: Fit to 2-Debye terms model and the experimental $n(\omega)$ and $\kappa(\omega)$.

The extraction of the fit-parameters correlated to the two polarization mechanisms involved in the relaxation of liquid water was successfully conducted. Despite discrepancies observed in the characteristic time of the high frequency relaxation, assumed to be due to the reduced available frequency range in reflection geometry, the global water behavior was well observed and quantified. One may now go a little further toward breast cancer assessment.

II.5 Phantoms

The environment of the actual woman breast to be encountered in pre-clinical studies is complex. Additionally, it is impossible to concomitantly address the dielectric specificity of a fresh tissue and to perform a battery of tests to determine its constituent fractions. Therefore, it is of paramount importance to investigate on samples that were specially designed to mimic the theoretical dielectric response of human tissues. As breast tissues are heterogeneous volumes whose elements exhibit dielectric characteristic spanning over the biological spectrum, mimicking-tissue requirements are of generic nature. One can for instance report the necessity to produce high-water-content samples to imitate cancer lesions and fibroglandular tissues, while adipose tissues would be mimicked by low-water-content specimen. Preliminary studies on anthropomorphic

phantoms are of great interest since it would provide some roots for the comprehension of actual tissues and terahertz fields interaction.

To this aim, an impressive work was accomplished on three-component phantoms by Caroline Reid in 2009, in her thesis manuscript **[37]**. The procedure for producing the three-component phantoms that were analyzed in her work is described in **[38]**. To mimic the structure tissues, phantoms were made of lipid emulsion, gelatin and water. She investigated both the transmission and reflection configuration and reported more accurate results in the first mode. It was also reported that the accuracy was lower than the one of less complex structures such as two-component phantoms. The author assumed that the decrease in accuracy could be due to more complex bonding within the phantoms. In fact, a greater number of variables being determined from the standard deviation measurements of the experimental system inherently induces poorer results.

II.5.1 Extracted Intrinsic Properties

Despite the difficulties reported by Reid, some investigations in reflection geometry towards three-components phantoms were conducted to face previously stated issues. The aim behind these investigations is to prepare the ground for further pre-clinical studies. To mimic the complexity of breast volume, sunflower oil, isolated soya protein and liquid water were used **[39; 40]**. Oil, proteins and water were homogenized to make an emulsion. Then, the phantoms were deposited into vacuum packages. Five different three-component phantoms with varying water-content were produced to consider the generic nature of the breast volume. Content in percent-by-weight (% wt) within each samples are given in Table II.7.

Table II.7: Fraction of protein, oil and water in each phantom in percent-by-weight.

# Phantom	Protein (% wt)	Oil (% wt)	Water (% wt)
# 1	13	17	70
# 2	13	19.5	67.5
# 3	13	24.5	62.5
# 4	13	27	60
# 5	25	75	0

Each of these samples were probed on a C-cut sapphire substrate by means of a terahertz time domain spectrometer in nitrogen controlled-atmosphere. Phantoms were measured five times in a row, whilst removing the sample from the sapphire top surface between each. Measurements were averaged a hundred times. The typical standard deviation between 200 GHz and 1 THz was by about 5% while at higher frequencies, data are not reliable due to the low signal-to-noise ratio. The extracted frequency dependent refractive index n and κ for each phantoms are shown in Figure II.28.

The first thing that is obvious is that the water-free phantom (#5) is totally aside from others. Indeed, both the refractive index and the extinction coefficient are much more smaller than their counterparts, *i.e.* than water-made phantoms. Another interesting observation can be drawn on both the refractive index and the extinction coefficient at low frequencies, namely between 200 GHz and 500 GHz. In fact, while the water-free phantom exhibits positive or null slopes on its optical properties, water-made samples display negative slopes at low frequencies. Such a behavior can be physically explained by the presence of hydrogen-bonded water molecules within the first cluster of phantoms. Hence, at low frequencies, water-made samples exhibit the characteristic relaxation of the bulk water network while on phantom #5 such a behavior is absent. Although within the actual breast volume, adipose tissues are not fully deprived from water, their water-content is quite low. In a nutshell, adipose tissues should be highly distinctive tissues among the entire tissue matrix. Moreover, in *Chapter I. Section I.4 Breast Cancer*, we stated that in the specific case of invasive breast carcinoma, cancer cells were found to invade the connective tissues that

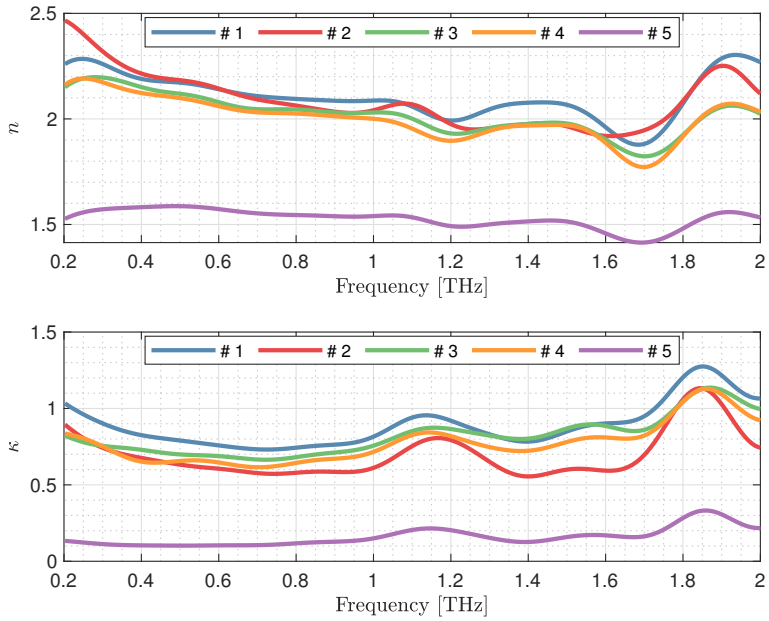


Figure II.28: Frequency dependence of the refractive index and the extinction coefficient of each samples.

surround the initial lesion site, including adipose tissues. Hence, for such a pathology, the water-content within adipose matrix may increase due to the presence of invading cancer cells bringing within their membrane additional water molecules. It is therefore reasonable to state that, adipose tissues whose optical properties exhibit specific trends corresponding to a significant presence of water may reflect an invasive breast carcinoma, either of lobular or ductal etiology.

Considering now the optical property behaviors of water-made phantoms, still in the trustworthy spectral range spanning from 200 GHz to 1 THz, one can observe their similar behavior: the refractive index decreases from low to high frequencies. Additionally, the extinction coefficient decreases from 200 GHz to roughly 600 GHz. These two considerations are in line with the behavior that was observed previously for liquid water.

When assessing the discrimination among the water-made phantom set, one can easily point out their similitude. The discrepancy between each sample is quite small and difficult to appreciate. Let's however notice that phantoms #1, #3 and #4 are in line with expectations. In fact, the higher water-content is expected to give arise to both the greater refractive index and the greater extinction coefficient. Expectations are fulfilled in the sense that samples are sorted from top to bottom from higher water-content to lower water-content. However, as stated earlier the standard deviation of these measurements was by about 5% which induces that the samples made of 62.5% and 60% of water are not reliably distinguishable. Nonetheless, one can, on the contrary, underline a significant discrimination when considering the 70% and the 60% water-content samples, out of the deviation range.

Last but not least, the phantom #2 does not respond to what was expected there. In fact, its refractive index exhibits some specific trends which are not in line with others. Moreover, such a refractive index is higher than the one extracted for phantom #1 made of a greater water-content. Concomitantly, the extinction coefficient derived from the measurement on sample #2 is clearly lower than others, which basically makes no sense. Therefore, it is worth to ask oneself what could give arise to such a tendency?

By thinking on the possible reasons, one may list: the different sources of noise previously mentioned; The inaccurate positioning between the reference signal and the sample one, inducing amplitude and phase discrepancy between the two signals; Intrinsic alteration within the phantom itself. While the first statement toward noise can be corrected using proper precautions, the second and last ones are more tedious. In fact, by replacing successively the sample on the top

surface of the sapphire substrate one may have, at a moment, slightly tilted the sapphire from its initial position. Consequently the position of the beam waist is modified and the beam focus at the photoconductive antenna is shifted from the referent location. By doing so, the extinction coefficient may decrease in a greater way than the one stipulated by its intrinsic properties. Considering now the greater refractive index of phantom #2, the modification of the optical path reflects in faster or smaller propagation velocity, inducing inconsistent extraction toward the refractive index. It is hence reasonable to assume that results on phantom #2 are inconsistent and reflect a shortcoming and a limitation in the experimental setup used heretofore. On the other side, we mentioned previously the intrinsic alteration of the sample itself. Effectively, let's remember that our phantoms were made of water, oil and proteins. Although attempting to homogenize the emulsions, water and oil simply do not mix properly together due to their respective polarities. As oil is lighter than water, it is unfortunately possible than it tends to move to the surface, denaturing the wished nature of the sample.

From samples exhibiting consistent optical behavior, namely #1, #3, #4 and #5, one may now try to fit their respective dielectric properties by deriving them first from the extracted optical components and subsequently find out the underlying physical parameters. It is worth noting that as these samples have been specifically designed and produced for the aim of this study, information concerning their dielectric behavior in the microwave region are therefore not available. Consequently, all the fitting parameters are free to be adjusted during the procedure. The task is particularly tedious as models and poles to be used to fit the data are unknown. Indeed, the different constituents may interact with each others, giving arise to specific relaxations which may not be simply exponential like Debye relaxations. Moreover the complexity increases proportionally to the number of constituent.

II.5.2 Relaxation Mechanisms

Multiple combinations of Debye, Cole-Cole, Davidson-Cole and Havriliak-Negami poles were successively tested. Nonetheless, none of the considered forms was suitable to describe the behavior of both the real and the imaginary part of the complex permittivity. One has still to determine a theoretical model to describe the behavior of the permittivity in the complex plane. According to Figure II.28, the most noticeable feature is the peak located around 270 GHz for water-made phantoms. A comparable increase is spotted for the free-water phantom around 492 GHz. The presence of these peaks was reported in the literature for real breast tissues but their origin is yet to be understood [**41**]. However, the increase of the permittivity in the low frequency range does not lead to a Debye response or any polynomial models. Throughout the literature, one can found Debye-like fittings on tissues and phantoms that match well the data. However, the fitting procedure is generally conducted from 400-500 GHz and hence, does not take into account the low frequency permittivity increase. Alternatively, one can assess the dielectric function behavior in frequency subranges and merge subsequently the results. However, such attempts are often cumbersome and inefficient.

In this work, an other approach is developed that allows to fit the biological data over the entire considered spectral band. Note that the model to be described was only tested in the limited terahertz bandwidth and does not aim to simulate the dielectric function over the entire electromagnetic spectrum.

As it was described in the first chapter, the dielectric function over a given spectral band can be approximated as a combination of dielectric mechanisms. The dielectric function can therefore be simulated through the generic shape of:

$$\hat{\epsilon}(\omega) = \epsilon_\infty + \sum_{m=1}^{M} \epsilon_m(\omega), \tag{II.30}$$

where ϵ_∞ is the permittivity at infinite frequency, and $\epsilon_m(\omega)$ is the dielectric mechanism of index m. Indeed, the dispersion of the dielectric function usually does not correspond to an unique

mechanism and therefore dielectric poles ϵ_m are distributed over the spectral band. The terms ϵ_m are of the form:

$$\epsilon_m(\omega) = \frac{\alpha_{m,0}}{\beta_{m,0} - i\omega\beta_{m,1} - \beta_{m,2}\omega^2}, \tag{II.31}$$

where α and β are coefficients that belong to $\mathbb{R}$ but are not necessary physical terms. However, such kind of formulation is quite constraining and is not always sufficient to properly match the data. The idea is therefore to provide more flexible pole. By introducing an additional term in the numerator of II.31 of the form $\alpha_{m,1}(-i\omega)$ and $-i\omega\beta_{m,1}$ in the denominator the model is overall more flexible. The introduced parameter in the numerator can be seen as a parameter depending on the time derivatives of the electromagnetic field interacting with the material. On the denominator, it corresponds to an electronic damping factor. Based on the foregoing, one can rewrite equation II.31 as:

$$\epsilon_m(\omega) = \frac{\alpha_{m,0} + \alpha_{m,1}(-i\omega)}{\beta_{m,0} - 2i\omega\beta_{m,1} - \beta_{m,2}\omega^2}, \tag{II.32}$$

we kept $\beta_{m,2}$ fixed to unity for simplicity and finally equation II.32 could be rewritten as:

$$\hat{\epsilon}(\omega) = \epsilon_\infty + \sum_{m=1}^{M} \frac{\alpha_{m,0} + \alpha_{m,1}(-i\omega)}{\beta_{m,0} - 2i\omega\beta_{m,1} - \omega^2}. \tag{II.33}$$

In the following, excluding ϵ_∞, the parameters are empirical and do not aim to underline the physical nature of samples. It is then worth to emphasize that these parameters should only be regarded as required parameters to obtain accurate fits. In its final shape, one can notice that the model described in equation II.33 is developed on the basis of Lorentz oscillators given in *I.3.2 Interaction Mechanisms*, equation I.28. Here, the Lorentz oscillators are aimed to compensate between each others to produce the dispersion profiles.

Remarkably, the dielectric response of phantoms was found to require four different poles (M = 4) while for the water-free phantom 3 poles (M = 3) were found to be sufficient. The parameters of the effective model previously described were therefore extracted and are reported for each phantoms in Table II.8 and the respective fitted dispersion profiles are fitted in Figure II.29.

Table II.8: Fitting parameters of each phantom with respect to equation II.33.

Phantom #1					Phantom #3				
$\epsilon_\infty = 2.69$	α_0	α_1	β_0	β_1	$\epsilon_\infty = 2.34$	α_0	α_1	β_0	β_1
$1^{st} - Pole$:	395.49	22.02	99.92	5.56	$1^{st} - Pole$:	1.08E3	149.65	100	13.91
$2^{nd} - Pole$:	119.67	64.75	9.28	5.02	$2^{nd} - Pole$:	108.30	29.91	14.51	4.00
$3^{rd} - Pole$:	3.93	2.83	0.04	0.03	$3^{rd} - Pole$:	3.66	~ 0	0.03	~ 0
$4^{th} - Pole$:	-40.23	-156.99	1.99	7.77	$4^{th} - Pole$:	-35.66	-272.18	5.94	45.32
	|Res|	0.4809				|Res|	0.3297		
Phantom #4					Phantom #5				
$\epsilon_\infty = 2.33$	α_0	α_1	β_0	β_1	$\epsilon_\infty = 1.85$	α_0	α_1	β_0	β_1
$1^{st} - Pole$:	257.73	1.22	99.72	0.47	$1^{st} - Pole$:	0.34	2.34	1.12	7.69
$2^{nd} - Pole$:	~ 0	11.21	~ 0	1.71	$2^{nd} - Pole$:	~ 0	6.44	~ 0	1.78
$3^{rd} - Pole$:	-14.62	-0.36	0.68	0.02	$3^{rd} - Pole$:	-0.08	-2.07	1.16	30.70
$4^{th} - Pole$:	-31.91	-39.30	2.95	3.63	$4^{th} - Pole$:				
	|Res|	0.6364				|Res|	0.2327		

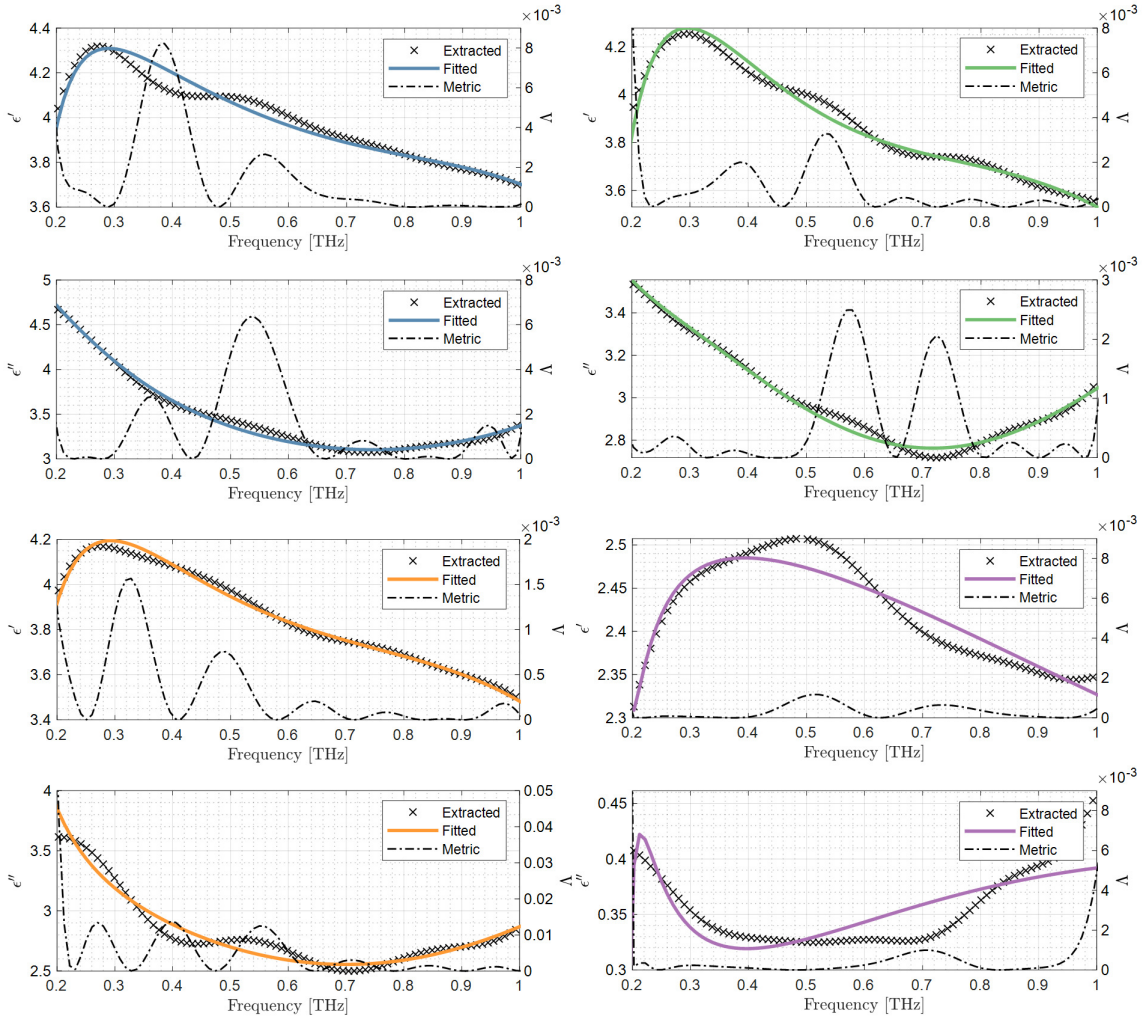


Figure II.29: Fitting of the dispersion profiles of each phantoms with respect to II.33. Blue: #1 phantom; Green: #3 phantom; Yellow: #4 phantom; Purple: #5 phantom. Black dashed line represents the Λ metric.

It is worth noting that as the model does not attempt to describe the physical properties of the samples, the parameters are not constrained during optimization. Concomitantly, one may have identified a drastic shortcoming in the model. In fact, it seems quite obvious that II.32 is hardly convex and therefore the minimization of the corresponding objective function would be very challenging for the standard optimization solver *fminsearch* without falling into saddle points. Such an issue was not treated in depth during this work. However, for suggestions, one may explore the structures of II.33 that are partly convex by the deriving the Hessian of the problem function ϕ and verifying analytically the following statement **[42]**:

$$\phi'' \geq 0. \tag{II.34}$$

One would also probably want to keep held $\beta_{m,0}$ and $\beta_{m,1}$ in a first instance and reexpress the problem in a more computational friendly fashion. It is aimed that overcoming the formulation of the model, it may provide accurate parameters that should reflect in a certain way the physical properties of biological materials.

Nevertheless, one can appreciate the accuracy of the model. It is certain that although being still at the development stage due to its formulation, the model may lead with further work to a robust theoretical basis toward the simulation of the dispersion profiles of anthropomorphic samples.

Additionally, it seems important to emphasize that the imaginary part of the dielectric function that was extracted for phantoms exhibit a particularly strange behavior. In fact, ϵ'' exhibit a

strong convex shape and is therefore barely consistent with the theoretical biological spectrum. In fact, an ϵ'' convex function would reflect the presence of two characteristic frequencies on both side of the minimum of ϵ''. However, we largely discussed on the fact that due to the complexity of biological tissues, the dielectric function was expected to be frequency-dependent in the terahertz range with no specific features. As phantoms aim to mimic tissue behaviors, the convex nature of ϵ'' is therefore inconsistent. Furthermore, such a frequency dependence is not in line with reported data throughout the literature. On the contrary, the real part ϵ' behaves similarly to those reported results. Roots for such an observation are probably related to the extracted optical parameter κ which is, in reflection configuration, less reliable than n. Indeed, as the focus terahertz beam only interacts with phantoms at the dielectric interface $sapphire \longrightarrow phantom$, it does not propagate within the material and the recorded signal is therefore mainly dictated by the refractive index difference between sapphire and sample. Hence, it is reasonable to assume that the small quantity of information carried about the extinction coefficient κ is significantly affected by the noise and may turn to be less solid. In view of these considerations on the doubtful reliability of κ it seems interesting to perform the optimization only on the real part of the permittivity. Such an action is supported by the fact that in the spectral range lying between 200 GHz and 500 GHz, the weak value in κ as a poor impact on ϵ' regarding the impact of n. Indeed, let's remind that $\epsilon'(\omega) = n^2 - \kappa^2$, hence on comparing n versus κ it is reasonable to state that ϵ' reflects most strongly the dispersion in n than in κ. The idea is then to perform the objective function minimization only toward the real part of residuals by modifying the definition II.29 respectively as:

$$|Res| = \sqrt{\sum_{\omega=0}^{\omega_{max}} \partial\epsilon'_r(\omega)^2}. \tag{II.35}$$

On taking the mathematical formulation of the model in II.33, one can see that it can be rewritten as a function of its real and imaginary part as:

$$Re\ \hat{\epsilon}(\omega) = \frac{\alpha_0(\beta_0 - \omega^2) + 2\alpha_1\omega^2\beta_1}{(\beta_0 - \omega^2)^2 + 4\omega^2\beta_1^2}, \tag{II.36a}$$

$$Im\ \hat{\epsilon}(\omega) = \frac{-\alpha_1\omega(\beta_0 - \omega^2) + 2\alpha_0\omega\beta_1}{(\beta_0 - \omega^2)^2 + 4\omega^2\beta_1^2}\, i. \tag{II.36b}$$

Therefore, the optimization will be proceed toward the definition in II.36a. Ultimately, on optimizing only on the real part of the permittivity, the sufficient number of poles to accurately fit the dispersion profiles was reduced to three (M = 3) for all phantoms. The extracted coefficient are given in Table II.9. The overall real part of the dielectric functions are better fitted than previously (Figure II.30). Such a trend is not surprising since the optimization is not performed in the complex plane. Therefore, it is easy to understand that parameters are slightly less constrained. However, it is worth mentioning that a fit that looks better is not obviously a better fit. Let's keep in mind the deliberate choice of assuming the weak contribution of κ to ϵ' and to shelve the optimization on the complex plane.

The effective model employed to simulate the response of the real part of the complex permittivity mimics very well the dielectric behavior of phantoms whereas even large number of other relaxation terms were inadequate and consistently failed to assess the low frequency permittivity increase. Interestingly, the norm of residuals increases as a function of water-content which may suggest that greater water content requires additional dielectric poles or supplemental $(-i\omega)^n$ terms. Nonetheless the accuracy of the fit, with respect to residuals, is clearly sufficient and further considerations are not substantial. However, despite the accuracy of the fitting, one has first to ensure that sensed phantoms are decently mimicking real breast tissues before any further steps. To this aim, a comparison between the model proposed by Truong *et al* in 2015 derived from a wide measurement campaign on 74 breast samples and the model employed here has to be performed using the $\Lambda(f)$ metric **[43]**. Let's underline that such a comparison remains purely

Table II.9: Fitting parameters of each phantom for (II.36a).

Phantom #1					Phantom #3				
$\epsilon_\infty = 3.35$	α_0	α_1	β_0	β_1	$\epsilon_\infty = 3.28$	α_0	α_1	β_0	β_1
$1^{st} - Pole:$	7.48	13.43	1.49	2.68	$1^{st} - Pole:$	2.34	3.70	1.21	1.91
$2^{nd} - Pole:$	-3.06	-10.23	1.29	4.31	$2^{nd} - Pole:$	0.09	~ 0	4.58	~ 0
$3^{rd} - Pole:$	-20.83	-0.12	3.54	0.02	$3^{rd} - Pole:$	-3.34	-0.52	1.50	0.23
	|Res|	0.2290				|Res|	0.1651		
Phantom #4					**Phantom #5**				
$\epsilon_\infty = 3.27$	α_0	α_1	β_0	β_1	$\epsilon_\infty = 2.24$	α_0	α_1	β_0	β_1
$1^{st} - Pole:$	1.30	5.57	0.50	2.14	$1^{st} - Pole:$	0.13	0.71	0.64	3.57
$2^{nd} - Pole:$	-0.13	-1.74	3.73	49.99	$2^{nd} - Pole:$	-0.38	-1E-3	7.34	0.02
$3^{rd} - Pole:$	-10.01	-1.49	1.95	0.29	$3^{rd} - Pole:$	-2.47	-0.05	2.46	0.05
	|Res|	0.1319				|Res|	0.0569		

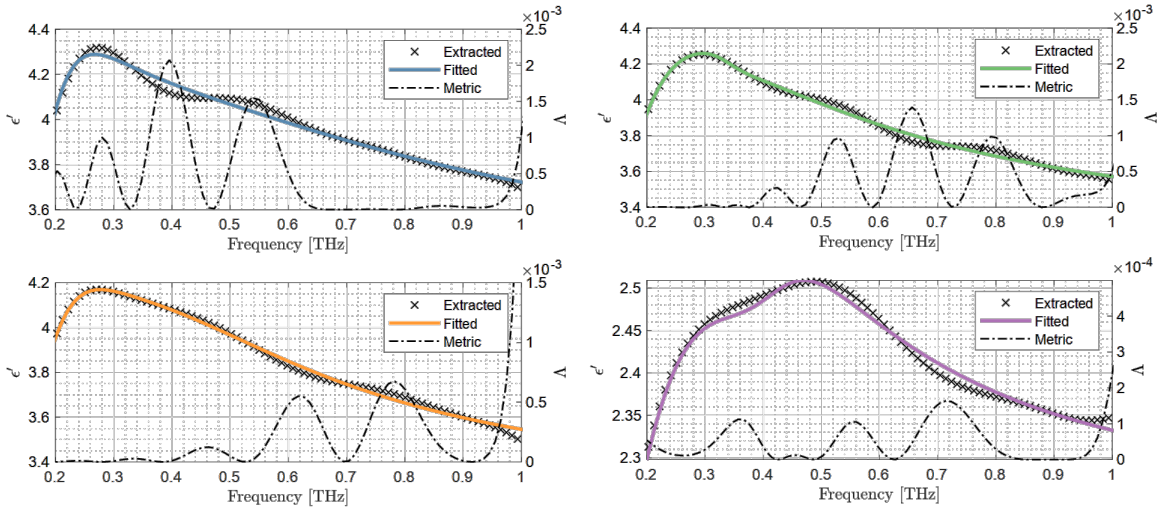


Figure II.30: Fitting of the real part of the dispersion profiles of each phantoms. Blue: #1 phantom; Green: #3 phantom; Yellow: #4 phantom; Purple: #5 phantom. Black dashed line represents the Λ metric.

qualitative since water, fat and protein fractions in fresh samples were not accessible. One can simply assume that phantoms #1, #3 and #4 should exhibit a dielectric behavior roughly similar to healthy fibroglandular and malignant tissues while sample #5 is expected to stay in line with adipose tissue. In their paper, Truong *et al* considered an empirical model based on the combination of non-Debye and Debye relaxations (equation II.37).

$$\hat{\epsilon}_r(\omega) = \epsilon_\infty + \frac{\omega\tau_1\Delta\epsilon_1 + \Delta\epsilon_2}{1 + (i\omega\tau_1)^\alpha} + \frac{\Delta\epsilon_3}{1 + i\omega\tau_2} + \frac{\sigma}{i\omega}, \tag{II.37}$$

here σ corresponds to the impact of the conductivity on the dielectric loss within tissues. The authors emphasized that $\Delta\epsilon_1$ and $\Delta\epsilon_2$ are more empirical parameters rather than being of physical nature and should therefore be considered as required parameters to obtain adequate fits. Extracted parameters from their study are reported in Table II.10.

The $\Lambda(f)$ metric aiming to compare the two model is given in Figure II.31.On reducing the water content among water-made samples, the dielectric function tends to be closer to that of normal tissues reported by Truong *et al*. On the other hand, the greater the water-content, the closer to cancer behavior it is. Concerning the adipose model and the one for water-free phantom,

Table II.10: Parameters of the model proposed by Truong *et al.* aR^2 is the adjusted R-square value denoting the square of the correlation between the response values and the simulated ones. When equal to unity, the adjusted R-square value denotes a perfect correlation while being equal to 0 it denotes a null correlation. BCC: basal cell carcinoma.

	ϵ_∞	$\Delta\epsilon_1$	$\Delta\epsilon_2$	$\Delta\epsilon_3$	σ	$\tau_1(ps)$	$\tau_2(ps)$	α	aR^2
Adipose	2.50	3.54	-1.89	0.47	0.39	9.67	0.13	1.49	0.996
Normal	3.34	1.33	-3.05	0.71	5.27	1.91	0.11	1.20	0.998
BCC	2.86	28.64	-3.45	1.59	8.33	3.58	0.10	2.91	0.997

both are quite similar as depicted by the metric. Overall, one can appreciate the correspondence between the two models even though, and unsurprisingly the two models do not match perfectly since compound fractions are not identical. Moreover, phantoms just aim to approximate tissue trends. In that sense, it is reasonable to confirm their anthropomorphic nature. Nevertheless one still has draw some interesting features belonging to mimicking breast tissue phantoms which are of cardinal importance to address the up-coming pre-clinical study on freshly excised breast tissues.

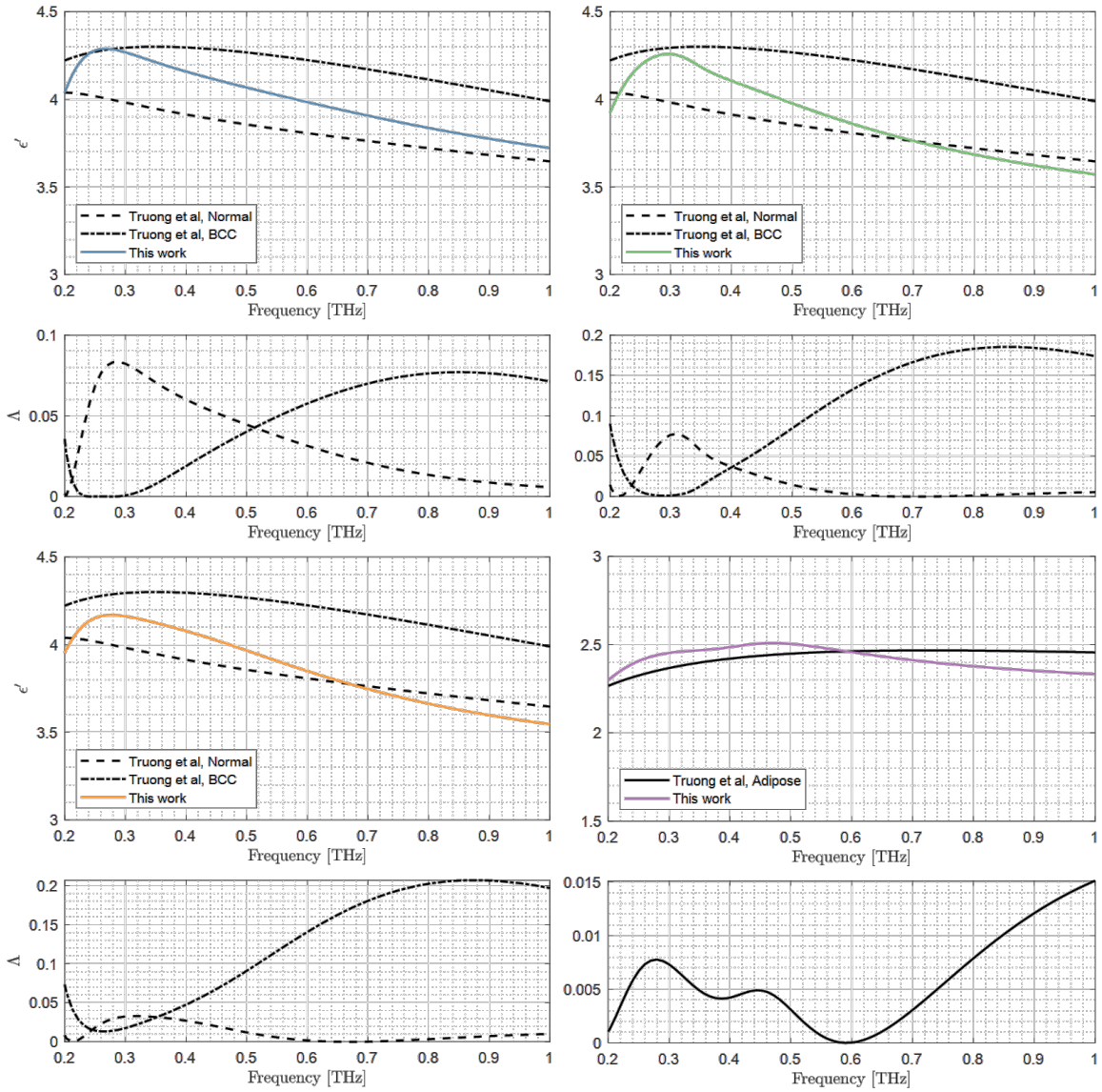


Figure II.31: Λ metric on the comparison between Truong *et al* model and the modified-Lorentz one for phantoms. Blue: #1 phantom; Green: #3 phantom; Yellow: #4 phantom; Purple: #5 phantom. Black dashed line represents the Λ metric.

II.6 Excised Breast Tissues

PRELIMINARY ETHICAL APPROVAL: HUMAN TISSUE ANALYSIS HAVE BEEN PERFORMED IN VIEW OF THE FUNDAMENTAL ETHICAL STIPULATED IN THE HELSINKI DECLARATION AND ITS LATER REVISIONS. SAMPLES WERE OBTAINED WITH THE WRITTEN APPROVAL OF EACH PATIENT UNDERGOING AN EXCISION PROCEDURE.

Investigations toward freshly excised breast tissues were conducted at the histology department of the Bergonié Institute, the oncology laboratory partner. To this aim, a TPS4000 was moved to the hospital. In this work, sixteen different excised tissues from eleven non-consecutive patients were used to collect a total of forty-eight waveforms, each being averaged over one hundred measurements. Excised tissue classification and preparation is described in the following section.

II.6.1 Tissue Registration & Preparation

To correctly assess variation across and within freshly excised breast samples, a classification of tissues regarding their biological compositions is obviously of capital importance. Each sample

was catalogued to belong to one or more of the subsequent groups: (A), (N) and (C), denoting respectively the presence of adipose, normal and cancerous tissues. Unless advised by a doctor, fibrous waveforms from breast conserving surgery were consistently put into the malignant category. Table II.11 shows the data classification among samples. Five samples were invasive lobular carcinoma, six samples were invasive ductal carcinoma, two were metaplastic breast cancer, one was breast phyllodes tumor and two were excised from a breast reduction. Measurements were performed on samples that have not been submitted to any chemical treatments. After surgery, the breast excision was carried to the histology room to be cut in slices of few millimeter thick and small enough to sit on the top of the sapphire substrate. To ensure a good contact with the sapphire window, sliced tissues were slightly pressed with another sapphire substrate on the top. A picture of sandwiched tissues (I.I and K.I samples) is shown in Figure II.32. A maximum of two hours elapsed between the excision procedure in the operating room and the acquisition, to ensure no drastic degradation of fresh state tissues. Samples that did not fulfill such a requirement were removed from the initial database. From each sample, several locations were probed and the corresponding waveform was recorded. After being measured, tissues were fixed in a formalin solution before conducting the classical histology hematoxylin and eosin stain process.

Table II.11: List of the fifteen freshly excised tissues analyzed in reflection spectroscopy. Samples having the same letter reference code belong to the same patient and have been excised during the same procedure. The diagnosis column refers to the pathology from which is suffering the patient. ILC: invasive lobular carcinoma; IDC: invasive ductal carcinoma; MBC: metaplastic breast cancer; BPT: breast phyllodes tumor; BR: breast reduction.

Patient	Sample	Classification	Diagnosis	Stade	Age
A	I.	(C)	ILC	2	74
-	II.	(C)	-	-	-
B	I.	(C)	IDC	3	82
-	II.	(A);(N)	-	-	-
C	I.	(C)	MBC	3	63
-	II.	(A)	-	-	-
D	I.	(C)	IDC	3	66
E	I.	(C)	IDC	2	80
F	I.	(C)	ILC	2	73
G	I.	(C)	IDC	2	55
H	I.	(C)	BPT	2	63
I	I.	(A);(N)	BR	-	48
-	II.	(A);(N)	-	-	-
J	I.	(A);(C)	ILC	3	82
K	I.	(C)	IDC	3	84
Histological distribution among the waveform set					
(A)		(N)		(C)	
11		11		26	

II.6.2 Extracted Intrinsic Properties

The physical properties lying behind each recorded waveform contained in the database were extracted. From the individual properties, the mean as well as the standard deviation were derived. The real and the imaginary part of the complex refractive index for each category are shown in Figure II.33.

The extracted intrinsic properties are generically in good agreement with other published data. Especially with the reference paper published by Ashworth *et al* **[41]**. However, globally the three

Figure II.32: I.I and K.I samples sandwiched between two sapphire substrates to prepare the tissue measurement.

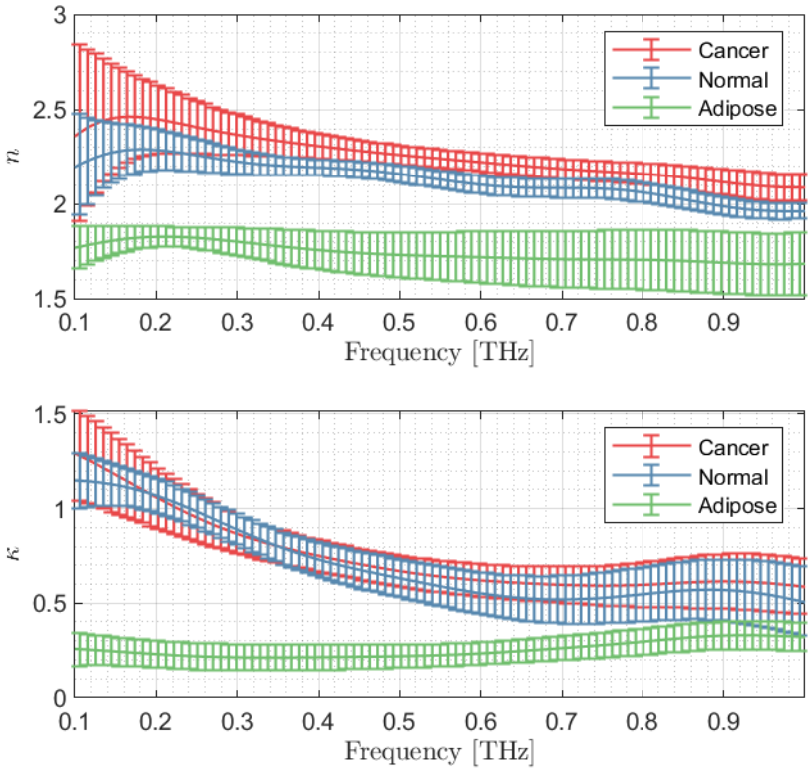


Figure II.33: Refractive index and extinction coefficient for malignant lesions, healthy fibrous and adipose tissues.

types of tissues exhibit -in this work- slightly greater real and imaginary parts. This is assumed to be due to the fact that in their paper, Ashworth and colleagues have frozen tissues before measurement to prevent from water release during acquisition. On our part, tissues were not frozen and were analyzed directly after being excised. The contributed surplus likely results from the excess of water within freshly excised tissues.

The error bars are globally larger for cancer and healthy fibrous at low frequencies. Such a trend is supposed to be due to the concomitant effect of both the strong terahertz attenuation within these materials and the low signal-to-noise ratio located at these frequencies.

The attenuation reflects both the extinction, globally due to generalized thermal effect, and

the internal scattering[6]. However, from the measurements that have been done within this work, it is clearly impossible to state on the impact of each of those. Nevertheless, the high water content found in freshly excised tissues leads us to believe that extinction is the major contribution toward attenuation.

A particularly interesting feature is exhibited by all tissue class. Although being already reported in the section on phantoms, the refractive index consistently shows an increase in the low frequency range before decreasing. The presence of this speak is not fully understood. Nevertheless, it is still possible to draw some assumptions. Such a feature is clearly absent from the dispersion profiles that were reported before for water. Therefore, it is reasonable to state that such trend is not solely correlated to water-content nor malignancy since the three groups exhibit it. However, each of these tissue category contains a certain amount of protein that may on interacting with bulk water molecules, produce the peak. Let's note that the refractive index increase is slightly shifted to lower frequencies in comparison to that of phantom mimicking tissues.

The overall tendency clearly indicates that adipose tissues are distinguishable from the two other types. Indeed, both the real and imaginary part of the optical function are consistently lower than those of healthy fibrous and lesions over the entire considered spectral range. The first reason that comes to mind for this specific behavior is that the overall water-content is drastically lower in adipose tissue than in others. Although being lower in term of optical properties than other tissues, the set of adipose tissue is also greater than the water-free phantom that was analyzed in the previous section, supporting the idea that water plays a substantial role. To illustrate the statement toward the discrimination of adipose regarding other tissue types, Figure II.34 depicts the refractive index and extinction coefficient difference as a percentage of the healthy fibrous properties. Fibrous tissues have been preferred over cancerous tissues as the mean complex refractive index of healthy fibrous tissues is globally lower than cancerous lesions. The error propagation was defined by logarithmic differentiation and was found to be:

$$\Delta(\Delta_{n,k}) = \Delta_{n,k}\left(\frac{\Delta_{n,k_{fibrous}} + \Delta_{n,k_{adipose}}}{n,k_{fibrous} - n,k_{adipose}} + \frac{\Delta_{n,k_{fibrous}}}{n,k_{fibrous}}\right). \tag{II.38}$$

Differences toward the two optical parameters are clearly located in the low part of the spectral band. On the contrary when looking at differences at higher frequencies, the gaps between the two groups decreases.

On the refractive index difference, one remarks that the difference between the refractive index of adipose tissues and the one of healthy fibrous tissues exhibits some fluctuations instead of being monotonous. These fluctuations are not attributed to structural features but rather to arise from the slight oscillations observed in the dispersion profile of the refractive index of normal tissues. Between 200 GHz and 900 GHz, the refractive index difference is in average ranging from 18% to 20% of fibrous real optical function. On considering the limits of distinction, the refractive index difference is at least of 10% in the 200 - 600 GHz window for the entire set of samples. Concomitantly, beyond 600 GHz, the two groups of healthy freshly excised samples are exhibiting a refractive index difference of at least 4%.

The discrimination between the two benign healthy groups is even more appreciable on comparing the extinction coefficient. Indeed, between 200 GHz and 600 GHz, the extinction difference is not less than 15% for the entire set. On the contrary, at higher frequencies the distinction is clearly not reliable.

Therefore, it seems interesting to underline that frequencies for discrimination between adipose tissues and other types are not appreciable on the same level. Particularly, frequencies ranging from 200 GHz to 600 GHz seem to provide clearer classification between adipose tissue and fibrous/cancer. Furthermore the intrinsic extinction is -in this range- a more efficient classifier. However, we previously stated that the refractive index is, in reflection geometry, more reliable than the extinction coefficient. Hence, on an advisory basis, it would be recommended to always

[6]The contribution of scattering to attenuation is however expected to be relatively small in comparison to extinction.

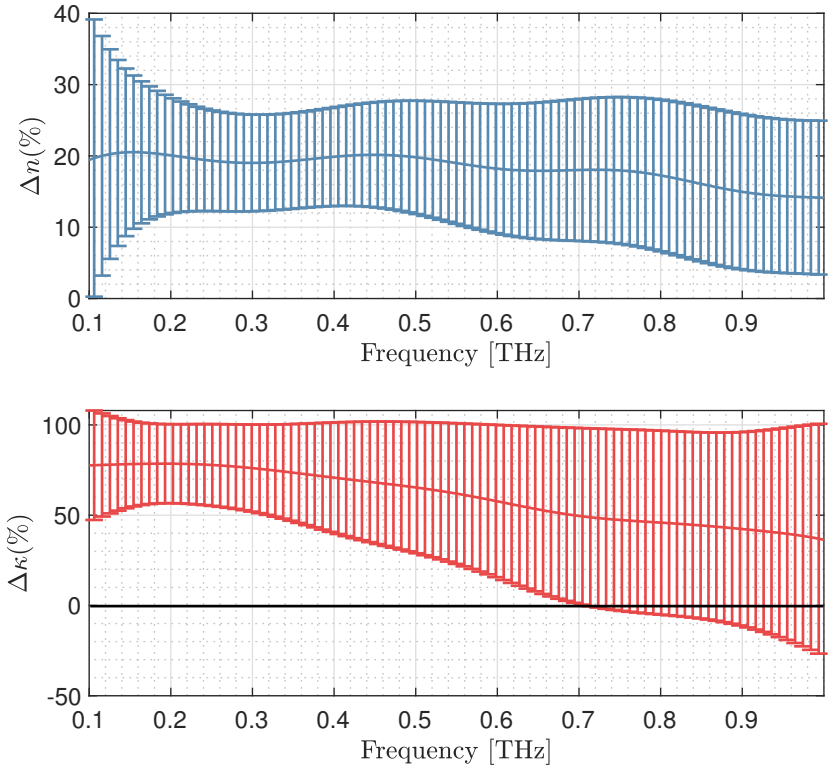


Figure II.34: Refractive index and extinction coefficient difference between adipose and healthy fibrous as a percentage of healthy fibrous characteristics.

check both the refractive and extinction index instead of working only with κ to discriminate adipose tissue from other tissue types.

In view of the foregoing, it seems reasonable to state that the use of terahertz spectroscopy allows one to discriminate adipose tissues that are located within the breast volume from other connective tissues.

It is nonetheless apparently more tricky to distinguish between healthy fibrous and cancer. The refractive index of cancer tissues and healthy fibrous are in average slightly distinguishable. However, it is clearly obvious that differences are not so clear and that their respective error bars slightly overlap at low frequencies. In Figure II.35, the optical property differences between normal and cancer tissues as a percentage of cancer characteristics are plotted. The error propagation was calculated in a similar way as defined in equation II.38.

Considering the refractive index difference, below 300 GHz as the errorbars overlap (Figure II.33), it is obvious that no distinction can be made between the two groups. It is hence reasonable to state that the biological structure of both healthy fibrous and cancer are quite similar and the signal-to-noise ratio does not allow to discriminate benign from malignant tissues in the low frequency range. After roughly 300 GHz, the errorbars do not overlap as depicted by the positiveness of the refractive index difference. Although the respective refractive indices do not overlap with each other, their demarcation is relatively weak. Indeed, over the set of sample that was analyzed, the refractive index difference between healthy and malignant tissues varies from 0% to 15%. Therefore, it is not possible, from a statistical point of view, to consistently distinguish with certitude healthy from malignant tissues.

On observing the behavior of the extinction coefficient, it is obvious that no distinction are feasible from the study of this property. In fact, the respective confidence intervals overlap over the entire band.

On comparing these results to the work of Ashworth *et al*, one can ask oneself why each study

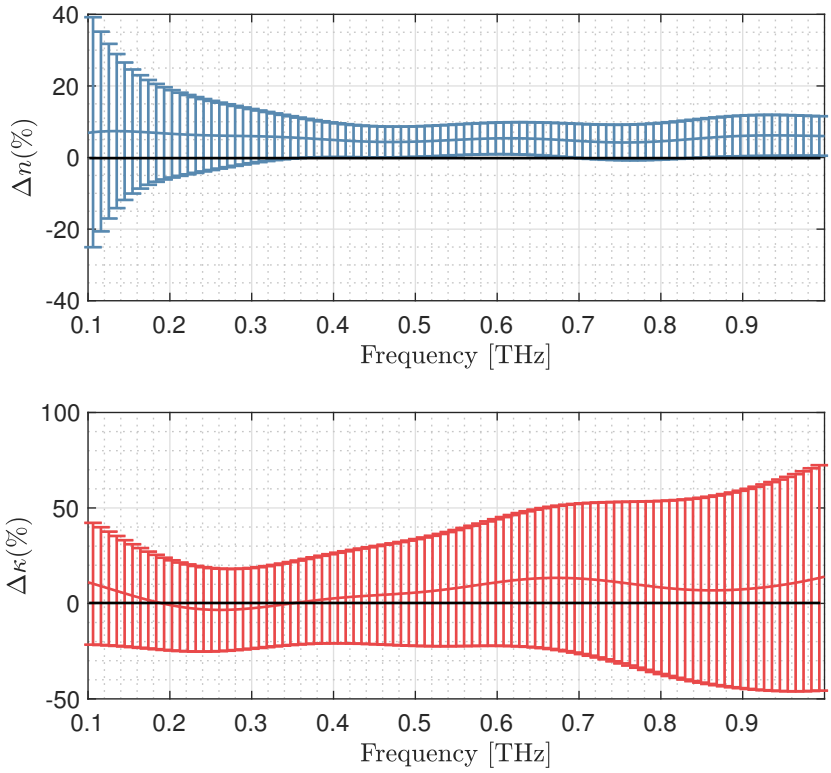


Figure II.35: Refractive index and extinction coefficient difference between healthy fibrous and cancer as a percentage of cancer characteristics.

came, *a priori*, to opposite conclusions? Indeed, in their work, Ashworth *et al* reported significant demarcations between healthy and cancerous tissues. However, heretofore, we explicitly stated on the particularly weak difference between healthy fibrous and malignant lesions properties. Although working toward the same goal, the two studies are actually quite different and **do not** reflect a shortcoming in the previous study **nor** in the present one. In their work, Ashworth *et al* have investigated over a large set of samples that were pre-selected as a function of the degree of homogeneity toward their histological structure. Tissues that were classified to belong to a tissue type were exhibiting a homogeneity toward this group of roughly 90%. In other words, cancer samples were presenting 90% of lesion over the analyzed surface. Similarly, samples classified as healthy fibrous and adipose tissue were exhibiting in average 80.8% and 89.7%, respectively, of the corresponding class. However, in the work reported through this manuscript, samples were attributed to the malignant category from the time that abnormal cells were present in the sample volume as mentioned in the *Tissue Registration and Preparation* section. Hence, the difference percentage is likely dictated by the fraction of healthy tissue to that of malignant lesion. On probing cancer region presenting a weak density of abnormal cells, the dielectric response is expected to mainly depict the one of fibrous tissue. On the contrary, sensing homogeneous pathological tissues would lead to higher refractive index. In that sense, it seems that terahertz radiations are not able to discriminate between healthy and malignant regions when facing weak cancer cell densities. Obviously, such a statement remains hypothetical and would require further works. As a suggestion, the worldwide bio-terahertz community has to focus on the quantification and qualification of the degree of sensitivity of terahertz radiations to lesion fraction in reflection configuration, that would be, in a second instance, a milestone to reach.

Concomitantly, one may underline the different pathologies of various etiologies that are, in this work, analyzed. Even though it is clearly impossible to state on the sensitivity of terahertz waves to specific pathologies, it still stands as a parameter that may interfere in the analysis. Un-

fortunately, the data set used in this work is not statistically significant to conduct investigations toward the possible different behavior to terahertz stimuli of the various pathologies gathered behind the term breast cancer. Reasonably, no features can be confirmed or negated to be due to a specific breast disease. Same conclusions are drawn towards patient age.

Although not being consistently distinguishable, healthy fibrous and cancer tissue still exhibit in average a significant difference in refractive index.

II.6.3 Relaxation Mechanisms

The model previously derived was then tested on each category. To this aim, the complex function of each group is derived from the optical function. Fitting parameters are given in Table II.12 and the respective graph are shown in Figure II.36.

Table II.12: Fitting parameters for the mean complex dielectric function of each tissue types for (II.33).

Cancer					Fibrous				
$\epsilon_\infty = 2.36$	α_0	α_1	β_0	β_1	$\epsilon_\infty = 2.15$	α_0	α_1	β_0	β_1
$1^{st} - Pole$:	3.34	23.72	0.61	4.75	$1^{st} - Pole$:	1.55	~ 0	0.50	~ 0
$2^{nd} - Pole$:	0.27	1.23	0.07	0.50	$2^{nd} - Pole$:	1.32	2.58	0.39	0.54
$3^{rd} - Pole$:	0.21	15.61	0.07	~ 0	$3^{rd} - Pole$:	0.85	21.20	0.16	4.77
$4^{th} - Pole$:	-0.56	-9.63	0.23	0.04	$4^{th} - Pole$:	~ 0	3.94	~ 0	0.04
	|Res|	0.1520				|Res|	1.3828		
Adipose									
$\epsilon_\infty = 1.89$	α_0	α_1	β_0	β_1					
$1^{st} - Pole$:	49.32	4.20	39.44	1.11					
$2^{nd} - Pole$:	21.01	17.99	4.28	4.53					
$3^{rd} - Pole$:	-26.25	-45.41	1.97	21.46					
$4^{th} - Pole$:	-39.21	-2.62	39.13	0.96					
	|Res|	0.6364							

The cancerous group is by far the class for which the model suits the best as indicated by residuals. The fitting for the mean complex dielectric function of malignant tissues is given in Figure II.36. Both the real and the imaginary part of the experimental dielectric function are well fitted. As depicted by the $\Lambda(f)$ metric, the error in the real and in the imaginary part is less than 10^{-4} and 10^{-3} respectively between 200 and 600 GHz. However, the error function clearly indicates that at frequencies beyond 800 GHz, the model does not match the data. Such an observation is assumed to be due to the fluctuation in κ that arises around 900 GHz. Although such a fluctuation is likely due to noise, it limits the accuracy of the fit at these frequencies. Remarkably, the slight increase in the real part of the permittivity is well retrieved. Additionally, the solely physical parameter ϵ_∞ that depicts the permittivity at infinite frequencies is greater than other tissue types but still remains lower than the one of water of 2.66 that was extracted by means of a two Debye poles in the corresponding section. Although such a trend was expected, it is still worth to underline it as freshly excised tissues nearly match *in vivo* circumstances and therefore are expected to exhibit a water-content in the range of 60-80%.

