TERAHERTZ TIME-DOMAIN SPECTROSCOPY AND IMAGING OF ARCHIVAL DOCUMENTS

Tiphaine BARDON

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor in Engineering

at

University College London

UCL EngD Centre in Virtual Environments, Imaging and Visualisation (VEIV)
UCL Institute for Sustainable Heritage
University College London

May 20, 2016
I, Tiphaine BARDON, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.
Abstract

This project explores the scope of the cutting-edge terahertz time-domain (THz-TD) spectroscopy and imaging as a potentially non-invasive diagnostic and imaging tool for the study of archival documents. While some previous research successfully uncovered drawings or texts from inaccessible layers in heritage artefacts using THz-TD imaging, inscriptions in some historical documents investigated in this project could not be revealed with this technology. This work therefore explores the parameters influencing contrast in terahertz images of historical documents, in order to objectively inform curators when terahertz technologies are likely to give well-contrasted images of inscriptions in a complex archival document and are a useful non-invasive alternative to current digitization processes. These parameters include the dispersive refractive indices and absorption coefficients of various archival materials, the physical structure of documents, as well as the choice of signal parameters and terahertz imaging systems or configurations used to produce the terahertz image. This investigation was conducted systematically using historically informed models and historical documents, with different THz-TD set-ups and accessories, and complementary analytical and imaging tools. THz-TD imaging was found to be a useful non-invasive digitization technology for documents containing lamp black, carbon black or graphite but not for documents containing bistre, sepia, verdigris or Prussian blue. While deconvolution of the terahertz signal helps in distinguishing between layers in a complex document, it can also decrease the image contrast of inscriptions. The results further reveal the potential and limitations of the use of THz-TD spectroscopy for material identification in archival collections. Minium, vermilion and corrosive iron gall inks with excess iron(II) sulfate have a characteristic spectral signature below 3 THz, which can be extracted from the surface of an inked paper or parchment sheet. The project was developed in collaboration with TeraView Ltd, Cambridge, and Nationaal Archief (Netherlands).
It is a great folly to wish to be wise alone, La Rochefoucauld.

I would like to testify my gratitude to the many people who have helped me through the completion of this thesis. The first is my principal supervisor, Prof. Matija Strlič, who has not only helped me blossom into a researcher, but also served as a mentor, thanks to his continued curiosity, enthusiasm, diplomacy and diligence, together with his useful feedbacks and judgements. I am grateful for the meetings I had with my second supervisor, Prof. Sir Michael Pepper, during which he shared his insight into terahertz science and electromagnetism. I would like to thank my industrial partner, TeraView Ltd., for enabling me to have access to their facilities, software tools and protocols, with my sincere appreciation going to the following staff members, Drs. Phil Taday, Rob May, Alessia Portieri and Jesse Alton, who were keen on supporting me with the data acquisition and interpretation. I am particularly grateful for the helpful feedbacks and literature that Drs. Taday and May provided during the revision of draft articles, posters and chapters. I was blessed to be in partnership with the Dutch National Archives, which financially supported this project and was eager to discover the potential of terahertz time-domain technology for the analysis and imaging of archives. I am particularly thankful to Gerrit de Bruin and Gabriëlle Beentjes, who raised my awareness on curatorial practices and interests, and shared relevant reports, images and samples. Finally, I cannot thank enough the UCL EngD Centre in Virtual Environments, Imaging and Visualisation (VEIV), for awarding me with their only studentship for international students.

My work has been influenced by the continued support provided by Dr Bianca Jackson, from the one-month internship I performed under her supervision in April - May 2012 at the Centre de Recherche et de Restauration des Musees de France (C2RMF), to the day-long experiments we performed at the Laboratoire de Recherche
Acknowledgements

des Monuments Historiques (LRMH) and at the University of Reading. Thanks to Dr. Michel Menu, C2RMF, the internship enabled me to have access to a time-domain set-up developed by Picometrix, Advanced Photonics, Inc. Access to this set-up was once again possible with the invaluable help of David Giovannacci, LRMH. Thanks to Dr John Bowen, University of Reading, access to a continuous waves terahertz set-up in transmission was also made possible. Drs. John Bowen and Gillian Walker should be specially thanked here, as they shared their invaluable knowledge on terahertz science and its application to cultural heritage. Prof Timothy Korter and his PhD student, Mark Ruggiero, computational chemists at Syracuse University, USA, showed a remarkable and repeated interest in this research project and I am grateful for the useful publications they produced, as it shined light into my results and broadened the readership of my research. Martin Vickers and Dr. Jeremy Cockcroft, UCL Department of Chemistry, also showed a keen interest in this research and performed Powder X-ray Diffraction on my samples. I was lucky to meet Dr Jennifer Griffiths, UCL Department of Medical Physics and Bioengineering, who instantly offered her help in acquiring X-ray images of my model ink dots with a micro-computed tomography system. She further made much efforts into contacting colleagues to have access to a clinical mammography system and a phase-contrast imaging system. I also had the chance to assist Dr. Dafydd Griffiths, UCL Institute of Archaeology, in acquiring scanning electron microscope images of my ink dots and energy dispersion X-ray spectra from pigments and inks. I am also indebted to Prof. Adrian Podoleanu and Dr Adrian Bradu, School of Physical Sciences at the University of Kent, for introducing me to their Optical Coherence Tomographic system, and for acquiring some images of my model ink dots. I would also like to thank Dr. Mark Underhill, technician at the UCL Institute for Sustainable Heritage, for his guidance in the use of laboratory equipments and in laboratory safety. Dr. Lik-Ren Tai, technician at the UCL Institute for Sustainable Heritage, provided helpful Fourier-transform infrared spectra from pigments to validate my preliminary results.

A mosaic of professional opportunities and discussions made me the researcher I am today. I was blessed to work at the UCL Institute for Sustainable Heritage with Prof. Cassar, Dr. Grau-Bové, Dr. Curran, Dr. Fouseki, and all the past and present PhD and MRes students.
I would also like to thank the VEIV doctoral centre, the UCL Grant museum of zoology, Dr. Antoniou (University of Peloponnese, Greece) and my co-workers Andrew Barnes and Dane Virk, for the opportunity to work on the exciting project *Micro-augmentations: situated calibration of a novel non-tactile, peripheral museum technology*. I am grateful to UCL, The Bartlett School of Graduate Studies and The Centre for Advancing Learning and Teaching, for giving me the opportunity to obtain a national recognition in teaching.

Despite much efforts, a research project often spills over personal life, so I would like to thank the many people who helped me stay (relatively) sane. The first is my husband, Anthony, whose love, humour, pragmatism and experience in research were daily solicited. I would also like to thank my mother, my grandmother, my family, Alaa and Céline, who were angels on earth during the writing up of this thesis, together with the PhD students from the department who contributed to making these 4 years unforgettable.
Contents

1 Introduction 29

2 Background 35
   2.1 Textual archives: conservation and research . . . . . . . . . . . . . . . 35
      2.1.1 Morphology of textual archives . . . . . . . . . . . . . . . . 35
         2.1.1.1 Supports . . . . . . . . . . . . . . . . . . . . . . 35
         2.1.1.2 Inks . . . . . . . . . . . . . . . . . . . . . . . . 38
      2.1.2 Depth-imaging technologies applied to textual archives . . . . . 41
      2.1.3 Degradation processes in textual archives and their analytical
           study . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
           2.1.3.1 Iron gall ink corrosion and crystal formation . . . . 43
           2.1.3.2 Denaturation of collagen in parchment . . . . . . . 44
      2.2 Terahertz science and technology . . . . . . . . . . . . . . . . . . . . 47
         2.2.1 Terahertz radiation and interaction with matter . . . . . . . . 47
         2.2.2 Broadband radiation . . . . . . . . . . . . . . . . . . . . . 55
            2.2.2.1 Spectroscopy . . . . . . . . . . . . . . . . . . . . . 55
            2.2.2.2 Imaging . . . . . . . . . . . . . . . . . . . . . . . 60
         2.2.3 Narrowband radiation . . . . . . . . . . . . . . . . . . . . . 65
      2.3 Terahertz applications in cultural heritage . . . . . . . . . . . . . . . 65
         2.3.1 Pioneers . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
         2.3.2 Past and current research . . . . . . . . . . . . . . . . . . . 67
         2.3.3 Latest trends . . . . . . . . . . . . . . . . . . . . . . . . . 73

3 Materials and instrumentation 75
   3.1 Terahertz pulsed technologies . . . . . . . . . . . . . . . . . . . . . . . 75
3.1.1 Terahertz pulse generation and detection .................................. 75
3.1.2 Terahertz time-domain spectroscopy ........................................... 79
  3.1.2.1 Spectroscopy in transmission mode ........................................ 79
  3.1.2.2 Software ............................................................................. 79
  3.1.2.3 Attenuated Total Reflectance (ATR) spectroscopy ..................... 80
3.1.3 Terahertz time-domain imaging ..................................................... 82
  3.1.3.1 TPS spectra 3000 equipped with a Reflection Imaging Module (RIM) ............................................................................. 82
  3.1.3.2 TPS spectra 3000 with a gantry system ..................................... 83
  3.1.3.3 Software ............................................................................. 83
3.1.4 Time-domain imaging with T-Ray 4000 (Picometrix) ......................... 84
  3.1.4.1 Software ............................................................................. 86
3.2 Samples ......................................................................................... 86
  3.2.1 Pigments and chemicals .............................................................. 86
    3.2.1.1 Carbon-based black ............................................................ 87
    3.2.1.2 Coloured pigments .............................................................. 89
    3.2.1.3 Binders ............................................................................. 91
    3.2.1.4 Iron gall inks .................................................................... 91
    3.2.1.5 Preparation ....................................................................... 93
  3.2.2 Historical paper and parchment sheets .......................................... 94
4 Metrology in terahertz time-domain spectroscopy and imaging ................. 95
  4.1 Terahertz time-domain spectroscopy ............................................... 95
    4.1.1 Beam size ............................................................................ 95
    4.1.2 Pellet preparation .................................................................. 96
    4.1.3 Calculation of optical parameters .......................................... 97
      4.1.3.1 Absorbance .................................................................. 98
      4.1.3.2 Refractive index .............................................................. 98
      4.1.3.3 Absorption coefficient .................................................... 101
    4.1.4 Determination of optical parameters ...................................... 103
      4.1.4.1 Precision .................................................................... 103
      4.1.4.2 Accuracy .................................................................... 113
4.1.4.3 Sample concentration ................................................. 121
4.1.4.4 Dynamic range ....................................................... 123
4.1.5 Apodization ............................................................... 124
4.1.6 Conclusions ............................................................... 127
4.2 Terahertz time-domain imaging ......................................... 128
4.2.1 Lateral spatial resolution .............................................. 128
4.2.2 Signal-to-noise ratio .................................................... 143
4.2.3 Detectability .............................................................. 143
4.2.4 Time measurement uncertainty ...................................... 146
4.2.5 Conclusions ............................................................... 148

5 Terahertz time-domain imaging of model and historical documents 151
5.1 Depth-imaging of a letter within an envelope .......................... 151
5.2 Aim of the terahertz Time-Domain (TD) imaging study ............... 155
5.3 Surface pulse amplitude imaging ........................................ 155
  5.3.1 Influence of the constant refractive index .......................... 155
    5.3.1.1 Measurement of the constant refractive index .............. 155
    5.3.1.2 Influence of pulse overlap on the maximum amplitude of
            the waveform: the need for deconvolution .................. 159
    5.3.1.3 Influence of double Gaussian filter high frequency
            cut-off on image contrast .................................. 162
    5.3.1.4 Influence of time subsampling on image contrast ............ 164
    5.3.1.5 Normalisation of images ...................................... 166
    5.3.1.6 Influence of step size ......................................... 167
    5.3.1.7 Comparison of terahertz image contrasts ..................... 169
    5.3.1.8 Repeatability .................................................... 172
    5.3.1.9 Comparison between different set-ups ........................ 172
  5.3.2 Influence of indentations and other topographic features ......... 174
    5.3.2.1 Model indentations on paper and parchment ................. 174
    5.3.2.2 Sieve lines ..................................................... 179
    5.3.2.3 Indentations on historical documents ........................ 182
5.4 Depth pulse amplitude imaging ........................................ 187
5.4.1 Ink dots in verso configuration .......................... 187
5.4.2 Inscriptions on verso of a parchment book cover .......... 192
5.4.3 Overpainting: the case of an Ethiopian codex ............... 196

5.5 Spectroscopic imaging ........................................ 201
5.5.1 Dispersive behaviour of optical parameters .................. 201
5.5.2 Influence of topographic features and support thickness on the
frequency-domain signal ........................................... 204
5.5.3 Extraction of spectral information from an ink inscription ... 206

5.6 Future work and trends ........................................ 215
5.6.1 Image processing tools to improve contrast .................. 215
5.6.2 Continuous Wave imaging ................................... 216
5.6.3 Algorithm to remove occlusions in complex artefacts .......... 218
5.6.4 Algorithm to extract the pulse reflected from the layer of inter-
est in complex layered systems ................................... 218
5.6.5 Expanding the library of terahertz optical properties to modern
archival materials ..................................................... 219
5.6.6 X-ray attenuation and phase contrast imaging ............... 223
5.6.7 Diversity of supplies and round robin tests ................... 226

6 Terahertz time-domain spectroscopy of supports and iron gall inks 227
6.1 Terahertz time-domain spectroscopy of supports .............. 227
6.2 Terahertz time-domain spectroscopy of iron gall inks .......... 234
6.3 Conclusions ....................................................... 242

7 Conclusions ......................................................... 243
7.1 Contribution to new knowledge ................................ 243
7.2 Impact: current and future ................................... 245

Appendices .......................................................... 253

A Complementary characterization and imaging 255
A.1 Fourier-transform infrared spectroscopy ...................... 255
A.2 Powder X-ray diffraction ...................................... 260
A.3 X-ray fluorescence (XRF) spectroscopy ....................... 264
## Contents

A.4 Scanning electron microscopy with energy dispersive X-ray spectroscopy 265
A.5 Scanning electron microscopy 266
A.6 3D microscopy 266
A.7 Optical coherence tomography 267
A.8 Profilometry 267
A.9 X-ray imaging 267

B Phase unwrapping algorithm 269
C Deconvolution algorithm 273
D Algorithm to extract peak amplitude from a layer of interest 275
E Algorithm for image normalisation 279
F Raw terahertz images from ink dots 283
G OCT and 3D images of model ink dots 287
H Near-infrared spectra from model and historical ink inscriptions 289
I Principal Component Analysis of historical parchment optical constants 291
J Terahertz spectral analysis of model inks in ATR mode 299

Bibliography 313
List of Figures

1.1 Examples of archival holdings ............................................. 32
2.1 Scheme of the fibrous structure of cellulose ................................36
2.2 Schematic diagram of the fibrous structure of collagen .................38
2.3 Schematic diagram of the structure of the iron gall ink complex ..........40
2.4 Scanning electron micrographs from parchments ..........................46
2.5 Dispersive behaviour of the refractive index and extinction coefficient 
as calculated using the Lorentz or the Drude model .......................51
2.6 Characteristic refractive index and absorption coefficient values for dif-
ferent families of solid materials ..............................................53
2.7 Schematic diagram representing the propagation of an ultrashort tera-
hertz pulse impinging on a slab surrounded by air ........................58
2.8 Schematic diagram describing terahertz time-domain imaging data col-
lected in reflection mode from two overlaid sheets of paper ..............61
2.9 Schematic diagram describing terahertz time-domain imaging data col-
lected in transmission mode from two overlaid sheets of paper ...........63
2.10 Schematic diagram describing terahertz time-domain imaging data col-
lected in reflection mode from two overlaid sheets of paper, with in-
scriptions written on both sheets ............................................64
2.11 Number of outputs related to terahertz and heritage or to terahertz and 
art conservation published each year within the latest 20 years ..........67
3.1 Schematic diagram of a photoconductive antenna ..........................77
3.2 Image and schematic diagram of the TPS Spectra 3000, TeraView Ltd ........80
3.3 Image of the attenuated total reflectance module by TeraView Ltd ....81
3.4 Images of the reflection imaging module by TeraView Ltd ..............83
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>Image of the gantry system by TeraView Ltd.</td>
</tr>
<tr>
<td>3.6</td>
<td>Image of the T-Ray 4000 by Picometrix</td>
</tr>
<tr>
<td>4.1</td>
<td>Terahertz time-domain waveforms and absorbance spectra of the signal transmitted through an iris diaphragm with different aperture sizes</td>
</tr>
<tr>
<td>4.2</td>
<td>Schematic diagram detailing the propagation of a terahertz beam through a sample with parallel faces</td>
</tr>
<tr>
<td>4.3</td>
<td>Terahertz time-domain waveforms transmitted through the empty sample compartment and through a pellet of bistre diluted in HDPE</td>
</tr>
<tr>
<td>4.4</td>
<td>Frequency-dependent refractive index and absorption coefficient from a single pellet of HDPE and bone black, for 6 subsequent signal acquisitions</td>
</tr>
<tr>
<td>4.5</td>
<td>Frequency-dependent refractive indices and absorption coefficients from pellets of sucrose powders with different grain sizes</td>
</tr>
<tr>
<td>4.6</td>
<td>Microscope and SEM images of the surface of Prussian blue pellets with different grinding pellets</td>
</tr>
<tr>
<td>4.7</td>
<td>Frequency-dependent absorption coefficients and refractive indices of Prussian blue pellets</td>
</tr>
<tr>
<td>4.8</td>
<td>Influence of the phase unwrapping process on the frequency-dependent refractive index</td>
</tr>
<tr>
<td>4.9</td>
<td>Triplicate refractive index measurements from a solid sheet of pure HDPE and pellets of HDPE</td>
</tr>
<tr>
<td>4.10</td>
<td>Terahertz time-domain waveforms with and without etalon effect and their associated frequency-dependent refractive index and absorption coefficient</td>
</tr>
<tr>
<td>4.11</td>
<td>Terahertz time-domain waveforms from a historical parchment sheet before and after deconvolution</td>
</tr>
<tr>
<td>4.12</td>
<td>Average absorbance spectra from triplicate pellets for bone black and iron sulfate heptahydrate, with different sample-to-HDPE mass ratios</td>
</tr>
<tr>
<td>4.13</td>
<td>Dynamic range of the TPS Spectra 3000, TeraView, in transmission</td>
</tr>
<tr>
<td>4.14</td>
<td>Original and apodized TD waveforms transmitted through a pellet made of 10% Prussian blue</td>
</tr>
</tbody>
</table>
4.15 Absorbance spectra from 2 sets of triplicate pellets made of 10% and 1% Prussian blue ...................................................... 127
4.16 Gaussian functions representing the 1D PSF from two point sources separated by different distances .......................... 130
4.17 ERF, LSF and MTF calculated from the maximum amplitude of the terahertz electromagnetic pulse reflected from an edge and imaged with the T-Ray 4000, Picometrix ................................................. 132
4.18 Terahertz image of the target obtained with the T-Ray 4000 set-up, Picometrix, from the maximum amplitude of the pulse reflected .......................................................... 133
4.19 Profile of the normalized maximum amplitude of the reflected pulse, using the TRay 4000, Picometrix .................................................. 134
4.20 Terahertz images of the USAF 1951 target produced with the maximum amplitude of the main reflected pulse, using the RIM ........................ 136
4.21 Profile of the normalized maximum amplitude of the reflected pulse, using the RIM ............................................................... 137
4.22 Terahertz images of the USAF 1951 target produced with the maximum amplitude of the main reflected pulse, using the gantry system ........ 137
4.23 Profiles of the normalized maximum amplitude of the reflected pulse, using the gantry system .................................................. 138
4.24 Profiles of the normalized maximum amplitude of the reflected pulse and the Frequency-Domain (FD) signal at 2 THz and 2.58 THz, using the RIM .................................................. 139
4.25 Terahertz images of the USAF 1951 target produced with the amplitude of the signal at 2 THz and 2.58 THz, using the RIM with different scanning step sizes ........................................ 141
4.26 Profiles of the amplitude of the FD signal at 2 THz and 2.58 THz, using the RIM ................................................................. 142
4.27 Terahertz spectroscopy in transmission of a stack of rag paper sheets with an increasing number of sheets ........................ 142
4.28 Terahertz images of the USAF 1951 target, alone and under 2 sheets of rag paper, using the gantry system ................................. 142
4.29 Dynamic ranges of the T-Ray 4000, Picometrix and the TPS Spectra 3000, TeraView, in reflection mode ........................................ 144
4.30 Schematic diagram highlighting the influence of several parameters on the resolution and detectability of object features .................. 145
4.31 Schematic diagram highlighting the influence of the size of the scanning step size with respect to the width of the ink line on the image contrast .......................................................... 146
4.32 Photograph of a letter typewritten with Prussian blue ink and associated terahertz images with a scanning step size of 0.2 mm and 0.075 mm .... 147
4.33 Terahertz images produced using the time of collection of the maximum amplitude of the waveforms reflected from a flat gold coated mirror, scanned with the RIM. a) Waveforms constituted of 4096 time points and b) 1024 time points .................................................. 148
5.1 Terahertz images from graphite inscriptions on the recto and verso of a document, obtained with raw terahertz signals .................... 152
5.2 Terahertz images of graphite inscriptions in a closed envelope .......... 154
5.3 Schematics showing the influence of pulse overlap on the maximum amplitude of the waveform .............................................. 160
5.4 Terahertz images of a vermilion and a minium ink dot, using different values of the high frequency cut-off of the double Gaussian filter in deconvolution of time-domain signals .................................. 163
5.5 Terahertz images of a vermilion ink dot produced with and without time subsampling ....................................................... 165
5.6 Terahertz image of a vermilion ink dot overlapped with a schematic diagram describing the boundaries used for the image normalisation and calculation of the contrast value .................................. 167
5.7 Terahertz images of vermilion and minium ink dots on rag paper, with different step sizes ................................................. 168
5.8 Terahertz images of all ink dots on original rag paper ................... 169
5.9 Terahertz images of all ink dots on additionally sized rag paper ........ 170
5.10 Terahertz images of all ink dots on parchment .......................... 171
List of Figures

21

5.11 Terahertz images of bistre and vermilion ink dots applied in triplicates
on rag paper, additionally sized rag paper and parchment . . . . . . . . 173
5.12 Terahertz images of ink dots on rag paper and additionally sized rag
paper, acquired using T-Ray 4000 . . . . . . . . . . . . . . . . . . . . 174
5.13 Terahertz images of surface indentations left by writing ”UCL” with a
clean nib pen on a sheet of print paper and a sheet of parchment, using
the raw signals . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 177
5.14 Terahertz images of surface indentations left by writing ”UCL” with a
clean nib pen on a sheet of print paper, using deconvolved signals . . . 178
5.15 Terahertz images of a sieve line at the surface of a rag paper sheet from
the 19th century . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 181
5.16 FTIR spectra from typewritten letters sent in July 1946 . . . . . . . . . 183
5.17 Terahertz images of the signature typewritten on letters sent by the
same person in July 1946, produced using the maximum amplitude of
the deconvolved waveform . . . . . . . . . . . . . . . . . . . . . . . . 185
5.18 Terahertz images of the signature typewritten on letters sent by the
same person in July 1946, produced using the maximum amplitude of
the raw waveform . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 186
5.19 Terahertz images of an iron gall ink dot scanned in recto and verso
configurations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 188
5.20 Terahertz images of the front and back surfaces of a letter from July
1946 written with a lead pencil . . . . . . . . . . . . . . . . . . . . . . 191
5.21 Photographs of historical book covers, schematic diagram explaining
the succession of layers used for the model study, and layout of the
model book cover during a scan acquisition . . . . . . . . . . . . . . . 193
5.22 Terahertz images of inscriptions written on the back surface of a parchment book cover . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 195
5.23 Terahertz images of a page with text in an Ethiopian codex . . . . . . . 197
5.24 Terahertz images of a page with a miniature in an Ethiopian codex . . . 199
5.25 NIR photograph and terahertz images of a miniature in an Ethiopian
codex . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 200


5.26 Frequency-dependent refractive indices from different dry pigments, inks and binders .................................................. 202
5.27 Frequency-dependent absorption coefficients from different dry pigments, inks and binders .................................................. 203
5.28 Terahertz images in the frequency-domain from indentations and sieve lines .............................................................. 205
5.29 Terahertz time-domain B-scans along a line across an ink dot on rag paper and associated raw time-domain waveforms ............... 207
5.30 Tail of the time-domain waveforms reflected from minium and vermilion ink dots on different supports ........................................ 208
5.31 Tail of the time-domain waveforms transmitted through minium and vermilion pellets ......................................................... 209
5.32 Time-domain and frequency-domain terahertz B-scans along a line crossing a minium and a vermilion ink dot on rag paper .............. 210
5.33 Time-domain and frequency-domain signals collected from a single sheet of support with and without a vermilion ink dot on its surface, using different set-ups and configurations .......................... 214
5.34 Terahertz image of the signature typewritten on letters sent in July 1946, processed using different image segmentation algorithms .... 216
5.35 Refractive indices and absorption coefficients from thin sheets of historical paper and Prussian blue pigment powders acquired from different suppliers ........................................................... 217
5.36 Terahertz images of an Islamic manuscript ........................................... 220
5.37 Near-infrared and ATR-FTIR spectra from the surface of an Islamic manuscript ................................................................. 222
5.38 X-ray absorption and phase-contrast images of carbon-based ink dots on parchment ....................................................... 225

6.1 Terahertz absorbance spectra from historic parchments and from historic and contemporary papers ........................................ 228
6.2 Terahertz absorbance spectra from historic parchments, sorted according to the thickness of each sheet ........................................ 228
6.3 Absorbance spectra and terahertz time-domain waveforms from different areas of a historic parchment sheet ........................................ 230
6.4 Constant refractive index and absorption coefficient from contemporary and historic paper sheets and parchment sheets ................. 233
6.5 Terahertz absorbance spectra of historically informed iron gall inks and of their individual constituents ............................. 235
6.6 Terahertz absorbance spectra of iron gall inks prepared with different molar and mass ratios of iron(II) sulfate to acid ...................... 236
6.7 Constant refractive indices and absorption coefficients of historically informed iron gall inks ............................................. 238
6.8 Frequency-dependent refractive indices and absorption coefficients of historically informed iron gall inks ........................... 239
6.9 Absorbance spectra from historic iron gall inks, acquired in ATR mode . 241
A.1 FTIR spectrum of carbon black, graphite and lamp black powders .... 256
A.2 FTIR spectrum of bone black ................................................. 256
A.3 FTIR spectrum of ivory black ................................................ 256
A.4 FTIR spectrum of bistre ........................................................ 257
A.5 FTIR spectrum of malachite .................................................... 257
A.6 FTIR spectrum of lapis lazuli .................................................. 257
A.7 FTIR spectrum of vermilion .................................................... 258
A.8 FTIR spectrum of minium ........................................................ 258
A.9 FTIR spectrum of verdigris ..................................................... 258
A.10 FTIR spectra of pigments in the Prussian blue family ................... 259
A.11 FTIR spectrum of Arabic gum ................................................ 259
A.12 FTIR spectrum of rabbit skin glue ........................................ 259
A.13 Powder X-ray diffractogram of Arabic gum ............................. 260
A.14 Powder X-ray diffractogram of gallic acid ............................. 261
A.15 Powder X-ray diffractogram of iron(II) sulfate ........................ 262
A.16 Powder X-ray diffractogram of copper sulfate ......................... 263
A.17 Powder X-ray diffractogram of tannic acid ............................ 264
List of Figures

F.1 Terahertz images of ink dots on rag paper, obtained using the raw waveforms ........................................ 283
F.2 Terahertz images of ink dots on additionally sized rag paper, obtained using the raw waveforms .................. 284
F.3 Terahertz images of ink dots on parchment, obtained using the raw waveforms ..................................... 285
F.4 Terahertz images of three different vermilion and bistre ink dots applied on different supports, obtained using the raw waveforms .......................................................... 285
G.1 OCT and 3D images of bistre and iron gall ink dots on rag paper and parchment ........................................ 287
H.1 NIR spectra of iron gall ink inscriptions on historic documents and of bistre, bone black, lamp black and sepia mixed with Arabic gum and applied on rag paper .................................................. 289
I.1 Photographs, FTIR spectra and material properties of historic parchment OR81 ........................................ 292
I.2 Photographs, FTIR spectra and material properties of historic parchment OR94 ........................................ 293
I.3 Biplots displaying the results of the principal component analysis on the set of parchment properties .............. 294
I.4 Photographs from five historic parchment samples before (top) and after (bottom) thermal degradation .......... 295
I.5 Terahertz absorbance spectra and associated time-domain waveforms from five historic parchment samples before and after thermal degradation ......................................................... 297
J.1 Triplicate ATR-THz spectra from three different ink dots made of 6 µL of iron gall ink models applied on parchment .......................................................... 300
J.2 Triplicate ATR-THz spectra from an ink dot made of 2 µL of IG3 ink applied on rag paper, filter paper and parchment .......................................................... 301
J.3 Triplicate ATR-THz spectra from iron gall ink dots made of 2µL and 6 µL applied on parchment and rag paper .......................................................... 301
List of Tables

2.1 Non-invasive depth-imaging technologies applied to historical documents 42

2.2 Summary of the published research works related to terahertz spectroscopy of textual documents 69

2.3 Summary of the published research works related to terahertz imaging of textual documents 70

3.1 Compositions of iron gall inks 93

4.1 Constant refractive index from a single pellet of HDPE and bone black, for 6 subsequent signal acquisitions 104

4.2 Average constant refractive index, slope of the absorption coefficient, observed time delay and measured difference in thickness between the sample (Prussian blue or sucrose) and reference pellets 108

4.3 Refractive index values reported in the literature for solid HDPE samples 116

4.4 Weights of sample powder added to 1440 mg of HDPE powder to obtain different sets of triplicate pellets with different sample-to-HDPE mass ratios 121

4.5 List of the average constant refractive indices from triplicate pellets of bone black and iron(II) sulfate heptahydrate with different sample-to-HDPE mass ratios 123

4.6 Summary of the lateral spatial resolution for different terahertz TD imaging set-ups and accessories 138

5.1 Constant refractive indices of different pigments, binders and supports 156

5.2 Contrast values extracted from terahertz images of different ink dots on different supports 168
5.3 Constant refractive indices from thin sheets of historical paper and
Prussian blue pigment powders acquired from different suppliers . . . . 182

A.1 Atomic percentage ratios of elements present in ivory black and bone
black pigment powders . . . . . . . . . . . . . . . . . . . . . . . . . . 265

I.1 Optical constants from five historic parchment samples before and after
thermal degradation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 296
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATR</td>
<td>Attenuated Total Reflectance</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>CW</td>
<td>Continuous Wave</td>
</tr>
<tr>
<td>ERF</td>
<td>Edge Response Function</td>
</tr>
<tr>
<td>EDS</td>
<td>Energy-Dispersive X-ray Spectroscopy</td>
</tr>
<tr>
<td>FD</td>
<td>Frequency-Domain</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier-Transform Infrared</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full Width at Half Maximum</td>
</tr>
<tr>
<td>HDPE</td>
<td>High Density Polyethylene</td>
</tr>
<tr>
<td>ISH</td>
<td>UCL Institute for Sustainable Heritage</td>
</tr>
<tr>
<td>LSF</td>
<td>Line Spread Function</td>
</tr>
<tr>
<td>MTF</td>
<td>Modulation Transfer Function</td>
</tr>
<tr>
<td>NIR</td>
<td>Near Infrared</td>
</tr>
<tr>
<td>OCT</td>
<td>Optical Coherence Tomography</td>
</tr>
<tr>
<td>PSF</td>
<td>Point Spread Function</td>
</tr>
<tr>
<td>PXRD</td>
<td>Powder X-Ray Diffraction</td>
</tr>
<tr>
<td>RIM</td>
<td>Reflection Imaging Module</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning Electron Microscope</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>TD</td>
<td>Time-Domain</td>
</tr>
<tr>
<td>THz-TDS</td>
<td>Terahertz Time-Domain Spectroscopy</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Archives are collections of records. These records may be on forms or supports, from microfilms, photographs, CD-ROMs, DVDs, audiotapes, as well as ink on paper, parchment and many other less commonly used supports. We will designate records constituted of ink on support as **textual archives**.

The Dutch Public Record Act (1995) [1] clearly states that the twofold legal obligation of authorities responsible of archival collections is "*to ensure that the records kept by them are in good condition, [...] and accessible*". Similarly, the National Archives (national archive for England, Wales and the United Kingdom) abide by the Public Records Act 1958 [2] and must safeguard all records selected for permanent preservation, but also abide by the Freedom of Information Act [3], which guarantees everyone the legal right to consult the information held by them.

Citizens should have free access to governmental records to ensure that any administrative or business operation follows agreed procedures, to look for documentation supporting a legal case, to search for genealogical evidence and learn more about their personal history, or more broadly, to carry out historical research.

The drive to ensure and improve free accessibility to governmental archival records has gained momentum with the advent of the digital era. Digitisation of records makes the archival information available remotely to a global public, with the possibility to add interactive tools (zooming options, embedded metadata) enriching the user’s experience. It also gives the opportunity to showcase the extent, richness and uniqueness of a collection to a broad audience. It can be argued that enabling access to a digital version of a record could aid the conservation of the original document, as it would reduce the number of users consulting the original and therefore reduce the wear
and tear due to handling. It has however been observed that access to a digital version has often triggered requests to consult the non-digitised documents related to the record digitally available, which may increase handling and degradation of those other documents [4]. What is certain, however, is that these digital versions of archival records represent valuable surrogates\(^1\) in case of disastrous events (fire, flood). Aware of the benefits of digitisation, the Dutch National Archives (DNA) have been asked “to identify which ten percent of the total national collections is of most interest to the public and should be made available in digital form over the next ten years” [5], and the digitisation process of the selected items has started in 2015. The selection of only 10% of the total archival records to be digitised is related to the extensive cost, time and staff that the digitisation process currently requires. The European Commission beckons EU member states to stimulate the digitisation and online availability of the collections from their libraries, archives, museums and audiovisual archives, and has set mass digitisation and online accessibility on the Europeana platform [6] as one of the challenges of the Digital Agenda for Europe [7]. Yet the Commission is aware of the obstacles to digitisation mentioned above and further acknowledges the lack of synergy between cultural and technology programmes to facilitate digitisation.

Current digitisation of textual archives is mostly performed by overhead cameras, scanners with a flat scanning bed or planetary book scanners [8, 9]. While these technologies are adequate for scanning flat loose sheets in a non-invasive way, their use to digitise bound documents would require the document to be opened at 180° and maintained flat, often clamping it and placing it under a glass window: this causes mechanical stress to the document and increases risks of damage. Such scanning procedures may also produce distorted images of the pages. Additionally, the operator would need to manually turn each page of the document to scan a complete book. In some cases, the decision to unbind or cut the bound document to flatten it to facilitate digitisation might be taken [10]. Scanning robots have now been developed to overcome these issues: the bound book is placed on a V-shape cradle and opened at a low angle, pages

\(^1\)The definition of "valuable surrogate" is subjective: David Arnold (University of Brighton) rightly pointed out in a lecture that "Surrogates need not be complete copies - they need only be fit for purpose in the role"; and this role, be it ensuring the legacy to future generations of the totality of the textual content or physical structure of the document with as much details as possible, or only revealing a few words (e.g. a signature, for authentication) with a resolution just sufficient to make them legible, would need to be evaluated first by the conservator before digitisation.
are turned by air flow and a prism head scans the pages of the open book [11]. These scanning robots open the way for fast mass digitisation of bound documents, including fragile books with pages at risk of mechanical damage if manually handled.

But the most optimal digitisation option adapted to very fragile bound or closed documents (i.e. books or letters) would be an in-depth scanning tool which could produce an image of each page of the entire document in one single non-invasive scan while the document remains closed. According to a study performed by Havermans et al. [12], items considered to be ”in bad condition, [of which] consultation will certainly shorten its durability [and of which] preservation is seriously threatened”, are categorized as unfit for use (“category 3” of the Consultation Profile), and represents 2,494 m of records out of the 69,572 m examined at the DNA in February 1999 (consisting of 44.5% of the total DNA holdings)\(^1\). A non-invasive in-depth scanning technology could potentially enable access to inaccessible archival content in records which have been altered in the past, such as records with pages glued together, painted over or with text scraped off to re-use the expensive parchment material. Another type of records which would benefit from this alternative technology could be folded documents, the folding pattern of which constitutes historical information in itself, as shown in Figure 1.1c, and which would benefit from remaining closed.