On fibrous tissues, the experimental real part of the permittivity exhibits certain oscillations that cannot from our study be solely attributed either to the noise nor internal processes. However, on comparing the behavior with the literature, it seems that the dispersion profile in ϵ' suffers from some experimental artifacts. Indeed, these features were not reported in any other studies. Although it is difficult to state on the origin of such fluctuations, let's remind that terahertz waves

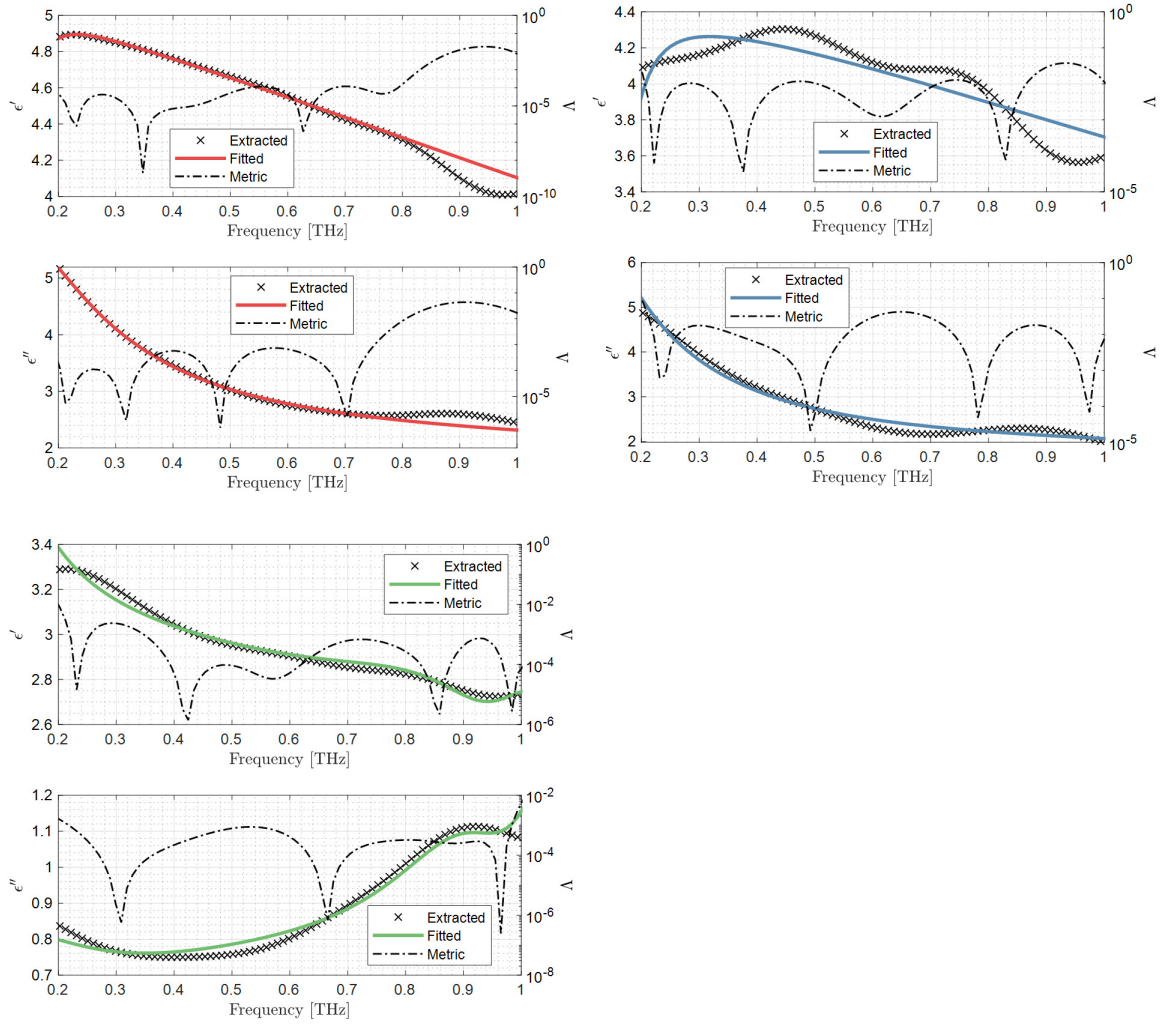


Figure II.36: Fitting of the dispersion profiles of the mean response of cancer (solid red), fibrous (solid blue) and adipose tissue (solid green). Dashed black line represents the Λ metric between measurement and fitting.

are expected to be particularly sensitive to water-content. Therefore, it was expected that the profile of the dielectric function of fibrous tissues follows a smooth tendency instead of presenting oscillations. Hence the fitted curve does not perfectly reflect the reported experimental behavior. In the used model, let's remind that the approximation of the dispersion is performed by the combination of oscillators that can compensate each others. In that sense, it would have been possible to strictly fit the experimental behavior. However, the less nonsmooth response among the imaginary part of the dielectric function allows to preserve the expected general behavior. Evidently, despite the fact that the imaginary part seems to limit numerical oscillator's stretching, the reliability in fitted profiles is affected. Nevertheless, the tendency of the adjusted ϵ' is close to the one shown by cancer tissues and should therefore remain approximately valid.

Last, adipose tissue fitting is quite tedious. In fact, the presence of an oscillation around 900 GHz that was attributed to the noise is of greatest impact on the extracted imaginary part. In fact, the naturally weak extinction in adipose tissues particularly suffers from the experimental noise that turns to blur the attenuation tissue response and ultimately, experimental fluctuations take over. Despite the good fit, it makes no doubts that the imaginary permittivity increase around 900 GHz is not caused by structural response as it was already pointed out for previous tissue types. Moreover, adipose tissues are expected to contain a lower amount of water and therefore sensing terahertz field should, globally be less sensitive to such type of tissue. Therefore, all these considerations may negatively impact the identicalness of the proposed model toward adipose

tissues. Of course, it would have been possible to adjust the dispersion profiles of adipose tissues between 200 GHz and 500 GHz to do not take into account the noisy contribution that takes place at higher frequencies. However, we would therefore consider a particularly short frequency band by about 300 GHz which, by no means, may traduce the dielectric behavior over the spectral band that we, heretofore, considered. Although the extracted dispersion profile in ϵ'' of adipose tissue seems weakly reliable at high frequencies, it would not be surprising that such a trend frequently appears when probing weakly attenuating samples. Therefore, although the permittivity increase may not be correlated to physical process but rather to noise, such a trend concomitantly depicts the weak intrinsic extinction in adipose tissues that still may be used as identification parameters.

Globally, the model that was reported in equation II.33 seems to be very sensitive to experimental artifacts and the different sources of noise. That is not surprising since the model was proposed to be more flexible. Flexibility turns to play against us when dealing with noisy signals and hence particularly weak absorbing material like adipose tissues. In fact, the model succeeds in adjusting dispersion curves that are *a priori* not depicting solely the intrinsic dielectric behavior but rather a summation of noise and tissue response, like it was shown for adipose tissues. In the same time fitting on the complex plane allows to maintain consistency as one less nonsmooth dielectric part may weight fluctuations exhibited by the other part, as shown for healthy fibrous samples. Finally, cancer curves were almost perfectly retrieved. This may be explained by the smooth dispersion profiles that were extracted and the sensitivity of terahertz waves to the theoretically higher water-content in malignant lesions.

The discussion in the foregoing, having eventually clarified the relevant facts of the fitting procedure using II.33, one is at that moment able to assess whether there are specific distributions among the sorted parameters that may help to discriminate between tissue types. First, let's consider the value of the permittivity at infinity frequency. The value of ϵ_∞ decreases from cancer to adipose tissues with fibrous tissues making the bridge between them. Overall the value is lower than the one that was reported for water. However, these values are quite different from the ones reported in **[43]** but are closer to the one reported by **[44]**. A brief comparison is given in Table II.13

Table II.13: Comparison of ϵ_∞.

	(A)	(N)	(C)
[43]	2.41	2.80	3.15
[44]	-	2.58	2.65
This work	1.89	2.15	2.36

Differences among these values may arise from various origins: first, tissue registration and classification among studies; Second, fitting model. **[43]** used the model described in equation II.37, **[44]** employed two Debye poles at frequencies beyond roughly 500 GHz and here, the model in II.33 was used; Third, the spectral band within which the procedure was performed; Last but not least, the choice and the implementation of the solving algorithm. All these factors may obviously lead to slightly different values. Nevertheless, the increase in ϵ_∞ from adipose tissue to malignant ones is common to all studies. Let's note however, that no value for adipose tissues was reported by **[44]**. In Figure II.37, is reported the distribution of the contribution of the parameters of each poles for the three types of tissues. Overall, the adipose tissues exhibit higher responses than other groups. Such a statement remains generically valid among the four fitting poles. On comparing fibrous and tumor, one can note the important response of α_{1m} for the four poles. Additionally, an important difference is reported between healthy and malignant tissues in α_{11}. Let's remind that α_{1m} was introduced to offer more flexibility to the model. Furthermore, it was introduced to add a time-dependent parameter, such as the probed region polarization as a function of the time derivative of the incident electromagnetic field. However, as the model was not specifically designed to trace back the internal physical processes, it would not be reasonable to state on the physical significance of the contribution distribution. It is, nonetheless, possible to draw an as-

sumption on the high specificity of α_{1m} toward the discrimination between healthy tissues and tumor.

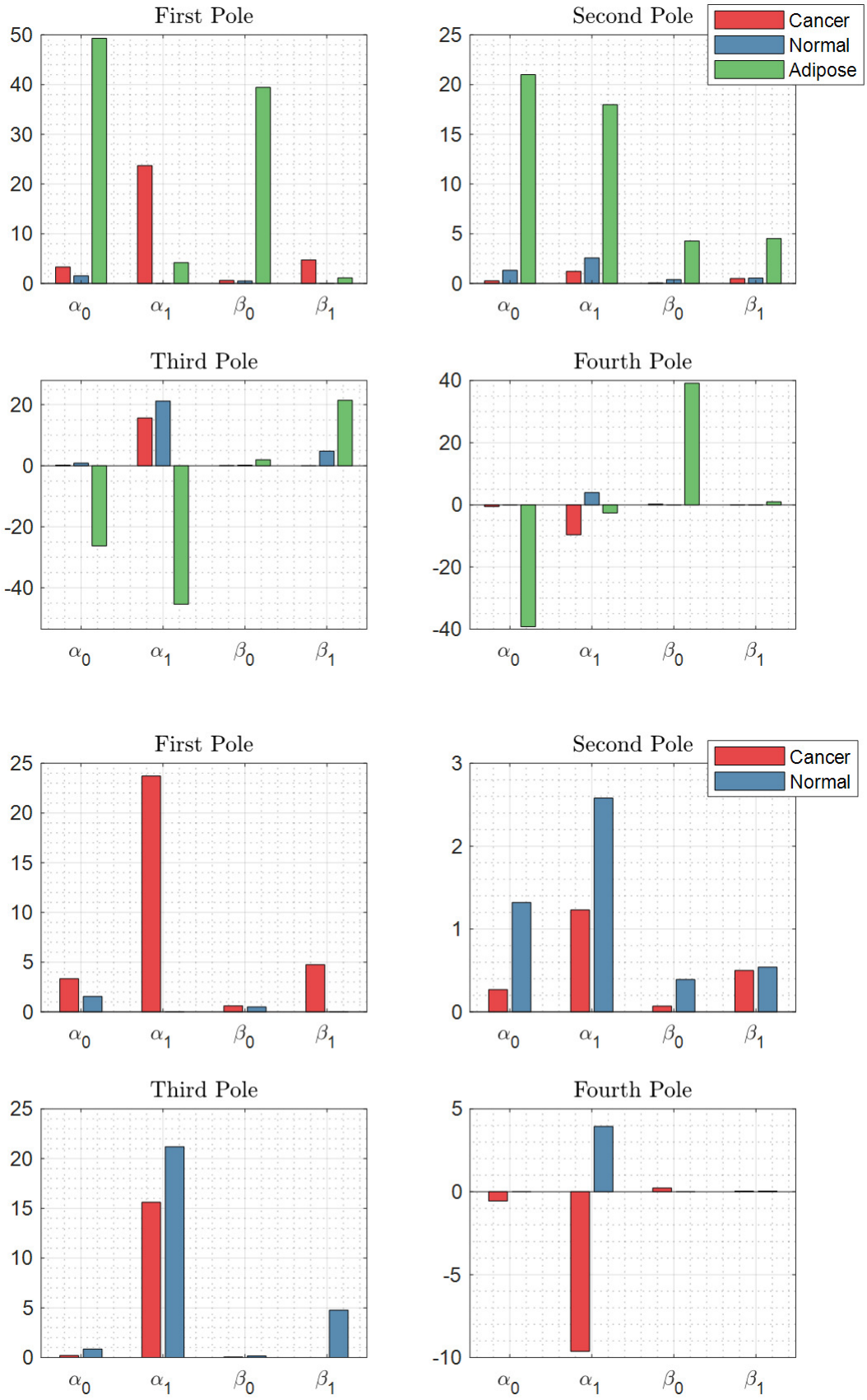


Figure II.37: Averaged distribution of model parameters of each poles for tissue types.

In conclusion, the pre-clinical study allowed to identify the key features of each tissue type to further discriminate them by means of terahertz pulses.

Overall, among the three tissue categories, adipose tissues are highly distinguishable from fibrous and cancer tissues. Both the refractive index and the extinction coefficient are significantly lower than the ones of healthy fibrous and malignant samples. The only parameter that can, from our study, explain such a trend is the low water-content found in adipose tissue in comparison to fibrous and cancer. Although such a statement towards water-content is valid, additional processes may play specific roles in lowering the dielectric response but are yet to be understood and identified.

On analyzing freshly excised human breast tissues, it was also reported that weak differences were observed between healthy fibrous tissues and malignant samples. At low frequencies, below 300 GHz, the errorbars overlap leading to the conclusion that no distinction can be made in this spectral window. Considering the refractive index beyond 300 GHz, it was observed that the difference between the real optical function of cancer and the one of fibrous was varying between 0% and 15%. We assumed that the range of variation was mostly dictated by the fraction of cancer cells to the one of healthy cells and connective tissues. Concluding that despite differences were in average observed in the refractive index, false identification diagnosis would frequently be reported.

The model described in equation II.33 was therefore employed to assess whether fitting parameters may be used as complementary identifier to distinguish between healthiness and malignancy. Despite minimization issues that were all along reported in the previous section and the present one, we identified the α_{1m} parameter to be particularly sensitive in comparison to other parameters. The high response of α_{1m} can support the idea of the necessity to introduce more flexible parameter to fit the overall dispersion profiles of freshly excised human tissues. Evidently, further works are required toward the development of the proposed model to lead its use to more robust analysis and especially when working towards breast cancer detection.

II.7 Cancerous Cells

In the foregoing, it has been argued that one of the most promising factor to contrast between healthy and cancer was the inner water-content. But, in reality such a statement is a shortcut. Let's first ask oneself how the inside of a cell looks like ? The inside of a cell is actually analogous to the *Champ-de-Mars* during summer: a particularly crowded place. As reported by R. John Elis in **[45]**, macromolecules occupy typically 20-30% of the intracellular volume, water making up the remainder. Therefore it was suggested that the inside of the cytoplasm is made of a jelly-like environment rather than being simply filled with a fluid **[46]**. Interestingly, one of the most astonishing advance toward cancer detection was realized with nuclear magnetic resonance **[47]**. The author reported a considerable increase in the relaxation time of the spin-lattice T_1 and of the spin-spin T_2 on comparing malignant Novikoff hepatoma and normal rat liver. The relaxation time increase was attributed to a less ordered intracellular water network. A decrease in the degree of ordering of inner water reflects the increase of water dynamics inside the cytoplasm. Shiraga and colleagues probed the dielectric response of DLD-1, HEK293 and HeLa human cancer cell monolayers by attenuated total reflection (ATR) **[48]**. For the three types of cell lines, Shiraga *et al* reported an increase in the real part of the dielectric function and a decrease in the imaginary one to those of distilled water. On the basis of the same technique, Grognot and Gallot investigated Madine-Darby Canine Kidney (MDCK) cell membrane permeabilization with a detergent, the saponin **[49]**. Saponin removes cholesterol chains that are present in the cell membranes making them permeable to antibodies or allowing them to release intracellular proteins **[50]**. Following the addition of saponin, they reported a time progressive decrease of the cell monolayer response to the incident terahertz field.

II.7.1 Concentration-Dependent Properties

In the following, investigations on Michigan Cancer Foundation - 7 (MCF7), a breast cancer cell line are reported. Different concentration of MCF7 cells were probed in transmission spectroscopy by means of a TPS3000 (TeraView Ltd, Camrbidge) to study the dielectric function behavior as a function of the cell density. Cells were grown in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% Fetal Bovine Serum, 1% Penicilin-Streptomicin and 1% L-Glutamine. When 80% confluence was reached, cells were detached with trypsin and subsequently immersed in standard cellular medium. Four different concentrations were analyzed (Table II.14).

Table II.14: MCF7 cell concentrations.

#	Concentration (c/mL)
#1	20,000
#2	1,000,000
#3	2,000,000
#4	3,300,000

MCF7 cell solutions were homogenized before being injected in an infrasil quartz sealed liquid spectrophotometer cell with rectangular aperture from International Crystal Laboratories® (Figure II.38). The sealed liquid cell is made up with two 4 mm thick infrasil quartz window separated by 109.1 μm. The calibration of the sealed liquid cell was performed by the International Crystal Laboratories® company. Each concentration was measured subsequently and independently five times. The measurement chamber was filled with nitrogen to minimize the absorption from water vapor along the optical path.

Figure II.38: Sealed liquid cell.

The extraction of the dielectric properties of each solution was performed in a similar way than in previous sections. It is worth mentioning that as the study was performed in transmission configuration, the transfer function to be inserted in the inverse electromagnetic problem objective function is the one depicted in equation II.20a instead of II.20b. The real and imaginary profiles of the dielectric function for each MCF7 cell concentration are given in Figure II.39.

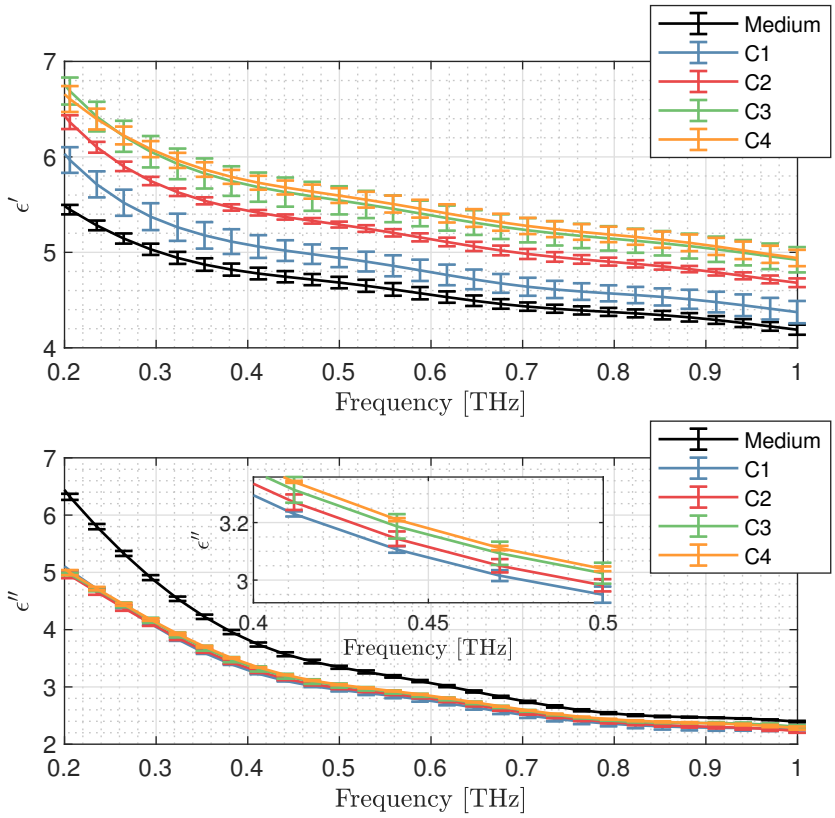


Figure II.39: Dispersion profiles of the real and imaginary part of the dielectric constant for each cell concentration.

As far as it is know, it is the first time that such studies on cancer cells in suspension are reported. As such, both the results and the assumptions that are drawn towards the dispersion profiles must be interpreted and considered with caution.

The lossless ϵ' increases with increased MCF7 cell concentration. The results are significant since the errorbars that represent the standard deviation over the independent measurements do not overlap for C_1 and C_2. However, C_3 and C_4 errorbars overlap over the entire frequency window. A reasonable assumption would state that the optical path through which passes the probing pulses is already saturated with cancer cells at 2,000,000 cells per milliliter and therefore no significant differences are observed on increasing the cell concentration to 3,300,000 cells per milliliter. The increase of the real part of the permittivity as a function of the concentration can be explained by the increase of the medium density and therefore the increase of the probability of interactions between cells and the sensing pulses. Consequently, terahertz pulses are progressively delayed in a rate proportional to the cell concentration.

On the lossy permittivity ϵ'', the significant difference between the medium and the different cell concentrations are mainly located between 200 and 700 GHz. On the other hand beyond 700 GHz, no noticeable variations are observed. Overall, with the addition of cancer cells, the solution permittivity decreases. That can be explained by the fact that strongly absorbing bulk and free water molecules found in the biological medium are replaced by intracellular hydrated water that form the hydration shell around macromolecules on interacting with macromolecules surface groups. Additionally, by increasing the cell concentration the permittivity seems to increase. Therefore one could ask oneself if the previous statement is still valid. Although it is impossible to confirm or infirm the previous assumption, one has still to keep in mind the possible contribution from scattering. Eukaryote cells have diameter that typically range from 10 to 50 μm. Furthermore cells may link to each others to form aggregates of few individuals that may increase even more the scattering. Therefore, one possible reason for such a global behavior is that on adding cell

into the medium, the absorbing water molecules are replaced with less movable hydrated water and therefore the part of the permittivity that corresponds to dielectric losses decreases. Concomitantly, the previous effect is counterbalanced by scattering. Remarkably, the two processes seem to roughly compensate each other since increasing the cell concentration does not lead to significant changes among dielectric losses. Interestingly, the addition of hydrated water into the solution does not lead to specific features between 200 GHz and 1 THz. Globally, the response of the dipole of water molecules in the hydration shell are retarded to a nanosecond timescale as it was reported by **[51–53]**. Therefore, the contribution of hydrated water is expected to be negligible in the terahertz window that is heretofore considered.

II.7.2 Influence on Relaxation Mechanisms

In order to determine and to ascertain the roots for these frequency dependent dispersion profiles, the dielectric constants that were extracted are analyzed by taking apart two relaxation mechanisms of the form:

$$\epsilon_m = \frac{\Delta\epsilon^m}{1 + i\omega\tau_m}, \tag{II.39}$$

where the first pole ($m = 1$) corresponds to the slow relaxation process of bulk water and the second ($m = 2$) denotes the fast relaxation of free and quasi-free water molecules. The two corresponding relaxation times were fixed to the respective values that were extracted for water in the related section, *i.e.*, $\tau_D = 8.34$ ps and $\tau_2 = 0.1$ ps. Holding fixed these two parameters allows to investigate the impact of cancer cells in the medium to the overall amplitudes of each polarization mechanisms. The relaxation amplitudes $\Delta\epsilon_1$ and $\Delta\epsilon_2$ as well as the permittivity at infinite frequencies ϵ_∞ are reported in Table II.15 for best fitted relaxation strengths. In Figure II.40 are depicted the three free parameters to those of the medium as a function of the cell concentration.

Table II.15: Debye parameter for the best fitted relaxation strengths.

	Medium	C_1	C_2	C_3	C_4
ϵ_∞	2.28	2.29	2.53	2.70	2.73
$\Delta\epsilon_1$	67.54	54.17	53.72	54.00	54.30
$\Delta\epsilon_2$	2.53	2.84	2.96	3.07	3.08

From the extraction of the parameters one can see that electron excitation as well as intramolecular vibrations that occur at infinite frequencies and denoted by ϵ_∞, are increasing alongside the cell concentration increase. Such a trend would likely denote supplemental intramolecular modes that appear on suspending cancer cells in the biological medium and that take place in the intracellular medium.

Interestingly, for all concentrations the slow relaxation process strength $\Delta\epsilon_1$ decreases with respect to that of biological medium. That may consequently explain the decrease observed in the imaginary part ϵ'' of the dielectric function although the variation of the parameter $\Delta\epsilon_1$ does not strictly follow the cell concentration for a reason that as yet to be understood. However, such a decrease in the relaxation strength $\Delta\epsilon_1$ clearly indicates the diminution of bulk water and therefore demonstrates that intracellular water molecules do not possess the same dynamics.

Remarkably, the strength $\Delta\epsilon_2$ increases as a function of the cell concentration, which may explain the increase in the lossless part of the dielectric function. The origin for such a behavior is not certain but we can draw three differents hypothesis: first, MCF7 cells naturally possess in average a greater number of free and quasi-free water molecules; Second, although the density of bulk water into cytoplasm are expected to be weak, the present hydrogen-bonded molecules may be less strongly bonded than those in liquid water and therefore may exhibit larger network reformation when exposed to terahertz radiations; Third, on irradiating cells with terahertz waves,

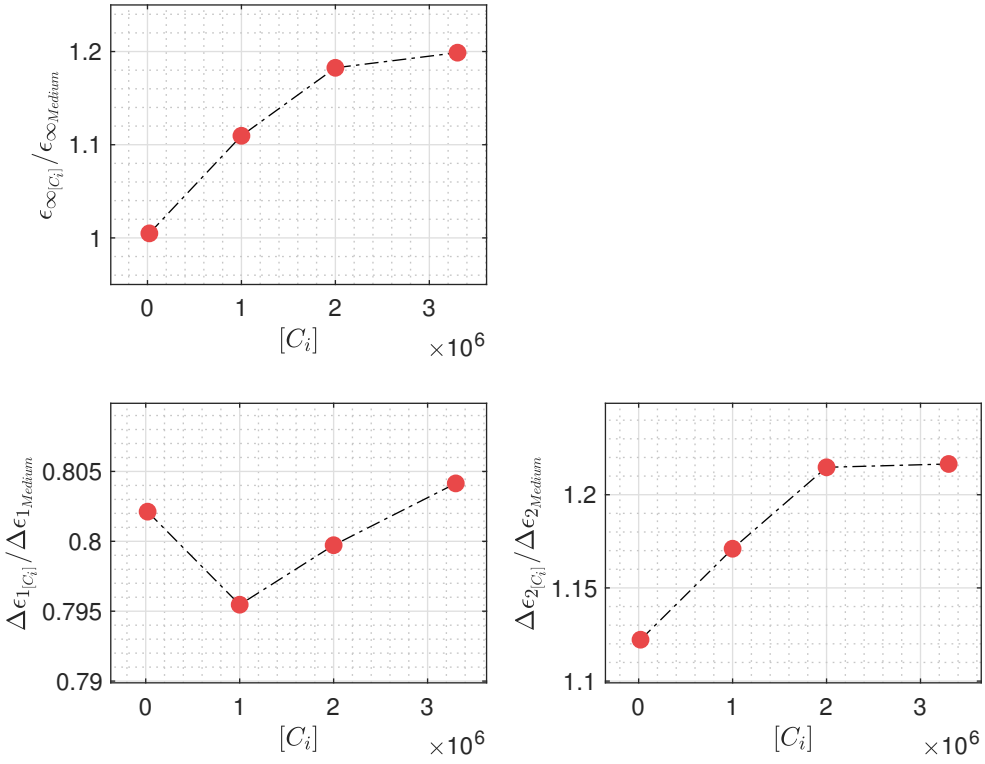


Figure II.40: Variations in ϵ_∞, $\Delta\epsilon_1$ and $\Delta\epsilon_2$ as a function of MCF7 cell concentration.

the hydration shell may release some hydrated water that transiently behave as free and quasi-free molecules.

The contribution to each of the mechanisms, namely ϵ_∞, ϵ_1 and ϵ_2 to the lossless and lossy part of the dielectric constant are depicted for each concentration in Figure II.41.

Overall, it is worth to mention that further studies to lead further considerations and discussions are required. However, on investigating the dielectric response of MCF7 cells that are a widely used cell model to mimic *in vivo* breast carcinoma cells, we raised important foundation to understand the contrast mechanism that appear when probing breast cancer by means of terahertz radiations. A supplemental step would consist in investigating the dielectric response of cardiofibroblast that may be seen as a good representation of healthy cells. Such a study would allow one to determine the contribution of free and quasi-free water molecules in differentiating healthy from malignant tissues that are likely, from this work, playing a substantial role.

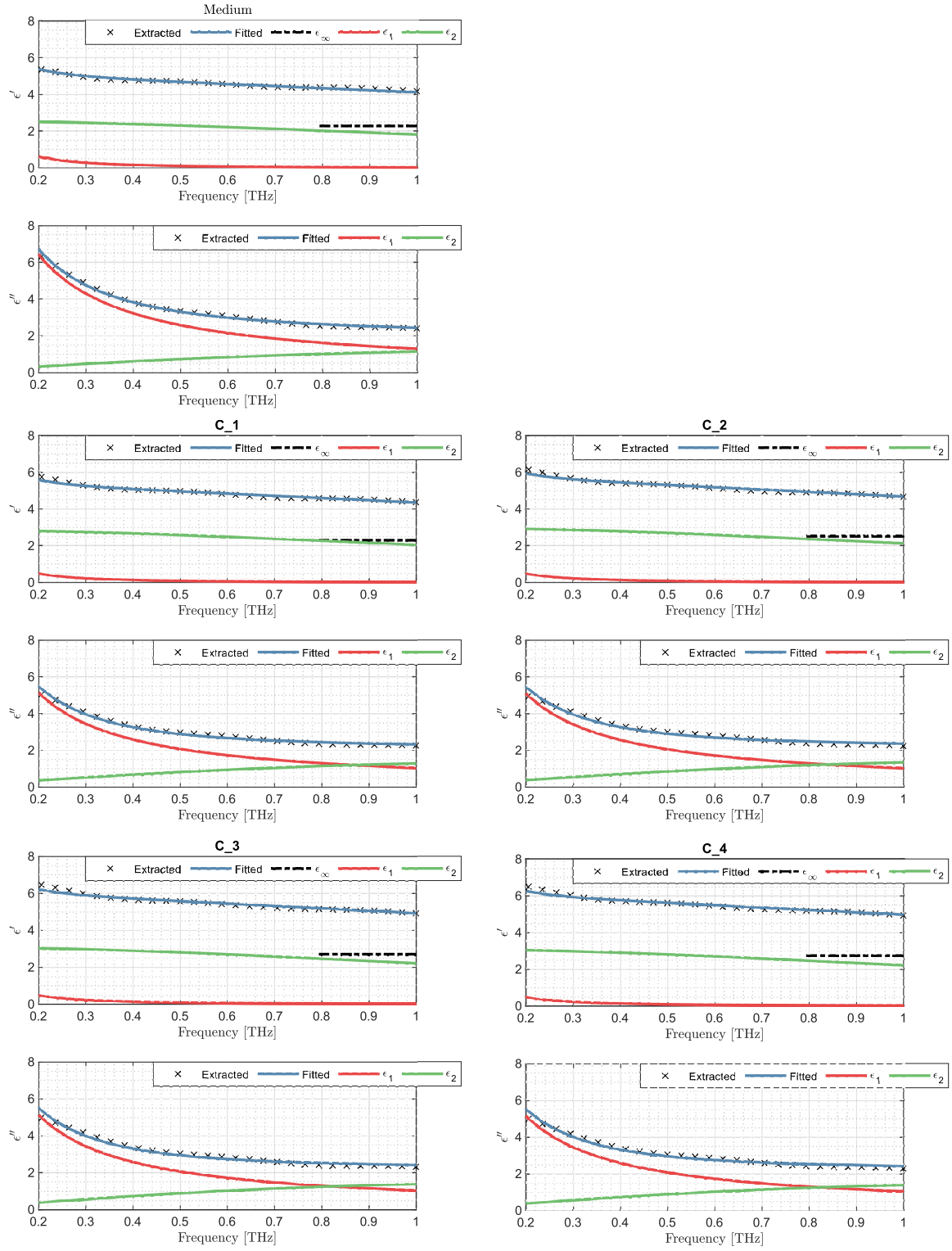


Figure II.41: Contribution of $\epsilon\infty$, ϵ_1 and ϵ_2 to the dielectric constant dispersion profiles.

II.8 Conclusion

Throughout this chapter, the reader went through the theoretical basis toward generating, detecting and using terahertz radiations to sense materials in the subpicosecond timescale.

After having chiefly described the different experimental setups that were used to conduct the present work, a mathematical model, namely ITA, that was developed within the scope of this study was presented. The equations that were here derived allow to define the complex transfer

function of any materials that may contain up to an infinite number of internal dielectric interfaces. Furthermore thanks to the mathematical design of the model, a parametric complex transfer function that enlightens the specific optical paths that contribute the most to the recorded signal is achievable. The effectiveness of the procedure was demonstrated by reconstructing the reflected signal from a stack made of four layers of tens of microns.

The numerical procedure that was used all along this work to extract the intrinsic dielectric and optical properties of probed samples was described. In a nutshell it consists of solving an inverse electromagnetic problem by minimizing an objective function that denotes the difference between simulated modulus and phase, and the recorded ones.

These preliminary numerical and mathematical works allowed to pursue toward biological investigations. As water is the predominant compound in living organisms, a specific section was dedicated to water and the extraction of its physical properties in the terahertz band. Subsequently, preliminary investigations on phantom mimicking phantoms were conducted to prepare pre-clinical studies on real freshly excised human breast tissues. A numerical model that is aimed to describe the complete dispersion profiles of the real and imaginary part of the dielectric function was presented. Although such a model requires further developments and rigorous additional considerations, it may lead to new routes toward the construction of a theoretical model to simulate the dielectric response of tissues.

Afterwards, investigations on mimicking tissues, a pre-clinical study at hospital were reported. A set of 48 terahertz waveforms arising from eleven different patient was collected. Important statement towards the possibility of discriminating between healthy and malignant tissues were drawn. We shown that except from adipose tissues that was consistently distinguished, healthy fibrous and cancer tissues were exhibiting close behaviors. We assumed that the ability of differentiating between the two tissue types was correlated to malignant cell density to that of healthy compounds. In average, the two types of tissues were distinguished at frequencies higher than 300 GHz while at lower frequencies, the errorbars were strongly overlapping.

Finally, we investigated the dielectric response of MCF7 cells in suspension that are a good model of breast carcinoma cells. Interestingly, we reported particular features that indicate that breast cancer cells possess either more weakly bonded water or more naturally free water molecules. It was therefore mentioned that such a trend could possibly state for the contrast mechanism that was -in average- observed on probing freshly excised cancer tissues and healthy ones.

In the next chapter, the reader will follow the work that was done towards terahertz imaging and image processing to localize malignant lesions and to compare terahertz pictures from those that are biding on medical parties.

II.9 Bibliography

[1] H Rubens and O Von Baeyer. On extremely long waves, emitted by the quartz mercury lamp. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 21(125):689–695, 1911. 54

[2] EF Nichols and JD Tear. Short electric waves. *Physical Review*, 21(6):587, 1923. 54

[3] A Glagolewa-Arkadiewa. Short electromagnetic waves of wave-length up to 82 microns. *Nature*, 113(2844):640, 1924. 54

[4] Masahiko Tani, Michael Herrmann, and Kiyomi Sakai. Generation and detection of terahertz pulsed radiation with photoconductive antennas and its application to imaging. *Measurement science and technology*, 13(11):1739, 2002. 54

[5] Nathan M Burford and Magda O El-Shenawee. Review of terahertz photoconductive antenna technology. *Optical Engineering*, 56(1):010901, 2017. 54

[6] Lionel Duvillaret, Frédéric Garet, J-F Roux, and J-L Coutaz. Analytical modeling and optimization of terahertz time-domain spectroscopy experiments, using photoswitches as antennas. *IEEE Journal of Selected Topics in Quantum Electronics*, 7(4):615–623, 2001. 54

[7] David H Auston. Picosecond optoelectronic switching and gating in silicon. *Applied Physics Letters*, 26(3):101–103, 1975. 55

[8] Kiyomi Sakai and Masahiko Tani. Introduction to terahertz pulses. In *Terahertz optoelectronics*, pages 1–30. Springer, 2005. 55

[9] PR Smith, DH Auston, AM Johnson, and WM Augustyniak. Picosecond photoconductivity in radiation-damaged silicon-on-sapphire films. *Applied Physics Letters*, 38(1):47–50, 1981. 55

[10] Peter R Smith, David H Auston, and Martin C Nuss. Subpicosecond photoconducting dipole antennas. *IEEE Journal of Quantum Electronics*, 24(2):255–260, 1988. 55

[11] Robert W Boyd. *Nonlinear optics*. Elsevier, 2003. 56

[12] C Winnewisser, P Uhd Jepsen, M Schall, V Schyja, and H Helm. Electro-optic detection of thz radiation in litao 3, linbo 3 and znte. *Applied Physics Letters*, 70(23):3069–3071, 1997. 57

[13] L Xu, X-C Zhang, and DH Auston. Terahertz beam generation by femtosecond optical pulses in electro-optic materials. *Applied Physics Letters*, 61(15):1784–1786, 1992. 57

[14] X-C Zhang, XF Ma, Y Jin, T-M Lu, EP Boden, PD Phelps, KR Stewart, and CP Yakymyshyn. Terahertz optical rectification from a nonlinear organic crystal. *Applied Physics Letters*, 61(26):3080–3082, 1992. 57

[15] Arno Schneider, Max Neis, Marcel Stillhart, Blanca Ruiz, Rizwan UA Khan, and Peter Günter. Generation of terahertz pulses through optical rectification in organic dast crystals: theory and experiment. *JOSA B*, 23(9):1822–1835, 2006. 57

[16] Qi Wu and X-C Zhang. Free-space electro-optic sampling of terahertz beams. *Applied Physics Letters*, 67(24):3523–3525, 1995. 57

[17] Benjamin S Williams. Terahertz quantum-cascade lasers. *Nature photonics*, 1(9):517, 2007. 57

[18] Ping Gu, Masahiko Tani, Kiyomi Sakai, and T-R Yang. Detection of terahertz radiation from longitudinal optical phonon–plasmon coupling modes in insb film using an ultrabroadband photoconductive antenna. *Applied Physics Letters*, 77(12):1798–1800, 2000. 57

[19] R Kersting, Karl Unterrainer, G Strasser, HF Kauffmann, and E Gornik. Few-cycle THz emission from cold plasma oscillations. *Physical Review Letters*, 79(16):3038, 1997. 57

[20] Abdul J Jerri. The shannon sampling theorem—its various extensions and applications: A tutorial review. *Proceedings of the IEEE*, 65(11):1565–1596, 1977. 57

[21] R. Bocquet, N. Breuil, L. Chusseau, J.-L. Coutaz, P. Crozat, J. Demaison, L. Duvillaret, G. Gallot, F. Garet, J.-F. Lampin, D. Lippens, J. Mangeney, P. Mounaix, G. Mouret, and J.-F. Roux. *Optoélectronique térahertz*. EDP Sciences, 2008. 59, 61

[22] Ronald Newbold Bracewell and Ronald N Bracewell. *The Fourier transform and its applications*, volume 31999. McGraw-Hill New York, 1986. 60

[23] Robert E Edwards. *Fourier series: a modern introduction*, volume 2. Springer Science & Business Media, 2012. 60

[24] Lionel Duvillaret, Frédéric Garet, and Jean-Louis Coutaz. Influence of noise on the characterization of materials by terahertz time-domain spectroscopy. *JOSA B*, 17(3):452–461, 2000. 61

[25] Withawat Withayachumnankul, Bernd M Fischer, Hungyen Lin, and Derek Abbott. Uncertainty in terahertz time-domain spectroscopy measurement. *JOSA B*, 25(6):1059–1072, 2008. 61

[26] Martin Van Exter, Ch Fattinger, and D Grischkowsky. Terahertz time-domain spectroscopy of water vapor. *Optics letters*, 14(20):1128–1130, 1989. 61

[27] Quentin Cassar, Adrien Chopard, Frederic Fauquet, Jean-Paul Guillet, Mingming Pan, Jean-Baptiste Perraud, and Patrick Mounaix. Iterative tree algorithm to evaluate terahertz signal contribution of specific optical paths within multi-layered materials. *IEEE Transactions on Terahertz Science and Technology*, 2019. 62

[28] S Krimi, J Klier, M Herrmann, J Jonuscheit, and R Beigang. Inline multilayer thickness sensing by using terahertz time-domain spectroscopy in reflection geometry. In *2013 38th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2. IEEE, 2013. 63

[29] Norbert Palka, Soufiene Krimi, Frank Ospald, Danuta Miedzinska, Roman Gieleta, Marcin Malek, and Rene Beigang. Precise determination of thicknesses of multilayer polyethylene composite materials by terahertz time-domain spectroscopy. *Journal of Infrared, Millimeter, and Terahertz Waves*, 36(6):578–596, 2015. 63

[30] Lionel Duvillaret, Frederic Garet, and J-L Coutaz. A reliable method for extraction of material parameters in terahertz time-domain spectroscopy. *IEEE Journal of selected topics in quantum electronics*, 2(3):739–746, 1996. 71

[31] Lionel Duvillaret, Frederic Garet, and Jean-Louis Coutaz. Highly precise determination of optical constants and sample thickness in terahertz time-domain spectroscopy. *Applied optics*, 38(2):409–415, 1999. 71

[32] Osman S Ahmed, Mohamed A Swillam, Mohamed H Bakr, and Xun Li. Efficient optimization approach for accurate parameter extraction with terahertz time-domain spectroscopy. *Journal of Lightwave Technology*, 28(11):1685–1692, 2010. 71

[33] Ioachim Pupeza, Rafal Wilk, and Martin Koch. Highly accurate optical material parameter determination with thz time-domain spectroscopy. *Optics express*, 15(7):4335–4350, 2007. 71

[34] Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998. 73

[35] Constantino Grosse. A program for the fitting of debye, cole–cole, cole–davidson, and havriliak–negami dispersions to dielectric data. *Journal of colloid and interface science*, 419:102–106, 2014. 74

[36] Uffe Møller, David G Cooke, Koichiro Tanaka, and Peter Uhd Jepsen. Terahertz reflection spectroscopy of debye relaxation in polar liquids. *JOSA B*, 26(9):A113–A125, 2009. 78, 79

[37] Caroline Reid. *Spectroscopic methods for medical diagnosis at terahertz wavelengths*. PhD thesis, UCL (University College London), 2009. 80, 81

[38] Mariya Lazebnik, Ernest L Madsen, Gary R Frank, and Susan C Hagness. Tissue-mimicking phantom materials for narrowband and ultrawideband microwave applications. *Physics in Medicine & Biology*, 50(18):4245, 2005. 81

[39] E Liakhov, O Smolyanskaya, A Popov, E Odlyanitskiy, N Balbekin, and M Khodzitsky. Fabrication and characterization of biotissue-mimicking phantoms in the thz frequency range. In *Journal of Physics: Conference Series*, volume 735, page 012080. IOP Publishing, 2016. 81

[40] OA Smolyanskaya, Q Cassar, MS Kulya, NV Petrov, KI Zaytsev, AI Lepeshkin, J-P Guillet, P Mounaix, and VV Tuchin. Interaction of terahertz radiation with tissue phantoms: numerical and experimental studies. In *EPJ Web of Conferences*, volume 195, page 10012. EDP Sciences, 2018. 81

[41] Philip C Ashworth, Emma Pickwell-MacPherson, Elena Provenzano, Sarah E Pinder, Anand D Purushotham, Michael Pepper, and Vincent P Wallace. Terahertz pulsed spectroscopy of freshly excised human breast cancer. *Optics express*, 17(15):12444–12454, 2009. 83, 90

[42] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 85

[43] Bao CQ Truong, Hoang D Tuan, Anthony J Fitzgerald, Vincent P Wallace, and Hung T Nguyen. A dielectric model of human breast tissue in terahertz regime. *IEEE Transactions on Biomedical Engineering*, 62(2):699–707, 2014. 86, 97

[44] Emma Pickwell. *Biological applications of terahertz pulsed imaging and spectroscopy*. PhD thesis, University of Cambridge, 2005. 97

[45] R John Ellis. Macromolecular crowding: obvious but underappreciated. *Trends in biochemical sciences*, 26(10):597–604, 2001. 99

[46] Jack T Trevors and Gerald H Pollack. Hypothesis: the origin of life in a hydrogel environment. *Progress in biophysics and molecular biology*, 89(1):1–8, 2005. 99

[47] Raymond Damadian. Tumor detection by nuclear magnetic resonance. *Science*, 171(3976):1151–1153, 1971. 99

[48] Keiichiro Shiraga, Yuichi Ogawa, Tetsuhito Suzuki, Naoshi Kondo, Akiyoshi Irisawa, and Motoki Imamura. Characterization of dielectric responses of human cancer cells in the terahertz region. *Journal of Infrared, Millimeter, and Terahertz Waves*, 35(5):493–502, 2014. 99

[49] Marianne Grognot and Guilhem Gallot. Quantitative measurement of permeabilization of living cells by terahertz attenuated total reflection. *Applied Physics Letters*, 107(10):103702, 2015. 99

[50] Maria Célia Jamur and Constance Oliver. Permeabilization of cell membranes. In *Immunocytochemical methods and protocols*, pages 63–66. Springer, 2010. 99

[51] Johan Qvist, Erik Persson, Carlos Mattea, and Bertil Halle. Time scales of water dynamics at biological interfaces: peptides, proteins and cells. *Faraday discussions*, 141:131–144, 2009. 102

[52] Biman Bagchi. Water dynamics in the hydration layer around proteins and micelles. *Chemical Reviews*, 105(9):3197–3219, 2005.

[53] Keiichi Yokoyama, Takashi Kamei, Hiroshi Minami, and Makoto Suzuki. Hydration study of globular proteins by microwave dielectric spectroscopy. *The Journal of Physical Chemistry B*, 105(50):12622–12627, 2001. 102

Chapter III

Terahertz Imaging: Far-field Survey

Who came up with the idea that telling the truth is easy? That's already a lie.

Dmitry Glukhovsky

Sommaire

III.1 Introduction . . . 110
III.2 Far-field Imaging: Principles . . . 110
III.2.1 Image & Limitation . . . 110
III.2.2 Effective Spatial Resolution . . . 112
III.3 Self-Reference Image Inversion . . . 114
III.3.1 Second Dielectric Interface Imaging . . . 115
III.3.2 Reference Retrieval . . . 115
III.3.3 Wavelet Denoising . . . 117
III.3.4 Image Inverse Electromagnetic Problem . . . 119
III.4 Image Registration . . . 121
III.4.1 Pathology Image Issues . . . 121
III.4.2 Registration Procedure . . . 122
III.5 Correlation . . . 124
III.5.1 Classification Models . . . 124
III.5.2 Confusion Matrices . . . 129
III.6 Study Cases . . . 131
III.6.1 NS1 Sample . . . 132
III.6.2 NS2 Sample . . . 137
III.6.3 NS3 Sample . . . 142
III.6.4 Discussion on Inversion & Segmentation . . . 146
III.7 Principal Component Analysis . . . 148
III.7.1 Mathematical Definition . . . 148
III.7.2 NS1 Sample . . . 150
III.7.3 NS2 Sample . . . 153
III.7.4 NS3 Sample . . . 156
III.7.5 Discussion on PCA . . . 158
III.8 Conclusion . . . 160
III.9 Bibliography . . . 161

III.1 Introduction

The third chapter *Terahertz Imaging: Far-field Survey*, review the work that was conducted towards breast carcinoma recognition, in the context of the Fraunhofer approximation.

As reported in the previous chapter, the different types of tissue respond differently to terahertz pulses. Based on these differences, diverse strategies are investigated to classify tissue regions that are imaged in accordance to their biological nature.

Beforehand, the operational characteristics of the imaging setup that is used in this work are reported, followed by the description of a self-reference method that was developed to improve sample characterization.

The refractive index map of a set of tissue samples was extracted by means of inverse electromagnetic problems. The collected maps are then segmented with different classification methods. Chiefly, it consists of defining a refractive index threshold above which, pixels are marked as malignant. However, as pixels are numerical entities rather than biological ones, such a type of rough segmentation are often hardly efficient. Therefore, supplemental strategies are worked out and are reported. The efficiency of each segmentation is assessed by filling the associated confusion matrix with regards to the corresponding pathology images, annotated by the pathologist. Sensitivity and specificity of each segmentation are derived.

Complementary to these studies, principal component analysis is investigated. Such a procedure does not aim to qualify principal component analysis efficiency for tissue classification, but rather to explore a possible correlation between the contrast exhibited from principal component analysis and pixel malignancy.

III.2 Far-field Imaging: Principles

Numerous techniques are currently available that allow to acquire the image of a real object either in 2D or in 3D. One of the most commonly used technique in medicine and in biology is the X-rays imaging **[1]**. Especially, in the case of breast cancer detection, senology and tomosynthesis are widely used X-rays based techniques towards early detection of breast carcinoma. The spatial resolution of such kind of techniques is limited by the Rayleigh scattering that is by about 50 μm for X-rays. Although the spatial resolution is quite comfortable, X-rays are obviously and concomitantly ionizing due to the high energy carried by X-photons.

Sometimes, in order to ascertain the diagnosis from X-rays mammography, a patient can pass a nuclear magnetic resonance (NMR) examination with or without the injection of contrast agent. NMR is particularly sensitive to the inner water-content **[2; 3]**. However, the spatial resolution is limited to 0.5 mm, approximately **[4]**. It is worth reporting additional imaging technique that present greater spatial resolution such as optical coherent tomography **[5; 6]**. Nevertheless although the spatial resolution is effectively greater by employing such an imaging technique, the penetration depth is drastically reduced in comparison to the above techniques as it ranges from few to a dozen of millimeters.

In view of the foregoing chapters, terahertz radiations are promising candidates to complete the set of tools that is used to detect, to remove and to cure the breast cancer that is still -at the time this manuscript is written- the first cause of death by cancer among women in the world. As a reminder, terahertz waves have shown specific sensitivity to water-content that reveals their usefulness on dealing with living organisms made up of roughly 70% of water. Therefore, tissues that exhibit varying water-content over their structure exhibit high contrast between areas.

III.2.1 Image & Limitation

Terahertz imaging is nothing else than recording waveforms from different and adjacent locations. By doing so, one has access to a volume of information that depicts for each location the interaction with the sensing terahertz field. A schematic drawing is given in Figure III.1 that illustrates the statement.

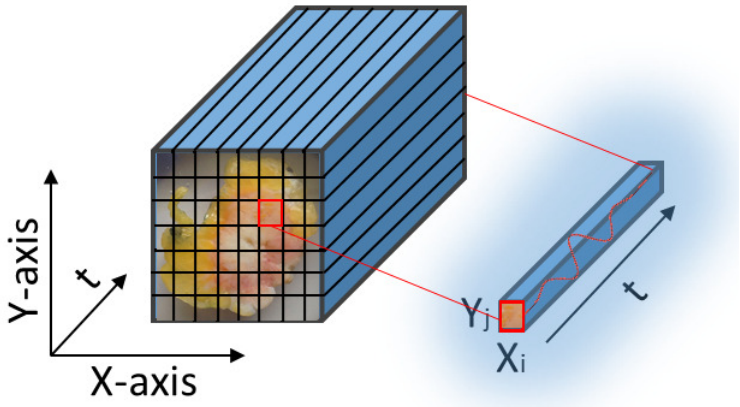


Figure III.1: Schematic representation of the information volume acquired during 2D material imaging. The third axis, t represents the time.

The first concern that arises when talking about imaging is the spatial resolution. Spatial resolution defines the ability of an imaging technique to distinguish between two locations that are physically separated by a given distance. Globally, the spatial resolution is linked to the smallest achievable focus of the sensing beam. Alternatively, it might be said to be dictated by the ability to concentrate the probing energy within the smallest surface or volume as possible. The smallest diameter it is possible to focus down a beam is defined by the Abbe's limit **[7]**:

$$d = \frac{\lambda}{2\mathrm{NA}}, \tag{III.1}$$

where d is the size of the focus diameter, λ is wavelength, and NA is the numerical aperture of the imaging lens. One can see that the Abbe's limit does not permit imaging systems to distinguish between two locations separated by a lateral distance that is smaller than half of the wavelength of the light that has been used. The numerical aperture of the imaging lens can alternatively be defined as:

$$\mathrm{NA} = n_0 \sin\theta_0, \tag{III.2}$$

where n_0 is the refractive index of the medium from which is observed the inspected sample and θ_0 the converge angle of the probing beam. Therefore, it is possible to reduce the spot size either by sensing with smallest wavelength and consequently with more energetic photons, or by immersing the imaging setup into a higher refractive index medium or finally by using imaging lens that has higher numerical aperture. Concomitantly we have underlined an important point: as terahertz pulses are used within this work, they carry broadband frequencies. Therefore, in view of the Abbe's limit, regardless of the numerical aperture and the refractive index of the medium, high frequencies that are contained in generated pulses will consistently exhibit better resolved images. To clarify the comment, the graph in Figure III.2 depicts the Abbe's limit factor d in free space as a function of the frequency ranging from 200 GHz to 4 THz, for a beam convergence of $\theta_0 = \frac{\pi}{6}$.

As shown in Figure III.2, for the given convergence angle, the resolution is limited to approximately 300 μm at 1 THz. Imaging systems whose spatial resolution is limited by the diffraction limit are referred to as far-field imagers and are subject to equations set out by Fraunhofer's approximation. Consequently, it is not possible to achieve cellular resolution (few microns) with such kind of far-field terahertz systems. However, it already offers a sufficient spatial resolution for some applications including pre-clinical studies towards breast cancer detection. On breaking the diffraction limit, it is possible to reach resolution by about $\frac{\lambda}{100}$ **[8]**. Imaging with a spatial resolution that is lower than the diffraction limit takes place in the context of near-field imaging and the correlated Fresnel's approximation. Further considerations are discussed in the subsequent

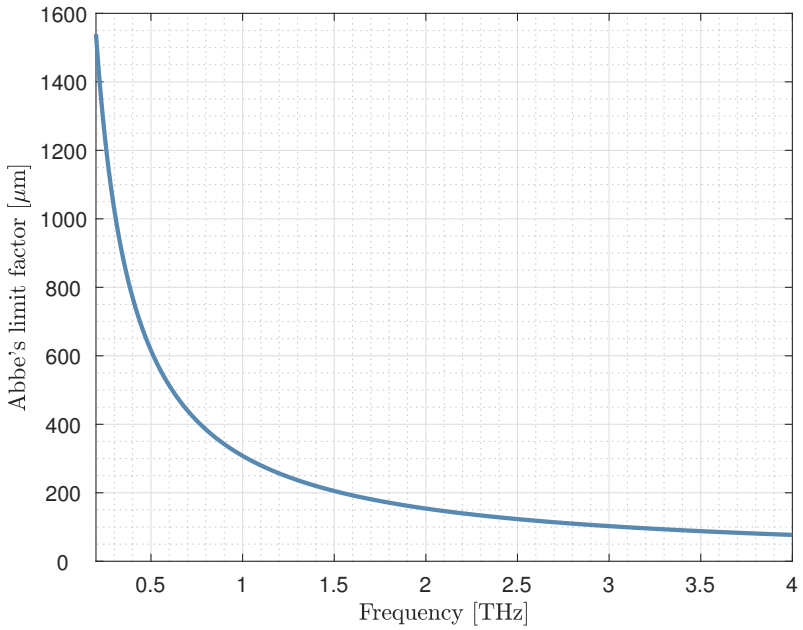


Figure III.2: Abbe's limit between 200 GHz and 1 THz in free space for $\theta_0 = \frac{\pi}{6}$.

chapter.

Overall, far-field terahertz imaging setups are equivalent to those that were reported in the chapter on spectroscopy. Probing different location requires -obviously- the implementation of motorized stage that aim to either displace the sample or the optical path and the related optics.

III.2.2 Effective Spatial Resolution

The spatial resolution of the Terapulse 4000 -used in this work- in reflection configuration was assessed with a USAF-1951 test chart whose generic design is provided Figure III.3

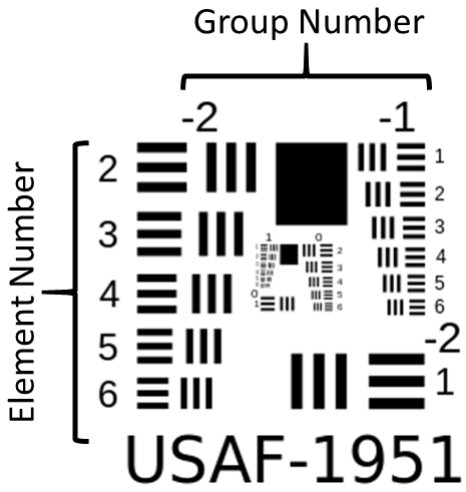


Figure III.3: Resolution test-target.

Typically, the resolution test chart is made of several sets of lateral and longitudinal lines that have specific distances that get smaller and smaller. For a scanner, a USAF 1951 chart is scanned and the effective resolution is the smallest distance for which the scanner can discriminate between different lines. It is worth mentioning that determining the effective resolution of imaging systems in such a way is overly subjective and rough. In fact, the effective resolution is ultimately

determined by the careful examination of the image by the analyst. Each element that belongs to a group possesses specific width (mm) and spatial frequency (line pair per mm) that are listed in Table III.1.

Table III.1: Thickness and spatial frequency of each element of the USAF-1951 test-chart.

Group	-2	-1	0	-2	-1	0
Element	Width (mm)			Frequency (LP/mm)		
1	2.00	1.0	0.50	0.25	0.50	1.00
2	1.78	0.89	0.45	0.28	0.56	1.12
3	1.59	0.79	0.40	0.32	0.6	1.26
4	1.41	0.70	0.35	0.35	0.70	1.41
5	1.26	0.63	0.31	0.40	0.79	1.59
6	1.12	0.56	0.28	0.45	0.89	1.78

Generally, optical system resolution determination are performed with a USAF 1951 test-chart made of glass. However glass drastically absorbs terahertz radiations and thus cannot provide the transparent material that is expected to make up the test-chart background. Indeed, the system resolution will be determined from the highest spatial frequency for which, the contrast between the background and element is greater than a given threshold value of 0.3. One of the common way to define the contrast interval C is **[9]**:

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}, \tag{III.3}$$

where I_{max} and I_{min} are the maximum and minimum intensity. It is worth mentioning that the scanner system that is used informs on the amplitude of the electric field rather than on its intensity. Therefore, one should consider the square of the recorded signals to express the contrast. The USAF 1951 test-chart was adapted for terahertz frequencies with a transparent polymer background and elements made of copper. Basically, it was realized from a USAF 1951 masked printed circuit board submitted to ultra-violet radiations.

With a temporal window by about 33 ps sampled every 8 fs and an imaged area of 70×70 mm^2 corresponding to 350×350 pixels (step in x and y coordinates of 0.2 mm), 501 760 000 points were recorded **[10]**. The corresponding sampling frequency F_s was therefore by about 30 GHz. In Figure III.4, is represented the minimum resolved spatial frequency in line pair per millimeter as a function of the frequency that satisfies the fixed value of 0.3 for the contrast.

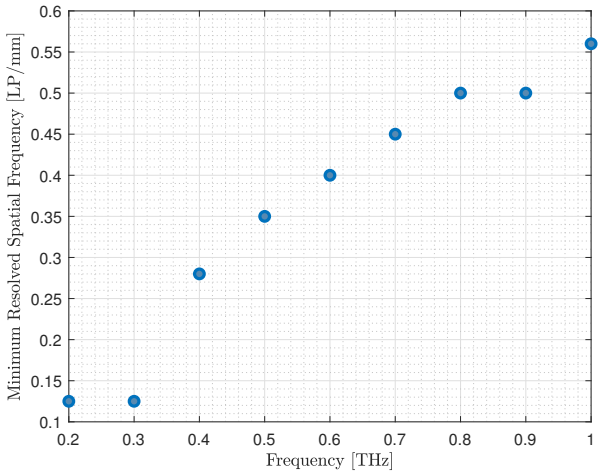


Figure III.4: Minimum resolved spatial frequency corresponding to at least a contrast of 30% as a function of the sensing terahertz frequency.

The total acquisition time was about twenty hours to record approximately a volume of 2 gigabytes. The adapted test-chart images obtained at each even hundred gigahertz are shown in Figure III.5. The minimum spatial frequency that provided a contrast of 0.3 at each frequency is depicted by a dashed red square.

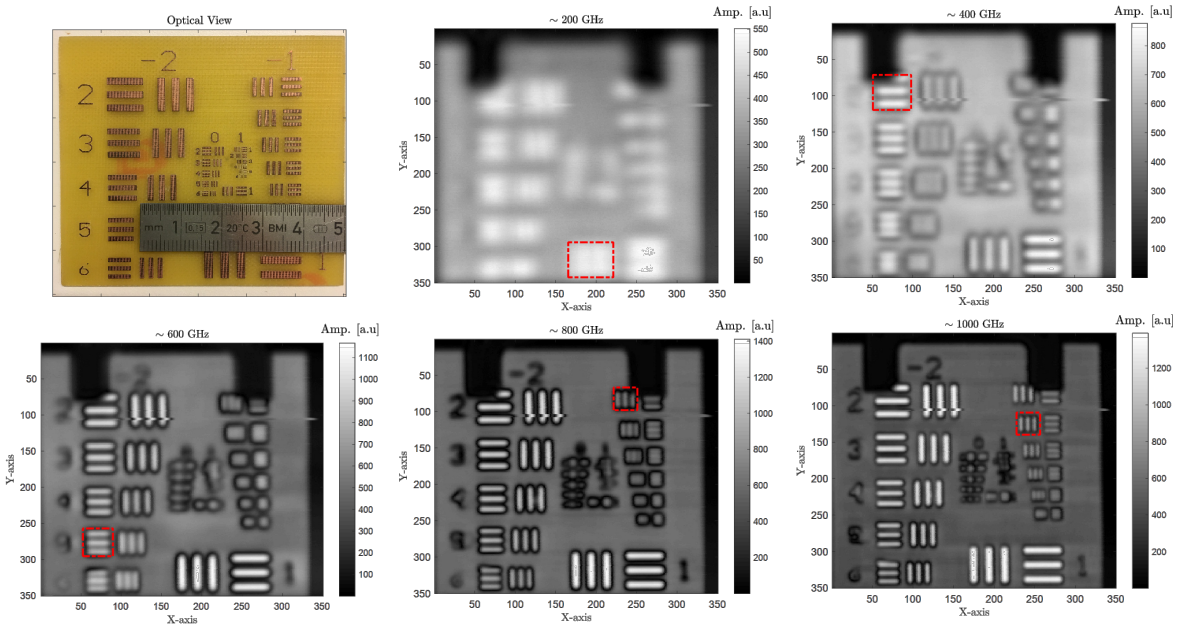


Figure III.5: Smallest resolved line pairs (dashed red square) for 200, 400, 600, 800 and 1000 GHz. The smallest resolved spatial frequency at 200 GHz was assumed to be half the spatial frequency of group -2, element 1.

III.3 Self-Reference Image Inversion

In *Chapter II* and in the beginning of the *Chapter III*, it was reported how difficult it is to ascertain the accurate location match between the sapphire substrate that supports the tissues to be inspected and the reference signal provided by a perfect mirror. Importantly, the mismatch between these two locations may alter the profiles and the values of the intrinsic properties that are extracted. Indeed, on slightly modifying the global optical path between the reference and the probed material, one concomitantly alters both the signal phase and the sensing energy flux that reaches a given location. Effectively on shortening or prolonging the optical path between reference and sample, one may attribute to the sample, properties that reflect either the slowdown or the speed of light. One way to correct the position mismatch would consist in aligning numerically the reference and the sample signals such that, the reflected peak from the mirror precisely matches the temporal position of the first reflected peak from the sapphire window. However, by doing so, one only corrects the phase error between the two signals. In fact, we previously stated that modifying the optical path also alters the flux of energy that reach each spatial location. Therefore, aligning numerically the two signals to “clean” the phase error would therefore transfer the position error to the signal amplitude, solely. It is hence not reasonable to proceed this way and is often inefficient. Some signal correction methods were proposed and employed to improve sample characterization. These methods mainly consist in correcting the sample signal based on the signal recorded from water (Figure III.6).

Variations between the recorded signal from water with the one derived from a simulation using the measured reference are therefore used to estimate a degree of signal correction and is subsequently applied to the sample signal **[11–14]**. However, such methods are still recording the reference signal from a perfect reflector.

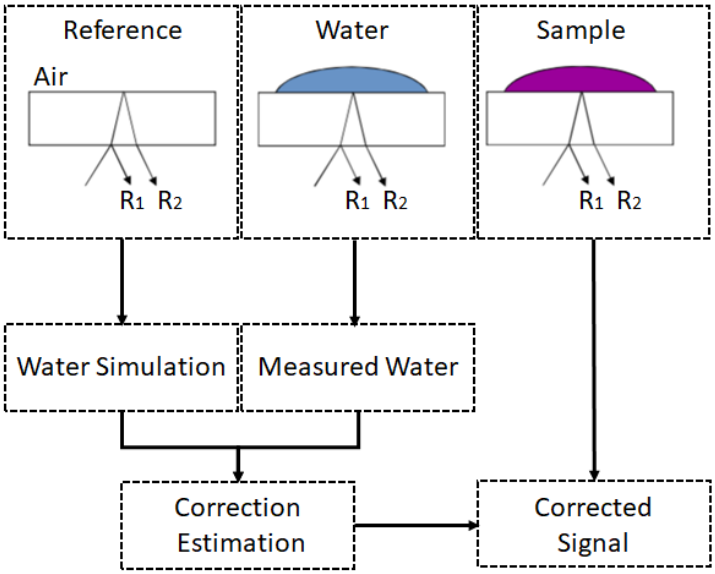


Figure III.6: Flow chart to illustrate the procedure to improve sample characterization. R_1 and R_2 depicts the lower and upper reflected pulses at dielectric interfaces.

To overcome the previously mentioned issues, a self-reference method that allows to determine the intrinsic properties from each pixel of a terahertz image without recording a reference signal was developed and is, hereafter, described. Globally, the self-referenced method to extract the dielectric properties that underlie each pixel can be decomposed in four steps: (i). to image the second dielectric interface between sapphire and either the air or the tissues (ii). to calculate from a sapphire/air pixel the theoretical reference signal; (iii). to denoise the calculated signal by Daubechie's wavelets; (iv). to input back the calculated reference into an inverse electromagnetic problem. We propose to go through all these steps to enlighten the key concepts behind the developed self-reference imaging.

III.3.1 Second Dielectric Interface Imaging

The overall idea that lies behind self-referencing, is that the necessary information to extract the dielectric properties of biological tissues is solely located at the second dielectric interface that is, either a sapphire/air or a sapphire/tissue interface. Figure III.7 shows the plane that is aimed to be sensed by the terahertz scanner. Obviously, as the sensing field is not monochromatic, the recorded signal arising from the second dielectric interface is not a straight line. Quite the contrary, such a signal is expected to broaden over few picoseconds. Hence, rather than just belonging to a plane, the wished information is comprised in a cuboid whose thickness corresponds to the acquisition time window divided by the sampling period. Globally, the entire time waveform is recorded over 6 picoseconds. Therefore, the terahertz image is a volume whose size is given by $N_X \times N_Y \times T$; with N_X and N_Y the number of steps along the x and y coordinates respectively, and where T depicts the time samples at the output of the THz system. The time-domain waveform of the imaged volume is then computed in frequency domain by means of fast Fourier transform algorithm.

III.3.2 Reference Retrieval

First, one wants to select a pixel that corresponds to an interface between the sapphire substrate and free space. Figure III.8 illustrates one adequate pixel that would allow the reference retrieval.

Theoretically, the mathematical formulation that allows one to express the waveform of the retrieval pixel is as follows:

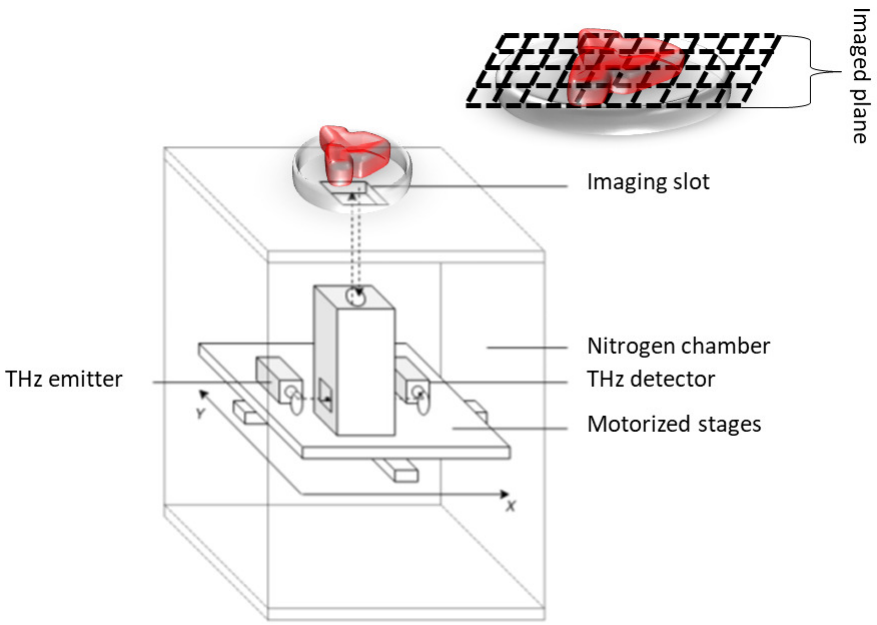


Figure III.7: Schematic illustration of the plane of interest that contains the information about the tissue sample.

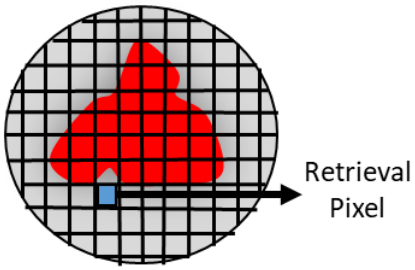


Figure III.8: Example of a typical pixel that allows the reference retrieval.

$$\frac{\mathrm{E_{RP}}(\omega)}{\mathrm{E}_{Ref}(\omega)} = t_{1,2}(\omega) \times r_{2,1}(\omega) \times t_{2,1}(\omega) \times exp^{\frac{-2i\omega\hat{n}_2 d_2}{c}}, \tag{III.4}$$

where $\mathrm{E_{RP}}$ is the waveform of the retrieval pixel, E_{Ref} is the reference waveform, t and r are the transmission and reflection Fresnel's coefficients respectively, ω is the angular frequency, $\hat{n}$ is the complex refractive index, d_2 is the sapphire substrate thickness, c stands for light velocity in free space and where the index 1 and 2 corresponds to the first and second medium, namely free space and sapphire. Therefore, one can derive the corresponding reference waveform that was emitted by the photoconductive antenna:

$$\mathrm{E}_{Ref}(\omega) = \frac{\mathrm{E_{RP}}(\omega)}{t_{1,2}(\omega) \times r_{2,1}(\omega) \times t_{2,1}(\omega) \times exp^{\frac{-2i\omega\hat{n}_2 d_2}{c}}}. \tag{III.5}$$

Unsurprisingly, considering a pixel that corresponds to an interface between sapphire and free space, the right-hand terms in equation III.4 are not unknown parameters if wisely expressed. Let's remind the expression of both the transmission and reflection Fresnel's coefficients at an interface

between a medium i and a medium j:

$$t_{i,j}(\omega) = \frac{2\hat{n}_i(\omega)\cos\theta_i}{\hat{n}_i(\omega)\cos\theta_i + \hat{n}_j(\omega)\cos\theta_j}, \tag{III.6a}$$

$$r_{i,j}(\omega) = \frac{\hat{n}_i(\omega)\cos\theta_i - \hat{n}_j(\omega)\cos\theta_j}{\hat{n}_i(\omega)\cos\theta_i + \hat{n}_j(\omega)\cos\theta_j}, \tag{III.6b}$$

where $\hat{n(\omega)}$ is still the angular-frequency-dependent complex refractive index and θ is the incidence angle. For simplicity, it is useful to introduce $X_i(\omega) = \hat{n}_i(\omega)\cos\theta_i$. Hence we can reexpress the three terms $t_{1,2}$, $r_{2,1}$ and $t_{2,1}$ that are present in equation III.4 in a lighter way:

$$t_{1,2}(\omega) = \frac{2X_1(\omega)}{X_1(\omega) + X_2(\omega)}, \tag{III.7a}$$

$$r_{2,1}(\omega) = \frac{X_2(\omega) - X_1(\omega)}{X_1(\omega) + X_2(\omega)}, \tag{III.7b}$$

$$t_{2,1}(\omega) = \frac{2X_2(\omega)}{X_1(\omega) + X_2(\omega)}. \tag{III.7c}$$

On remarking that the frequency-independent complex refractive index of free space is $\hat{n}_1 = 1 - 0i$, and thanks to Snell's law that states $n_i \sin\theta_i = n_j \sin\theta_j$, we can easily show that:

$$X_1 = \cos\theta_1 \tag{III.8a}$$

$$X_2(\omega) = \hat{n}_2(\omega)\cos(\sin^{-1}(\frac{\sin\theta_1}{n_2(\omega)})) \tag{III.8b}$$

$$X_3(\omega) = \cos(\sin^{-1}(\sin(\sin^{-1}(\frac{\sin\theta_1}{n_2(\omega)}))\hat{n}_2(\omega))) \tag{III.8c}$$

It is worth mentioning that θ_1 can be determined from the data sheet of the experimental setup. In this work, $\theta_1 = 8°$. Therefore, knowing, θ_1 and $\hat{n}_2$ (that was determined in the previous chapter) and the sapphire thickness d_2, one can retrieve the theoretical electric field that sensed the tissue by solving equation III.4. Eventually, one can additionally compute the first reflected peak from the sapphire substrate.

III.3.3 Wavelet Denoising

The reference signal that is calculated back may be noisy. Indeed, deriving the waveform of the sensing field, we concomitantly propagated the noise that affects the retrieval pixel waveform. However, it is worth to mention that the retrieval pixel is impacted by the noise that arises from the terahertz detector which should not impact the emitted terahertz wave especially above 1 THz. Therefore, it is advised to approximate a denoised reference. To do so, we assume that the denoised reference $E_{Ref}(\omega)$ can be modeled as:

$$E_{Ref}(\omega) = E'_{Ref}(\omega) + G, \tag{III.9}$$

where $E'_{Ref}(\omega)$ is the noisy calculated reference and G is the noise contribution, arising from the presence of noise in the experimental reference pixel signal. It is important to note that Gaussian white noise remains Gaussian and white for all orthogonal basis. The idea behind denoising is to extract E_{Ref} without the contribution from G. The reference signal denoising was performed with wavelets using thresholding estimators **[15–18]**. A soft thresholding rule was preferred compared to the hard one to ascertain the regularity of the denoised signal. The threshold value for

feature estimators was preferably set to a low value to limit sample information removal as best as possible. Daubechies wavelet with six vanishing moments (db6) -and the corresponding scaling function- were chosen as they best correlate with the waveform being analyzed [**19**]. Let's underline that there is no fundamental rules to select the mother wavelet to be used. The db6 was selected as it overall matches with the input signal waveform from eyeball inspection. Neighboring wavelets may also have been selected as mother wavelets for signal denoising. The scaling function and the shape of db6 are given in Figure III.9.

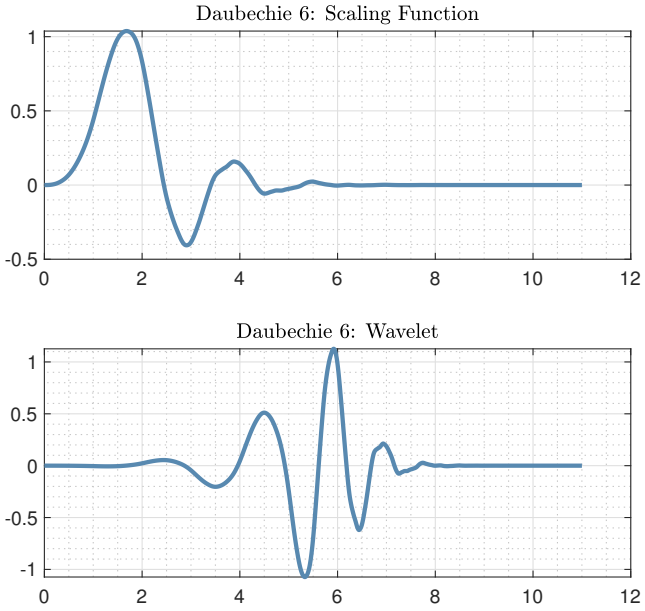


Figure III.9: Scaling function and shape of the db6 wavelet.

Before inputting back the reference into an inverse electromagnetic problem to extract the physical properties, the exact same denoising procedure is applied to the entire set of waveforms contained in the volume $N_X \times N_Y \times T$. To summarize the first three steps, Figure III.10 emphasizes the global numerical procedure.

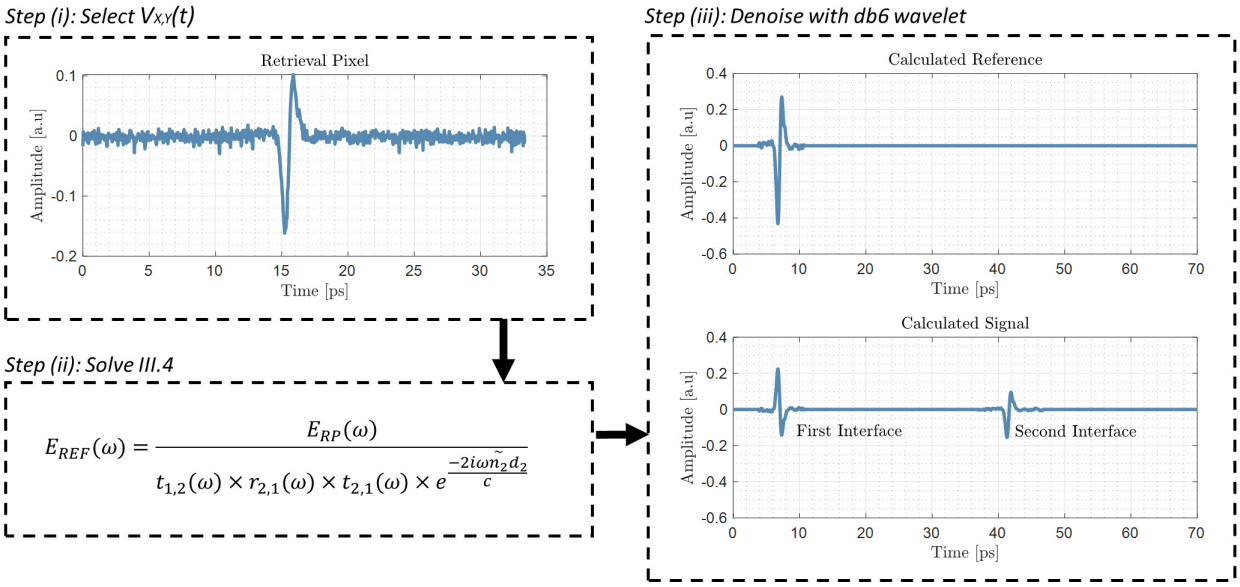


Figure III.10: Block diagram of the self-reference method. (i). A retrieval pixel is selected; (ii). Signal inversion is performed to calculate the reference signal; (iii). The reference signal is denoised by means of db6 wavelet. Eventually, the complete signal of the retrieval pixel arising from the interaction of the calculated reference with the two dielectric interfaces is retrieved. Note that the time-scale is relative rather than absolute.

III.3.4 Image Inverse Electromagnetic Problem

In view of the quantity of information that is contained in an image, it is wise to limit the region over which has to be solved the inverse electromagnetic problem. Indeed, only pixels that correspond to an interface between the sapphire and the tissue are of interest. In other words, it is -obviously- not necessary to perform the inversion for pixels that are related to a sapphire/air interface.

In order to limit the region over which one wants to perform the inverse electromagnetic problem, a contouring algorithm is necessary to define the tissue region on the image. Contours of a 2D scalar field are called isolines and are made of a set of points for which the scalar field s as a constant value c:

$$x \in \Re^2 : s(x) = c. \tag{III.10}$$

For example, on considering a pixel that can be represented as a quadrangle cell, one can define four locations that are called pointel: $(0;0), (1;0), (0;1), (1;1)$ where the scalar field takes the respective values $S_{00}, S_{10}, S_{01}, S_{11}$ (Figure III.11). The pointels are elements of dimension 0 (point) and are linked with each others by linels that are elements of dimension 1 (segment).

From the value taken by S_{00}, S_{10}, S_{01} and S_{11}, one can estimate the value of the scalar field at each location within the quadrangle cell by interpolating. Rather than just interpolating in one direction between the nearest neighbors, one has to interpolate in the two directions as the scalar field depends both on x and y coordinates. Such a bidirectional interpolation is often referred to as bilinear interpolation but has however a quadratic form:

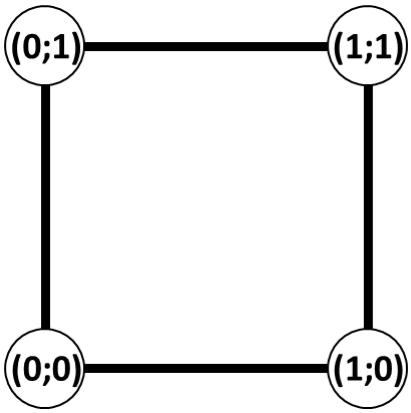


Figure III.11: Representation of a pixel with the four associated pointel (0;0), (0;1), (1;1), (1;0) and the four linels joining them.

$$s = xy(\mathrm{S}_{11} + \mathrm{S}_{00} - \mathrm{S}_{10} - \mathrm{S}_{01}) + x(\mathrm{S}_{10} - \mathrm{S}_{00}) + y(\mathrm{S}_{01} - \mathrm{S}_{00}) + \mathrm{S}_{00} \tag{III.11}$$

$$= \mathrm{A}xy + \mathrm{B}x + \mathrm{C}y + \mathrm{D}. \tag{III.12}$$

Obviously, if A = 0, the corresponding contour is a line. Nevertheless, it is rarely the case and most of the time, contours are hyperbola whose formulation is dictated by the coefficient in equation III.12:

$$s = \mathrm{A}(x + \frac{\mathrm{C}}{\mathrm{A}})(y + \frac{\mathrm{B}}{\mathrm{A}}) - \frac{\mathrm{BC}}{\mathrm{A}} + \mathrm{D}. \tag{III.13}$$

Additionally to those mathematical statements, one should keep in mind that isolines are always orientable, non-intersecting and closed.

An infinite number of contours can be determined using such kind of approach. The choice of the isoline that will be used to define the region within which the physical properties are going to be extracted remains subjective. Indeed, on carefully comparing the original terahertz image and the different contour levels, the analyst has to determine the one that suits the most, *i.e.* that best corresponds to the visualized tissue area. An example of the decreased amount of pixels to be considered in further inverse electromagnetic problem, thanks to contouring procedure, is given in Figure III.12.

The pixels that are comprised within the selected level are submitted to the inverse electromagnetic problem. On the other hand, the optical constant $n - i\kappa$ for pixels that are not comprised within the chosen contour is set to $1 - i0$. Globally, the objective function to be minimized during the upcoming procedure is equivalent to the one that was defined in *Chapter II*, section related to inverse electromagnetic problem, that is:

$$\chi(\omega) = \delta\mathrm{M}^{\mathrm{T}}(\omega)\delta\mathrm{M}(\omega) + \zeta\delta\Phi^{\mathrm{T}}(\omega)\delta\Phi(\omega), \tag{III.14}$$

where $\delta\mathrm{M}(\omega)$ and $\delta\Phi(\omega)$ are the amplitude and phase penalty respectively, and ζ is a regularization parameter weighting phase penalty regarding amplitude.

However, since the propagating pulses that are recorded are not expected to travel through the tissue before being detected, but rather to be simply reflected at the second dielectric interface and ultimately to be detected, it is reasonable to state that the propagation velocity of terahertz pulses are not going to be highly affected by the dielectric properties of tissues. Therefore, the objective function to be minimized was reduced to:

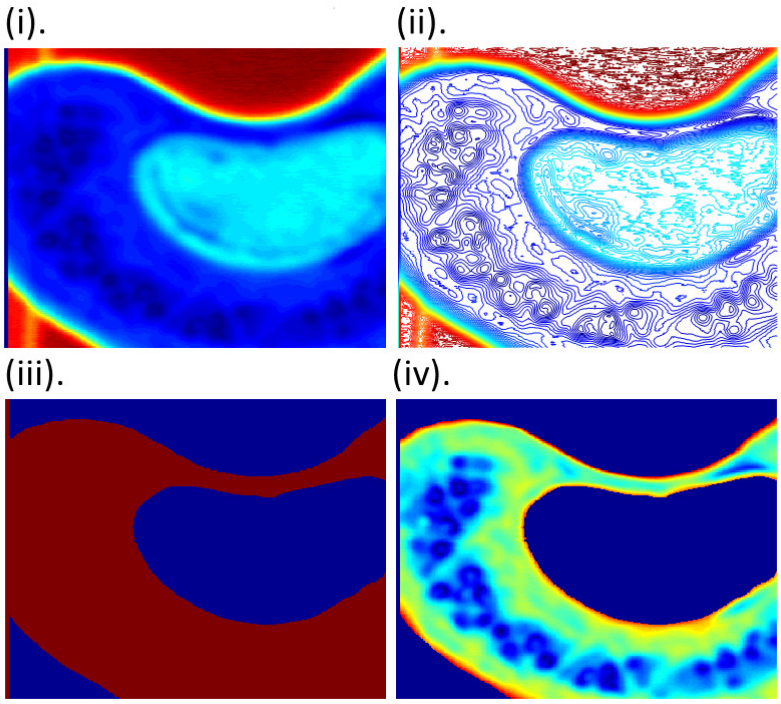


Figure III.12: Illustration of the contouring procedure to reduce the number of pixels treated by the inverse electromagnetic problem. (i). Raw terahertz image of a sample; (ii). Contour levels extraction; (iii). Enclosed surface associated to a selected contour level; (iv). Delineation of the region to be treated by inverse electromagnetic problem. In the example, the number of pixel to be considered was reduced from 101,440 pixels to 56,296 pixels.

$$\chi(\omega) = \delta \mathrm{M}^{\mathrm{T}}(\omega)\delta \mathrm{M}(\omega). \tag{III.15}$$

Furthermore, it was shown in the previous chapter that differences between healthy fibrous tissues and cancerous ones in extinction were nonexistent. One can still state that interesting demarcations between adipose tissues and other tissue types were pointed out using the extinction coefficient. However, the refractive index was also exhibiting differences between these tissues. For these reasons, investigations on property mapping are focused on the real part of the optical constant, *i.e.* the refractive index.

III.4 Image Registration

Property maps that are derived from the procedure heretofore described must demonstrate their compliance to the histological examination. In other words, the maps that are aimed to classify tissue regions as a function of the exhibited refractive index must be conform to the technique that is approved by the medical profession. Therefore, it is compulsory to compare between pathology images and terahertz images.

III.4.1 Pathology Image Issues

Terahertz images and histological ones are obtained from the same excised tissue. However image acquisition is performed under different modalities. Indeed, while the terahertz images are obtained by scanning freshly excised tissues, the pathology cliché is obtained after the histological treatment chain. To obtain the pathology image the tissue is first paraffin embedded then is sliced in sections of few micrometers, immersed subsequently in different alcohol baths and is

finally colored with hematoxylin and eosin. Ultimately, the inked section is evaluated under optical microscope. Hence, the terahertz images may differ from their histological alter ego. Indeed, the embedding, the sectioning and the desiccation alter the global structure of tissues. There is actually a lot of issues that have been reported to affect the accuracy of histology slides. These alterations are collectively named as artefacts. Artefacts include, loss of tissue area and details, folds and wrinkles, cracks and holes. These alterations could result in misinterpretation as they are modifying the morphology tissue structures. In his thesis, Abiodun Ogunleke who worked toward 3D pathology image creation from 2D pathology images **[20]**, reported the most commonly encountered issues when working with 2D pathology slides (Figure III.13).

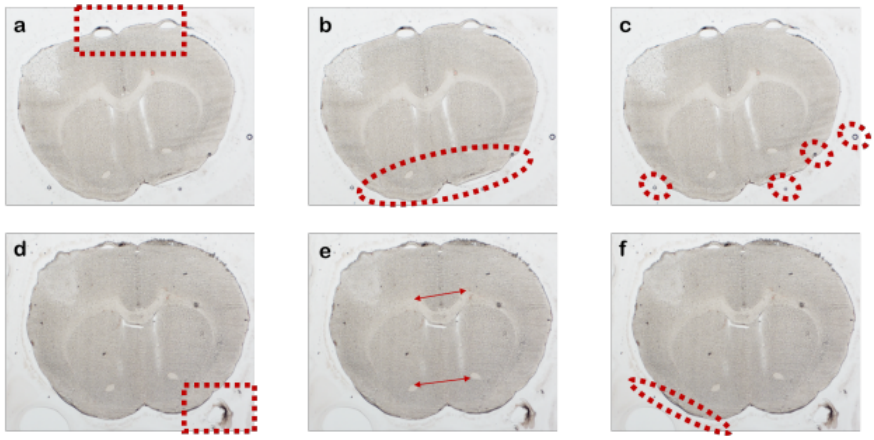


Figure III.13: Problems associated with 2D histological images. (a) Presence of tears and holes. (b) Global shape deformation and missing information. (c) Artefacts in and around. (d) Glue deposited on sample slide. (e) Local misalignment of tissue substructures. (f) Vignette arising from difference of thickness around tissue edges. Reproduced with the courtesy of Abiodun Ogunleke **[20]**.

Ultimately, these alterations may drastically limit the evaluation of the terahertz classification compliance. However, histology slides are the only available reference clichés that allow one to establish the sensitivity of its classification method. Overall, there are two ways to overcome these issues: to correct the histology slides at risk of adding artificial information and altering even more the global structure or to compare directly the terahertz image with the raw pathology image at risk of underestimating the accuracy of the techniques used for classification. The first way would require to morph the pathology image to correspond to the terahertz picture: image pre-processing to resolve the problems introduced by histological sample preparation, image information retrieval process where useful information such as anatomical landmark is determined and extracted in the histological images, post-process where the extracted anatomical landmarks are refined for reconstruction. Some numerical procedures to do so were already reported in the literature. However, these methods are cumbersome and the evaluation of the reconstruction is often complicated as no perfect reference pathology image exists. As terahertz technology is quite new towards breast cancer detection, it was preferred to follow the second approach at risk of underestimating the accuracy of tissue classification. In fact, the idea of this work is not to promote terahertz techniques as a marvellous tool for cancer detection but rather to provide some preliminary statements on terahertz capabilities. Therefore, the compliance evaluation between refractive index maps and pathology slides is performed directly with altered pathology slides.

III.4.2 Registration Procedure

In addition to previous items, let's remind that we stated in the foregoing that the resolution of any imaging system is intrinsically dictated by the sensing field wavelength. It is clear that the resolution with optical microscope is by far greater than the one that can be achieved using millimeter radiations.

Last but not least, the tissue sample orientation in both the terahertz image and the pathology image are expected to be different as their are not evaluated at the same angle and location. Thus, the histological image exhibits a tissue sample that is twisted and shifted compared to the terahertz one.

Therefore, pixel by pixel comparison is not possible as it stands. In fact, one has two images of $N_X \times N_Y$ and $M_X \times M_Y$ pixels, of different size, resolution and orientation. To evaluate the compliance of the terahertz images one has therefore to mask the pathological image, to resize the derived mask to match the terahertz cliché and to align the two pictures by translation and rotation. Importantly, the process has to be performed in view and only in view of the tissue contour of the two images and not the internal features to avoid unintentional bias in the morphing results that would alter the compliance evaluation. The different steps that allow to ultimately compare the histological diagnosis and the terahertz picture are hereafter described.

Contouring

The contouring of the pathology image is performed in the exact same manner than the one described to define the contour of the terahertz image (*Chapter III.3.4 Image Inverse Electromagnetic Problem*). For instance, the function *contour(·)* or *imcontour(·)* from MATLAB® can be used.

Resizing

As mentioned, the resolution of the pathology images is by far greater than the ones obtained with terahertz imaging. Therefore, it is substantial to reduce the resolution of these clichés. To do so, one reduces the resolution down to the resolution of terahertz images, by performing a bicubic interpolation. Contrary to the bilinear interpolation, that was chiefly described in the foregoing where interpolation is approximated based on the four nearest pixels, bicubic interpolations takes into account a neighborhood of sixteen pixels. Therefore, using bicubic interpolation, the subsampled histology slide is overall less nonsmooth than using bilinear interpolation. Moreover, cubic refers to the degree of equations that govern the definition of the contour lines. Importantly, the bicubic interpolation was not developed during this work. The function *imresize(·)* from MATLAB® was chosen.

Reorientation

The alignment of the morphed pathology and the terahertz image is realized to ascertain that comparison actually compares the same regions. First, the contour of the morphed pathology and the terahertz image are computed. The contour of the histology mask is manually and progressively shifted in x and y axis to bring it closer to the terahertz mask contour. Once the coordinates approximately match, the pathology contour is iteratively rotated within a loop that calculates the correlation between the two matrices. Basically, it consists in determining the Pearson's linear correlation coefficients. The rotation angle providing the higher positive correlation (close to unity) is selected and the terahertz image is correspondingly rotated.

An imaged flow chart of these three pre-treatment, namely contouring, resizing and reorientation for image registration is provided in Figure III.14.

Pathology Masking

Since histological slides are stained with hematoxylin and eosin, the pathology image is colored and therefore is obviously not binarized. The classification of tissue sample regions in either benign or malignant group is established by the careful examination of the pathologist. The pathologist draws the contour of the malignant area that often comprises the stroma (tumor nutrient medium). In this work, histology slides were scanned and evaluated with either NDP.view2 software from Hamamatsu® or Aperio ImageScope software from Leica®. Both these softwares are

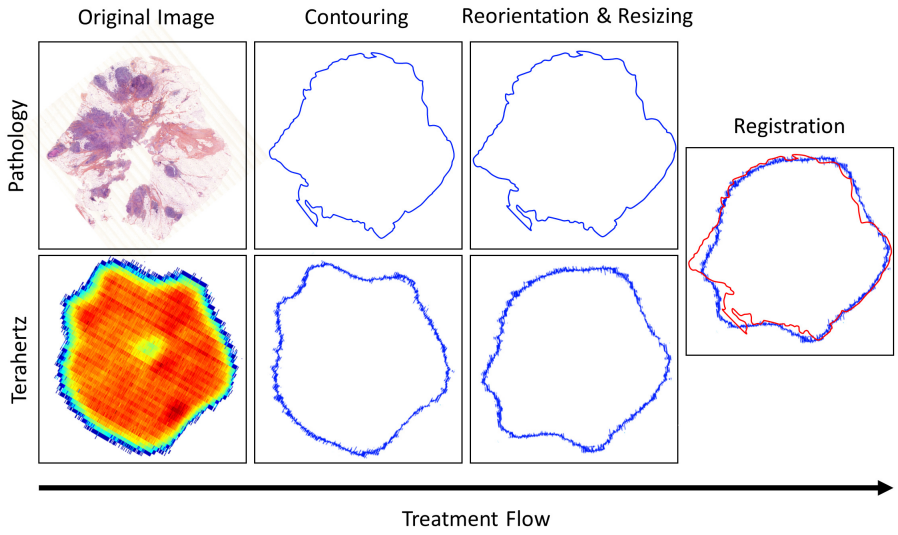


Figure III.14: Representation of the registration process. The terahertz and pathology original images are treated by a contouring algorithm to extract the edges of each image. The pathology image resolution is reduced to match the lower resolution of the terahertz image. Then, the terahertz image is reoriented with respect to the pathology image until the best correlation is achieved. These modifications are therefore applied to the refractive index map and the pathology mask.

pathology slide viewers on which it is possible for the pathologist to delimit manually tissue sample regions and to annotate them. From the interpretation of the pathologist, one can therefore binarize the histological image where each pixel is either classified as benign or malignant.

III.5 Correlation

The regions examined by the pathologist on the histology slide have been classified to depict either benign or malignant tissues. However, refractive index maps are obviously not binary maps but are rather expected to provide a wide variety of refractive index value for each frequency image. Therefore, one has to define classification models such that terahertz pixels are classified in one of the possible categories, namely malignant and benign. From these classifications, one can compare the reference image (the histology one) and the terahertz binary refractive index maps at each frequency.

III.5.1 Classification Models

From the refractive index maps, one can define a wide variety of decision chains that aim to classify the terahertz pixels either as malignant or as benign. In this work, two classification rules have been designed to evaluate the efficiency of the refractive index as a malignancy identifier.

It is important to acknowledge that other classification techniques may have alternatively been used to segment terahertz images, such as Markov chain Monte Carlo **[21]**, k-mean clustering **[22]**, support vector machine **[23; 24]**, probabilistic neural network **[25]** or deep neural network **[24]**. However, the need for a large amount of training samples make their application complicated, cumbersome and periodically inconsistent.

Hard Thresholding Approach

The simplest classification that can be realized is a hard thresholding. The analyst defines a threshold value t that halves the refractive index vector. Thereafter, pixels whose refraction index is higher or equal to the defined threshold are marked as malignant. Inversely, pixels that show a refractive index lower than the threshold are categorized as benign. Ultimately, it provides the corresponding binary map.

For instance, let $n_{x,y}$ be the refractive index of pixel (x, y) that was determined by solving the associated inverse electromagnetic problem and let t be the considered refractive index threshold. The global structure of the underlying algorithm, applied on image I whose height is I_h and width is I_w, is given as:

Algorithm 1 Hard Thresholding Classification

Require: $t \in \mathbb{R}$
  for (y = 1; y = I_h; y++) **do**
    for (x = 1; x = I_w; x++) **do**
      if $n_{x,y} \geqslant t$ **then**
        $(x, y) \equiv$ malignant
      else
        $(x, y) \equiv$ benign
      end if
    end for
  end for

It should be noted that **Algorithm 1** is not given in any specific programming language but rather describes in a schematic way the -overall simple- binarization process. Alternatively, Figure III.15 aims to depict such a binarization.

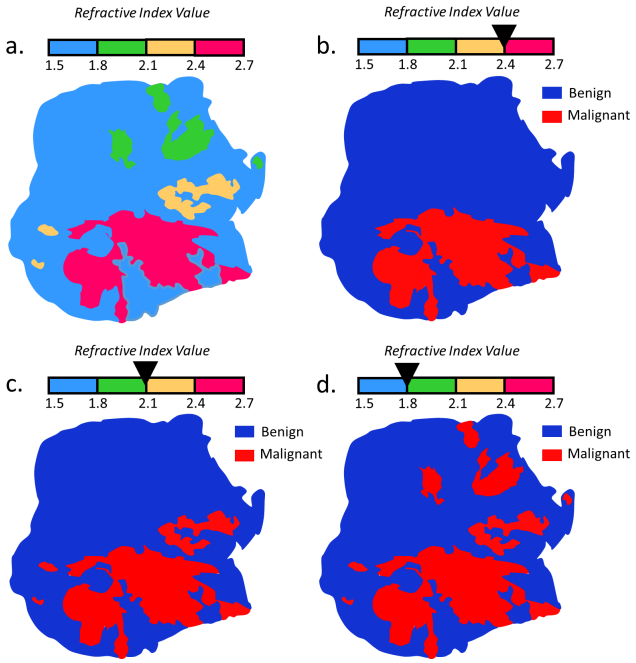


Figure III.15: Representation of the refractive index thresholding. a. Schematic refractive index map. b. Refractive index thresholding for $t = 2.4$. c. Refractive index thresholding for $t = 2.1$. d. Refractive index thresholding for $t = 1.8$.

Class Overlapping

Hard thresholding classification may however be less efficient than the way it appears on Figure III.15. In fact, a reader that is familiar with imaging would have pointed out that image scales (intensities, refractive index, etc...) are continuous rather than being discrete. Therefore, at the edge between two tissue types, a refractive index gradient should be observed, rather than a "strict jump" from one tissue type to another. This is partly due to the point spread function (PSF) that is intrinsic to the imaging system **[26]**. Chiefly, the PSF describes the impulse response of the imaging system to a punctual source. In other words, PSF dictates the deformation of an object

when imaged with an imaging system. Figure III.16 schematically represents the deformation of a point due to the PSF of a given imager.

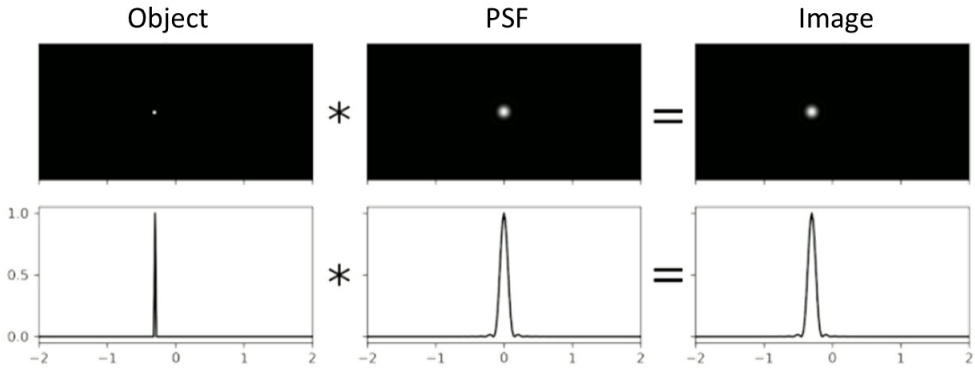


Figure III.16: Spreading of an object due to PSF. From left to right: punctual object, PSF and spread object image.

But why does PSF interfere in our case ? Let's imagine that we have two distributions of pixels, C_1 and C_2 that depict the amplitude modulus of malignant and benign pixels, respectively, at random frequency. Rather than having two well separated distributions, the two distributions, C_1 and C_2, likely overlap. Commonly, such an overlap is called class overlapping. Figure III.17 aims to depict such a fact.

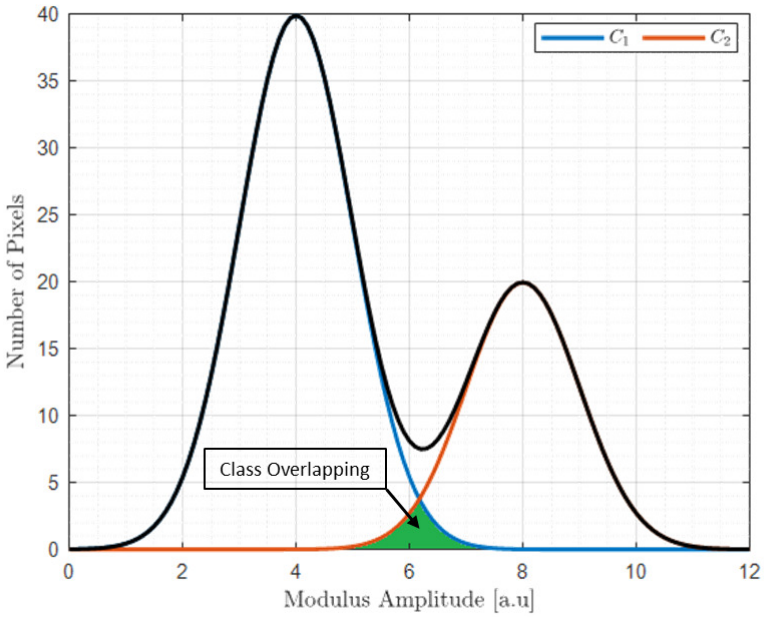


Figure III.17: Representation of the class overlapping for the modulus amplitude distribution C_1 (blue) and C_2 (orange). Green fills the overlapping region where it is not possible to clearly define pixel's belonging. Black curve represents the sum of the two distributions and therefore the observed global behavior.

And ? Well, the modulus amplitude (as well as the phase) of each pixels are used to derive the refractive index maps of samples. Therefore, replacing the modulus amplitude by the refractive index in the x-axis of Figure III.17 should give some clues: the class overlapping does not allow to determine to which category, either C_1 or C_2, pixels belong. We can reformulate as follows, are pixels located at the edges between a malignant and a benign area, benign or malignant ? Furthermore, how to define the edge span between such regions ? Rather than going through complex and cumbersome definitions of tissue region delineation, an alternative was found in region dilation. The idea lying behind is to consider the proximity between pixels in addition to their intrinsic refractive index.

Dilation for Connected Thresholding

To overcome the issue previously reported a constrained image dilation was employed. Roughly, the idea consists in classifying as malignant all pixels that are close to a pixel that was marked as malignant due to a refractive index higher than a defined threshold, even though they were previously defined as benign. Mathematically, let's consider X the ensemble of pixels that belong to the tissue sample imaged. Hence, we can consider that the coordinated (x, y) of each pixels are such as $x, y \in \mathrm{X}$. The region morphological dilation $\delta_\Sigma(\mathrm{X})$ with a structuring element Σ is given as:

$$\delta_\Sigma(\mathrm{X}) = \left\{x + \epsilon, y + \epsilon | \epsilon \in \Sigma, x \in \mathrm{X}, y \in \mathrm{X}\right\}. \tag{III.16}$$

It is therefore more appropriate to talk about diagnosis dilation rather than image dilation. As pointed out in (III.16) the dilation directly depends on the structuring element Σ. The most simple way to define Σ is the *cross* defined by:

$$\Sigma = \{(-1,0),(0,-1),(0,0),(0,1),(1,0)\}. \tag{III.17}$$

In the following, we will refer to this definition of the structuring element Σ as Σ^1, a structuring element of connectivity rank one. Two other structuring elements have been considered, one being of connectivity rank two, Σ^2 and one being a rank three, Σ^3:

$$\begin{aligned}
\Sigma^1 &= \{(-1,0),(0,-1),(0,0),(0,1),(1,0)\},\\
\Sigma^2 &= \{(-2,0),(-1,0),(1,0),(2,0),(0,-1),\ldots\\
&\quad (0,-2),(0,0),(0,1),(0,2),(-1,1),\ldots\\
&\quad (-1,-1),(1,-1),(1,1),\\
\Sigma^3 &= \{(-3,0),(-2,0),(-1,0),(0,0),(1,0),\ldots\\
&\quad (2,0),(3,0),(0,-1),(0,-2),(0,-3),\ldots\\
&\quad (0,1),(0,2),(0,3),(-1,1),(1,1),\ldots\\
&\quad (1,-1),(-1,-1),(-2,1),(-1,2),(1,2),\ldots\\
&\quad (2,1),(-1,2),(-2,1),(-1,-2),(-2,-1).
\end{aligned} \tag{III.18}$$

To describe these structuring elements more easily, Figure III.18 schematically represents both Σ^1 and Σ^2 structures.

Therefore, depending on the structuring element rank, a pixel is attributed to the malignant group if at least, one of the coordinates ϵ of the structuring element Σ^N reports a pixel whose refractive index is higher than the defined threshold t. Simply, pixels that are located in the influence area of a malignant pixel are in turn categorized as malignant. It is however of paramount importance to underline that the process is not iterative and is based on the hard thresholding diagnosis. Hence, the diagnosis dilation is not reiterate from the newly established binarization.

For instance, let $n_{x,y}$ be the refractive index of pixel (x, y) that was determined by solving the associated inverse electromagnetic problem and let t be the considered refractive index threshold. Let additionally consider another pixels whose coordinates are (x', y') and a structuring element Σ^N of rank N whose structuring components are ϵ^N. The global structure of the underlying algorithm, applied on image I whose height is I_h and width is I_w, is given in **Algorithm 2**.

Again, it has to be noted that **Algorithm 2** is not given in a specific computing language. Overall this process aims to minimize the ambiguity that stands on the class overlapping region and therefore on edges between tissue types where refractive index gradients are expected. Obviously, despite the potential increase towards cancer recognition, such a process may concomitantly assimilate benign pixels as malignant pixels. While the increase of correct cancer classification is reported by the sensitivity, the incorrect classification of benign pixels as malignant ones is given by specificity. Sensitivity and specificity are deeper addressed in the following.

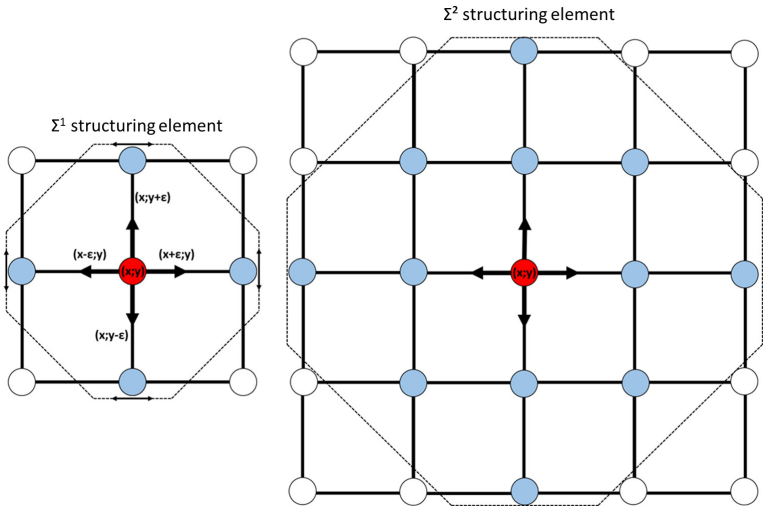


Figure III.18: Schematic representation of Σ^1 and Σ^2 structuring elements. Left: Σ^1 structuring element that may also be named local connectivity with four connectivity. Right: Σ^2 structuring element that is not solely local as it consists of twelve connectivity pixels.

Algorithm 2 Dilation for Connected Thresholding Classification

Require: $t \in \mathbb{R}$, $N \in N$
  for (y = 1; y = I_h; y++) **do**
    for (x = 1; x = I_w; x++) **do**
      if $n_{x,y} \geqslant t$ **then**
        $(x, y) \equiv$ malignant
      else
        $(x, y) \equiv$ benign
      end if
    end for
  end for

  for ($y' = 1$; $y' = I_h$; y'++) **do**
    for ($x' = 1$; $x' = I_w$; x++) **do**
      if $(x', y') \subseteq (x, y) + \epsilon^N$ && $(x, y) \equiv$ malignant **then**
        $(x', y') \equiv$ malignant
      else
        $(x', y') \equiv$ benign
      end if
    end for
  end for

Before going further, it is important to clarify an additional point. What is going on if malignant pixels have been identified at the edges of the tissue section ? Does it induce an increase of the tissue surface ? Well, hopefully no. Let's remind that rather than talking about dilation we talked about constrained dilation. In order to avoid the tissue section to be expanded, the dilation process is limited to an area that is comprised inside the tissue. To do so, from the refractive index map, a logical mask is derived by a contour algorithm that is similar to what was previously reported in *III.4.2 Registration Procedure, Pathology Masking*. Therefore, after the dilation, the newly established diagnosis map is basically multiplied by the mask. Pixels that have increased the tissue section surface during the dilation process are consequently erased and the original shape is preserved giving arise to what we call connected thresholding in the following. Importantly, the term "connected" must not be mistaken with the special programming language in which it solely

refers to a local connectivity, *i.e*, Σ^1. It is worth mentioning that here it refers to the entire set of Σ^1, Σ^2 and Σ^3.

As an image is often more explicit than a lot of words, Figure III.19 proposes to summarize the overall process, from the raw data image to the diagnosis dilation and through the hard thresholding.