In parallel to the search for a tool adapted to the digitisation of documents ”in bad condition”, we should also postpone the time when archival holdings reach the state of ”unfit for purpose” (where ”purpose” could be physical consultation, digitisation with commonly used technologies, display). The research agenda of archival institutions therefore also includes identification, understanding and monitoring of degradation mechanisms in archival collections and slowing down their rate [13, 14].

The DNA have the desire to follow the action plan set by the European Commission and plainly fulfil their legal role as national archival repositories by setting goals for the future, aiming at ensuring access to the totality of the archival content in their collection, and one day, have it all available to the online audience. They therefore actively investigate whether the above described digitisation technologies exist or could realistically be developed and implemented in their conservation practice. The DNA also actively promote research on degradation in archival collections [15, 16].

---

\(^1\)Information provided by the Dutch National Archives
Terahertz time-domain technology first emerged in the mid-1990s [17], but its use is now increasing through its commercialization. TeraView Ltd is one of the leading companies designing and manufacturing technologies involving terahertz radiation (0.1 - 10 THz). Their goal is to provide terahertz set-ups and data processing adapted for use in industrial, security and medical applications. They have expertise in using terahertz TD technology for non-invasive depth-imaging of laminar samples such as pharmaceutical tablets and car paint coatings [18], and have collaborated with the UCL Institute for Sustainable Heritage (ISH) in a study of historic plastics [19]. This study illustrated a twofold potential of terahertz TD technology for non-invasive characterisation of heritage artifacts, as it provided information on the chemical composition of the plastic sample, through the use of Terahertz Time-Domain Spectroscopy (THz-TDS), and produced images of the plastic bulk revealing the presence of bubbles, cracks and delamination, through the use of terahertz TD imaging.

Figure 1.1: a) A document from the Dutch National Archives, The Hague, showing extensive iron gall ink corrosion: the document is exposed to a risk of significant loss of textual and structural information when manually turning its pages; b) A document from the Dutch National Archives, The Hague, Verenigde Oostindische Compagnie (VOC). The parchment used for the book cover is originally an older manuscript containing a map. Although this parchment cover has detached from the board, other parchment covers still glued to the board might also contain inscriptions on their verso; c) Unopened letter from the collection of Dutch sailors’ letters stored at The National Archives in Kew, England. The paper of the letter itself was folded in a characteristic way, so that no envelope was needed. To preserve this characteristic folding, none of these letters have ever been opened and have been kept in the same condition as they were found. Credit: Gabrielle Beentjes, Dutch National Archives.
Aim
The aim of this doctoral project is to explore the interactions between pulsed terahertz radiation and materials found in textual archives, in order to determine whether and how terahertz TD technology could be used by archival institutions as a non-invasive characterisation tool for preventive conservation and as an alternative digitisation tool for fragile and/or complex documents. The research approach adopted is not "pure applied research”, but rather "use-inspired basic research”, described by Stokes [20] as Pasteur’s quadrant. It does not focus on a few case studies from archival collections, but is based on a systematic study of many model and historical documents, and contributes to both the understanding and the application of the technology and the science behind it.

Objectives
The project had the following objectives:

• To measure absorbance spectra, refractive indices and absorption coefficients from pure archival materials and explore their accuracy and precision.

• To relate these optical values to the observed image contrast from model and historical documents and provide guidance on optimisation of terahertz scanning conditions and data processing to ensure successful characterisation and imaging of inscriptions in complex documents.

• To relate these optical values to the composition of model and historical iron gall inks and assess whether terahertz TD technology could be used to determine the corrosive nature of an ink inscription.

• To relate these optical values to the composition and structure of historical paper and parchment sheets and assess whether terahertz TD technology could be used to determine the degradation state of such supports in collections.
Chapter 2

Background

2.1 Textual archives: conservation and research

2.1.1 Morphology of textual archives

While historical inscriptions engraved on tablets and stones reporting decrees, such as the famous Rosetta stone, can be considered as textual archives, the most common forms of textual records found in archival collections constitute printed, painted or handwritten information in the form of letters, leaflets, books, scrolls or loose sheets. The nature and composition of the materials constituting such typical textual archives will be explored below, together with their development through history.

2.1.1.1 Supports

Paper

Paper is prepared from cellulosic fibrous materials such as hemp, cotton or linen rags, or wood fibres diluted with water and beaten in a vat, before being poured onto a grid typically made of brass wires fixed on a wooden frame (although in Asia, the first grid moulds were made of a coarsely woven cloth, and, later on, of woven silk, laced bamboo stripes or laced grass). The mould is then shaken horizontally to remove the excess of water and distribute the fibres. The wet sheet of paper is transferred to a piece of felt or cloth, is then flattened under a press and finally cut to the required size [21, 22].
Figure 2.1: Scheme of the fibrous structure of cellulose in plants used for paper production, adapted from [26] and reused with permission of the author.

The fibres contain macrofibrils, with a highly ordered internal arrangement [23] as shown in Figure 2.1. A macrofibril contains several microfibrils, mainly constituted of cellulose polymer chains. Each chain consists of glucose monomers held together along the chain by ether covalent bonds and transversally cross-linked by hydrogen bonds. Water molecules can infiltrate between chains and microfibrils. Both crystalline and amorphous arrangements of cellulosic chains can be found in a microfibril, with crystalline regions being defined as regions possessing a 3-dimensional order sufficient to give a regular repeating system of hydrogen bonds [24]. Paper is made of a 3-dimensional network of randomly arranged fibres, with voids (or pores) with a radius of 30 to 50 µm [25].

In the early 18th century, with the increased production of printed books and newspapers, the demand for paper increased and other cellulosic materials such as wood, esparto and straw were used as well as rags [22]. Paper made from wooden pulp contains a natural polymer called lignin which exacerbates the photo-yellowing of paper [27].

To prevent ink from spreading on the paper, most paper sheets are sized. From 700
AD, sizing of rag paper consisted of dipping the paper sheet into a solution of gelatine and alum, or starch. From the early 19th century, after the invention of paper-making machines, sizing was made by adding a diluted solution of rosin (resin from pinewood) to the cellulose pulp and precipitating it onto the fibres by addition of alum [22]. In contemporary paper production, methylcellulose, polyvinyl alcohol and styrene acrylic copolymers are used as sizing agents [28].

**Parchment**

Parchment is prepared from wet unhaired skin soaked in a lime solution (Ca(OH)$_2$), before being dried and thinned at ambient temperature. During this process, it is constantly maintained under tension on a stretching frame [29]. Skins of small animals such as calves, sheep and goats were often preferred for writing. Parchment is constituted of collagen fibres with a very ordered and compact structure [30, 31], as illustrated in Figure 2.2. Each fibre is formed by fibrils and microfibrils held together by intermolecular bonds. Each microfibril is made of transverse collagen molecules. Each collagen molecule represents a triple-helical structure of polypeptide chains: each helical polypeptide chain constitutes a regular succession of amino-acids connected by carbon-nitrogen peptide bonds, and the three helical chains are held together via hydrogen bonds and water molecules. Collagen molecules connect to other triple-helical structures (i.e. other collagen molecules) via aldol cross links between two carboxylic ends of side-chains.

After the removal of hair from the limed skin, the drying stretched skin is rubbed with a pumice stone to smoothen its surface. Powdered chalk (CaCO$_3$), calcium sulfate dihydrate (CaSO$_4$, 2H$_2$O) and quicklime (calcium oxide, CaO) can be applied on the surface of the parchment to remove the excess of grease and moisture, to improve the adhesion of ink on the fatty and porous parchment skin and to prevent it from running [29]. These calcium salts can also neutralize the acidity caused later on by ink or by the ambient environment. The above steps are essential stages of the preparation of parchment, but further details and variations are described by Reed [29].
2.1.1.2 Inks

Historic black inks

*Carbon black and bistre inks*

The black colour of carbon black inks comes from the dispersion of soot particles from burnt organic materials in a mixture of water and binder [33]. Soot for carbon black inks mainly came from burnt oil (lamp black pigment) and burnt bones or ivory (bone black or ivory black pigment). A brown ink called bistre ink was also made from soot particles, obtained from carbonized vegetable materials like wood. A commonly used binder was Arabic gum, a resin mainly obtained from *Acacia arabica* Willd.

Carbon black inks are known to be chemically inert. Yet, such ink penetrates into fibres of the support only to a limited extent, so that scraping or wiping of the surface of the document with a wet cloth can remove the ink. These inks can be found in ancient
Egypt or Greek and Roman documents [33], and were used in the production of Dead Sea scrolls [34]. They were also largely used in Asia (e.g. Chinese ink was obtained from the soot of pine wood and resin, but also from the soot of oil).

Although also used in Europe in the medieval period, their low penetration into the support led to the preferential use of iron gall inks for the production of European archival documents from the 13th century onwards [33].

**Iron gall inks**

Iron gall inks were produced from the reaction of iron(II) with gallic acid moieties in tannin resulting in a dark blue complex upon oxidation, which degrades into brown-black over time [35]. The source of iron(II) was vitriol (mainly FeSO₄), the purity of which varied largely and would often contain copper sulfate and traces of other metal salts. The source of gallic acid was crushed gall nuts (products of oak trees under the influence of insects) either fermented in cold water or cooked in hot water [33, 35]. Because iron gall inks are solutions and not a dispersion of pigments, and because they are so chemically reactive, they adhere firmly to the support surface and are water resistant. The complex structure formed in iron gall inks and the chemical processes leading to the formation of the dark colour of the inks are still under debate.

Due to the large variety of ink recipes and possible environmental conditions during preparation, storage and even analysis of the ink, several structures of the complex and chemical reaction paths might exist. A common feature of all the suggested structures of iron gall ink is that it has a highly ordered arrangement. Krekel [36] suggested the structure shown in Figure 2.3 (left): Fe(II) iron ions from the iron sulfate are oxidized into Fe(III) and two of these iron ions are bound to phenol groups of two gallic acid (GA) moieties to form a 1:1 GA:Fe complex. Wunderlich et al. [37] suggested Fe(III) - O bonding through both carboxylic and phenol groups of the gallic acid moieties, leading to the formation of a 1:2 GA:Fe complex, as illustrated in Figure 2.3 (right).
Figure 2.3: Schematic diagram of the structure of the iron gall ink complex suggested by Krekel [36] (left, image reused with permission of the author) and by Wunderlich et al. [37] (right).

**Sepia ink**

Sepia ink was used very early and is found in Greek and Roman documents [33]. It is produced by cuttlefish and owes its colour to melanin, a natural polymer. It can be used alone or mixed with other inks. The brownish colour of sepia ink and bistre ink (made from the soot of burnt plant materials) often makes them difficult to visually discern from degraded iron gall ink.

**Historic coloured inks**

Red pigments and dyes commonly found in manuscripts are minium (Pb₃O₄ or 2PbO; PbO₂), vermilion -also called cinnabar, (HgS)- cochineal (containing carmine acid), Brasillian wood, red madder and red clays (containing iron oxides). The most often used blue pigments and dyes were lapis-lazuli ((Na,Ca)₈(AlSiO₄)₆(S, SO₄, Cl)₁₋₂) for ultramarine blue, azurite (Cu₃(CO₃)₂(OH)₂), indigo and copper sulfates [38, 39]. The green colour was either obtained from natural copper minerals such as malachite, or from the reaction of copper with vinegar to obtain copper acetate -also known as verdigris- or finally by mixing orpiment with a blue mineral, such as ultramarine or indigo [39].
Contemporary inks

The compositions of contemporary inks are not very well referenced or disclosed by the producers. Still, black pigments in modern printing inks are often the same as in ancient inks [40]: type-writing machines commonly use carbon-based inks [41] and toners for laser printers contain carbon black particles and iron oxides (mars black). Other dry dyes and inks historically and currently used for printers, typewriters and fountain pens are, amongst others, aniline, nigrosine, methylene blue, and Prussian blue [41].

2.1.2 Depth-imaging technologies applied to textual archives

A variety of non-invasive depth-imaging technologies have been used for the analysis of painted cultural heritage artifacts, as reported by Janssens et al. [42], and can also be applied to sub-surface analysis of textual archives, as described in Table 2.1.

Synchrotron light sources are often used in these imaging experiments in order to penetrate deeper into matter, reach better signal-to-noise ratios and sometimes reach better lateral spatial resolution to see finer detail, due to the highly focused and intense beam used. Yet access to such facilities is limited, often demands a long waiting time and requires documents to travel.

2.1.3 Degradation processes in textual archives and their analytical study

Apart from ensuring access to archival records, the Dutch Public Record Act also states that national archives authorities should keep their records in "good condition". It is the role of conservators to manage degradation, its impact and to treat the affected records.

A variety of degradation processes affect archival collections, threatening the survival of information. These diverse degradation processes are very well described and illustrated in a book published by Metamorfoze, the Netherlands National Programme for the Preservation of Paper Heritage [54]. While damage to binding and text block as well as mechanical damages due to handling have been briefly mentioned before, attention will be placed here to a few specific chemical degradation processes:
Non-invasive depth-imaging technologies applied to historical documents.

### Table 2.1: Non-invasive depth-imaging technologies applied to historical documents

<table>
<thead>
<tr>
<th>Technology</th>
<th>Advantages</th>
<th>Drawbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray attenuation imaging combined with CT</td>
<td>Sub-micrometer lateral resolution. Can map the distribution of specific elements.</td>
<td>Requires access to a synchrotron; not portable.</td>
</tr>
<tr>
<td>Edge-illumination X-ray Phase Contrast</td>
<td>Image contrast based on phase difference influenced by changes in refractive index and thickness of the ink layer.</td>
<td>Requires a high level of expertise in computer science and is time-consuming.</td>
</tr>
<tr>
<td>X-ray fluorescence imaging</td>
<td>Can be portable.</td>
<td>Requires access to a synchrotron; not portable.</td>
</tr>
<tr>
<td>Near-infrared reflectography</td>
<td>Portable.</td>
<td>Requires access to a synchrotron; not portable.</td>
</tr>
<tr>
<td>Optical Coherence Tomography (OCT)</td>
<td>Ease of access and often affordable.</td>
<td>Requires access to a synchrotron; not portable.</td>
</tr>
</tbody>
</table>

Examples from the literature:

- Figure adapted from [43] and reused with permission from the author. The thickness of the ink layer can be estimated, but the assessment of the state of degradation of the parchment underneath using this image requires a high level of expertise in computer science and is time-consuming.
- Figure adapted from [47] and reused from the author. Archeologists at Ein Gedi, Israel, revealing verses from the Book of Leviticus written with an ink containing metal.
- Image contrast is based on the increased X-ray attenuation with the presence of a thin layer of metal (such as gold).
- Ionisation of the object which may induce further damage (although a lower exposure dose can be obtained using a fast acquisition time).
- Drawbacks of in situ ionisation include highly energetic beam which may induce further damage (although a lower exposure dose can be obtained using a fast acquisition time). A mobile instrument for X-ray micro-CT can be used.
- Drawbacks of portable X-ray micro-CT include portable but not movable. The synchrotron source is not movable, and archival repository cannot be moved.

### Table 2.2: Conditions for image quality

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sufficient Energy</td>
<td>The escape depth from which fluorescent X-rays can be detected is limited (in the 10^(2) - 10^(3) range for parchment).</td>
</tr>
<tr>
<td>Depth of Penetration</td>
<td>High penetration depth, yet sufficient to reveal many underdrawings.</td>
</tr>
<tr>
<td>Access to Object</td>
<td>Ease of access and often affordable.</td>
</tr>
<tr>
<td>Amount of Data</td>
<td>Instantaneous, no scanning. How ever, to reach a good lateral resolution, a large amount of data is needed.</td>
</tr>
</tbody>
</table>

Note on Near-infrared-based X-ray-based imaging:

In near-infrared-based X-ray-based imaging, the penetration depth is lower than 300 \(\mu\)m (range of 100 - 500 \(\mu\)m). A mobile instrument for X-ray micro-CT can be used.

Critical aspects of the application of non-invasive depth-imaging technologies to historical documents include:

- The potential for damage to the document.
- The need for high-resolution images.
- The requirement for portable equipment.
- The need for a high level of expertise in computer science.

For a full list of references, please refer to the literature cited in the text.
2.1. Textual archives: conservation and research

2.1.3.1 Iron gall ink corrosion and crystal formation

**Mechanism**

Iron gall inks corrode the support they are written on, particularly paper. Yet, “the mechanisms that cause ink corrosion are only partly understood. It is striking that in the case of identical objects, such as copies, one object will be affected by ink corrosion while the other remains undamaged” [54]. What is currently known is that it causes cellulose depolymerization in paper through two processes:

- Acidic iron gall inks (pH ≃ 2) catalyze the hydrolysis of the ether bonds between the monomers of cellulosic chains [35].

- Iron gall inks also catalyze the oxidation of cellulose, via a Fenton-type reaction: the free metal ions react with $\text{H}_2\text{O}_2$ (resulting from the recombination of $\text{O}_2^-$, product of the oxidation of terminal aldehyde units in cellulose by molecular oxygen) or organic hydroperoxides of cellulose, which produces a highly reactive oxidant, hydroxyl radical: $\text{HO}^-$ [35, 55].

The consequences of the corrosive nature of iron gall ink are dark-coloured degradation products, or even loss of degraded support in and around the inscriptions (see Figure 1.1a), rendering the documents difficult to handle and read.

**Methods of identification and monitoring**

Identification and monitoring of the chemical processes of iron gall ink corrosion are directly related to studies of the composition and chemical processes occurring within the ink itself. For example, Ferrer *et al.* observed the formation of white crystals on ink lines indicating on-going corrosion of the support underneath and a low pH of 3.5, while these crystals were not observed in historical ink inscriptions with no signs of corrosion [56]. Fourier-Transform Infrared (FTIR) spectroscopy enabled them to identify these crystals as iron sulfate. La Camera *et al.* also observed the formation of crystals on the surface of some historical inscriptions and drawings where iron gall ink had been densely applied and they also identified these crystals as sulfate, mostly iron sulfate, using X-ray fluorescence, scanning electron microscopy with electron dispersive spectroscopy, FTIR spectroscopy and Raman spectroscopy. They further reported
that crystals seem to preferentially form when iron sulfate is introduced in excess to gallic acid and Arabic gum during the preparation of the ink [57, 58].

X-ray diffraction is a crucial tool for the analysis and identification of crystalline structures, and it has been used by Wunderlich [37] and Feller et al. [59] on experimental iron gall inks prepared with different processes, using only iron(III) chloride (Fe(III)Cl$_3$) and gallic acid, and (Fe(II)Cl$_2$), gallic acid and KOH, respectively. The results of both X-Ray diffraction experiments were identical: the 3D structure and coordination formula determined by Feller et al. with the aid of a software tool was the same as determined by Wunderlich et al. (Figure 2.3 left), and consisted of a crystalline iron(III) complex. These same crystallographic and X-ray diffraction data were further used by Zaccaron et al. [60] to estimate the electronic structure of different iron gall complexes with the aid of Density Functional Theory (discussed in the next Section). They found that the most probable (i.e. energetically favourable) structure is a crystalline iron(II) complex. Most importantly, their study showed that “the oxidizing potential of iron gall inks is low. [...] the direct oxidation of substrates of interest in the field of cultural heritage by iron gall inks is an improbable pathway for the commonly observed corrosion reactions”. This would support the idea of a Fenton-type reaction triggered by free (i.e. not bound to gallic acid) iron ions.

These free iron ions can be easily and routinely probed thanks to the use of iron(II) indicator strips containing bathophenanthroline, which forms an intensely magenta-coloured complex in the presence of free iron(II) ions within an ink line. However, as its name suggests, it only probes free iron(II), not iron(III) ions. Besides, this test requires the strip to be humid and applied in direct contact with the ink line: this is therefore an invasive test which may cause further degradation of ink.

2.1.3.2 Denaturation of collagen in parchment

Mechanism

As explained earlier, parchment is made of collagen. A molecule of collagen has the form of a triple helix of polypeptide chains held together by intermolecular hydrogen bonds. If those bonds are lost, the helical structure collapses and the ordered molecule of collagen transforms into disordered gelatine. This phenomenon of gelatinisation is observed in the presence of water molecules, which would enable the formation
of new hydrogen bonds leading to their adsorption on a polypeptide chain [61]. It is also observed when exposing parchment to temperatures high enough to break the hydrogen bonds [62]. Additionally, it was further explained above that amino-acids in a polypeptide chain are connected by carbon-nitrogen peptide bonds. Oxidation can break those bonds, and is triggered by free radicals formed in the process of splitting a water molecule into $\text{H}^+$ and $\text{OH}^-$ under the influence of UV radiation or by atmospheric pollutants in a humid environment. This latter can also cause acidic hydrolysis, responsible for dissociation of the ketone bound within an amino-acid into a part with a carboxylic end and a part with a methyl end. Iron gall ink can also participate in both oxidation and hydrolysis of parchment. The result of these two phenomena is fragmentation of collagen and loss of long-range order [61]. It should be noted that different mechanisms or different combinations of these mechanisms can be found in different areas of the same parchment sheet [63].

Gelatinisation, oxidation or hydrolysis induce the loss of ordered structure of the collagen molecular arrangement, and allow for collagen molecules and, more generally, collagen fibrils, to move and form new conformations, with the possible formation of new cross-links [61]. These conformational changes cause parchment to loose its elasticity and mechanical strength: if no cross-links are formed, parchment becomes gel-like and glassy, while if cross-links are formed between fragments of collagen molecules and fibrils, parchment becomes rigid and brittle [64, 65]. “However, within one collection there may be parchments with similar levels of damage but different stability and resistance against ageing” [66].

**Methods of identification and monitoring**

The *IDAP* (Improved Damage Assessment of Parchment) project [67] investigated both qualitatively and quantitatively the mechanisms of deterioration of parchment. To do so, a variety of analytical tools was used, such as visual analysis, Scanning Electron Microscope (SEM), FTIR spectroscopy, UV-Visible-Near Infrared (NIR) spectroscopy, Raman spectroscopy, unilateral- Nuclear Magnetic Resonance Spectroscopy (NMR) and X-ray scattering.
Figure 2.4: Scanning electron micrographs (300 000x) showing typical morphologies of fibres in different stages of degradation. Left: goat parchment (2005, shrinkage temperature $T_s = 69.1 \, ^\circ C$), middle: and right: historical parchments from the reference collection of the ISH (middle) 1853, $T_s = 49.9 \, ^\circ C$ and right) 1554, $T_s = 33.8 \, ^\circ C$). The scale bar denotes 200 nm. [68].

SEM proves to be particularly useful to highlight swelling, glassiness and loss of ordered structure in fibres of collagen resulting from natural or thermal denaturation of collagen [62, 68] (see Figure 2.4).

Small Angle X-ray Scattering (SAXS) studies on parchment associate the increase of diffuse scatter due to an increased proportion of non-crystalline material related to the amount of degraded collagen in parchment: the authors explain that collagen in historical parchment shows some crystallinity which is lost upon denaturation [69].

Finally, two thermal analysis methods exist to observe the thermal denaturation of collagen and assess the deterioration state of parchment: shrinkage measurements by the Micro Hot Table (MHT) method, which relates the deterioration state of parchment to the temperature $T_S$ at which the majority of fibres in a $\simeq 0.3$ mg sample of fibres placed in water shrink simultaneously and continuously, and Differential Scanning Calorimetry (DSC), which determines the energy necessary to unfold collagen (loss of fibril assembly and uncoiling of the triple helix) by measuring the temperature at which the heat flow from fibres in water decreases [66, 70]. Since these methods require sampling of parchment fibres and are therefore destructive, the potential of NIR spectroscopy combined with partial least square analysis (PLS) has been investigated as a tool to non-invasively assess the temperature at which the collagen fibres shrink (higher temperature indicating more stable fibres with less degradation), and replace the MHT method [15]. The model built with PLS to estimate $T_S$ showed an uncertainty of $\pm 7 \, ^\circ C$ (average error between the measured and estimated shrinkage temperature of samples not used in the calibration of the model), which could be considered too large.
if the intention is to monitor the state of degradation of a parchment document over time.

The recurrent features of materials commonly found in textual archives are crystallinity or long-range order in hierarchical structure, and its loss or loss of appearance upon ageing. The studies also confirm that there is still much to learn about the detailed processes of denaturation of collagen in parchment, the mechanisms of iron gall ink corrosion on paper or parchment, and the chemical structure of the ferro-gallic complex in iron gall ink. Among the analytical tools employed, none could directly probe and identify the weak intermolecular forces responsible for the cohesion in molecular crystals: accessing this information could reinforce the findings of some published studies or bring forth new evidence on the nature of archival materials and their degradation pathways.

### 2.2 Terahertz science and technology

#### 2.2.1 Terahertz radiation and interaction with matter

Terahertz radiation is an electromagnetic oscillation with frequencies ranging from 0.1 to 10 THz. Like all electromagnetic waves, its propagation through most materials is predominantly influenced by the interaction of its electric field with charged particles. A detailed description of the propagation of electromagnetic waves in matter is provided in a reference textbook by Griffiths [71] and lecture notes by Kuzel [72]. More recent textbooks, published between 2008 and 2013, further describe the propagation of terahertz radiation in matter [73–77]. The main conclusions from these works are compiled below in order to identify which materials are opaque, transparent or reflective in the terahertz region.

<table>
<thead>
<tr>
<th>Frequency (THz)</th>
<th>0.1</th>
<th>3</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (µm)</td>
<td>3000</td>
<td>100</td>
<td>60</td>
<td>30</td>
</tr>
<tr>
<td>Wavenumber (cm⁻¹)</td>
<td>3</td>
<td>100</td>
<td>167</td>
<td>334</td>
</tr>
<tr>
<td>Energy (meV)</td>
<td>0.4</td>
<td>12</td>
<td>21</td>
<td>41</td>
</tr>
</tbody>
</table>

In interaction with a solid material of a specific thickness in air, electromagnetic radiation is partially reflected and partially transmitted at the air-material and material-air interfaces. The response of a material to an electromagnetic wave depends on its
intrinsic optical parameters: the (real) refractive index $n$ and the extinction (or absorption) coefficient $\kappa$. The larger the real refractive index of the material, the more strongly the incident radiation will be reflected at the air-material interface. The larger the extinction coefficient of the material, the more strongly the amplitude of the propagating radiation will be attenuated. A material is opaque if no radiation is transmitted through it, and transparent if the incident radiation is barely affected by the presence of the material. An opaque material therefore either has a very large refractive index (most of the incident radiation is reflected), and/or a very large extinction coefficient (most of the incident radiation is absorbed), while a transparent material has both a low refractive index and a low extinction coefficient. While $n$ and $\kappa$ are referred to as optical constants, they are in reality dependent on the frequency of the incident electromagnetic wave. Studies of frequency-dependence of $n(\omega)$ and $\kappa(\omega)$ highlight spectral regions where a material is opaque or transparent to an electromagnetic wave.

An electromagnetic wave is composed of an oscillating electric field and an oscillating magnetic field. The response of a solid material to an external electric field can be distinguished in terms of the mobility of its electrons, i.e. its electrical conductivity. Therefore $n(\omega)$ and $\kappa(\omega)$ are related to electrical conductivity of the material.

- Conductors can be described by the Drude model, i.e. as having electrons unbound to atoms and free to move under the influence of the driving force $-e \cdot \vec{E}$ resulting from the applied electric field $\vec{E}$. The motion of these free electrons trapped in the metallic material, accelerated by an applied field ($\vec{F}_{\text{driving}}$) and slowed down by random collisions ($\vec{F}_{\text{damping}}$), can be approximated using the laws of classical mechanics:

$$m \cdot \frac{d^2 \vec{r}}{dt^2} = \vec{F}_{\text{driving}} + \vec{F}_{\text{damping}} = -e \cdot \vec{E} - m \cdot \gamma \cdot \frac{d\vec{r}}{dt} \quad (2.1)$$

where $m$ is the mass of electrons and $\gamma$ is the damping factor equal to the inverse of time between two collisions.

- Dielectrics (or insulators) have all their electrons bound to atoms, and the Lorentz model describes such bonds as a mass-less spring with a resonance frequency $\omega_0$. Some electrons can experience a displacement due to the applied electric field, but are restored to their original position by the binding force of the spring
(\vec{F}_{\text{binding}}). In a time-periodic electric field, the motion of the electrons can therefore be approximated using the harmonic oscillator model and the laws of classical mechanics:

\[
m \cdot \frac{d^2 \vec{r}}{dt^2} = \vec{F}_{\text{driving}} + \vec{F}_{\text{damping}} + \vec{F}_{\text{binding}} = -e \cdot \vec{E} - m \cdot \gamma \cdot \frac{d\vec{r}}{dt} - m \cdot \omega_0^2 \cdot \vec{r} \tag{2.2}
\]

For a harmonic oscillator, the motion of electrons varies as \( r = r_0 e^{-i \omega_0 t} \) with the same frequency \( \omega \) as the electric field, and the solution of Equation 2.2 shows that the displacement of valence electrons in a dielectric is a function of the electric field:

\[
\vec{r}(\omega) = \frac{-e \vec{E}}{m} = \frac{-e \vec{E}_0}{m} \cdot e^{-i \omega_0 t} \tag{2.3}
\]

In a conductor, \( \omega_0 \) is simply set to 0 in Equation 2.3.

The refractive index \( n \) of a material is a “response function […], a macroscopic manifestation of the microscopic response of matter to a periodic driving force” [78]. \( N \) electrons per unit volume with an electric charge \(-e\) and microscopic displacement \( r(\omega) \) forced by an external electric field \( \vec{E} \) induce macroscopic manifestations such as polarisation of matter \( \vec{P} = -N.e.\vec{r} \) in dielectrics, and current \( \vec{j} = -N.e.\frac{d\vec{r}}{dt} \) in conductors, linked to the applied electric field by the refractive index:

\[
\vec{P} = \epsilon_0 (n^2 - 1) \cdot \vec{E} \tag{2.4}
\]

\[
\vec{j} = i \cdot \omega \cdot \epsilon_0 (n^2 - 1) \cdot \vec{E} \tag{2.5}
\]

With \( \epsilon_0 \) the dielectric constant of vacuum (8.9 \( \cdot \) 10\(^{-12}\) F.m\(^{-1}\)). It should be noted that the response of a material to an electric field is not instantaneous, but happens after a certain time delay, which can be represented by a phase offset in the case of periodic fields and motions. As a result, the refractive index is better described by a complex number \( \tilde{n} \), with an amplitude and phase information.

\[
\tilde{n}(\omega)^2 = [n(\omega) + i \kappa(\omega)]^2 = 1 + \frac{N.e^2}{m \epsilon_0} \frac{1}{-\omega^2 - i\omega \cdot \gamma + \omega_0^2} \tag{2.6}
\]
Defining the plasma frequency $\omega_p$ such that:

$$\omega_p^2 = \frac{N \cdot e^2}{m \cdot \varepsilon_0} \quad (2.7)$$

it results that:

$$n(\omega)^2 - \kappa(\omega)^2 = 1 + \frac{\omega_p^2(\omega_0^2 - \omega^2)}{\omega_0^2 - \omega^2 + \gamma^2 \cdot \omega^2} \quad (2.8)$$

$$2 \cdot n(\omega) \cdot \kappa(\omega) = \frac{\omega_p^2 \cdot \gamma \cdot \omega}{\omega_0^2 - \omega^2 + \gamma^2 \cdot \omega^2} \quad (2.9)$$

The real refractive index $n(\omega)$ relates to the speed of propagation of the electromagnetic wave in the material and the reflectivity at the air-material interface, while the extinction coefficient $\kappa(\omega)$, imaginary part of the refractive index, relates to the attenuation of amplitude of the electric field during propagation through the material, as mentioned earlier.

For metals under the influence of a weakly energetic terahertz electric field (0.4 - 40 meV), only free electrons undergo a displacement. Metals can therefore be considered as perfect conductors in the terahertz region and are well described by the Drude model ($\omega_0$ is set to zero) [79]. Using this model, it was reported that for metals, $\omega_p \approx$609 - 3570 THz and $\gamma \approx$2.19 - 48.9 THz [79]. At 1 THz, $\omega < \gamma << \omega_p$: it results from Equations 2.8 and 2.9 that $n(\omega)^2 - \kappa(\omega)^2$ is negative and $2 \cdot n(\omega) \cdot \kappa(\omega)$ is very large, which can also be expressed as $n < \kappa$ and both $n$ and $\kappa$ are large: metals are therefore considered opaque in the terahertz region. For copper, $n(1 \text{ THz}) \approx$730, and 99.7% of the incident terahertz radiation is reflected, with the rest of the radiation penetrating the material over a distance of 66 nm before its field amplitude decays by $1/e$ [75]. In comparison, X-rays can penetrate up to 20 $\mu$m in copper [80, 81].

Statistically, intrinsic semiconductors have, at room temperature, few electrons in their conduction band: these free electrons can also be described by the Drude model and will interact with the terahertz electric field. Yet, the number of free electrons per unit volume $N$ is low compared to metals, and it follows that the plasma frequency of semiconductors lies in the microwave or infrared region [82]. In the terahertz range, $\omega > \omega_p$ for some intrinsic semiconductors. It results that the extinction coefficient and
the refractive index are low and intrinsic semi-conductors can therefore be considered transparent in the terahertz region.

Solid materials with no free charges are better described by the Lorentz model. Yet valence electrons bound to atoms in matter have a resonant frequency $\omega_0$ located in the ultraviolet or visible spectral ranges ($\approx 100 - 1000$ THz), i.e. not in the terahertz range. Between 0.1 - 10 THz, we are therefore in a spectral domain where $\omega << \omega_0$ (with $\gamma << \omega_0$ for dielectrics): it results from Equations 2.8 and 2.9 that $\text{Re}(\tilde{n}^2) = n(\omega)^2 - \kappa(\omega)^2 \approx 1 + \frac{\omega_0^2}{\omega^2}$ and $0 < \text{Im}(\tilde{n}^2) = 2n(\omega)\kappa(\omega) \approx \frac{\omega_0^2}{\omega}\gamma(\omega) << 1$. Using the expressions of $n$ and $\kappa$ as a function of $\text{Re}(\tilde{n}^2)$ and $\text{Im}(\tilde{n}^2)$ described in [75], it follows that $n(\omega \to 0) \to \sqrt{1 + \frac{\omega_0^2}{\omega^2}}$ and $\kappa(\omega \to 0) \to 0$: dielectrics have a low refractive index and a low extinction coefficient, and they are often described as "transparent" to terahertz radiation.

![Figure 2.5: Dispersion behaviour of the refractive index (solid line) and extinction coefficient (dashed line) as calculated using the Lorentz or the Drude model.](image)

The relative displacement between two isolated atoms or groups of atoms inducing rearrangement of the density of charges can be described as the vibration of charged particles and is influenced by an external electric field. This displacement resonates at a frequency $\omega_0$ in the mid-infrared range.

In a crystalline solid, although atoms or groups of atoms are arranged in a fixed structure, they are not static. It is as if the atoms were connected by mass-less springs. Yet, if one atom or group of atoms moves relatively to another, it forces its neighbours to move, and this relative displacement is passed on to the next atom until the static
boundaries of the material are reached. This propagation of local deformation can be described as a standing wave, or by a quasi-particle called phonon. If the movement of a positively charged atom is out of phase with the movement of its negatively charged neighbour, this creates an oscillating separation of charges which can interact with an electric field: this is an optical phonon. Under the influence of an oscillating electric field, all atoms in the crystalline network can start vibrating collectively at the same frequency as the applied field. Many of these collective vibrations resonate in the terahertz frequency range. It should be noted that many solid polymers can be crystalline and can therefore have resonance frequencies in the terahertz region: this is the case of α-lactose monohydrate [83], cellulose [84] or sucrose (see Chapter 4), to name a few.

Although the Lorentz model describes the vibration of an isolated system [negatively charged particle - positively charged particle] in response to an external electric field, it can be adapted to description of collective vibrations, using reduced mass \( \frac{1}{m} = \frac{1}{m_1} + \frac{1}{m_2} \) instead of the mass of an electron, with \( N \) in Equation 2.7 becoming the number of oscillators per unit volume. At the resonance frequency \( \omega_0 \), the amplitude of vibration of the Lorentzian oscillator increases significantly, which leads to a peak in the extinction coefficient, as illustrated in Figure 2.5 (left). The refractive index undergoes what is called an “anomalous” dispersion: it rapidly increases with frequency before the resonant frequency, and rapidly decreases afterwards. There is little absorption of terahertz radiation by crystalline solids resonating in the terahertz range, except within specific spectral windows around resonance frequencies where the material becomes highly absorbent and therefore opaque. The width of these spectral windows increases with the damping factor \( \gamma \) (Equation 2.2).

When an electromagnetic wave propagates through matter together with a phononic standing wave as was described above, the combination of these two waves is called a polariton wave [85]. Absorption or scattering of such a wave can occur at a specific frequency in the terahertz range and is associated with attenuation of the terahertz radiation electric field amplitude. This results in a broad peak in the extinction coefficient of the material, similar to the peak observed for resonant behaviour [86].

As a general rule, for a material with a low density, the number of charged particles per unit volume \( N \) is reduced: \( \omega_p \) is therefore smaller and so are the refractive index and the extinction coefficient.
Water is a polar molecule: it has partially positive and negative charges, and therefore possesses a permanent dipole moment. The direction of these dipole moments in liquid water is random due to the random orientation of water molecules with respect to one another. The application of an electric field would align all these dipoles along the direction of the field, and once the field is discontinued, the dipoles would recover their random orientations over a relaxation time of a few ps (\textit{i.e.} few THz). An electric field oscillating at terahertz frequencies will therefore strongly interact with these dipoles and, as in a resonant system, will have its amplitude strongly attenuated. Liquid water and, by extension, materials with a high water content (including air with a high relative humidity), are therefore considered highly absorbent / opaque to terahertz radiation.

The optical properties in the terahertz region of different families of solids mentioned above are summarized in Figure 2.6.

![Figure 2.6: Characteristic refractive index and absorption coefficient values for different families of solid materials. The position of metals is only indicative here (their refractive index and absorption coefficient extend to much larger values). Schematic diagram adapted from [73] and reused with permission of the author.](image)

Strong absorption due to the resonant behaviour of the crystal lattice or polariton absorption / scattering are characteristic of a unique crystalline arrangement of the material and can therefore be used for identification. In terahertz spectroscopy, the frequency dependence of the extinction coefficient is observed: if, in a sample of an unknown composition, strong extinction of the electric field is systematically observed at the same frequencies as for another pure specific compound, then it can be deduced that the latter compound is present in the sample.
But many other spectroscopic tools are already used for material identification: the novelty of terahertz spectroscopy is that the presence of strong absorption bands indicates that the material is crystalline, of a specific crystal structure allowing collective vibrations to resonate in the terahertz range. Linking the frequencies at which resonant behaviour is observed to the relative atomic displacements in the crystal would enable us to learn more about the material structure and some of its inter-molecular vibrations.

**Assignment of resonant frequencies**

In mid-infrared spectroscopy, the identification of the frequency at which the vibration of a covalent bond between two atoms resonates was at first largely experimental rather than theoretical. Several compounds with exactly the same covalent bond present in their molecular structure were analysed and the frequency at which an increased absorption of the field was systematically observed for all compounds was identified as the resonant frequency of this specific bond [87]. The same approach is, alas, not possible when it comes to assigning spectral features in terahertz spectra: the same motion of an atom or group of atoms (e.g. the rotation of sulfate ions) in two compounds with exactly the same composition but with different crystalline structures (also called polymorphs) or in two hydrates, will lead to different collective vibrations, as it will induce a different displacement of its neighbour atoms. Polymorphs and hydrates would therefore have different resonance behaviours and different spectral responses in the terahertz region. Since a polymorphic change would lead to some changes in interatomic distances, the force constants of some intra-molecular bonds would be modified, and so would the vibrational frequencies of these bonds in the mid-infrared region: polymorphs therefore have different mid-infrared spectral signatures, as well as different terahertz spectral signatures. Still, White and Roy [88] observe that if the polymorphic transformation involves only minor structural rearrangements, the mid-infrared spectra of the two polymorphs are very similar. As a result, identification of polymorphs with mid-infrared spectroscopy, although possible, may not be as easy as with terahertz spectroscopy. The drug compound ranitidine hydrochloride, for example, has two polymorphs with very similar mid-infrared spectra, differing by the presence of a weak band at 1045 cm$^{-1}$ in the spectrum of one polymorph [89], while their terahertz spectra are very clearly different [90].
Determination of the crystal structure and identification of collective vibrations leading to spectral absorption bands due to resonant behaviour requires the calculation of the total energy of the crystalline network for different atomic coordinates and different electronic distributions (or densities). The most stable and favourable crystal structure and electronic density is the one for which the total energy is minimal. This can be identified using the quantum mechanical computational method Density Functional Theory [91].

2.2.2 Broadband radiation

2.2.2.1 Spectroscopy

Spectra can be determined using different approaches. Taking inspiration from the different designs of mid-infrared spectrometers, terahertz spectrometers can send a polychromatic wave onto a sample and decompose the radiation which has not been absorbed by the sample into a series of signals at each frequency, either through the use of a two-beam interferometer such as a Michelson type, combined with Fourier transform signal processing (Fourier-transform spectroscopy [73]), or through the use of dispersive elements such as a prism, a grating system or a Fabry-Prot (dispersive spectroscopy [92]).

As for the mid-infrared region, polychromatic radiation (also called broadband radiation) can originate from:

- A heated material irradiating electromagnetic waves with a broad energetic spectrum and the maximum energy emission with a spectral position and amplitude depending on the temperature, as defined by Planck’s law. Globar, a silicon carbide rod heated above 1400 K, is often used [75,93], as well as the hot plasma of a mercury arc lamp [73,94,95].

- Deviation of the linear trajectory of a charged particle, using a synchrotron instrument [96].

Alternatively, a broad frequency range $\Delta f$ (or bandwidth) can be obtained from the generation of a radiation over a very short time interval $\Delta t$, as stated by the time-bandwidth theorem [75]:

$$\Delta f \cdot \Delta t \geq 1$$ (2.10)
To generate terahertz radiation within the frequency range of \( \simeq 0.1 - 3 \) THz, the electromagnetic field should therefore be emitted during only \( \Delta t = 500 \) fs or less, also defined as an *ultrashort pulse*. If \( \Delta t = 1 \) \( \mu \)s, \( \Delta f \geq 1000 \) Hz = \( 10^{-9} \) THz and the radiation can be considered monochromatic, or *narrowband*. The emitted electromagnetic field should also have a central frequency below 3 THz, *i.e.* an energy below 12.4 meV, “the total energy of the terahertz pulse is found by integrating [the electric field] over both the area of the terahertz focal spot and the duration of the terahertz pulse” [97].

Several technologies have been developed to generate such terahertz pulses: photoconductive antenna (PCA), surface-field emitter, photo-Dember emitter, or optical rectification. The first three technologies rely on the use of a near infrared laser sending an ultrashort pulse less than 100 fs long, which carries enough energy to generate charge supports (electron - hole pairs) at the surface of a semi-conductor, but only for a short period of time. These short-lived charge supports are separated either by an electric field (internal to the semiconductor or externally applied) or due to their different diffusion rates. This progressive separation of the charges with time results in the creation of a time-dependent electric dipole. The work of Maxwell and Hertz showed that a transient dipole radiates electromagnetic waves [98]. For these emitters, the generated electromagnetic waves interact constructively and result in a wave with a maximal amplitude (*i.e.* a quasi resonant effect) when the frequency of the wave lies in the terahertz range below 3 THz. But these waves only last for a period of time that depends on the lifetime of the charge supports, and their time-bandwidth is typically below 500 fs (so a frequency range broader than 2 THz, most commonly up to 4 THz). Optical rectification, however, does not rely on photogeneration of charge supports and creation of short-lived dipoles. It uses an intense optical (*i.e.* visible) pulsed laser impinging on a crystal with nonlinear optical properties. This crystal converts incident electromagnetic waves oscillating at visible frequencies within the pulse spectral range into electromagnetic waves oscillating at a frequency in the infrared region, but with an amplitude envelope oscillating at a frequency in the terahertz region. The spectral range of the resulting terahertz pulse can reach several tens of THz with a central frequency below 15 THz, but Huber *et al.* have succeeded to reach a central frequency of 41 THz with a spectral range broader than 20 THz (terahertz pulse shorter than 50 fs) [99]. More information on these emission devices is provided in textbooks by Lewis [75]
Spectroscopy with pulsed sources works as follows (please refer to Figure 2.7):

1. The emitted terahertz pulse propagates into ambient or purged air until reaching the surface of a sample with a refractive index $n_S > n_{\text{air}}$.

2. At the air - front surface interface of the sample, the pulse is partially reflected and partially transmitted through the sample. Because $n_S > n_{\text{air}}$, the reflected pulse undergoes a phase shift of $\pi$, while the transmitted pulse undergoes no phase shift.

3. Upon transmission through the sample, the amplitude of the pulse decreases due to absorption by the sample, and possibly due to scattering effects. Besides, its speed of propagation $v$ decreases due to the increased refractive index of the sample compared to air ($n = c/v$, with $v$ being the phase velocity). Since most materials are dispersive, $n_S$ varies depending on the frequency of the propagating electromagnetic wave: terahertz waves at certain frequencies will propagate in the sample more slowly than others, and this will cause pulse broadening. At the sample - air interface, the pulse is again partially reflected and partially transmitted. This time, because $n_{\text{air}} < n_S$, the reflected pulse undergoes no phase shift, and neither does the transmitted pulse.

4. The pulse reaching the detecting device on the other side of the sample than the source (transmission geometry) has a lower amplitude, is broader and is delayed compared to a pulse which simply travels through air (as illustrated in Figure 2.7b). The receiver detects the amplitude of the electric field of the pulse (and not its intensity or irradiance) at successive points in time, in order to retrieve its time evolution and reconstruct the full pulse. Terahertz spectroscopy in transmission mode consists of comparison of the frequency-dependence of the magnitude of the complex Fourier transform of pulses transmitted through the sample with those transmitted through air. Whenever collective vibration causes absorption at a specific frequency, the magnitude of the signal at this frequency will be lower.
than the one from the air reference.

Terahertz TD or pulsed spectroscopic instruments are now widely used due to the recent commercial developments. They represent an alternative to Fourier-transform spectroscopy when investigating the spectral response of a material below 3 THz due to improved signal-to-noise ratio in this spectral region, while Fourier-transform spectroscopy is preferred for analysis above 5 THz [94].

Figure 2.7: Schematic diagram representing the propagation of an ultrashort terahertz pulse a) within a system consisting of a material of refractive index $n_S > n_{\text{air}}$ (grey rectangle) surrounded by air; b) through air; c) within a system consisting of a metallic plate (textured rectangle) surrounded by air. The end signal measured by the receiver, either in transmission or reflection geometry, is framed. Transmitted pulses are represented by dashed curves, and reflected pulses by dotted curves.
The second difference between THz-TDS and Fourier transform spectroscopy is that it provides information about time evolution of the electric field. The time delay and pulse broadening of the signal transmitted through the sample compared to the signal transmitted through air are associated with the frequency-dependent refractive index of the sample, inducing a frequency-dependent phase velocity of the pulse during propagation through the sample.

The difference in time of collection and pulse shape between the pulses transmitted in the presence and absence of a sample can therefore enable the calculation of the refractive index $n_S$ of the sample, while their difference in amplitude enables the calculation of the absorbance or absorption coefficient $\alpha_S$ of the sample (please refer to the Chapter 4). While it was already possible to recover both of these optical parameters with Fourier transform spectroscopy, the two parameters were not simultaneously recorded and required the use of the Kramers-Kronig dispersion relationship to derive the value of $n_S$ at the frequency $\omega'$ from the measured $\alpha_S$, or extinction coefficient $\kappa_S$ ($\alpha_S(\omega) = 2.\kappa_S(\omega).\frac{\omega}{c}$ [100]):

$$n_S(\omega') - 1 = \frac{2}{\pi} P \int_0^\infty \frac{\omega \kappa_S(\omega)}{\omega^2 - \omega'^2} d\omega$$

(2.11)

With the symbol $P$ referring to the calculation process involving the Cauchy principal value [101]. However, Kramers-Kronig relations need absorption information over the whole spectral range (0 to infinity), while only a limited spectral range can be accessed experimentally (e.g. 0.1 to 10 THz). Therefore, the refractive index determined with THz-TDS is more accurate than the one calculated from the absorption coefficient measured with Fourier-transform spectroscopy.

The signal reaching the detecting device on the same side of the sample as the source (reflection geometry) consists of two successive pulses corresponding to the first and second interface, air - sample and sample - air, respectively. If the sample is too thick or too absorbent to transmit the terahertz pulse to the sample - air interface, transmission spectroscopy as described above is not possible and the reflected signal only consists of a single pulse. Frequency-dependent refractive index and absorption coefficient can still be extracted from this signal, as explained by Blampey et al. [102], using the single pulse reflected from a mirror as the reference (see Figure 2.7c). Yet for
materials with absorption, reflectivity and thickness low enough to transmit a detectable pulse through the sample and to the detector, the transmission geometry is preferred for spectroscopy as it yields more precise values [102].

2.2.2.2 Imaging

We have so far implicitly discussed the signal collected from one single point or area of the sample, where the electromagnetic wave impinges on. By raster-scanning the sample or the antennas (both source and detector), signals from many adjacent points can be collected. In reflection geometry, a two-pulse signal is collected for a layer of a transparent material forming two interfaces on each of its side, due to the mismatch in refractive index between the sample material and the adjacent material (air in Figure 2.7a). If materials with different refractive indices are overlaid, the reflected signal contains as many pulses as there are interfaces with different refractive indices. This is the case of two overlaid thick sheets of paper separated by a layer of air, as represented in Figure 2.8a: the succession of air - paper / paper - air / air - paper / paper - air interfaces results in a reflected signal consisting of four pulses. At each scanning position, such a TD signal is acquired. A dataset resulting from a scan therefore consists of amplitude values of the terahertz electric field along 3 dimensions: the lateral and vertical scanning positions (X and Y) and the time of acquisition (t).

This 3-dimensional dataset can be reduced to a 2-dimensional dataset (i.e. an image) by selecting and fixing a value along one dimension. If selecting a specific time of acquisition \( t_i \), corresponding to an interface of interest, an image from this interface can be produced by mapping \( X(t_i) \) against \( Y(t_i) \). Such an image is called a C-scan. In Figure 2.8, the selected time of acquisition \( t_i \) corresponds to the air - front surface interface of the bottom sheet of paper, which contains an ink inscription. Provided that the ink has a different refractive index to paper, the refractive index difference between air and paper is different to the one between air and ink: this results in a terahertz electric field reflected from the ink with a different amplitude to the one reflected from the surrounding paper. This difference in amplitude between the electrical fields reflected from ink and paper translates into an image contrast in the C-scan: the larger (negative) pulse amplitude from the ink will result in darker pixels on the C-scan (Figure 2.8b).
Figure 2.8: a) Schematic diagram representing two overlaid sheets of paper separated by a thin layer of air, with the second sheet showing an ink inscription at its surface. The 3 dimensions of the data acquired during an imaging scan are represented. $X$ and $Y$ represent the scanning directions, with scanning steps forming a grid of pixels. $t$ represents the acquisition time of the reflected electric field, with the successively measured amplitudes of the field forming a waveform (in red), also displayed in c). b) Schematic of a C-scan at the time $t_i$ coinciding with the air - front surface interface of the second sheet of paper. d) and e) Schematic of B-scans, i.e. virtual cross-sections along the horizontal line $Y_i$ and along the vertical line $X_i$ respectively.
If selecting a specific scanning position $Y_i$, the 3-dimensional dataset is reduced to amplitude values of the electric field along $X$ and $t$, forming an image called a B-scan (see Figure 2.8d). Such an image represents a virtual cross-section of the scanned multilayer object: each depth interface along the horizontal line $Y=Y_i$ is recognizable from its different pixel colour, with an interface corresponding to an increase in refractive index (e.g. air - paper) associated to negative amplitudes and dark pixel colours, while an interface corresponding to a decrease in refractive index (e.g. paper - air) is associated to positive amplitudes and light pixel colours. The presence of the ink at the front surface of the second paper sheet results in larger (negative) amplitudes of the electric field and pixels associated to the presence of ink appear darker than the rest of the pixels on the same time line. A virtual cross-section can also be produced along the other scanning direction by simply selecting $X=X_i$ and mapping the amplitude values of the electric field along $Y$ and $t$. If the ink is more absorbent than paper, the terahertz beam transmitted through the ink will have a lower amplitude than the beam transmitted through paper. Electric field amplitudes collected at times of acquisition after $t_i$ will therefore be lowered for scanning positions corresponding to the presence of ink: this is observed on the B-scan as an "occlusion" effect for time lines acquired later than $t_i$.

Pulsed imaging in transmission configuration is also possible, as illustrated in Figure 2.9. Keeping the example of Figure 2.8a of two overlaid sheets of paper, the transmitted signal at each scanning position would consist of a main pulse collected first, and subsequent pulses collected later in time, resulting from multiple reflections on internal interfaces but eventually propagating in the forward direction. The first acquired pulse has passed through all the different interfaces without undergoing any reflection within the sample. It represents an accumulation of the absorption properties from all the different materials it has transmitted through. In the example of Figure 2.9a, the only difference to the overall absorption properties of the stack of sheets comes from the presence of ink at the front surface of the second sheet of paper. Thus, by selecting the time $t_0$ of acquisition of the first and main transmitted pulse to produce a C-scan, the resulting image 2.9b is very similar to Figure 2.8b: the overall sample absorption is increased at scanning positions where the ink is present, which translates into darker pixels in the C-scan image.
2.2. Terahertz science and technology

Figure 2.9: a) Schematic diagram representing two overlaid sheets of paper separated by a thin layer of air, with the second sheet showing an ink inscription at its surface. \( t \) represents the acquisition time of the transmitted electric field. The main signal of the full waveform is collected first and is the pulse which has transmitted through the entire sample, not undergoing reflection. Subsequent pulses collected at later time points are pulses which have transmitted through the entire sample but have also undergone reflection(s) on interfaces internal to the sample (only two are represented as an example); b) Schematic of a C-scan at the time \( t_0 \) coinciding with the main transmitted pulse. c) Schematic diagram representing two overlaid sheets of paper, with both the first and the second sheets showing an ink inscription at their surface. d) Schematic of a C-scan at the time \( t_0 \) coinciding with the main transmitted pulse.

When ink inscriptions are present on both sheets, as illustrated in Figure 2.9c, the main transmitted electric field amplitude is altered by the absorption from the two ink inscriptions. The resulting C-scan at \( t_0 \) (Figure 2.9d) shows a superposition of the two inscriptions, rendering the textual information from each sheet far less legible.
Figure 2.10: a) Schematic diagram representing two overlaid sheets of paper, with both the first and the second sheets showing an ink inscription at their surface. \( t \) represents the acquisition time of the reflected electric field. 

b) Schematic of a C-scan at the time \( t_0 \) coinciding with the pulse reflected from the air - front surface interface of the first sheet. 

c) Schematic of a C-scan at the time \( t_i \) coinciding with the pulse reflected from the air - front surface interface of the second sheet.

The same sample imaged in reflection configuration (Figure 2.10a) would also lead to a superposition of the two inscriptions on the C-scan at \( t_i \) (Figure 2.10c), but the inscription from this specific interface can easily be recovered by subtracting the values from the C-scan at \( t_0 \) (Figure 2.10b). This is the reason why the reflection configuration is often preferred when imaging multilayered samples. Yet there might be cases (e.g. text written on one side of an unfolded sheet of paper within an envelope) for which the transmission configuration can yield similar or better results than the reflection configuration.
2.2.3 Narrowband radiation

Many developed terahertz sources do not generate a broadband beam but a (quasi) monochromatic, \textit{i.e.} narrowband beam. For many of these sources, the terahertz beam is continuously generated, as is the case of thermal and synchrotron broadband sources. They are therefore called Continuous Wave (CW) technologies, as opposed to pulsed or TD (TD) technologies. Amongst those, a source generating electromagnetic waves in the millimetre or microwaves range (such as a Gunn diode) combined with a frequency multiplier can generate narrowband CW at a frequency below 2 THz [77]. Terahertz lasers also exist: the most common are the molecular (gas), the germanium and the quantum cascade laser. These three types of lasers operate in CW mode (but can be modified to also operate in pulsed mode [77]). Additionally, they all have a tunable frequency of emission. Yet the spectral width of an emission band is often too broad and the spectral gaps between two emission bands are often too large to reach the right continuous spectral range and resolution of use in spectroscopy. However, all these sources can be used for imaging purposes.

With CW narrowband sources, the evolution of the electric field with time is lost, and the collected signal is a single irradiance value influenced by the accumulation of the absorption (reflective) properties from all the layers within the sample in transmission (reflection) configuration, much like X-ray imaging. The output of an imaging scan is therefore a 2-dimensional dataset. To recover 3-dimensional information, tomography (\textit{i.e.} rotation of the object or the source to acquire a set of 2D images at multiple angles) can be used [77, 103], as is the case of X-ray based technologies, and the internal structure of large 3D objects such as clay vessels can be investigated [104,105].

Additionally, real-time (\textit{i.e.} video rate) imaging can be performed with narrowband sources (either continuous or pulsed) as the sensors associated with these waves (\textit{e.g.} microbolometers) can be engineered in an array: raster-scanning the sample is not necessary any more to acquire a 2-dimensional image, and 60 frames per second can be acquired [77].

2.3 Terahertz applications in cultural heritage

To prove the relevance of this project, it is crucial to understand the research approach, motivation and terahertz set-ups previously used for analysis of documents and, more
broadly, cultural heritage materials that are also used in the production of documents, to be able to build on the current state-of-the-art.

2.3.1 Pioneers

Because Fourier-transform spectrometers expanding data acquisition to the terahertz range are commercially available, “affordable”, and allow very fast acquisition of spectra, they were first used (and are still widely used) by the pioneers in the study of art materials with terahertz spectroscopy [106]. This Japanese team was the first one to create a free online terahertz spectral database dedicated to art materials (within the NICT-THz project and jointly with RIKEN laboratories [107]). As their intention was the recognition of terahertz spectroscopy as a routine tool for art analysis, they hoped that this spectral database would be integrated in the widely used IRUG (Infrared and Raman Users Group) database [108, 109]. Their works highlight very interesting features of various art materials. Fukunaga et al. [106, 110] first revealed that most inorganic pigments have fingerprint spectra in the terahertz region 0.6 - 13 THz, presenting a few FT-THz spectra of inorganic pigments and fillers in transmission (e.g. aureolin, Prussian blue, ultramarine, lead white, titanium white, calcium carbonate). They then used these spectral signatures to demonstrate that two pigments of similar colour could be discerned with terahertz spectroscopy (such as titanium white and lead white), and that a mixture of a pigment and a binder would show the spectral characteristics of all the separate constituents. They also suggest that some art materials do not have specific terahertz spectral signatures, such as some organic dyes and earth pigments, while many binding media do.

With regards to imaging using terahertz radiation, the first images ever obtained were by Hu and Nuss in 1995 at AT&T Bell Laboratories, using TD technology in transmission [17]: they exploited the strong absorption of terahertz radiation by polar water molecules and imaged the relative concentration of liquid water in a leaf freshly cut and in the same leaf 48 h later, highlighting its dehydration process. The pioneers of terahertz imaging for heritage applications also used TD technology in transmission: in 2006 Kohler et al. imaged a portrait painted on canvas [111]. They produced 2D images using the pulse delay and the amplitude of the Fourier-transform of the transmitted signal within 0.1 - 0.2 THz and 0.3 - 0.4 THz. At the time, they acknowledged that they
had too little knowledge on the interaction of terahertz radiation with the materials and the layered structure of the artefact, to fully interpret those images. Yet, they showed the potential of this technology in the field of heritage science: they showed for the first time that terahertz beams could transmit with relatively little attenuation through a canvas painting, and they also showed that the canvas texture could be revealed. They also rightly identified research avenues worth following: revealing hidden wall paintings with imaging in reflection mode, identifying paint substances by investigation of their terahertz spectra, determining the structure and thickness of paint and coating layers. Finally, they pointed out knowledge gaps that would need to be explored in order to interpret those images: understanding terahertz transmission spectra, evaluating and differentiating the effects of absorption, scattering and Fabry-Perot type effects in terahertz spectra and images, gaining information on the refractive index of paint.

Figure 2.11: Citation report obtained from Web of Science™ publication database in October 2015, showing the number of outputs published each year within the latest 20 years with a topic related to terahertz and heritage or to terahertz and art conservation (search entry: TS=(terahertz AND heritage) OR TS=(terahertz AND ”art conservation”).

2.3.2 Past and current research

Some research teams focussing on development and use of terahertz technologies or, more broadly, imaging or non-destructive technologies, showed continued interest in heritage applications, and transferred the knowledge they developed to the field of heritage science when a connection with a heritage institution was possible and when fund-
ing for such projects was available. It is therefore fair to say that "terahertz heritage science" has evolved almost at the same pace as "terahertz science". As an example, imaging of the dehydration process in a leaf at two points in time separated by 48 h shown by Hu and Nuss was pushed further by Fukunaga et al. to real-time (60 images per second) imaging of the evolution of the distribution of water in a sheet of paper over 10 s [112] using a portable terahertz camera developed only a year earlier by NEC and NICT, made with of a focal plane array of micro-bolomoters, and a quantum cascade laser (QCL) source operating at 3.1 THz [113]. These real-time images undeniably showed the potential of this technology to monitor aqueous conservation treatments of heritage artefacts.

These research teams are active in France, Japan, United Kingdom, Denmark, Italy, Greece, Germany and often collaborate quite closely [114]. Together, they now form a research community publishing about 5 papers each year (see Figure 2.11), meeting at dedicated workshops (e.g. THz-ARTE, 2014 [114]) and enabling the field to gain international recognition with technical sessions in large international conferences and symposia now dedicated to terahertz heritage science (e.g. "Terahertz and Allied Methods for Cultural Heritage" session organized at the 2015 Eastern Analytical Symposium [115]). This community shows a real motivation to provide technical solutions to curatorial questions, and is aware that i) this requires more than occasional investigation of heritage artefacts, that ii) this technology is currently difficult to implement in a heritage institution, partly due to its cost, and partly due to the expertise often required to run such set-ups and interpret results, and that iii) the heritage community is often unaware of the existence of this technology, the variety of existing instruments and their potential. For this reason, the gap between terahertz science research teams and conservators in heritage institutions is bridged through more long-term PhD and post-doctoral projects, either co-funded by the scientific and heritage institutions or funded by European research programme [116]. It should therefore be mentioned that before the start of my research project, there has been four PhD projects related to the application of terahertz science to heritage [93, 103, 117, 118], and during the course of my project, another PhD project emerged (Corinna Koch-Dandolo at DTU Fotonik, Denmark [119, 120]). All have brought extremely valuable and complementary findings, for both the terahertz research community and for conservators, as we will see
2.3. Terahertz applications in cultural heritage

throughout this thesis.

A very good survey produced by Jackson et al. describes and links publications related to terahertz applications in cultural heritage and conservation science until 2011 [95], and the program of the THz-ARTE workshop is a good indicator of the research trends and progresses since then [114]. To better serve the purpose of our research project, we will focus on research work associated with textual archives and their constituting components, summarized in Tables 2.2 and 2.3.

<table>
<thead>
<tr>
<th>Spectroscopy</th>
<th>Continuous Waves</th>
<th>Broadband</th>
<th>Pulsed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pigments, inks, and binders</td>
<td>Absorbance / transmittance / reflectance spectra: “carbon black” [121], lamp black [107], vine black [107], ivory black [107], sepia [107], bistre [107], iron gall ink [121], Prussian blue [107], minium (red lead) [107, 122], vermilion (cinnabar) [93, 107], cochinéal [107, 121], red or rose madder [107], red clay (Terrà di Siena) [93], lapis-lazuli [107], azurite [107, 122], indigo [107, 122], burnt verdigris [107], malachite [122] [107], rabbit skin glue [107], materials used in coal tar dyes (naphthalene, anthracene, phenanthrene, pentacene, as well as some of their halogenated compounds) [124], “red ochre” [125], “carbon black” [125], burnt sienna [117], mixed carbon iron oxide (black) [117], calcium carbonate [126].</td>
<td>Absorbance / transmittance / reflectance spectra: lamp black [107], vine black [107], ivory black [107], Prussian blue [107], minium (red lead) [107, 122], vermilion (cinnabar) [107], red or rose madder [107], lapis-lazuli [107], azurite [107, 122], indigo [107, 122], burnt verdigris [107], malachite [122] [107], rabbit skin glue [107], materials used in coal tar dyes (naphthalene, anthracene, phenanthrene, pentacene, as well as some of their halogenated compounds) [124], “red ochre” [125], “carbon black” [125], burnt sienna [117], mixed carbon iron oxide (black) [117], calcium carbonate [126].</td>
<td>Refractive index and/or absorption coefficient: minium (red lead) [122], azurite [122], indigo [122], malachite [122], burnt sienna (bulk refractive index) [117], mixed carbon iron oxide (black) (bulk refractive index) [117].</td>
</tr>
<tr>
<td>Supports</td>
<td>Absorbance / transmittance / reflectance spectra: “paper” [107], filter paper [107], wall paper [107], “copier” paper [127], blue-lined paper [127], newsprint paper [127], cellulose [127]</td>
<td>Absorbance / transmittance / reflectance spectra: alkali lignin [128], hydrolytic lignin [128], gelatine [128], cellulose [84, 128–130], Whatman filter paper (made of pure cellulose) [128], ground-wood containing paper [128], 250 historical paper sheets dating from ≃1500 to 2006 [128], printer paper [129], wall paper [107].</td>
<td>Refractive index or absorption coefficient: cellulose [129, 130], “photocopy” or “printer” or “copyer” paper [129, 131, 132, 134], “fine” paper [131], 21 µm thick lens paper [135], weighing paper [134], letterhead paper [129], glossy paper [129], thin card [129], tissue paper [132], magazine paper [132], Italian paper from 1430 [136], modern paper artificially aged [136].</td>
</tr>
</tbody>
</table>

Table 2.2: Summary of the published research works related to the spectroscopy of textual documents and their constituting components using terahertz technologies. References in bold relate to works published during the course of this research project (2011 - 2015).
Chapter 2. Background

Narrowband

| Imaging | Narrowband | Images of model "carbon black" inscriptions handwritten on overlaid papyrus sheets [125].
|         | Real-time imaging of the evolution of the distribution of water in a sheet of paper over 10 s, using a QCL source at 3.1 THz [112].
|         | Real-time image of the inscriptions MIT handwritten with a "pencil" on the inside of a paper security envelope, using a QCL source at 4.3 THz [137].
|         | Image of 2 mm-width letters "T", "H" and "Z" handwritten with an "ordinary" pencil on the 2nd, 25th, and 50th sheet of paper within a stack of 50 Post-it notes, using a CW source at 540 GHz and a tomosynthesis imaging system (similar to X-ray CT) [138].

Broadband (pulsed)

| Images of texts printed on the recto and verso of two overlaid sheets of paper (i.e. 2 double-side printed sheets) [142].
| Images of inscriptions on a sheet of paper within an envelope, handwritten with a "pencil", a "ball point pen" and laser printed [142].
| Topographic image of the surface of a sheet of paper with inscriptions handwritten with different grades of graphite pencil [143].
| Image of a butterfly drawn by hand on "standard" paper with a 4B graphite pencil and recovered with a "black ink felt-tip pen" [143].
| Image of an inscription handwritten with a charcoal pencil and recovered with a "black paint" [144].
| Inscriptions handwritten on paper with a 4H lead (graphite) pencil covered by a model honey-based seal [145].

Table 2.3: Summary of the published research works related to the imaging of textual documents using terahertz technologies. References in bold relate to works published during the course of this research project (2011 - 2015).

A few things to notice about Tables 2.2 and 2.3 are:

- A large variety of terahertz technologies, with different generation and detection systems, can and have been used for the study of documents and their components.

- Paper has been extensively studied with terahertz spectroscopy, particularly pulsed spectroscopy, while the spectral response of parchment in the terahertz range has not been investigated yet. There are three reasons for this. The paper industry, a prominent industrial sector, shows a great interest in non-invasive and fast analytical tools which could be implemented in-line and would give access to a combination of material properties, which would otherwise require independent analysis, sometimes destructive or time-consuming. THz-TDS seems to move in this direction [131, 133]. Secondly, paper is a crucial material when it comes to mail screening for security reasons: it is important to distinguish the spectral signature of drugs or explosive substances enclosed in an envelope from the signature of paper. Finally, the chemical composition, physical properties and mechanisms of degradation of paper have been more intensively investigated than
2.3. Terahertz applications in cultural heritage

those of parchment, even in the heritage science sector (according to the results of the search entry [TI=(parchment AND characterization) OR TI=(parchment AND study)] in the Web of Science\textsuperscript{TM} publication database, 73 outputs have been published over the past 20 years, which is in a big contrast with 16,721 for the search entry [TI=(paper AND characterization) OR TI=(paper AND study)]). There is therefore more information on paper than on parchment to help with the interpretation of terahertz spectroscopic results. This might be due to the fact that there are often more paper-based than parchment-based documents in libraries, archives and museum collections, and that parchment is a more durable material, more resistant to environmental degradation and ink corrosion than paper [146]. Many studies focus on the terahertz spectral response and optical parameters of \textit{in vivo} or fresh skin, with a water content close to 80%. This response is therefore likely to be dominated by terahertz absorption from bound water, and small variations in moisture content / hydration state of the skin can be probed with terahertz pulsed spectroscopy and have enabled researchers to distinguish between different layers of skin or healthy and cancerous tissues (e.g. [147–149]). These studies are not informative of the terahertz response of parchment, as it represents dry and processed skin: scattering processes on the highly ordered fibrous structure and collective vibrations arising from inter-molecular bonds responsible of the cohesion of this ordered structure are not masked by water absorption any more and may now be observed in the spectra of optical parameters.

- There was no terahertz spectral study of iron gall ink when our research project started, although it is a crucial component of textual archives, as discussed before. Interestingly, neither iron(II) nor iron(III) sulfate compounds had ever been studied with terahertz spectroscopy, while they are widely used in pharmaceutical products and industrial processes. As La Camera indicates, crystalline iron sulfate can form at the surface of iron gall ink inscriptions [58], and investigation of these compounds with terahertz spectroscopy would be of use to both the cultural heritage field and to many more research and industrial sectors.

- Until the work of Hong \textit{et al.} [122] published in 2014, there was very little information on optical parameters (refractive index and absorption coefficient)
of pigments, binders and inks used in historical documents. It is interesting to note that many publications about imaging use graphite (lead) pencil or "carbon black" ink for their model inscriptions, materials which are known to be highly reflective and absorbent in the terahertz region as will be discussed in Chapter 5. Since terahertz waves are sensitive to differences in refractive indices or absorption coefficients between adjacent layers, such pigment selection ensured a good image contrast of the ink on support, and focussed the research on development of novel terahertz signal and image processing approaches. When imaging the letters "THZ" written (with an undisclosed writing tool) on a folded piece of paper within an envelope, Wallace et al. state “Rather than resulting in a change of thickness, the letters visibility is more likely due to an increase in absorption and/or refractive index” [139]. Yet they also admit that they “do not know the refractive index of ink on paper”. So how could they have been sure that the image contrast did not result from indentations left by the writing tool? Abraham et al. recognized that “the amplitude of the terahertz wave can vary depending on the pen pressure” (unfortunately without an experimental proof) [143]. Mittleman et al. used a charcoal pencil for their case study as they stipulated that “Not surprisingly, if the inscription is written with conventional ink, one cannot detect it in the terahertz image, because the dye molecules in the ink are non-polar and thus do not absorb strongly in the terahertz range” [144]. So if the indentations are not responsible for the image contrast observed by Wallace et al. or by Taday et al. using a ball point pen [142], and if the dye potentially present in the ink is not responsible either, what makes the inscriptions legible in the terahertz image? Terahertz depth-imaging requires sophisticated and expensive equipment, with potentially lengthy post-processing of the acquired data, and it is crucial to objectively explore the parameters leading to the image contrast. In this way, scientists and conservators can together assess the likelihood of a document to be legible with specific terahertz technology and the value of an imaging experiment with this technology. To do so, more information about refractive indices and absorption coefficients of pigments, inks, binders and supports below 4.5 THz (the frequency range mostly used for terahertz imaging) needs to be gathered.
• The terahertz spectral signature or frequency-dependence of optical parameters of pigments is often not supported by complementary analysis to confirm the nature of the pigment acquired from art suppliers. Additionally, the reproducibility of the spectral data is often not discussed (error bars representing standard deviation or standard errors are not systematically provided). Since it is not a functional group which gives rise to a characteristic absorption feature in the terahertz range, and since each crystalline material has its own specific spectral signature, it is difficult to assess the validity of a published spectrum from a pigment if it represents the only published spectrum of this pigment, with no error bars provided.