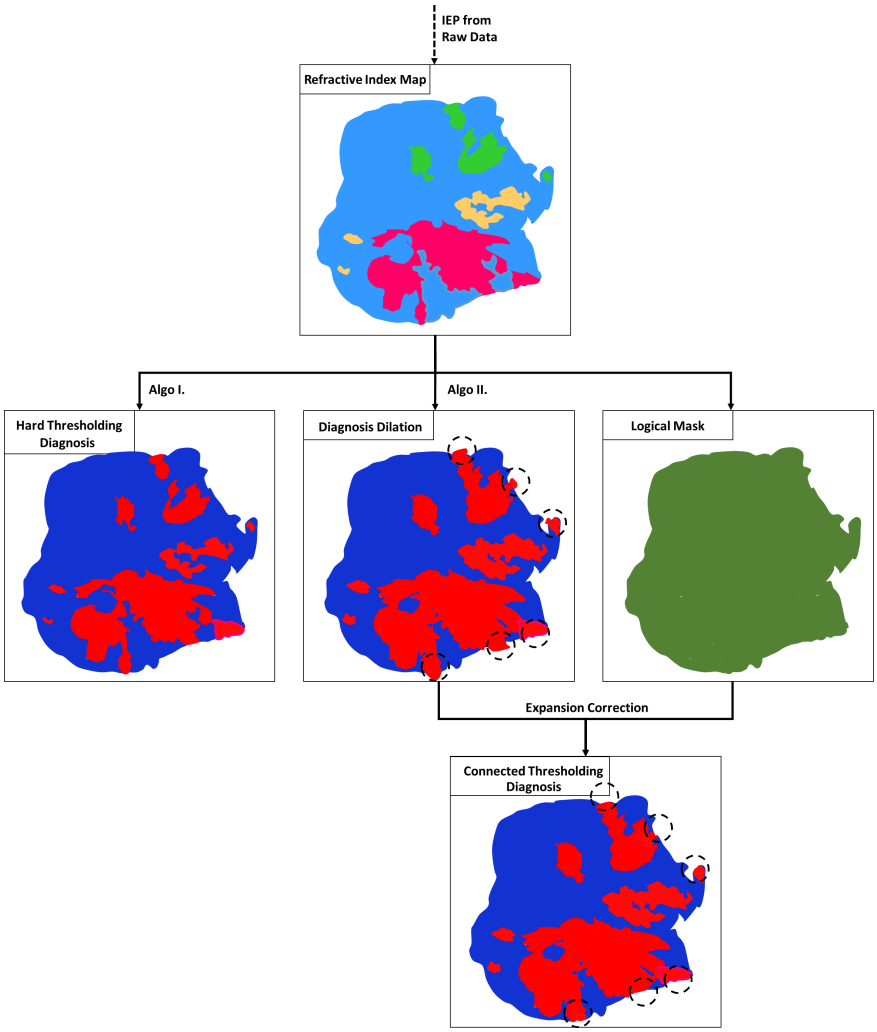


Figure III.19: Diagram of the classification process from the refractive index map to both the hard and connected thresholding diagnosis. The black dashed arrow on the top of the diagram refers to the inverse electromagnetic problem (IEP) and upstream data treatment performed to derive the refractive index map. Note that it is possible to think about connected thresholding as an expansion of the hard thresholding rather than an independent process. For clarity purposes, the two strategies have been organised in two distinct arms. Black dashed circles highlights

III.5.2 Confusion Matrices

Once the two images have the same orientation, the same resolution, and are both exhibiting binary information, one has to determine the degree of compliance between them regarding the classification method that is used. It is worth mentioning that the pathology image prevails in case of discrepancy.

As the classification is binary: benign or malignant, one can distinguish four different cases on comparing the terahertz and pathology diagnosis: (i). both images diagnose a malignant pixel: true positive; (ii). both images diagnose a benign pixel: true negative; (iii). terahertz image diagnoses a malignant pixel while pathology picture states for a benign pixel: false positive; (iv). terahertz image diagnoses a benign pixel while pathology picture states for a malignant pixel: false

negative. The agreement between the two images is similar to a XOR gate and can be summarize as depicted in Table III.2.

Table III.2: Compliance decision rule between the terahertz image and the pathology one.

True Positive: Tumor→ Tumor	False Positive: Tumor→ Benign
False Negative: Benign→ Tumor	True Negative: Benign→ Benign

Instead of being overwhelmed with each confusion matrices that contain the classification of pixels for each threshold and connectivity distance, the effectiveness of a diagnostic technique is often realized on plotting the true positive rate (TPR) as a function of the false positive rate (FPR) for each classification rule. Such a representation is conventionally called receiver operating characteristic (ROC) curve. Alternatively it can be seen as the representation of the sensitivity of the method as a function of unity minus the specificity. In such a representation, the y-axis corresponds to the proportion of malignant pixels that have been correctly classified as abnormal. On the other hand, the x-axis represents the proportion of benign pixels that were incorrectly classified as malignant and are therefore false positives. The true positive rate and the false positive rate are calculated as follows:

$$\text{True Positive Rate} = \text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}, \tag{III.19}$$

$$\text{False Positive Rate} = 1 - \text{Specificity} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}. \tag{III.20}$$

Depending on the number of false positives that one is willing to accept, one can progressively increase the number of malignant pixel that are correctly classified. In fact, a ROC curve is always monotonous and increasing. In other words, a ROC curve tells us the price to pay in terms of false negatives to increase the proportion of correctly classified a larger area of the disease.

Now, if one wants to compare different classifying methods, the area under the curve (AUC) is commonly used. It basically consists in calculating the area that is under the ROC curve. The higher the AUC is, the greater the method is towards classification.

In the following, we compare the two methods, namely hard thresholding and connected thresholding for three freshly excised tissues. The refractive index threshold that are investigated are ranging from 1.5 to 3.0. The selected frequency to perform the investigation is 560 GHz. Overall 560 GHz belongs to the spectral band that was giving arise to the best dielectric demarcation between tissue types. Additionally, the mean discrepancy between the refractive index of healthy fibrous and tumour was shown to be roughly constant from 300 GHz to 1 THz. Therefore, we can state that the overall contrast is not expected to drastically change from one frequency to another within that spectral band. We will see in the next chapter that the choice of this frequency is also partly dictated by another parameter.

Overall, the global procedure to be applied on freshly excised breast tissue samples is schematically represented through the diagram given in Figure III.20. The procedure can be organised in two arms that aim to evaluate the terahertz far-field approach for breast cancer recognition. The first arm consists on extracting the generic knowledge of tissue type responses to terahertz stimuli in reflection configuration. The studies reported throughout *Chapter II* of this work allowed to draw some of the breast tissue characteristics and especially the overall refractive index profiles of tissue types. This knowledge is used in *Chapter III* to define the threshold values that are *a priori* suitable to discriminate breast carcinoma from healthy volume. Thresholding is applied on the section refractive index map that is derived from a self-reference imaging and an inverse electromagnetic problem solved over the entire tissue image. Two main thresholding strategies

are investigated: hard and connected thresholding, giving arise to binarized maps depicting benign and malignant area. On adjusting the pathology image with respect to the binarized terahertz diagnosis, the degree of correlation between the two diagnosis is established. The quantification of the terahertz diagnosis compliance with respect to the pathologist's one (histology annotation) is assessed by the calculation of the receiver operating characteristics (ROC) and the associated area under the curve (AUC). The best thresholds for each classification strategies are subsequently determined as the ones that allow the greater correctly classified malignant pixel rate (sensitivity) while preserving the healthy area from a misdiagnosis (specificity). Finally, results arising from the spectroscopic studies in *Chapter II* and the terahertz diagnosis procedure described heretofore in *Chapter III*, should indicate the potential of far-field terahertz towards breast carcinoma detection.

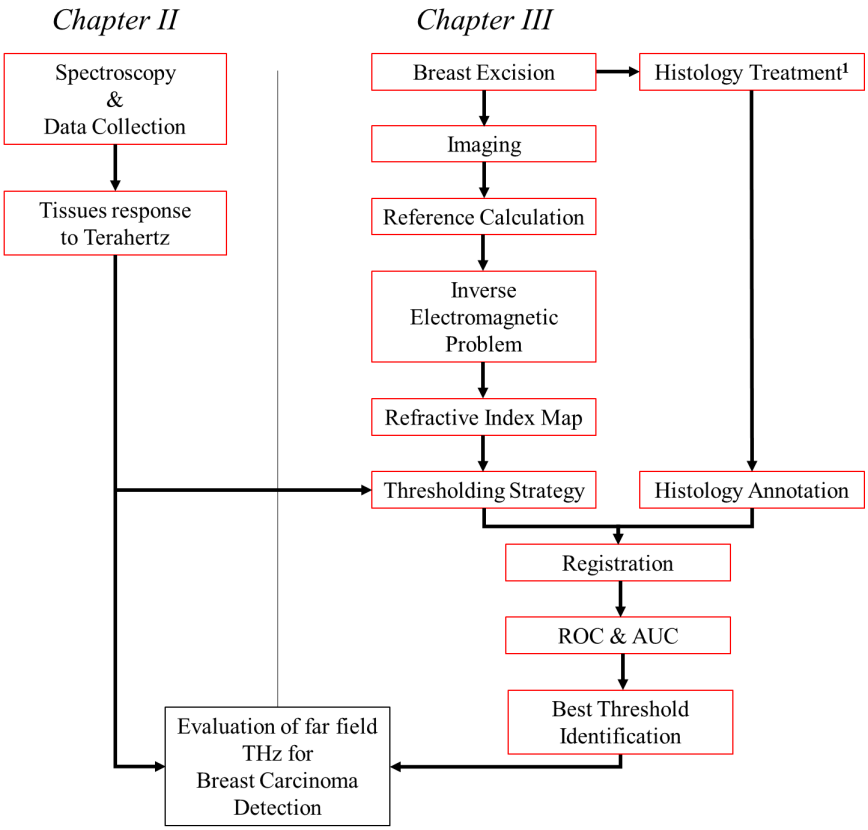


Figure III.20: Flow chart of the method followed to classify tissue section terahertz pixels. [1]: Note that the histology treatment is performed after the terahertz imaging.

III.6 Study Cases

Three different tissue samples illustrate, in the following, the self-reference numerical procedure to ultimately classify pixels belonging to the terahertz image of freshly excised breast carcinoma as malignant or as benign based on the refractive index calculated at 560 GHz. The results are presented as follows: first, a generic view of the sample is given with the photograph, the pathology image, the masked pathology, the raw terahertz image at 560 GHz, the terahertz mask, the refractive index map derived at 560 GHz, the retrieval pixel signal and the reconstructed reference and signal; Second, the ROC curves for the four different classification methods namely hard thresholding (HT), Σ^1 connected thresholding (CT1), Σ^2 connected thresholding (CT2) and Σ^3 connected thresholding (CT3), for each refractive index thresholds are given; Third, the related AUC values are listed; Finally, the pixel classification images for the two threshold values that offer the best trade-off between TPR and FPR for each method are shown. These results are subsequently commented.

III.6.1 NS1 Sample

NS1 sample is an invasive ductal carcinoma from which was suffering a 67 year old woman. The initial tumor site was found to be by about 100 mm^2. On Figure III.21 are reported the pathology image and some enlightened pathology area, the pathology mask, the terahertz image at 560 GHz, the derived terahertz mask, the refractive index map at 560 GHz, the raw and denoised retrieval pixel signal and the denoised reconstructed signal. The pathology image as well as the pathology mask exhibit the presence of a hole where no tissue is found. The lack of tissue in the middle of the section is not natural and enlightens the issues that have been previously reported towards pathology images. Additionally, another hole is exhibited but this time in the terahertz image. This gap is not expected to be due to the tissue response itself but rather due to mechanical constrains within the tissue that tense the tissue and induces a not perfect match between the sapphire window and the sample.

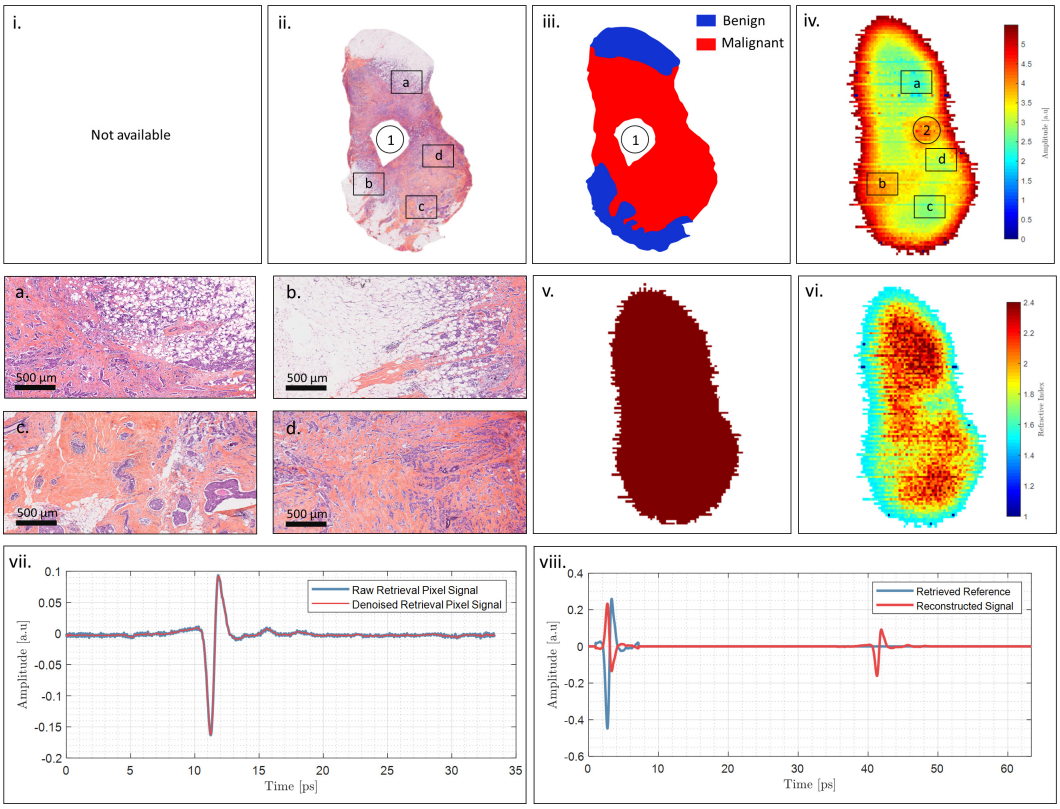


Figure III.21: Overview of the results obtained via the procedure for the NS1 sample: i. not available photograph; ii. pathology image with four enlightened regions. A gap is shown in the middle of the section that is due to the histology chain; iii. Pathology mask derived from the pathologist diagnosis. iv. Terahertz image of the tissue sample at 560 GHz; v. Terahertz mask derived from the 560 GHz raw image; vi. Refractive index map calculated from the self-reference procedure and the inverse electromagnetic problem; vii. Raw and denoised retrieval pixel signals used to retrieve the reference signal - relative time-scale; viii. Reconstructed signal from the retrieval denoised pixel signal - relative time-scale. a, b, c and d are pathological view of the regions related in the pathology image (ii) and the raw terahertz image (iv). (1) highlights a gap in the pathology section and (2) points out a gap in the terahertz image that originates from a gap in the tissue sample due to mechanical strain.

Overall, the refractive index ranges from 1.6 to 2.4 within the tissue. The physical property map exhibits higher refractive index in regions where the density of cell is higher, as denoted for instance by the four rectangular boxes. Clearly, the high density of cancer cells shown in region (a.) is well pointed out in the refractive index map. The same observation can be made for regions (c.) and (d.). On the other hand, region (b.) that mainly consists of adipose tissue, presents a lower refractive index that is roughly between 1.6 and 1.9. Still, in this specific region, some pixels present

higher refractive indices that may be due to the presence of cancerous agglomerates. One can observe that the contour of the tissue sample is not well retrieved and present a low refractive index. It is assumed that the exhibited value does not express the biological structure of the contour but rather the concomitant reflection from air pixels and contour tissue sample ones.

Pixel classification was assessed with the two methods described in the foregoing: the hard thresholding and the connected one. The connectivity ϵ was set to 1, 2 and 3 successively. In Figure III.22 are reported the ROC curve for each classification methods.

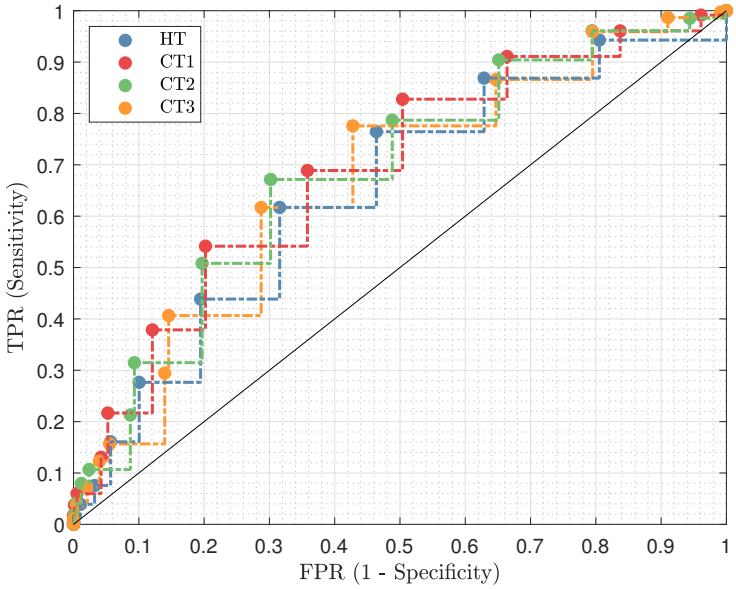


Figure III.22: Receiver operating characteristic for the different classification methods, at 560 GHz. The black line represents TPR = FPR.

Since all ROC curves are to the left of the TPR = FPR line, we can state that the proportion of well classified samples that are malignant (true positives) is greater than the fraction of samples that were incorrectly classified as malignant (false positives) for any decision rules. However, clearly the hard thresholding is the less efficient classification method as the related TPR(FPR) function is lower than for other methods. Such an observation is not that surprising since the hard thresholding method only considers the calculated refractive index value and does not take into account pixel proximity. Therefore, pixels that exhibit a refractive index lower than the threshold although being surrounded by pixels classified as malignant are not considered as sick areas despite their obvious belonging to the tumor site. For this reason, connected thresholding is expected to provide better results. However, on the other hand, connected thresholding may rapidly misdiagnose pixels at the edges between two tissue types and for low refractive index thresholds.

To quantify the effectiveness of each classification methods, the AUC was calculated by trapezoidal integration for each decision rule ROC curve and is reported in Table III. 3.

Table III.3: Area under the curve for each classification method.

	Thresholding Method			
	Hard	Connected		
ϵ	-	1	2	3
AUC	0.6976	0.7307	0.7264	0.7127

Observations towards the classification accuracy of the hard thresholding method from the ROC curve are unsurprisingly in accordance with the AUC value. Overall, the AUC is greater for the connected thresholding CT1 and would therefore indicate that such a classification method

is greater than others. However differences in AUC are relatively small and therefore indicate that CT1 may not consistently be the best classifier. Basically, the choice of the classifier depends on the price one is willing to pay to detect more and more malignant pixels. To select the best refractive index thresholds for each segmentation method, one plot the difference between TPR value and FPR value as a function of the considered thresholds. The best trade-off between TPR and FPR are given by the thresholds providing a TPR – FPR value as close as possible to unity. In Figure III.23 is plotted the TPR – FPR distribution for each method at specific thresholds.

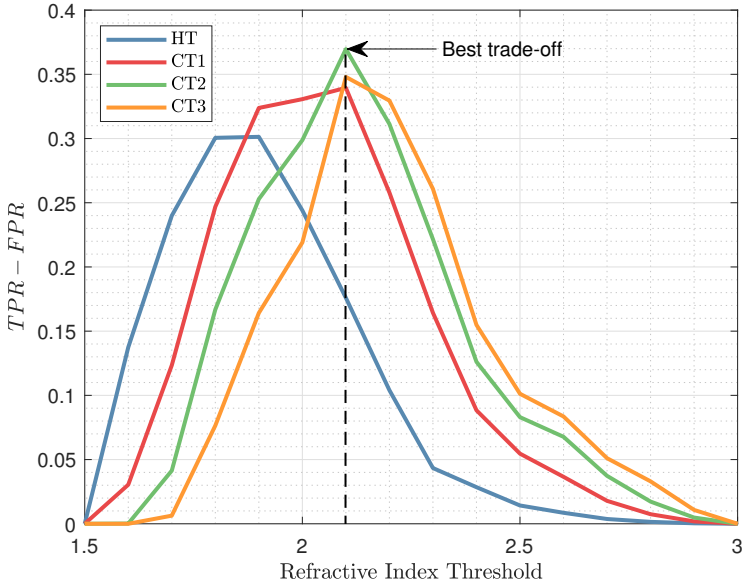


Figure III.23: TPR – FPR as a function of the refractive index threshold for HT, CT1, CT2 and CT3, at 560 GHz.

The first two best thresholds for each methods and their corresponding TPR – FPR value, sensitivity and specificity for the NS1 sample case are listed in Table III.4.

Table III.4: First two best refractive index threshold characteristics considering the trade-off between TPR and FPR.

Method	1^{st} Best Threshold	TPR-FPR	Sensitivity (%)	Specificity (%)
HT	1.9	0.30131	62	68
CT1	2.1	0.33947	54	80
CT2	2.1	0.36965	67	70
CT3	2.1	0.34798	78	57
Method	2^{nd} Best Threshold	TPR-FPR	Sensitivity (%)	Specificity (%)
HT	1.8	0.30055	76	54
CT1	2.0	0.33067	69	64
CT2	2.2	0.31132	51	80
CT3	2.2	0.32953	62	71

From Figure III.23 and Table III.4 it seems clear that the best trade-off between TPR and FPR is provided by the refractive index threshold at 2.1. Such a threshold maximizes both the sensitivity and the specificity of CT1, CT2 and CT3 methods. From the values listed in Table III.4, depending on the parameter that one wants to maximize, namely the sensitivity or the specificity, each CT method offers different capabilities. While CT1 gives at $t = 2.1$ the highest specificity (80%), CT3 offers a better sensitivity by about 78% and CT2 is an interesting trade-off between those two. Interestingly, the second best threshold is lower than the first best one for HT and CT1 classifications while it is higher for CT2 and CT3. Such a behavior is expected to be due to the fact that HT

and CT1 methods slowly segment tissue area on diminishing the threshold value. On the opposite, CT2 and CT3 rapidly spread over the entire image. Therefore, decreasing the refractive index threshold for those methods rapidly lead to a significant decrease towards specificity. Hence, for CT2 and CT3, better secondary trade-off between TPR and FPR are located upstream the first best threshold.

Obviously, it is possible to tune the refractive index threshold to lower value than the one that was previously estimated as the best compromise between TPR and FPR. Such a manoeuvre would drastically increase the sensitivity but concomitantly would affect the specificity. As already pointed out, the choice of the threshold value has to be done depending on the price one is willing to pay to correctly classify all malignant pixels.

Evidently, one could state that the most important would be to correctly identify all malignant pixels. However, it is worth remembering that this work takes place in the context of breast conserving surgery. Therefore, the goal is to minimize the fraction of healthy breast volume to be removed during surgical procedures. Hence, low refractive index thresholds, although increasing the rate of pixels correctly classified as malignant, would concomitantly increase the healthy tissue area that is incorrectly classified and marked to be removed. In a nutshell, the best choice would consist in a threshold value that maximises the true positive rate while minimizing the false positive rate.

For the two best classifying thresholds of each method, the pixel classification images are given in Figure III.24. The images show the superimposition of the resized, translated, rotated and masked pathology image with the refractive index map on which is applied the thresholding. Due to the different treatments that gave arise to these two images, the pathology mask and the terahertz refractive index map at 560 GHz are not expected to match precisely at each location. When the pathology mask is oversized in comparison to the refractive index map, it appears in strong purple color. On the contrary, when the terahertz image is oversized compared to the pathology mask, it appears in deep green color. It is worth to underline once again that regions that do not match are not taken into account in the confusion matrix entries. When the two images match, the four different cases that belong to the confusion matrix, namely true positives (white), false negatives (pink), true negatives (grey) and false positives (green), are depicted in specific faded colors.

Overall, HT method requires low thresholds to increase the number of correctly classified malignant pixels. On comparing the first two best thresholds for this specific classifier, 1.9 and 1.8, with the empirical values that were extracted in the previous chapter, we clearly see that these two thresholds are comprised between the refractive index profiles of healthy fibrous and healthy adipose tissues. Therefore, it seems reasonable to state that the overall density of cancer cells is not high in comparison to the healthy fraction. Additionally, it may also be pointed out that the lesion does not seem to be condensed and limited at a specific region but rather spreads over the entire tissue section. It is worth mentioning that these statements can not be rigorously validated or denied only with this sample, even though the pathology image clearly indicates a weak cell density and a not precisely located lesion.

On their parts, connected methods provide better results for higher refractive index thresholds, that are globally ranging between the previously extracted profiles of malignant tissues and healthy fibrous (between 2.1 and 2.5). Obviously, it was expected that higher connectivity ϵ would offer a best trade-off between TPR and FPR at higher refractive index threshold since these techniques spread more rapidly. However, it is still interesting to note that overall the sensitivity increases while keeping the false positive rate at a reasonably low level.

Moreover, although it seems that connected methods may provide better results towards the correct classification of malignant pixels for the NS1 sample, it has to be underline that further increase of the connectivity ϵ may rapidly lead to inaccurate classification.

Last but not least, it seems that a way to improve the classification methods would be to implement a segmentation that marks as malignant, pixels that are classified as healthy but are completely surrounded by pixels classified as malignant ones. Such an implementation would likely

slightly increase the true positive rate while unchanging the false positive rate.

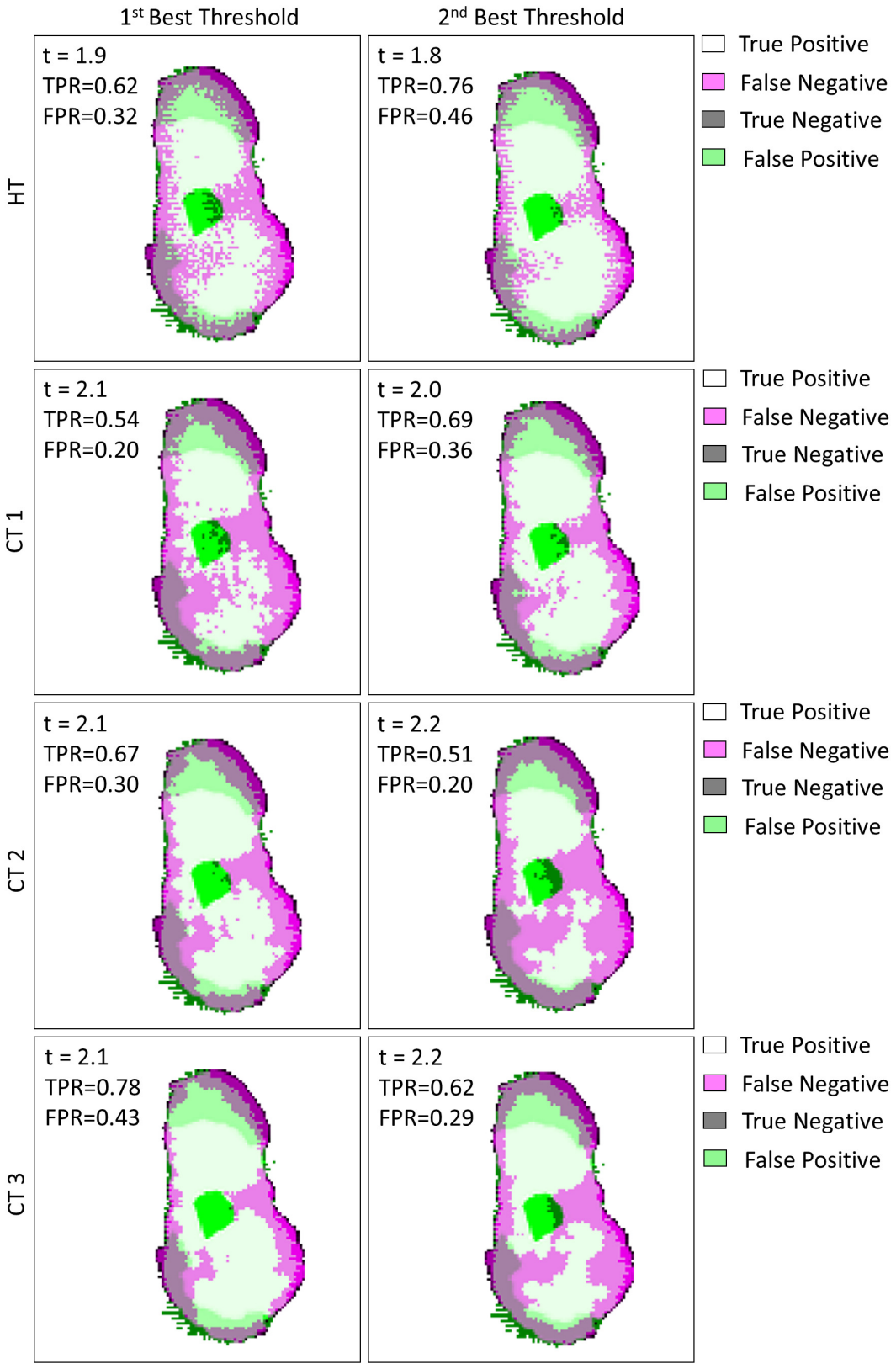


Figure III.24: NS1 tissue sample classification map at 560 GHz for HT, CT1, CT2 and CT3 and their respective first two best thresholds, at 560 GHz.

III.6.2 NS2 Sample

The NS2 sample is an invasive ductal carcinoma. On Figure III.25 are reported the pathology image and some enlightened pathology area, the pathology mask, the terahertz image at 560 GHz, the derived terahertz mask, the refractive index map at 560 GHz, the raw and denoised retrieval pixel signal and the denoised reconstructed signal for the NS2 tissue sample.

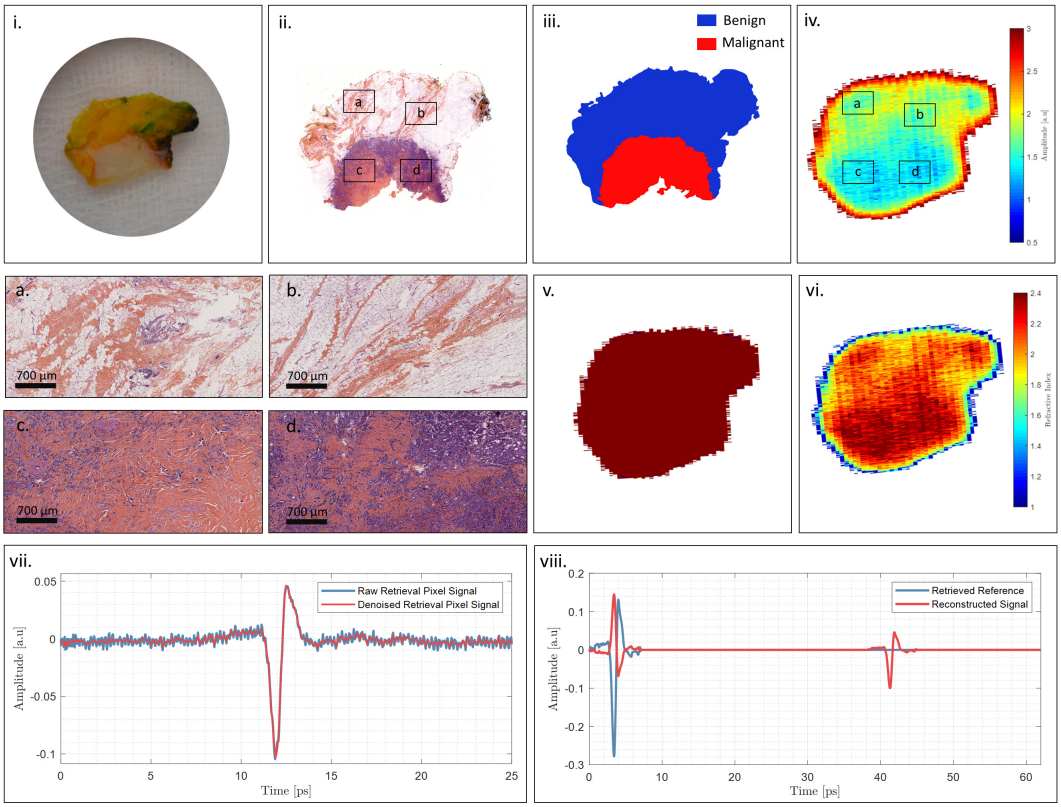


Figure III.25: Overview of the results obtained via the procedure for the NS2 sample: i. photograph; ii. pathology image with four enlightened regions. A gap is shown in the middle of the section that is due to the histology chain; iii. Pathology mask derived from the pathologist diagnosis. iv. Terahertz image of the tissue sample at 560 GHz; v. Terahertz mask derived from the 560 GHz raw image; vi. Refractive index map calculated from the self-reference procedure and the inverse electromagnetic problem; vii. Raw and denoised retrieval pixel signals used to retrieve the reference signal - relative time-scale; viii. Reconstructed signal from the retrieval denoised pixel signal - relative time-scale. a, b, c and d are pathological view of the regions related in the pathology image (ii) and the raw terahertz image (iv).

Clearly, both the raw terahertz image at 560 GHz as well as the refractive index map at the same frequency are exhibiting particular features that correspond to the pathology image. Indeed, region (a.) and (b.) on the pathology image correspond to fibrous that are expected, from the previous spectroscopic investigations, to slow down terahertz light more than adipose tissue but less than tumor ones. These fibrous tissue are roughly localized on the refractive map at 560 GHz while being not observable on the photograph. Concerning the lesion region that is pointed out on histology image region (c.) and (d.), the raw terahertz image depicts lower signal in this specific region that reflects a refractive index close to the one of the sapphire substrate. Concomitantly, the refractive index map denotes the presence of highly refractive tissue.

The terahertz image is more similar to the photograph than the histology image. Overall the malignant lesion region was slightly ripped off in the center due to mechanical strain. Importantly, we distinguish on the right of both the photograph and the pathology image tissues that tend to be darker than others. The hue does not depict specific tissue or pathology. In reality, it is ink that is used to facilitate the slicing of the excised tissue volume.

Similarly to the foregoing, the sensitivity and the specificity of each classification methods was

assessed with ROC curves (Figure III.26).

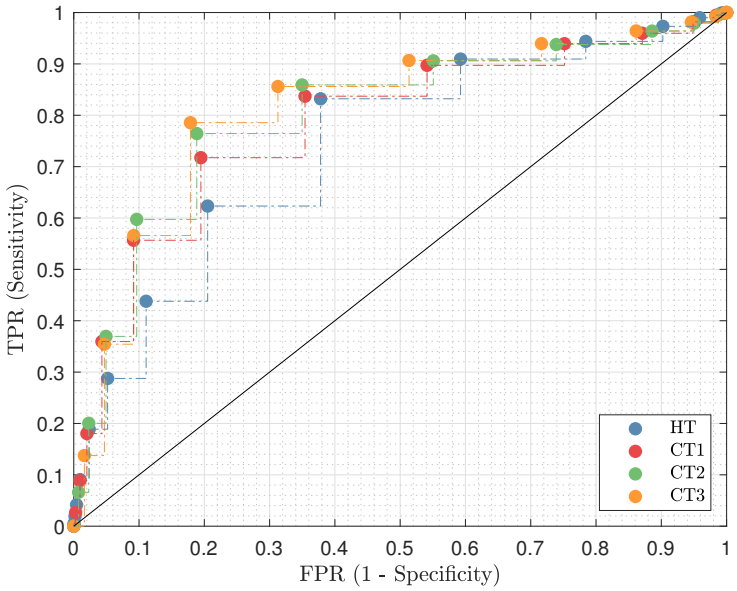


Figure III.26: Receiver operating characteristic for the different classification methods, at 560 GHz. The black line represents TPR = FPR.

Again, all ROC curves are to the left of the TPR = FPR line. Hence, the fraction of true positives is greater than false positives. However, the classification methods are overall more sensitive to abnormal tissue as the ROC curves are less rounded than for the previous study case NS1. Similarly to the previous sample, the hard thresholding classification is the less efficient among the set of tested classifiers. Nevertheless, it is not surprising as the tumor region is well located in the tissue matrix and pathology propagation based classifiers are therefore more powerful. The calculation of the AUC for each method, reported in Table III.5, indicates that the connected method with CT3 is the best classifying method although the values are very close to each other.

Table III.5: Area under the curve for each classification.

	Thresholding Method			
	Hard	Connected		
ϵ	-	1	2	3
AUC	0.7804	0.8149	0.8285	0.8360

The choice of the first two best threshold values was defined in a similar fashion than for the previous sample. The two thresholds that provide the greater TPR − FPR value are selected as they offer the best trade-off between sensitivity and specificity. The TPR−FPR graph for the NS2 sample is given in Figure III.27. The first two best thresholds for each methods and their corresponding TPR − FPR value, sensitivity and specificity for the NS2 sample case are listed in Table III.6.

In comparison to the previous sample, the best trade-offs between true positive rate and false positive rate are provided by higher refractive indices. The shift towards the most efficient thresholds can be explained by the high density of cancer cells in the cancerous lesion that is concomitantly well delimited at the bottom part of the sample section. Therefore, it is expected that on probing the tissue section surface with terahertz pulses, the contribution from the stroma is overall weaker than cancer cells. Consequently, a higher refractive index is expected to be exhibited in such regions.

Remarkably, both the sensitivity and the specificity are hugely greater for the NS2 case than for the NS1 one. The first best threshold among all is provided by the CT3 method for $t = 2.6$. For

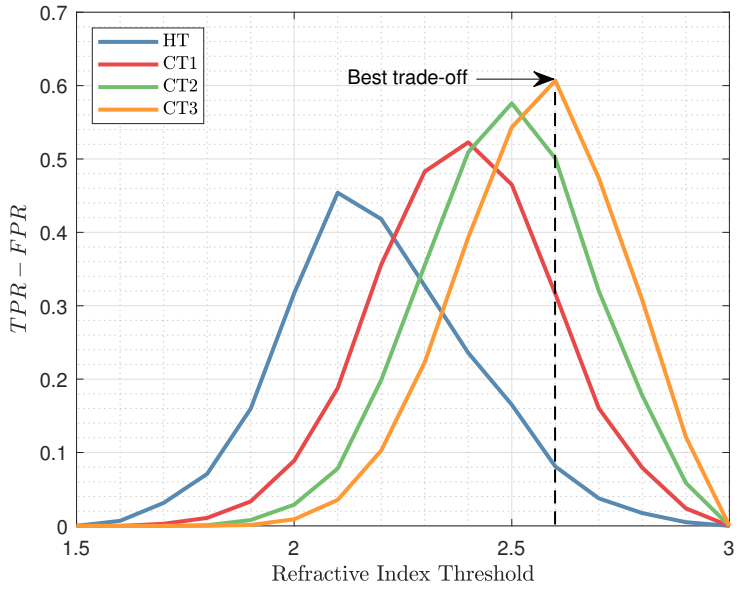


Figure III.27: TPR – FPR as a function of the refractive index threshold for HT, CT1, CT2 and CT3, at 560 GHz.

Table III.6: First two best refractive index threshold characteristics considering the trade-off between TPR and FPR.

Method	1^{st} Best Threshold	TPR-FPR	Sensitivity (%)	Specificity (%)
HT	2.1	0.45402	83	62
CT1	2.4	0.52270	72	81
CT2	2.5	0.57593	76	81
CT3	2.6	0.60683	79	82
Method	2^{nd} Best Threshold	TPR-FPR	Sensitivity (%)	Specificity (%)
HT	2.2	0.41813	62	79
CT1	2.3	0.48288	84	65
CT2	2.4	0.50928	86	65
CT3	2.5	0.54328	86	69

such a thresholding, the classification reaches a sensitivity of 79% and a specificity of 82%. In other words for CT3 at $t = 2.6$, eight malignant pixels over ten are correctly classified as malignant while preventing more than eight healthy pixels over ten to be misdiagnosed as malignant. While the two others connected thresholding method CT1 and CT2 are less efficient than CT3 towards both sensitivity and specificity, the HT classification provides a slightly greater sensitivity for $t = 2.1$ by about 83%. However, the gain of 4% in sensitivity from CT3 to HT costs at the same time 20% in specificity. Reasonably, such a gain is not worth considering the drastic decrease in specificity. Alternatively, if ones wants to increase the sensitivity while maintaining specificity at a reasonable level, second best thresholds may offer a promising substitute. On using the second best threshold provided by CT3 classification, an increase of 7% towards sensitivity conjointly reduces specificity by about 12 %. By doing so, one reaches a sensitivity of 86%. Once again, it would be possible to increase even more the sensitivity by lowering the threshold value. However it will, in fact, reduce the specificity as already pointed out. Another way would probably stand in classifying automatically as malignant pixels that are surrounded by malignant zones. Overall, it is still interesting to note that CT1 and CT2 do not offer better intermediary alternatives.

Furthermore, it is also important to notice that the accuracy of the method is limited by the pathology image deformations due to the histology treatment. Therefore, both the sensitivity and

the specificity reported for the NS2 sample, although being promising towards breast cancer detection, are relative to the best registration achieved between the terahertz image at 560 GHz and the pathology image and not an absolute registration.

The classification maps for each classifier with their related first two best thresholds are given in Figure III.28.

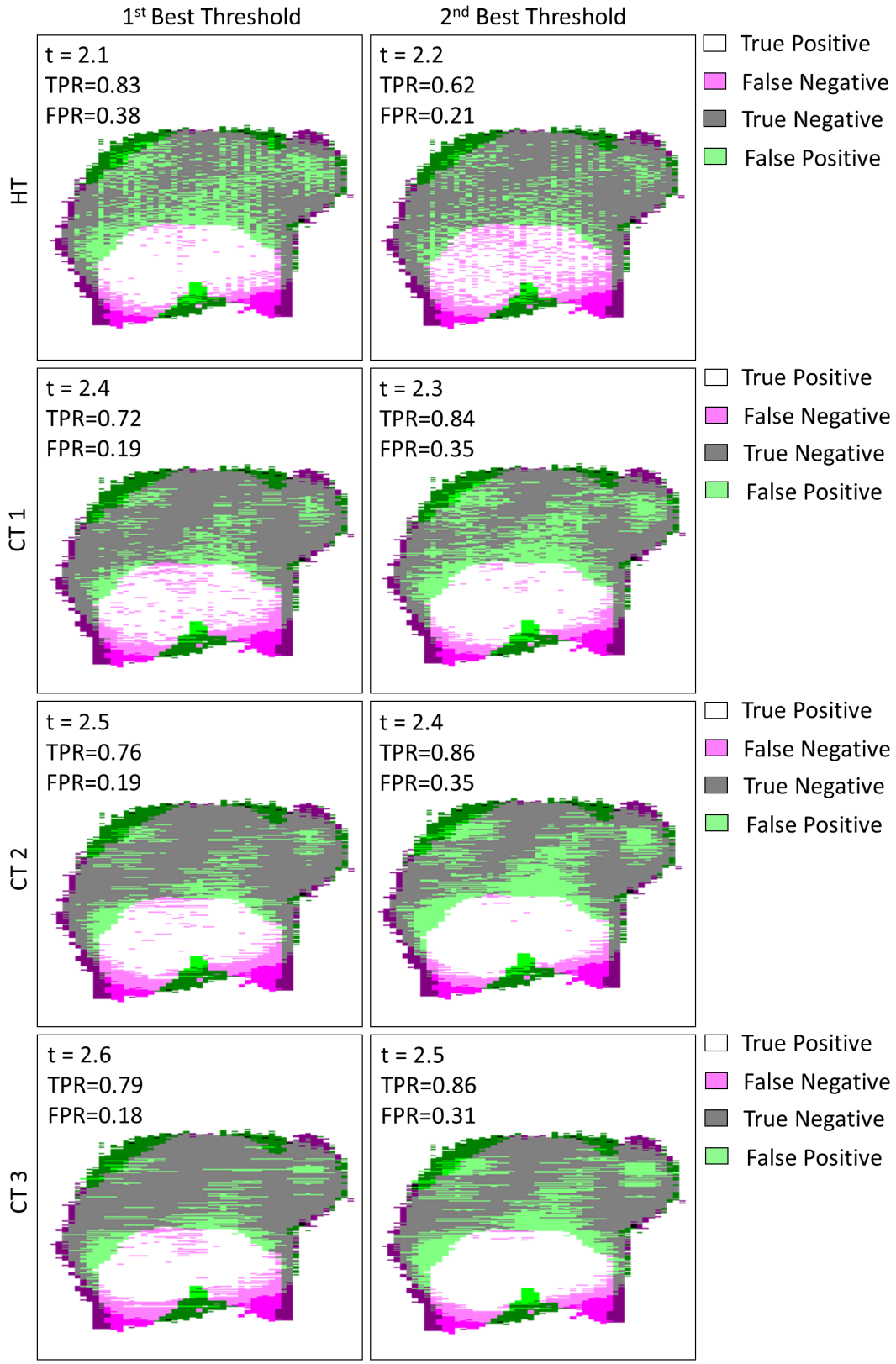


Figure III.28: NS2 tissue sample classification map at 560 GHz for HT, CT1, CT2 and CT3 and their respective first two best thresholds, at 560 GHz.

III.6.3 NS3 Sample

The NS3 sample is a section of an invasive lobular carcinoma from which was suffering a 86 year old woman. According to the physician diagnosis, none lymph nodes were affected by the disease. On Figure III.29 are reported the pathology image and some enlightened pathology area, the pathology mask, the terahertz image at 560 GHz, the derived terahertz mask, the refractive index map at 560 GHz, the raw and denoised retrieval pixel signal and the denoised reconstructed signal for the NS3 tissue sample.

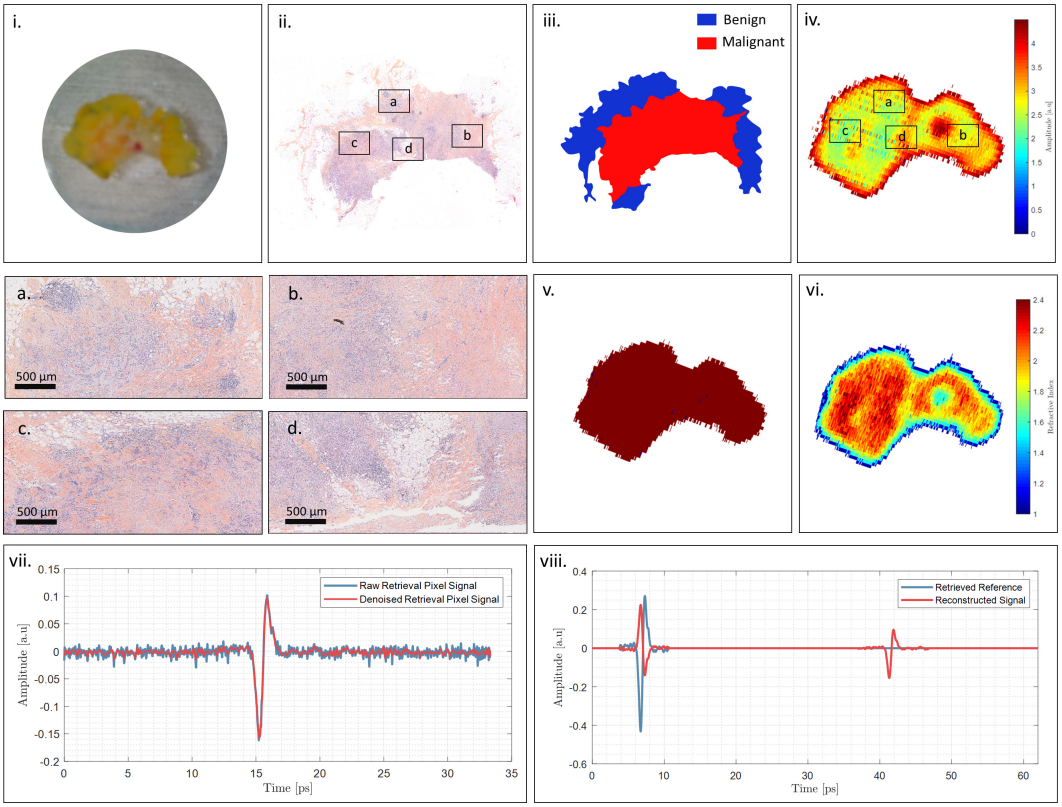


Figure III.29: Overview of the results obtained via the procedure for the NS3 sample: i. photograph; ii. pathology image with four enlightened regions. A gap is shown in the middle of the section that is due to the histology chain; iii. Pathology mask derived from the pathologist diagnosis. iv. Terahertz image of the tissue sample at 560 GHz; v. Terahertz mask derived from the 560 GHz raw image; vi. Refractive index map calculated from the self-reference procedure and the inverse electromagnetic problem; vii. Raw and denoised retrieval pixel signals used to retrieve the reference signal - relative time-scale; viii. Reconstructed signal from the retrieval denoised pixel signal - relative time-scale. a, b, c and d are pathological view of the regions related in the pathology image (ii) and the raw terahertz image (iv).

The shape of the terahertz image is compliant with the photograph. A gap is observed in the terahertz image that corresponds to a local contraction of the tissue within the malignant lesion. Therefore, it is assumed that such a feature does not indicate either the presence or the absence of malignancy. In addition, it seems clear that the pathology image and therefore its mask have been affected by important deformation during the histology treatment. Consequently, it is expected that the correlation and the accuracy estimation of the classification will be impacted.

The ROC curves for HT, CT1, CT2 and CT3 at each refractive index thresholds for the NS3 sample are given in Figure III.30.

Although the correlation between the pathology mask and the refractive index map at 560 GHz is expected to be tricky and weakly accurate, the ROC curves are still all to the left of the TPR = FPR line therefore indicating that the number of correctly classified malignant pixels is higher than the number of healthy pixels classified as malignant. However, it must be underline that among the three tested samples, the NS3 shows ROC curves that are much closer to the TPR = FPR line than

other cases. Such an observation may be attributed to two concomitant facts that are the inaccurate correlation and the important gap seen on the sample section. It is worth to underline that these issues do not arise from the procedure described heretofore but rather point out the difficulties encountered when experimenting on freshly excised tissues. Importantly, it does not even underline the inefficiency of terahertz technology towards breast detection. In fact, the destructive histology treatment as well as the fresh tissue intrinsic mechanical constrains are assumed to be the principal causes of the reported poor performances of the classification. Nevertheless, the mechanical constrains that are assumed to cause the gap in the terahertz image are issues that can not be omitted on probing fresh tissues as they may often be encountered. Therefore it is compulsory to take into account such obstacles rather than omitting them when evaluating terahertz technology and the suggested numerical procedure. On its part, histology treatment remains the image that rules the medical truth. Hence, we have deliberately chosen to present the results arising from the NS3 case that reflect some limitations.

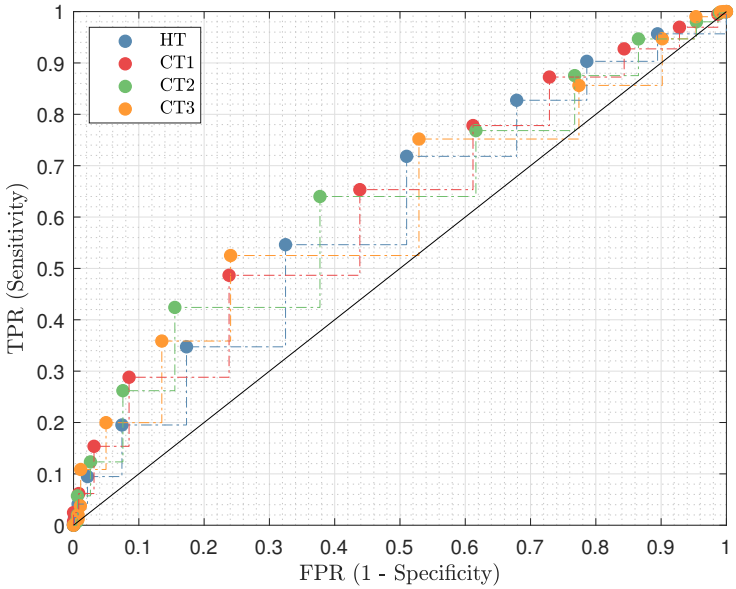


Figure III.30: Receiver operating characteristic for the different classification methods. The black line represents TPR = FPR, at 560 GHz.

The calculation of the AUC for each method, are reported in Table III.7. As observed on the ROC curves, none of the classification methods set itself apart from the others. Overall we can state that both the CT2 and CT3 are equivalent and may give arise to the most accurate classification.

Table III.7: Area under the curve for each classification.

	Thresholding Method			
	Hard	Connected		
ϵ	-	1	2	3
AUC	0.6478	0.6631	0.6695	0.6693

The TPR – FPR distributions for sample NS3 are given in Figure III.31.

The most efficient classifying strategy is the CT3 with a threshold of $t = 2.3$ on the refractive index. The TPR – FPR function s by about 0.28 with a sensitivity of 53% and a specificity of 76%. As already indicated, these characteristics are far below the efficiency reached for other cases and origin for such low performances were already pointed out. CT2 strategy offers a greater specificity of 84% but simultaneously concedes 11% upon sensitivity, that drops below the critical threshold of half the number of malignant pixel correctly classified. Overall, it seems that in such a config-

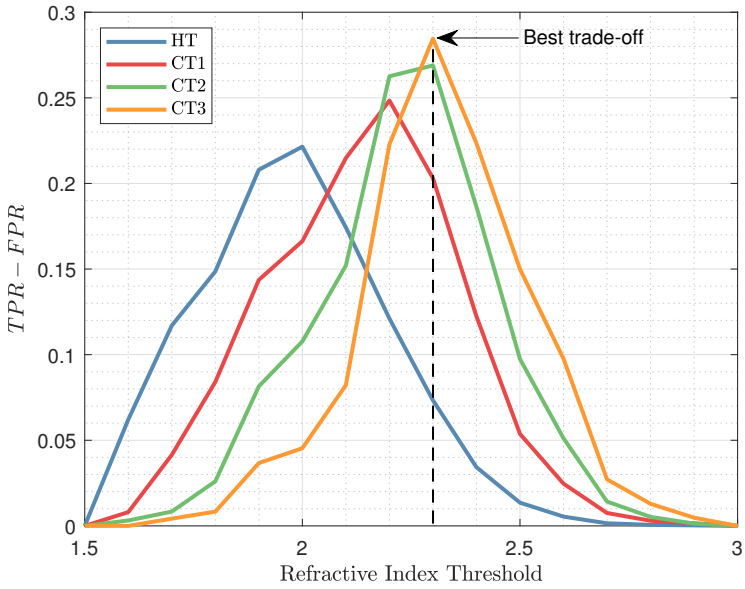


Figure III.31: TPR – FPR as a function of the refractive index threshold for HT, CT1, CT2 and CT3, at 560 GHz.

Table III.8: First two best refractive index threshold characteristics considering the trade-off between TPR and FPR.

Method	1^{st} Best Threshold	TPR-FPR	Sensitivity (%)	Specificity (%)
HT	2.0	0.22146	55	68
CT1	2.2	0.24843	49	76
CT2	2.3	0.26891	42	84
CT3	2.3	0.28454	53	76
Method	2^{nd} Best Threshold	TPR-FPR	Sensitivity (%)	Specificity (%)
HT	1.9	0.20801	72	49
CT1	2.1	0.21481	65	56
CT2	2.2	0.26258	64	62
CT3	2.4	0.2234	36	86

uration, the second best threshold in terms of trade-off between TPR and FPR, provided by HT classification is a better choice. In fact, it provides a sensitivity of 72% and a specificity of 49%. Nevertheless, due to the weak accuracy between the pathology image and the terahertz one, these considerations are, overall, questionable. Even though the efficiency of the classification are debatable for the NS3 sample, it is still worth to underline that for the three connected classifying strategies, the best thresholds are located between the refractive index profiles of fibrous and tumor that were extracted in the previous chapter. The classification images for the first two best thresholds are given in Figure III.32.