2.3.3 Latest trends

It is interesting to note that in 2008, Wallace et al. concluded their terahertz imaging experiment of inscriptions enclosed in an envelope by “due to resolution and signal-to-noise constraints, it is unlikely to be an important enough application to be developed further”. Yet the majority of the imaging experiments listed in Tables 2.2 and 2.3 were published after 2008. Besides, apart from our research project, a research team at the Sapienza University, Rome, is designing novel FD imaging technology “to read what is unreadable at infrared and visible wavelength e.g. [...] pages are sticking to each other, or the ink has disappeared [...] or cartouches were glued by the ancient writers on top of some hidden text” [150]. Additionally, the LOMA terahertz research group in Bordeaux, France, now collaborates with the Groupe La Poste in France and the German Deutsche Post [151] to further develop mail screening, and is also involved in the collaborative project TeraBook, dedicated to development of terahertz technologies and knowledge on radiation-matter interaction in order to improve document digitisation [152]. The INSIDDE project (INtegration of cost-effective Solutions for Imaging, Detection, and Digitisation of hidden Elements in paintings) [153], funded by the European Union Seventh Framework Programme, launched in January 2013 and is due to end in December 2015. It investigates digitisation of paintings and 3D artworks with terahertz technologies, but will also include terahertz spectral analysis of art materials, expanding and enriching the current online database thzdb.org [107]. Finally, the last recent work to discuss is related to ”seeing through” metal with terahertz imaging tech-
nology. As observed in Table 2.1, a layer of gold was sometimes added to a page and covered underlying text. The following question was received during our project from a conservator: “is it possible to read carbon ink under a layer of gold? In old manuscripts that changed owner, [...] miniatures colofons were often covered with gold. It would be really nice if you could read the text underneath using terahertz”. As discussed above, metals are highly opaque in the terahertz region, with terahertz radiation penetrating less than 1 $\mu$m into the material before its amplitude substantially decays. Yet, recent work has shown that gold layers used in some artworks can be of a thickness less than or comparable to the penetration depth of terahertz radiation in gold and imaging of the layer underneath is made possible with TD imaging in reflection mode, such as tree rings of a wooden panel [154].
Chapter 3

Materials and instrumentation

3.1 Terahertz pulsed technologies

3.1.1 Terahertz pulse generation and detection

The two terahertz time-domain set-ups used in this project (designed by TeraView, Ltd and Picometrix) use laser gated photo-conductive semiconductor antennas for terahertz pulse generation and detection.

A photo-conductive antenna is represented in Figure 3.1 and consists of a semiconductor substrate with two metallic electrodes at its surface, spaced 5 to 10 $\mu$m apart. When the area between the electrodes is illuminated with a focused electromagnetic wave of larger energy than the semiconductor band gap, electron-hole pairs are generated at its surface. When illumination stops, the surface charge carriers either recombine or are trapped in an energy state between the valence and conduction bands of the semi-conductor created by a defect in its crystal structure. Metallic electrodes are connected to a voltage source through contact electrodes: they represent an electric potential difference and induce a permanent electric field across the illuminated area. Surface charge carriers, photo-generated by an illuminating wave are accelerated and separated by this electric field, before they are trapped or recombined. The electrodes separated by a surface with time-dependent charge carrier densities and speeds can be described as an electric dipole of very small spatial dimensions with a dipole moment oscillating in time.

The works of Maxwell and Hertz showed that an oscillating dipole generates electromagnetic waves, comprised of a fluctuating electric and magnetic field [98]. These waves are linearly polarized, with the electric field oriented in the direction parallel to...
Chapter 3. Materials and instrumentation

the dipole axis [155]. The time-dependent amplitude of this electric field $E_{\text{emitted}}(t)$ observed at a position $r$ from the position of the source dipole $r'$ was expressed by Jefimenko as [156]:

$$
\vec{E}_{\text{emitted}}(\vec{r}, t) = \frac{1}{4\pi\varepsilon_0} \int \rho(\vec{r}', t - \frac{r - r'}{c}) \cdot \vec{u}_r + \frac{\dot{\rho}(\vec{r}', t - \frac{r - r'}{c})}{c \cdot (r - r')} \cdot \vec{u}_r - \frac{\vec{J}(\vec{r}', t - \frac{(r - r')}{c})}{c^2 \cdot (r - r')} \, d\tau'
$$

(3.1)

with $\rho$ the density of charge carrier, and $J$ the current density created by the displacement of charge carriers. $J$ is linked to $\rho$ by:

$$
\vec{J}(\vec{r}', t) = q \cdot \rho(\vec{r}', t) \cdot \vec{v}(\vec{r}', t)
$$

(3.2)

with $q$ and $v$ the electrical charge and speed of the charge carriers, respectively; $v$ depends on the mobility $\mu$ of the charge carrier in the substrate, and on the amplitude of the electric field $E_{\text{voltage}}$ induced by the voltage source and accelerating the carriers to the electrodes. Considering that electrons have larger mobility than holes, the influence of holes in the generation of the current density can be neglected:

$$
\vec{J}(\vec{r}', t) = e \cdot \rho_e - \vec{v}(\vec{r}', t) \cdot \mu_e - \vec{E}_{\text{voltage}}
$$

(3.3)

It follows from Equation 3.1 that the duration and amplitude of the electric field of the wave emitted by the dipole, $E_{\text{emitted}}(t)$, depends on the power, wavelength and time-duration of the illuminating wave, as well as the width of the illuminated area, the nature of the substrate and the voltage applied to the electrodes, since all these parameters influence the generated number, lifetime and acceleration of the charge carriers ($\rho(t)$, $\dot{\rho}(t)$ and $\dot{v}(t)$).

TeraView technology for terahertz pulse generation uses a Ti: Sapphire laser emitting in the near-infrared range with a central wavelength at $\approx 800$ nm (i.e. 1.55 eV) and an average power of $\approx 300$ mW. This laser is mode-locked in order to generate ultrashort pulses of near-infrared radiation with a bandwidth limited pulse of less than 100 fs. The substrate is a Gallium Arsenide (GaAs) semiconductor, with a band gap of 1.42 eV (smaller than the energy of the incident near-infrared pulse). The illuminated area is $\approx 5 \mu$m wide and the applied voltage uses an AC source of 50 V, mod-
ulated at 33.3 kHz. The resulting emitted electromagnetic wave has a time-duration of less than 1 ps and its maximum amplitude falls in the low-terahertz spectral region (< 3 THz): it is a terahertz pulse. This pulse is radiated out into $4\pi$ steradians, and the portion that is transmitted forward through the bulk of the semiconductor is collected and directed into free space via the use of a hyper hemispherical silicon lens.

**Figure 3.1:** Schematic diagram explaining the generation of a terahertz pulse from a laser-gated photoconductive antenna: the grey area represents the surface of the semiconductor illuminated by the near-infrared incident pulse and crossed by accelerated surface charge carriers (electrons - plain circle - and holes - blank circle). It should be noted that while terahertz pulses are generated in both the specular direction of the incident near-infrared pulse and through the semiconductor substrate [157], only the latter is often used and represented here.

Detection of the terahertz pulse uses the same principles as generation. However, no voltage is applied to the electrodes of the detecting photoconductive antenna: it is the terahertz electromagnetic pulse that is detected and directed with the silicon lens onto the back of the semiconductor substrate which creates an electric field across the surface between the electrodes and accelerates surface charge carriers generated by the illuminating near-infrared pulse. The photocurrent $J$ created by the displacement of the charges is then amplified and detected. From Equation 3.3, it follows that this detected current is proportional to the amplitude of the electric field of the terahertz pulse to be detected.

Yet, the detection only lasts for $\sim 200$ fs, corresponding to the lifetime of the charge carriers in the semiconducting substrate of the detecting antenna. Therefore, the measured current only corresponds to a sample of the full time-dependence of the terahertz electric field, during the time $\delta t$ when the detection “gate” is open. To recover
the entire terahertz pulse, the electric field should be sampled at different discrete points in time. To do so, the same near-infrared pulsed laser source is used to illuminate the semiconductor surface of both the emitting and receiving photoconductive antennas at a repetition rate of 80 MHz, using a beam splitter: the two subsequent near-infrared beams are often referred to as *pump* and *probe* beams, respectively. A series of terahertz pulses is therefore generated every 12500 ps by the emitting antenna, while the detecting antenna is also illuminated every 12500 ps. Each of the generated terahertz pulses has its electric field detected at a different sampling time by delaying the detection time by $\delta t$ (e.g. 12500 ps + $\delta t$, 12500 ps + 2$\delta t$ etc.). This progressively increasing delay of the detection time is obtained by increasing the travelling path of the *probe* near-infrared pulse with the help of an optical delay line, constituted of mobile mirrors. The full time-dependence of the terahertz electric field is therefore reconstructed from the successive current measurements performed for different lengths of the delay line.

It should however be noted that terahertz radiation naturally and constantly emitted by the environment at room temperature (thermal background) also impinges on the detecting photoconductive antenna. A previous study [158] showed that the terahertz background power level at room temperature on a 0.5 mm diameter spot is 9.9 $\mu$W. Considering a diameter spot of 5 $\mu$m corresponding to the gap between the electrodes, the area is reduced by a factor $10^4$ and so is the power: the terahertz thermal background therefore illuminates the gap on the detecting antenna with a power of $\approx$0.99 nW. Since the power of the terahertz pulse generated by the emitting antenna is also in the order of nanowatts, a lock-in amplifier is used to extract the current corresponding to the terahertz pulse from the current corresponding to the thermal background and considered as random noise. This amplifier uses the modulation frequency of the voltage source connected to the emitting antenna: as the voltage switches from 50 V to 0 V every 33.3 kHz, no terahertz pulses are generated and therefore detected every 33.3 kHz. The lock-in amplifier amplifies the global signal evolving at the same frequency as the modulation frequency, *i.e.* 33.3 kHz, and attenuates all other signals evolving at different frequencies.

Piao et al. observed that the major portion of the terahertz radiation emitted by a photoconductive antenna is due to the ultra fast change of the carrier density, *i.e.* $\dot{\rho}(t)$, with the carrier acceleration $\dot{v}(t)$ only accounting for a minor portion of the emit-
terahertz radiation. Upon illumination of the antenna, charge carriers are created and therefore $\dot{\rho}(t) > 0$. When the illumination stops, charge carriers recombine or are trapped, and $\dot{\rho}(t) < 0$. It follows from Equation 3.1 that the emitted terahertz electromagnetic field has a positive amplitude during and just after the illumination, and then becomes negative. Piao et al. also observed that detection of the terahertz radiation with an other photoconductive antenna distorts the shape of the signal from the emitter. The resulting detected waveform therefore has a positive peak and a negative peak, and the convention followed by TeraView is to define the larger of these two peaks as positive. This explains the shape of the time-domain signals presented in the following chapters.

### 3.1.2 Terahertz time-domain spectroscopy

#### 3.1.2.1 Spectroscopy in transmission mode

Spectroscopic data was acquired with the TPS spectra 3000 from TeraView Ltd, in transmission mode. The picture and a schematic diagram of the set-up are shown in Figure 3.2. The generated terahertz pulses are focused with gold coated ellipsoidal mirrors and the sample to be analyzed is placed at this focal plane with the help of a sample holder. The sample compartment is closed and purged with nitrogen during a scan acquisition to minimize terahertz absorption by ambient water vapour. 1800 scans were collected in rapid-scan mode, i.e. with a delay line from 0 to 6 mm at a rate of 30 scans/s (giving a total acquisition time of 1 min), and the resulting signals were averaged to maximize the Signal-to-Noise Ratio (SNR). The TD signal was decomposed into its frequency components and thus transformed into a frequency-domain spectrum through the use of a Fast Fourier Transform (FFT) algorithm built-in the data acquisition software. The resulting spectral resolution of each acquired spectrum was 1.2 cm$^{-1}$.

#### 3.1.2.2 Software

Data was acquired and processed with the TPS Spectra software Version 1.17.6 developed by TeraView Ltd.
Chapter 3. Materials and instrumentation

Figure 3.2: a) Image (credit: TeraView Ltd) and b) schematic diagram of the TPS Spectra 3000, TeraView Ltd, used for THz-TDS in transmission. The grey rectangle represents the purged sample compartment, and the dashed and dotted lines represent the probe and pump beams, respectively. PCA refers to photoconductive antennas.

3.1.2.3 ATR spectroscopy

The ATR mode allows the incident terahertz pulse to travel through a medium with a refractive index $n_1$ larger than the index of the sample to be analysed $n_2$. At the interface of the two media, the beam can be totally reflected if the angle of incidence is superior to the critical angle ($\theta_c$) defined as:

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right)$$  \hspace{1cm} (3.4)

This phenomenon is known as total internal reflection. When a beam is totally reflected at an interface, a non-propagating wave normal to the surface penetrates the less dense medium. This wave is known as an evanescent wave, and penetrates the less dense medium at a penetration depth defined by:

$$d_p = \frac{\lambda}{2\pi \sqrt{n_1^2 \sin^2(\theta) - n_2^2}}$$  \hspace{1cm} (3.5)

Over the penetration depth of this evanescent wave, the sample partially absorbs the wave. As a result, the terahertz pulse totally reflected at the interface and detected will not have the exact same amplitude as the incident pulse, which enables the con-
An ATR module designed by TeraView and composed of a silicon crystal \((n \approx 3.4)\) and a hand-tightened anvil was fitted into the sample compartment of the TPS Spectra 3000, as represented in Figure 3.3. The angle of incidence was set to 35°. For a sample with a refractive index of \(\approx 1.3\), the penetration depth ranges from \(\approx 10\) to \(100 \ \mu m\) over the terahertz range \(10 - 100 \ \text{cm}^{-1}\) (3 - 0.3 THz).

For measurements of the spectral signature from ink lines on historical documents, the ATR spectrometer is preferred over the transmission spectrometer since:

- Absorption from minute amounts of sample (as low as 1 mg) can be detected with this set-up, making this acquisition \textit{a priori} suitable for detection of the spectral signature from components of a thin line on the surface of a writing support.

- Spectral information is extracted solely from the area in contact with the crystal, the diameter of which is \(\approx 6\ \text{mm}\) and is smaller than the width of the terahertz beam above \(40 \ \text{cm}^{-1}\) for the transmission set-up (see next chapter): the support surrounding an ink line would therefore contribute less to the collected signal, leading to an \textit{a priori} better signal collected for the ink.

- Considering that a sheet of paper or parchment is on average thicker than \(70 \ \mu m\), the ATR mode enabled extraction of information mostly from the front surface of the document \((\text{recto})\), without any interference from the back surface \((\text{verso})\).

The same scanning parameters and software tool were used as for the transmission mode.

\[ \text{Figure 3.3: ATR module designed by TeraView Ltd and fitted in the TPS Spectra 3000 sample compartment (credit: TeraView Ltd).} \]
3.1.3 Terahertz time-domain imaging

3.1.3.1 TPS spectra 3000 equipped with a RIM

Terahertz images were acquired in reflection mode with the TPS spectra 3000 equipped with a RIM fitted in the nitrogen purged sample compartment, as shown in Figure 3.4. This module comprises of a software controlled motorized scanning table with a set of ellipsoidal mirrors underneath, focusing the generated terahertz pulses onto the holed scanning stage at an angle of 30°. Samples were placed on the scanning table and some were maintained steady and flat by placing them between two ring-shaped magnets taped to the stage. The stepper motor could perform scanning steps as low as 50 µm. For each scanning step, terahertz pulses reflected from different sample interfaces were redirected through ellipsoidal mirrors towards the photoconductive detector and altogether constituted a TD waveform. A total of 4096 time points were sampled over 47.7 ps to reconstruct the waveform, which corresponded to a 7-mm optical delay:

$$\text{Time}_{\text{Optical delay}} = 2 \cdot \frac{\text{Length}_{\text{Optical delay}}}{c}$$

(3.6)

To reduce the file size, only one data point every 4 sampling points could be stored, resulting in 1024 saved time points. For each scanning position, two waveforms were acquired at a rate of 15 Hz (one during the up and one during the down sweep of the delay line) and averaged, resulting in an integration time of ≃130 ms: this integration time could not be modified by the user. Finally, the entire dataset of a terahertz TD imaging scan consisted of a 3D matrix with sampling time, scanning position x and scanning position y as a basis.
3.1. Terahertz pulsed technologies

3.1.3.2 TPS spectra 3000 with a gantry system

Some documents were larger than 100 mm x 100 mm and could not be placed on the scanning table of the RIM. In such cases, a gantry system was used, consisting of metallic arms with photoconductive antennas and focusing hemispherical TPX lenses (made of poly 4-methyl-pentene-1, with a 25 mm diameter and a 30-mm focal length) at both ends oriented at 60°, connected to the near-infrared pulsed laser via fibre optic cables. The sample was placed on a software controlled motorized stage under the emitting and receiving heads, as illustrated in Figure 3.5.

3.1.3.3 Software

Data was acquired with the ScanAcquire software, and processed with the DataProcess software, both developed by TeraView Ltd. Data were stored as .tvl files only readable with TeraView software tools, but images produced with the DataProcess Software could be exported as .dat file, and raw data could also be exported in a format suitable for Matlab. For a scanned area of 4 cm by 0.9 cm and a scanning step size of 0.2 mm, the acquisition time was $\approx$20 min and the generated .tvl file size was $\approx$36 MB.

Origin Pro 8.6 and Matlab R2008a were used to further process and display the acquired terahertz data.

**Figure 3.4**: a) Image of the RIM attached to the TPS Spectra 3000 during the scan of a letter, and b) Image of the module alone (credit: TeraView Ltd).
3.1.4 **Time-domain imaging with T-Ray 4000 (Picometrix)**

To explore the advantages and challenges related to different instrumentation manufactured by different suppliers, and form a more informed opinion on the potential of terahertz pulsed technologies for archival applications, a second terahertz pulsed instrument was used at the premises of the C2RMF, Paris, and later at the premises of the LRMH, Paris, in collaboration with Bianca Jackson.

The commercially available T-Ray 4000 (Picometrix) was used, represented in Figure 3.6. This terahertz pulsed instrument also relies on laser-gated photoconductive semiconductor antennas. The mode-locked near-infrared Ytterbium laser operates at a central wavelength of $\approx 1064$ nm (1.16 eV) and generates pulses less than 100 fs at a repetition rate of 50 MHz, with an average power of $\approx 400$ mW. Photoconductive antennas are designed with a bespoke multi-layer semiconductor structure grown on a GaAs substrate, the resulting bandgap of which is equal to or less than 1.16 eV [159].

The main differences with the RIM connected to the TPS Spectra 3000 from TeraView are:

- The emitting and receiving antennas are movable and fibre-coupled, enabling the user to adopt various configurations, such as scans in transmission or reflection modes, with normal or oblique incidence. Reflection mode at normal incidence was used for all images presented in this thesis.
3.1. Terahertz pulsed technologies

- Terahertz pulses are focused onto the surface of the sample via the use of aspheric High Density Polyethylene (HDPE) lenses, not ellipsoidal mirrors. The lens used during image acquisition had a diameter of 1.5” (38 mm) and a 1” focal length (25.4 mm), but could be replaced by a 3” or 6” focal length lens. The low (though not zero) absorption of HDPE in the terahertz region, specifically at high frequencies, leads to a reduced bandwidth, and therefore poorer spatial resolution compared to ellipsoidal mirrors, as will be discussed in the next Chapter.

- The Ti:Sapphire laser is large (length ≃ 50 cm) and requires water cooling, while the Ytterbium laser is compact and air-cooled. The small weight of the main unit of the T-Ray 4000 system (25 kg [160]), the movable fibre-launched antennas and the possibility to change the focusing lens to another lens with a different focal length and therefore a different depth of focus makes this system easily transportable and flexible, which represents an advantage for in situ measurements. Yet, this also meant that measurements were not performed in a purged environment, leading to strong water vapour absorption of the terahertz pulse and thus a poorer SNR and an increased complexity of the signal.

- The number of waveforms to be averaged for each scanning position could be defined by the user, and was set to 10 in all images presented in this thesis.

- The lowest scanning step size obtained with the stepper motor was 100 µm.

- A rotating optical delay line operating at 100 Hz was used, as opposed to the set of mobile mirrors linearly displaced by a stepper motor for the TPS Spectra 3000 from TeraView. 4096 time points were acquired over a time range of 320 ps. The resulting waveform therefore had a lower time-resolution than waveforms acquired with the 7 mm optical delay line within the TPS Spectra 3000 set-up (0.078 ps, against 0.046 ps for 1024 data points collected with the RIM, or 0.012 ps for 4096 data points collected with the RIM), but offered the potential to collect signals from deep interfaces in thick samples (i.e. signals with large time delays).
3.1.4.1 Software
Zaber (Vancouver) motorized linear slides were used for the scanning stage and were computer controlled via a Labview interface provided by Zaber. Data were acquired and processed using a custom Labview interface developed by Bianca Jackson [161]. An image data set consisted of a folder containing a series of .txt files, one for each scanning position, containing the TD data. For a scanned area of 3 cm by 0.8 cm and a scanning step size of 0.2 mm, the generated folder size was ≃63 MB and the acquisition time was 40 min, for a step speed of 0.5 mm/s.

3.2 Samples
3.2.1 Pigments and chemicals
The selection of pigments and inks used in this study was based on the frequency of their use in archival documents [33, 162, 163], their predisposition to degrade the carrier, the availability of their components, and the feasibility of their production in a laboratory. Owing to the minimal sample preparation it requires, the ease of access to the instrument and the large number of spectral databases available for identification, ATR-FTIR spectroscopy was used as a characterization tool to confirm the nature of pigments acquired from art suppliers and identify additives or fillers. When possible and when of interest to collaborators, Energy-Dispersive X-ray Spectroscopy (EDS),
X-ray fluorescence (XRF) spectroscopy and Powder X-Ray Diffraction (PXRD) were used to provide further insight into the composition and structure of the investigated compounds. Spectra and information about these characterization tools can be found in the Appendix A.

3.2.1.1 Carbon-based black

Following the convention proposed by Winter (1983), the term carbon-based black is used to denote a group of black or brown pigments composed primarily of carbon [162].

**Lamp black**

Lamp black belongs to the group of flame carbons as it was historically produced from the soot of burnt oil and resin (including tar or resinous wood) [162]. The pigment was purchased from Cornelissen and Son, London.

**Carbon black**

Carbon black belongs to the group of flame carbons and is "produced by incomplete combustion of gaseous hydrocarbons" [162]. The pigment was purchased from Cornelissen and Son, London. It should be noted that the surface of carbon black has been reported to be paracrystalline (i.e. locally crystalline, without long-range ordering of the lattice) [164].

**Graphite**

Graphite is a crystalline form of carbon, composed of overlaid sheets of hexagons of carbon atoms, and is found in lead pencils. The high mobility of the electrons makes it a semi-metal (i.e. a conductor along the plane of each sheet) [165]. The pigment was purchased from Cornelissen and Son, London.

Bands at 375 - 650 (broad), 876, 1385, 1431, 1630 and 3100 - 3700 cm\(^{-1}\) (broad) are observed for KBr pellets containing lamp black, carbon black and graphite, as well as for the blank KBr pellet used as a reference. These bands originate from water molecules adsorbed and trapped within the KBr and pigment pores (375 - 650, 1630 and 3100 - 3700 cm\(^{-1}\)) [166], NO\(^-\) in KNO\(_3\) (1385 cm\(^{-1}\)) and inorganic carbonates (714, 876 and 1431 cm\(^{-1}\)), contaminants in KBr salts [167, 168]. Bands at 2854 and 2923 cm\(^{-1}\) are characteristic of -CH\(_3\) and -CH\(_2\) structures, suggesting the evidence of some aliphatic species on the carbon [169]. The multiple bands between 1032 - 1270 cm\(^{-1}\) can be assigned to ether-like, epoxide and phenolic structures [169].
black, lamp black and graphite, although mostly constituted of pure carbon, therefore also contain organic matter.

**Bone black**

Bone black belongs to the group of cokes as it was historically produced from burnt bones, the organic phase (collagen) forming a black to brown coke on an inorganic hydroxylapatite substrate [162]. The pigment was acquired from Cornelissen and Son, London, as "ivory black", but a discussion with the supplier revealed that it was more likely to have been produced from calcined bones than ivory. Its ATR-FTIR spectrum revealed spectral features characteristic of bone and hydroxyapatite, with vibrations from $\text{PO}_4^{3-}$ in calcium phosphate (469, 560, 599, 960, 1022, 1085 cm$^{-1}$ [170–172]) and $\text{CO}_3^{2-}$ in calcium carbonate (873, 1409 and 1452 cm$^{-1}$ [170, 172]), together with vibrations attributed to cyano stretches in cyanamide, characteristic of burnt bone (697 and 2012 cm$^{-1}$ [173, 174]). EDS results further indicated an atomic percentage ratio of 22.95% of carbon, 50.91% of oxygen, 8.12% of phosphorus and 15.71% of calcium.

**Ivory black**

Ivory black belongs to the group of cokes as it was historically produced from burnt ivory. The pigment was acquired from Cornelissen and Son, London, as "genuine ivory black". Its ATR-FTIR spectrum is identical to the one of bone black, except for the absence of the cyano stretches bands at 697 and 2012 cm$^{-1}$. EDS results further indicated an atomic percentage ratio of 17.75% of carbon, 56.02% of oxygen, 9.79% of phosphorus and 14.53% of calcium. It should be noted that while the fraction of calcium to carbon was not found to be statistically different for bone black and ivory black, ivory black had a fraction of phosphorus to carbon larger than bone black (0.55% ± 0.03% against 0.35% ± 0.07%).

**Bistre**

Bistre belongs to the group of flame carbons as it was historically produced from the soot of burnt wood. After the soot was collected close to the flames, it was treated with hot or boiling water; when the particles had settled out, the supernatant liquors were decanted and the sediment taken to dryness [162]. The pigment was acquired from Kremer Pigmente, Munich (Product number: 12100). Its FTIR spectrum shows spectral features characteristic of chars from burnt lignin (a polymeric aromatic organic substance in wood) [175].
3.2. Samples

**Sepia**

Sepia ink was made by diluting 4 g of cuttlefish ink (acquired from a local fishmonger) in 10 mL of deionised water.

3.2.1.2 Coloured pigments

The following pigments were commonly found in illuminated manuscripts [163].

**Malachite**

Malachite is a green pigment constituted of copper carbonate, \( \text{Cu}_2\text{CO}_3(\text{OH})_2 \), forming crystals with a monoclinic symmetry [176]. Its FTIR spectrum is identical to the FTIR spectrum of pure malachite available on the RRUFF online database (RRUFF ID: R050508) [177].

**Lapis lazuli**

Lapis lazuli, or simply lapis, is a blue rock composed of a mixture of minerals, usually calcium carbonate, pyrite and lazurite, with the latter being the dominant constituent, and the composition of lazurite is \((\text{Na,Ca})_8[(\text{Al,Si})_{12}\text{O}_{24}]\text{(S,SO}_4\text{)}\). The structure of lazurite itself is complex and several polymorphs (e.g. cubic, modulated and orthorhombic) are known; these polymorphs can be found within a single fragment of the mineral [162]. The pigment was acquired from Cornelissen and Son, London, and the supplier was not able to attest the provenance of this pigment (Chilean or Afghan). Its ATR-FTIR spectrum shows intense overlapping bands at 960 - 1000 cm\(^{-1}\) also found in the lazurite sample from the Forbes Pigment Database [178], but shows no band associated to calcium carbonate (RRUFF ID: R040070; 712, 872, 1400 cm\(^{-1}\)) or pyrite (RRUFF ID: R050190; 816, 1012, 1076, 1122, 3232 cm\(^{-1}\)) from the RRUFF online database [177]. It can therefore be assumed that the supply is largely composed of pure lazurite.

**Vermilion**

Vermilion (or cinnabar) is constituted of crystalline red mercury(II) sulfide (HgS) with a trigonal symmetry [162]. The pigment was acquired from Cornelissen and Son, London. Neither its ATR nor transmission FTIR spectra showed spectral feature, which is consistent with the literature [174, 179]. However, Derrick et al. remind us that the red pigments cadmium red and realgar also show no spectral band in the mid-infrared region [174]. Still, the terahertz spectral signatures of these three red pigments are
Chapter 3. Materials and instrumentation

different [107]: vermilion shows sharp spectral bands around 1.1, 1.3 and 2.6 THz, which is in agreement with the terahertz spectrum of the vermilion supply presented in Section 5.5.1.

Minium
Minium is a crystalline red lead(II,IV) mixed-valence oxide, \( \text{Pb}_3\text{O}_4 \), with a tetragonal symmetry. The pigment was acquired from Cornelissen and Son, London. The transmission FTIR spectrum of minium showed intense spectral features at 387, 453 and 532 cm\(^{-1}\), as previously reported for minium [93, 180], together with weak bands at 685 and 1410 cm\(^{-1}\) originating from minute amounts of lead carbonates, as reported in the online NIST Chemistry WebBook [181] and further discussed by Guo [182]. Yet, lead(II, IV) oxide is rarely pure commercially, more typical contents being in the range 85 - 98%, with the rest being made up by lead(II) oxide (i.e. PbO) [162]. PbO has two polymorphs: litharge (tetragonal), showing a broad absorption band at \( \approx 465 \text{ cm}^{-1} \), and massicot (orthorombic), showing a broad absorption band at \( \approx 505 \text{ cm}^{-1} \) [88], which may explain the presence of overlapping weak and broad bands around 489 and 511 cm\(^{-1}\) in the spectral signature of the supply.

Verdigris
The term verdigris has a loose definition. It applied to corrosion products formed from copper and copper alloys following exposure to the atmosphere and more particularly to organic acids (particularly acetic and formic acid) [162]. A crystallised form of the green pigment verdigris is copper(II) acetate monohydrate (\( \text{Cu(CH}_3\text{COO})_2\cdot\text{H}_2\text{O} \)), with a monoclinic symmetry [183]. The pigment was acquired from Cornelissen and Son, London. Its ATR-FTIR spectrum is identical to the spectrum of copper acetate monohydrate reported in the online NIST Chemistry WebBook [181].

Prussian blue
Prussian blue is regarded as the first modern synthetic pigment and was widely used in artworks and documents due to its availability and ease of use [41, 184]. Prussian blue is a term which might reasonably be considered applicable to any of the blue hexacyanoferrate(II) pigments (i.e a metal ion forming a complex with the \([\text{Fe(II)(CN)}_6]_4^-\) ion) which also contain Fe(III) [162]. Due to its high tinting properties, the pure pigment can appear as almost black and extender may be added (starch, calcium salts, barytes, kaolin, alumina). The resulting mixture is often called Antwerp blue [184].
3.2. Samples

Prussian blue has also been observed to be prone to colour fading or darkening, although the chemistry underlying these processes is still subject to research [184, 185]. To account for the possible variations in chemical compositions, four different supplies were acquired: ”Prussian blue” and ”Antwerp blue” from Cornelissen and Son, London, ”Milori blue” (another trade name for Prussian blue) from Kremer Pigmente, Munich, and ”Prussian blue” from Zecchi, Florence. The ATR-FTIR spectra of all supplies were identical and showed the strong absorption at 2080 cm\(^{-1}\) characteristic of the cyano group. The bands at 605, 1611 and 3256 cm\(^{-1}\) were also reported elsewhere for Prussian blues of analytical grade or laboratory-prepared [186,187], while the bands at 834, 982, 1046, 1118 and 1413 cm\(^{-1}\) could indicate the presence of calcium carbonate and kaolinite. ”Antwerp blue” showed strong absorption at 633, 982, 1066, 1171, 2855, 2925 and 2955 cm\(^{-1}\) attributed to barium sulfate [181], a common extender.

3.2.1.3 Binders

Arabic gum

Arabic gum is a naturally occurring substance made of polysaccharides. It was purchased from Kremer Pigmente, Munich, and was ground into powder in a coffee grinder. Its ATR-FTIR spectrum shows bands at 1018, 1600 and 3280 cm\(^{-1}\) characteristic of polysaccharides [174] and is identical to the spectrum of Arabic gum of industrial grade reported elsewhere [188]. XRF spectroscopy revealed the presence of potassium (17153 ppm) and calcium (7095 ppm) in this supply, but PXRD further showed that the gum was amorphous (no Bragg peak observed and presence of a strong scattering background): it can therefore be concluded that calcium and potassium present in the gum do not form crystalline salts.

Rabbit skin glue

Rabbit skin glue was purchased from Cornelissen and Son, London, as small solid aggregates and was ground into powder in a coffee grinder. Its ATR-FTIR spectrum shows amide I, II and III bands at 1628, 1528 and 1447 cm\(^{-1}\) respectively, characteristic of protein [174].

3.2.1.4 Iron gall inks

The preparation of iron gall inks (IG) was based on nine historical recipes [35]. Although originally made from crushed gall nuts, IG inks were prepared in this study
using pure tannic or gallic acids (from Chinese gall nuts, Sigma Aldrich) to increase control over the experiments. Gall nuts contain almost 70% tannic acid [35], therefore the ratio of \( \approx 70\% \) of the mass of nuts indicated in the recipe was used to determine the mass of pure acid needed for the four different recipes of iron gall ink: IG1, IG2, IG3 and IG5. Historic recipes suggest that crushed gall nuts can either be decanted or boiled in water [33], thus increasing the hydrolysis of tannic acid into gallic acid. Both pure acids were used for the preparation of inks in this research, tannic acid from Chinese gall nuts (Product number: T0200, Sigma Aldrich) and gallic acid (Product number: G7384, Sigma Aldrich). The purity and amount of iron sulfate contained in the vitriol historically used by ink makers may have varied a lot, therefore, different ratios of iron and copper sulfates (p.a., Sigma Aldrich) were investigated. Similarly, the ratio of iron sulfate and acid in the ink also varied significantly, and additional inks with different mass or molar ratios than indicated in the reference recipe [35] were prepared. Table 3.1 presents the mass of the different ingredients for each iron gall ink, including Arabic gum (Kremer, Munich), mixed with 50 mL of deionised water.

Gallic acid exists in several polymorphs and hydrated forms. The supply used for the experiments represented two crystalline polymorphs, as revealed by PXRD (see Appendix A): an anhydrate form (ijumeg02), with a crystal space group C2/c corresponding to the polymorph AH-II described elsewhere [189], and a monohydrate form (kontiq02), with a crystal space group P2/n corresponding to the form II described elsewhere [190]. The supply of tannic acid showed no crystalline structure with PXRD. PXRD of the supply of iron(II) sulfate revealed the presence of two hydrates: the heptahydrate form, with a crystal group P2_1/c, and the tetrahydrate form, with a crystal group P2_1/n. PXRD of the supply of copper sulfate revealed the presence of anhydrous copper sulfate, crystallizing in the orthorhombic Pnma space group, and monohydrate copper sulfate, crystallizing in the triclinic P -1 space group [191]. No significant scattering background was observed in the diffractogram, suggesting that there was no amorphous copper sulfate present.
Table 3.1: Compositions of iron gall inks, for preparation with 50 mL of deionised water. \textit{rmass} stands for mass ratio of iron(II) sulfate to acid, and \textit{rmolar} stands for molar ratio of the same.

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Tannic acid</th>
<th>Gallic acid</th>
<th>Iron(II) sulfate</th>
<th>Copper(II) sulfate</th>
<th>Arabic gum</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG1</td>
<td>1700 mg</td>
<td>-</td>
<td>2500 mg</td>
<td>25 mg</td>
<td>1250 mg</td>
</tr>
<tr>
<td>IG2</td>
<td>1700 mg</td>
<td>-</td>
<td>2500 mg</td>
<td>250 mg</td>
<td>1250 mg</td>
</tr>
<tr>
<td>IG3</td>
<td>-</td>
<td>1700 mg</td>
<td>2500 mg</td>
<td>25 mg</td>
<td>1250 mg</td>
</tr>
<tr>
<td>IG5</td>
<td>1700 mg</td>
<td>-</td>
<td>2500 mg</td>
<td>-</td>
<td>1250 mg</td>
</tr>
<tr>
<td>IG1/rmass = 0.25</td>
<td>1700 mg</td>
<td>-</td>
<td>425 mg</td>
<td>4.25 mg</td>
<td>1250 mg</td>
</tr>
<tr>
<td>IG1/rmass = 0.1</td>
<td>1700 mg</td>
<td>-</td>
<td>170 mg</td>
<td>1.70 mg</td>
<td>1250 mg</td>
</tr>
<tr>
<td>IG3/rmolar = 0.1</td>
<td>-</td>
<td>340 mg</td>
<td>56 mg</td>
<td>-</td>
<td>250 mg</td>
</tr>
<tr>
<td>IG3/rmolar = 0.2</td>
<td>-</td>
<td>340 mg</td>
<td>111 mg</td>
<td>-</td>
<td>250 mg</td>
</tr>
<tr>
<td>IG3/rmolar = 0.5</td>
<td>-</td>
<td>340 mg</td>
<td>278 mg</td>
<td>-</td>
<td>250 mg</td>
</tr>
<tr>
<td>IG3/rmolar = 0.9</td>
<td>-</td>
<td>340 mg</td>
<td>500 mg</td>
<td>-</td>
<td>250 mg</td>
</tr>
<tr>
<td>IG3/rmolar = 2</td>
<td>-</td>
<td>340 mg</td>
<td>1112 mg</td>
<td>-</td>
<td>250 mg</td>
</tr>
<tr>
<td>IG3/rmolar = 5</td>
<td>-</td>
<td>340 mg</td>
<td>2780 mg</td>
<td>-</td>
<td>250 mg</td>
</tr>
</tbody>
</table>

3.2.1.5 Preparation

For terahertz spectroscopic analysis, dry powders were analysed, and the solutions of sepia ink and iron gall inks were therefore left to dry on a ceramic plate, scraped off and ground into a powder with a pestle and mortar before analysis.

For terahertz imaging experiments, ink solutions were prepared by dispersing the pigment, with no further grinding, in a solution of Arabic gum. Although a broad range of ratios of binder over water can be found in the literature, the same proportion recommended by Kolar \textit{et al.} [35] was consistently used (i.e. 625 mg of Arabic gum with 25 mL of deionised water heated at 70 °C and left to cool down before the addition of the pigment), as it gives suitable viscosity and increases control over the experiments. Inks were prepared with a mass ratio of pigment to gum of 1:2. Mass ratios were preferred over volume ratios, as the volume of a pigment would be difficult to measure accurately due to its inter- and intra-particle pores, and as the penetration properties of the binder into the pigment pores are not known. To demonstrate the influence of ink on image contrast in a document, any indentations left by handwriting with a quill, nib pen or pen should be avoided, as they can influence the image contrast, as will be demonstrated later. Application with a brush or using screen printing was also avoided as the volume of ink and pigment concentration applied could not be uniformly distributed, which would have increased the complexity of the experiment and interpretation. Application with an ink jet was not considered, due to the colloidal nature of the pigment suspensions, which would have risked clogging the ink jet heads. For these
reasons, 2 µL of each fresh ink was applied on historical rag paper and parchment sheets using pipettes. The resulting ink deposits had a diameter of 2 mm ± 0.5 mm, depending on the affinity of the ink to the carrier. Although the diameter of ink inscriptions written with most contemporary pens was usually smaller than 2 mm, some inscriptions written with quills, nib pens or brushes can be wider than 2 mm: the diameter of these ink droplets is therefore not unrealistic with regards to the width of archival ink inscriptions. Additionally, the ink spot is significantly larger than the step size and associated spatial resolution of the imaging set-up (see Chapter 4), which ensured that observation of the dot in the terahertz image was possible.

Despite the care taken in controlling the amount of pigment and ink applied on the carrier, it should be noted that the colloidal nature of these ink suspensions often led to a non-uniform deposition of particles during evaporation of water from the ink drop, creating ring-shaped solid residues (also known as "coffee ring effect" [192]). Additionally, application of ink with a pipette sometimes introduced microscopic bubbles in the droplet, creating local depletion of particle concentration. Finally, ink may penetrate non-uniformly into fibres constituting the carrier. All these phenomena led to the non-uniform distribution of ink thickness and pigment concentration within some ink dots: these variations of pigment concentration and topography have been recorded by acquiring micro-photographs of the ink dots (displayed alongside terahertz images of the same ink dots in the following Chapters) and accounted for in the interpretation of terahertz images.

### 3.2.2 Historical paper and parchment sheets

Six rag paper sheets from the 19th century and one parchment sheet from 1752, from the ISH Historic Reference Material Collection, were used as supports. To observe the influence of paper coating on terahertz images, three paper sheets were sized, i.e. immersed in a gelatine solution for less than 10 s, then placed between two sheets of polyester and pressed between wooden boards. The preparation of the size followed the instructions from Kolar and Strlíč [35], mixing 12500 mg of gelatine from bovine skin (Product number: G9382, Sigma Aldrich) and 1425 mg of alum (aluminium sulfate hydrate, Product number: 368458, Sigma Aldrich) in 0.5 L of deionised water.
Chapter 4

Metrology in terahertz time-domain spectroscopy and imaging

4.1 Terahertz time-domain spectroscopy

4.1.1 Beam size

When analysing a sample in transmission, it is important to know the diameter of the terahertz beam spot at the focal plane where the sample is placed as the collected spectrum represents the average spectral information from all materials present in a sample across the beam spot. If analysing an ink line on paper, spectral information from both the ink and the surrounding support is simultaneously collected. The spectral signature from the ink will depend on the thickness of the ink line with respect to the diameter of the beam spot.

However, the diameter of a beam spot, and more largely the beam intensity profile, depends on the frequency of the radiation. For example, in the case of a Gaussian or Bessel beam with a transverse intensity profile forming an Airy disc or concentric Airy patterns, the diameter of the beam spot can be defined as the diameter of the first dark Airy ring, and it is linked to the frequency of the radiation $\nu$ by the following equation [193]:

$$\text{Beam diameter} = 2 \cdot r_{\text{Rayleigh}} = 2 \cdot \frac{1.22 \cdot f \cdot c}{D \cdot \nu} \quad (4.1)$$

with $f$, the focal length and $D$, the aperture diameter of the imaging system (see Section 4.2.1 below). The diameter of the beam spot is therefore inversely proportional to the frequency (or wavenumber).
To estimate the frequency-dependent diameter of the beam spot at the focal plane of the TPS Spectra 3000 spectrometer in transmission, an iris diaphragm was placed across the beam in lieu of the sample holder. The aperture of the iris was progressively reduced. When the iris aperture matched the beam diameter of a specific frequency component of the pulse, the TD signal transmitted through the iris was less intense than the reference signal transmitted through the empty sample compartment. In the FD, this blockage of the outer spread of the beam governed by the lower frequency components of the pulse induced that the amplitude ratio \( \frac{|E_{\text{transmitted through iris}}(\nu)|}{|E_{\text{reference}}(\nu)|} \) was not zero for the lower frequencies. Figure 4.1b indicates that for an incident beam at 20 cm\(^{-1}\) (0.6 THz), the signal is partially blocked (i.e. it is \( > 0 \)) from iris aperture ‘40’ onwards, corresponding to a hole diameter of 9.5 mm, while for an incident beam at 40 cm\(^{-1}\) (1.2 THz), the signal is partially blocked from iris aperture ‘30’ onwards, corresponding to a diameter of 6.5 mm. We can therefore conclude that the diameter of the spot size at the focal plane is approximately equal to 3.5 mm at 100 cm\(^{-1}\), 6.5 mm at 40 cm\(^{-1}\) and 9.5 mm at 20 cm\(^{-1}\).

![Figure 4.1](image)

**Figure 4.1:** a) Terahertz TD waveforms of the signal transmitted through an iris diaphragm with different aperture sizes, placed at the focal plane of the terahertz beam, in the empty purged sample compartment. b) Associated frequency dependent ratio \( \frac{|E_{\text{transmitted through iris}}(\nu)|}{|E_{\text{reference}}(\nu)|} \).

### 4.1.2 Pellet preparation

Pigments, inks and binders were analyzed in their dry form, as liquid water strongly absorbs terahertz radiation. Powders were obtained by grinding the samples with an agate pestle and mortar. The sample powder was then mixed with a low-absorbing
4.1. Terahertz time-domain spectroscopy

HDPE powder (Inducos 13/1, particle size < 80 µm) and pressed into duplicate or triplicate pellets using a hydraulic press (Specac, P/N 15011, Orpington, UK) with a pressure of 2 tons (spread over an area of 132.73 mm² for a 13-mm die) for 2 min. For a powder mixture with a mass ratio of sample to polyethylene of 10% and a target pellet of 400 mg, 40 mg of sample powder was mixed with 360 mg of HDPE. To avoid diffraction during spectroscopic analysis (see Section 4.1.1), the powder mixture was pressed into pellets of 13 mm diameter using the Atlas Evacuable pellet die (P/N GS03000). Spectroscopic measurements were performed at least 12 h after preparation of the pellets, to allow for their expansion, as is usual practice [194].

4.1.3 Calculation of optical parameters

1800 scans were acquired for each sample at a scan rate of 30 Hz, and averaged to minimize random noise. When analysing pellets, the signal transmitted through a 360-mg pellet of pure HDPE powder was set as a reference, to only account for the influence of the 40 mg of sample on the optical path of the terahertz pulse. When analysing sheets of paper or parchment, the signal transmitted through the empty sample compartment was used as a reference.

Figure 4.2: Schematic diagram detailing the propagation of a terahertz beam through a sample with parallel faces, a) at random incidence with multiple internal reflections, and b) at normal incidence with no internal reflection.
Chapter 4. Metrology in terahertz time-domain spectroscopy and imaging

4.1.3.1 Absorbance

The TPS Spectra software calculates the absorbance as the logarithmic ratio of the magnitudes of the complex Fourier transforms from the sample and reference waveforms, $\hat{S}(\nu)$ and $\hat{R}(\nu)$ respectively:\n
$$\begin{align*}
A(\nu) = -\log\left(\frac{|\hat{S}(\nu)|}{|\hat{R}(\nu)|}\right) = -\log(|\hat{T}(\nu)|) \quad (4.2)
\end{align*}$$

where $\hat{T}(\nu)$ is the complex transmittance ratio.

4.1.3.2 Refractive index

Figure 4.2 details the propagation of a terahertz beam in a sample with parallel faces (i.e., a pellet, or a sheet of paper or parchment) surrounded by air.

Most terahertz TD spectrometers operate at normal incidence ($\theta_i = 0$):

$$\begin{align*}
\hat{E}_{R1} &= \hat{E}_i \cdot \bar{r}_{\text{air-sample}} \\
\hat{E}_{T1} &= \hat{E}_i \cdot \bar{t}_{\text{air-sample}} \cdot e^{-i\frac{n_{\text{sample}} \nu D_{\text{sample}}}{c}} \cdot \bar{t}_{\text{sample-air}} \\
\hat{E}_{R2} &= \hat{E}_i \cdot \bar{r}_{\text{air-sample}} \cdot \bar{r}_{\text{sample-air}} \cdot \left(e^{-i\frac{n_{\text{sample}} \nu D_{\text{sample}}}{c}}\right)^2 \cdot \bar{t}_{\text{sample-air}} \\
\hat{E}_{T2} &= \hat{E}_i \cdot \bar{t}_{\text{air-sample}} \cdot (\bar{r}_{\text{sample-air}})^2 \cdot \left(e^{-i\frac{n_{\text{sample}} \nu D_{\text{sample}}}{c}}\right)^3 \cdot \bar{t}_{\text{sample-air}} \quad (4.5)
\end{align*}$$

with, for $\bar{n}_{\text{purged air}} \cong 1$:

$$\begin{align*}
\bar{t}_{\text{air-sample}} &= \frac{2}{1 + \bar{n}_{\text{sample}}} \quad (4.7) \\
\bar{r}_{\text{air-sample}} &= \frac{1 - \bar{n}_{\text{sample}}}{1 + \bar{n}_{\text{sample}}} \quad (4.8) \\
\bar{t}_{\text{sample-air}} &= \frac{2\bar{n}_{\text{sample}}}{1 + \bar{n}_{\text{sample}}} \quad (4.9) \\
\bar{r}_{\text{sample-air}} &= \frac{\bar{n}_{\text{sample}} - 1}{\bar{n}_{\text{sample}} + 1} \quad (4.10)
\end{align*}$$

corresponding to the complex reflection and transmission coefficients at the different interfaces. $\bar{n}_{\text{sample}}$ is the complex refractive index of the sample that can be decomposed in terms of real refractive index, $n_{\text{sample}}$, and extinction coefficient, $\kappa_{\text{sample}}$:

\footnote{The symbol $\bar{}$ indicates a complex number.}
\[\tilde{n}_{\text{sample}} = n_{\text{sample}} - i \cdot \kappa_{\text{sample}} \]  

(4.11)

The total transmitted signal can be expressed as the sum of the transmission signals:

\[
\tilde{E}_T = \tilde{E}_i \cdot \tilde{t}_{\text{air-sample}} \cdot \tilde{t}_{\text{sample-air}} \cdot \left(e^{-i\tilde{n}_{\text{sample}}vD_{\text{sample}}/c}\right) \\
\left[1 + \left(\tilde{r}_{\text{sample-air}} \cdot e^{-i\tilde{n}_{\text{sample}}vD_{\text{sample}}/c}\right)^2 + \left(\tilde{r}_{\text{sample-air}} \cdot e^{-i\tilde{n}_{\text{sample}}vD_{\text{sample}}/c}\right)^4 + \ldots\right] 
\]

(4.12)

Assuming no internal reflections [100] (we will see that this assumption may not be valid any more when analysing thin sheets of paper or parchments), the total transmitted signal becomes:

\[
\tilde{E}_T = \tilde{E}_i \cdot \tilde{t}_{\text{air-sample}} \cdot \left(e^{-i\tilde{n}_{\text{sample}}vD_{\text{sample}}/c}\right) \cdot \tilde{t}_{\text{sample-air}} 
\]

(4.13)

which can be re-written as:

\[
\tilde{E}_T = \tilde{E}_i \cdot \frac{4\tilde{n}_{\text{sample}}}{(1 + \tilde{n}_{\text{sample}})^2} \cdot \left(e^{-i\tilde{n}_{\text{sample}}vD_{\text{sample}}/c}\right) \\
\simeq \tilde{E}_i \cdot \frac{4n_{\text{sample}}}{(1 + n_{\text{sample}})^2} \cdot \left(e^{-\kappa_{\text{sample}}vD_{\text{sample}}/c}\right) \cdot \left(e^{-in_{\text{sample}}vD_{\text{sample}}/c}\right) 
\]

(4.14)

In the case of the reference signal transmitted through the empty sample compartment, \(\tilde{n}_{\text{sample}} = \tilde{n}_{\text{purged air}} \simeq 1\), and:

\[
E_{\text{reference}} = \tilde{E}_i \cdot \left(e^{-ivD_{\text{sample}}/c}\right) 
\]

(4.15)

The transmittance ratio discussed in 4.2 becomes:

\[
\tilde{T}(v) = \frac{\tilde{S}(v)}{\tilde{R}(v)} = \frac{\tilde{E}_T(v)}{E_{\text{reference}}(v)} = \frac{4n_{\text{sample}}}{(1 + n_{\text{sample}})^2} \cdot \left(e^{-\kappa_{\text{sample}}vD_{\text{sample}}/c}\right) \cdot \left(e^{-in_{\text{sample}}vD_{\text{sample}}/c}\right) \\
= T_0 \cdot e^{i\phi(v)} 
\]

(4.16)
The frequency-dependent refractive index can therefore be calculated using the phase \( \phi(\nu) \) of the complex transmittance ratio \( \tilde{T}(\nu) \) in the FD:

\[
n(\nu) = 1 - \frac{\phi(\nu)}{2\pi \nu \ cm^{-1} D_{\text{sample, cm}}} \tag{4.17}
\]

When considering a reference signal transmitted through a pellet of pure HDPE with a thickness \( D \) and sample signal transmitted through the pellet with a thickness \( D+d \) and containing sample powder diluted in the same mass of HDPE as the reference pellet, the transmittance ratio should be written as:

\[
\tilde{T}(\nu) = \frac{\bar{S}(\nu)}{\bar{R}(\nu)} = \frac{4 \tilde{n}_{\text{diluted pellet}} e^{-i \tilde{n}_{\text{diluted pellet}}(D+d) \nu}}{(1+\tilde{n}_{\text{diluted pellet}})^2} \frac{4 \tilde{n}_{\text{pellet of HDPE}} e^{-i \tilde{n}_{\text{pellet of HDPE}}D \nu}}{(1+\tilde{n}_{\text{pellet of HDPE}})^2} \tag{4.18}
\]

To say that the attenuation due to transmission through the bulk HDPE can be accounted for as a factor \( X(\nu) \) in both Equations 4.15 and 4.14 would be an approximation: in such a case, the transmittance ratio and the refractive index remain the same as in Equations 4.16 and 4.17, with \( D_{\text{sample, cm}} = d \), the thickness difference between the reference and sample pellets. In conclusion, the TPS spectra software uses, for all sample and reference configurations, the following equation:

\[
n(\nu) = 1 - \frac{\phi(\nu)}{2\pi \nu \ cm^{-1} d_{\text{cm}}} \tag{4.19}
\]

If the sample and reference TD waveforms are not aligned on the time axis, a phase offset \( \Delta \Phi(\nu) \) proportional to the initial time difference is subtracted from \( \Phi(\nu) \) for the calculation of the refractive index.

When a beam propagates in a material of refractive index \( n \), its speed changes \( (n = c \nu) \), and so does the time of collection of the transmitted pulse. The constant (i.e. frequency independent) refractive index is calculated using the time difference \( \Delta t \) (in mm\(^1\)) between the maxima of the waveforms collected for the sample and the reference (see Figure 4.3), and the difference in thickness \( d \) (in mm) between the sample and the

\(^1\)Time measured in mm refers to the displacement of the optical delay line in the spectrometer that is required for the near-infrared pump and probe beams to have the same optical path and to detect the amplitude of the terahertz electromagnetic field at a specific point in time.
4.1. Terahertz time-domain spectroscopy

reference, using the following equation:

\[ n_{\text{Constant}} = 1 + \frac{2\Delta t}{d} \]  

(4.20)

Figure 4.3: Terahertz TD waveforms transmitted through the empty sample compartment (black curve) and through a pellet made with bistre diluted in HDPE.

Coefficient 2 arises from the fact that the measured time delay \( \Delta t \) corresponds to the change in position of the optical delay line, which induces a delay in the light path of the pump beam of \( 2\Delta t \). In order to detect the probe beam transmitted through the sample, the light paths of pump and probe beams must be equal: the time delay introduced by the sample is therefore \( 2\Delta t \).

While this value is a good and easy to compare indicator of the delay of the pulse due to changes in speed of propagation rather than to changes in pellet thickness, this value would only truly be a constant if the speed of propagation were constant, i.e. did not depend on frequency, with no broadening of the pulse observed after transmission through the pellet. This is often not the case. Frequency-dependent refractive indices were therefore also calculated.

4.1.3.3 Absorption coefficient

Attenuation of the amplitude of the terahertz electric field transmitted through the sample pellet compared to the electric field transmitted through the reference pellet arises from three phenomena: absorption by the sample, Fresnel reflection losses at
the surface of the pellet, and internal scattering from powder particles and air pores. Frequency-dependent extinction coefficients were calculated from Equation 4.19 using the following equations correcting for the effects of Fresnel reflection losses [100]:

\[ T_0 = \frac{|\tilde{S}(\nu)|}{|\tilde{R}(\nu)|} = \frac{4n(\nu)}{(n(\nu)+1)^2} \cdot e^{-\frac{\kappa_{\text{sample}}(\nu)vd}{c}} \]  

(4.21)

\[ \kappa_{\text{sample}}(\nu) = -\frac{c}{\nu \cdot d} \ln\left(\frac{|\tilde{S}(\nu)|}{|\tilde{R}(\nu)|} \frac{(n(\nu)+1)^2}{4n(\nu)}\right) = \frac{c}{\nu \cdot d} \left[\ln\left(\frac{4n(\nu)}{(n(\nu)+1)^2}\right) - \ln\left(\frac{|\tilde{S}(\nu)|}{|\tilde{R}(\nu)|}\right)\right] \]  

(4.22)

While the extinction coefficient represents the attenuation of the collected amplitude of the electromagnetic field compared to the reference amplitude of the electromagnetic field, the absorption coefficient \( \alpha \) represents the attenuation of the intensity (or irradiance, square of the amplitude of the electromagnetic field) over a certain distance called depth of penetration:

\[ I_0(d_{\text{penetration}}) = I_0(d = 0) \cdot e^{-\alpha(\nu)vd_{\text{penetration}}} \]  

(4.23)

As a result, the absorption coefficient can be written as:

\[ \alpha(\nu) = 2 \cdot \kappa_{\text{sample}}(\nu) \cdot \frac{\nu}{c} = \frac{2}{d} \left[\ln\left(\frac{4n(\nu)}{(n(\nu)+1)^2}\right) - \ln\left(\frac{|\tilde{S}(\nu)|}{|\tilde{R}(\nu)|}\right)\right] \]  

(4.26)

with \( n_{\text{Ref}} \) set to 1. Equation 4.26 does not correct for any scattering losses originating from within the pellet. Therefore, \( \alpha(\nu) \) is indicative of the total extinction of the terahertz beam, and encompasses any absorption by the sample powder, as well as scattering from powder particles and air pores.

\[ \text{TPS Spectra software uses the following equation:} \]

\[ \alpha(\nu)_{\text{TPS Spectra}} = \frac{2}{d} \left[\log_{10}\left(\frac{|\tilde{R}(\nu)|}{|\tilde{S}(\nu)|}\right) - \log_{10}\left(\frac{4n(\nu)}{(n(\nu)+1)^2}\right)\right] = \frac{2}{d} \left[-\log_{10}\left(\frac{4n(\nu)}{(n(\nu)+1)^2}\right) - \log_{10}\left(\frac{|\tilde{S}(\nu)|}{|\tilde{R}(\nu)|}\right)\right] \]  

(4.24)

The first term is negative while it should be positive, and the decimal logarithm was calculated instead of the Napierian logarithm. Absorption coefficients presented in this thesis have therefore been corrected by calculating:

\[ \alpha(\nu) = \ln(10) \cdot \alpha(\nu)_{\text{TPS Spectra}} + 2 \cdot \frac{2}{d} \cdot \log_{10}\left(\frac{4n(\nu)}{(n(\nu)+1)^2}\right) \]  

(4.25)
4.1.4 Determination of optical parameters

4.1.4.1 Precision

Sample thickness measurements

Using Equation 4.20, it can be observed that the error in refractive index $\Delta n$ caused by an error $\Delta d$ in the reading of the thickness of the pellets is a function of $d$:

$$\Delta n = |n(d + \Delta d) - n(d)| = \frac{\Delta d \cdot 2 \cdot \Delta t}{(d + \Delta d)}$$  \hspace{0.5cm} (4.27)

For a sucrose pellet with a difference in thickness between the sample and reference pellets of $d = 164 \mu m$, a time difference $\Delta t = 54.8 \mu m$ was observed. Uncertainties in readings of $d$ of 1 and $10 \mu m$ would correspond to uncertainties in refractive index of

$$\frac{\Delta n_{0.01}}{n_{\text{constant}}} = \frac{0.0384}{1.6683} = 2.3\%$$  \hspace{0.5cm} (4.28)

for $\Delta d = 10 \mu m$, and

$$\frac{\Delta n_{0.001}}{n_{\text{constant}}} = \frac{0.0041}{1.6683} = 0.2\%$$  \hspace{0.5cm} (4.29)

for $\Delta d = 1 \mu m$

For the same thickness difference of $d = 164 \mu m$, a material with a larger refractive index than sucrose would lead to a longer optical path length $n.d$, and therefore to a larger recorded time delay. For such material, the variation of $\Delta n$ due to the thickness accuracy $\Delta d$ would therefore be even larger than the values reported here. To minimize uncertainties in $n$ due to thickness measurements, a caliper with an accuracy of $\pm 1 \mu m$ was used, and calculations of $n$ were made using the average of triplicate thickness measurements on each pellet.

Signal fluctuations

It takes approximately 5 min to insert a new pellet in the sample holder, place the sample holder in the sample compartment of the TD spectrometer and wait for the compartment to purge. It takes another minute to acquire 1800 scans (scan rate of 30 scans/s). If a reference signal were to be acquired before acquiring every sample signal, the total time of analysis for each sample would be doubled ($\approx 12$ min). To increase the number of samples which could be analysed in a day, the compromise was to acquire a new reference signal only after analysis of five different samples. During the $\approx 30$ min separ-
rating each reference acquisition, the signal can fluctuate. To estimate the uncertainty in the refractive index value induced by such signal fluctuation, the following experiment was performed. Two sample pellets were prepared as described in Section 4.1.2: one with a low-absorbing material (HDPE) and one with a high-absorbing material (bone black). A pellet made of 360 mg HDPE was used as a reference. After acquisition of the reference signal, one sample pellet was placed in the sample compartment: each 5 min the pellet was taken out of the compartment and placed back in (as if a new sample pellet was inserted in the compartment), and the signal from the pellet was acquired once the purge of the compartment was complete. A total of five subsequent signal acquisitions were thus performed from the same pellet, and are reported as test 1 to test 5 in Figure 4.4. A reference pellet was then inserted in the compartment, and a new reference signal was acquired. Test 6 is the signal from the same sample pellet acquired after the new reference acquisition. The same experiment was then performed with the other sample pellet. Constant refractive indices acquired from test 1 - test 6 for both bone black and HDPE are listed in Table 4.1. The largest uncertainty in refractive index value induced by the signal fluctuating between two reference acquisitions (test 1 - test 5) is equal to $\Delta n = 0.01284 \simeq 0.01$ for HDPE and $\Delta n = 0.008397 \simeq 0.01$ for bone black, while the uncertainty in refractive index value due to random error (i.e. difference between test 1 and test 6) is equal to $\Delta n = 0.009338 \simeq 0.01$ for HDPE and 0.007634 $\simeq 0.01$ for bone black. When displaying constant refractive index values with two significant digits, the influence of signal fluctuation can therefore be discarded.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Test 5</th>
<th>Test 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% HDPE</td>
<td>1.458</td>
<td>1.462</td>
<td>1.445</td>
<td>1.445</td>
<td>1.445</td>
<td>1.467</td>
</tr>
<tr>
<td>10% bone black</td>
<td>2.088</td>
<td>2.085</td>
<td>2.092</td>
<td>2.085</td>
<td>2.096</td>
<td>2.095</td>
</tr>
</tbody>
</table>

Yet the signal does not solely fluctuate in terms of its time delay: its shape and magnitude can also undergo changes during the 30 min between two reference acquisitions, and this can influence the frequency-dependent refractive index and absorption coefficient, as observed in Figure 4.4. However, differences in refractive index values from test 1 to test 5 in Figure 4.4a remain below 0.02, which is similar to the largest
difference observed between test 1 and test 6. The same behaviour was observed for bone black. In conclusion, the acquisition of a new reference signal every 30 min does not seem to introduce more error than if acquiring a reference signal before each sample acquisition.

![Graph](image)

**Figure 4.4:** a) and b) Refractive indices and c) absorption coefficients from a pellet of HDPE (black curves) and a pellet of bone black (grey curves) analysed five times subsequently with THz-TDS without re-acquisition of a reference signal (test 1 - 5) and one time just after re-acquisition of a reference signal (test 6).

### Particle size and grinding process

Previous studies have investigated the influence of the particle size of granular materials on the spectral signature (absorbance, transmittance, absorption or attenuation coefficient spectra) of the material in the terahertz region, observing changes in the monotonic increase of the baseline, anomalous spectral features and distortions of vibrational absorption bands [195–200]. A study has recently suggested a curve-fitting equation to distinguish the absorption behavior of the material from scattering-induced effects in heterogeneous and scattering specimens [200]. Yet, no study has so far been made of the influence of particle size on refractive index values extracted using THz-TDS.
ditionally, few studies present data from duplicate or triplicate measurements (or more) and discuss the precision of the optical constants [130, 201–206]. Many substances commonly studied with THz-TDS are white, therefore the distribution of their particles within the white polyethylene matrix is difficult to observe.

The study of optical properties of pigments in the terahertz region has enabled us to observe the influence of particle size as well as of the mixing procedure of the granular material on the homogeneity of spectroscopic pellets and on the overall value and precision of the optical constants. These observations are reported and discussed below, together with best practice of pellet preparation.

Sucrose is a compound readily available in large quantities and has a distinct spectral signature in the terahertz region 0.15 - 3 THz [207]. Sucrose powder is therefore a convenient material for investigating the influence of particle size on the measured optical constants, as large amounts can be processed to obtain sufficient quantities of each size range of interest to spectroscopic analysis. Prussian blue is regarded as the first modern synthetic pigment and was widely used in artworks and documents due to its availability and ease of use [184]. The strong contrast in colour between the white polyethylene matrix and Prussian blue facilitates the observation of the effects of sample preparation on the homogeneity of the pellet.

Powders of sucrose with different grain sizes were prepared as described elsewhere [195] in five categories: 53 - 75, 75 - 106, 106 - 150, 150 - 250, and > 250 µm. Powders from each size category were then mixed with a low-absorbing HDPE powder with a mass ratio of sucrose to polyethylene of 10%.

Prussian blue pigment was dispersed on a glass slide and examined under optical microscope: individual particles could not be resolved with the optical microscope, as observed by Eastaugh et al. [162], but the average size of particle aggregates was found to be below 13 µm. Prussian blue was mixed with HDPE powder with a mass ratio of pigment to polyethylene of 10%, using 3 different procedures. Procedure a consisted of further grinding the pigment powder in an agate mortar before mixing it with HDPE powder. Procedure b consisted of mixing the pigment powder as received with HDPE (the same procedure as the one used for the preparation of sucrose pellets). Procedure c consisted of moderately grinding the powder mixture obtained from Procedure b using a pestle and mortar until a homogeneous blue powder was obtained. Procedures
4.1. Terahertz time-domain spectroscopy

and therefore have the same particle size of pigment powder before mixing with HDPE powder, but the mixing procedure is different. Each sample powder mixture was pressed into three 400 mg pellets using a hydraulic press.

Figure 4.5: Frequency-dependent refractive indices (left) and absorption coefficients (right) from pellets of sucrose powders with different grain sizes, calculated using Equations 4.19 and 4.26. Solid, dashed and dotted lines in each graph represent data collected from triplicate pellets.
Table 4.2: Average constant refractive index, slope of the absorption coefficient, observed time delay and measured difference in thickness between the sample and reference pellets, and their associated standard deviations for each set of triplicate pellets. The constant refractive index and the slope of the absorption coefficient were calculated using Equation 4.20 and the linear fit to Equation 4.26 in the range 5 - 25 cm\(^{-1}\) ($R = 0.98 \pm 0.01$).

Walther et al. [207] reported that the refractive index of a solid spectroscopic pellet obtained from milled sucrose powder diluted with polyethylene is 1.8 on average over the frequency range 0.2 - 4 THz. While Figure 4.5 and Table 4.2 report similar values, they also indicate that the average value and precision of the refractive index measurements from such pellets depends on the particle size range. While the average value and standard deviation of the refractive index is lower for particles with sizes between 75 - 106 µm, both are seen to increase with larger particle sizes, until reaching a standard deviation of 0.1 for particles larger than 250 µm. A similar phenomenon is observed for the absorption coefficients in Figure 4.5: for particles larger than 75 µm, the increasing baseline is more pronounced and less reproducible with increasing particle size. This is particularly evident when approximating the power-law frequency dependence of the baseline of the absorption coefficient to a linear dependence at low frequencies: the slope of the linear fit of the absorption coefficient between 5 - 25 cm\(^{-1}\) increases with increasing particle size, together with its standard deviation, as reported in Table 4.2.

Additionally, for both refractive index and absorption coefficient values, spectral features at 48, 55, 60, 74, 85 and 94 cm\(^{-1}\) characteristic of sucrose at 300 K [207] are well observed in all triplicate spectra for particles with sizes below 150 µm. However, for larger particles these spectral features broaden and are not consistently observed. For particles larger than 250 µm, additional features are also detected. When re-analyzing pellets with a sucrose particle size larger than 150 µm, rotating them in
the plane perpendicular to the direction of propagation of the beam, these additional bands were observed not to be repeatable. Such spectral artifacts and distortions, together with increasing baselines, were also observed by Shen et al. [195] for sucrose particles with sizes more than 250 µm in a powder cell, and by Franz et al. [196] for phenyllactic acid particles with sizes between 400 - 600 µm, diluted with HDPE and compressed into pellets. These studies describe spectral artifacts and distortions as a result of scattering, induced by the refractive index discrepancy between a sample particle with a size similar to the incident wavelength (100 - 2000 µm in dry air) and the surrounding polyethylene matrix, also known as the Christiansen effect [196]. For particles smaller than the incident wavelength, the media constituting the pellet can be considered homogeneous, and the terahertz beam is not scattered. Such scattering can also influence the time delay $\Delta t$ between the terahertz pulses transmitted through the reference and sample pellets: a terahertz pulse experiencing an internal reflection on a large particle will escape the spectroscopic pellet at a later time than a pulse experiencing no internal reflection, and the measured $\Delta t$ would be larger. Yet, if multiple internal scattering occurs on various large particles, the incident signal would be separated into different scattered signals escaping the same pellet at different times, and with different intensities. These signals would overlap and the resulting waveform would have a Full Width at Half Maximum (FWHM) and a time position of the maximum amplitude (i.e. a time delay) depending on the number and degree of overlap of these scattered signals.
This is the reason why the average time delays for sucrose pellets reported in Table 4.2 do not display a systematic decrease with smaller particle size, while their standard deviations do. The change in the degree of overlap of the scattered signals would change the shape of the waveform, and therefore its spectral dependency in the frequency-domain: this is why additional spectral features may be observed for pellets containing particles with sizes similar to or larger than the wavelength of the incident pulse. When re-analyzing pellets with a sucrose particle size larger than 150 \( \mu \text{m} \), rotating them about the axis of propagation of the pulse, the orientation of the sucrose particle HDPE particle interfaces is changed with respect to the polarized incident beam. Therefore, the incident pulse may not undergo the same multiple internal scattering and the number and degree of overlap of the scattered signals may change: the shape of the resulting waveform may differ and as a result, the spectral features observed once for this same pellet are not observed again. Therefore, we can consider that pellets constituted of large particles of sucrose are birefringent; however this does not mean that sucrose itself is birefringent.
4.1. Terahertz time-domain spectroscopy

It should be noted that the standard deviation for both the refractive index and the absorption coefficient is larger for particle sizes comprised between 53 - 75 µm than for particle sizes between 75 - 106 µm. This could be explained by the fact that at a critical particle size, the active surface area of a particle is such that surface forces (e.g. electrostatic or van der Waals) start to be significant and sample particles may have a preferred affinity to bind to other sample particles, rather than with HDPE particles. This would create clusters of sample particles that can be considered as large sample particles and would lead to scattering effects similar to those mentioned above.

Prussian blue fine particles are known to form aggregates [162] and unlike white sucrose particles are intensely coloured, which makes it possible to observe clusters of particles in the white HDPE matrix. Figure 4.6a shows an example of such a cluster at the surface of a spectroscopic pellet: the fine Prussian blue powder was further ground prior to its mixing with HDPE and this resulted in the formation of aggregates which can be larger than 500 µm. Mixing the Prussian blue powder without additional grinding resulted in aggregates smaller than 150 µm, as observed in Figure 4.6b. The use of mechanical forces through moderate grinding of the powder mixture with a pestle and mortar to counterbalance the electrostatic forces at the surface of small Prussian blue particles resulted in an intricate binding of Prussian blue particles with HDPE particles, as observed in Figures 4.6d and e. This procedure produced a homogeneous powder, and the resulting spectroscopic pellet could be considered as a homogeneous medium with respect to the wavelength of the incident beam, as observed in Figure 4.6c. Refractive indices and absorption coefficients from Prussian blue pellets produced using the different procedures are listed in Table 4.2 and Figure 4.7. Standard deviations reported in these Figures clearly indicate that optical constant values extracted from pellets produced with Procedure c are significantly more reproducible than values extracted from pellets produced with Procedure a. Furthermore, as observed for sucrose with increasing particle sizes, the increasing baseline due to scattering with the Christiansen effect is also more pronounced in absorption coefficients extracted from pellets resulting from Procedure a and b. Finally, data in Figure 4.7b and Table 4.2 indicate that the average refractive indices for pellets resulting from Procedure a and b have an offset of 0.3 to 0.4 compared to refractive indices from homogeneous pellets.
Figure 4.7: Frequency-dependent a) absorption coefficients and b) refractive indices of Prussian blue spectroscopic pellets, prepared following Procedure a, b and c (black, gray and light gray lines, respectively). The lines represent the average of data from triplicate pellets, and the error bars represent the standard deviation.