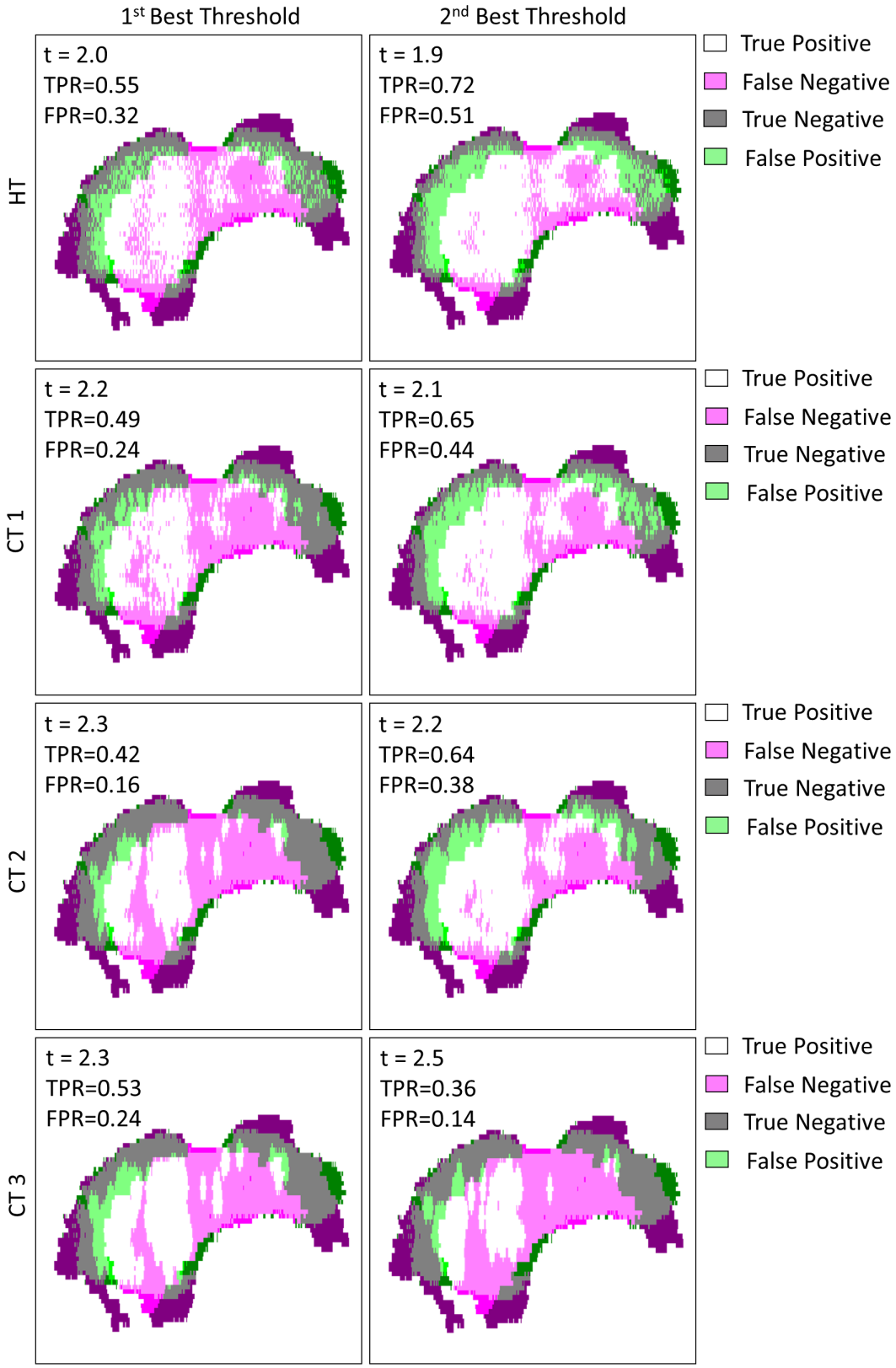


Figure III.32: NS3 tissue sample classification map at 560 GHz for HT, CT1, CT2 and CT3 and their respective first two best thresholds, at 560 GHz.

III.6.4 Discussion on Inversion & Segmentation

Heretofore, a numerical approach was described to derive a map of the refractive index of freshly excised breast tissues. The procedure is performed on a reduced set of information as the reference signal and the first dielectric interface signal between the air and the sapphire substrate are *a priori* unknown. Both these signals are retrieved and reconstructed from the imaging of the second dielectric interface, on which stands the tissue sample to be inspected. From a pixel that reflects terahertz pulse from sapphire to air (second dielectric interface), an inverse propagation is realized, knowing the properties of the sapphire substrate and the incidence angle of the probing terahertz pulses.

Based on the calculated reference, an inverse electromagnetic problem is solved at each location of the tissue sample to extract the refractive index within these specific regions. One way to enhance the self-reference procedure would likely lie in calculating the reference signal for each pixel that belongs to the second dielectric interface, from the first dielectric interface between the air and the sapphire substrate. By doing so, one may totally erase issues arising from a possible slight tilt of the air-sapphire substrate interface, and therefore one will improve the accuracy of the refractive index extraction.

The 560 GHz frequency was chosen to purse further investigations as it lies in the spectral band that offered the most promising discrimination between tissue types in terms of refractive index. Obviously, it would be interesting to investigate additional frequencies while keeping in mind that one would therefore face a tremendous amount of data. On facing big data, one would consider specific numerical implementations, that were not treated in this work, to proceed calculations in a reasonable time. As pixels are *a priori* independent from the imaging point of view, it seems quite indicated to parallelize the inversion task on central processing units or -better- on graphical processing units.

From the refractive index map of each sample, pixels were classified as benign or as malignant using different refractive index thresholds. A wide range of refractive index threshold was investigated based on the refractive index profiles that have been extracted in *Chapter II*, section *II.6.2 Extracted Intrinsic Properties* for each tissue types. Concomitantly to these different thresholds that were considered, two segmentation strategies were tested. While the first classified pixels depending only on their exhibited refractive index, the second method additionally takes into account pixel proximity. The proximity between pixels was introduced with a parameter ϵ that defines the space of cross dependence between neighbor pixels. Hence, pixels that were belonging to the space of length-ϵ of a pixel classified as malignant were therefore automatically marked as belonging to the lesion. The second strategy that was named connected thresholding in opposition to the first method, the hard thresholding, was investigated for three different structuring element Σ^1, Σ^2 and Σ^3. It would be of great interest to study the efficiency of additional segmentation techniques or to improve the previous ones. As a suggestion, one may improve the thresholding segmentation techniques that were in the foregoing described by classifying pixels marked as healthy but surrounded by malignant ones to as belonging to the lesion. Furthermore, watershed method, based on the calculation of the gradients within the image or clustering approach that aims to divide the image into homogeneous clusters could be investigated [**27**]. Additionally, at the time this manuscript is written, the field of artificial intelligence and machine deep learning is burgeoning, and therefore indicating future promising capabilities for medical applications such as tissue segmentation. However, a huge quantity of data is necessary to train efficiently neuronal network. Another way would to study principal component analysis performances. This will be addressed in the following section. Last but not least, alternative characteristics or even a set of multiple parameters could be used for segmentation. In fact, this work mainly focused on the refractive index that is obtain with relative computational complexity, but, more simpler features could be defined such as the time-domain peak-to-peak amplitude, the peak full-width at mid-half, the maximum amplitude of the frequency-domain modulus or the phase of the signal at a given frequency. All of these at different times and frequencies. In this sense, Fitzgerald *et al* conducted in 2012 an interesting work on ten different parameters tested on 31,236 pulses from

normal tissues and 30,405 pulses from tissues with tumor . They reported that using the support vector machine classification algorithm, the parameter that shown the best performances towards correct classification was the time width of the pulse from the time at which the pulse amplitude is initially at the fraction of 0.3 of the minimum amplitude E_{min} to the time T_{Emin}. Such a parameter provided a correct classification by about 70%. In addition, different combinations of parameters were investigated. The overall parameter performances can be found in **[28]**.

Each segmentation strategy was subsequently evaluated by comparing the pathologist's diagnosis provided by a hematoxylin and eosin stain pathology image and the refractive index map. The comparison between the two diagnosis was performed by first assessing the important differences between the two images. Indeed, it was reported that the two images were presenting different resolution and orientation. The pathology image resolution was reduced to the resolution of the terahertz image by bicubic interpolation and the orientation registration was achieved by maximizing the Pearson correlation coefficients between the edges of the two clichés. From the adjusted pathology image, a binary pathology mask was derived that consists into regions suffering from the cancer and the healthy ones. Despite all of these special cares, correlation was not expected to be perfect since the histology treatment that allows the pathologist to deliver his diagnosis may drastically deform and affect the global structure of the tissue sample section. Hence, some concerns were reported on the accuracy of the correlations between the pathology image and the terahertz image and subsequently on the evaluation of the segmentation techniques that were employed to delineate between sick and healthy regions within tissue samples. Some specific solutions were raised by Chavez *et al* in **[29]** that employed Delaunay triangulation to correct pathology images.

Segmentation strategy evaluation was assessed by filling confusion matrices denoting the compliance of the terahertz refractive index segmentation to that of the pathology diagnosis. Receiver operating characteristic curves were hence obtained by plotting the true positive rate as a function of the false positive rate. The best thresholds for each method were defined as the ones providing the best trade-off between true positive rate and false positive rate. Globally, the connected thresholding strategies offered quite better sensitivity and specificity in comparison to the hard thresholding performances. Even though the number of samples reported in this work does not allow to state on a generic operating factor to classify consistently in a correct manner pixels for all carcinoma cases, it was observed that low thresholds were suitable for lesions that tend to spread over the entire section due to the important contribution of the stroma to the overall terahertz signal. On the contrary, high thresholds were offering promising performances for well delimited lesions were the stroma contributes less to the recorded signal. For instance, CT3 segmentation offered a sensitivity of 79% and a specificity of 82% for the NS2 sample with the threshold providing the best trade-off between true positive rate and false positive rate. The sensitivity was increased to 86% with the second best trade-off threshold by conceding a specificity of 69%.

These results have to be compared with other existing techniques. For instance one can report the automated segmentation of masses within mammograms with Fisher's linear discriminant that reaches a maximum sensitivity of 65% and a specificity of 80% with constrained region growing **[30]**. The study reported above was conducted on a much larger set of samples. Moftah *et al* achieved an accuracy of more than 85% towards breast cancer recognition on magnetic resonance clichs by developping an adaptative k-means clustering algorithm **[31]**. Accuracy is slightly different from the operating receiver used in this work as it depicts the number of correct assessments by the total number of assessments. Interesting results were also reported from the domain of ultra-sound imaging where solid breast tumors were correctly identified with a sensitivity of more than 88% and a specificity of 82.5% **[32]**. Finally, Rouhi and colleagues obtained both sensitivity and specificity higher than 90% by training a neuronal network with mammograms from international databases **[33]**. Various other techniques have to be acknowledged but would require an entire additional manuscript which is not the goal of the manuscript. Overall, it is important to mention that all these techniques are not aimed to be employed under the same circumstances. While some are expected to detect cancer on its early stages like mammography, other techniques

and especially terahertz are anticipated to be used within the operating room to evaluated the excision margins. It is often stated that these techniques are in competition between each others. However, it sounds -on a personal note- not to be the case. Rather than competing with each others, these various methods are ultimately aimed to work together to lower as far as possible the rate of breast cancer death among women, worldwide.

III.7 Principal Component Analysis

To enlarge the field of inquiry, principal component analysis (PCA) was investigated. PCA also referred to as Karhunen-Loève expansion, was first formulated in statistics by Karl Pearson **[34; 35]**. PCA is often employed for different goals like data reduction **[36]**, outlier detection **[37]**, classification **[38]** and prediction **[39]** as it extracts the dominant patterns in a collected data set.

III.7.1 Mathematical Definition

PCA consists of an orthogonal transformation that outlines a set of uncorrelated values with each successive principal component having the highest possible variance under orthogonality constraint **[40]**. In a nutshell, PCA tends to determine the subspace in which, the orthogonal axis best describe the discrepancy within the data set **[41; 42]**. In Figure III.33 is schematically represented an example of data reduction on a simple 2D data set, using principal component analysis.

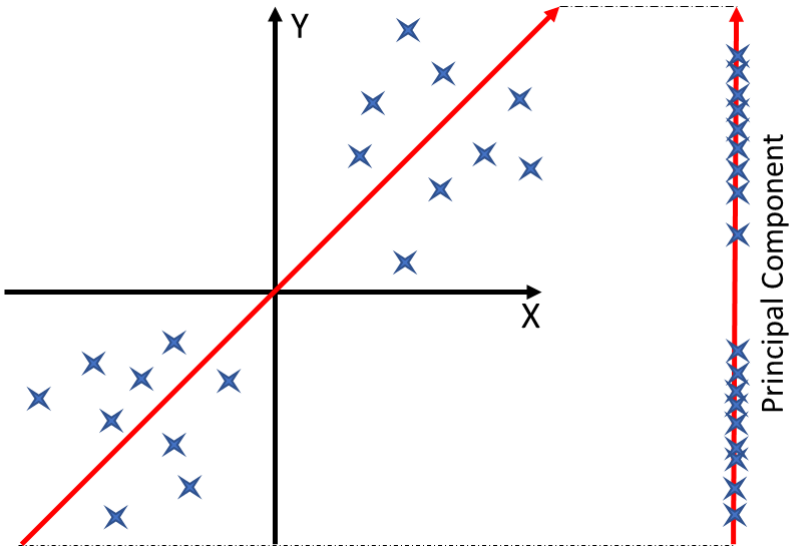


Figure III.33: Data compression: from 2D to 1D data set.

Mathematically, the projection of a data set X such as $X \in \mathbb{C}^N$, down on a subspace by means of a matrix of projection P' provides the data coordinates in the plane. Let this plane termed T. The columns in T are the score vectors s and the rows in the projection matrix P' are called loading vectors l. Score and loading vectors are strictly orthogonal that means $s'_i s_j = 0$ and $l'_i l_j = 0$ for $i \neq j$. The projected coordinates in the newly defined plane may deviate from the original coordinates and are termed residuals. Residuals are collected in a dedicated matrix R. It is therefore possible to express the PCA in a matrix form:

$$X = TP' + R. \tag{III.21}$$

Another way to think about score and loading vectors is to regard matrix P' as a measure of the influence of variables found in the original data set X on T that in turn describes how the different observations are related to each other. Furthermore, it can be seen as a mathematical transformation whose projection core depends on the input data. There is multiple algorithms that

can be employed to extract the principal components and their related score and loading vectors. Among them the most widely used are the nonlinear iterative partial least squares (NIPALS) algorithm **[43; 44]** and the singular value decomposition (SVD) one **[45]**. Both these algorithm sort the principal components. However NIPALS algorithm processes the principal components sequentially while SVD sort them out simultaneously. In this work, the $pca(\cdot)$ function from MATLAB® that uses SVD was employed.

First and foremost, one has to clearly identify the variables and the observations on which the PCA has to be performed. In this specific work, variables are the frequencies which are comprised in the predefined confidence interval ranging from 200 GHz to 1.0 THz. Observations are either the modulus or the phase of each pixel belonging to the tissue sample at these given frequencies. Hence, rows of the data set to be projected on a subspace from principal component analysis, are the spectra or the phase that were recorded for each location on the sample. On the other hand, columns are the specific values of either the modulus or the phase at a given frequency.

Additional proper precautions should be taken considering pixels that have to be taken into account during the PCA. Indeed, the shape of a matrix is necessarily rectangular. However, it would be remarkable if a tissue sample would also be rectangular. Globally tissues are of random shape. Therefore, pixels that belong to the background must be ignored in order to not affect the score and loading vectors arising from the analysis. To do so, the tissue sample area is determined using a similar masking procedure that the one described in the previous section. As the tissue mask is nothing else than a logical entity that either take 0 or 1 as value at each pixel, it is therefore easy to delimit the region of interest onto which has to be performed the PCA routine. Moreover, despite the fact that our data set are 3D images, PCA usually takes 2D data set as entries. Therefore, 3D images have to be reshaped in 2D matrices. Pixels submitted to PCA have therefore to be arranged in a vector. Basically, $reshape(\cdot)$ function from MATLAB® can be used to do so. Figure III.34 shows an example of a final matrix to be inserted into the PCA routine.

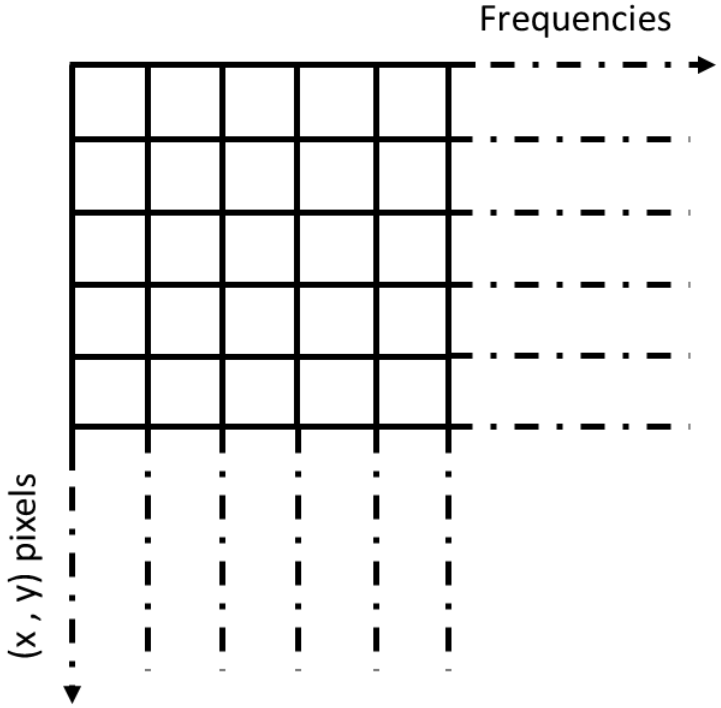


Figure III.34: Schematic representation of a data matrix to be inserted into the PCA routine.

Furthermore, one has to clearly understand what PCA returns after processing. First, it returns a matrix of coefficients, the loadings, for the $n-$by$-p$ data matrix X. The returned loading matrix is a $p-$by$-p$ matrix. Each value of a given column corresponds to the loading of the related variable (frequency). Therefore, if we consider the loadings of a given principal component to be l_i, one can obtain the projection from this principal component by:

$$\sum_{i=1}^{i=\mathrm{N}} l_i \times \mathrm{O}_i, \tag{III.22}$$

where O_i is the value of the observable (here modulus or phase) of a specific pixel at variable (frequency) i.

Evidently, one therefore solely obtain weighted observable. To obtain the final image arising from the principal component one has to integrate the observable along the variable (frequency) vector length.

Each of the principal component explain by itself a certain percentage of the total variance within the original image. Overall, the very first principal components are expected to provide by themselves the overall variance explanation. Let's however underline that as far as the process is unsupervised, it is almost impossible to identify the physical meaning of the projection. Moreover, in many cases, the projection on a subspace corresponds to a weighted combination of various parameters. Therefore, PCA is investigated and performed to exhibit possible interesting features rather than providing physical operating factors. In that respect, correlation with pathology images would make no sense.

In the following, results obtained from PCA on the study cases NS1, NS2 and NS3 are presented and commented.

III.7.2 NS1 Sample

Principal component analysis was performed on both the modulus and the phase exhibited from each pixels within the frequency band lying from 200 GHz to 1 THz. In Table III.9 are reported the contribution to variance of each principal component for both the modulus and the phase.

Table III.9: Contribution to variance (%) for the first six principal components based on the modulus and the phase of each pixel between 200 GHz and 1 THz.

	Observable: Modulus	Observable: Phase
Principal Component	Contribution to Variance (%)	Contribution to Variance (%)
PC 1	98.88	97.37
PC 2	0.66	2.40
PC 3	0.12	0.11
PC 4	0.07	0.05
PC 5	0.05	0.03
PC 6	0.02	0.02

In view of the contribution to variance of the first principal components that are shown in Table III.9, only images that corresponds to at least the first three principal components will be shown. Indeed, it is expected that higher principal component may partly described the variance found in the experimental noise rather than highlighting tissue discrepancies. In Figure III.35 are sorted the corresponding images for the NS1 sample.

Clearly, the first principal component PC1 almost entirely explains the variance within spectra recorded in the image over the investigated frequency range. The three frequencies that influence the most the first principal component from the modulus data set are, in descending order, 360 GHz, 330 GHz and 480 GHz. Overall, the first principal component is relatively similar to the raw terahertz image at 560 GHz that was shown in the previous section. Therefore, it is reasonable to suspect that the first principal component mainly refers to the amplitude of the modulus over the frequency range. The second principal component seems to give information that are kinda similar to the first principal component. However, it is interesting to note that demarcation within the sample are overall better delineated. Obviously such a statement should take into account the small contribution to variance of the second principal component that is by about 0.7%. The last principal component to be shown does not reflect similar features but rather possible new

information. However, it is difficult to identify the origin of the contrast. Moreover, the particularly low contribution to variance of this principal component it is reasonable to not drag it out.

Concerning the analysis on the phase, the first principal component also almost entirely explain the variance. However it is interesting to point out that the contribution from the second principal component is not that low as it reflects by about 2.5% of the total variance. The first principal component from the phase data set is mainly influenced by 540 GHz, 630 GHz and 720 GHz frequencies while the second principal component is mainly influenced by 540 GHz, 810 GHz, and 900 GHz. It is interesting to underline that the second principal component of the phase data set seems -subjectively- to be the inverse of the third principal component of the modulus data set. It is still, however, not possible to relate the contrast to specific features. As already stated, it is possible that these features arise from noise fluctuations rather than biological differences within the tissue section. Similarly, features observed from the third principal component calculated for the phase data set do not offer additional information.

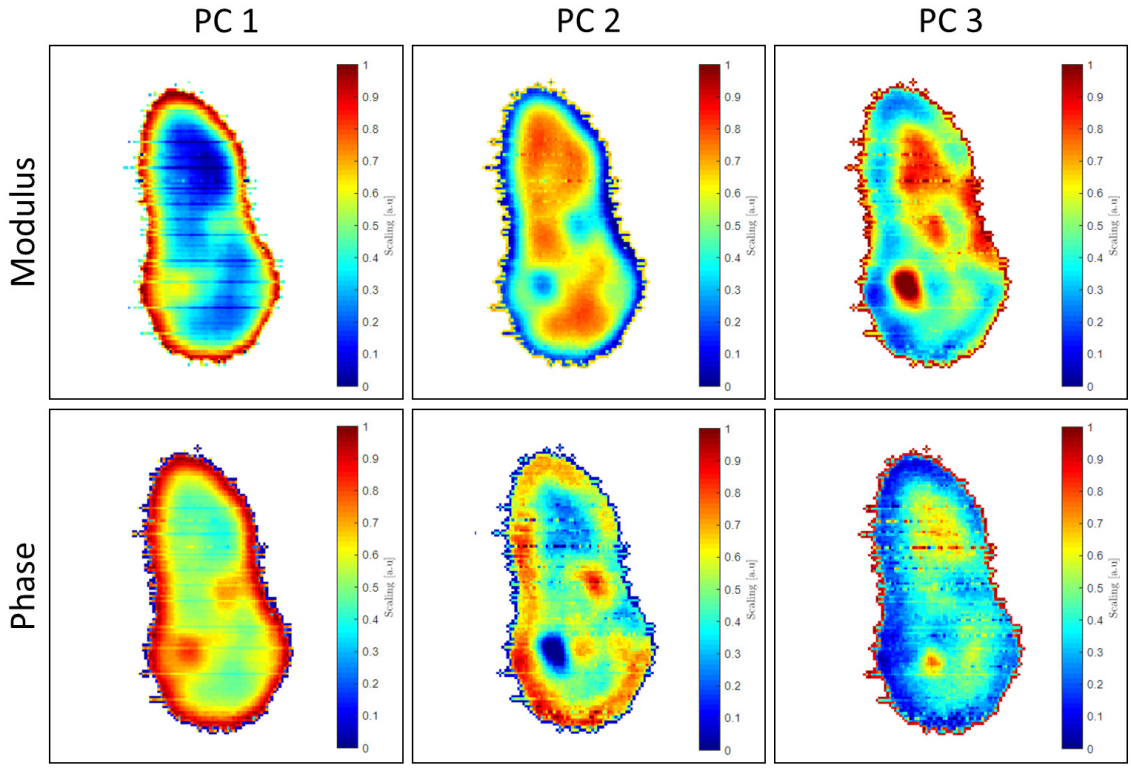


Figure III.35: Integrated loading weighted frequencies for modulus and phase data set for the first three principal components. The reader is referred to Figure III.21 for eyeball comparison.

Overall, it is difficult to precisely understand the origin of the contrast that are depicted in images sorted from the principal component analysis. Therefore it is almost impossible to determine a method for classification. However, information that was extracted from the first two principal component of the spectra and from the first component of the phase data set seem overall suggest similar areas than the one suggested by the refractive index mapping in the previous section.

To better understand the origin of the contrast, one has to plot the scores, that is to project the pixels on the derived new subspaces based on the newly defined coordinates (scores). These pixels can not be attributed either to the malignant or the benign group solely from PCA. Therefore, to highlight the possible existence of an unsupervised classification during PCA, we will, in the following, distinguish two groups: first, pixels that were classified as malignant from the best method and threshold in the previous section; Second, pixels that were classified as benign from the best method and threshold in the previous section. It is worth to underline, that such a procedure does not aim to qualify PCA for tissue classification, but rather to establish a possible correlation between the contrast exhibited from principal component analysis and pixel malignancy.

As a reminder, in the previous section the best trade-off between true positive rate and false positive rate for the NS2 sample was given by the CT2 strategy with a refractive index threshold set to 2.1. This specific segmentation gave arise to a sensitivity of 67% and a specificity of 70%.

On Figure III.36 are given the pixel projection on the first three component space for the modulus.

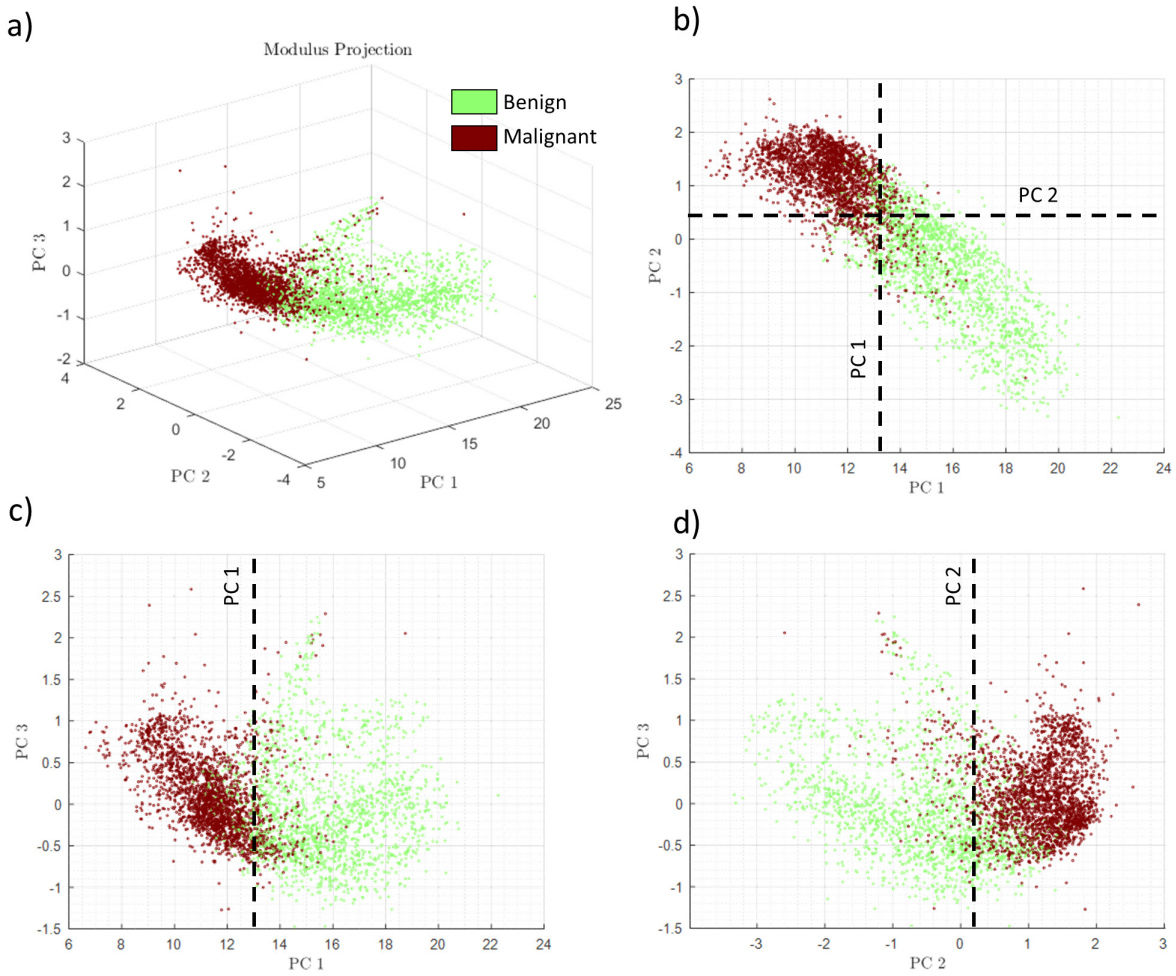


Figure III.36: Modulus data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection.

The projection on the 3D space (Figure III.36 a) whose axis are PC1, PC2 and PC3 clearly shows two distinct groups of pixels. On the left side of the projection, pixels hat are present were previously classified as malignant while on the right part of the projection stand pixels that were identified to belong to the healthy category. Obviously, there is a merger between the two clusters where malignant and healthy pixels are mixed with each others. However, the overall demarcation remains remarkably evident.

On analyzing 2D projections, that is one principal component versus another (Figure III.36 b, c and d), it appears that the first principal component PC1 as well as the second one PC2 contribute to express the biological structure variance among the set of pixels on analyzing the modulus. On Figure III.36 b, c and d are drawn (black dashed lines) estimated thresholds that successfully isolate the malignant pixel cluster from the rest, using PC1 and PC2. On its part, PC3 does not provide any remarkable demarcation between the two groups of pixels.

Overall, the first two principal component give arise to a classification that is compliant with the one that was described in the previous section. Let's underline that such an observation does not imply that PCA offers similar performances towards classification but rather depict its potential in assessing tumor delineation.

Additionally to the analysis on the modulus, the phase was also investigated. On Figure III.37 are given the phase data set projections thanks to PCA.

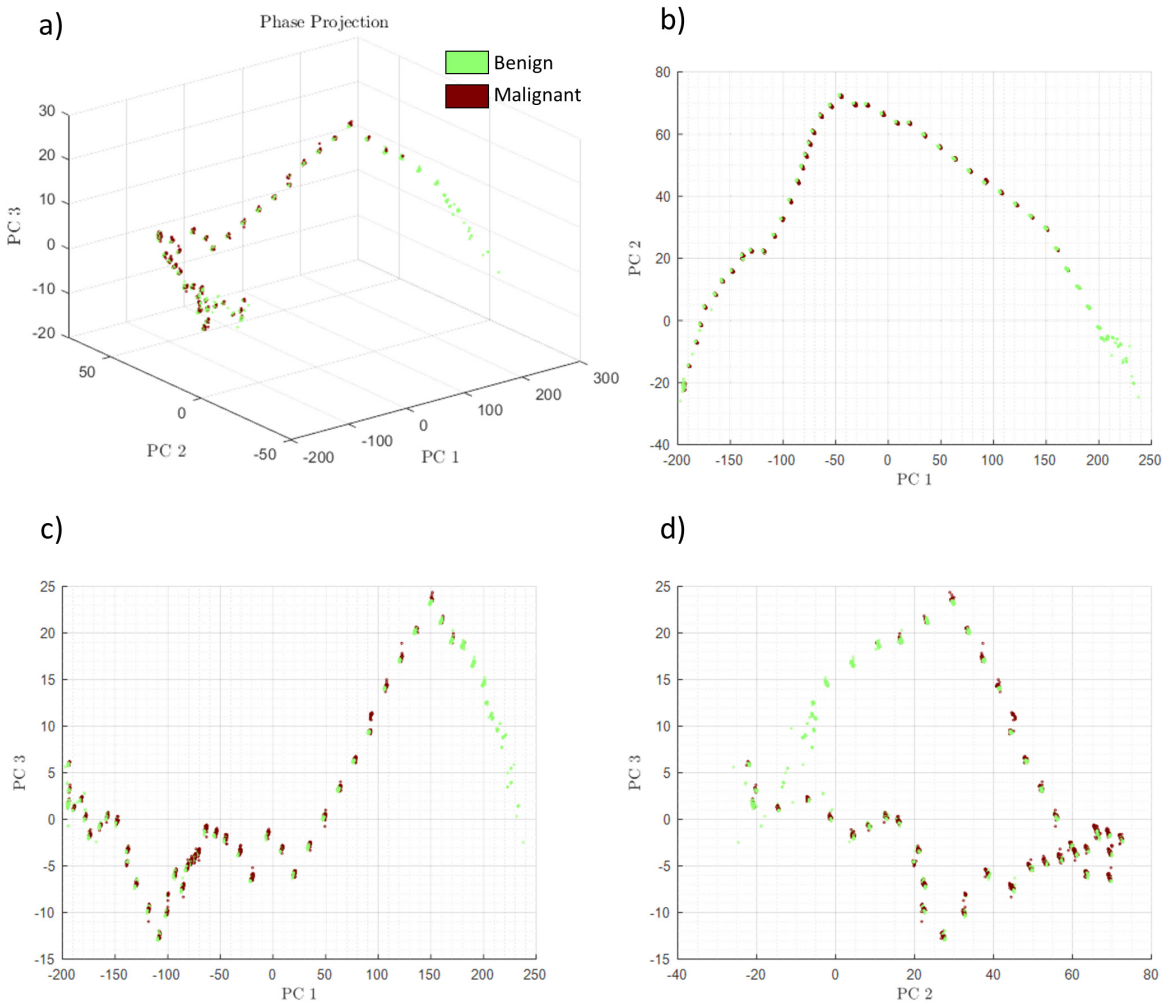


Figure III.37: Phase data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection.

Globally, the analysis towards the phase data set does not reveal specific clustering (Figure III.37 a). Considering the plot of PC2(PC1), PC3(PC1) and PC3(PC2) it is clear that the phase does not categorize pixels as a function of their biological nature. However, this observation does not indicate that phase is not suitable at all for demarcation, but rather that in this precise study case, phase analysis failed to classify pixels in accordance to their belonging to each category.

III.7.3 NS2 Sample

Similarly to the previous sample, the first six principal component for the data set of the modulus and the phase of NS2 sample were calculated and their respective contribution to variance are reported in Table III.10.

Again, the first principal component explains by itself by about 95% of the variance observed within the spectra data set and the phase data one. Let's note that the contribution from the second principal component is not that weak as it expresses roughly 2% of the total variance among the modulus set and by about 3.7% among the phase set. Similarly to the previous study case, further principal component are not contributing that much to the total variance within the sets and are expected to be mainly affected by noise fluctuations.

The first two principal component of each data set are not influenced by the same frequencies. Remarkably, the principal components arising from the analysis of the modulus data set are more

Table III.10: Contribution to variance (%) for the first six principal components based on the modulus and the phase of each pixel between 200 GHz and 1 THz.

	Observable: Modulus	Observable: Phase
Principal Component	Contribution to Variance (%)	Contribution to Variance (%)
PC 1	95.34	95.63
PC 2	1.94	3.67
PC 3	0.83	0.69
PC 4	0.20	0.12
PC 5	0.13	0.10
PC 6	0.11	0.04

greatly impacted by lower frequencies while the phase data set principal components are mostly weighted by high frequencies. Indeed, the first principal component from modulus set is mostly impacted, in descending order, by 330 GHz, 390 GHz and 510 GHz frequencies. Roughly, the same frequency windowing is impacting the modulus second principal component. On the other hand, frequencies that impact the most the weighting among phase principal components are ranging from 690 GHz to 990 GHz.

Images obtained from the trapezoidal integration of loading weighted spectra and phase for the first three principal components are given in Figure III.38.

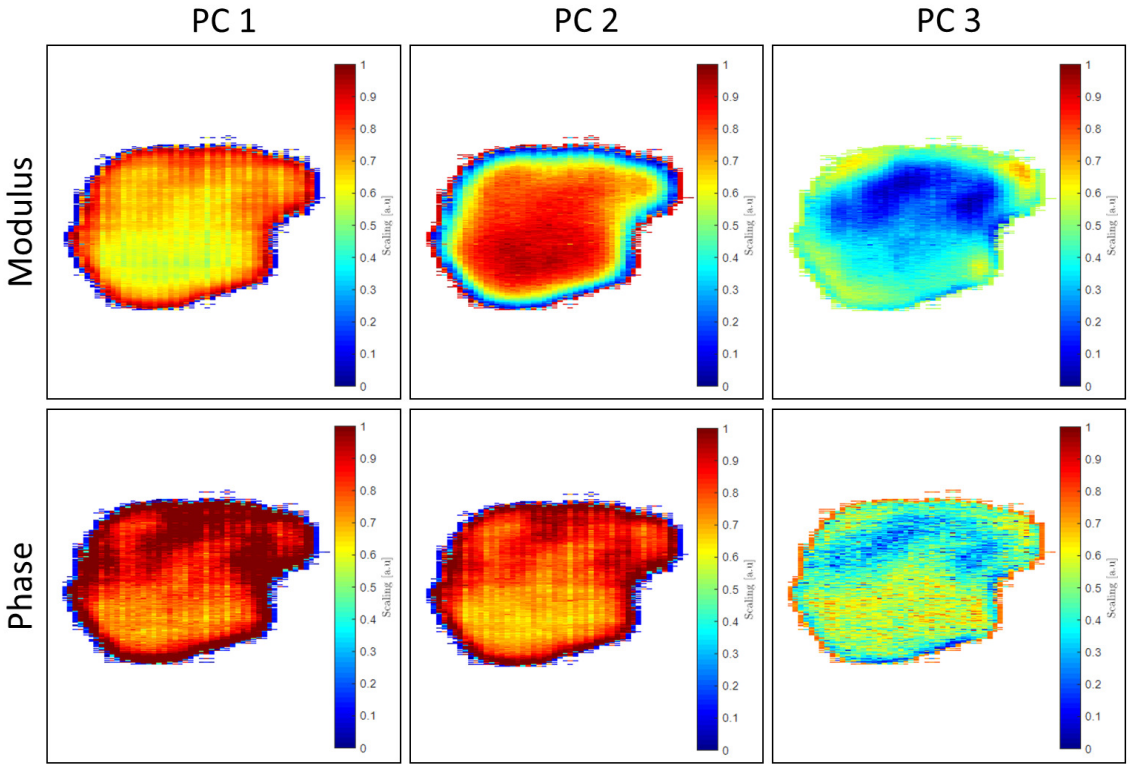


Figure III.38: Integrated loading weighted frequencies for modulus and phase data set for the first three principal components. The reader is referred to Figure III.25 for eyeball comparison.

Overall, interesting contrast are depicted for the first two principal components. Additionally, some contrast are exhibited by the third principal components. However, it is worth mentioning that the third principal component that was calculated for the phase data set seems to be widely affected by noise.

In order to better understand the contrast that is exhibited on each images, it is suitable to plot the scores related to each pixels. In other terms, it consists in projecting the pixels in the calculated subspace based on their newly defined coordinates. These pixels can not be classified as

malignant or benign solely from PCA. Therefore, to highlight the possible existence of an unsupervised classification during PCA, we will, in the following, distinguish two groups: first, pixels that were classified as malignant from the best method and threshold in the previous section; Second, pixels that were classified as benign from the best method and threshold in the previous section. It is worth to underline, that such a procedure does not aim to qualify PCA for tissue classification, but rather to establish a possible correlation between the contrast exhibited from principal component analysis and pixel malignancy.

As a reminder, in the previous section the best trade-off between true positive rate and false positive rate for the NS2 sample was given by the CT3 strategy with a refractive index threshold set to 2.6. This specific segmentation gave arise to a sensitivity of 79% and a specificity of 82%.

On Figure III.39 are given the pixel projection on the first three component space for the modulus.

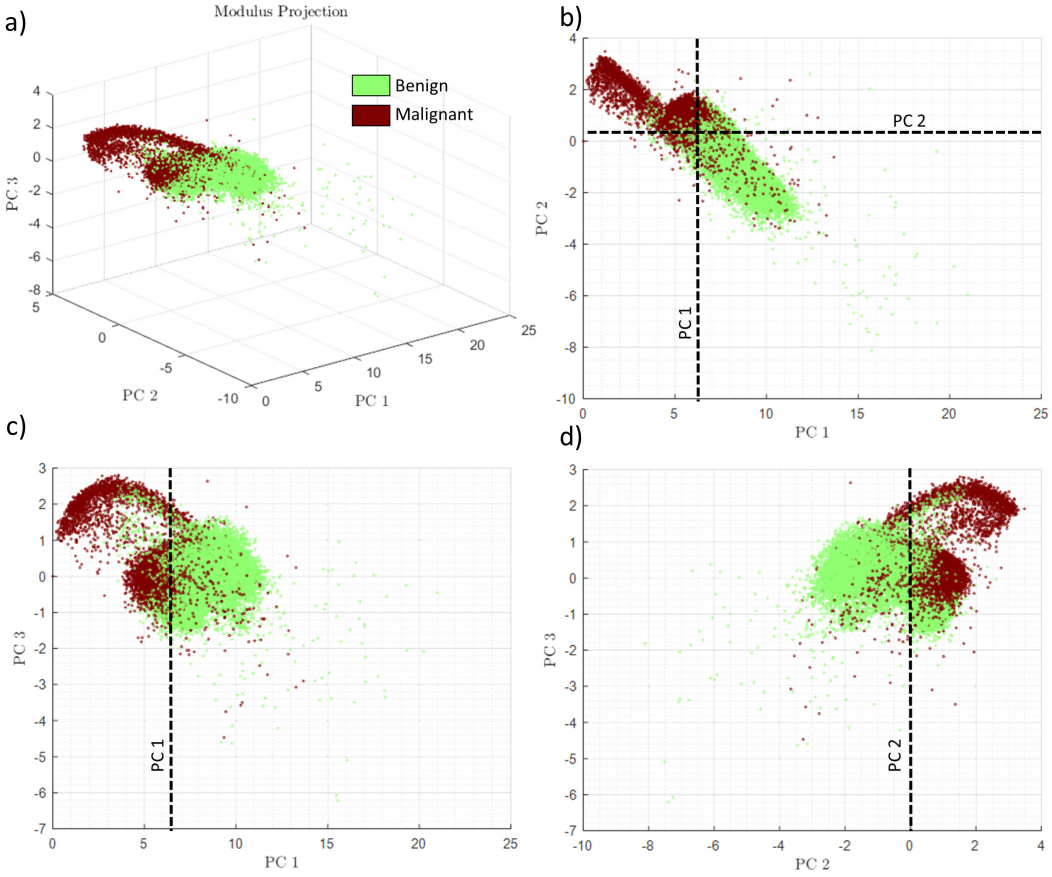


Figure III.39: Modulus data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection.

The 3D projection of the spectra data set clearly highlights a main lobe where a wide quantity of pixels is present. Most of these pixels were previously classified as malignant. However, a non-negligible amount of pixels that is comprised within the lobe were previously marked as benign. Especially, on the south-east part of the lobe one can observe a great density of benign pixels (Figure III.39 a). Remarkably, apart from the lobe, one can see a tail-like shape that mainly comprised pixels belonging to the benign category. Overall, plotting PC2(PC1) (Figure III.39 b) well described the previous observation. It is therefore reasonable to state that PCA on pixel spectra may have some potential towards breast tissue classification. However it is important to keep in mind that such a statement remains -at this stage- solely hypothetical. In a interesting manner, the second principal component seems to actively participate in the separation of benign pixels from malignant ones as depicted by the black dashed lines on Figure III.39 b, despite it's relatively weak contribution to variance ($\sim 2\%$). Plotting PC3(PC1) and PC3(PC2) (Figure III.38 c and d) does

not give arise to specific additional clustering. That validates the idea that the third component would mainly be dictated by noise fluctuations. Subsequently, further principal components are therefore not aimed to explain the spectra set variance in a biological way.

PCA capabilities were also inspected on the phase data set. The phase projections are given in Figure III.40.

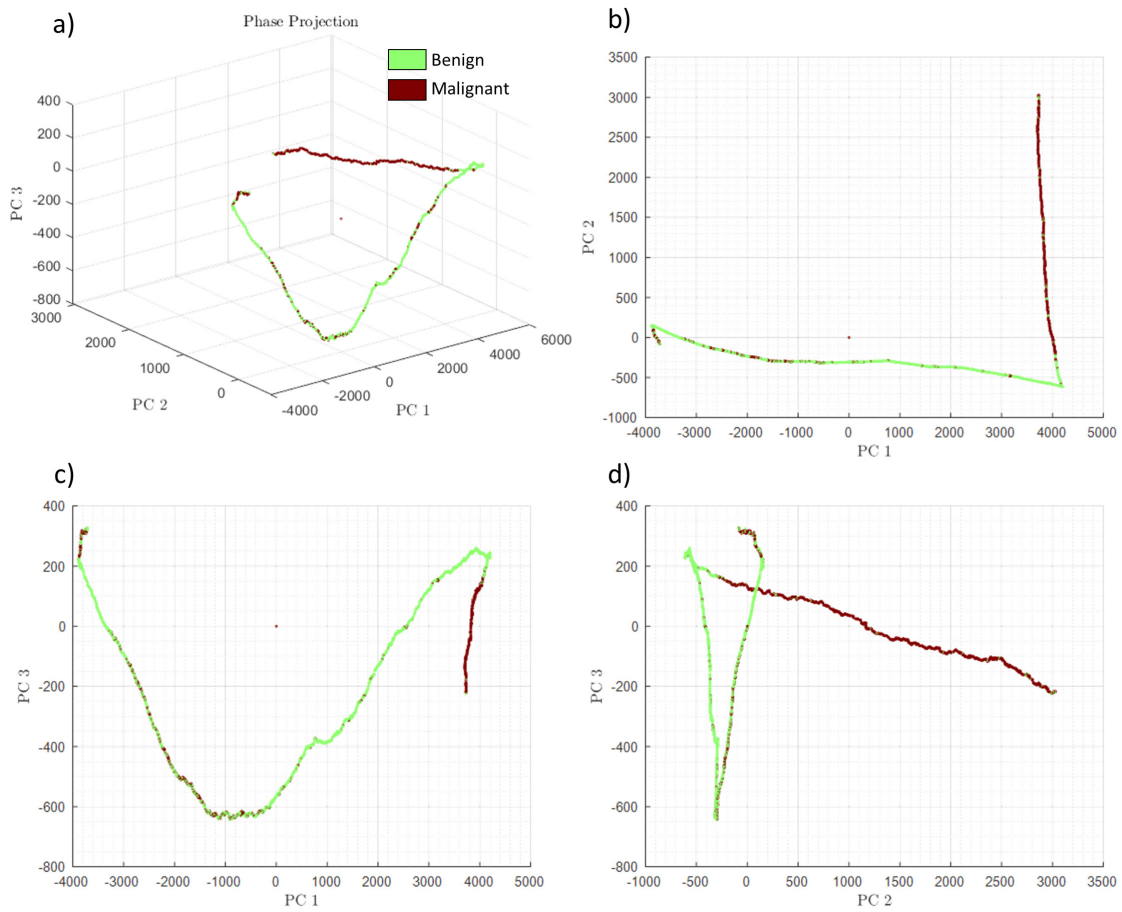


Figure III.40: Phase data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection.

Although it seems from the 3D projection (Figure III.40 a) that some distinctions between malignant and healthy pixels are feasible, these features do not appear anymore on plotting the projection in PC2(PC1), PC3(PC1) and PC3(PC2) (Figure III.40 b, c and d). The reason behind it is overall quite simple. Due to the wide number of pixels that is considered, dots that represent the projection of these pixels are superimposed. Hence, it is not possible -from this study- to state on the capabilities of principal component analysis on the phase data set to discriminate between healthiness and malignancy.

III.7.4 NS3 Sample

Last to be reported, the NS3 sample that was previously pointed out to suffer from the lack of accuracy concerning the pathology - terahertz image correlation. The first six principal component contribution to variance for NS3 sample are given in Table III.11.

Overall, the same tendency is shown that is, the first principal component almost entirely explains the variance within the modulus set. However, it appears that such an observation is not valid towards the phase data set. Indeed, the contribution to variance of the first principal component drops below 83% while the second and the first principal component have a contribution that reaches approximately 10.3% and 6.3%, respectively. Nevertheless, we have observed in the foregoing, that the phase was -overall- poorly contributing to express biological demarcation within

pixels. Therefore, one can anticipate that this different contribution to variance distribution is not due to specific biological features.

Table III.11: Contribution to variance (%) for the first six principal components based on the modulus and the phase of each pixel between 200 GHz and 1 THz.

	Observable: Modulus	Observable: Phase
Principal Component	Contribution to Variance (%)	Contribution to Variance (%)
PC 1	97.70	82.97
PC 2	0.51	10.27
PC 3	0.44	6.33
PC 4	0.11	0.21
PC 5	0.10	0.08
PC 6	0.08	0.06

Overall, the most notable influential frequencies on modulus data set are located from 300 GHz to 600 GHz with highest weight for 330 GHz, 510 GHz, 480 GHz and 540 GHz which are all almost equivalent in terms of influence. On the contrary, important weighting towards the phase data set are generically observed between 660 GHz and 990 GHz. Images that correspond to the integration of the weighted frequency band are given in Figure III.41.

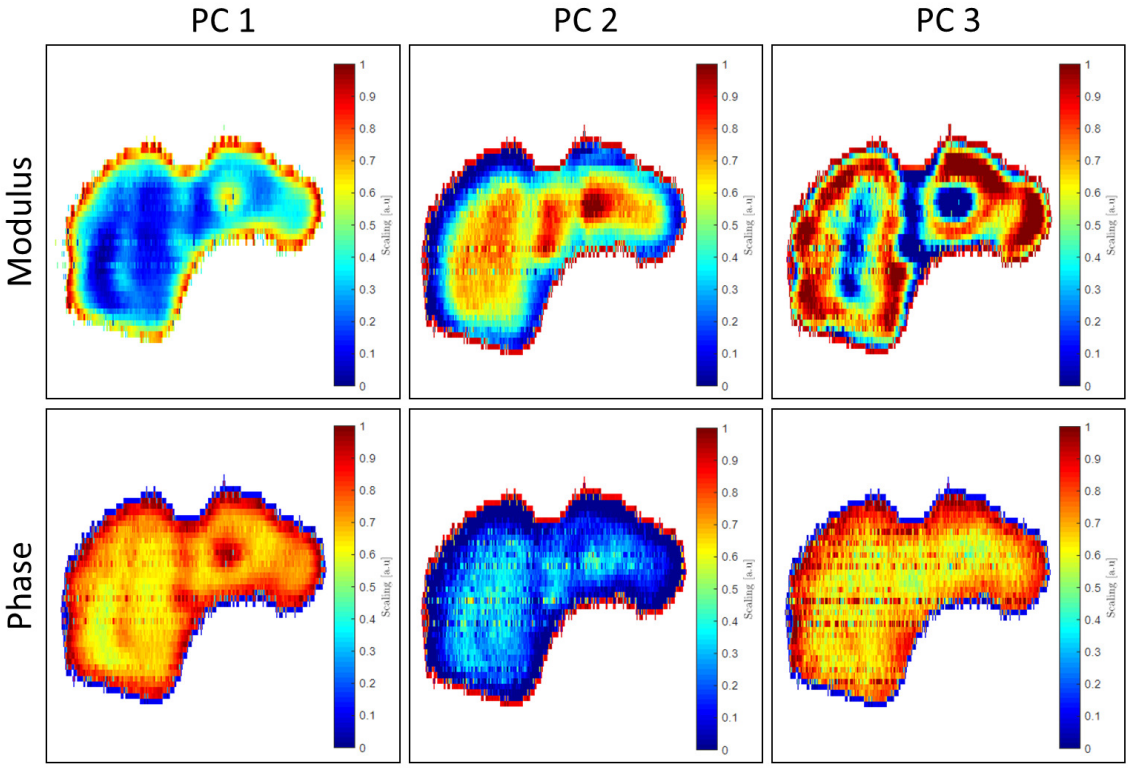


Figure III.41: Integrated loading weighted frequencies for modulus and phase data set for the first three principal components. The reader is referred to Figure III.29 for eyeball comparison.

Globally, some contrast clearly appears on each image that results from the integration of loading weighted frequencies. However, as already pointed out, to explore the possible existence of a correlation between the observed contrast and tissue malignancy, it is favorable to analyze the distribution of original pixels once projected into the newly derived space. The score plots for the modulus data set are given in Figure III.42.

Overall, one can still observe a rough demarcation between pixels that were classified as malignant and benign (Figure III.42 a). Remarkably, the first principal component still shows an approximate clustering as depicted by the black dashed lines on Figure III.42 b and c. Nevertheless,

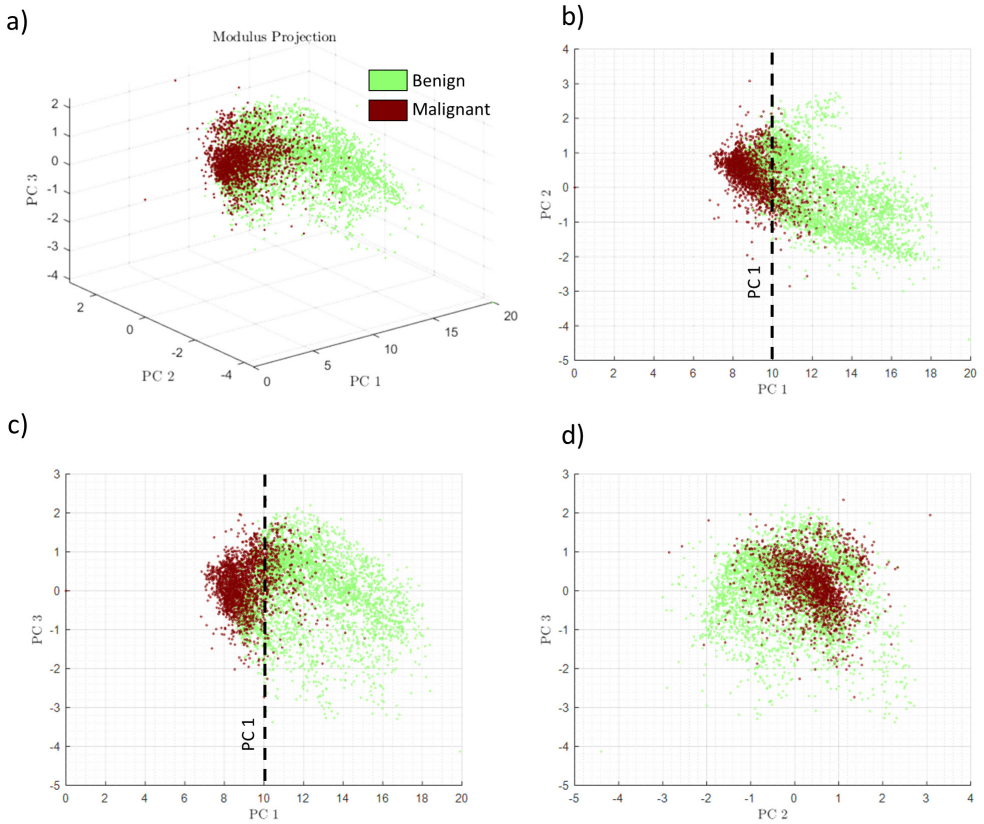


Figure III.42: Modulus data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection.

in comparison to those of the two previous sample NS1 and NS2, the second principal component of the present tissue sample does not provide clear delineation. Indeed, it is not possible to approximately determine a threshold in PC2 that would separate malignant from healthy pixels (Figure III.42 c and d). It is worth to underline that the second principal component contribution to variance is -in this case- lower than for other samples.

Concerning the phase data set analysis, the projections are given in Figure III.43. Similarly to other samples, the phase data set analysis through PCA does not give any information at all. In the conclusion dedicated to principal component analysis, we will address this specificity and try to comprehend the reasons why phase does not seem to be efficient towards tissue segmentation via PCA.

III.7.5 Discussion on PCA

In the section dedicated to principal component analysis, we have investigated the potential of such a transformation towards breast carcinoma detection. As PCA projects the original data set on a subspace that aims to highlight data collection variance, one can therefore reasonable expect that it would enhance the contrast originally observed within the tissue section.