Table 4.2 also indicates that pellets resulting from Procedure c were found to be significantly thicker than pellets with a poor pigment distribution. One would expect that this would increase their associated time delays, yet Table 4.2 reports decreased time delays. Analysis of the time domain waveforms (data not shown) reveals that the signals from Procedure a and b are more attenuated and have a broader FWHM with respect to their reference signals than the signals from Procedure c. These three signs imply that the optical paths for the inhomogeneous pellets prepared with Procedure a and b are increased due to internal scattering, compared to homogeneous pellets prepared with Procedure c. Multiple internal scattering within the pellet would induce multiple time delays collected for the same incident pulse, which increases the FWHM of the resulting signal. Prussian blue samples from other suppliers have also been analyzed (see Table 5.3 in Chapter 5). Using Procedure c, homogeneous pellets of similar thickness difference and time delays as those shown for ‘Prussian blue c’ in this study were obtained. As a result, refractive index and absorption coefficient values were found to be comparable. This strengthens the argument that internal scattering on clusters of particles with a size similar to or more than the wavelength of the incident pulse, together with difference in pellet thickness due to different response to compression or expansion depending on particle distribution, lead to the large difference in optical properties between the pellets prepared following different procedures.

This section introduces a simple experimental procedure aimed at minimizing scattering from particles so as to improve the precision (i.e. reduce random errors)
of the optical constants extracted from diluted spectroscopic pellets. Its efficiency is particularly noticeable for powders with strong surface forces and prone to aggregation, such as Prussian blue as reported here. Yet scattering may also occur due to the presence of air pores of sizes similar to the incident wavelength and embedded in a homogeneous matrix. Different powders would present different mechanical properties once diluted with HDPE and pressed into a pellet, and the number and size of the pores within a pellet may differ depending on the mixing procedures (as observed here for Prussian blue) and on the nature of the sample material (its compressive strength and expansion properties, its hygroscopicity, the nature of its surface forces). More generally, optical constants extracted from porous pellets may contain an offset compared to those extracted from solid materials (see Sample morphology on p.116), the size of which would depend on the size and number of pores within the pellet. The use of Kramers-Kronig equations [208] and effective medium theory [198, 201] to account for scattering from air pores within the pellet could further help to obtain accurate and comparable optical constants from spectroscopic pellets. Finally, it should be noted that this procedure may not be used for materials prone to phase transformation during grinding or blending with a pestle and mortar. The above study has been accepted for publication [209].

4.1.4.2 Accuracy

Phase unwrapping

The frequency-dependent refractive index calculated from Equation 4.17 uses the phase angle $\phi$ of the complex transmittance ratio (Equation 4.16). Yet, the phase angle extracted from this complex ratio using any programming function similar to the function ’angle’ in Matlab returns a discontinuous value in the interval $\pm \pi$ (see Figures 4.8a and d). To obtain a refractive index spectrum without discontinuities, the phase needs to be unwrapped, i.e. an offset of $\pm 2\pi$ needs to be added to the phase $\phi(i)$ whenever $\|\phi(i-1) - \phi(i)\| > 2\pi$. The TPS Spectra software progressively unwraps the phase starting from $\omega = 0$ cm$^{−1}$. However, below $\approx 3.5$ cm$^{−1}$, the SNR is reduced: spectral artefacts occur at both ends of the FD signal, and arise from the fact that the Fourier transform is performed on a TD signal finite in time. Although these artefacts can be attenuated by using an appropriate window function (see Section 4.1.5), they cannot be
eliminated. Therefore, the unwrapped phase in the noisy low-frequency region may be erroneous and this phase error could propagate to higher frequencies.

Withayachumnankul and Naftaly [100] suggest another phase unwrapping approach to bypass this error propagation. They recommend to start unwrapping the transmittance phase from a spectral region with a good SNR onwards, and then linearly extrapolate the phase to the lower frequency region. Figure 4.8 shows the transmittance phase for a pellet of 10% minium and a pellet of 10% bone black, as calculated with the Matlab programming code emulating the steps followed by the TPS Spectra software, and as calculated using a version of the code which I adapted using the phase unwrapping process described in [100] (available in Appendix B). The latter calculation consisted of:

- Unwrapping the phase from 14.1 cm$^{-1}$ onwards
- Performing a linear fit between 14.1 - 21.6 cm$^{-1}$ (corresponding to the spectral region with the largest signal to noise region, as described in the Section 4.1.4.4)
- Extrapolating the linear fit to the region 0 - 14.1 cm$^{-1}$
- Offsetting the phase to force the intercept to the point (0, 0)

The associated refractive indices are displayed in Figure 4.8c and f. The inset in Figure 4.8d shows that the wrapped phase of the minium pellet starts at 0 and does not undergo a $2\pi$ jump at low frequencies. As a result, the refractive index of the minium pellet is not significantly affected by the selected phase unwrapping process. However, the inset in Figure 4.8a shows that the wrapped phase of the bone black pellet starts at 3.1416, and jumps to the opposite polarity by $2\pi$ for the next frequency. As a result, the refractive index of the bone black pellet calculated with the phase unwrapping process used by the TPS Spectra software shows a strong asymptotic behaviour at low frequency. This asymptotic behaviour is no longer observed when using the extrapolated phase algorithm.
Figure 4.8: Wrapped angle of the transmittance ratio from a 10% pellet of a) bone black and d) minium; b and e) associated unwrapped phase and c and f) frequency-dependent refractive index calculated using Equation 4.17. Black solid lines correspond to results obtained with Matlab programming code emulating the steps followed by the TPS Spectra software; black dashed lines correspond to results extracted directly from the TPS Spectra software (Boxcar window function); grey solid lines correspond to results obtained with the adapted version of the programming code (Appendix B), using the phase unwrapping process as described by Withayachumnankul and Naftaly [100].
When observing an asymptotic behaviour of the refractive index at low frequencies, it is therefore important to change the phase unwrapping process, to determine whether or not this frequency-dependence accurately reflects the dispersive behaviour of the material. It should also be noted that the refractive index values extracted from the extrapolated phase range between 2.065 - 2.080, while the constant refractive index of the pellet is \(n_{\text{constant}} = 2.164\). There may therefore be a low frequency signal influencing the optical delay of the transmitted pulse, which is discarded when linearly extrapolating the phase.

**Sample morphology: pellet and sheet**

To evaluate the influence of the sample morphology on the extracted optical parameters, a solid sheet of HDPE from the reference collection of the UCL Institute for Sustainable Heritage (HS57, thickness: 1.897 mm) was analysed with THz-TDS, together with three pellets made of 400 mg of compressed HDPE powder. Triplicate measurements from the sheet were acquired, with rotation of the sheet between measurements, to account for any possible birefringence of the material [214]. Figure 4.9 indicates that the frequency-dependent refractive index of a solid sheet of pure HDPE is larger than the refractive index of pellets made of compressed HDPE powder. The same phenomenon was also observed for the constant refractive index (\(n_{\text{sheet}} = 1.52 \pm 0.00\), \(n_{\text{pellet}, \text{ref pellet}} = 1.47 \pm 0.01\), \(n_{\text{pellet}, \text{ref empty compartment}} = 1.44 \pm 0.00\)), indicating that this difference does not arise from a phase unwrapping effect (see paragraph on Phase unwrapping on p.113). Table 4.9 lists some refractive index values reported in the literature for solid HDPE samples. These values encompass the values extracted from the HDPE sheet using the TPS Spectra 3000 spectrometer and the TPS Spectra software: this demonstrates the accuracy of the spectrometer from TeraView.

**Table 4.3:** Refractive index values reported in the literature for solid HDPE samples.

<table>
<thead>
<tr>
<th>Sample morphology</th>
<th>Reported refractive index</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDPE sheet</td>
<td>1.54 ± 0.01</td>
<td>[210]</td>
</tr>
<tr>
<td>HDPE sheet</td>
<td>1.54</td>
<td>[211]</td>
</tr>
<tr>
<td>HDPE bar</td>
<td>1.54</td>
<td>[212]</td>
</tr>
<tr>
<td>50-mm thick slice from a HDPE cylindrical rod</td>
<td>1.51</td>
<td>[205]</td>
</tr>
<tr>
<td>6.17 mm thick HDPE sheet</td>
<td>1.53</td>
<td>[213]</td>
</tr>
</tbody>
</table>
The difference in the refractive index as observed for the sheet and the one for the pellet originates from the porous and therefore inhomogeneous nature of the pellet, with air voids and sometimes clusters of sample particles (see paragraph on Particle size and grinding process on p.105), creating local differences in the refractive index and leading to internal scattering of the terahertz beam ([208], [201]). When using the signal transmitted through a pellet of 360 mg of HDPE as the reference signal, some of this scattering is already accounted for, while when using the signal transmitted through the empty sample compartment as a reference, none of the scattering effect is taken into account. As a result, the refractive indices from pellets referenced against the 360 mg pellet of HDPE are closer to the real refractive index value of the solid sheet than indices from pellets referenced against the empty sample compartment.

While solid sheets appear to be the ideal sample morphology to accurately extract the required optical parameters, Equation 4.12 demonstrates that when the sheet is too thin, additional pulses resulting from multiple internal reflections on the sheet surfaces may start to be detected in the transmitted waveform, as illustrated in Figure 4.10a: such phenomenon is often called a Fabry-Perot or 'etalon' effect [215]. These subse-
sequent pulses collected later in time due to multiple internal reflections have a main peak which has the same sign as the main peak of the first transmitted pulse. This is due to the fact that the first collected pulse in transmission does not undergo any reflection, and that the second, third and subsequent collected pulses (the etalon effect) undergo reflections at the sheet - air interface, corresponding to a change from a larger to a lower refractive index value: such interface induces no phase shift of $\pi$ and therefore the main peak of the pulse does not change sign.

To observe the influence of such additional TD features on the frequency-dependent optical parameters extracted from the material, a 'corrected' waveform with no apparent etalon effect was designed by simply deleting any TD data within the range where the etalon feature was observed, and shifting any data above this range to lower scanning positions. Zero values were also added at the end of the waveform to obtain the same time length as the original waveform (zero padding). The refractive index and absorption coefficient for both waveforms were calculated with a Matlab programming code emulating the steps followed by the TPS Spectra software (yet, with corrections brought to the calculation of the absorption coefficient). Figures 4.10b and c indicate that the etalon effect introduces broad spectral bands at low frequencies (below 40 cm$^{-1}$).

Additionally, peaks and troughs are observed in the TD waveform transmitted through an empty and purged sample compartment (Figure 4.11a, grey): due to the overall response of the system, the detected signal is not a narrow Gaussian-like pulse as would theoretically be produced by a terahertz radiation generation device, but a distorted broader signal with negative and positive peaks. Therefore, when the solid sheet to be analysed is too thin, the signal corresponding to a single transmission through the sheet can overlap with the signal caused by multiple internal reflections and transmissions through the sheet, i.e. the signal resulting from the etalon effect. This overlap can change the position and amplitude of the maximum peak of the detected waveform. Due to the etalon effect, the refractive index and absorption coefficient extracted from the sample waveform would therefore be different than the refractive index and absorption coefficient intrinsic to the material of the sheet (see Equation 4.20).
\textbf{Figure 4.10: }a) TD waveforms with and without etalon effect and b) their associated frequency-dependent refractive index and c) absorption coefficient (no apodization) calculated using data collected from a historical parchment sheet from the reference collection of the ISH (sample ID: OR90, 172-µm thick). The arrow in a indicates the signal transmitted after internal reflections on the front and back surfaces of the parchment sheet.

To prove this, the sample waveform which would have been detected if there was no response from the system needs to be recovered. This is also called the \textit{Impulse Response Function (IRF)} and it is obtained by deconvolution [76, 141]. FD deconvolution consists of the inverse Fourier transform of the transmittance ratio (see Equation 4.2). Yet, due to division by the reference spectrum \( R \) in the calculation of the transmittance ratio, any noise in the sample waveform \( s \) passing on to the sample spectrum \( S \) is amplified.
Figure 4.11: a) Experimental TD waveforms from the reference (grey) and the historical parchment sheet from the reference collection of the ISH (black, sample ID: OR90); b) Associated deconvolved TD waveforms calculated using Equations 4.30 and 4.31, using HF=0.03 and LF=200.

To circumvent this, a filter $f$ can be applied to the sample waveform. A double Gaussian filter was used here, as recommended by Zeitler and Chen [76].

\[
IRF = FFT^{-1}[\frac{FFT(s)}{FFT(r)} \cdot FFT(f)]
\]  
(4.30)

\[
f(t) = \exp(-t^2/\text{HF}^2) - \exp(-t^2/\text{LF}^2)
\]  
(4.31)

A double-Gaussian deconvolution tool is already built-in in the TVL Imaging Suite software from TeraView, designed to process imaging data. However no such tool is built in the TPS Spectra software, designed to process spectroscopic data. A programming code was therefore developed to perform double-Gaussian deconvolution on data acquired from transmission spectroscopy, and is provided in Appendix C.

The reference and sample waveforms resulting from deconvolution are shown in Figure 4.11b: the signal resulting from the etalon effect (at 21.3232 mm) is no longer overlapping with the main transmission signal (at 21.0917 mm). The constant refractive index extracted from those waveforms using Equation 4.20 is equal to 1.50, compared to 1.53 using the detected waveforms shown in Figure 4.11a: the presence of an etalon effect overlapping with the signal transmitted once through the sheet introduces error in the calculation of the refractive index from a solid sheet.
4.1. Terahertz time-domain spectroscopy

<table>
<thead>
<tr>
<th>Sample concentration</th>
<th>Sample weight</th>
<th>Pellet weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1%</td>
<td>1.44 mg</td>
<td>360.4 mg</td>
</tr>
<tr>
<td>0.25%</td>
<td>3.6 mg</td>
<td>360.9 mg</td>
</tr>
<tr>
<td>0.5%</td>
<td>7.2 mg</td>
<td>361.8 mg</td>
</tr>
<tr>
<td>1%</td>
<td>14.5 mg</td>
<td>363.5 mg</td>
</tr>
<tr>
<td>2.5%</td>
<td>37 mg</td>
<td>369 mg</td>
</tr>
<tr>
<td>5%</td>
<td>75.8 mg</td>
<td>379 mg</td>
</tr>
<tr>
<td>10%</td>
<td>160 mg</td>
<td>400 mg</td>
</tr>
<tr>
<td>20%</td>
<td>360 mg</td>
<td>450 mg</td>
</tr>
</tbody>
</table>

Table 4.4: Weights of sample powder added to 1440 mg of HDPE powder in order to obtain different sets of triplicate pellets with different sample-to-HDPE mass ratios. All resulting pellets have the same bulk weight of HDPE as the reference pellet (i.e. 360 mg of HDPE).

4.1.4.3 Sample concentration

Pellets with different sample-to-HDPE mass ratios were prepared to investigate the influence of the sample concentration within a pellet on the optical parameters extracted with THz-TDS. The different mixing ratios are further detailed in Table 4.4.

**Absorbance spectrum**

Figure 4.12 displays absorbance spectra for pellets of bone black (a) and iron(II) sulfate heptahydrate (b) for different sample-to-HDPE mass ratios. It can be observed that the slope of the overall frequency-dependence increases with concentration for both amorphous (bone black) and crystalline (Fe(II)SO₄) samples, and that the amplitudes of the spectral features at ≃50 and 64 cm⁻¹ for Fe(II)SO₄ also increase with concentration. Additional spectral features observed in the spectrum of Fe(II)SO₄ with a mass ratio of 20% are not reproducible (i.e. are not systematically observed in each triplicate pellet) and their number and shape largely depend on the apodisation function used: they can therefore be regarded as spectral artefacts. Some materials are very strongly absorbent in the terahertz region, and absorb all the incident terahertz pulse even when diluted in HDPE with a mass ratio of 10%: no transmitted signal can therefore be detected. Using pellets with smaller mass ratios would enable the measurement of their spectral signatures. For iron sulfate heptahydrate, the spectral signature is still distinct down to a mass ratio of 0.5%, corresponding to 7.2 mg of iron sulfate. At 10% mass ratio, vermilion and minium pigments show similar absorbance values as iron sulfate heptahydrate and are therefore expected to reach a similar limit of detection.
Figure 4.12: Average absorbance spectra from triplicate pellets for a) bone black and b) iron sulfate heptahydrate sample sets, with different sample-to-HDPE mass ratios (20%, 10%, 5%, 2.5%; and 1%, 0.5%, 0.25%, 0.1% in the inset). Error bars indicate their associated standard deviation.

Refractive index and absorption coefficient

Refractive indices and absorption coefficients are parameters intrinsic to the material and should therefore not vary when increasing the concentration of material in the pellet while keeping the mass of HDPE constant. However, Section 4.1.4.2 indicates that the porous nature of the pellet induces undesirable scattering effects, which have a noticeable impact on the extracted refractive index values. HDPE is a plastic material with very good thermal expansion properties, therefore its volume concentration is likely to influence expansion properties of the pellet, which would have an impact on its resulting porosity: pellets with a low sample concentration (i.e. a large concentration of HDPE) will behave like pellets of pure HDPE and are likely to show good expansion properties, while pellets with a large sample concentration (i.e. a low concentration of HDPE) are expected to show poorer expansion than a pellet of pure HDPE with the same mass. The constant refractive index values reported in Table 4.5 show an apparent increase with increasing sample concentrations (and decreasing HDPE concentrations). This increase is likely due to a reduced porosity of the pellet, leading to reduced scattering effects on internal voids. It should be noted that for a sample concentration equal to or below 2.5%, the pellet thickness is very close to or identical to the thickness of the reference pellet: the measured thickness difference $d$ can therefore become inaccurate and may lead to erroneous refractive index values when using Equation 4.20.
4.1. Terahertz time-domain spectroscopy

<table>
<thead>
<tr>
<th>Sample concentration</th>
<th>Bone black</th>
<th>Fe(II)SO₄, 7H₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5%</td>
<td>2.76 ± 0.09</td>
<td>2.07 ± 0.19</td>
</tr>
<tr>
<td>5%</td>
<td>2.11 ± 0.02</td>
<td>1.84 ± 0.05</td>
</tr>
<tr>
<td>10%</td>
<td>2.17 ± 0.01</td>
<td>1.96 ± 0.03</td>
</tr>
<tr>
<td>20%</td>
<td>2.29 ± 0.03</td>
<td>2.23 ± 0.07</td>
</tr>
</tbody>
</table>

Table 4.5: List of the average constant refractive indices (see Equation 4.20) and associated standard deviations from triplicate pellets of bone black and iron(II) sulfate heptahydrate with different sample-to-HDPE mass ratios.

4.1.4.4 Dynamic range

The SNR in the FD is frequency-dependent: Figure 4.13a indicates that the signal transmitted through an empty sample compartment reaches its maximum value between 10 - 20 cm⁻¹, and then decreases with frequency until reaching a noise floor. To determine the working spectral range of the terahertz TD spectrometer and be able to distinguish spectral features arising from the vibrational behaviour of the sample from spectral artefacts, the notion of dynamic range is sometimes used [100, 194, 216]:

\[
\log(Dynamic\ Range(\nu)) = -\log\left(\frac{E_{\text{noise}}(\nu)}{E_{\text{reference}}(\nu)}\right) \quad (4.32)
\]

\(E_{\text{noise}}\) corresponds to the root-mean-square of the noise floor. In the non-noisy working spectral range, we have:

\[E_{\text{sample}}(\nu_i) > E_{\text{noise}}(\nu_i)\quad (4.33)\]

Therefore, using Equations 4.2 and 4.32, the working spectral range becomes the spectral range within which the absorbance values are smaller than the logarithm of the dynamic range values. The working spectral range is also often defined as the spectral range within which the absorption coefficient values \(\alpha(\nu)\) are smaller than \(\alpha_{\text{max}}\) [100, 194, 216]:

\[
\alpha_{\text{max}}(\nu) = \frac{2}{d} \ln \left( Dynamic\ Range \cdot \frac{4n(\nu)}{(n(\nu) + 1)^2} \right) \quad (4.34)
\]

The working spectral range therefore depends on the sample under investigation (its constituting material(s) and the concentration of its absorbing material(s)). The dynamic range, however, solely depends on the set-up used and on environmental conditions.
When calculating the dynamic range, care should be taken when selecting \( E_{\text{reference}} \) and \( E_{\text{noise}} \). If the absorbance spectrum of the sample was calculated using a reference signal transmitted through the empty chamber, then for the calculation of the dynamic range, \( E_{\text{reference}} \) should correspond to \( E_{\text{empty chamber}} = E_{100\% \text{transmission}} \). If the absorbance spectrum was calculated using a reference signal transmitted through a 360-mg pellet of HDPE (as is often the case when analysing pellets), then \( E_{\text{reference}} \) should correspond to \( E_{\text{HDPE pellet}} \). For \( E_{\text{noise}} \), the noise floor of \( E_{\text{reference}} \) was used as recommended in [100]. The background noise of the spectrometer, corresponding to the signal detected when the incident terahertz beam is blocked with a piece of aluminium, shows lower values than the noise floor of \( E_{\text{reference}} \) (see Figure 4.13a) and has been found to lead to overestimated values of the dynamic range. Figures 4.13b and c present experimental data from a pellet made of 10% Prussian blue, with \( E_{\text{reference}} = E_{\text{HDPE pellet}} \). The noise floors of the signals transmitted through the sample pellet and the reference pellet show similar values from 120 cm\(^{-1}\) onwards: therefore, \( E_{\text{noise}} \) was defined as the root-mean-square of the noise floor from \( E_{\text{reference}} \) between 120 - 275 cm\(^{-1}\). The resulting dynamic range (grey curve in Figure 4.13c) spans from 5 to 120 cm\(^{-1}\) and peaks at the maximum of 7.9 absorbance units at 16 cm\(^{-1}\) (i.e. 79 dB at 0.48 THz). This dynamic range crosses the absorbance curve from Prussian blue at 73 cm\(^{-1}\). The working spectral range for a 10% Prussian blue pellet is therefore between 5 and 73 cm\(^{-1}\) and any data displayed above this frequency reaches the detection limit and is considered as unreliable.

### 4.1.5 Apodization

A terahertz TD signal, like all TD signals, can be considered as a continuous sum of sine and cosine waves with different amplitudes and frequencies, accessible using Fourier transform. Sine and cosine waves are infinitely repeating periodic signals. The waveform measured with a terahertz TD spectrometer is thus only a section of the full periodic signal between the observation times \( t_{\text{start}} \) and \( t_{\text{end}} \) (e.g. \( t_{\text{start}} = 19 \text{ mm} \) and \( t_{\text{end}} = 26 \text{ mm} \) in Figure 4.14). The measured waveform \( s_{\text{measured}} \) is therefore in essence a different infinite signal than the original signal \( s_{\text{periodic}} \) it has been sampled from, and can be described as:

\[
s_{\text{measured}}(t) = s_{\text{periodic}}(t) \cdot w(t)
\]  

(4.35)
4.1. Terahertz time-domain spectroscopy

Figure 4.13: a) FD signal (average of 6 consecutive signal acquisitions) collected for an empty sample compartment (100% transmission), for a reference pellet of HDPE, and in the absence of transmitted signal, using a piece of aluminium blocking the incident beam (0% transmission); b) FD signal collected for a reference pellet of HDPE and a pellet of 10% Prussian blue; c) Absorbance spectrum of a pellet of 10% Prussian blue and dynamic range of the TPS Spectra 3000, TeraView.

with $w(t)$ a rectangular window function (also called Boxcar function) such that:

- for $t < t_{\text{start}}$, $w(t) = 0$
- for $t \in [t_{\text{start}} - t_{\text{end}}]$, $w(t) = 1$
- for $t > t_{\text{end}}$, $w(t) = 0$

$s_{\text{measured}}$ is called the apodized signal. The FD signal $S_{\text{measured}}$ of the measured TD waveform $s_{\text{measured}}$ and displayed by the TPS Spectra software is therefore not the Fourier transform of the true signal TD $s_{\text{periodic}}$, but is the Fourier transform of the apodized signal:

$$S_{\text{measured}}(\nu) = S_{\text{periodic}}(\nu) \ast W(\nu)$$

(4.36)

To obtain a FD signal which is a good estimate of the true periodic signal $s_{\text{periodic}}$, the window function should resemble a Dirac function in the FD (i.e. the neutral element of the convolution product), equivalent to a sinusoidal function infinite in time in the TD [217]. Therefore, the use of a rectangular function $w(t)$ as described above would not lead to a faithful representation of the true signal $s_{\text{periodic}}$. However, window functions finite in time and resembling the sinusoidal function exist: the Blackman-Harris function is an example of such a window function and has often been used for THz-TDS [83, 194, 218].
Figure 4.14: Original and apodized TD waveforms transmitted through a pellet made of 10% Prussian blue: the apodized waveform results from the product of the original waveform and the Blackman-Harris 3-term window function.

Figure 4.14 shows the shape of the original waveform collected from a pellet made of 10% Prussian blue, together with the apodized waveform resulting from the product of the original waveform with a Blackman-Harris 3-term window function. Figure 4.15 shows the associated FD absorbance spectra. In the latter figure, a sharp and intense spectral band is observed at $\simeq 71 \text{ cm}^{-1}$ in the absorbance spectrum of the waveform apodized with the Blackman-Harris function, but is not clearly observed in the absorbance spectrum of the original waveform. It is important to note that this intense band was observed for triplicate pellets of 10% Prussian blue, and was further observed when preparing and analysing another set of triplicate pellets with 10% Prussian blue. Yet, Section 4.1.4.4 indicates that at this frequency, the limit of detection is reached and no signal intrinsic to the vibrational behaviour of the material can be clearly distinguished from noise any more.

To evaluate whether this spectral band is a spectral information arising from the vibrational behaviour of Prussian blue or simply is a spectral artefact arising from the apodization process, the amplitudes of the absorbance values in this frequency range therefore need to be reduced so that they do not reach the limit of detection. In accordance with Section 4.1.4.3, a new set of pellets was made with only 1% Prussian blue. Figure 4.15 reveals that although the spectral range 70 - 100 cm$^{-1}$ presents larger absorbance values than at other frequencies, the intense and narrow band at $\simeq 71 \text{ cm}^{-1}$ is no longer observed for either the spectrum of the apodized waveform or the spectrum of the original waveform.
It can therefore be concluded that this band is a spectral artefact due to the apodization process.

### 4.1.6 Conclusions

- The refractive index extracted from a sample material diluted in HDPE and pressed into a pellet is likely to be lower than the true refractive index of the solid material, due to potential scattering effects occurring on air voids within the porous pellet. However, it provides a good estimate of the true index and it can be quantitatively compared to refractive indices from other materials, as long as the same sample concentration and pellet preparation is used for all materials. Refractive indices from pellets cannot, however, be quantitatively compared to refractive indices from solid sheets of paper or parchment.

- A mass ratio of sample-to-HDPE of 10% was used as it enabled clear observation of most spectral features from materials without reaching the detection limit in relation to the dynamic range.

- The sample−HDPE powder mix must be ‘gently’ (*i.e.* low applied pressure) blended using a pestle and mortar to prevent the formation of large particle clusters which could lead to additional scattering effects, and influence the precision.

**Figure 4.15:** Absorbance spectra from 2 sets of triplicate pellets made of 10% and 1% Prussian blue: the light grey lines correspond to absorbance values calculated from the Fourier transforms of the original waveforms (framed within a Boxcar, rectangular, window), and the black lines correspond to absorbance values calculated from the Fourier transforms of waveforms windowed with a Blackman-Harris 3-term function.
and accuracy of the extracted optical parameters.

- The use of the Blackman-Harris 3-term apodisation function enables to produce a FD signal approaching the Fourier-transform of the true signal. This apodisation function can however introduce additional spectral artefacts when absorbance values are close to the limit of detection of the set-up.

- Asymptotic behaviour at low frequencies should be treated with caution, as it may arise from the default phase unwrapping process used by the software. A constant refractive index is often a good point of reference to demonstrate accuracy of the observed frequency-dependent index, as it involves little signal processing.

- Multiple internal reflections of the terahertz pulse observed for thin sheets of paper and parchment introduce broad spectral bands at low frequencies, unrelated to the vibrational behaviour of the material. The internal reflections also influence the maximum amplitude and time delay of the main pulse from the transmitted TD waveform, and therefore introduce errors in the calculation of the constant and frequency-dependent refractive index and absorption coefficient from a thin sheet. To extract optical parameters from different sheets which can be quantitatively compared, deconvolution of the transmitted TD signals should be performed before calculating the constant refractive index.

### 4.2 Terahertz time-domain imaging

The quality of a spatial image in relation to the captured object can be characterized by two main criteria, regardless of the imaging device used: the image resolution and the SNR, or 'detectability'.

#### 4.2.1 Lateral spatial resolution

We define spatial resolution as the minimum distance between two separate features on a captured object leading to discernible features in the image of this object. If the human eye is employed as the sensing element making the decision of whether or not two features are discernible, a theoretical law to estimate the spatial resolution should be based on the performances of the imaging system in the visible range. The
4.2. Terahertz time-domain imaging

Rayleigh criterion is the most widespread theoretical law defining spatial resolution in the visible range. It is based on the image acquired from measurements of the visible light intensity emitted by two separate incoherent point sources. The 2D distribution of the light intensity from each point source (also called the Point Spread Function (PSF)) is defined in spherical coordinates as [193]:

\[ I(\theta) = I(0) \cdot \left[ \frac{2 \cdot J_1(k \cdot a \cdot \sin(\theta))}{k \cdot a \cdot \sin(\theta)} \right]^2 \] (4.37)

with \( J_1 \) representing the Bessel function of order 1, \( k = 2 \pi / \lambda \) the wavenumber, \( a \) the aperture radius of the imaging device, and \( \theta \) the angular position of the image point. Its optical image represents concentric Airy patterns [193]. The limit of resolution is defined by Rayleigh as the distance between the point sources when the maximum intensity of the PSF of one point source (i.e. the centre of an Airy disc) coincides with the first minimum of the PSF of the other point source (i.e. the first dark Airy ring). In other terms, the angular resolution is defined as the angular radius of the first dark Airy ring, also corresponding to the position of the first minimum of the first-order Bessel function (equal to 3.83).

\[ k \cdot a \cdot \sin(\theta) = 3.83 \] (4.38)

which can be re-written as:

\[ \sin(\theta) = \frac{3.83 \lambda}{2\pi a} = \frac{1.22 \lambda}{2a} \] (4.39)

The angular position \( \sin(\theta) \) can be described in terms of the radial position \( r \) and the distance between the object and the detector of the imaging system, i.e. the focal length \( f \). With \( D \) the aperture diameter of the imaging system, the limit of resolution can further be defined as:

\[ r_{Rayleigh} = \frac{1.22 \cdot f \cdot \lambda}{D} \] (4.40)

At the limit of resolution defined by Rayleigh, the composite intensity distribution, sum of the PSF of the two point sources, exhibits a central depression corresponding to a decrease in intensity of \( 1 - \frac{8}{\pi^2} = 19\% \) [193].
Figure 4.16: Gaussian functions representing the 1D PSF from two point sources separated with a distance equal to a) the FWHM of the Gaussian curve, b) the double of the standard deviation of the Gaussian curve (the two curves intersect at their inflection point) and c) the standard deviation of the Gaussian curve.
The selection of the Rayleigh criterion to define the spatial resolution of a terahertz TD imaging system raises the following issues:

- It refers to the detection of light intensity, better defined as irradiance, while detection of a terahertz pulse with a photoconductive antenna is a measure of the amplitude of electromagnetic field. Yet, the amplitude of electromagnetic field is proportional to the square root of the irradiance.

- It considers irradiance from incoherent point sources [193], while terahertz pulses reflected from the surface of an object can be coherent. The PSF of such point sources may therefore be more accurately defined by a Gaussian function, rather than a Bessel function with zeros.

The Buxton’s criterion uses the amplitude of electromagnetic field, rather than irradiance, and defines the limit of resolution as the distance between point sources at which the two PSF intersect at their inflection points [219]. If the PSF are better described by Gaussian function, as discussed above, this inflection point corresponds to the standard deviation $\sigma$ of the curve. Figure 4.16 summarizes different configurations of overlapping PSF depending on the distance separating the point sources.

If the PSF is the image collected of a single dot, it is difficult to accurately measure it, as the SNR of the image of the dot will be dependent on the width of the dot in relation to the pixel size, and on its position in relation to the pixel grid (see Section 4.2.3). Assuming that the PSF obtained with a terahertz TD imaging set-up is a Gaussian function, the image of a line, called a Line Spread Function (LSF), will also be a Gaussian function, and will have the same standard deviation. Moreover, the response from a line is the 1st derivative of the response from an edge. The advantages of an edge is that it can easily be manufactured and imaged with a good SNR.

The edge of a flat piece of aluminium held at $\approx 20$ cm above the optical table was imaged in reflection mode with the T-Ray 4000 (Picometrix), with a step size of 0.1 mm. The signal collected from the edge, also called the Edge Response Function (ERF), corresponds to the maximum amplitude of the electromagnetic pulse reflected from the edge and is shown in Figure 4.17 along both the $x$ and $y$ scanning directions.
Figure 4.17: ERF, LSF and MTF calculated from the maximum amplitude of the terahertz electromagnetic pulse reflected from an edge (a flat piece of aluminium) and imaged with the T-Ray 4000 set-up, Picometrix.

The LSF was calculated from the 1st derivative of the ERF, using third polynomial Savitzky-Golay smoothing and a window size of 10 points: this improved the SNR of the Gaussian shape curve, while keeping its width and height. The curve was then fitted with a Gaussian function to accurately extract its standard deviation. It can be concluded that, for a scanning step size of 0.1 mm, two lines are resolved with the T-Ray 4000 imaging set-up in reflection if they are, according to the Buxton’s criterion, at least spaced by 720 µm along the x direction and 640 µm along the y direction.

To observe how the composite distribution of two LSFs evolves for different separating distances, i.e. how the image contrast of two lines evolves with different line spacings, the MTF is produced from the amplitude of the Fourier transform of the LSF [220], and represents the drop in contrast for different line spacings, expressed as line pairs per mm (a line pair corresponds to a full ‘striped’ and a full ‘white’ line shown in Figure 4.16b).
The relationship between the number of line pairs per mm, $f$, and the line width $x$ is:

$$f = \frac{1}{2x}$$  \hspace{1cm} (4.41)

The MTF displayed in Figure 4.17 indicates that for a line pair of $f = 1/2 \sigma$ (equal to 0.69 and 0.78 in the $x$ and $y$ directions respectively), the contrast value reaches 0%.

To confirm the limit of resolution of the T-Ray 4000 imaging set-up (Picometrix), let us consider the image of an object with line patterns of different frequencies, using the same parameter (the maximum amplitude of the terahertz pulse reflected from the front surface of the object) and the same scanning step size as above (0.1 mm): the limit of resolution is reached when the image contrast between two lines drops to 0%. Figure 4.18 shows such image for a target with patterns made of a material very reflective in the terahertz range (e.g. gold) on a low-absorbing background material ($z$-cut quartz). This image has been normalized so that the maximum signal from the target is set to 100% and the minimum signal is set to 0%.

The image contrast recorded for several line spacings is further reported in Figure 4.19. It can be observed that lines separated by a distance of 700 to 600 µm are still resolved (central dip of $\approx 20\%$ for a line spacing of 700 µm). The limit of resolution
appears to lie at 340 µm and 260 µm along the x and y directions respectively, when the contrast at the central decrease is equal to zero. This limit of resolution is far less than the value predicted with the knife-edge measurement using the Buxton’s criterion, i.e. reading the standard deviation of the LSF of the electromagnetic field amplitude $\sigma$ or its associated MTF.

It can therefore be concluded that the use of a knife-edge measurement to determine the LSF or the MTF of the incident terahertz beam and of this information to estimate the limit of resolution cannot be applied to terahertz TD imaging systems, as little is known about the spatial distribution of a pulsed terahertz beam: the use of a target with frequential patterns as shown in Figure 4.18 does not require any assumption on the nature of the pulsed terahertz beam and seems therefore more appropriate to determine the limit of resolution.

![Figure 4.19](image)

**Figure 4.19:** Profile of the normalized maximum amplitude of the reflected pulse displayed in Figure 4.18 along a line perpendicular to the line pattern of different elements. The scanning step size is 0.1 mm.

It has been demonstrated above that the use of the LSF of the electromagnetic field amplitude $\sigma$ or its associated MTF combined with a criterion such as the Buxton criterion does not accurately predict the terahertz image from close parallel lines, and should therefore not be used to define the resolving power of a terahertz TD imaging set-up. Using the terahertz image from a test target to define resolution, although more realistic, still requires the use of a criterion to set the limit of resolution. While the limit set by Rayleigh of a central depression in amplitude of 19 % may still be used, it should be reminded that this limit was determined satisfactory by Rayleigh because it matched what the eye could detect through telescopes or microscopes. Lord Rayleigh
himself described his criterion in the following manner: "We may consider this to be about the limit of closeness at which there could be any decided appearance of resolution. [...] This rule is convenient on account of its simplicity; and it is sufficiently accurate in view of the necessary uncertainty as to what exactly is meant by resolution. Perhaps in practice somewhat more favourable conditions are necessary to secure a resolution that would be thought satisfactory" [221]. Today, as there are computational image processing and statistical techniques able to emphasize contrast and improve the distinction between close features (see Section 5.6.1), it seems more appropriate to set the limit of resolution as a central depression in amplitude of 0 %.

A 1951 USAF resolution test target (Product ID: R3L3S1P and R3L3S1N, Thorlabs Inc) was used to determine the limit of resolution of the TPS Spectra 3000 from TeraView with the RIM or the gantry system. Lasch and Naumann [222] advise to work with normalized images of the target to determine the limit of resolution. Yet, scanning the entire target with a resolution of 0.1 mm would be a time-consuming process and would generate large files. Therefore, an alternative was to scan the full frame containing groups 0 and 1 with a larger step size (e.g. 0.5 or 0.2 mm), and to normalize the image obtained from the maximum amplitude of the main reflected pulse so that the maximum signal from the target, corresponding to the signal reflected from the plain square between groups 0 and 1, is set to 100%. A subregion of interest is then scanned again with a finer resolution (0.10 or 0.05 mm), and is normalized using the same maximum and minimum signals extracted from the scan of the full frame, as indicated in Figure 4.20. The amplitude profile of the signals along a line across parallel line pairs is displayed in Figure 4.21. This profile indicates that for line pairs from group 1 and element 4, there is no longer a clear central decrease and signals from the three line pairs are not resolved. Therefore, the limit of resolution is at a line spacing of 177 µm for a scanning step size of 0.1 mm with the RIM.

The same scanning procedure and imaging process was repeated for the gantry system connected to the TPS spectra 3000, and the results are shown in Figures 4.22 and 4.23. The limit of spatial resolution is at a line spacing corresponding to Group 1 and Element 1, i.e. 250 µm, when no central decrease is observed between the signals reflected from the glass lines.
Figure 4.20: a) Photograph of the USAF 1951 target made with metallic lines deposited on a flat glass substrate and b-e) terahertz images of the target produced with the maximum amplitude of the main reflected pulse, using the RIM and different scanning step sizes.
4.2. Terahertz time-domain imaging

Figure 4.21: Profile of the normalized maximum amplitude of the reflected pulse displayed in Figure 4.20c along a line perpendicular to the line pairs of Group 1. The scanning step size is 0.1 mm.

Figure 4.22: a) Photograph of the USAF 1951 target made with a metallic negative pattern deposited on a flat glass substrate and b-c) terahertz images of the target produced with the maximum amplitude of the main reflected pulse, using the gantry system in reflection with different scanning step sizes, b) full frame and c) Group 0.

The imaging performances of the different set-ups and accessories are summarized in Table 4.6. The RIM connected to the TPS Spectra 3000 is the imaging set-up which achieves the best performances in terms of lateral spatial resolution: this set-up provides a lower loss of signal due to the use of ellipsoidal gold mirrors to focus the beam, instead of HDPE lenses, and to a nitrogen purged environment reducing absorption from ambient water vapour.
Figure 4.23: Profiles of the normalized maximum amplitude of the reflected pulse displayed in Figure 4.22c along a line perpendicular to the line pairs of Group 0, as well as Group 1 (figure not displayed). The scanning step size is 0.1 mm.

Table 4.6: Summary of the lateral spatial resolution for different terahertz TD imaging setups and accessories, using images produced with the maximum amplitude of the reflected pulses and a scanning step size of 0.1 mm.

<table>
<thead>
<tr>
<th>RIM, TPS Spectra 3000 Gantry system, TPS Spectra 3000</th>
<th>T-Ray 4000 (Ellipsoidal mirrors)</th>
<th>(f-number: 1.2)</th>
<th>(f-number: 0.66)</th>
</tr>
</thead>
<tbody>
<tr>
<td>180 µm</td>
<td>250 µm</td>
<td>x: 340 µm</td>
<td>y: 260 µm</td>
</tr>
</tbody>
</table>

These limits of resolution can be further reduced by reducing the scanning step size, as illustrated in Figures 4.20 and 4.24: the limit of resolution for images produced with the maximum amplitude of the reflected pulse corresponds to Group 1 and Element 6, i.e. a line spacing of 140 µm with a scanning step size of 0.05 mm, instead of 177 µm for a step size of 0.1 mm.
Figure 4.24: Profiles of the normalized maximum amplitude of the reflected pulse displayed in Figure 4.20e, and the amplitude of the FD signal at 2 THz and 2.58 THz displayed in Figure 4.25c and f, along a line perpendicular to the line pairs of Group 1. The scanning step size is 0.05 mm.

The limit of resolution also depends on the parameter of the terahertz signal used to produce the image. Equation 4.40 suggests that the limit of resolution is proportional to the wavelength $\lambda$ of the incident beam, and is therefore inversely proportional to its frequency. Using terahertz signals in the FD, obtained from Fourier transform of the TD waveforms, it is therefore possible to improve spatial resolution and thus the quality of the image by selecting high frequency signals to produce the image. Figure 4.25 shows the images of the target produced from the amplitude of the FD signals at 2 and 2.58 THz, with a scanning step size of 0.05 mm.
The profiles shown in Figure 4.24 indicate that all elements of Group 1 are now resolved when using the signal at 2 THz. However, at 2.58 THz, Figure 4.25d and f shows a lot of noise (i.e. the signal from the background shows an intensity similar to the signals from the reflective lines), and this noise is also evident in the profile in Figure 4.24. This is because the performance of the set-up is maximum around 0.48 THz, and then progressively declines with increasing frequency, as discussed in Section 4.1.4.4 and evidenced in Figure 4.13. In reflection mode, it therefore seems that the dynamic range lies below 2.5 THz. The limit of resolution reachable with a step size of 0.05 mm and a frequency of 2 THz corresponds to Element 3 in Group 2, as illustrated in Figure 4.26. Therefore, the minimum reachable spatial resolution at surface is 100 µm with the RIM by TeraView: finer patterns may be resolved with a higher frequency (e.g. 2.58 THz), but the image becomes difficult to interpret due to the poor SNR.

So far we discussed spatial resolution at surface and observed that selecting higher frequency components of the signal enabled better resolution of the terahertz image. Yet, we will later show that the absorbance of paper and parchment sheets increases with frequency. The signal transmitted through a stack of paper sheets is therefore more strongly attenuated at higher frequencies, and this attenuation increases with the number of sheets, as illustrated in Figure 4.27. Therefore, when imaging the surface of a USAF target covered with paper sheets, the signal reflected on the surface of the target and reaching the detector is more attenuated at high frequencies than at low frequencies. As a result, the SNR of the image at 2 THz from the surface of the target covered with a stack of paper sheets is poorer than for the target alone. This can be observed by comparing Figures 4.28a and b.

In conclusion, the minimum limit of resolution reachable with the imaging set-ups available during the doctoral project is approximately 100 µm at surface, with a step size of 0.05 mm and using the high frequency components of the TD signal. When imaging through sheets of a document, the SNR will be more or less deteriorated depending on the absorption properties of the material constituting the sheets, their thickness and number. As for the image obtained at 2.58 THz from two features spaced by 100 µm at surface, the image at 2 THz from two features spaced by 100 µm and covered with sheets may become difficult to interpret due to the poor SNR.
While ‘resolution’ refers to the notion of being able to distinguish between two close features, it is not the only possibility of defining the quality of an image. Indeed, a feature finer than the smallest attainable resolution (*i.e.* 100 μm) may still be observed in a terahertz image: this depends on the SNR of the image, and is further discussed in the Section below.
Figure 4.26: Profiles of the amplitude of the FD signal at 2 THz and 2.58 THz displayed in figure 4.25a and d, along a line perpendicular to the line pairs of Group 2. The scanning step size is 0.05 mm.

Figure 4.27: THz-TDS in transmission of a stack of rag paper sheets with an increasing number of sheets: a) TD waveforms and b) associated Fourier-transform spectra.

Figure 4.28: Terahertz images of the first elements of Group 1 on the USAF 1951 target, a) alone and b) under 2 sheets of rag paper, produced with the gantry system in reflection using the amplitude of the FD signal at 2 THz and a scanning step size of 0.05 mm. Images were acquired subsequently on the same day, and normalized.
4.2.2 Signal-to-noise ratio

As shown in Figure 4.29, the reference TD signal reflected from a gold coated mirror acquired with the TPS 3000 equipped with a RIM has a FWHM of 0.33 ps, while the reference TD signal reflected from a flat aluminium sheet acquired with the TRay-4000 has a FWHM of 1.17 ps\(^1\). It follows that images acquired with the TPS 3000 and presented in this thesis have a finer depth resolution than those acquired with the TRay-4000. It also follows that the spectral bandwidth is larger for the TPS 3000 (\(\simeq 3\) THz) than for the TRay 4000 (\(\simeq 1\) THz). The SNR is described here as the maximum value of the dynamic range, calculated as calculated for Figure 4.13. The SNR that could be obtained on the day of acquisition with the TRay 4000 was \(\simeq 19\) dB, as opposed to \(\simeq 54\) dB for TPS 3000. Images acquired with the TPS 3000 and presented in this thesis therefore have a better SNR than those acquired with the TRay 4000\(^2\).

4.2.3 Detectability

A pixel is the basic unit of the image: by definition, the spatial resolution cannot be smaller than the pixel size. Its size is limited by the size of the detector. Contrary to imaging devices equipped with a camera consisting of an array of sensors (\(e.g.\) CCD, CMOS, MCT), detection in terahertz TD set-ups used in this study is performed through the use of a photoconductive antenna with a detection gap of a few square microns and a motorized stage enabling step-by-step scanning of the object as described above. The pixel size is therefore equal to the step size of the scanning stage.

While a pair of two separate features on a captured object may not be separable with an imaging device due to their small separation width, a single feature of a width smaller than the spatial resolution, or even smaller than the pixel size, may still be detectable with the imaging device.

---

\(^1\)Software tools developed by TeraView Ltd set the sign of the main reflected pulse as positive, while it is set as negative for data acquired with the set-up from Picometrix. This is simply a convention.

\(^2\)It should be noted that this is not a definitive comparison of the set-ups from TeraView and Picometrix. In fact, Irl Durling, Picometrix, indicated that the SNR of the TRay 4000 in reflection and normal incidence is often around 50 dB. A technical report submitted in the framework of the CHARISMA project [161] indicates that a maximum SNR of 35 dB could be obtained with the same TRay 4000 set-up as used for our experiments, averaging over a large number of waveforms. Bianca Jackson, University of Reading, further reported that the set-up had suffered from optical pump power loss during its use on an archaeological site in Turkey. She also pointed out that periodic signals after the main pulse are due to resonances in the beam splitter pellicle, and that their subtraction using the background noise signal could potentially improve the SNR.
Figure 4.29: a) Signal reflected from an aluminium sheet and acquired in reflection mode with normal incidence using the T-Ray 4000, Picometrix; b) Signal reflected from a gold coated mirror and acquired in reflection mode with an oblique incidence using the TPS Spectra 3000 equipped with a RIM; c) and d) FD dynamic range obtained from the fast Fourier transform of a) and b), respectively, and divided by the root mean square of the values between 1 - 1.7 THz for c) and between 3 - 10 THz for d).

Figure 4.30 illustrates that features A and B are not resolved, *i.e.* cannot be discerned, but the same feature alone (*i.e.* feature C or D) is detectable. Two parameters influence the detectability and contrast of the feature: the position of the feature with respect to the pixel grid or scan steps, and the difference in optical properties between the material of the feature and the material of its surrounding background. As shown in Figure 4.30, the feature is more likely to be detected and show a good image contrast if it is fully inscribed within a single pixel, rather than overlapping two or more adjacent...
4.2. Terahertz time-domain imaging

pixels. The same figure also displays features of the same width and on the same background, but with a different optical property (e.g. different performances in absorbing or reflecting the incoming signal): a large contrast in optical property between the materials of the feature and its surrounding background leads to an intense signal collected from the feature and a weak signal collected from the background, resulting in a good SNR and therefore to an improved detectability and contrast.

![Figure 4.30](image)

**Figure 4.30:** Schematic diagram highlighting the influence of several parameters on the resolution and detectability of object features: the width of the feature in relation to the pixel size, the position of the feature in relation to the pixel grid and the optical property of the feature in relation to the optical property of the background.

When using a scanning step size (i.e. a pixel size) half the width of the ink line, there is always a pixel which is fully inscribed within the line and faithfully renders the signal from the ink, so shows the maximum contrast, regardless of the position of the line with regards to the pixel grid (see Figure 4.31). Using a scanning step size below half the width of the ink line does not improve the contrast of the inscriptions, as shown in Figures 4.31 and 4.32a and b, even for ink lines showing little difference in optical properties with the surrounding support. However, selecting a smaller step size enables us to render the width of ink lines more accurately, as illustrated in Figure 4.31. Considering that a small scanning step size leads to longer acquisition times and larger files, if the goal of scanning is to read the inscription, it may be sufficient to use
a scanning step size approximately half the width of the inscription. In practice, when
trying to read through a document with a terahertz TD set-up, it can be recommended
to start to image a small area with a larger step size (e.g. 0.5 mm), and to progressively
reduce the scanning step size until there is no change in contrast observed in the in-
scription: the corresponding scanning step size can then be used to image a larger area
of the document.

Figure 4.31: Schematic diagram highlighting the influence of the size of the scanning step size
\( a \) (i.e. pixel size) with respect to the width of the ink line \( d \) on the image contrast.

4.2.4 Time measurement uncertainty

As described in Section 3.1.3.1, a time-domain waveform is constituted of 4096 col-
lected time points, corresponding to the 4096 steps that the optical delay line marks
over 7 mm. The associated time resolution is 0.012 ps. To reduce the file size, only
1 every 4 time points may be saved, resulting in a waveform of 1024 time points, and
a time resolution of 0.046 ps. During the translation of the optical delay line, any jit-
ter or drift may cause uncertainty in the measurement of the time of collection of the
signal. To determine this uncertainty, a flat gold coated mirror was scanned with the
RIM and the time of collection of the maximum amplitude of the waveforms reflected
from it was measured. Figure 4.33 shows that this maximum can be collected between
the 1970\(^{th}\) and the 1977\(^{th}\) time point for a waveform constituted of 4096 time points
(Figure 4.33\(a\)), and between the 496\(^{th}\) and the 495\(^{th}\) time point for a waveform of 1024
time points (Figure 4.33\(b\)). The time measurement uncertainty is therefore equal to
\[ \approx \pm 3.5 \times 0.012 \text{ ps} = 0.042 \text{ ps} \text{ or } \approx \pm 1 \times 0.046 \text{ ps} = 0.046 \text{ ps}, \text{ and can be rounded up to} \]
0.05 ps.
As discussed in Section 5.3.2, terahertz images of inscriptions and topographic features at the surface of a support can be produced using the time of collection of the pulses reflected from these inscriptions and features. It follows that any topographic feature at the air-front surface interface smaller than 15 \( \mu \text{m} \) would result in a difference in time of collection between the pulses reflected from the surface with and without topographic features smaller than the time measurement uncertainty of 0.05 ps. Consequently, such features cannot be systematically detected nor reliably attributed to topographic features intrinsic to the document.

**Figure 4.32:** a) Photograph of a letter typewritten with Prussian blue ink and dated July 1946; b and c) associated terahertz images produced with the RIM using the maximum amplitude of the main reflected pulse and a scanning step size of 0.2 mm and 0.075 mm, respectively.
Figure 4.33: Terahertz images produced using the time of collection of the maximum amplitude of the waveforms reflected from a flat gold coated mirror, scanned with the RIM. a) Waveforms constituted of 4096 time points and b) 1024 time points.

4.2.5 Conclusions

• The lowest spatial resolution reachable with the TPS Spectra 3000 equipped with the RIM is 100 $\mu$m. Yet, this value can only be reached if imaged materials lead to a good contrast, i.e. materials with reflection properties very different to those of its surrounding support (such as metal on glass - like the USAF target - or metal on paper). When imaging ink lines on paper or parchment, the contrast would be poorer and the image of ink lines spaced by 100 $\mu$m would become difficult to interpret.

• When imaging an ink line with optical properties similar to those of the surrounding support, the scanning step size should be equal to half the width of the ink line to increase the likelihood to observe the ink line in the terahertz image. Such inscriptions are less likely to be revealed with the T-Ray 4000 set-up used in our study, due to its lower SNR compared to the TPS Spectra 3000.

• When imaging an ink line with optical properties very different to those of the surrounding support, the scanning step size can be equal to or even larger than the width of the ink line, provided that it remains equal to or lower than half the space between ink lines, to increase the likelihood to discern between lines.
The Section has shown the influence of the difference between the optical properties of the ink and the surrounding support on the image contrast. The next Chapters will therefore present and discuss optical properties from various inks and supports.
Chapter 5

Terahertz time-domain imaging of model and historical documents

5.1 Depth-imaging of a letter within an envelope

To familiarize ourselves with extraction of textual information in a multi-layer document with terahertz TD reflection imaging by TeraView Ltd described in Chapter 3, a letter from 1946 from the reference collection of the ISH was used as a case study. This letter consisted of a single sheet of paper handwritten with lead pencil (*i.e.* a mixture of clay and graphite), folded three times (6 overlaid layers of paper) and enclosed in an envelope.

Firstly, the single sheet of paper was taken out of its envelope, unfolded, placed on the scanning stage of RIM and scanned independently. This simple configuration was used to compare the shape of the signal reflected from an area with inscription and an area without. Imaging an opened single sheet also enabled us to determine the scanning step size which ensured good legibility of inscriptions in terahertz images while limiting file size and acquisition time. This step size was found to be 0.2 mm. During scan acquisition, the sheet was maintained flat using a round magnet, allowing an area of 1.4 cm x 1.1 cm to be imaged in its centre. Signals reflected from areas with inscriptions on the *recto* or the *verso* of the sheet, and from an area without inscriptions are displayed in Figure 5.1a. It can be observed that the maximum amplitude of the overall signal is larger for pixels where there is a graphite inscription on the *recto*. 


Graphite is a material known for its large reflectivity and absorption in the terahertz range [143, 223, 224]. The thin layer of graphite locally increases the reflection coefficient of the front surface, which leads to a larger amplitude of the reflected signal. As a result, a well-contrasted image of the graphite inscriptions on the recto can be produced by displaying the maximum amplitude of the reflected signal for each pixel using the **TVL Imaging Suite** software from TeraView, as shown in Figure 5.1d. Interestingly, a well-contrasted image of graphite inscriptions on the verso can also be produced by displaying the minimum amplitude of the reflected signal for each pixel, as shown in Figure 5.1f: inscriptions on the verso of the page are therefore legible using the data acquired from the scan of the recto, without the need to turn the page over and repeat a scan.

The letter was then folded back into its original shape and inserted into its envelope. The document was placed on the scanning stage of RIM, with no additional weight or magnetic holder used to maintain the document flat: this was to preserve
any air gaps present in between paper sheets within the document and improve the time-separation between signals reflected from each sheet. A 1.8 cm x 2 cm area was scanned with a scanning step size of 0.2 mm and the signal reflected from the central scanning position is shown in Figure 5.2a. The layer represented by peak 4 was placed in focus during the scan acquisition, with layers represented by peak 3 and peak 5 mostly contained within the focal plane. These 3 peaks represent the air - front surface of the paper sheet interfaces for the first folds of the letter. For each scanning position, the amplitude of peak 3, 4 and 5 was extracted and used to produce a terahertz image of the front surface of the first, the second and the third letter fold, using the Matlab code developed specifically for this purpose and attached in Appendix D. These terahertz images are displayed in Figure 5.2c, d and e, and reveal legible inscriptions written on each fold. Due to absorption by the graphite lines, the amplitude of the signal transmitted through the inscriptions on the first fold is attenuated compared to the signal transmitted through the blank paper. This leads to a less intense signal reflected from the second and third fold, which is observed in Figures 5.2d and e as a shadow of the inscriptions on the previous layer (also called 'occlusions').

These occlusions can easily be removed by selecting a specific pixel within the shadowed area in the image of the second (or third fold), and adding the weighted value of the same pixel from the image of the fold(s) above: in order to no longer observe shadow inscriptions, the result of the sum should equal the result of the same sum for pixels corresponding to blank paper on all folds. Figure 5.2i and j show the result of this operation. Finally, some texturing effect and speckle artefacts are observed in the terahertz images: these features are not associated with the document under investigation nor with the image acquisition process, but arise from the Delaunay triangulation used to display the value from each pixel point (please refer to the code in Appendix D). Further development of the code to prevent these image artefacts could be explored in future applications.
Figure 5.2: a) Terahertz signal reflected from the closed envelope: peak 1 corresponds to the air - front surface of the envelope interface, peak 2 to the back surface of the envelope - air interface, peak 3 to the air - front surface of the first layer of the folded letter interface, peak 4 to the air - front surface of the second layer of the letter interface, in focus, and peak 5 to the air - front surface of the third layer of the letter interface; b) B-scan showing the different layers within the envelope; c), d) and e) terahertz images produced from the amplitude of peak 3, 4 and 5 respectively; f), g) and h) photograph of the associated layer; i) and j) terahertz images d and e with occlusions removed.
5.2. Aim of the terahertz TD imaging study

Inscriptions handwritten with a lead pencil on the first three layers of a letter folded three times and enclosed in an envelope were revealed with terahertz TD imaging in reflection mode, with the terahertz beam focused on the second fold. While the 4th, 5th and 6th fold appear to be out of focus (see Figure 5.2b), the whole letter could be simply turned around and a second imaging scan could be performed to access inscriptions written on these other folds. This experiment therefore shows the potential of terahertz TD imaging to uncover the textual content from an entire closed letter. Yet, the image contrast using the amplitude of each positive peak strongly relied on the large differences in reflection properties between graphite inscriptions and the surrounding paper.

5.2 Aim of the terahertz TD imaging study

In order to objectively inform archival conservators and curators of the potential of this depth-imaging technique to access textual content in complex or fragile archival documents, this research project will not be limited to the imaging of disparate case studies, as they may not be representative of the content of archival collections. It will instead investigate the optical properties in the terahertz region of various pigments, inks and supports commonly found in archives and advise on the materials likely to lead to a good image contrast.

5.3 Surface pulse amplitude imaging

5.3.1 Influence of the constant refractive index

5.3.1.1 Measurement of the constant refractive index

A selection of pigments, dry binders and historical supports described in Chapter 3 was independently analysed with THz-TDS in transmission to determine their refractive index and absorption coefficient in the terahertz region, using TPS Spectra 3000 set-up. Arabic gum or rabbit skin glue were ground to obtain a fine powder. Pure powders of binders or pigments were pressed into three 400 mg pellets with a 10% mass concentration of sample to HDPE powder, following the optimal procedure described in Chapter 4.
**Table 5.1:** Constant refractive indices of different pigments, binders and supports. Rows highlighted in red, white and green correspond, respectively, to materials unlikely, likely and very likely to lead to a well-contrasted terahertz image when applied on support.

<table>
<thead>
<tr>
<th>Sample</th>
<th>( n_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic gum</td>
<td>1.74 ± 0.06</td>
</tr>
<tr>
<td>Bistre</td>
<td>1.77 ± 0.01</td>
</tr>
<tr>
<td>Bone black</td>
<td>2.17 ± 0.01</td>
</tr>
<tr>
<td>Carbon black</td>
<td>No transmitted signal</td>
</tr>
<tr>
<td>Genuine Ivory black</td>
<td>1.83 ± 0.07</td>
</tr>
<tr>
<td>Graphite</td>
<td>No transmitted signal</td>
</tr>
<tr>
<td>Iron gall ink (IG1)</td>
<td>2.01 ± 0.05</td>
</tr>
<tr>
<td>Lamp black</td>
<td>No transmitted signal</td>
</tr>
<tr>
<td>Lapis lazuli</td>
<td>2.02 ± 0.02</td>
</tr>
<tr>
<td>Malachite</td>
<td>2.20 ± 0.14</td>
</tr>
<tr>
<td>Minium</td>
<td>1.95 ± 0.01</td>
</tr>
<tr>
<td>Rabbit skin glue</td>
<td>1.46 ± 0.06</td>
</tr>
<tr>
<td>Sepia</td>
<td>1.70 ± 0.01</td>
</tr>
<tr>
<td>Verdigris</td>
<td>1.45 ± 0.02</td>
</tr>
<tr>
<td>Vermilion</td>
<td>1.93 ± 0.10</td>
</tr>
<tr>
<td>Original rag paper</td>
<td>1.29 ± 0.02</td>
</tr>
<tr>
<td>Additionally sized rag paper</td>
<td>1.33 ± 0.06</td>
</tr>
<tr>
<td>Parchment</td>
<td>1.71 ± 0.02</td>
</tr>
</tbody>
</table>

For the analysis of sample pellets, the reference signal was set as the signal transmitted through a 360 mg pellet of pure HDPE. For the analysis of paper and parchment sheets clamped in a magnetic sample holder, the reference signal was set as the signal transmitted through the empty sample compartment. Deconvolved signals were used to calculate the constant refractive index of supports, as recommended in Chapter 4. It is interesting to observe that for the thinner support, the original rag paper sheet with a thickness of 104 ± 6 \( \mu \)m, the constant refractive index calculated from the raw waveform (1.33 ± 0.02) is different than the one using the deconvolved waveform (1.29 ± 0.02); while for the thicker support, the parchment sheet with a thickness of 173 ± 18 \( \mu \)m, there is no difference (1.72 ± 0.02 versus 1.71 ± 0.02). Deconvolution is therefore not needed to accurately calculate the refractive index of the support if the thickness of the sheet allows a good separation of the pulses reflected from the front and back surface of the sheet.

Constant refractive indices from all pellets and supports are displayed in Table 5.1. Interestingly, there seems to be a correlation between the refractive index values from
5.3. **Surface pulse amplitude imaging**

Pigments and binders with their electrical conductive properties and their densities. According to Figure 2.6 in Chapter 2, dry dielectrics have a low refractive index. Paper and parchment are dielectrics. Arabic gum constituted of carbohydrates (as in paper) and rabbit skin glue constituted of collagen (as in parchment) can also be considered as dry dielectrics. Metals have a high refractive index. Graphite is a semi-metal, as discussed in Chapter 3, and therefore has a high conductivity along the plane of each sheet of carbon hexagons. It is thus expected to have a high refractive index, strongly reflect terahertz radiation and transmit over less than 1 µm: this would explain why no signal is transmitted through a pellet of graphite with a 10% mass concentration. Carbon black comes from the incomplete combustion of gaseous hydrocarbons, and, as pointed out by Hauptman et al., "consists of graphite-like crystalline and amorphous carbon domains". The authors also indicate that "the major factor that determines the electrical conductivity of carbon black particles is the graphitic character giving delocalized π-electrons" [225]. Other pigments resulting from the combustion of carbonaceous materials may as well contain some graphitic and some amorphous carbon.

While the graphitic (crystalline) content of carbon black and lamp black cannot be assessed using their mid-infrared spectra, their spectral signatures indicate that they are constituted of pure carbon and organic compounds. However, ivory black and bone black contain, in addition to organic compounds, inorganic calcium phosphate and calcium carbonate: they would therefore have a lower mass fraction of carbon than graphite, carbon black or lamp black within the 10% mass concentration pellet. As a result, they would contain a lower mass fraction of graphitic carbon too, and would be less conductive, so would have a lower refractive index. This would explain why a transmitted signal is measured through pellets of ivory black and bone black. Bistre contains aromatic carbons connected to functional groups containing hydrogen and oxygen atoms, resulting from the charred lignin [175]: there is therefore an even smaller mass fraction of pure carbon to potentially form graphitic structures in bistre pellets than in ivory black or bone black pellets. It could therefore be assumed that bistre is less conductive than ivory black and bone black, which would be why it shows a lower refractive index in Table 5.1.

Minium (Pb₃O₄) and vermilion (or cinnabar, α-HgS) can be considered as semiconductors (bandgap of 2.1 eV at 300 K [226, 227]). According to Figure 2.6 in Chap-
ter 2, semiconductors have a refractive index of \( \approx 2 \) and more. These pigments also exhibit very high densities (reported as 8.2 and 8.1 \( \text{g/cm}^3 \), respectively [228]). Minium and vermilion therefore have a larger refractive index (close to 2 as reported in Table 5.1) than dielectrics such as Arabic gum or rabbit skin glue.

It should be noted that while optical constants extracted from pellets with the same mass concentration of sample to HDPE can be quantitatively compared, they cannot be quantitatively compared with values extracted from solid sheets of paper or parchment, as illustrated by the difference in refractive index observed between a pellet of HDPE and a solid sheet of HDPE in Chapter 4. However, a qualitative comparison can be made between the constant refractive indices from pigments, inks and binders and the constant refractive indices from supports. It was discussed in Chapter 2 that a large refractive index leads to a large amplitude of the reflected pulse. Therefore, when using the amplitude of the pulse reflected from a specific layer to obtain a terahertz image from this layer, as illustrated in Figure 5.2, the image contrast of ink inscriptions on support is based on the difference between the refractive index of the ink and the refractive index of the support. Inks with a refractive index similar to the refractive index of the surrounding support are unlikely to give a good image contrast, and are highlighted in red in Table 5.1.

Inks with a refractive index much larger than the refractive index of the surrounding support are very likely to give a good image contrast, and are highlighted in green in Table 5.1. Carbon black and lamp black pigments appear to have similar absorption properties as graphite in the terahertz range since no signal could be transmitted through pellets with a 10% mass concentration. Due to the similarities of the mid-infrared spectral signatures of graphite, lamp black and carbon black, it is therefore reasonable to assume that, like the graphite inscriptions observed in Figure 5.1, both carbon black and lamp black inscriptions largely reflect terahertz radiation and are very likely to lead to a well-contrasted terahertz image when applied on support. All other inks with refractive indices larger than the refractive indices of supports, but still allowing the signal to be transmitted through the pellet to be detected, are likely to give a well-contrasted terahertz image.

This categorization of pigments and inks based solely on the semi-quantitative comparison of their refractive indices with those from supports may appear subjec-
5.3. Surface pulse amplitude imaging

tive and needs to be confirmed by observing the contrast in terahertz images from ink inscriptions made with those same pigments and binders and applied on those same supports. Ink dots prepared as described in Chapter 3 were therefore imaged with TPS Spectra 3000 with RIM, and a terahertz image from the dot was produced by mapping the maximum amplitude of the reflected waveforms, corresponding to the air - front surface of the support interface, *i.e.* the face of the support where the ink has been applied (*recto*).

5.3.1.2 Influence of pulse overlap on the maximum amplitude of the waveform: the need for deconvolution

A sheet of support constitutes two interfaces with a difference of refractive indices: the air - front surface of the sheet interface and the back surface of the sheet - air interface. As a result, the incident pulse is reflected on both interfaces, with the pulse reflected on the back surface of the sheet - air interface being collected at a later time than the first air - front surface of the sheet interface. The pulse reflected from the first interface corresponds to a change from a lower to a larger refractive index \( n_{\text{air}} \simeq 1 \) and \( n_{\text{support}} > 1 \): it therefore undergoes a phase shift of \( \pi \) compared to the incident pulse. However, the second interface corresponds to a change from a larger to a lower refractive index, so there is no phase shift. As a result, in reflection mode, the amplitude of the pulses reflected from the two interfaces of the sheet would have opposite signs, which is different from the case of the etalon effect observed for the waveform from a single sheet collected in transmission mode. Different research groups use different conventions with respect to the sign given to the main peak of the first collected pulse. TeraView Ltd uses a positive main peak for the first collected pulse (as illustrated by the black line in Figure 5.3), and the second reflected pulse is therefore displayed with a negative main peak (gray line in Figure 5.3).

The overlap between the two pulses and the maximum amplitude of the resulting signal depends on the refractive index, the absorption coefficient and the thickness of the support sheet, together with the width of the incident terahertz pulse:

- A large refractive index for the support will induce a large amplitude of the first reflected pulse and a large delay between the first and the second reflected pulse.
Figure 5.3: Left: schematics showing theoretical pulses reflected from the air - front surface of the sheet interface (black line) and from the back surface of the sheet - air interface (gray line), together with the resulting waveform which would be collected (red line), and represents the sum of the two overlapping pulses. Right: waveform after deconvolution, using the same deconvolution routine as described in Chapter 4.
5.3. **Surface pulse amplitude imaging**

- Large absorption by the support will result in a lower amplitude of the second collected pulse. The same applies to a large thickness of the support.

- A broad incident pulse will lead to more overlap between the first and the second reflected pulse.

When applying ink on support, if the ink has completely penetrated into the support fibres, it can still be considered that there are only two interfaces (air - front surface of the sheet and back surface of the sheet - air). Yet the presence of ink locally modifies the optical properties of the sheet close to its surface: the reflected two pulses described above would have different amplitudes and time delays than when no ink is present at the surface. If the ink constitutes a distinct additional layer on the support, three interfaces corresponding to differences in the refractive index would have to be considered: the air - ink, ink - support and support- air interfaces. As a result, three reflected pulses would be observed. The second pulse associated to the ink - support interface would correspond to a change from a large to a low refractive index, and its amplitude would therefore have a sign opposite to the first reflected peak and identical to the last reflected peak. Due to the overlap of all pulses, the presence of three pulses can rarely be directly observed when analysing raw reflected waveforms.

Any dispersive behaviour of the refractive index from either the ink or the support, and any scattering occurring within the sheet would also influence the overlap and the maximum amplitude of the collected waveform. Dispersion causes different frequencies of the pulse to propagate at different speeds within the ink or the support layer, and therefore results in a broader second collected pulse. Scattering increases the optical path of the pulse as it propagates through the sheet, which results in an attenuated second reflected pulse.

Using the maximum amplitude of the reflected waveforms without further signal processing to objectively compare the contrast in terahertz images from different ink dots on different supports would therefore be unsound. The comparison should rather be based on imaging data obtained from waveforms processed to minimize the influence of the overlap between pulses. As described in Chapter 4, deconvolution attempts to retrieve the theoretical impulse response function and therefore improves the separation between the two (or three) pulses. Figure 5.3 shows that deconvolu-
tion of the collected waveform enables the retrieval of signal of which the maximum amplitude is little or no longer affected by the overlap. Terahertz images from the ink dots were therefore produced using the maximum amplitude of the deconvolved signals calculated for each scanning position (i.e. each pixel), using the TVL Imaging Suite software, TeraView Ltd.

5.3.1.3 Influence of double Gaussian filter high frequency cut-off on image contrast

As highlighted by Zeitler and Chen [76] when discussing the high frequency (HF) and low frequency (LF) cut-offs of the double-Gaussian filter used in the deconvolution process, a lower value of HF and a higher value of LF improve signal preservation due to an overall broader filter bandwidth, thus it is the best combination for the deconvolution. LF was therefore set to the maximum value available with the software (2048). However, an HF value too low introduces noise in the deconvolved TD waveforms, which could influence the image contrast. Different values for HF have therefore been tested, and both the deconvolved TD waveforms and the image contrast from a vermilion and a minium ink dot on rag paper were examined. Results are displayed in Figure 5.4. It can be observed that lower HF values lead to a better image contrast, along with sharpening of the image features. The improved image contrast is due to the improved separation of pulses reflected from the recto and verso of the sheet, particularly for the waveforms reflected from the inked area. From HF = 60, an additional negative feature is even observed between the first and the last reflected pulse, corresponding to a reflection at the ink - paper interface. This sharpening is due to the increased filter bandwidth, leading to a participation of higher frequencies to the image, associated with better spatial resolution. At HF = 50, the waveforms become quite noisy. To objectively compare the image contrast between images from different ink dots, it is important to keep the same HF value for the deconvolution of all acquired image data. HF = 60 was selected, as it resulted in a good image contrast, good sharpening, and good SNR. This coincided with a FWHM of the first collected pulse of \( \simeq 0.3 \) ps, corresponding to the FWHM of the raw reference waveform reflected from a gold-coated mirror (see Figure 4.29 in
Chapter 4).