Heretofore, PCA was performed onto the raw modulus and phase of pixels for frequencies ranging between 200 GHz to 1 THz. Remarkably, frequencies that contribute the most to the overall variance are not similar for the modulus data set than for the phase one. Influential frequencies are globally comprised between 300 GHz and 600 GHz for the modulus while they stand from 660 GHz to 990 GHz for the phase. Instead of merely reporting such an observation, could we not construct a reasonable hypothesis that may explain it? Regarding the frequencies that mostly affect the variance towards the modulus, let's remember that in the previous chapter we have underlined the frequency band [200 - 600] GHz to present the highest contrast between the refractive index of adipose tissue and fibrous/cancer tissue. The amplitude of the modulus is dictated by

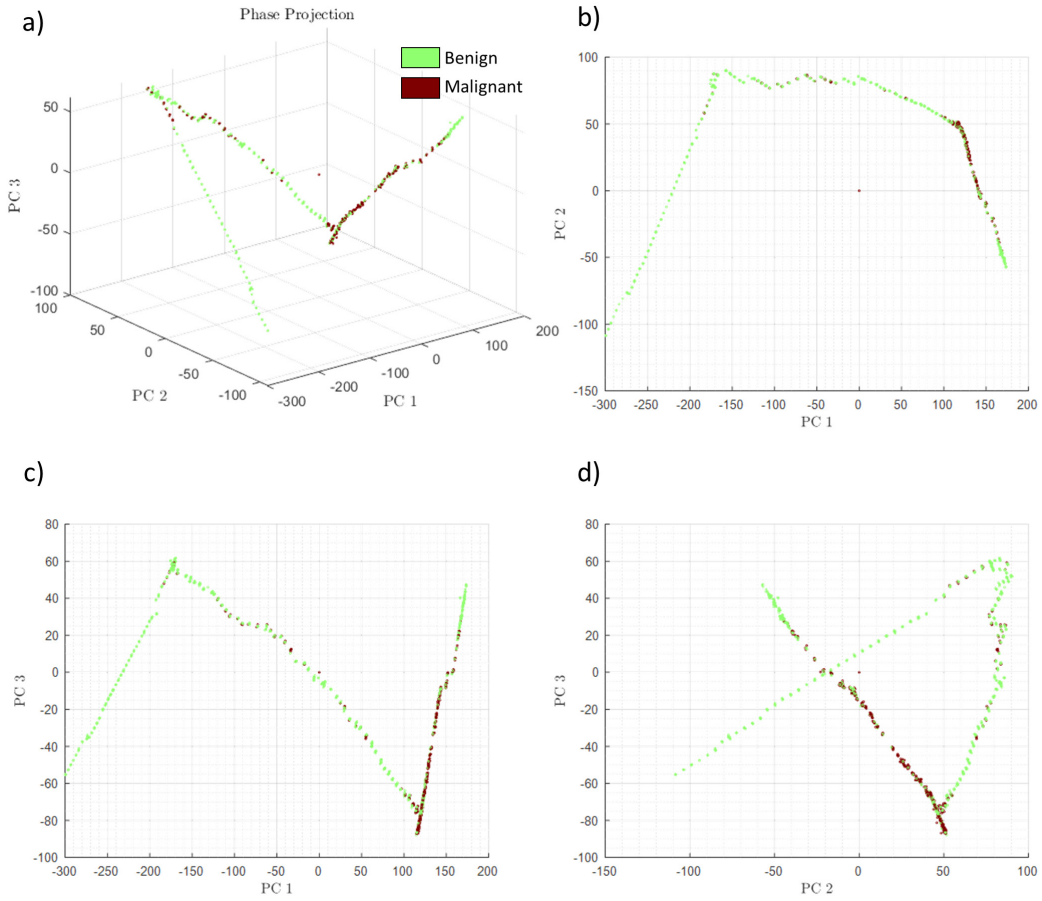


Figure III.43: Phase data set projection on the PC1-PC2-PC3 space. a) 3D projection on PC-PC2-PC3 space. b) PC2(PC1) projection. c) PC3(PC1) projection. d) PC3(PC2) projection.

the proportion of photons that are reflected at the dielectric interface for each frequency. In turn, the fraction of reflected photons directly depends on the refractive index difference at the dielectric interface. As tissues exhibit the highest refractive index difference between 200 GHz and 600 GHz, it is reasonable to state that consequently the highest modulus variance takes place within the same frequency window. Moreover, as the signal-to-noise ratio was shown to be greater at low frequencies, it is clear that such a variance does not reflect noise fluctuation but rather intrinsic biological differences within tissue types. Concerning the phase, it is also important to mention that it is impacted by the refractive index difference at the dielectric interface. Therefore, on solely considering the refractive index difference one could expect that the same frequencies than the modulus would contribute to the global variance. However, it is of paramount importance to have in mind that to study the phase data set, the original phase was unwrapped. Therefore, each pixel exhibits a phase that either increases or decreases in a monotonous way. Let's note that the sign of the phase only depends on the sign of the Fresnel's coefficient at the interface and is, here, not of interest. Therefore, it is obvious that higher frequencies are expected to contribute more than lower frequencies. In a nutshell, the global variance observed towards the analysis of the unwrapped phase data set may not solely reflect biological differences within tissue sections but concomitantly phase direction due to phase unwrapping.

Globally, the first principal component either for the modulus or for the phase was almost entirely explaining the discrepancies among the respective data set. We still can mention that some contribution to variance were observed from the second principal component but drastically lower than the first component. Beyond the first two components the contribution to variance was particularly low and was consequently mainly impacted by internal noise fluctuation. Such a statement was further demonstrated by the inefficiency of the third principal component to roughly cluster malignant and benign pixels.

In order to better understand the contrast that was exhibited by images resulting from the PCA, each pixel was projected on the PC space. To visualize the possible existence of a demarcation between malignant and healthy pixels, each pixel was beforehand marked as malignant or as benign depending on the classification that was established from the best threshold and segmentation method in the previous section. Therefore, one could observe the potential sensitivity of PCA to pixel pathology. For each case, the first principal component derived from the modulus data set has shown rough demarcation between healthy and sick pixel group. It is however important to report that all pixels were not correctly classified or at least were not fully in accordance with the diagnosis established by the previous methodology. Nevertheless, it clearly indicates that the principal component analysis may have some potential towards breast cancer detection, when applied on the terahertz modulus. On the contrary, despite the artistic features described by the projection of the phase data set on the PC space, it was not possible to determine specific features that would have tend to express any demarcation between healthy and malignant pixels.

III.8 Conclusion

The third chapter was devoted to the evaluation of pulsed-terahertz technology towards the identification of malignant lesions within breast sections.

As reflection configuration systems are often reported to be extremely sensitive to the phase, a self-reference approach was proposed that aims to improve sample characterization or at least to diminish reference-sample phase mismatch that globally limits property extraction accuracy. The suggested method consists of calculating the theoretical signal generated by the terahertz emitter from the signal that is recorded at a dielectric interface between a sapphire substrate and the air. Indeed, as far as sapphire and air properties are known, it is possible to derive the signal shape that gave arise to the recorded signal. Hence, the recording of a reference signal from a perfect reflector beforehand sample measurement is not required and consequently it erases the error due to position mismatch between the perfect reflector and the sapphire substrate that carries the sample.

Insofar as the different types of tissue were reported to exhibit distinct response to the terahertz stimuli, specific contrast was expected to be displayed by sample section terahertz images, depending on tissue region biological nature. Evidently, the level of contrast was concomitantly expected to be dictated by the density of cancer cells reported within the breast section.

Hence, electromagnetic problems were solved to express image contrast as a function of the refractive index level. Importantly, either extinction or absorption coefficient mapping were not assessed as the differences between healthy fibrous and tumor were previously reported to be null. The reader is invited to keep in mind that solely the maps at 560 GHz and related subsequent investigations are -in this manuscript- reported.

Based on the refractive index contrast, different strategies towards the section segmentation were tested. In fact, it would not be relevant to simply state on the observation of a contrast. Therefore, tissue section images were binarized depending on the value of the calculated refractive index for each pixel. To do so, various refractive threshold were considered that determine the level above which pixels are classified as malignant and below which, they are marked as benign. Nevertheless, as reported in the dedicated sections, pixels are numerical objects rather than carrying a biological nature. In other words, despite their numerical separation, the presence of malignant cell within a pixel should influence the diagnosis on neighbor pixels. Hence, in order to take into account such an influence, connected segmentations that consider pixel proximity in addition to pixel refractive index to classify pixel in either the malignant or the healthy pixel cluster.

The efficiency of each segmentation was assessed by filling the associated confusion matrix with regards to the corresponding pathology images, annotated by the pathologist. The resulting receiver operating characteristics corresponding to each segmentation were derived. Overall, results highlight that connected thresholding is more sensitive and specific than hard thresholding.

Remarkably, the study on one of the inspected sample reported that connected thresholding correctly localized malignant lesion in approximately 80% of cases while maintaining a specificity in the range of 82% in comparison to its pathology image. However, this statement has to be taken with precautions since the correlation between refractive index map and pathology images are often tricky due to resolution and orientation mismatch between the two pictures and due to tissue deformations that occur during histology slide preparation.

Supplemental studies on principal component analysis were thereafter conducted. Both the modulus and the phase of pixels were projected on subspaces that aim to highlight the discrepancies among the data set. We reported that the first principal component was often explaining most of the variance contained in sets. Pixel distribution in principal component spaces was compared with the diagnosis established upstream with refractive index thresholding. On comparing the refractive index thresholding diagnosis with principal component analysis performed on the modulus data set, it was immediately apparent that the two processes were giving similar results towards pixel categorization. On the contrary, the absence of particular features was reported regarding the phase data set analysis.

Despite all the precautions that have to be taken on considering the results, the different investigations that were conducted all lead to the same idea -that is not anymore a simple hypothesis-: terahertz technology has the potential to localize malignant lesions within breast tissue sections.

Well, that is right but with a certain accuracy. As stated in the beginning of this chapter, the main limitation of far-field imaging lies in its resolution. Hence, to improve even further the efficiency of such a tool toward breast carcinoma recognition and segmentation, resolution below this theoretical limit has to be achieved. To push further the development of terahertz technology as an effective tool to be implemented during breast conserving surgeries, the following and last chapter is devoted to the development and the use of a near-field silicon-based terahertz subwavelength imaging matrix operating at the singular frequency of 560 GHz **[46; 47]**. The operating frequency was determined from three key points: first, the frequency belongs to the spectral band within which the demarcation between cancer and healthy fibrous is expected to be the highest; Second, the results on the diverse numerical treatments at 560 GHz provided effective demarcation between cancer and normal tissue; Last, such a frequency matches new BiCMOS technology capabilities and thus opens up the opportunity for near-field terahertz imaging using such kind of device.

III.9 Bibliography

[1] Samuel K Moore. Better breast cancer detection. *Ieee Spectrum*, 38(5):50–54, 2001. 110

[2] Xin Li, Wei Huang, Elizabeth A Morris, Luminita A Tudorica, Venkatraman E Seshan, William D Rooney, Ian Tagge, Ya Wang, Jingang Xu, and Charles S Springer. Dynamic nmr effects in breast cancer dynamic-contrast-enhanced mri. *Proceedings of the National Academy of Sciences*, 105(46):17937–17942, 2008. 110

[3] Olaf Beckonert, Jürgen Monnerjahn, Ulrich Bonk, and Dieter Leibfritz. Visualizing metabolic changes in breast-cancer tissue using 1h-nmr spectroscopy and self-organizing maps. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, 16(1):1–11, 2003. 110

[4] Winfried Kuhn. Nmr microscopy—fundamentals, limits and possible applications. *Angewandte Chemie International Edition in English*, 29(1):1–19, 1990. 110

[5] Freddy T Nguyen, Adam M Zysk, Eric J Chaney, Jan G Kotynek, Uretz J Oliphant, Frank J Bellafiore, Kendrith M Rowland, Patricia A Johnson, and Stephen A Boppart. Intraoperative evaluation of breast tumor margins with optical coherence tomography. *Cancer research*, 69(22):8790–8796, 2009. 110

[6] Stephen A Boppart, Wei Luo, Daniel L Marks, and Keith W Singletary. Optical coherence tomography: feasibility for basic research and image-guided surgery of breast cancer. *Breast cancer research and treatment*, 84(2):85–97, 2004. 110

[7] Ariel Lipson, Stephen G Lipson, and Henry Lipson. *Optical physics*. Cambridge University Press, 2010. 111

[8] Hou-Tong Chen, Roland Kersting, and Gyu Cheon Cho. Terahertz imaging with nanometer resolution. *Applied Physics Letters*, 83(15):3009–3011, 2003. 111

[9] Eli Peli. Contrast in complex images. *JOSA A*, 7(10):2032–2040, 1990. 113

[10] Jean-Baptiste Perraud. *Fast 3D terahertz imaging*. PhD thesis, Bordeaux, 2018. 113

[11] Shengyang Huang, Philip C Ashworth, Kanis WC Kan, Yang Chen, Vincent P Wallace, Yuanting Zhang, and Emma Pickwell-MacPherson. Improved sample characterization in terahertz reflection imaging and spectroscopy. *Optics express*, 17(5):3848–3854, 2009. 114

[12] Amel Al-Ibadi. *Terahertz imaging and spectroscopy of biomedical tissues: application to breast cancer detection*. PhD thesis, Bordeaux, 2018.

[13] A Al-Ibadi, Q Cassar, T Zimmer, G MacGrogan, L Mavarani, P Hillger, J Grzyb, UR Pfeiffer, JP Guillet, and P Mounaix. Thz spectroscopy and imaging for breast cancer detection in the 300–500 ghz range. In *2017 42nd international conference on infrared, millimeter, and terahertz waves (IRMMW-THz)*, pages 1–1. IEEE, 2017.

[14] Quentin Cassar, Amel Al-Ibadi, Laven Mavarani, Philipp Hillger, Janusz Grzyb, Gaëtan MacGrogan, Thomas Zimmer, Ullrich R Pfeiffer, Jean-Paul Guillet, and Patrick Mounaix. Pilot study of freshly excised breast tissue response in the 300–600 ghz range. *Biomedical optics express*, 9(7):2930–2942, 2018. 114

[15] David L Donoho and Jain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455, 1994. 117

[16] David L Donoho. De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3):613–627, 1995.

[17] David L Donoho, Iain M Johnstone, Gérard Kerkyacharian, and Dominique Picard. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(2):301–337, 1995.

[18] Ronald R Coifman and David L Donoho. Translation-invariant de-noising. In *Wavelets and statistics*, pages 125–150. Springer, 1995. 117

[19] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. Siam, 1992. 118

[20] Abiodun Ogunleke. *3D chemical imaging of brain tumors*. PhD thesis, Bordeaux, 2019. x, 122

[21] Tanny Chavez, Nagma Vohra, Jingxian Wu, Keith Bailey, and Magda El-Shenawee. Breast cancer detection with low-dimension ordered orthogonal projection in terahertz imaging. *IEEE Transactions on Terahertz Science and Technology*, 2019. 124

[22] Mohamed Walid Ayech and Djemel Ziou. Segmentation of terahertz imaging using k-means clustering based on ranked set sampling. *Expert Systems with Applications*, 42(6):2959–2974, 2015. 124

[23] Yury V Kistenev, Alexey V Borisov, Maria A Titarenko, Olga D Baydik, and Alexander V Shapovalov. Diagnosis of oral lichen planus from analysis of saliva samples using terahertz time-domain spectroscopy and chemometrics. *Journal of biomedical optics*, 23(4):045001, 2018. 124

[24] Xianhua Yin, Wei Mo, Qiang Wang, and Binyi Qin. A terahertz spectroscopy nondestructive identification method for rubber based on cs-svm. *Advances in Condensed Matter Physics*, 2018, 2018. 124

[25] Haishun Liu, Zhenwei Zhang, Xin Zhang, Yuping Yang, Zhuoyong Zhang, Xiangyi Liu, Fan Wang, Yiding Han, and Cunlin Zhang. Dimensionality reduction for identification of hepatic tumor samples based on terahertz time-domain spectroscopy. *IEEE Transactions on Terahertz Science and Technology*, 8(3):271–277, 2018. 124

[26] Kurt Rossmann. Point spread-function, line spread-function, and modulation transfer function: tools for the study of imaging systems. *Radiology*, 93(2):257–272, 1969. 125

[27] Dilpreet Kaur and Yadwinder Kaur. Various image segmentation techniques: a review. *International Journal of Computer Science and Mobile Computing*, 3(5):809–814, 2014. 146

[28] Anthony J Fitzgerald, Vincent P Wallace, Sarah E Pinder, Anand D Purushotham, Padraig O'Kelly, and Philip C Ashworth. Classification of terahertz-pulsed imaging data from excised breast tissue. *Journal of biomedical optics*, 17(1):016005, 2012. 147

[29] Tanny Chavez, Tyler Bowman, Jingxian Wu, Keith Bailey, and Magda El-Shenawee. Assessment of terahertz imaging for excised breast cancer tumors with image morphing. *Journal of Infrared, Millimeter, and Terahertz Waves*, 39(12):1283–1302, 2018. 147

[30] Alfonso Rojas Domínguez and Asoke K Nandi. Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recognition*, 42(6):1138–1148, 2009. 147

[31] Hossam M Moftah, Ahmad Taher Azar, Eiman Tamah Al-Shammari, Neveen I Ghali, Aboul Ella Hassanien, and Mahmoud Shoman. Adaptive k-means clustering algorithm for mr breast image segmentation. *Neural Computing and Applications*, 24(7-8):1917–1928, 2014. 147

[32] Ruey-Feng Chang, Wen-Jie Wu, Woo Kyung Moon, and Dar-Ren Chen. Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. *Breast cancer research and treatment*, 89(2):179, 2005. 147

[33] Rahimeh Rouhi, Mehdi Jafari, Shohreh Kasaei, and Peiman Keshavarzian. Benign and malignant breast tumors classification based on region growing and cnn segmentation. *Expert Systems with Applications*, 42(3):990–1002, 2015. 147

[34] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. 148

[35] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987. 148

[36] Nandakishore Kambhatla and Todd K Leen. Dimension reduction by local principal component analysis. *Neural computation*, 9(7):1493–1516, 1997. 148

[37] Donald A Jackson and Yong Chen. Robust principal component analysis and outlier detection with ecological data. *Environmetrics: The official journal of the International Environmetrics Society*, 15(2):129–139, 2004. 148

[38] Craig Rodarmel and Jie Shan. Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science*, 62(2):115–122, 2002. 148

[39] SIV Sousa, Fernando Gomes Martins, MCM Alvim-Ferraz, and Maria C Pereira. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, 22(1):97–103, 2007. 148

[40] Joyce Bou Sleiman. *Imagerie et spectroscopie térahertz : application aux problématiques de défense et de sécurité*. PhD thesis, Bordeaux, 2016. 148

[41] Q Cassar, A Al-Ibadi, L Mavarani, P Hillger, J Grzyb, G MacGrogan, UR Pfeiffer, T Zimmer, JP Guillet, and P Mounaix. Studies on pca for breast tissue segmentation. In *2018 43rd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2. IEEE, 2018. 148

[42] A Al-Ibadi, J Bou Sleiman, Q Cassar, Gaëtan Macgrogan, Hugo Balacey, T Zimmer, Patrick Mounaix, and Jean-Paul Guillet. Terahertz biomedical imaging: From multivariate analysis and detection to material parameter extraction. In *2017 Progress In Electromagnetics Research Symposium-Spring (PIERS)*, pages 2756–2762. IEEE, 2017. 148

[43] Herman Wold. Estimation of principal components and related models by iterative least squares. *Multivariate analysis*, pages 391–420, 1966. 149

[44] Herman Wold. Path models with latent variables: The nipals approach. In *Quantitative sociology*, pages 307–357. Elsevier, 1975. 149

[45] Gene H Golub and Christian Reinsch. Singular value decomposition and least squares solutions. In *Linear Algebra*, pages 134–151. Springer, 1971. 149

[46] Laven Mavarani, Philipp Hillger, Thomas Bücher, Janusz Grzyb, Ullrich R Pfeiffer, Quentin Cassar, Amel Al-Ibadi, Thomas Zimmer, Jean-Paul Guillet, Patrick Mounaix, et al. Nearsense–advances towards a silicon-based terahertz near-field imaging sensor for ex vivo breast tumour identification. *Frequenz*, 72(3-4):93–99, 2018. 161

[47] Ullrich R Pfeiffer, Philipp Hillger, Ritesh Jain, Janusz Grzyb, Thomas Bucher, Quentin Cassar, Gaetan MacGrogan, Jean-Paul Guillet, Patrick Mounaix, and Thomas Zimmer. Ex vivo breast tumor identification: Advances toward a silicon-based terahertz near-field imaging sensor. *IEEE Microwave Magazine*, 20(9):32–46, 2019. 161

Chapter IV

Breast Cancer Imaging: Near-field Survey

The nice thing about doing a crossword puzzle is, you know there is a solution.

Stephen Sondheim

Sommaire

IV.1 Introduction . . . 166
IV.2 Beating the Diffraction Limit of Resolution . . . 166
IV.2.1 Far-field versus Near-field Regimes . . . 166
IV.2.2 Terahertz Near-field Probing . . . 168
IV.2.3 Challenges for Biomedical Application: Sensor Integration . . . 172
IV.2.4 Resonator-Based Probing . . . 173
IV.3 Towards a Silicon-Based Terahertz Near-field Imaging Matrix . . . 175
IV.3.1 Split-Ring Resonators (SRR) . . . 175
IV.3.2 Shift Tracking . . . 177
IV.3.3 Cross-Bridged Dual-Square SRR . . . 177
IV.3.4 Sensor Response Simulation . . . 178
IV.3.5 Matrix Implementation . . . 186
IV.3.6 Pixel Readout . . . 187
IV.3.7 Encapsulation and Packaging . . . 187
IV.3.8 Performances . . . 188
IV.3.9 Experiments on Liquids . . . 189
IV.3.10 Biomedical Imaging: a Premise . . . 192
IV.4 Conclusion . . . 194
IV.5 Bibliography . . . 195

IV.1 Introduction

The last chapter, *Terahertz Imaging: Near-field Survey* reports the development of a near-field array imager to break the diffraction limit of resolution at terahertz frequencies. The imager has been designed based on breast tissue behaviors to external terahertz stimuli that were reported throughout the previous chapters. Its design ultimately aims to provide an intra-operative device to assess the freshly excised breast tissue margin cleanliness with subwavelength resolution.

Beforehand, an overview of the fundamental concepts to sense the near-field is provided, whereupon different approaches and existing applications are reported.

The last section of this chapter is dedicated to the basic theoretical physics behind the implementation of such a sensor, the description of the approaches for its development, the evaluation of its theoretical response under various conditions, and ultimately, its use towards breast cancer recognition.

IV.2 Beating the Diffraction Limit of Resolution

Anyone who has ever used a microscope or an imaging system has wished that it had a slightly greater resolution. However, as stated in *Chapter III.2.1 Image and Limitation*, the resolution of any light-based microscope and imager is limited by around $\lambda/2NA$ wherein λ is the considered wavelength and NA is the numerical aperture. In fact, when light is focused via an objective, rays do not converge to an infinitely sharp focal point but rather form a hazy spot of finite size. The profile of a spot, which defines the point spread function (PSF) previously defined, has roughly the same width as that of the focal spot aforementioned. Therefore, two points separated by a distance lower than the width of the PSF will appear as a unique object, consequently making them indistinguishable from each other. For instance the spatial resolution of an imaging system that probes at 500 GHz is about 0.3 mm if considering $NA = 1$ (free space being the propagation "medium" and the aperture being hemispherical). Hence, the profile response of biological entities that are separated by a distance less than 0.3 mm will be averaged. On comparing the resolution at 500 GHz of such an imager with the size of many sub-cellular structures like proteins, DNA or carbohydrates, it follows that they are unresolvable by far-field terahertz spectroscopy and imaging. Considering now that the typical diameter of eukaryote cells is by about tens of microns, such a system does not either manage to resolve at the cellular level.

One direct way to increase the resolution is to increase the refractive index of the propagation medium. In light microscopy, oil-immersion is a commonly employed technique. However, let's remind that absorption would drastically limit the propagation path to short distances in the case of terahertz waves. Interestingly, some liquids as paraffin could however be of interests towards increasing terahertz far-field imager resolution as its partial transparency in the frequency band was reported **[1]**. Instead of liquid-immersion, solid-immersion of terahertz waves was successfully reported that enhanced the spatial resolution by a factor of 2.43 **[2]**.

But overall, is it really breaking the diffraction limit ? It rather turns to tame it than beating it. The diffraction limit impacts light that has propagated for a distance that is substantially larger than its intrinsic wavelength. Hence, one way to bypass the limitation is to place the sample to be inspected close to the excitation source or detection probe **[3]**. Such an approach is referred to as near-field sensing and the prevalent method to collect information about the near-field are known as near-field scanning optical microscopy (NSOM).

IV.2.1 Far-field versus Near-field Regimes

In light of the above, two different propagation regimes have to be distinguished, namely the far-field and the near-field regimes. So how are these distinguished from one another ? Actually the singular value of $\lambda/2\pi$ distinguishes these two, and is called the transition regime. Let us consider an electromagnetic wave emitter located at random position. While anything that propagates for

a distance smaller than $\lambda/2\pi$ from this source is regarded as the near-field, anything that is propagating for a greater distance is governed by far-field phenomena (Figure IV.1). That's it for the definition, but what does it imply on wave characteristics ?

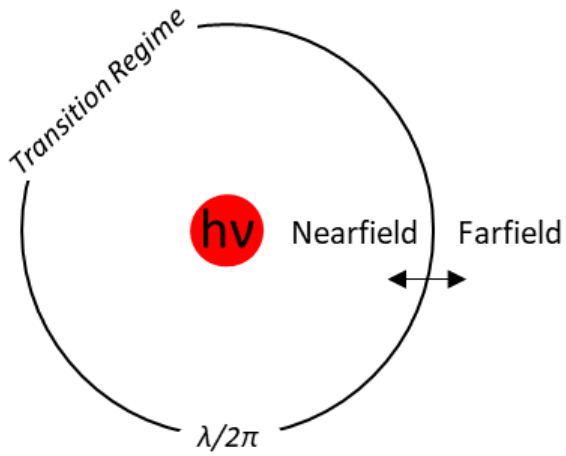


Figure IV.1: Delineation between the near-field domain and the far-field regime. The near-field regime is defined as the sphere of radius $\lambda/2\pi$ whose center is the punctual electromagnetic source. Beyond the near-field sphere, interactions with matter should be regarded as far-field effects.

The first thing that may come to mind is the shape of the electric and magnetic wave, $\vec{E}$ and $\vec{H}$ respectively. In fact, the diffraction limit impacts waves that propagate for an infinite distance from a punctual source. To ensure the propagation of a maximum of rays to the desired location, convex lenses are used to collect and to direct them. Consequently, $\vec{E}$ and $\vec{H}$ components of the propagating wave in far-field regime are strictly planar. On the contrary, as waves are not collected in near-field regime, both $\vec{E}$ and $\vec{H}$ are either spherical or cylindrical waves.

The second point that is of interest is the electromagnetic wave impedance $\eta(\omega)$ that is usually defined as the ratio of the strength of the electric and magnetic fields:

$$\eta(\omega) = \sqrt{\frac{j\omega\mu(\omega)}{j\omega\epsilon(\omega)+\sigma}}, \tag{IV.1}$$

where, j is the imaginary unit, ω is the angular frequency, $\mu(\omega)$ is the frequency-dependent magnetic permeability of the propagating medium, σ is the medium conductivity and $\epsilon(\omega)$ stands for the frequency-dependent electric permittivity of the propagating medium. It is worth to underline that here, $\mu(\omega) = \mu_0\mu_r(\omega)$ with μ_0 being the permeability of free space and μ_r being the relative permeability of the medium. In an equivalent manner, $\epsilon(\omega) = \epsilon_0\epsilon_r(\omega)$. On considering a perfect dielectric, that is a null conductivity, we end up with:

$$\eta(\omega) = \sqrt{\frac{\mu(\omega)}{\epsilon(\omega)}}. \tag{IV.2}$$

As stated in the foregoing, in far-field regime, both $\vec{E}$ and $\vec{H}$ are planar waves, which therefore induces that they are in phase with each other. Consequently, the nature of the wave in far-field regime is mainly dictated by the propagation medium. On the other hand, in near-field $\vec{E}$ and $\vec{H}$ are not in phase with each other. Hence, the electromagnetic wave impedance $\eta(\omega)$ depends on the source characteristics rather than the propagation medium. Therefore in near-field, the effects caused by the light-matter interaction will mainly depend on the predominant electromagnetic component, that is either $\vec{E}$ or $\vec{H}$, emitted by the electromagnetic energy source. In a nutshell, while the nature of the wave in near-field regime depends on the source, in far-field it is mainly dictated by the propagation medium.

IV.2.2 Terahertz Near-field Probing

Surpassing the diffraction limit of resolution within the terahertz band was focused by numerous research teams during the last decades. Often based on existing methods derived from optic wavelengths, some of the most promising near-field terahertz technologies successfully provide spatial resolution as high as $\lambda/3000$ **[4]**. The state-of-the-art of the super-resolution terahertz probing is hereafter briefly discussed.

Before carrying this on, it is worth noting that there is a consistent and inherent quest to any imaging that tends to reduce the developped technique solely to its spatial resolution.

Overall, three main ways are commonly employed: first, the determination of the smallest line pair element, of an USAF 1951 test chart, that is resolved; Second, the measurement of the imaging system response to a sharp object standing across the optical path, also referred in the literature to as the line spread function; Last, the edge response that is globally similar to the line spread function but slightly differs as it consists in recording the response from and edge rather than a line. Here, the pursued goal is not to debate on these methods and their associated relevance. Therefore, spatial resolution of the techniques that are to be reported are given as stipulated by authors. Ultimately, the reader will make his/her own idea about their respective accuracy.

Source in Close Vicinity

The most intuitive way to satisfy near-field sensing requirements is likely a "in-contact" approach. Lecaque *et al* have developed the concept of generating terahertz directly at the sample surface **[5]**. To this aim, the sample to be inspected is placed on a 500 µm-thick second order non-linear $\chi^{(2)}$ ZnTe crystal. A typical femtosecond laser is focused at the rear of the crystal to drive terahertz radiation generation. Detection was achieved with a bolometer and a typical spatial resolution of $\lambda/3$ was achieved (Figure IV.2 a). Later on, the same group employed a thinner ZnTe crystal (~ 200 µm) allowing to achieve a $\lambda/6$ spatial resolution when combined with a specific deconvolution algorithm **[6]**. Overall, it is obvious that performing so has the advantage of recording a terahertz near-field signal that is not related to the sample surface profile.

Alternatively, Yi *et al* coupled a InAs crystal to the tip of an optical fiber through which is passes through the pump beam **[7]**. The sample to be inspected adjoins the InAs crystal (Figure IV.2 b). A spatial resolution by about $\lambda/3$ at 500 GHz was reached by the authors.

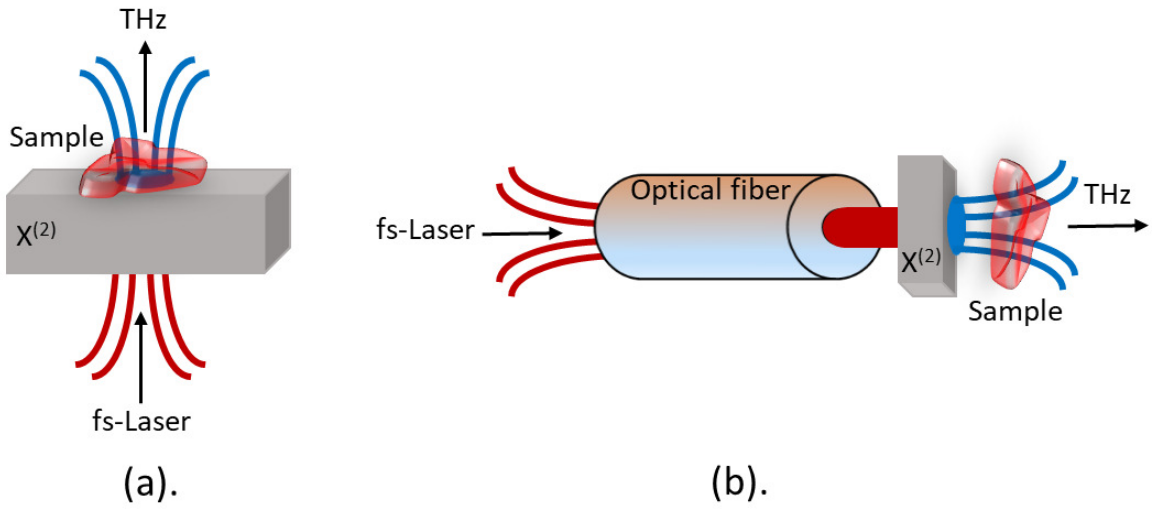


Figure IV.2: Near-field sensing with terahertz source near the sample to be inspected. a: approach developed by Lecaque *et al* in **[5; 6]**. The sample is placed on the emitter. b: coupling between an optical fiber and a $\chi^{(2)}$ crystal as reported by **[7]**

.

A similar approach was adopted by Okada *et al* in **[8]** where they reported the development

of a scanning laser terahertz near-field reflection imaging system using a fiber-coupled femtosecond laser source for high spatial resolution measurements without any sub-wavelength probes. In their setup, terahertz radiations, that are generated at the laser focusing spot in a non-linear optical crystal, interact with a sample set on the crystal top-surface. Using this system, authors obtained terahertz images in reflection configuration of metallic structures with a resolution of $\lambda/34$ at 441 GHz. The imaging setup was further employed to acquire images of pathology sections of tens of microns.

Detector in Close Vicinity

Even though it seems instinctive, instead of positioning the terahertz energy source close to the sample, one can place the detector near the object. By doing so, the resolution can also be enhanced, provided the detector is small enough.

The collaborative work conducted by the group of Planken and Kim has successfully captured the vectorial electrodynamics ($\vec{E}$, $\vec{H}$ and Poynting vector $\vec{S}$) near the slits of a metal foil **[9]** (Figure IV.3 a). To achieve such a description of the electromagnetic field, the incoming terahertz beam is focused on the object, that stands on an electro-optic crystal. The terahertz field induces a refractive index change (referred as the Pockels effect in the foregoing) in the electro-optic crystal. The refractive index change is probed by a counter-propagating fs-laser that is focused on the front surface of the electro-optic material (Figure IV.3 1). The terahertz radiations that emerged from the metal foil slits were measured with a subwavelength resolution by about 50 µm.

On their part, Bitzer *et al* replaced the electro-optical crystal by a silicon layer onto which are patterned H-shaped electrodes (Figure IV.3 b). The H-shaped electrodes face the back-side of the sample to be inspected. The probe fs-laser beam is focused into the photoconductive gap of the detector. Using this strategy, authors obtained a slightly greater spatial resolution that in the previous study by about 20 µm at a frequency of 0.7 THz ($\sim \lambda/20$). Authors reported that the measurement is quite sensitive but however can not be performed directly "in-contact" with the sample as the distance between detector and sample is limited from the experiment to 35 µm. In **[10]**, the authors improved even further the achievable resolution, up to 5 µm and stated that they could enhance even more such a resolution by reducing the probe-sample distance and by lateral down scaling the size of the photoconductive gap employing nanostructuring techniques.

Quasi Near-field

Someone who reads those lines may have wonder, why don't we bring both the electromagnetic energy source and detector in close vicinity to the sample of interest ? Actually, such a method was investigated by Chakkittakandy and colleagues **[12]**. In their so called "quasi near-field" setup, authors employed two electro-optic crystals positioned close to each other to generate and to detect terahertz radiations (Figure IV.4). The proximity between the two electro-optic layers offered a particularly high signal-to-noise ratio. However, authors could not image any samples as the two crystals are clamped and adjusted with respect to the pump and probe fs-laser beam so that they are, ultimately, immobile.

Apertures Near-field Scanning Optical Microscopy (A-NSOM)

Another way to achieve super-resolution is to reduce the size of the detector. Obviously, at some point, that approach will rapidly find actual available technologies as obstacles. An alternative way consists of placing a sub-wavelength aperture across the propagating path. Therefore, only what comes in and comes out the aperture will consequently be detected. Therefore, the spatial resolution is scaled down to the size of the aperture that stands in the optical path. On considering a terahertz beam that has the shape of a Gaussian, one can consider the aperture as a high-pass filter. In fact, while low frequencies are theoretically located on the Gaussian periphery, high frequencies are located near the Gaussian center. The associated cut-off frequency can be determined

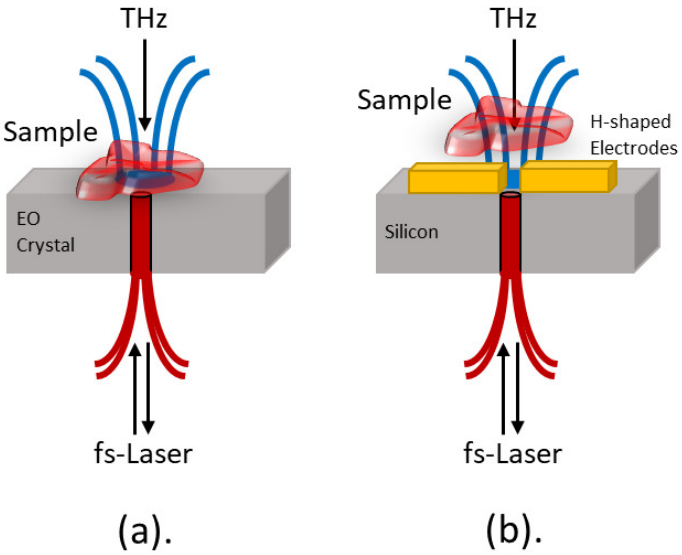


Figure IV.3: Near-field sensing with terahertz detector near the sample to be inspected. a: approach developed by **[9]**. The terahertz field induces a refractive index change in the electro-optic crystal onto which is mounted the sample. Polarization ellipticity changes in the counter-propagating probe fs-laser beam are recorded. b: alternative idea suggested by **[11]** where the electro-optic crystal is replaced by H-shaped electrodes patterned on a silicon layer.

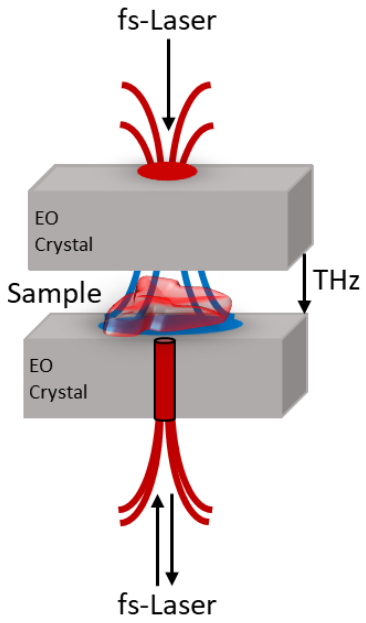


Figure IV.4: Near-field sensing with both terahertz electro-optic source and detector close to the sample to be inspected **[12]**.

experimentally from waveguide theory as it was shown by Gallot *et al* in **[13]**.

As terahertz is a recently investigated part of the electromagnetic spectrum, most of the implementations are coming from other regions that have already been explored for decades. In 1972, Ash and Nicholls published the first concept of a super-resolution aperture scanning microscope **[14]**. Originally, this study was conducted towards aperture implementation for microwave microscopy. Based on the same principle, Mitrofanov *et al* successfully reported the implementation of a 50 × 50 μm^2 aperture made of gold and patterned on a GaAs substrate **[15]**. Detection of the radiations that are transmitted through the aperture and near it was achieved with a photoconductive antenna that was placed 4µm below the aperture and right in its center (Figure IV.5 a). Later

on, the same research team improved a little bit further their conceptual design by implementing a silicon lens on the top of the aperture to better collect terahertz radiations **[16]**.

Different aperture geometries may lead to different effect. Ishihara and colleagues designed a bow-tie shaped aperture surrounded by concentric periodic structures in a metal film **[17]** (Figure IV.5 b). A bow-tie aperture is known to exhibit particularly interesting electromagnetic field enhancement at the sharp tips of metal **[18; 19]**. Moreover, the concentric grooves, commonly named bull's eye configuration, allows enhanced transmission beyond the diffraction limit thanks to the resonant excitation of surface waves **[20]**. Authors obtained a spatial frequency by about $\lambda/17$ at the wavelength of 207 µm.

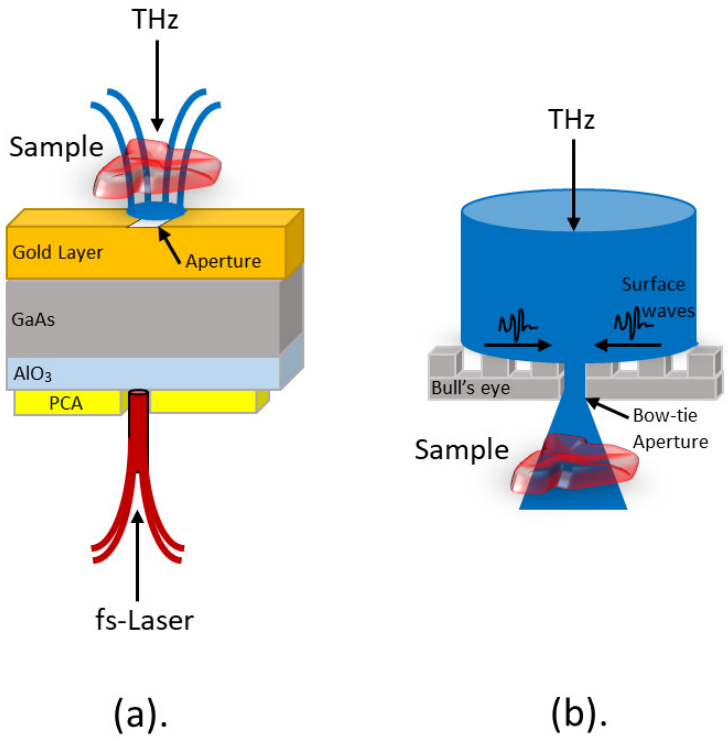


Figure IV.5: Near-field sensing with illuminated sub-wavelength aperture. Note that the sample may either be upstream or downstream the sub-wavelength aperture. a: the terahertz beam passes through the sample and subsequently through a sub-wavelength gold aperture. The electric component of the transmitted field is detected by a low temperature grown GaAs photoconducting antenna **[15]**. b: the incident terahertz beam is sent onto a bull's eye structure that activates surface waves. The resonant excitation of those waves allow to increase the amplitude of the electric field that is detected. Terahertz passes through a bow-tie aperture and interacts with the investigated sample **[17]**.

One would therefore ask oneself until which limit the size aperture can be scaled down ? While some responses can be found in **[21; 22]**, it is worth mentioning that the smaller the aperture is and the lower the electromagnetic field strength is.

In **[16; 22]**, researchers have shown that there is a linear decrease of the electromagnetic field amplitude with the aperture size in the region below the center of the aperture that does not go further than the aperture size. In other words, letting d being the aperture size and z the distance from which one observes the electromagnetic amplitude, the linear dependence remains valid for $z \in \mathscr{R}, z < d$.

On the other hand, for distances from the aperture center that are larger ($d < z$), the model suggested by Bouwkamp in **[22]** describes a cubic power dependence instead of a linear one. It is therefore worth to underline that the aperture size and the distance from which is observes the diffracted electromagnetic field have to be tuned in a concomitant way.

Hence, how far from the aperture center the terahertz detector should be positioned ? Mair *et al* have reported in **[23]** that the signal which is transmitted through a considered aperture decays

in an exponential manner with the distance. Although such a statement seems evident, it however only remains valid for an electromagnetic field that is observed in the close vicinity of the aperture, that is $z << \lambda$, wherein λ is the wavelength. In fact, for larger distances, the electromagnetic field strength decays as d^3/z^2.

It is overall interesting to note that in this case, the definition we gave in the beginning of this chapter about near-field sensing is not that accurate since the near-field depicts a region within the characteristics dimensions of an object rather than within a typical wavelength threshold. As written by Adam in **[24]**, "typically, if you want to measure the near-field from a 5 µm hole, your detector should be placed at least, 1 µm from the hole to detect the proper near-field, whatever your wavelength is in the case it is larger than the hole diameter".

Scattering Near-field Scanning Optical Microscopy (S-NSOM)

Metal tips are well known to scatter the light that is focused on their apex. The apex of a sharp tip can be seen as an infinitely small sphere which is an ideal scatterer. Hence, placing the apex close to a sample that is illuminated would allow one to scatter the near-field. However, the scattered near-field is expected to be relatively weak in comparison to the field that irradiates the inspected sample. As a consequence, to achieve a reasonable signal-to-noise ratio, the collection of the near-field information has to be performed in the apex direct proximity. To detect the near-field scattered by a sharp tip apex of subwavelength dimension, van der Valk *et al*, held a (100) oriented electro-optic GaP crystal in close vicinity **[25]**. A synchronised fs-laser is focused onto the GaP electro-optic crystal that detects the component of the probed field that is perpendicular to the (100) oriented surface. Authors stated that as the incoming terahertz polarization is parallel to the electro-optic material, the perpendicular component only exists close to the tip apex, "which functions as a lighting rod where field lines are bent and concentrated". To bid even further on the choice of the electro-optic crystal orientation, the apex is naturally going to scatter the incoming light in every direction which leads to the generation of all possible polarization. However, a part of this field is expected to be non-propagating, that is evanescent waves. The electro-optic crystal orientation allows to detect only the non-propagating components, and therefore only the near-field. In their study, van der Valk and colleagues reported that they achieved a subwavelength resolution by about $\lambda/110$ at the frequency of 150 GHz. Let's note that the resolution was determined from the beam spot size directly under the tip.

Instead of using judiciously oriented electro-optic crystals to collect the near-field information submerged within the strong far-field, Mittleman's group employed a mechanically modulated tip probe and lock-in methods **[26]**. The tip goes perpendicularly back and forth to the surface of the sensed sample with an amplitude by about 750 nm (Figure IV.6). The frequency of the measured signal is concomitantly locked to this specific frequency. Authors reported a resolution at the micron scale on imaging metallic domains in vanadium dioxide thin films.

A further study even implemented cantilever tungsten tip usually employed in atomic force microscopy **[27]**, as scattering tip, enabling imaging with a super-resolution of 40 nm, corresponding to $\lambda/3000$ with a continuous wave gas laser source generating at 2.54 THz **[4]**.

IV.2.3 Challenges for Biomedical Application: Sensor Integration

It seems that either A-NSOM or S-NSOM techniques may be profitable to the field of biomedicine to collect tissue response to an external terahertz stimuli, with a subwavelength resolution. *Yes, but no.* Let's remind that although very high spatial resolution can be reached with such strategies, they remain quite sophisticated implementations that is, lowly integrated setups. But especially because, the overall extremely weak collected near-field information is strongly shadowed by the far-field background clutter, requiring therefore a particularly long integration time to obtain a reasonable signal-to-noise ratio. Hence, such kind of imagers would be incapable to provide real-time visualization of tissue response. This incapacity towards offering in real-time the near-field components is obviously not an issue for laboratory uses. However, in the context of surgical

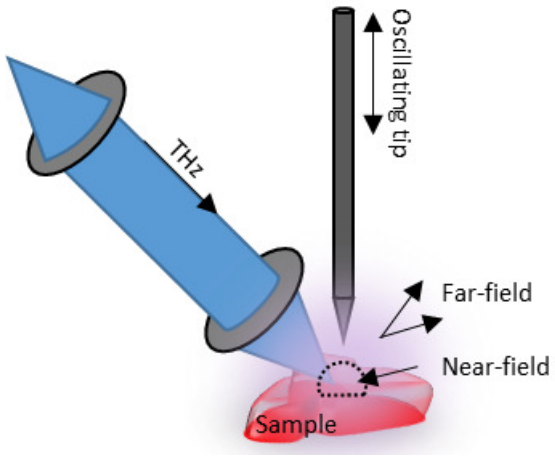


Figure IV.6: Scheme of a S-NSOM setup. A terahertz beam is focused onto the sample surface by means of lenses. The evanescent field is probed with an oscillating scatterer that vibrates at the frequency of the field to be recorded. The measurement of the near-field suffers from the inavoidable presence of a strong far-field background.

procedures, let's remind that it is not possible to let open a body more than thirty minutes. Consequently, these techniques are simply not suitable for intra-operative assessment of breast margin cleanliness purposes.

Several attempts were conducted towards enabling real-tissue visualization of the near-field components and upgrading near-field scanning optical microscopy integration in the terahertz band. All of this with the aim of addressing the challenges for biomedical applications.

To date, terahertz field detector were usually located remotely from the probe which, ultimately, drastically limit the sensitivity. Moreover, the influence of the far-field background remains problematic.

To overcome these issues, Kawano and Ishibashi reported in 2008 on the development of an integrated near-field terahertz detection device which consists of a cryogenically cooled semiconductor chip onto which are designed a subwavelength dimension aperture, a probe and a detector **[28]**. They reported a spatial resolution of 9 µm, that is $\lambda/24$ at 1.40 THz.

The group of Mitrofanov integrated 3µm aperture in a terahertz near-field probe **[29]**. Here it is worth to remind the reader that we previously reported on the fact that for distances higher than the aperture size, the transmitted evanescent field decreases in a cubic manner, which means that there is a field intensity dependence of $I \propto z^{-6}$, wherein I is the field intensity and z the distance from the aperture. However, for closer distances from the aperture, the transmitted near-field significantly deviates from the cubic law; Using this deviation, the authors successfully achieved to integrate these small apertures and to obtain a spatial resolution of 3µm corresponding to $\lambda/100$ at 1 THz. Nevertheless, the overall dynamic range remained at this stage very low.

We also stated that the integration time as an issue. Blanchard *et al* developed a real-time terahertz near-field imager, working at 700 GHz and providing a spatial resolution by about 14 µm **[30]**. To do so, authors combined a non-linear crystal located at the sample vicinity with a high-intensity source that entirely irradiated the sample under inspection whose area was about 370 × 740 μm^2.

IV.2.4 Resonator-Based Probing

Similarly to what is observed in far-field, the intrinsic dielectric properties of materials are expected to modify the near-field components from a reference state, usually free space.

Different approaches are popular towards the characterization of permittivity-induced near-field component changes, in the presence of a dielectric sample. These approaches globally fall into two different categories: the non-resonant methods and the resonant ones **[31]**. While the

non-resonant methods are used to generically characterize materials over a wide frequency window, the resonant strategy is rather employed to get an accurate quantification of the sample dielectric properties at unique wavelength.

Non-resonant methods aim to determine the dielectric behavior of a sample from both the material impedance and the velocity of radiations that travel through it. The sensing electromagnetic energy is directed towards the material to be characterized. Either the reflected or the transmitted or the reflected and transmitted energies are used to derive sample characteristics. All types of transmission lines may theoretically be used to carry and to collect the sensing radiations. For instance, one can mention coaxial lines, dielectric waveguides or even free space.

On the other side, resonant methods that have greater sensitivities and accuracies are based on the perturbation theory and the resonant method. In essence, it consists in recording the resonant frequency and the quality factor of the resonator that are, under the presence of a dielectric sample, expected to be changed. Therefore from these shifts towards the natural wavelength resonance and the quality factor, the intrinsic properties of a sample can be determined.

One has therefore to opt for one of these approach. The Nearsense project that was raised in collaboration between the circuit design research group of Prof. Dr. rer. nat. Ullrich Pfeiffer at Wuppertal, the pathology department of Bergonié Institute led by Dr. Gaëtan MacGrogan and the terahertz and material research group of Dr. Patrick Mounaix at Bordeaux, aimed to design an integrated terahertz near-field sensor to qualify the margin cleanliness of freshly excised breast cancer section for intra-operative purposes. In view of tissue response in far-field investigations, reported throughout this manuscript, there is a clear evidence, that no specific features are exhibited among the different tissue types. Rather than observations of tissue-dependent features, distinct permittivity profiles were reported. Adipose tissues have been consistently demarcated from healthy fibrous and cancer tissues within the frequency band spanning from 300 GHz to 800 GHz. Healthy fibrous and cancer tissues were however, on their part, not consistently distinguished. Although we stated at that time that the cancer cell density and additional factors would likely play a role towards strict demarcation, differences between those tissue types were reported, in average, on the refractive index between 300 GHz and 600 GHz.

In view of all these considerations, the Nearsense project ultimately focused on the development of an integrated terahertz near-field sensor based on the resonant method and working at the unique frequency of 560 GHz for breast tumor identification. A conceptual design of such a super-resolution solid-state imaging device is given Figure IV.7.

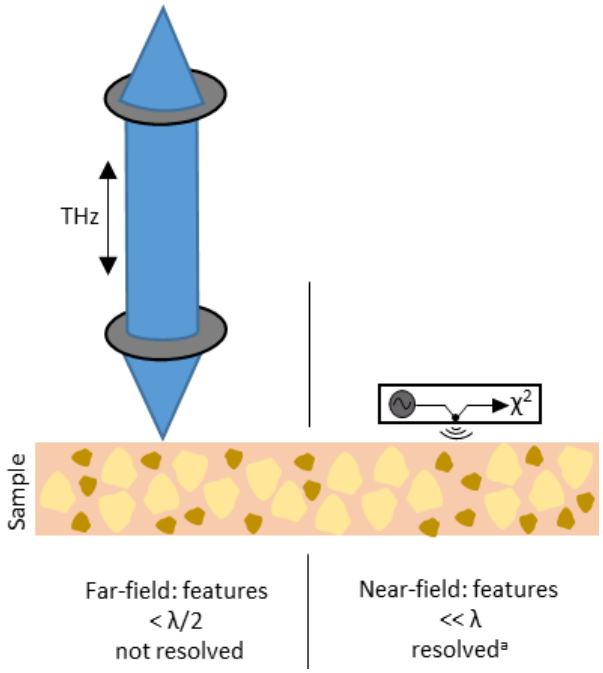


Figure IV.7: Conceptual design of a near-field device that senses the evanescent field arising from the interaction of terahertz radiations with a sample (right). The device is aimed to consist of an illumination source, a near-field surface sensor and a high-responsivity terahertz power detector, with all mounted on an unique chip. Thanks to subwavelength resolution, features whose size is smaller than the probing wavelength can be resolved. [a]: the smallest size of features than can be resolved depends on the near-field device performances. In comparison, far-field probing (left) does not allow to resolve sample features smaller than $\lambda/2$.

IV.3 Towards a Silicon-Based Terahertz Near-field Imaging Matrix

As discussed earlier, the diffraction limit of resolution has been successfully beaten and sub-wavelength terahertz imager were reported throughout the literature. However, although those imaging approaches can offer super-resolved images, they remain at that time moderately used in the field on biomedical imaging due to their overall weak sensitivity, high sophistication and low integration level.

In the following, we are about to address the operation principle and the overall design of the resonant terahertz evanescent-field imager in silicon technology that has been engineered and designed specifically for the purposes of the Nearsense project **[32; 33]**.

IV.3.1 Split-Ring Resonators (SRR)

The sensing method adopted is based on split-ring resonators (SRRs). A split-ring resonator is a pair of concentric loops etched in a dielectric substrate with splits in them at opposite ends. The shape of these loops can have various form: circular, square or U-shaped. Some of the most common geometries are illustrated hereafter, in Figure IV.8.

When a time-varying magnetic field penetrates through the metallic ring, it induces an electromotive force. The electromagnetic force induction induces a rotating current within the concentric loops. Such a behavior is pointed out by one of the four Maxwell's equations that rule classical electromagnetism theory. Indeed, the Maxwell-Faraday equation, states that a time-changing magnetic field accompanies a spatially-changing non-conservative electric field:

$$\oint_{\delta S} \mathrm{E}.dl = -\int_{S} \frac{\delta \mathrm{B}}{\delta t}.d\mathrm{A}, \tag{IV.3}$$

where S is the surface bounded by the close contour δS, E is the induced electric field, $\frac{\delta B}{\delta t}$ is

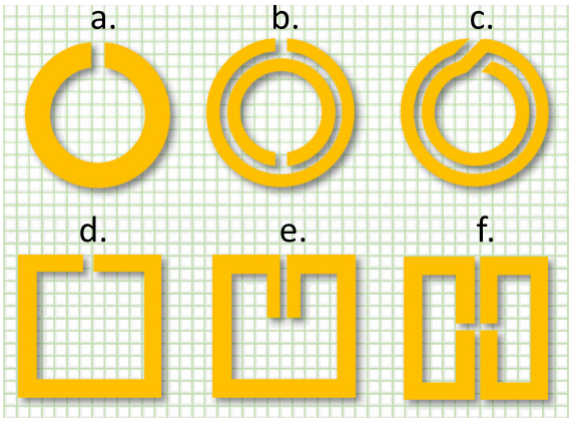


Figure IV.8: Representation of different split-ring resonator geometries. a: circular SRR; b: double circle SRR; c: complementary spiral SRR; d: square SRR; e: U-shaped SRR; f: dual-square SRR.

the time-varying magnetic field, δl is an infinitesimal vector element of the contour δS and δA is an infinitesimal vector element of surface S. Eventually, the electromotive force acts hence as a voltage. Therefore, the split-ring resonator can be seen as a simple electric circuit that has an inductance-capacitance (LC) behavior. The resonance of such a configuration occurs when an external energy source drives the LC-system at the angular frequency ω_0:

$$\omega_0 = \frac{1}{\sqrt{LC}}, \tag{IV.4}$$

where L is the inductance and C the capacitance. Between those two parameters, solely L, the inductance, remains fixed as it is defined by the intrinsic metallic ring properties. On its part, the capacitance C is expected to change depending on the local dielectric permittivity. Indeed, the presence of an electric-sensitive object increases the dielectric permittivity at gap vicinity (Figure IV.9). Consequently, the capacitance of the gap increases and ultimately leads to a slighter resonance frequency. In a nutshell, the gap capacitance acts as a permittivity sensor whose change induces the characteristic frequency of the LC-circuit to be shifted down **[34]**.

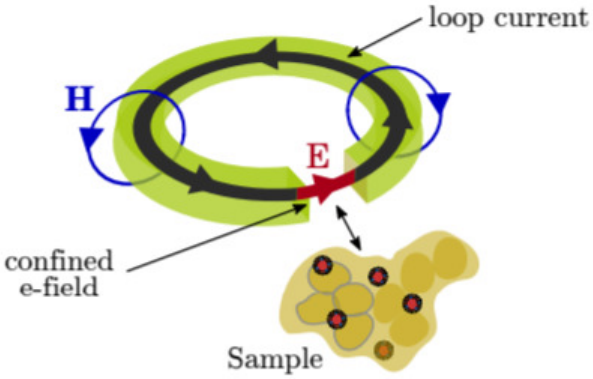


Figure IV.9: Schematic representation of a split-ring resonator probe. The resonator behaves like a LC-component whose characteristic frequency is about to be shifted in the presence of a sample, at the neighborhood of the resonator gap.

Overall, the nominal resonance frequency of a split-ring resonator is dictated by the intrinsic properties of the material used to design the resonator, the gap structure and the metallic loop geometry.

IV.3.2 Shift Tracking

Rather than measuring directly the resonance frequency shift from ω_0 to ω_0' due to permittivity disturbances at the split-ring resonator gap, the conceptual idea of the Nearsense collaboration, is to interpret an induced frequency shift as a change in transmission power through a band-rejection filter. Chiefly, such kind of filter only attenuates frequencies that are strictly comprised within a particular window. To avoid any confusion, it is worth to underline that it is the LC-circuit itself that acts as a rejection-band filter. Evidently, close to the resonance the transmission is drastically attenuated. On the contrary, far from the resonance, the LC-band-stop filter passes most frequencies unaltered.

A free-running on-chip oscillator irradiates the split-ring resonator with an angular frequency ω_{osc} slightly above ω_0. Upon loading the split-ring resonator with an object whose dielectric permittivity is greater than that of free space, the resonance frequency decreases as the capacitance increases and therefore the transmitted power through the band-rejection filter at ω_{osc} increases.

The changes towards the transmitted power through the band-rejection filter are monitored by a SiGe heterojunction bipolar transistor terahertz power detector whose response voltage V_0 is proportional to the incident power P_i. Detector response difference between free space (V_0) and a loaded sample (V_0') can be expressed as **[35]**:

$$\Delta V_{0,0'} = P_{osc} R_V \Delta T_{0,0'}, \tag{IV.5}$$

where P_{osc} is the power emitted by the free-running on-chip oscillator, R_V is the detector voltage responsivity and $\Delta T_{0,0'}$ the transmitted power difference between free space and loaded sample measurements. In summary, instead of assessing frequency shift in the frequency domain, the shifts are translated into the amplitude domain. While the band-rejection filter is partially opaque to $\omega_o sc$ in absence of sample, the bandstop notch becomes progressively transparent to ω_{osc} with smaller resonant frequencies ω_0 induced by higher permittivity disruptions.

It is worth mentioning that therefore, the sensor sensitivity is not solely dictated by permittivity disturbances but is supplementary governed by the band-rejection filter characteristics that are, its depth and its slope. Hence, in order to maximize the sensor sensitivity, a high modulation depth is required, that is, the transmitted power amplitude difference between the resonance and a non-rejected frequency. The overall conceptual working principle of the permittivity-induced shift resonance tracking is depicted in Figure IV.10.

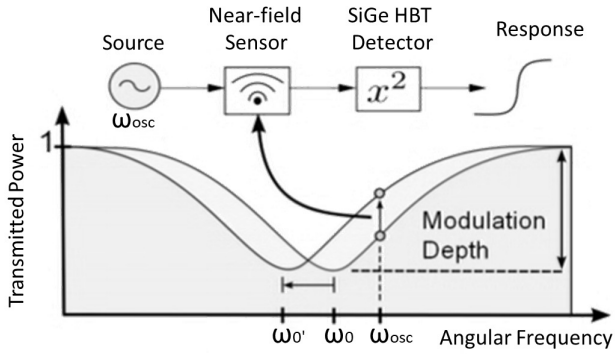


Figure IV.10: Block-diagram of the measurement operation. The resonance shift is interpreted as amplitude transmission changes from opaque to lucent at the terahertz power detector output.

IV.3.3 Cross-Bridged Dual-Square SRR

The working principle of SRRs and frequency shift tracking was previously described. Two of these SRRs are actually used to build up a complete sensor. The two SRRs are patterned symmetrically on each side of a transmission line, to which the SRRs are magnetically coupled. The respective SRRs gap are standing on opposite directions, such that they do not directly face the transmission

line. Each of the SRR is driven by a current i such that $i_1 = -i_2$. Two parallel strips connect the SRRs thereby forming a shared split-gap without modifying the ring respective current distributions. The strips are driven with antiparallel currents, due to the induction, to support the generation of an electric-dipole type sensing field. On the top of the sensor is deposited a thin layer of 2µm SiO_2 that acts as a protective layer. In fact, let's remind that ultimately, the sensor is aimed to probe soft tissues that may release organic fluids that would alter the overall electronics. The global structure of the described SRR is illuminated with a free-running on-chip oscillator at 0.56 THz. Figure IV.11 illustrates the geometry of a complete sensor.

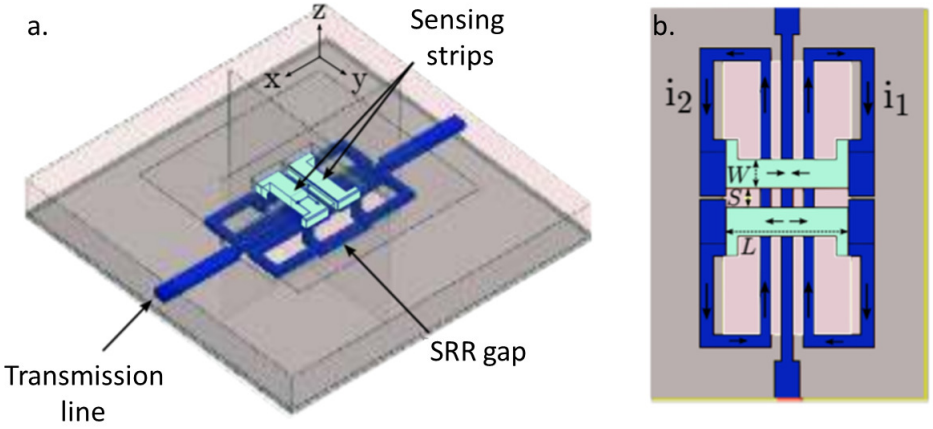


Figure IV.11: Sensor scheme. a: 3/4 view of the complete sensor made of a transmission line, and cross-bridged dual-square SRR; b: top view of the resonator with L = 19 µm, W = 4.5 µm, and S = 3 µm.

IV.3.4 Sensor Response Simulation

Various parameters are expected to influence the sensor response: first, as investigations are conducted towards sensing the near-field, the distance from the SiO_2 protective layer, that is deposited onto the sensor, is expected to be of paramount importance; Second, the real part of the dielectric permittivity of a sensed sample will modify the capacitance of the SRRs gap and consequently their respective resonances; Last, the lossy part of the dielectric permittivity may alter the quality factor of the sensor and therefore decreases the transmitted power through the transmission line. Hence, the sensor response is subject to the concomitant influences of these three external parameters. To better define the impact of each of these parameters that are the distance from the surface, the real part of the permittivity and its lossy part, the sensor theoretical responses to different configurations have been assessed by means of the CST Studio Suite ®, based on the finite integration technique (FIT) of electromagnetic fields. In the following, the simulations results are successively provided and discussed.

Sample-free Sensor

First, the sensor response to free space, that is without being loaded with a sample, was simulated. 182,000 meshcells were used in the simulation to derive the power transmitted through the transmission line. The power was injected from a port #1 located at one extremity of the transmission and was recorded at the opposite extremity at port #2. The profile of the S_{21} parameter, which is the ratio between the transmitted and the supplied power through the transmission line, depicts the presence of a nominal resonance at 540.0 GHz (Figure IV.12). It is interesting to underline that this specific resonance frequency is *de facto* the highest resonance that one can expect. In fact, as stated in the foregoing, upon loading the sensor with an object whose permittivity is higher than that of free space, it induces an increase in the SRR gap capacitance and ultimately shift the resonance to lower frequencies due to the linear dependence between the capacitance and the characteristic frequency of the LC-system.

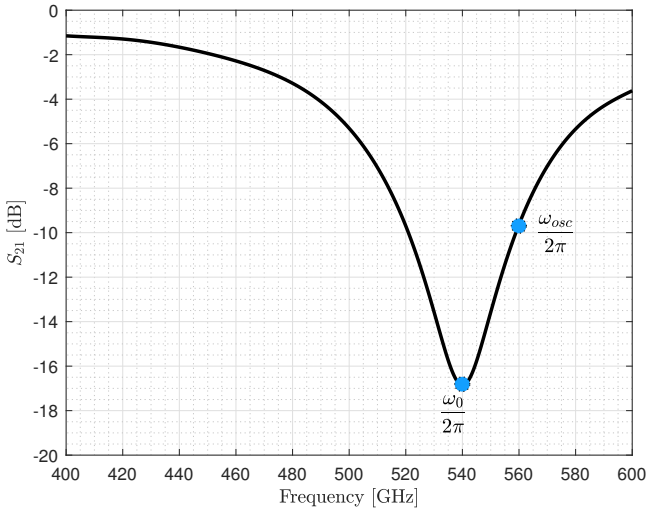


Figure IV.12: S_{21} parameter profile between 400 GHz and 600 GHz highlighting the presence of a nominal resonance frequency ($\frac{\omega_0}{2\pi}$) at 540 GHz. The sensor illumination ($\frac{\omega_{osc}}{2\pi}$) is proceeded at 560 GHz.

Altitude-dependence Response

To study the behavior of the sensor response as a function of the distance from the sensor surface, a perfect electric conductor (PEC) was chosen whose dielectric characteristic are $\epsilon' = \infty$ and $\epsilon'' = 0$. It is interesting to underline that therefore at null distance from the surface, such a material gives the highest possible resonance frequency shift and consequently the resonance frequency lowest value. The distance band investigated spans from 0 µm to 2.0 µm.

On being in contact with the sensor surface, the resonance frequency is shifted to a value by about 489.6 GHz (Figure IV.13 left). On comparing this value to that of free space, one can state that the resonance frequency interval among which the sensor is operating spans theoretically[1] over 50 GHz. The S_{21} parameter value decreases with increasing distance. It is evident from Figure IV.13 (left) that the altitude dependence is not linear. The dependence of the S_{21} parameter as a function of the distance is reported in Figure IV.13 (right) at 560 GHz (at $\frac{\omega_{osc}}{2\pi}$). On fitting the S_{21} dependency profile, a bi-exponential form was found to match the tendency (Figure IV.13 right). The response profile as a function of the distance $S_{21}(z)$ is of generic form:

$$S_{21}(z) = \alpha_1 e^{\alpha_2 z} + \beta_1 e^{\beta_2 z}, \tag{IV.6}$$

where α_1, α_2, β_1 and β_2 are fitting parameters and z stands for the distance from the surface. Let's underline that fitting parameters depend on both the frequency and sample properties, **but should over be regarded as necessary parameters to obtain accurate fit rather than carrying physical meaning**.

The exponential behavior was overall expected since non-propagating-field intensity decays exponentially with distance z from the source. However, we have shown that the sensor response to distance z was better modeled by a bi-exponential function rather than a single exponential.

It is worth to underline, that from the previous investigations one can define the interval within which the S_{21} parameter may vary at 560 GHz. Taking into account that S_{21} at 560 GHz for a PEC at $z = 0$ µm is by about -2.18 dB and by about -9.64 dB for free space, the transmitted power through the transmission line that is defined as $P_{0'} = 10^{\frac{(S_{21})}{10}} \times P_{in}$, increases at maximum by about 5.5 times from that of free space[2].

[1] Within the meaning of computation.

[2] $P_{0'} = 10^{\frac{(-2.18+9.64)}{10}} \times P_0 \approx 5.5 \times P_0$.

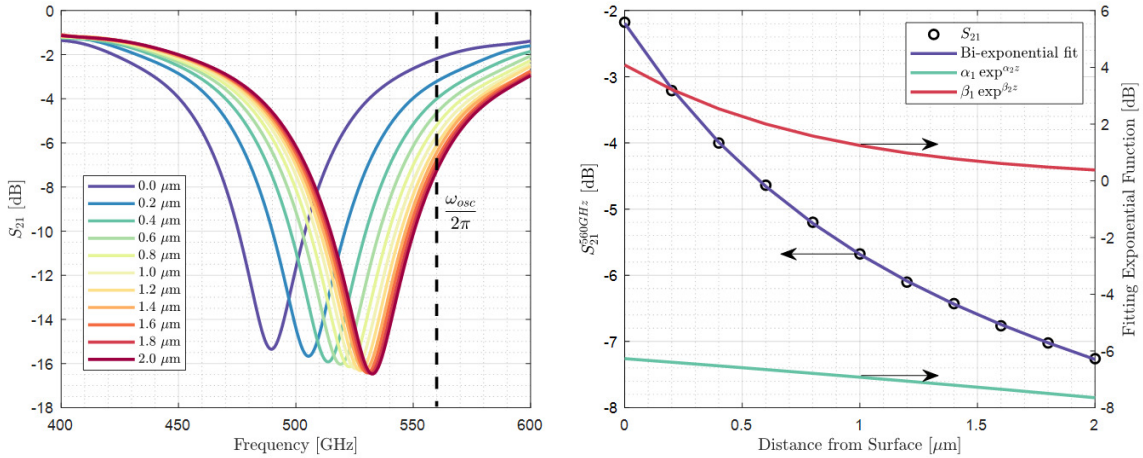


Figure IV.13: Profile of S_{21} as a function of the PEC distance from the surface. Left: Highlight of the induced frequency shift due to the PEC relative altitude. Right: Bi-exponential behavior of S_{21} as a function of the PEC relative distance at 560 GHz, with α_1 = -6.273, α_2 = 0.099, β_1 = 4.082 and β_2 = -1.189.

Response to ϵ'

Similarly to the investigations towards varying distance from the sensor surface, the sensor response to various real-permittivity part ϵ' was studied. The simulation was conducted for $\epsilon' \in [1;10]$ with a step size equal to unity. On its part, the lossy part of the relative permittivity ϵ'' was set to 0. The simulated behavior of the sensor as a function of ϵ' is depicted in Figure IV.14. As observed in Figure IV.14 (left) the frequency shift is directed towards lower frequencies with increasing ϵ'. Such a behavior is not a surprise as we endlessly pointed out the linear dependence between the resonance frequency ω_0 and the permittivity-induced shifted LC-circuit capacitance. It is interesting to note that as ϵ' increases, S_{21} increases on the right-hand side of the resonance frequency but decreases on its left-hand side. Overall, the frequency shifts from the nominal resonance ($\epsilon' = 1$) of 540 GHz to 511.8 GHz ($\epsilon' = 10$), *i.e.*, ~ 30 GHz.

In a like manner than that of the sensor z-dependence response, the real permittivity dependence of the sensor is better modeled by the sum of two exponential function (Figure IV.14 right). We have seen in previous chapters that biological tissues do not exhibit real permittivities that exceed $\epsilon' = 10$ between 200 GHz and 1 THz. Therefore, considering the maximum value of S_{21} given for $\epsilon' = 10$ that is by about -3.98 dB and its minimum value for $\epsilon' = 1$ which is of -9.64 dB, the change in transmitted power through the transmission line for biological tissues that are in contact with the sensor should at maximum be 3.70 times greater than that of free space[3].

Response to $\tan\delta$

Simulations towards the influence of the lossy part of the relative permittivity ϵ'' have been additionally conducted. To do so, we pick the definition of the loss angle defined as $\tan\delta = \frac{\epsilon''}{\epsilon'}$ in Chapter I. The loss angle tangent was investigated for values ranging from 0.0 to 0.5 with 0.05 being the increment size. To ensure that the sensor response is solely due to the lossy part ϵ'', ϵ' was fixed to unity. Sensor responses in such conditions are depicted in Figure IV.15.

From Figure IV.15 it is quite obvious that the increase in ϵ'' does not play a significant role towards shifting the resonance frequency. In fact, the increase in transmitted power is by about 1.02 times for $\tan\delta = 0.5$ compared to that of free space, *i.e.* $\tan\delta = 0$[4]. Nevertheless, it is worth mentioning that frequency shifts are not null and therefore lead to an ambiguity towards the sensor response. In fact, as far as the sensor response depends on both the real and imaginary permittivi-

[3] $P_{0'} = 10^{\frac{(-3.98+9.64)}{10}} \times P_0 \approx 3.7 \times P_0$.
[4] $P_{0'} = 10^{\frac{(-9.54+9.64)}{10}} \times P_0 \approx 1.02 \times P_0$.

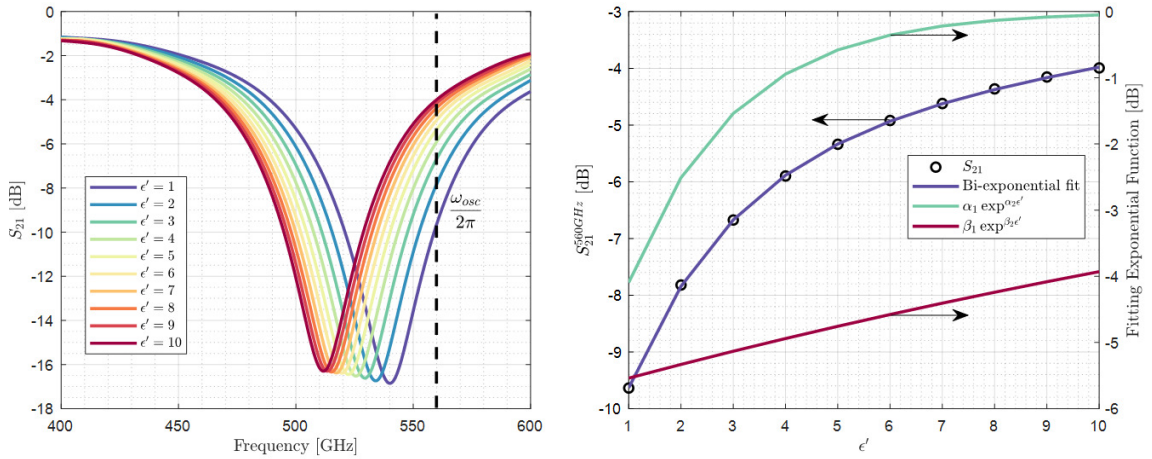


Figure IV.14: Profile of S_{21} as a function of different ϵ'. Left: Highlight of the induced frequency shift due to the progressive increase of ϵ' . Right: Bi-exponential behavior of S_{21} as a function of the ϵ' relative permittivity at 560 GHz, with α_1 = -6.675, α_2 = -0.488, β_1 = -5.755 and β_2 = -0.038.

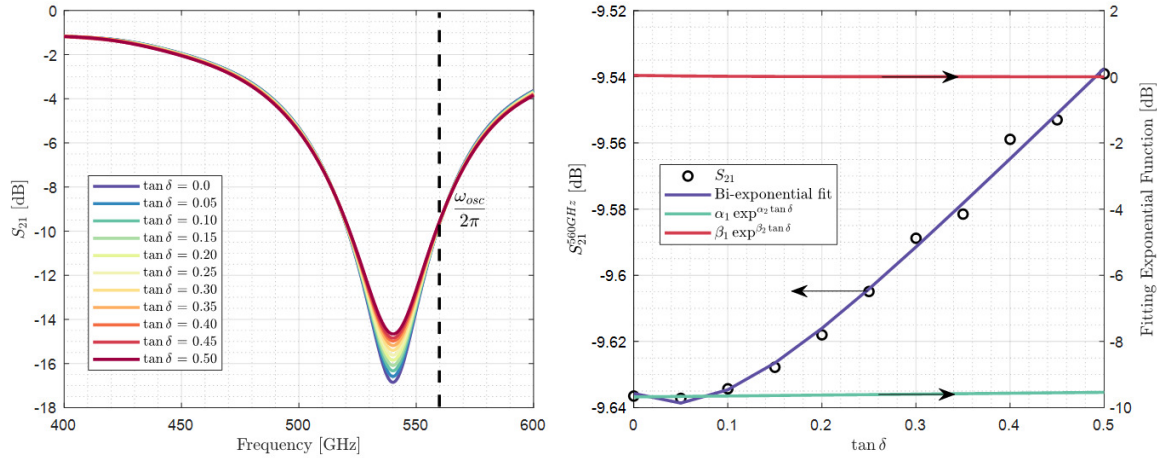


Figure IV.15: Profile of S_{21} as a function of different ϵ''. Left: Highlight of LC-circuit response due to the progressive increase of ϵ'' . Right: Bi-exponential behavior of S_{21} as a function of the $\tan\delta$ relative loss angle at 560 GHz, with α_1 = -9.677, α_2 = -0.029, β_1 = 0.041 and β_2 = -10.55.

ties, the two of those are concomitantly working towards shifting the resonance, without knowing precisely the influence of each. Furthermore, this effect is amplified for increasing $\tan\delta$ as the dependence is bi-exponential rather than linear.

It is however interesting to note that the slope changes due to varying $\tan\delta$ between the points at 540 GHz and 560 GHz is not negligible (Figure IV.16). The slope being defined as:

$$Slope = \frac{S_{21}^{560\text{GHz}} - S_{21}^{540\text{GHz}}}{560 - 540}. \tag{IV.7}$$

Therefore the sensitivity of the sensor that is directly linked to the slope of the bandstop notch, is also affected by lossy materials rather than solely small changes towards resonance frequency shifts. Consequently, the higher $\tan\delta$ is, the higher the frequency shift is but simultaneously, the lower the sensitivity is to these frequency shifts.

Variation in ϵ' and $\tan\delta$

As in the foregoing ϵ' was fixed to unity, a further study was conducted towards higher value for ϵ' with varying $\tan\delta$. The results that are to be reported (Figure IV.17) are for $\epsilon' \in [1;6]$ and $\tan\delta \in$

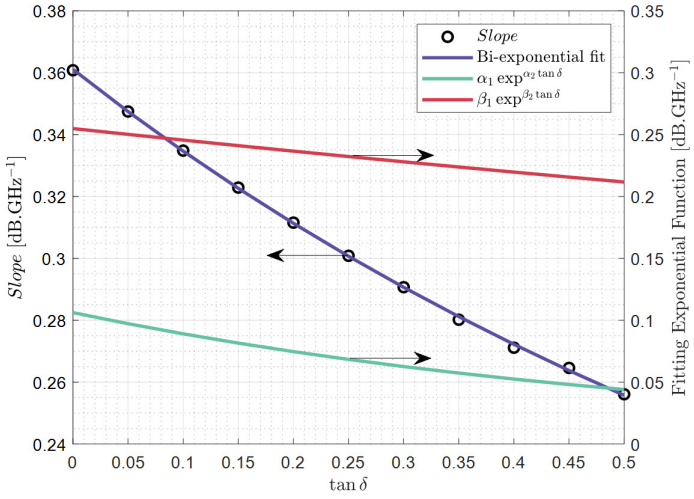


Figure IV.16: Slope as defined in equation IV.7 as a function of $tan\delta$ with α_1 = 0.106, α_2 = -1.761, β_1 = 0.255 and β_2 = -0.371.

[0;0.5] with respective increment size of $\Delta\epsilon' = 1$ and $\Delta \tan\delta = 0.05$.

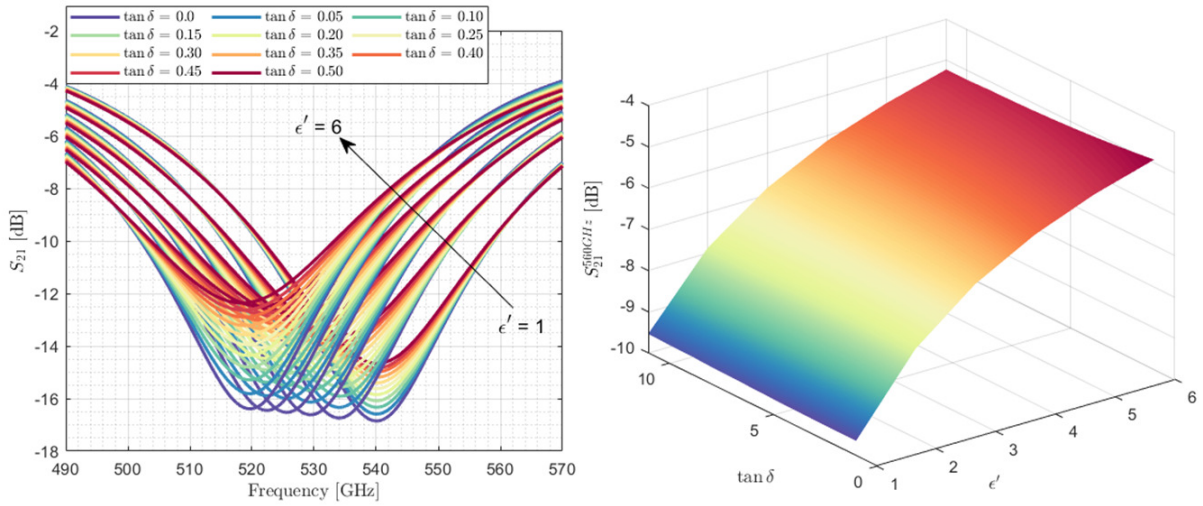


Figure IV.17: Profile of S_{21} as a function of ϵ' and ϵ''. Left: Highlight of LC-circuit response due to changes in ϵ' and ϵ'' . Right: Map of the S_{21} value at 560 GHz for $\epsilon' \in [1;6]$ and $\tan\delta \in [0;0.5]$.

Overall, on increasing the real part of the permittivity, an important frequency shift is observed. On its part, the imaginary part of the permittivity participates to a lesser degree to the resonance shift. Moreover, the resonance shift induced by the imaginary part ϵ'' of the dielectric constant is more pronounced for high $\tan\delta$. Additionally, we reported that $\tan\delta$ has an impact on the slope of the rejection-band filter that is directly linked to the sensor sensitivity. In an interesting manner, the decrease of the slope is more pronounced when the real part of the permittivity is concomitantly reaching high values. As it was anticipated from previous investigations on separated parameters (*Response to* ϵ' and *Response to* $\tan\delta$) the global behavior tends to be of bi-exponential surface rather than a plane. Obviously, due to the high impact of the real part of the permittivity in comparison to that of the imaginary part, the bi-exponential behavior towards $\tan\delta$ is lowly visible.

Liquids

Although previous studies highlighted valuable knowledge on the theoretical response of the sensor to different combination of ϵ' and ϵ'', these permittivity sets do not represent *a priori*, real

materials. To pursue investigations, the previous couples of (ϵ', ϵ'') were replaced by the dispersion profiles of real liquids. To this aim, the dispersion profiles of water, methanol, ethanol and isopropanol were simulated based on their Debye parameters. While the dispersion of water was assessed with two Debye poles, the dispersion profiles of alcohols were simulated by three Debye poles **[36]**. Table IV.1 sorts the respective Debye parameters for each of those simulated liquids.

Table IV.1: Debye parameter of water, methanol, ethanol and propanol as used to simulate their respective dispersion profiles.

Liquid	ϵ_∞	$\Delta\epsilon_1$	$\Delta\epsilon_2$	$\Delta\epsilon_3$	$\tau_{D}[ps]$	$\tau_2[ps]$	$\tau_3[ps]$
H_2O	2.66	73.19	2.24	-	8.34	0.10	-
MeOH **[36]**	2.1	27.28	1.98	1.27	48	1.25	0.16
EtOH **[36]**	1.93	20.2	1.43	0.79	161	3.3	0.22
PrOH **[36]**	1.93	17.01	1.06	0.79	316	2.9	0.2

The Debye parameters for water were directly taken from this work. The parameters of alcohols were taken from **[36]**. It is important to note that from terahertz time-domain spectroscopy, it is not possible to accurately fit the properties of these alcohols. Indeed, on observing the values of the characteristic relaxation times τ reported by **[36]**, it is evident that some of these relaxations do not belong to the terahertz domain but rather to the microwave one to which, one does not have access from terahertz time-domain spectroscopy. For this reasons, we kindly borrowed the extracted values from the microwave domain research field. Figure IV.18 depicts the simulated behavior of the LC-circuit with respect to the considered liquids.

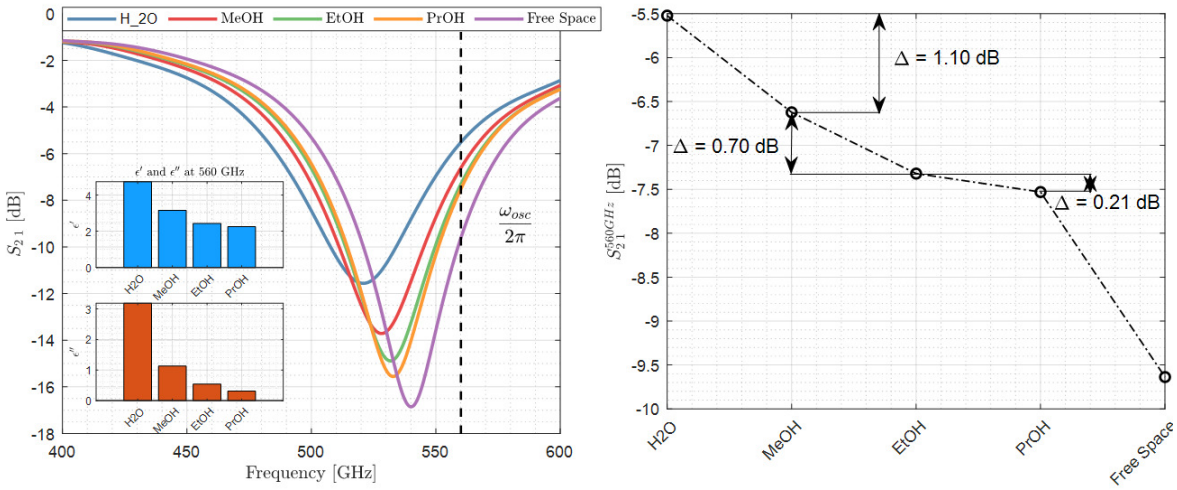


Figure IV.18: Profile of S_{21} for water methanol, ethanol and propanol. Left: Highlights of the LC-circuit response to polar liquids. The inset depicts the values of ϵ' and ϵ'' for each sample at 560 GHz as derived from the parameters in Table IV.1. Right: Behavior of S_{21} at 560 GHz as a function of the considered liquids. Simulated differences in S_{21} between the different samples are highlighted.

The response of the sensor is in accordance with the previous simulations; the higher ϵ' and ϵ'' are, the higher is the resonance shift to low frequencies. Hence, the transmitted power through the transmission line increases from propanol to water. Overall water is well separated from alcohols (Figure IV.18 left). Among the set of alcohols, methanol is the one that gives arise to the highest frequency shift. On their part, ethanol and propanol profiles, although they both show significant resonance shift to that of free space, are quite similar. The observations are supported by the differences in S_{21} parameter between these two polar liquids (Figure IV.18 right).

Naturally, on observing the theoretical response of the sensor to ethanol and propanol, one may legitimately ask oneself: is the sensor accurately able to distinguish between these two ? In other terms, is the sensor sensitivity great enough to differentiate between ethanol and propanol ?

In the foregoing, we already slightly pointed out that the sensitivity of the sensor is closely related to the slope of the band-rejection filter. Overall, the sensor sensitivity is determined by the minimum detectable change in transmitted power. On considering low relative permittivity differences like in the case of ethanol and propanol ($\hat{\epsilon}_{EtOH}^{560GHz} = 2.43 - 0.54j$; $\hat{\epsilon}_{PrOH}^{560GHz} = 2.25 - 0.31j$; see Figure IV.18 inset) the relationship between the power transmitted through the host line P_t and the permittivity $\hat{\epsilon}$ can be assumed to be roughly linear instead of bi-exponential as previously underlined. The minimum detectable change in permittivity $\Delta\hat{\epsilon}^{min}$ can be expressed as **[35]**

$$\Delta\hat{\epsilon}^{min} = \frac{V_{n,int}}{P_{osc} R_V \frac{\delta P_t}{\delta\hat{\epsilon}}}, \tag{IV.8}$$

where $V_{n,int}$ is the total noise integrated over the sensor and P_{osc} is the power of the sensor illumination. From simulations, it is obviously not possible to state on the experimental ability to differentiate between ethanol and propanol. However, in latter experimental studies, we will see that the sensor is effectively able to distinguish between these two permittivities. In Table IV.2 are summarized the principal simulated features for each liquid.

Table IV.2: Simulated sensor response to water, methanol, ethanol and propanol. Values are given with respect to that of free space. $P_{0'}/P_0$ denotes the gain in transmitted power for a particular sample with respect to that of free space.

Liquid	$\Delta\frac{\omega_0}{2\pi}$ [GHz]	ΔS_{21} [dB]	$P_{0'}/P_0$
H_2O	-19.20	4.11	2.58
MeOH	-11.80	3.01	2.00
EtOH	-8.20	2.32	1.71
PrOH	-7.20	2.13	1.63

Breast Tissues

Now that we raised valuable insights towards the theoretical sensor response to different combinations of permittivity and to real materials such as liquids, similar investigations are of interest concerning breast biological tissues.

It is important to remind the reader that from the previous investigations and equation (IV.7), we have highlighted that the overall sensitivity of the sensor to the lossless part of the permittivity is reducing with increasing lossy part. As biological tissues are mainly made of water, that is concomitantly the strongest absorbing material within the breast, such a trend may subject the sensor to ambiguity towards distinguishing the different tissue types.

To simulate the tissue permittivity profiles, the dielectric model that was reported in **[37]** was used. One would ask oneself, why not using the model derived during this work ? The model that was designed is currently (at the time this manuscript is written) in its early stages and its overall development requires additional investigations and larger set of samples. Therefore, as the model in **[37]** has shown robust and adequate modeling of breast tissues over a significant number of samples, it is reasonable to simulate the tissue dielectric profiles from such a model rather than the one designed in *Chapter II*.

The employed model is defined as:

$$\hat{\epsilon}_r(\omega) = \epsilon_\infty + \frac{\omega\tau_1\Delta\epsilon_1 + \Delta\epsilon_2}{1 + (i\omega\tau_1)^\alpha} + \frac{\Delta\epsilon_3}{1 + i\omega\tau_2} + \frac{\sigma}{i\omega}, \tag{IV.9}$$

where σ corresponds to the impact of the conductivity on the dielectric loss within tissues. The authors emphasized that $\Delta\epsilon_1$ and $\Delta\epsilon_2$ are more empirical parameters rather than being of physical nature and should therefore be considered as required parameters to obtain adequate fits. Extracted parameters from their study are reported in Table IV.3.

Table IV.3: Parameters of the model proposed by Truong *et al.* aR^2 is the adjusted R-square value denoting the square of the correlation between the response values and the simulated ones. When equal to unity, the adjusted R-square value denotes a perfect correlation while being equal to 0 it denotes a null correlation.

	ϵ_∞	$\Delta\epsilon_1$	$\Delta\epsilon_2$	$\Delta\epsilon_3$	σ	$\tau_1(ps)$	$\tau_2(ps)$	α	αR^2
Adipose	2.50	3.54	-1.89	0.47	0.39	9.67	0.13	1.49	0.996
Normal	3.34	1.33	-3.05	0.71	5.27	1.91	0.11	1.20	0.998
BCC	2.86	28.64	-3.45	1.59	8.33	3.58	0.10	2.91	0.997

The corresponding simulated sensor response are given in Figure IV.19. It is evident that adipose tissue is theoretically well identified by the sensor. The induced capacitance upon loading the sensor with adipose tissue gives arise to a frequency shift, with respect to the nominal frequency of 540 GHz, that is clearly smaller than those induced by tumor and healthy fibrous. Nevertheless, the first thing that comes to mind is the proximity between the profiles induces by abnormal and normal tissues. Such a proximity may be an issue as the sensor is less sensitive to high permittivities in presence of high losses. Indeed, as the profile of BCC and normal tissue are relatively close, since the real and the imaginary part of the dielectric constant of cancer are higher than those of healthy tissue, cancer may simply be indistinguishable from normal tissue. Once again comes the question of the sensor sensitivity, that will be addressed in deeper details soon.

In Table IV.4 are listed the principal features for breast tissues compared to that of free space.

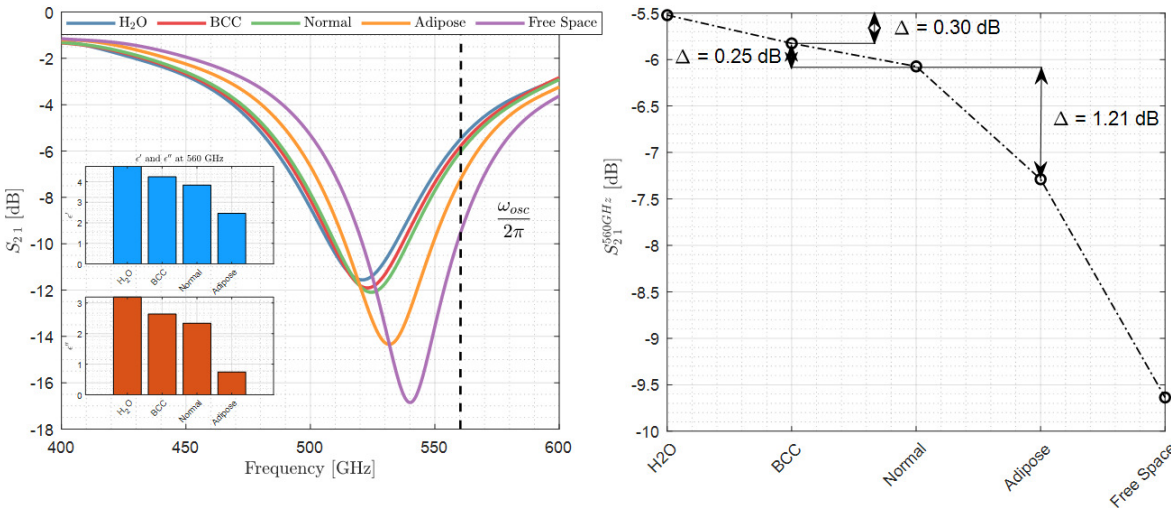


Figure IV.19: Profile of S_{21} for water, BCC, normal and adipose tissue. Left: Highlights of the LC-circuit response to biological tissues. The inset depicts the values of ϵ' and ϵ'' for each tissue types and water at 560 GHz as derived from the parameters in Table IV.3. Right: Behavior of S_{21} at 560 GHz as a function of the considered tissues. Simulated differences in S_{21} between the different samples are highlighted.

Table IV.4: Simulated sensor response to basal cell carcinoma (BCC), normal fibrous and healthy adipose. Values are given with respect to those of free space. $P_{0'}/P_0$ denotes the gain in transmitted power for a particular sample with respect to that of free space.

Liquid	$\Delta\frac{\omega_0}{2\pi}$ [GHz]	ΔS_{21} [dB]	$P_{0'}/P_0$
BCC	-17.20	3.81	2.40
Normal	-15.40	3.56	2.27
Adipose	-8.40	2.35	1.72

Rather than differentiating these tissues from free space, it is the distinction between them that is of interest. Therefore, the above table has to be converted to a more suitable space. The induced frequency shift, the S_{21} and the transmitted power through the host line are expressed in

Table IV.5 with respect to that of basal cell carcinoma.

Table IV.5: Simulated sensor response to water, normal fibrous and healthy adipose. Values are given with respect to those of **basal cell carcinoma** (BCC). $P_{0'}/P_{BCC}$ denotes the change in transmitted power for a particular sample with respect to that of BCC.

Liquid	$\Delta\frac{\omega_0}{2\pi}$ [GHz]	ΔS_{21} [dB]	$P_{0'}/P_{BCC}$
H_2O	-2.00	0.30	1.07
Normal	1.80	-0.25	0.94
Adipose	8.80	-1.46	0.72

Overall, one could state that differences are quite low. However, as such statement has no meaning in physics, the question of the sensitivity will be addressed in a later section upon experimental characterization of a real sensor.

IV.3.5 Matrix Implementation

Challenges towards imaging excised breast tissue during the surgical act are to acquire accurately the image in a short time scale. Therefore the implementation of several SRRs on the same sensing area is almost compulsory. Hence, on designing such a sensing cell, one has to achieve dense imaging object coverage while reducing as far as possible the dead zones that may lie between sensors. As there is no such thing as a free lunch, the increase of imaging density may in turn beget pixel cross-coupling effect that may alter sensor sensitivity. Therefore, one would suggest to dedicate a unique oscillator to a single sensor. Unfortunately due to the size mismatch between illumination sources and the SRRs, such a task is not feasible. Indeed, the SRRs are commonly of subwavelength size while available illuminators require transmission line that are rather of wavelength scale. Consequently, several SRRs have to share a collective illumination source so that it makes possible the dense integration of SRRs into the same sensing cell.

Collaborators at Wuppertal University successfully integrated 128-SRRs on a sensing area that spans over 3.2 mm. It consists of two rows of 64-pixels that are in turn chopped 0.56 terahertz triple-push oscillator using a power splitter network. The SRRs in each group are patterned with a 50 µm pitch, that is selected as a compromise between the fill factor and pixel cross coupling through the near field. Overall, filling factor by about 48% was achieved which corresponds to a density of around a hundred pixel per inch. A micrograph of the system on-chip is provided in Figure IV.20.

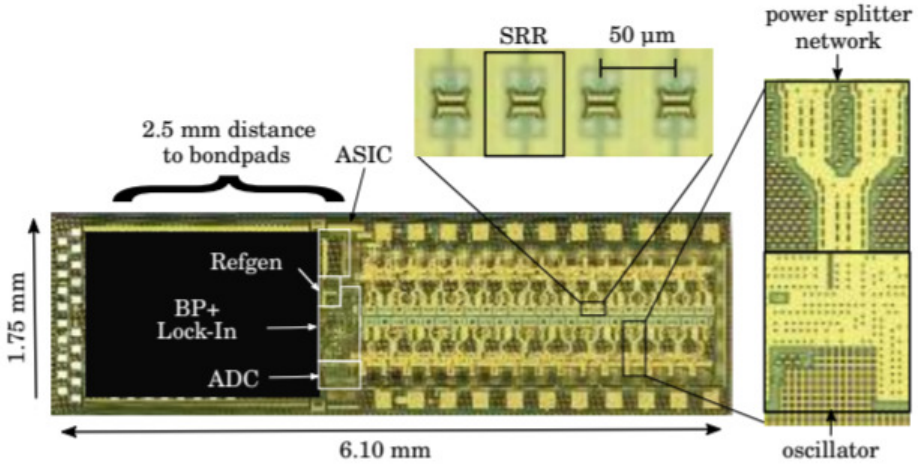


Figure IV.20: Micrograph of the near-field sensing array. The sensing strip length is 3.2 mm long. The black region contains additional circuits that are not associated to the near-field sensing. ASIC: application-specified integrated circuits. ADC: analog-to-digital converted. Refgen: reference generator. BP: bandpass.

IV.3.6 Pixel Readout

Generally, two main strategies are available towards capturing the array response upon loading it with a sample. The whole set of pixels may be simultaneously turned on and then are sequentially read out. Alternatively, pixels and their associated illumination sources can be turned on one at a time and pixel responses are captured through rapid electronic scanning across the sensing strip. The first strategy is commonly referred to as global shutter while the last is referred to as rolling shutter. A global shutter strategy is obviously faster than rolling shutter since the entire matrix is operated simultaneously. Nevertheless, although being slower, rolling shutter benefits from lower power consumption since pixels are turned on one by one. The time consumption associated to rolling shutter intrinsically depends on the time during which, the pixel is illuminated. Due to the power consumption that would be induced by a global shutter method, a rolling shutter strategy has been chosen to avoid temperature increase at the sensing surface to which, imaging may potentially be highly sensitive. The overall array readout architecture is given in Figure IV.21.

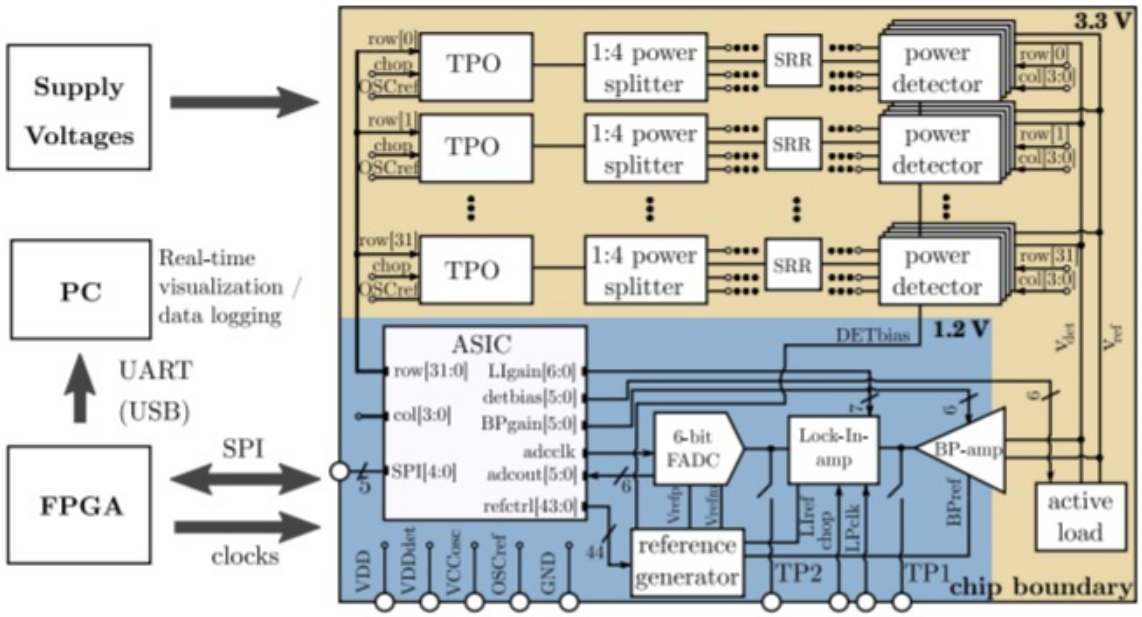


Figure IV.21: Block diagram of the system on a chip with the required external components. The chip contains 2 rows of 64 pixels, each divided into 16 subarrays of 4 entities illuminated by a common oscillator and an on-chip lock-in amplifier read-out. Pixel selection (row,col)) is achieved via an application-specific integrated circuit (ASIC). adlc$_{clk}$: analog-to-digital converter clock. FADC: flash analog-to-digital converter. GND: ground. LIref: lock-in reference. OSC: oscillator. TP: test point. UART: universal asynchronous receiver/transmitter. VDD: digital supply voltage. VCC: analog supply voltage.

IV.3.7 Encapsulation and Packaging

The reader of this manuscript must not forget that the chip has to operated under singular conditions for electronics. Indeed, upon loading the sensing line with freshly excised breast sections, biological fluids such as blood, lymph or water may be released. To handle those samples without being altered, the sensing array must be appropriately encapsulated. The sensing strip is covered with a 2-µm thick SiO_2 protective and passive layer. The sensing chip is assembled with a printed-circuit board by means of wire bonding. The wire bonds are encapsulated within standard epoxy resin. The chip takes place into an imaging module that is controlled by a field-programmable gate array (FPGA) via a serial peripheral interface bus clocked at 10 MHz. Ultimately, the required energy to drive the imaging module on is supplied by a standard USB port delivering a 5 volt supply. A photograph of the imaging module is hereafter given (Figure IV.22).

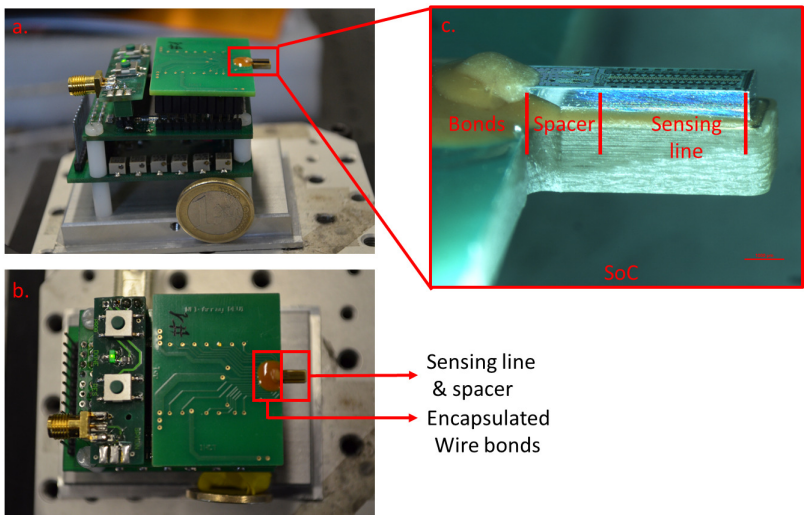


Figure IV.22: Photograph of the packaged imaging module. a: 3/4 view of the module with 1-euro coin for scale. b: top view of the module, highlighting the sensing line, the spacer and the wire bonds encapsulated in epoxy resin. c: zoom on the system on-chip (x20).

IV.3.8 Performances

The ability of the imaging module to differentiate between the different breast tissue types and more general samples is directly dictated by various performance metrics. These metrics are hereafter given and estimated.

*Nota: Overall, the performance metric estimations were performed at the Institute for High-Frequency and Communication Technology in Wuppertal. In the following, a brief description of these performances are reported. A reader who would like a less compact view of the sensor metrics is kindly referred to the thesis manuscript of Philipp Hillger **[38]** and to **[35]**. Both these documents relate accurately on the procedures that were undertaken to estimate the array performances.*

Dynamic Range

The dynamic range (DR) is usually defined as the ratio of the width of the interval within which the sensor response varies to the noise integrated over the sensor. In the foregoing, the response of a sensor to a perfect electrical conductor was simulated. Due to its ideal dielectric characteristics ($\hat{\epsilon} = \infty - 0j$), such a material provides the sensor maximum response $\Delta V_{0,max}$. However, the detectable value $\Delta V_{0,max}$ is intrinsically limited by the global noise arising from the operating sensor $V_{n,int}$. As the detectable $\Delta V_{0,max}$ can be shown to be equal to $P_{osc}(P_t^{PEC} - P_t^0)R_V$, wherein P_{osc} is the illumination power supplied by the oscillator, P_t are the transmission factors for PEC and free space, and R_V is the sensor voltage responsivity, the dynamical range of the sensor can be approximated as:

$$DR_{[dB]} = 20log_{10}\frac{P_{osc}(P_t^{PEC} - P_t^0)R_V}{V_{n,int}}, \tag{IV.10}$$

Finally, depending on the readout mode, that is either digital or analog, the dynamic range may significantly change. In a nutshell, the chip shown a dynamic range by about 63.8 dB in analog mode with a external lock-in detection at 1 ms time constant, and a dynamic range in digital readout mode at a frame rate of 28 frame-per-second of around 38.5 dB.

Sensitivity

In previous section we pointed out the importance of the sensor sensitivity to permittivity changes to distinguish between samples whose permittivity are relatively close. We defined the minimum

perceivable permittivity variation as:

$$\Delta\hat{\epsilon}^{min} = \frac{V_{n,int}}{P_{osc} R_V \frac{\delta P_t}{\delta \hat{\epsilon}}}. \tag{IV.11}$$

Alternatively, one can define the sensitivity to as the maximum possible contrast within an image. Such a contrast is impacted by various parameters other than the sensor dynamic range. For instance, sharing one illumination source for four sensors leads to an uncertainty in the detected response that scales with sample intrinsic permittivity. To estimate the uncertainty in relative permittivity, colleagues in Wuppertal estimated that the cross-coupling level is overall ranging from -12 dB to -16 dB. This leads, in a worst-scenario, to an approximate absolute uncertainty towards relative permittivity of around 0.33.

Spatial Resolution

Here comes one additional and elementary parameter. The need for greater resolution towards breast tissue demarcation has motivated the development of such an array. Then, what about it? The spatial resolution of the sensing area was estimated by scanning the sensor response to a nickel grid exhibiting a 50 µm bar width and a 250 µm bar pitch, along two orthogonal directions. By doing so, the resolution was estimated to be equal or larger than 10-12µm along each axis. This corresponds to around $\lambda/71$ at the operating frequency. Consequently, the device developed within the Nearsense project successfully beat the diffraction limit of resolution.

IV.3.9 Experiments on Liquids

The most simple and direct measurement that can be performed with the array is on liquids. Indeed, the topology of the array makes it difficult to analyze stiff materials. The sensing line shows hills (sensors) and valleys (pixel spacing). Therefore, on probing any sample, one has to perfectly control the distance of the sample to be inspected from the sensing surface. Such a task may be quite tricky while liquids naturally shape to cover the whole sensing line and concomitantly ensure a perfect contact with sensors. The surface profile of the sensing line is highlighted in Figure IV.23.

Therefore, liquids were the first samples to be experimentally inspected with the near-field sensing array. Due to the ambiguity regarding the influence of the real and imaginary part of the permittivity on sensor response, a careful selection of permittivity-known liquid samples is required. In the foregoing, we simulated the theoretical response of water, methanol, ethanol and propanol. It was demonstrated that the power transmitted through the transmission line of sensors is expected to increase from propanol to water. In addition to those, paraffin that is particularly transparent to terahertz radiations and exhibit smaller permittivity lossless part was simultaneously investigated. To verify the compliance with simulations, a drop of each of those liquids was deposited onto the array surface, as shown in Figure IV.24. Eventually, each of those sample was measured 20 times in a row with different droplets. Between each different liquid measurements, the matrix surface was cleaned with propanol to ensure no contamination between samples. Few seconds elapsed after cleaning to ensure propanol evaporation.

In Table IV.6 are briefly reported the real and the imaginary part of the permittivity for each of the inspected liquids at around 560 GHz.

The response of the near-field matrix as a function of the different liquids is given in Figure IV.25. An offset signal was recorded from free space over twenty measurements. The mean offset was subsequently subtracted from the mean response (over twenty independent measurements) exhibited by each pixel in response to the presence of a droplet.

Overall, results are in good agreements with previous simulations. Indeed, the matrix response increases from propanol (lowest permittivity; Figure IV.25 d) to methanol (highest permittivity; Figure IV.25 b) through ethanol (intermediary permittivity; Figure IV.25 c). Such a behavior was

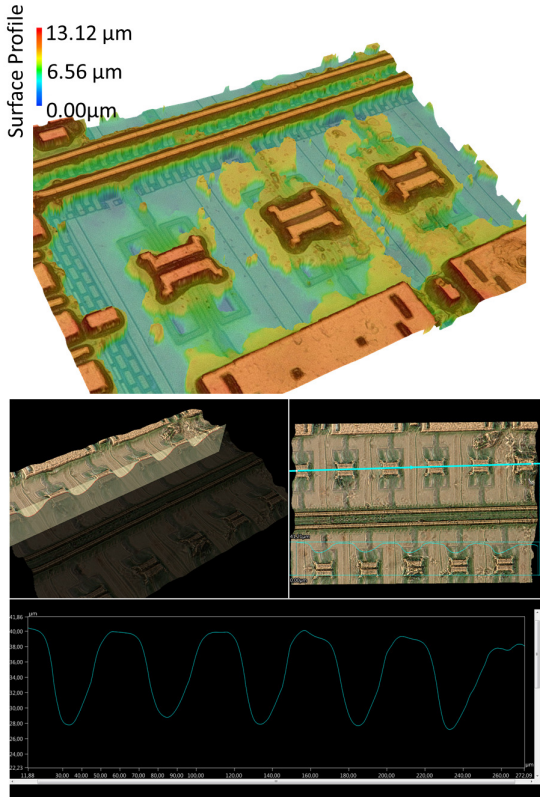


Figure IV.23: Surface profile analysis of a small fraction of the sensing strip. Strip lines are elevated by about 12µm. Such a topology makes it difficult to accurately sense stiff material as a perfect control of the altitude and the flatness with respect to the sensing line are required.

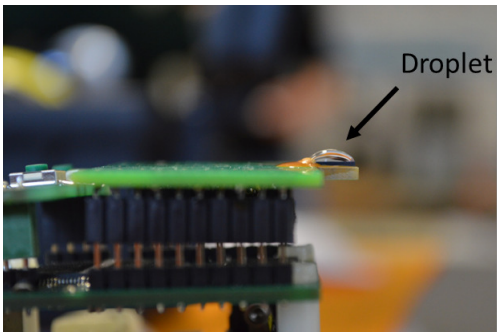


Figure IV.24: Droplet deposited onto the top of the sensing line.

Table IV.6: Permittivity lossless and lossy part for water, methanol, ethanol and propanol at 560 GHz from **[36]** and this work.

Liquid	ϵ'	ϵ''
H_2O	4.74	3.19
MeOH	3.16	1.13
EtOH	2.43	0.54
PrOH	2.25	0.31

anticipated by simulations. It is also interesting to remind the reader that simulations were suggesting a strong proximity between the response to propanol and to ethanol. From Figure IV.25 c and d, it is evident that the global matrix response to these liquids is particularly close. However, the results tend to suggest that the matrix is overall able to differentiate between those two alcohols despite their vicinity in terms of dielectric profile. An alternative representation of the matrix

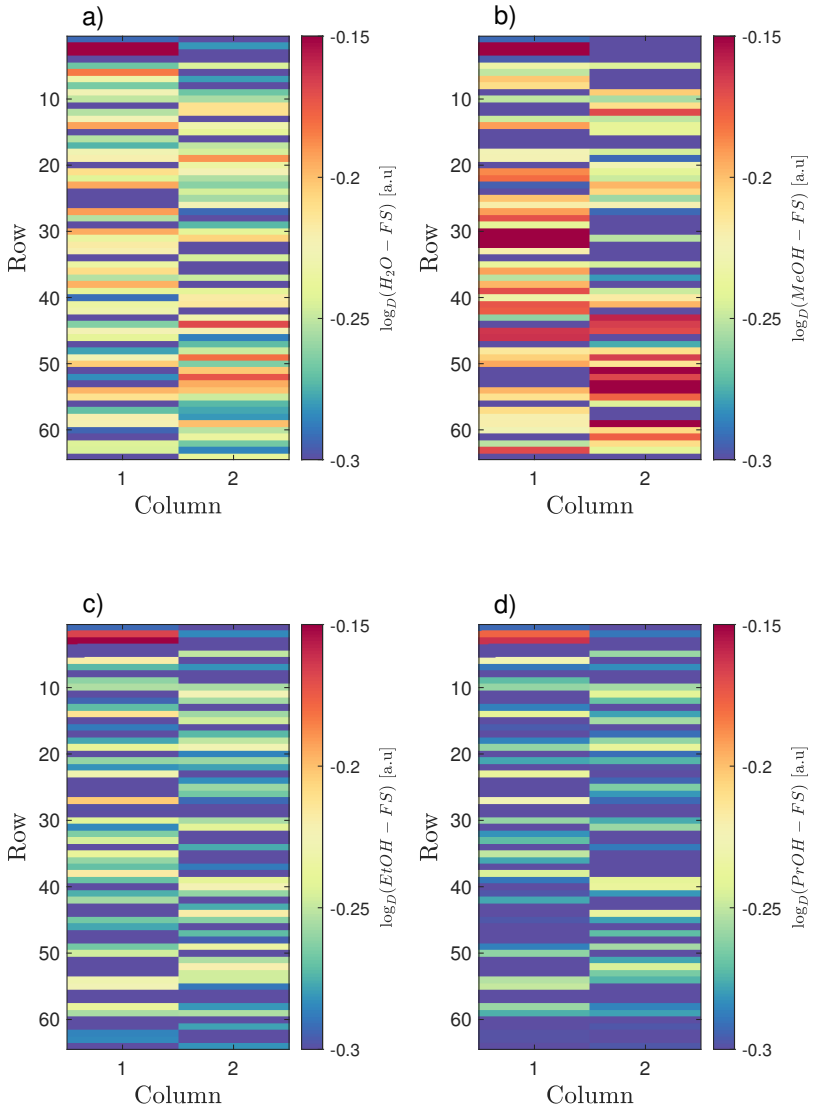


Figure IV.25: Response of the near-field array pixels as a function of the deposited liquids to that of free space. a: global matrix response to water. b: global matrix response to methanol. c: global matrix response to ethanol. d: global matrix response to propanol. Logarithmic scale in arbitrary units. The reader is invited to note that the scale is logarithmic rather than linear to highlight matrix response difference between liquids.

response is given in Figure IV.26.

However, -and it is likely the first thing that strikes- the global matrix response to water is not in agreement with simulations. In fact, due to its high real and imaginary permittivities, one would -reasonably- expect that water gives arise to the highest response. Previous simulations at least seemed to indicate that water should have lead to a greater response than that from methanol. It is therefore necessary to wonder about the practical origins of such a behavior.

The hypothesis that should prevail is the drastic impact of the lossy part of water permittivity. A careful reader would point out that simulations were already taken into account dielectric losses. That is right. Nonetheless, there is a significant difference between the simulations and the measurements. To derive the theoretical matrix response the different liquid samples were *deposited* only onto the top of the sensing strips. On the contrary, during measurements, the entire

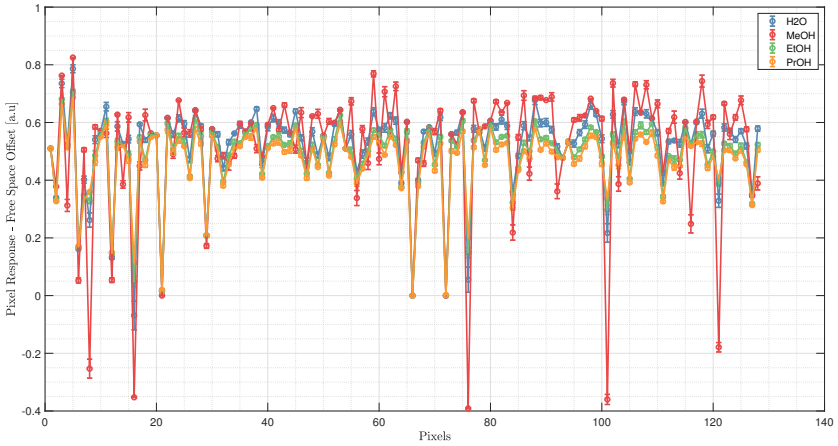


Figure IV.26: Pixel response to water (blue), methanol (red), ethanol (green) and propanol (yellow). The mean level of propanol is slightly lower than that of ethanol. The mean level of the matrix response to methanol is higher than for other liquids. Water stands in the middle as a result of the impact of its lossy part of permittivity.

top surface of the near-field array is filled with liquid sample rather than solely standing on strip lines. Hence, in simulation, the impact of the intrinsic dielectric properties of samples were underestimated. As alcohols have globally lower permittivities, the sensitivity of the matrix is weakly altered and their analysis are compliant with respect to simulations. However, in the specific case of water, covering the matrix entire surface with a droplet leads to a decrease in sensor response although simulations suggested an increase in response. Obviously, other parameters may -on their scale- be responsible. Simulations did not take into account neither cross-coupling effect, neither possible interactions with the module circuitry.

IV.3.10 Biomedical Imaging: a Premise

Different experiments towards imaging biological samples were conducted within the scope of the collaboration: a human finger was imaged with a 1D continuous scan and a 1D raster scan of a deparaffinized pathology section was acquired.

Fingerprints

The dielectric permittivity of the skin in the terahertz range was widely investigated and was reported to be of around $\hat{\epsilon} = 4.6 - 2.2j$ **[39; 40]**. Hence the global response of the matrix is expected to be significant. Moreover we reported the drastic sensitivity of the matrix to the distance of the sample from the chip top surface. Finger surface is made of ridges and hollows: the fingerprints. Overall, the ridges of the fingerprints ridges size between 30µm to 80µm. Hence, the sensing chip is expected to be able to provide the biometric fingerprint data. A human finger was scanned in contact with the near-field array without high accuracy mechanics setups. The finger was scanned in the direction normal to sensing line. A surface of 3.2 mm × 12 mm with a pixel pitch of 25 µm in Y-direction and 14.25 µm in X-direction which almost matches the maximum sensor resolution. The acquisition time was by about 30 seconds. In Figure IV.27 is given the acquired image from the finger. Globally, the near-field array succeeded in resolving the topology of the finger.

Pathology Section

A 560 GHz near-field image of a deparaffinized breast tissue section was also obtained at Wuppertal. As such kind of sample is overall rigid, the planarity of the sample surface with respect to that

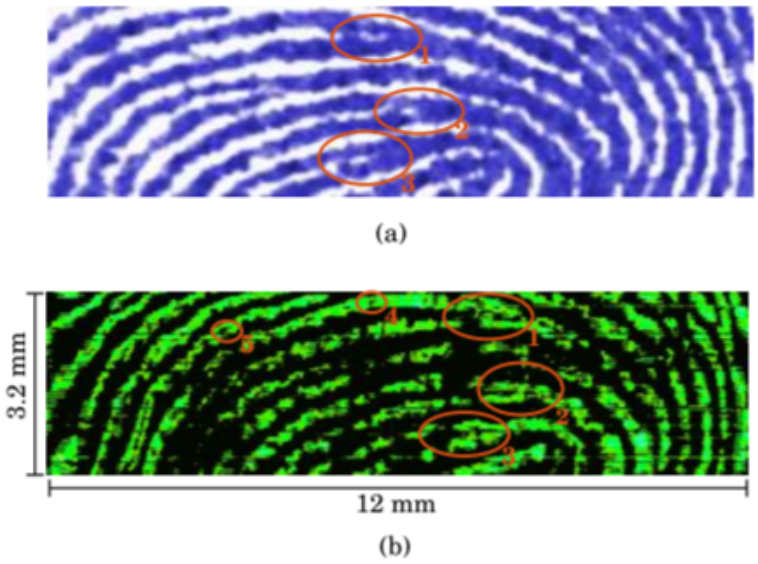


Figure IV.27: Biometric fingerprint data comparison **[35]**. a: ink-and-paper fingerprint. b: 560 GHz near-field image of the equivalent finger region. Points 1, 2 and 3 depict the presence of ridge splitting that are well retrieved from ink-and-paper fingerprint to terahertz image.

of the sensing lines has to be perfectly adjusted. Indeed, let's remind that sensor response is particularly sensitive to the distance of a sample from its surface. Therefore, a slight tilt between the two surfaces would induce the acquisition of a topological image rather than pathological. Hence, our collaborators equipped with a manual tilt adjustment stage the z-axis and added with an objective holder the x,y-direction to enable sample planarization through optical focus alignment. The whole acquisition setup that was adjusted in Wuppertal is described in the PhD manuscript of Philipp Hillger. The first image that was acquired from a breast tissue -although being fixed- is given in Figure I.28, where the sample was assumed to be in contact with the sensing line.

The terahertz near-field image clearly shows some features that are highlighted in term of pixel intensity in Figure IV.28. The generic morphology of the deparaffinized section is also compliant with the obtained picture. However, at this stage, it is not possible to state on the efficiency of the matrix to distinguish between cancer and normal tissues. In previous chapter we endlessly pointed out that an important number of samples are required to draw more statistical meaningful conclusions. Here, at this time, one unique sample was successfully imaged. However, from the topological image that is associated (Figure IV.28 b), we can underline that the near-field image does not particularly exhibit the same features. Nevertheless, due to the sensitivity of the matrix to the sample distance, the contrast observed in the terahertz picture may, in part, arise from the sample topology rather than solely pathological trends. Although the reasons for such a contrast are yet to be fully understood, the terahertz near-field resonator-based array may provide solid foundation towards further developments for intra-operative margin cleanliness assessment. In fact, even though far-field imaging techniques and the previously reported numerical developments demonstrated some abilities towards breast malignancy identification, their associated resolution, acquisition time, staggering amount of recorded data, phase mismatch issue and low integration level are all major obstacles that are, overall, overcame with the near-field matrix. Obviously, as there is nothing such a free lunch, the use of the matrix induces ambiguity towards the influence of the real and imaginary part of the permittivity in the global sensor response, and implies a perfect management of the sample distance from the sensing surface. It is worth to mention that these issues, towards the use of the near-field sensor, are raised solely in view of the results obtained from a fixed breast section. While the distance management may be easily overcame for soft tissue with in-contact measurement, the response ambiguity due to the dielectric constant may over be, on the contrary, be increased. Finally, unlike as compared to far-field imaging and spectroscopy, where multiple classifiers can be derived for tissue classification, the near-field ma-

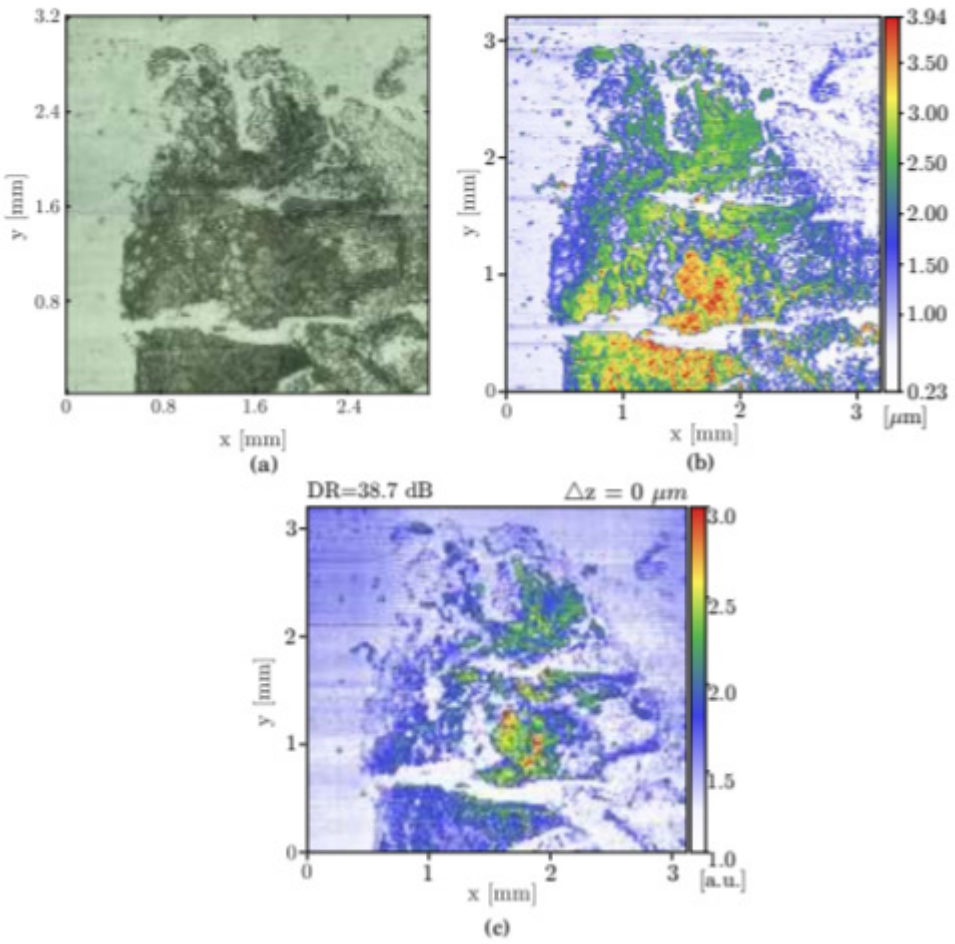


Figure IV.28: Terahertz near-field imaging of a deparaffinized breast section **[35]**. a: micrograph of the deparaffinized imaged breast section; b: height profile of the sample; c: terahertz near-field image at 560 GHz with the section in contact with the array.

trix solely measures a unique wavelength. Therefore other identification parameters, that would require further technical developments, need to be determined and evaluated.

IV.4 Conclusion

The last chapter of this manuscript presented the advances towards the development of a silicon-based THz sub-wavelength imager usable in life science applications, especially for tumour margin identification.

The operating frequency of the terahertz near-field sensor was conditioned by prior far-field investigations led on freshly excised breast cancer tissues. These studies implied that frequencies ranging from 300 GHz to 600 GHz were, globally, able to provide some demarcations between tissues. Specifically, investigations at 560 GHz in Chapter III, demonstrated from different numerical treatment (inverse electromagnetic problem and principal component analysis) that this frequency provides contrast between healthy and malignant tissue, and therefore, is well-suited to be employed in an integrated near-field imaging sensor to break the diffraction limit of resolution while being sensitive to breast lesion. Indeed, in contrast to the resolution of regular (*i.e.* far-field), near-field system provides a resolution closer to the typical eukaryote cell diameter, and thus might enable enhanced malignant margin recognition.

A wide part of *Chapter IV* was consequently dedicated to numerical investigations aiming to outline the theoretical mechanisms of interactions between such kind of resonator-based sensors with biological objects. The conceptualized system shown from finite integration treatment, im-

portant sensitivity to the lossless part of the permittivity. The sensor response was estimated to increase in a bi-exponential fashion to increasing permittivity lossless part. These numerical investigations also allowed to identify the influence of the lossy part of the permittivity to the global sensor response. The impact of the lossy part was demonstrated to increase with lossless part increase and to alter the sensor sensitivity. Consequently, the sensor is prone to response ambiguity between the real and imaginary part of an object's permittivity. Although not being fully investigated during this work, pixel cross-coupling is also expected to desensitize even more the sensor to an object. Indeed, the sensor response of a given pixel may be altered with the object's permittivity located at neighboring pixels. Such a parasitic coupling may consequently lead to an ambiguity regarding the sensor response to a given surface as it may account for local permittivity rather than punctual permittivity.

Furthermore, comparison between simulations performed on liquids and related experiments outline valuable insight towards matrix response. From simulations, water sample was expected to give arise to the higher matrix response in comparison to methanol, ethanol and propanol. Although experiments have shown that alcohols were succesfully sorted from lower to higher relative permittivity, sensor response to water was not matching expectations as the response to water was standing lower than that of methanol. We draw the hypothesis that the high lossy part of water permittivity was disturbing the global behavior of the matrix rather than solely the local permittivity at resonator gap vicinity. Further work are consequently required to simulate the entire array, with the entire pixel setting and associated power-divider network.

Distinction between cancerous and healthy fibrous tissue have been reported to be challenging as the ratio in transmitted power between those two tissue types was reported to be by about 0.94. The subtle difference was however expected, as the relative permittivity of these two tissue types are remarkably close. Let's underline that such a proximity in dielectric profile was pointed out in previous chapters and overall conditions demarcation by means of terahertz radiations. As the specific difference of 1.5% in transmitted power is exclusively related to the oscillator supply illumination at 560 GHz, additional parameters have to be found to ascertain the classification. One possible way would be to measure the rising slope in the neighborhood of the oscillator excitation frequency, by detuning it. Indeed, we have shown that such a slope was mainly dictated by the lossy part of permittivity and therefore may provide a supplementary characterization factor.

The array was fabricated and its performances was characterized at the Institute for High-Frequency and Communication Technology in Wuppertal. Our collaborators reported a spatial resolution of around $\lambda/71$ at the operating frequency, succesfully breaking the diffraction limit of resolution at terahertz frequencies. The dynamic range was estimated to be by about 63.8 dB in analog mode and 38.5 dB in digital one. Last but not least, the absolute uncertainty towards observed relative permittivity was estimated by about 0.33.

Finally, the chapter ends on the first image worldwide of fixed breast tissue section, using a terahertz near-field resonator-based array. Even though the contrast mechanisms are expected to be due to the concomitant effect of the sample topology and relative permittivity, it highlights the consequent advances that have been accomplished within the scope of the Nearsense project. Although this chapter ends and consequently and with it this manuscript, it is clear that a huge amount of work remains. From the identification of additional classification parameters to the implementation of new sensor design to enable better characterization and through statistical analysis on numerous freshly excised breast tissue, the work that has to be done will require further efforts to ultimately find an intra-operative device to beat breast cancer.

IV.5 Bibliography

[1] JB Perraud, J Bou Sleiman, B Recur, H Balacey, F Simoens, JP Guillet, and P Mounaix. Liquid index matching for 2d and 3d terahertz imaging. *Applied optics*, 55(32):9185–9192, 2016. 166

[2] Nikita V Chernomyrdin, Aleksander O Schadko, Sergey P Lebedev, Viktor L Tolstoguzov, Vladimir N Kurlov, Igor V Reshetov, Igor E Spektor, Maksim Skorobogatiy, Stanislav O

Yurchenko, and Kirill I Zaytsev. Solid immersion terahertz imaging with sub-wavelength resolution. *Applied Physics Letters*, 110(22):221109, 2017. 166

[3] EdwardH Synge. A suggested method for extending microscopic resolution into the ultramicroscopic region. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(35):356–362, 1928. 166

[4] Andreas J Huber, Fritz Keilmann, J Wittborn, Javier Aizpurua, and Rainer Hillenbrand. Terahertz near-field nanoscopy of mobile carriers in single semiconductor nanodevices. *Nano letters*, 8(11):3766–3770, 2008. 168, 172

[5] Romain Lecaque, Samuel Grésillon, Nicolas Barbey, Romain Peretti, Jean-Claude Rivoal, and Claude Boccara. Thz near-field optical imaging by a local source. *Optics communications*, 262(1):125–128, 2006. xii, 168

[6] Romain Lecaque, Samuel Grésillon, and Claude Boccara. Thz emission microscopy with subwavelength broadband source. *Optics express*, 16(7):4731–4738, 2008. xii, 168

[7] Minwoo Yi, Kanghee Lee, Jongseok Lim, Youngbin Hong, Young-Dahl Jho, and Jaewook Ahn. Terahertz waves emitted from an optical fiber. *Optics Express*, 18(13):13693–13699, 2010. xii, 168

[8] Kosuke Okada, Kazunori Serita, Zirui Zang, Hironaru Murakami, Iwao Kawayama, Quentin Cassar, Gaetan Macgrogan, Jean-Paul Guillet, Patrick Mounaix, and Masayoshi Tonouchi. Scanning laser terahertz near-field reflection imaging system. *Applied Physics Express*, 12(12):122005, 2019. 168

[9] MA Seo, AJL Adam, JH Kang, JW Lee, SC Jeoung, Q Han Park, PCM Planken, and DS Kim. Fourier-transform terahertz near-field imaging of one-dimensional slit arrays: mapping of electric-field-, magnetic-field-, and poynting vectors. *Optics express*, 15(19):11781–11789, 2007. xii, 169, 170

[10] Markus Wächter, Michael Nagel, and Heinrich Kurz. Tapered photoconductive terahertz field probe tip with subwavelength spatial resolution. *Applied Physics Letters*, 95(4):041112, 2009. 169

[11] Andreas Bitzer and Markus Walther. Terahertz near-field imaging of metallic subwavelength holes and hole arrays. *Applied Physics Letters*, 92(23):231101, 2008. xii, 170

[12] Reshmi Chakkittakandy, Jos AWM Corver, and Paul CM Planken. Quasi-near field terahertz generation and detection. *Optics Express*, 16(17):12794–12805, 2008. xii, 169, 170

[13] Guilhem Gallot, SP Jamison, RW McGowan, and D Grischkowsky. Terahertz waveguides. *JOSA B*, 17(5):851–863, 2000. 170

[14] EA Ash and G Nicholls. Super-resolution aperture scanning microscope. *Nature*, 237(5357):510, 1972. 170

[15] O Mitrofanov, I Brener, R Harel, JD Wynn, LN Pfeiffer, KW West, and John Federici. Terahertz near-field microscopy based on a collection mode detector. *Applied Physics Letters*, 77(22):3496–3498, 2000. xii, 170, 171

[16] Oleg Mitrofanov, M Lee, JWP Hsu, LN Pfeiffer, KW West, JD Wynn, and JF Federici. Terahertz pulse propagation through small apertures. *Applied Physics Letters*, 79(7):907–909, 2001. 171

[17] Kunihiko Ishihara, Keishi Ohashi, Tomofumi Ikari, Hiroaki Minamide, Hiroyuki Yokoyama, Jun-ichi Shikata, and Hiromasa Ito. Terahertz-wave near-field imaging with subwavelength resolution using surface-wave-assisted bow-tie aperture. *Applied physics letters*, 89(20):201120, 2006. xii, 171

[18] Eric X Jin and Xianfan Xu. Obtaining super resolution light spot using surface plasmon assisted sharp ridge nanoaperture. *Applied Physics Letters*, 86(11):111106, 2005. 171

[19] K Şendur and W Challener. Near-field radiation of bow-tie antennas and apertures at optical frequencies. *Journal of Microscopy*, 210(3):279–283, 2003. 171

[20] Kunihiko Ishihara, Gen-ichi Hatakoshi, Tomofumi Ikari, Hiroaki Minamide, Hiromasa Ito, and Keishi Ohashi. Terahertz wave enhanced transmission through a single subwavelength aperture with periodic surface structures. *Japanese journal of applied physics*, 44(7L):L1005, 2005. 171

[21] Hans Albrecht Bethe. Theory of diffraction by small holes. *Physical review*, 66(7-8):163, 1944. 171

[22] CJ Bouwkamp. On the diffraction of electromagnetic waves by small circular disks and holes. *Philips Research Reports*, 5:401–422, 1950. 171

[23] S Mair, B Gompf, and M Dressel. Spatial and spectral behavior of the optical near field studied by a terahertz near-field spectrometer. *Applied physics letters*, 84(7):1219–1221, 2004. 171

[24] Aurele Joseph Louis Adam. Review of near-field terahertz measurement methods and their applications. *Journal of Infrared, Millimeter, and Terahertz Waves*, 32(8-9):976, 2011. 172

[25] NCJ Van der Valk and PCM Planken. Electro-optic detection of subwavelength terahertz spot sizes in the near field of a metal tip. *Applied Physics Letters*, 81(9):1558–1560, 2002. 172

[26] Hui Zhan, Victoria Astley, Michael Hvasta, Jason A Deibel, Daniel M Mittleman, and Yong-Sik Lim. The metal-insulator transition in vo 2 studied using terahertz apertureless near-field microscopy. *Applied Physics Letters*, 91(16):162110, 2007. 172

[27] Daniel Rugar and Paul Hansma. Atomic force microscopy. *Physics today*, 43(10):23–30, 1990. 172

[28] Yukio Kawano and Koji Ishibashi. An on-chip near-field terahertz probe and detector. *Nature Photonics*, 2(10):618, 2008. 173

[29] Alexander J Macfaden, John L Reno, Igal Brener, and Oleg Mitrofanov. 3 μ m aperture probes for near-field terahertz transmission microscopy. *Applied Physics Letters*, 104(1):011110, 2014. 173

[30] F Blanchard, A Doi, T Tanaka, H Hirori, H Tanaka, Y Kadoya, and K Tanaka. Real-time terahertz near-field microscope. *Optics express*, 19(9):8277–8284, 2011. 173

[31] Lin-Feng Chen, CK Ong, CP Neo, Vasundara V Varadan, and Vijay K Varadan. *Microwave electronics: measurement and materials characterization*. John Wiley & Sons, 2004. 173

[32] Laven Mavarani, Philipp Hillger, Thomas Bücher, Janusz Grzyb, Ullrich R Pfeiffer, Quentin Cassar, Amel Al-Ibadi, Thomas Zimmer, Jean-Paul Guillet, Patrick Mounaix, et al. Nearsense–advances towards a silicon-based terahertz near-field imaging sensor for ex vivo breast tumour identification. *Frequenz*, 72(3-4):93–99, 2018. 175

[33] Ullrich R Pfeiffer, Philipp Hillger, Ritesh Jain, Janusz Grzyb, Thomas Bucher, Quentin Cassar, Gaetan MacGrogan, Jean-Paul Guillet, Patrick Mounaix, and Thomas Zimmer. Ex vivo breast tumor identification: Advances toward a silicon-based terahertz near-field imaging sensor. *IEEE Microwave Magazine*, 20(9):32–46, 2019. 175

[34] Philipp Hillger, Janusz Grzyb, Ritesh Jain, and Ullrich R Pfeiffer. Terahertz imaging and sensing applications with silicon-based technologies. *IEEE Transactions on Terahertz Science and Technology*, 9(1):1–19, 2018. 176

[35] Philipp Hillger, Ritesh Jain, Janusz Grzyb, Wolfgang Förster, Bernd Heinemann, Gaëtan MacGrogan, Patrick Mounaix, Thomas Zimmer, and Ullrich R Pfeiffer. A 128-pixel system-on-a-chip for real-time super-resolution terahertz near-field imaging. *IEEE Journal of Solid-State Circuits*, 53(12):3599–3612, 2018. xiv, 177, 184, 188, 193, 194

[36] JT Kindt and CA Schmuttenmaer. Far-infrared dielectric properties of polar liquids probed by femtosecond terahertz pulse spectroscopy. *The Journal of Physical Chemistry*, 100(24):10373–10379, 1996. xvi, 183, 190

[37] Bao CQ Truong, Hoang D Tuan, Anthony J Fitzgerald, Vincent P Wallace, and Hung T Nguyen. A dielectric model of human breast tissue in terahertz regime. *IEEE Transactions on Biomedical Engineering*, 62(2):699–707, 2014. 184

[38] Philipp Hillger. *Silicon-based Systems for Microscopic and Volumetric Terahertz Imaging*. PhD thesis, Universität Wuppertal, Fakultät für Elektrotechnik, Informationstechnik und . . . , 2020. 188

[39] Bao CQ Truong, Hoang Duong Tuan, Ha Hoang Kha, and Hung T Nguyen. Debye parameter extraction for characterizing interaction of terahertz radiation with human skin tissue. *IEEE Transactions on Biomedical Engineering*, 60(6):1528–1537, 2013. 192

[40] Emma Pickwell, Anthony J Fitzgerald, Bryan E Cole, Philip F Taday, Richard J Pye, Thomas Ha, Michael Pepper, and Vincent P Wallace. Simulating the response of terahertz radiation to basal cell carcinoma using ex vivo spectroscopy measurements. *Journal of Biomedical Optics*, 10(6):064021, 2005. 192

General Conclusion

This work debates on the ability of terahertz radiations to discriminate between healthy breast tissues and malignant breast lesions. Ultimately, this manuscript aims to propose elements and solutions to enable greater accuracy, thanks to terahertz waves, in breast tumor margin delineations, intraoperatively.

Regular terahertz scanners have been pointed out to be drastically impacted by the diffraction limit of resolution that restrict their spatial resolution to approximately $\lambda/2$, which is in the terahertz band far from the typical size of an eukaryote cell. Therefore, to beat the diffraction limit is essential to achieve cellular resolution and thereby to provide unambiguous information on tumor edges.

To equip surgery rooms with such a sub-wavelength system, preliminary studies with regular far-field terahertz scanning systems were conducted at hospital to characterize the generic behavior of breast volume to frequency-varying terahertz stimuli. Led on a total of forty-eight tissue waveforms, introductory spectroscopic investigations allowed to draw the mean dielectric profiles of adipose, fibrous and cancer tissues between 200 GHz and 1 THz. Overall, frequencies spanning from 300 GHz to 600 GHz were reported to give arise to the highest dielectric contrast between the different breast tissue types. Adipose tissues made of oil, fat and steroids exhibit optical constant that are consistently lower than other breast tissues. The mean demarcation between adipose and fibrous -that is the nearest tissue type within the meaning of dielectroscopy- is by about 20% in refractive index and ranges from 50% to 75% in extinction coefficient within the optimal contrast frequency band, making them easily identifiable by means of terahertz radiations. The mean demarcation between fibrous tissues and malignant lesions is however more subtle. While no distinction are observed towards the extinction coefficient, the refractive index difference varies from 0% to 10%. The intensity of the refractive index difference is assumed to be dictated by the cancer cell density present within the tissue section. Nevertheless, the set of samples may remain relatively small to draw statistical conclusion and a greater number of sections would allow to better anticipate the overall dispersion trends of the different breast tissues to terahertz stimuli. No specific features have been identified to depend on the patient age, the pathology etiology neither the tumor stage and grade.

On another note, it seems possible to conduct equivalent studies on different organs. Specifically on stratified organs such as the skin as the reported iterative tree algorithm (ITA) is aimed to provide a full set of parameters that depicts the optical path within structures. First and foremost, the ITA is aimed to be employed for non-destructive testing of 4D printed skins.

To better appreciate the mechanisms behind contrast, examinations on MCF7 cells have been conducted and results indicate that cancer cells possess a large relaxation mechanism attributed to a great number of quasi-free water molecules. Alternatively, the results suggest that the strength of bulk-water bonds in cancer cells -although being in small amount- is weaker than in liquid water. However, in order to ascertain these assumptions, further studies a required, and particularly on healthy cells. Moreover, the evaluation of the scattering effect depending on the size of cell aggregates would allow to clarify these statements.

As the difference in fibrous and cancer are relatively small, a self-reference imaging method is described to improve sample characterization. Such a method combined with inverse electromagnetic problem and refractive index classification rules has provided a best trade-off between sensitivity and specificity by about 80% sensitivity for 82% specificity. Additional surveys with

principal component analysis are overall giving arise to an equivalent pixel classification. In this field, a lack of image registration techniques has to be pointed out. The numerous differences between pathology clichés and terahertz images make the correlations tricky, cumbersome and possibly inaccurate.

Based on these findings, an operating frequency of 560 GHz for the near-field imager was defined. More than solely standing in the optimal dielectric contrast window, such a frequency matches with BiCMOS technology capabilities. This work reports on the successful integration of 128 terahertz near-field pixels on a sensing line spanning over 3.2-mm, within the scope of the Nearsense project. Different simulations have been performed to assess the theoretical response of the matrix to a panel of dielectric configuration. The resonator-based sensing line is sensitive to local permittivity disturbances that results in resonance frequency shift in each individual sensor. The matrix response is overall greatly sensitive to the real part of the permittivity. The influence of the lossy part of permittivity is on its part increasing with lossless part increase. The lossy part of permittivity simultaneously participates in shifting to low frequencies the resonance frequency of resonator and in reducing the overall matrix sensitivity. Hence the near-field sensor response is prone to ambiguity between the influence of the lossless and lossy part of permittivity. Cross-coupling effect between neighbor pixels is also underlined as a potential parameter altering the compliance of the near-field response. Further work towards the simulation of the entire imaging module are required to strip away the ambiguity. Some additional perspectives are to conduct measurements for different oscillator supply voltages (different frequencies) in order to see an impact. Theoretically, the real and imaginary part of permittivity can be deduced by measuring the whole tuning curve of the oscillator and thereby the curvature of the resonance slope. Ultimately, such a process may allow to calibrate the near-field response to real and imaginary parts. The difference in resonance frequency shift between fibrous tissue and malignant lesion has been anticipated to be particularly subtle and therefore the distinction is expected to be challenging. The ratio in transmitted power through the transmission line between those two tissues has been approximated to be by about 0.94 which corresponds to a difference of around 0.4 in real permittivity and 0.3 in imaginary permittivity at 560 GHz, while the absolute uncertainty towards relative permittivity of the matrix response has been estimated to be around 0.33. The terahertz near-field array shows a spatial resolution of $\lambda/71$, with which objects separated by a distance by about ten microns should be resolved. A first image, with this unique sensor worldwide, of a deparaffinized breast tissue section is reported. The contrast in the image is reported to be due to both the sample topology and its relative permittivity changes over the surface.

The second round of the Nearsense project will focus on the careful calibration of the near-field imager to precisely determine the influence of the lossless and lossy part of permittivity to image contrast. Then, a wide campaign of image acquisition from numerous fixed breast tissues should be conducted to draw further statistical background. The obtained images are aimed to be compared to pathology images onto which cell counts will be performed for precise correlation. Once that is complete, intra-operative experimentation will be required to address real surgical conditions. Ultimately the geometry of the matrix is expected to be slightly translated to a square matrix that would drastically facilitate imaging.

It is time to close this manuscript. However, before this happens, a reader may ask him/herself: *Terahertz radiations for breast cancer recognition ?* A battery of results seems to indicate that there is effectively a contrast in the terahertz band between breast tissues and especially between normal and abnormal ones, even subtle. As a consequence, it suggests that terahertz radiations may potentially provide useful information towards improving the accuracy of breast tumor margin delineation. However, the field still stands in its early stages and numerous questions have been raised rather than being answered. Accordingly, and on a personal note, the question mark remains.

Author Publication List

†: *Peer-reviewed papers.*
‡: *Proceedings and Conferences.*

[1] †Cassar, Quentin, Amel Al-Ibadi, Laven Mavarani, Philipp Hillger, Janusz Grzyb, Gaëtan MacGrogan, Thomas Zimmer, Ullrich R Pfeiffer, Jean-Paul Guillet, and Patrick Mounaix. Pilot study of freshly excised breast tissue response in the 300–600 ghz range. *Biomedical optics express*, 9(7):2930–2942, 2018.

[2] †Cassar, Quentin, Adrien Chopard, Frederic Fauquet, Jean-Paul Guillet, Mingming Pan, Jean-Baptiste Perraud, and Patrick Mounaix. Iterative tree algorithm to evaluate terahertz signal contribution of specific optical paths within multilayered materials. *IEEE Transactions on Terahertz Science and Technology*, 9(6):684–694, 2019.

[3] †Ullrich R Pfeiffer, Philipp Hillger, Ritesh Jain, Janusz Grzyb, Thomas Bucher, Cassar, Quentin, Gaetan MacGrogan, Jean-Paul Guillet, Patrick Mounaix, and Thomas Zimmer. Ex vivo breast tumor identification: Advances toward a silicon-based terahertz near-field imaging sensor. *IEEE Microwave Magazine*, 20(9):32–46, 2019.

[4] †Laven Mavarani, Philipp Hillger, Thomas Bücher, Janusz Grzyb, Ullrich R Pfeiffer, Cassar, Quentin, Amel Al-Ibadi, Thomas Zimmer, Jean-Paul Guillet, Patrick Mounaix, et al. Nearsense–advances towards a silicon-based terahertz near-field imaging sensor for ex vivo breast tumour identification. *Frequenz*, 72(3-4):93–99, 2018.

[5] †Kosuke Okada, Kazunori Serita, Zirui Zang, Hironaru Murakami, Iwao Kawayama, Cassar, Quentin, Gaetan Macgrogan, Jean-Paul Guillet, Patrick Mounaix, and Masayoshi Tonouchi. Scanning laser terahertz near-field reflection imaging system. *Applied Physics Express*, 12(12):122005, 2019.

[6] †Mingming Pan, Cassar, Quentin, Frederic Fauquet, Georges Humbert, Patrick Mounaix, and Jean-Paul Guillet. Guided terahertz pulse reflectometry with double photoconductive antenna. *Applied Optics*, in press, 2020.

[7] †G Cagnoli, M Lorenzini, E Cesarini, F Piergiovanni, M Granata, D Heinert, F Martelli, R Nawrodt, A Amato, Cassar, Q, et al. Mode-dependent mechanical losses in disc resonators. *Physics Letters A*, 382(33):2165–2173, 2018.

[8] †E Cesarini, M Granata, V Fafone, Cassar, Q, F Martelli, LO Mereni, CB Hurtado, J Dickmann, A Amato, S Kroker, et al. The virgo coating collaboration: a detailed study on thermoelasticity in crystalline materials and other research lines. *PoS*, page 006, 2018.

[9] ‡Cassar, Q, CL Koch-Dandolo, JP Guillet, M Roux, F Fauquet, and Patrick Mounaix. Ancient painting on copper substrate inspected by terahertz spectroscopy-imaging. In *2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–1. IEEE, 2019.

[10] ‡Cassar, Q, CL Koch-Dandolo, JP Guillet, M Roux, F Fauquet, and P Mounaix. Varnishes of painting material studied by terahertz spectroscopy. In *2018 43rd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2. IEEE, 2018.

[11] ‡Cassar, Q, A Al-Ibadi, L Mavarani, P Hillger, J Grzyb, G MacGrogan, UR Pfeiffer, T Zimmer, JP Guillet, and P Mounaix. Studies on pca for breast tissue segmentation. In *2018 43rd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2. IEEE, 2018.

[12] ‡A Al-Ibadi, Cassar, Q, T Zimmer, G MacGrogan, L Mavarani, P Hillger, J Grzyb, UR Pfeiffer, JP Guillet, and P Mounaix. Thz spectroscopy and imaging for breast cancer detection in the 300–500 ghz range. In *2017 42nd international conference on infrared, millimeter, and terahertz waves (IRMMW-THz)*, pages 1–1. IEEE, 2017.

[13] ‡Nikolay S Balbekin, Cassar, Quentin, Olga A Smolyanskaya, Maksim S Kulya, Nikolay V Petrov, Gaetan MacGrogan, Jean-Paul Guillet, Patrick Mounaix, and Valery V Tuchin. Terahertz pulse time-domain holography method for phase imaging of breast tissue. In *Quantitative Phase Imaging V*, volume 10887, page 108870G. International Society for Optics and Photonics, 2019.

[14] ‡OA Smolyanskaya, Cassar, Q, MS Kulya, NV Petrov, KI Zaytsev, AI Lepeshkin, J-P Guillet, P Mounaix, and VV Tuchin. Interaction of terahertz radiation with tissue phantoms: numerical and experimental studies. In *EPJ Web of Conferences*, volume 195, page 10012. EDP Sciences, 2018.

[15] ‡A Al-Ibadi, J Bou Sleiman, Cassar, Q, Gaëtan Macgrogan, Hugo Balacey, T Zimmer, Patrick Mounaix, and Jean-Paul Guillet. Terahertz biomedical imaging: From multivariate analysis and detection to material parameter extraction. In *2017 Progress In Electromagnetics Research Symposium-Spring (PIERS)*, pages 2756–2762. IEEE, 2017.

[16] ‡A Chopard, J Bou Sleiman, Cassar, Q, P Fauché, JP Guillet, P Mounaix, M Pan, JB Perraud, and A Susset. Contactless terahertz paint thickness measurements: specificity of aeronautics industry. 2019.

[17] ‡M Picot, H Ballacey, JP Guillet, Cassar, Q, and P Mounaix. Terahertz paint thickness measurements: from lab to automotive and aeronautics industry. In *Proceedings of the 15th Asia Pacific Conference for Non-Destructive Testing (APCNDT2017), Singapore*, pages 13–17, 2017.

[18] ‡L Mavarani, P Hillger, J Grzyb, Cassar, Q, A Al-Ibadi, T Zimmer, G MacGrogan, JP Guillet, P Mounaix, and UR Pfeiffer. A novel approach of aqueous solution analysis using a fully-integrated terahertz near-field sensor. In *2017 42nd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2. IEEE, 2017.

[19] ‡Olga A Smolyanskaya, Maksim S Kulya, Cassar, Quentin, Olga V Kravtsenuk, Patrick Mounaix, Jean-Paul Guillet, Kirill I Zaytsev, and Nikolay V Petrov. Reconstructed thz phase image of the two-component numerical model of breast cancer tissue. In *2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–1. IEEE, 2019.

[20] ‡Olga A Smolyanskaya, Valery N Trukhin, Polina G Gavrilova, Evgeniy L Odlyanitskiy, Anna V Semenova, Cassar, Quentin, Jean-Paul Guillet, Patrick Mounaix, Kamil G Gareev, and Dmitry V Korolev. Terahertz spectra of drug-laden magnetic nanoparticles. In *Colloidal Nanoparticles for Biomedical Applications XIV*, volume 10892, page 108920L. International Society for Optics and Photonics, 2019.

Introduction en Français

Les cancers mammaires sont les cancers les plus fréquemment diagnostiqués chez la femme. En 2018, plus de 58 000 nouveaux cas ont été dénombrés en France métropolitaine. Le taux de survie nette sur âge standardisé à 5 ans pour les patientes diagnostiquées entre 2005 et 2010 s'élevait à 87%. Chaque année, le taux de mortalité est amené à diminuer en raison de l'amélioration graduelle des techniques de dépistage et des thérapies de plus en plus adaptées au niveau de risque associé à chaque patiente. Le traitement thérapeutique envisagé dépend généralement de trois critères d'évaluation carcinologique : (i). le stade tumoral ; (ii). le grade tumoral ; (iii). l'immunophénotype tumoral. Les thérapies varient d'une patiente à l'autre en raison de la drastique hétérogénéité clinique et moléculaire de la pathologie. En effet, le cancer du sein n'est pas une pathologie unique : l'étude anapathologique révèle que la dénomination dénote en réalité une vingtaine de sous-catégories carcinologiques. Le plus fréquemment, l'initiation et le développement des lésions mammaires mastosiques ou cancéreuses prennent origine au niveau des cellules épithéliales qui tapissent les canaux lactifères et les unités terminales ducto-lobulaires, fonctionnelles durant la lactation. Ainsi que pour tout type de lésion épithéliale cancéreuse, sont distingués les cancers in-situ des cancers infiltrants. Les carcinomes in-situ se caractérisent par une prolifération cellulaire sans franchissement de la membrane basale. Dans le cas des cancers in-situ, le tissu conjonctif avoisinant le site original de la dégénérescence n'est ainsi pas touché et les métastases inexistantes. Le cancer in-situ est généralement associé à un stade précoce d'un cancer en devenir invasif. Toutefois, le cancer lobulaire in-situ ou néoplasie lobulaire in-situ est plus régulièrement considéré comme un facteur de risque à développer un cancer du sein plutôt qu'un précurseur de ses précurseurs directs. Les cancers infiltrants se distinguent par une prolifération des cellules épithéliales malignes en dehors du site original par diffusion progressive dans les tissus voisins. En conséquence, sont distinguées les thérapies dîtes locales des thérapies systémiques. Les thérapies locales tentent de prémunir contre la progression et la propagation des cancers précoces, en opposition des procédures systémiques, telle que la mastectomie, dont le recours survient généralement dans le traitement des maladies micro ou macro métastatiques. L'étude rapportée dans ce manuscrit intervient dans le cadre des thérapies locales et plus précisément pour la chirurgie conservatrice du sein.

La chirurgie conservatrice mammaire consiste en le retrait du tissu cancéreux tout en évitant une ablation totale de l'organe. Ce type de chirurgie suivie par une radiothérapie est la procédure standard adoptée pour soigner les cancers de premiers stades. En 2005, la conférence internationale de Milan sur un consensus autour de la chirurgie conservatrice du sein a défini cette dernière comme « [...] le retrait total de la lésion maline du sein entourée d'une marge périphérique saine, réalisé de manière cosmétique dans la mesure du possible et suivi d'une radiothérapie ». Le succès de cette procédure chirurgicale étant ainsi dicté par la précision avec laquelle les marges latérales d'exérèse sont définies. Bien qu'il n'existe aucune définition générique de ce qu'est une marge idéale, il est commun de considérer une marge périphérique saine de 1 cm. Les marges sont par la suite qualifiées de positives ou de négatives au regard de la présence ou de l'absence de cellules cancéreuses en périphérie. Afin de dresser ce constat, les tissus excisés sont successivement fixés dans des solutions de formaline, intégrés dans des blocs de paraffine, coupés en de fines sections de quelques microns, immergés dans différents bains à différentes concentrations d'alcool et finalement colorés avec de l'hématoxyline et de l'éosine. L'hématoxyline est un marqueur cationique, colorant les éléments basophiles et particulièrement les noyaux cellu-

laires. L'éosine de par sa nature anionique teint les éléments cellulaires chargés positivement tels que la matrice extracellulaire et les cytoplasmes. Actuellement, à la suite d'une inspection anapathologique, jusqu'à 20 % des volumes excisés présentent des marges positives. Les raisons de l'échec de la délimitation des bords de la tumeur sont souvent dues à un prélèvement aléatoire pour la biopsie, à la propagation discontinue de la tumeur à partir du site de la maladie d'origine ou à une excision inappropriée pendant l'opération.

Dans le but d'aborder la question de la netteté des marges, ce travail fait état d'enquêtes sur une fenêtre spécifique du spectre électromagnétique, la bande térahertz, en vue de déterminer son potentiel de démarcation entre les tissus mammaires sains et malins. Il a été rapporté que les radiations térahertz sont particulièrement sensibles à la teneur en eau. En effet, l'eau présente des mécanismes de relaxation dans la bande térahertz à environ quelques dizaines de gigahertz et 900 GHz, respectivement. Notez que ces valeurs peuvent légèrement varier d'une étude à l'autre. Simultanément, l'eau est le fluide prédominant chez les organismes vivants. L'unification des concepts autour de la cancérogenèse suggère que l'hydratation des cellules pourrait être en première ligne. Tous ces rapports avancent fortement que les ondes térahertz peuvent aider les chirurgiens à définir avec précision les marges des tumeurs du sein en per-opératoire, en se basant à la fois sur les radiations térahertz et les spécificités des tumeurs. Les ondes térahertz couvrent des énergies de photons qui sont des ordres de grandeur plus petits (0,4-40 meV) que le spectre visuel. En outre, la faible énergie transportée par ces ondes indique qu'elles n'ont pas d'effet ionisant et sont globalement considérées comme faiblement dangereuses pour les cellules, les organes et les organismes.

Ce manuscrit tente de répondre à des questions fondamentales sur la capacité de la bande térahertz à gérer la démarcation entre les tissus mammaires anormaux et normaux. Le titre de ce travail de thèse devrait être interrogatif plutôt que simplement affirmatif : *Les radiations térahertz pour la reconnaissance des tumeurs du sein ?* Tout au long de ce manuscrit, différentes stratégies et méthodologies sont ensuite examinées, décrites, appliquées et discutées pour tenter de répondre à une telle question.

Le premier chapitre *Contexte et concepts clés*, établit la physique qui s'avère utile pour étudier l'interaction des ondes électromagnétiques dans la gamme submillimétrique avec les entités biologiques, après quoi les mécanismes d'interaction d'un point de vue sub-cellulaire sont discutés. À cet égard, les candidats les plus prometteurs pour l'interaction sont présentés et un aperçu du domaine de recherche est fourni. La discussion porte ensuite sur la réactivité au niveau supercellulaire aux stimuli térahertz, où les principaux résultats et l'application biomédicale sont énumérés, de manière aussi exhaustive que possible. Enfin, le chapitre se termine sur une section consacrée au cas spécifique du carcinome du sein, de l'examen de l'anatomie du sein au traitement du cancer, en passant par la cancérogenèse.

Le deuxième chapitre *Spectroscopie térahertz* commence par présenter les différents concepts et techniques de génération et de détection des ondes térahertz. Ensuite, le dispositif expérimental utilisé pour mener les recherches sur des échantillons biologiques est présenté. Un tel système permet de mesurer la fraction de l'énergie électromagnétique qui est soit transmise à travers l'échantillon examiné, soit réfléchie par celui-ci. Par conséquent, des concepts importants pour extraire les propriétés diélectriques intrinsèques des différents types de tissus présents dans le volume des seins sont abordés. Certaines contributions spécifiques au domaine sont rapportées et le concept de problème électromagnétique inverse est décrit car il est omniprésent tout au long de ce travail. Ces outils sont ensuite utilisés pour extraire le profil diélectrique de l'eau et les relaxations associées sont abordées. Comme l'environnement réel du sein est complexe et certainement pas homogène, des études préliminaires sont menées sur des fantômes de tissus imitant le sein. Ces échantillons visent à fournir des informations génériques sur le comportement théorique des tissus réels, en imitant la composition chimique réelle. Des tendances spécifiques sont signalées, qui tendent à suggérer que le térahertz est globalement sensible à la fraction de protéines, de graisses et d'eau présente dans les tissus. Dans une tentative de fournir une formulation générique pour décrire les profils diélectriques globaux, les fondements d'un modèle mathématique sont discutés.

Puis, sur la base des caractéristiques présentées par des fantômes de tissus imitant le sein, une campagne de mesures sur des tissus mammaires fraîchement excisés, menée à l'Institut Bergonié (Bordeaux, FRANCE) est rapportée. Les profils diélectriques globaux des tissus adipeux, fibreux et malins sont rapportés. Des preuves évidentes de la démarcation entre le tissu adipeux et d'autres types de tissus sont signalées. Des changements plus subtils des constantes optiques sont également soulignés entre les tissus fibreux sains et les tissus malins. Enfin, pour comprendre la source du contraste entre les types de tissus, des recherches sur les différentes concentrations de cellules cancéreuses en solution sont signalées et certaines explications possibles du contraste entre les tissus sont avancées.

Le troisième chapitre *Terahertz Imaging : Far-field Survey* est consacré à l'acquisition d'images térahertz de tissus mammaires fraîchement excisés en champ lointain. Avant tout, les performances de l'installation d'imagerie sont identifiées et explicitées. Ensuite, une procédure d'inversion d'image par auto-référence est décrite, qui vise à améliorer la caractérisation de l'échantillon. A partir des images brutes, du processus décrit et du problème électromagnétique inverse, la carte de l'indice de réfraction des différents tissus est extraite. L'indice de réfraction dérivé sur la surface de l'échantillon est ensuite utilisé pour segmenter l'image en térahertz. L'enregistrement des images et la corrélation avec les images de pathologie sont détaillés. Par la suite, différentes stratégies de classification sont étudiées et leurs performances relatives en termes de sensibilité et de spécificité sont rapportées. Une carte de conformité entre les images de pathologie et les classifications térahertz est fournie. Enfin, pour compléter le champ d'investigation, l'analyse en composantes principales, c'est-à-dire une analyse de données non supervisée, est étudiée comme un outil supplémentaire potentiel vers la segmentation. Les résultats rapportés sur l'analyse en composantes principales sont en bon accord avec la classification de l'indice de réfraction.

Le dernier chapitre *Breast Cancer Imaging : Near-field Survey* fait état du développement d'une matrice d'imagerie à l'état solide en champ proche de type térahertz fonctionnant à une fréquence identifiée dans les chapitres précédents comme étant prometteuse pour la différenciation des tissus. Le chapitre commence par souligner la limite de diffraction de la résolution qui affecte gravement les systèmes d'imagerie en champ lointain. L'état de l'art dans le domaine de la recherche en champ proche est abordé, puis le principe de fonctionnement conceptuel d'un imageur en champ proche à sous-longueur d'onde sensible à la permittivité, basé sur un résonateur, est décrit. Un tel dispositif vise à briser la limite de résolution de la diffraction et à fournir une résolution d'image qui est globalement à l'échelle de la taille typique des cellules eucaryotes. Une batterie de simulations est rapportée qui soutient l'idée qu'un tel dispositif pourrait effectivement briser la limite de diffraction tout en restant sensible au comportement diélectrique des tissus. La tâche s'annonce cependant difficile car on ne prévoit que des changements subtils en réponse à un tel imageur entre les personnes en bonne santé et les personnes atteintes de cancer. Le processus de fabrication et l'architecture globale du réseau à semi-conducteurs sont décrits et ses performances intrinsèques en matière d'imagerie sont fournies. Des études préliminaires avec le capteur d'imagerie en champ proche de sous-longueur d'onde sont rapportées et une image d'une section de sein déparaffinée conclut ce manuscrit.

Les travaux rapportés dans ce manuscrit participent à des projets de collaboration : les projets *Far and Near-sense*, respectivement financés par la Région Nouvelle-Aquitaine et la Fondation allemande pour la recherche dans le cadre du programme prioritaire ESSENCE (SPP 1857). Trois entités ont participé à ces projets communs et doivent être reconnues. Des études sur les mécanismes d'interaction entre les radiations térahertz et les tissus biologiques ont été menées à l'Université de Bordeaux (FRANCE) sous la supervision du Dr. Rrch. Patrick Mounaix (DR). Le développement de l'imageur en champ proche a été abordé par l'Université de Wuppertal (ALLEMAGNE) sous la direction du Prof. Dr. rer. nat. Ullrich Pfeiffer. Enfin, le contexte médical de l'étude a été assuré par le département de pathologie de l'Institut Bergonié (Bordeaux, FRANCE) sous la direction du Dr. Gaëtan MacGrogan.