Figure 5.4: Left: deconvolved waveforms extracted from the same pixels within the terahertz reflection scan from a vermilion ink dot: a pixel within the inked area (black line) and a pixel within the surrounding support (gray line). Centre and right: terahertz images of a vermilion and a minium ink dot, respectively, acquired with a 75 µm scanning step and corresponding to a 3.6 x 3.6 mm² scanned area. All images have been normalized (see Section 5.3.1.5) and are displayed on the same grayscale. Each row of this table corresponds to a different value of the high frequency cut-off \( (HF) \), while the low frequency cut-off is kept constant. \( C \) values refer to contrast values, as explained in Figure 5.6.

\[1\] Different set-ups of the same TPS 3000 equipment were used throughout the project, with different updates to the image acquisition software. It was observed that the \( HF \) value leading to a FWHM of \( \approx 0.3 \text{ ps} \) for the first collected pulse and to the introduction of very little noise in the deconvolved waveform was not equal to 60 for all set-ups and versions of the software. To acknowledge this bug, the value leading to a FWHM of \( \approx 0.3 \text{ ps} \) was always selected as the reference value, to be used in deconvolution.
5.3.1.4 Influence of time subsampling on image contrast

Walker et al. [141] commented that in order to resolve features of a given width in the impulse response function, the Nyquist sampling theorem indicates that the TD sampling interval needs to be less than or equal to half that width. In the case of a waveform reflected from a vermilion ink dot on paper and deconvolved such that the FWHM of the main pulse is 0.22 ps (to try to be as close as possible to the impulse response function, and to obtain an accurate reading of the time position of pulses), the centre-to-centre spacing between the maximum of the pulse corresponding to the air-ink interface and the maximum of the pulse corresponding to the ink-paper interface is 20 time points (0.24 ps). The Nyquist limit to resolve these two peaks therefore corresponds to sampling the data once every 10 time points. While the first scans of ink dots were acquired with no time sampling, it was later decided to sample data only once every 4 time points to reduce file size. The full waveform constituted of 4096 time points between 0 and 47.378 ps (time resolution of 0.012 ps) then becomes a waveform constituted of 1024 time points between 0 and 47.378 ps (time resolution of 0.046 ps). Figure 5.5 shows that sampling every 4 time points does not lead to any significant difference in the amplitude of the deconvolved waveform (using the same double Gaussian filter for the deconvolution). Therefore, the terahertz image produced from the maximum amplitude of the deconvolved waveforms with or without sampling every 4 time points shows very similar contrasts. Equally, to reduce the file size, the full length of the collected waveforms could be reduced and further sampled: a full waveform constituted of 2048 time points between 0 and 20.831 ps (time resolution of 0.010 ps) then becomes, after a time subsampling of 4, a waveform constituted of 512 time points between 0 and 20.831 ps (time resolution of 0.041 ps). This time resolution is similar to that described earlier (i.e. 0.046 ps for a waveform of 1024 time points between 0 and 47.378 ps), so the contrast in the terahertz image produced with the maximum amplitude of the resulting deconvolved waveforms should not be significantly affected.
Figure 5.5: a) Raw waveform reflected from the centre of the vermilion ink dot on paper discussed in 5.4, with no time subsampling (black curve), corresponding to a time resolution of 0.012 ps (total length of 47.378 ps, 4096 time points), and with time sampling every 4 time points (grey curve), corresponding to a time resolution of 0.046 ps (total length of 47.3784 ps, 1024 time points). b) Associated deconvolved waveform, calculated with Matlab using the same double Gaussian filter and high and low frequency cut-offs. c) and d) Terahertz images produced with the Delaunay triangulation function from Matlab, using the maximum amplitude of the calculated deconvolved waveforms with and without time sampling, respectively. The two images show similar contrasts, so the influence of time sampling every 4 time points on the image contrast can be neglected.
Normalisation of images

To objectively compare the image contrast between ink dots, the maximum amplitude value $M(i,j)$ obtained for each pixel was normalized using the following formula:

$$M_{stand} = \frac{M_{reduced}(i, j)}{sSUBavg}$$  \hspace{1cm} (5.1)

with $M_{reduced}$, the cropped image to obtain an image size of 3.6 x 3.6 mm$^2$ (area within the user-defined red rectangle in Figure 5.6) and $sSUBavg$ the average of the pixel values corresponding to the support (area outside of the user-defined yellow circle in Figure 5.6). All grayscales were set from 1 (black) to 1.5 (white): while the carrier always appears dark in the normalized terahertz images, the more reflective the ink dot is in the terahertz region, the brighter it appears in the image.

To enable the simultaneous comparison of contrast between terahertz images of 43 different ink dots on supports, a contrast value $C$ was calculated from all terahertz images and summarized in Table 5.2:

$$C = \frac{sROIavg}{sSUBavg}$$  \hspace{1cm} (5.2)

with $sROIavg$ the average of the pixel values corresponding to the inked area (area inside the user-defined blue circle in Figure 5.6). Regions of interest excluded the vicinity of the ink dot edges, as upon drying, some ink droplets created ring-shaped solid residues inducing unpredictable scattering of the terahertz beam. Large differences of reflectivity between the ink and the carrier lead to large $C$ values. The algorithm used for image normalisation and calculation of the contrast value $C$ is provided in Appendix E.
5.3. Surface pulse amplitude imaging

Figure 5.6: Terahertz image of a vermilion ink dot obtained, after deconvolution, from the maximum amplitude of the reflected waveform. The image is overlaid by a schematic diagram describing the boundaries used for the image normalisation and calculation of the contrast value $C$. Red lines correspond to the cropped area to obtain a 3.6 x 3.6 mm$^2$ image (corresponding here to 48 pixels, with one pixel corresponding to a 75-µm scanning step). All pixels within the blue circle, the diameter of which corresponds to the minimum of the distances between blue horizontal or vertical lines, correspond to the Region Of Interest within the ink dot. All pixels outside the yellow circle, the diameter of which corresponds to the maximum of the distances between yellow horizontal or vertical lines, correspond to the Region Of Interest of the support.

5.3.1.6 Influence of step size

Scans of ink dots were at first acquired with a very fine step size (75 µm). While this enabled appreciation of variations in amplitude due to changes in topography and/or concentrations, it lead to long acquisition times and large files. Since the calculation of the contrast value uses the average of amplitudes from pixels within and outside the ink dot, the step size selected for scan acquisition should have no significant influence on the calculated contrast value, provided this scanning step size is less than half the width of the ink dot ($\simeq$2 mm), as discussed in Chapter 4. Figure 5.7 shows terahertz images of the same vermilion and minium ink dots on rag paper, but acquired with two different step sizes: 75 and 250 µm. The difference in calculated contrast values is 0.05. As reported in Table 5.2 and in Section 5.3.1.8, this difference is not considered to be significant, as it is similar to or less than differences in image contrast recorded for the same ink applied in triplicate dots on the same support. As a result, scans were acquired with a step size between 75 and 150 µm.
Figure 5.7: a) and d) Photographs of vermilion and minium ink dots on rag paper, respectively; b, c, e and f) associated terahertz images acquired in reflection mode using RIM, with a step size of 250 μm for b and e, and 75 μm for c and f, using data deconvolved with the same double Gaussian filter.

Table 5.2: Contrast values C extracted from terahertz images of different ink dots on different supports. Images acquired in reflection mode and produced by mapping the maximum amplitude of the deconvolved waveform for each pixel. Numbers in bold indicate C values extracted from terahertz images acquired with the TRay-4000, Picometrix, and produced using raw reflected waveforms.
5.3.1.7 Comparison of terahertz image contrasts

The first thing to observe when reading Table 5.2 and comparing Figures 5.8, 5.9 and 5.10 is that lamp black ink dots systematically lead to a very good image contrast. Additionally, for a given support, sepia, bistre and verdigris ink dots systematically belong to the category of lower image contrast values. The other ink dots show image contrasts between these two extremes. Categories listed in Table 5.1 are therefore confirmed when looking at image contrast, and conclusions about the likelihood of different pigments leading to a good image contrast are reinforced.

Additionally, comparing Figures 5.8 and 5.10, it can be observed that image contrasts are systematically poorer when imaging ink dots on parchment compared to rag paper sheets. This can be explained by the fact that the refractive index of the parchment sheet used as a support is larger ($1.71 \pm 0.02$) than the refractive indices of rag paper sheets, with or without additional sizing ($1.29 \pm 0.02$ and $1.33 \pm 0.06$).
Figure 5.9: Photographs of ink dots on additionally sized rag paper and the associated terahertz images acquired in reflection mode using RIM attached to TPS Spectra 3000, with a step size between 75 and 150 µm. Terahertz images were produced using the maximum amplitude of the reflected waveforms after deconvolution. All images were normalized as described in Section 5.3.1.5 and are displayed on the same grey scale.

The differences in refractive index values for inks and the surrounding support are therefore reduced in the case of parchment, and it follows that the differences in the amplitude of the pulse reflected at the air-front surface of the sheet interface with or without ink present is lower for parchment.

As for rag paper sheets with and without additional sizing, no significant difference in the constant refractive index is observed. However, ink dots applied on the additionally sized rag paper sheet tend to have a broader diameter than ink dots on the original rag paper sheets, while the applied volume of the ink remained the same (2 µL). This implies that sizing modifies the surface properties of paper in such a way that the angle of contact between the ink and the sheet is increased. With the exception of bone black (which seems to have undergone no or little spread in the presence of sizing), contrast values tend to be poorer for the rag paper sheet with additional sizing than for the original rag paper sheet.
5.3. Surface pulse amplitude imaging

Figure 5.10: Photographs of ink dots on parchment and the associated terahertz images acquired in reflection mode using RIM attached to TPS Spectra 3000, with a step size between 75 and 150 µm. Terahertz images were produced using the maximum amplitude of the reflected waveforms after deconvolution. All images were normalized as described in Section 5.3.1.5 and are displayed on the same grey scale.

Pigment concentration therefore plays a role in the image contrast: a larger concentration of pigment on the paper surface leads to an increased amplitude of the reflected waveform. This is further confirmed by the variations of contrast within an ink dot (mostly) coinciding with variations of pigment concentration, as clearly illustrated by the minimum ink dot on additionally sized paper in Figure 5.9.

When not caused by variations in concentration, variations in contrast within an ink dot can arise from scattering. The configuration of the optics within RIM is set up such that the illuminating source is at 30° from the normal to the surface, and a detector is also placed at 30° from the normal to the surface. An edge can be considered as a diffuse reflector: the beam would be reflected in all directions, but the intensity of the radiation (and therefore also the amplitude of the electric field) reflected from this edge would be larger in the direction of the normal to the surface, according to Lambert’s cosine law. As a result, the amplitude of the electric field collected from an edge at an
oblique angle would be reduced compared to a specular reflection. This would appear dark in the terahertz image produced from the amplitude of the pulse reflected from the front surface of the sheet. This phenomenon is particularly observable for bone black on any support, or for iron gall ink and bistre on the original rag paper: due to scattering, the coffee-stain like ring left upon drying appears darker than the centre of the ink dot, even though it may represent an area with larger pigment concentration.

5.3.1.8 Repeatability
To acknowledge the difference between variations of $C$ values inherent to minor variations in shape, concentration or topography of the dots, and variations of $C$ values caused by the different refractive properties of different inks, standard deviation values have been calculated after applying the same ink suspensions of vermilion (colour pigment likely leading to a good image contrast) and bistre (black pigment unlikely to lead to a good image contrast) in triplicate droplets on the same support and imaging them with terahertz TD imaging.

It can be observed that all bistre ink dots systematically lead to a poor image contrast, while the contrast of vermilion ink dots can vary greatly between triplicates on the same support, likely due to scattering and distribution of ink concentration. The largest recorded standard deviation was for vermilion on the original rag paper sheet: 0.17 as reported in Table 5.2.

This is the reason why the influence of the refractive index from pigments and supports on the image contrast can only be qualitatively, and not quantitatively, assessed using trends and categories rather than comparing the contrast from ink dots pairwise.

5.3.1.9 Comparison between different set-ups
To observe the influence of the terahertz imaging set-up and environment (open-air or purged) used for scan acquisition on the image contrast, T-Ray 4000, Picometrix, was also used, set in reflection mode with normal incidence, in ambient air. The deconvolution routine used previously for TPS Spectra 3000, TeraView, could not be used: much noise was introduced in the deconvolved signal and clearly separated pulses could not be achieved.
5.3. Surface pulse amplitude imaging

Figure 5.11: Photographs of three different bistre and vermilion ink dots applied on rag paper, additionally sized rag paper and parchment, and the associated terahertz images acquired in reflection mode using RIM attached to TPS Spectra 3000, with a step size between 75 and 150 µm. Terahertz images were produced using the maximum amplitude of the reflected waveforms after deconvolution. All images were normalized as described in Section 5.3.1.5 and are displayed on the same grey scale.

This phenomenon could be explained by the fact that under the conditions of image acquisition on the day of acquisition of the imaging data, T-Ray 4000 showed a lower SNR and a broader pulse width than TPS Spectra 3000 (19 dB and 1.17 ps against 54 dB and 0.33 ps for TPS Spectra 3000 with RIM), as discussed in Chapter 4. Redo-Sanchez et al. [229] recently suggested that under low SNR conditions, frequency-based deconvolution routines such as the one used in this study are not efficient, and another pulse separation approach should be used (see Section 5.6.4 further below).

Yet, when producing images from the maximum amplitude of the raw reflected waveforms, it is interesting to note that the same contrast categories based on the optical properties of pigments can still be observed: lamp black ink dots on rag paper with and without additional sizing lead to a very good contrast, bistre ink dots lead to a very poor contrast, and vermilion, bone black and minium can be more or less easily distinguished. Another interesting thing to note is that no dark edges are observed. This is not simply due to the larger step size used (200 µm) as these edges could still be observed with RIM attached to TPS Spectra 3000 using a step size of 250 µm. Instead, this is more likely due to the use of illumination at normal incidence: the amplitude of the diffusely reflected electric field is larger in the direction normal to the surface, so there is still much radiation collected from the edge.
Deconvolution can either enhance or decrease the resulting image contrast, depending on the pulse number, amplitude and overlap. Terahertz images obtained from the maximum amplitude of the raw reflected waveforms (no deconvolution) are attached in Appendix F for comparison.

**Figure 5.12:** Photographs of ink dots on rag paper (top) and additionally sized rag paper (bottom) and the associated terahertz images acquired using T-Ray 4000 set in reflection mode, Picometrix, with a step size of 200 µm. Terahertz images were produced using the minimum amplitude of raw reflected waveforms (the main and first reflected pulse corresponding to the air-front surface of the sheet interface is a negative pulse, following the convention used by Picometrix). All images were normalized as described in Section 5.3.1.5 and are displayed on the same grey scale.

### 5.3.2 Influence of indentations and other topographic features

#### 5.3.2.1 Model indentations on paper and parchment

Having imaged characters handwritten with lead pencils of different grades with a TD spectrometer, Abraham *et al.* observed a difference in the amplitude of the transmitted signal at 2 THz from one character to the other, which they attributed to the difference in graphite content within the pencils used for each separate character [143]. However, using the time delay of the main transmitted pulse, all characters appear similarly
contrasted in the terahertz image: they attribute this phenomenon to the local thinning ($\simeq 30 \, \mu m$) of paper caused by the indentations left by the pen. Theoretically, if indentations are present at the surface of the paper sheet, assuming that the refractive index $n_{\text{paper}}$ of paper is not significantly locally modified by any compression of the paper fibrous structure caused by the pressure applied by the pen, the thickness of the paper sheet $d_{\text{paper}}$ is locally decreased to $d_{\text{paper}} - d_{\text{indentation}}$ where the indentations are, and the optical path decreases from $n_{\text{paper}}d_{\text{paper}}$ to $n_{\text{air}}d_{\text{indentation}} + n_{\text{paper}}(d_{\text{paper}} - d_{\text{indentation}})$. It follows that the signal transmitted through the indentations caused by drawing with lead pencils should therefore be detected sooner in time than the signal transmitted through the blank unaltered paper sheet, since it travels through a lower optical path. Yet this conclusion seems to be in opposition with another statement made within the same article, when comparing the TD signals transmitted through a square graphite area (1 cm x 1 cm) drawn with different pencil leads on a 80 $\mu m$-thick white sheet paper. The results show that the main pulse of the signals transmitted through such squares of graphite on paper is delayed, and not accelerated, compared to the main pulse of the signal transmitted through a white sheet of paper: the authors claim that this temporal shift originates from the variations of the optical path due to the increasing sample thickness. There therefore seems to be a confusion in understanding the role of indentations on the time delay of the transmitted signal, which is all the more difficult to understand when analysing indentations covered by a layer of graphite and clay (assumed to be less than a few micrometers) and comparing raw waveforms (i.e. without any deconvolution applied to reduce or remove the influence of a possible etalon effect). We therefore suggest to analyse indentations left by handwriting with a clean nib pen, with no ink or other applied writing material. Since we did not have access to a set-up in transmission configuration, and since the study of ink dots was performed in reflection mode, we will observe the influence of indentations on the reflected signal, rather than the transmitted signal.

The letters ”UCL” were handwritten with a clean nib pen on a print paper and on a historical parchment sheet (the same sheet as used for the study with ink dots). Figures 5.13a and g display the surface profile of the indentations, using the TraceiT profilometer. For paper, there is a $\simeq 15 \, \mu m$ difference between the average paper surface and the depth of the inscriptions. Using the deconvolved signals, a delay of the
front surface of the paper sheet can clearly be observed on the B-scan shown in Figure 5.14a where the inscription is present and causes an indentation (arrow 1) compared to where there is no inscription (arrow 2). Based on the discussion in Section 4.2.4, the depth of the indentations is just large enough to reliably attribute the observed time delay of the front surface to the presence of indentations. Using this time delay, an image of the inscriptions can be reconstituted as shown in Figure 5.14c: indentations left by inscriptions appears in white or light grey as they are more delayed (i.e. their time delay is larger) than the surrounding paper, which appears in dark grey. However, if the terahertz set-up used for imaging has a larger time measurement uncertainty than the one discussed in Section 4.2.4, these indentations would not be reliably detected and distinguished from the noise in time position, and may not lead to a legible text on the image produced with the terahertz time delay.

No inscriptions can be observed in the terahertz image produced from the maximum amplitude of the deconvolved waveforms. This implies that the slight increase in optical path due to the presence of an indentation (the beam travels through an additional \(\approx 15 \mu\text{m} \) of air) does not change the amplitude of the pulse reflected at the air - surface of the paper interface, and that the indentation causes no noticeable local increase of the refractive index at the paper surface.

However, when using raw waveforms to produce a terahertz image with the maximum amplitude, the inscriptions are very clearly legible, as shown in Figure 5.13b. Comparing the deconvolved waveforms (Figure 5.14b) and the raw waveforms (Figure 5.13f) collected from the same scanning points, it can be observed that the local thinning of paper caused by indentations brings closer the main positive pulse corresponding to the air - paper interface and the main negative pulse corresponding to the paper - air interface. In the deconvolved waveforms, this proximity does not induce any overlap between the two main pulses, and therefore does not change the amplitude of the main pulse. However, with the raw waveforms, an overlap already exists between the two main pulses for the pixel with no indentation, and this overlap increases with indentation: while the overlap without indentation (black curve) creates a shoulder on the main pulse (indicated with a star), this shoulder merges with the main pulse with indentation (grey curve) and the amplitude of this main pulse increases. This is the phenomenon leading to the contrast observed in Figure 5.13b.
5.3. Surface pulse amplitude imaging

Figure 5.13: a) and g) Depth profile of the surface indentations left by writing “UCL” with a clean nib pen on a sheet of print paper and a sheet of parchment, respectively, obtained using the TRACEiT profilometer; b) and h) Terahertz images of the indentations scanned in reflection mode with RIM and a step size of 75 µm, obtained using the maximum amplitude of the raw waveform (no deconvolution); c) and i) Terahertz images of the indentations obtained using the time position of the maximum amplitude of the raw waveform; d) and j) Terahertz images of the indentations obtained using the amplitude of the raw waveform at the time position giving the best image contrast; e) and k) B-scan along the dotted horizontal line shown in c) and i); f) and l) terahertz waveforms extracted from pixels at the positions indicated with arrows in e) and k): the grey waveform corresponds to the presence of a surface indentation, represented by the grey arrow numbered “1”, and the black waveform, to the absence of surface indentation, represented by a black arrow numbered “2” (this waveform is displayed with an offset of +4).
Figure 5.14: Terahertz signals and images obtained after deconvolution, acquired from the scan in reflection mode of surface indentations left by writing "UCL" with a clean nib pen on a sheet of print paper, using RIM with a step size of 75 µm. a) B-scan along the dotted horizontal line shown in c) and d). The horizontal axis of the B-scan corresponds to the pixel position along the dotted line in c) and d), and the vertical axis corresponds to the optical delay from 22 to 27 ps. b) Deconvolved terahertz waveforms extracted from pixels at the positions indicated with arrows in a): the grey waveform corresponds to the presence of a surface indentation, represented by the grey arrow numbered "1", and the black waveform, to the absence of surface indentation, represented by a black arrow numbered "2" (this waveform is displayed with an offset of +0.03); c) and d) Terahertz images produced from the time delay and amplitude of the maximum positive peak of the deconvolved waveforms.

As with the deconvolved data, a delay is still observed at the paper surface in the presence of an indentation on the B-scan (arrow 1 in Figure 5.13e) and this delay can be used to produce a terahertz image of the inscriptions (Figure 5.13c). Interestingly, the pulse overlap is such that at the time position indicated with the star (23.41 ps ± 0.05 ps), the difference in amplitude between the waveforms with and without indentations is maximum. Therefore, by mapping the amplitude of the TD signals at this specific time position, a terahertz image with legible inscriptions can also be produced, as shown in Figure 5.13d.

Using raw waveforms instead of deconvolved waveforms therefore seems to enable a better observation of indentations at the surface of a paper sheet. Would this conclusion be applicable to a parchment sheet? Two main differences exist between these two supports. Firstly, a parchment sheet is thicker (170 µm) than a print paper sheet (110 µm). As a result, while the indentation is still clearly observable on the
5.3. Surface pulse amplitude imaging

B-scan (arrow 1), the local thinning of the sheet due to this indentation is small compared to the thickness of the sheet, and the slight change in overlap between the pulses reflected from the front and back surfaces of the parchment sheet does not induce a noticeable change in amplitude of the main pulse. It follows that the terahertz image produced using the maximum amplitude of the reflected waveform (Figure 5.13h) does not show any legible inscriptions. Secondly, the surface of the parchment sheet is more irregular (rougher and less flat) than the print paper, as observed on the profilometry image in Figure 5.14g and on the B-scan in Figure 5.14k. As a result, the terahertz image produced using the time delay of the main pulse (i.e. the front surface of the parchment sheet) does not show any legible inscriptions, as shown in Figure 5.14i: the depth of indentations is similar to the variations of surface height of the parchment sheet itself, and is also similar to the depth measurement uncertainty associated with the time measurement uncertainty described in Section 4.2.4.

The small change in time delay induced by the indentation leads to a noticeable change in amplitude at the time position 23.04 ps ± 0.05 ps (corresponding to the time position of the maximum amplitude for some of the main pulses): as a result, the terahertz image produced from the amplitude of the signal at 23.04 ps ± 0.05 ps shown in Figure 5.13j is the only image in which the handwritten inscriptions become legible.

5.3.2.2 Sieve lines

A sieve line is a topographic feature in the paper sheet formed during its manufacture: it is left by the threads forming the sieve when covered with wet paper pulp and left to dry. For sieve lines on a historical rag paper sheet ≃100-µm thick, the same phenomena as those observed for the indentations on the sheet of print paper are observed. The thinning of paper along a sieve line makes the pulses reflected from the air-paper and paper-air interfaces come closer together. In the raw waveforms, this merges the shoulder peak indicated with the star in Figure 5.15f with the main and first reflected peak, and increases its amplitude.

However, in the deconvolved waveforms, this proximity does not lead to a significant overlap and does not change the amplitude of the first reflected pulse. It follows that sieve lines can be observed in Figure 5.15b while they cannot in Figure 5.15h. As for images produced with the time delay or with an amplitude at a specific time position,
sieve lines are also well observed, and the use of the raw or deconvolved waveforms matters little as long as paper thinning induced by a sieve line is more pronounced than surface roughness or curvature of the paper sheet.

To an extent, indentations left by a writing tool or sieve lines cause a contrast in the image produced using the amplitude of the pulse reflected from the front surface of support, but only if using the raw waveforms and if the local thinning of the sheet leads to merging of the shoulder peak with the pulse of interest. Better image contrast of these topographic features could be obtained using the amplitude of the signal (raw or deconvolved) at a specific time slice, but this would only be the case for relatively flat surfaces: in the above study, all sheets were maintained flat using a magnetic sample holder. For a stack of paper sheets as curved as those shown in the B-scan in Figure 5.2, a specific time slice would not necessarily coincide with the position of the pulse reflected from the paper sheet of interest throughout the whole scanned area.
Figure 5.15: a) and g) Photograph of a sieve line at the surface of a rag paper sheet from the 19th century, obtained using the signal transmitted through the sheet; b) and h) terahertz images of the sieve lines scanned in reflection mode with RIM and a step size of 150 µm, obtained using the maximum amplitude of the raw and deconvolved waveforms, respectively; c) and i) terahertz images of the sieve lines obtained using the time position of the minimum amplitude of the raw and deconvolved waveforms, respectively; d) and j) terahertz images of the sieve lines obtained using the amplitude of the raw and deconvolved waveform, respectively, at the time position leading to the best image contrast; e) and k) B-scan along the dotted horizontal line shown in d) and j); f) and l) raw and deconvolved waveforms, respectively, extracted from the pixels at the positions indicated with arrows in e) and k); the grey waveform corresponds to the presence of a sieve line, represented by a grey arrow, and the black waveform, to the absence of sieve line, represented by a black arrow.
Table 5.3: Constant refractive indices from thin sheets of historical paper calculated using the deconvolved waveforms (thicknesses of the paper sheets are given between brackets), and constant refractive indices of Prussian blue pigment powders acquired from different suppliers, diluted with HDPE and pressed into pellets.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Constant refractive index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter 12.07.46 (115 ( \mu m ))</td>
<td>1.27 ± 0.08</td>
</tr>
<tr>
<td>Letter 13.07.46 (90 ( \mu m ))</td>
<td>1.41 ± 0.00</td>
</tr>
<tr>
<td>Letter 27.07.46 (70 ( \mu m ))</td>
<td>1.35 ± 0.00</td>
</tr>
<tr>
<td>Prussian blue, Cornelissen &amp; Son</td>
<td>1.48 ± 0.02</td>
</tr>
<tr>
<td>Antwerp blue, Cornelissen &amp; Son</td>
<td>1.33 ± 0.02</td>
</tr>
<tr>
<td>Prussian blue, Kremer Pigmente</td>
<td>1.49 ± 0.01</td>
</tr>
<tr>
<td>Prussian blue, Zecchi</td>
<td>1.47 ± 0.01</td>
</tr>
</tbody>
</table>

5.3.2.3 Indentations on historical documents

Three letters from 1946 from the reference collection of the ISH, were typewritten with a dark blue ink on three types of different paper. As indicated in Figure 5.16, analysis of the inked area with ATR-FTIR spectroscopy showed strong absorption at 2088 \( \text{cm}^{-1} \) characteristic of the cyano group found in Prussian blue (see Chapter 3). Analysis of the blank paper sheets with ATR-FTIR spectroscopy revealed that the paper used for the letter dated the 12\(^{th}\) of July 1946 contains lignin (coinciding with the slightly yellow colour of paper), while the other papers contain kaolinite, the mineral most extensively used as a filler and for coating of paper [230]. The more intense peaks of kaolinite in the normalized\(^1\) spectra of the document from the 13\(^{th}\) of July 1946 seems to indicate that this paper has a higher kaolinite content than the paper from the 27\(^{th}\) of July.

It can be observed from Table 5.3 that the refractive indices of Prussian blue and Antwerp blue pigments are similar to the refractive indices of historical paper sheets. It is interesting to observe that the refractive index of the paper sheet with a larger content of kaolinite (13.07.46) is larger than those of the other paper sheets, and that the refractive index of Antwerp blue, which contains barium sulfate as an extender, is lower than the refractive indices of other Prussian blues.

\[^1\] Normalisation to 0 - 1 was performed by applying the following Equation to each absorbance value \( y \):

\[
y_{\text{Normalized}}[i] = \frac{y[i] - \text{min}}{y[1029\text{cm}^{-1}] - \text{min}}
\]

with \( \text{min} \) the minimum absorbance value over the spectral range under study.
Figure 5.16: a), b) and c) ATR-FTIR spectra (in the mid-infrared region) from typewritten letters sent by the same person on the 12th, 13th and 27th of July 1946, respectively: black lines represent spectra acquired from an area with blank paper and grey lines represent spectra acquired from an area with ink; d) FTIR spectra acquired from areas with blank paper, for each letter: k indicates the spectral features characteristic of kaolinite and l of lignin.

From the study of ink dots presented above, the similarity of the refractive indices of support and pigments would imply that the contrast in terahertz images of inscriptions typewritten with Prussian blue, produced using the maximum amplitude of the deconvolved reflected waveform, would be poor. This is confirmed by the terahertz images in Figure 5.17b, h and n, with no observable inscriptions. However, the letters were typewritten: this could have left some faint indentations on the surface of the paper which could be used to improve the image contrast, as discussed above. These indentations are observable on the B-scans from the deconvolved data in 5.17e, k and q, and from the raw data in 5.18e, k and q, indicated with grey arrows. As with the sieve lines and the inscriptions left with a clean nib pen, the indentations left by the typewriter lead to a local thinning of the paper, associated with a decrease in time separation between the signals reflected from the air - paper and paper - air interfaces. This proximity does not lead to an overlap between the deconvolved pulses reflected from
the front and back of the sheet. Therefore the difference in amplitude between the main pulses with and without indentation observed in Figures 5.17f, l and r is not significant.

However, analysing the raw waveforms in Figures 5.18f and l, the shoulder peak indicated with the star merges with the main peak in the presence of an indentation, and the resulting amplitude of the main peak is significantly increased. As a result, the terahertz images produced with the maximum amplitude of the raw TD waveforms show a legible "a" in Figure 5.18b and a legible "A...D...L" in Figure 5.18h. The letter dated from 27\textsuperscript{th} July 1946 is thinner than the other letters (70 µm). As a result, there is no shoulder peak observed for the raw signal from the surface without indentation since the overlap between pulses reflected from the front and back surfaces is already significant. Besides, it appears that the indentations left by the typewriter affect both the front and back surfaces of the sheet: it follows that pulses reflected from the front and back surfaces are both delayed, and there is no decrease in time separation between pulses with the local thinning. Consequently, the amplitude of the main pulse remains the same with or without indentation, and no inscription can be read in the terahertz image produced with the maximum amplitude of the raw waveform (Figure 5.18n).

As for the study of inscriptions with a clean nib pen and sieve lines, an improved contrast can be obtained using the time delay of the main pulse (c, i and o), while the best contrast is obtained using a time slice corresponding to the maximum amplitude difference between the waveforms reflected from areas with and without indentations (d, j and p). No significant difference in contrast is observed if using deconvolved data rather than raw data.

So far, the influence of the type of ink and support, as well as of topographic features, on the amplitude of the pulse reflected at the air - front surface of the sheet interface has been investigated. The following Section discusses configurations where the ink layer of interest is located at the back surface of the sheet - air interface, and discusses any differences in data interpretation this may cause.
5.3. Surface pulse amplitude imaging

Figure 5.17: a), g) and m) Photographs of the signature typewritten on letters sent by the same person on the 12th, 13th and 27th July 1946, respectively, using different types of paper; b), h) and n) the associated terahertz images acquired in reflection mode with RIM and a step size of 75 µm; the images were produced using the maximum amplitude of the deconvolved waveform; c), i) and o) terahertz images produced using the time position of the maximum amplitude of the deconvolved waveform; d), j) and p) terahertz images produced using the amplitude of the signal at the time position giving the best image contrast; e), k) and q) B-scan along the red dotted line indicated in d), j) and p); f), l) and r) waveforms from pixel positions indicated with arrows in the B-scans: grey lines and arrows correspond to the presence of an indentation, while black lines and arrows correspond to the absence of indentation.
Figure 5.18: a), g) and m) Photographs of the signature typewritten on letters sent by the same author on the 12th, 13th and 27th of July 1946, respectively, using different types of paper; b), h) and n) the associated terahertz images acquired with TPS Spectra 3000 in reflection mode and produced using the maximum amplitude of the raw waveform; c), i) and o) the terahertz images produced using time position of the maximum amplitude of the raw waveform; d), j) and p) the terahertz images produced using amplitude of the signal at the time position giving the best image contrast; e), k) and q) B-scan along the red dotted line indicated in d), j) and p); f), l) and r) waveforms from pixel positions indicated with arrows in the B-scans: grey lines and arrows correspond to the presence of an indentation, while black lines and arrows correspond to the absence of indentation.
5.4 Depth pulse amplitude imaging

5.4.1 Ink dots in *verso* configuration

When the ink is on the *verso* of a sheet during scan acquisition, the incident beam impinges first on the air - support interface, then travels through the support and reaches the support - ink interface, before being reflected on the ink - air interface. The collected waveform is similar to the *recto* configuration, but "reversed", as illustrated by comparing grey curves in Figures 5.19e and i. The first interface corresponds to a change from a lower (air) to a larger (support) refractive index and is represented by a positive peak (number 1 in i). The second interface corresponds to a change from a lower (support) to a larger (ink) refractive index and is also represented by a positive peak (number 2 in i). The last interface corresponds to a change from a larger (ink) to a lower (air) refractive index and is represented by a negative peak (number 3 in i). It is possible to obtain an image from the ink dot by mapping the amplitude of the last reflected pulse, which is the minimum amplitude of the collected waveform, as illustrated in Figures 5.19f and g. Yet, conclusions on the image contrast discussed above and linking the difference in pulse amplitude reflected from the ink and from the blank paper solely to the difference in refractive index between the ink and the support should be reviewed. In this configuration, the pulse reflected by the ink dot has travelled through a layer of ink before it is reflected at the ink - air interface. It has therefore undergone absorption from the ink layer before being reflected back towards the incident direction. As a result, although the coefficient of reflection at the ink - air interface may be larger than at the support - air interface if the refractive index of the ink is significantly larger than the index of the support, the difference in amplitude between the pulse reflected from the dot (on the ink - air interface) and the pulse reflected from the surrounding support (on the support - air interface) may not be as large as in *recto* configuration, due to absorption from the ink layer. This is what is observed in Figure 5.19: the amplitude ratio between the pulses 1 and 4 in e is larger than between 3 and 5 in i, and the contrast of the normalized terahertz image in c (contrast value $C = 1.64$, as calculated using Equation 5.2) is larger than the contrast in g ($C = 1.21$). Otherwise, in *verso* configuration, the scattering phenomenon is still observed on the edges of the ink dot, leading to a less intense reflected pulse from the ink - air interface.
Figure 5.19: a) Photograph of an iron gall ink dot on rag paper; b - e) the dot scanned in recto configuration, with the incident beam impinging on the surface with ink dot; f - i) the dot scanned in verso configuration, with the incident beam impinging on the surface without ink dot. Terahertz images acquired with RIM, with a step size of 75 µm, and produced using b and c) the maximum amplitude or f and g) the minimum amplitude of the deconvolved reflected waveforms, before and after normalization respectively. d and h) B-scans along the red dashed line indicated in b) and f). e and i) Deconvolved waveforms reflected from the centre of the ink dot, and from an area with blank paper. j) Virtual cross-section of the ink dot obtained using optical coherence tomography: image displayed on the same scale as terahertz images, and also displayed with a larger scale and tilted to improve visualisation of the surface features on the flat paper surface.
5.4. Depth pulse amplitude imaging

Analysing the B-scan and waveforms in recto configuration (Figures 5.19d and e), it can be observed that the surface of the ink is at a similar height as the surface of the support, as the first reflected pulses from both waveforms are detected at similar times: 23.72 ps ± 0.05 ps for the centre of the ink dot (peak 1 in e) and 23.70 ps ± 0.05 ps for the paper (peak 4 in e). This is confirmed by the virtual cross section of the ink dot obtained using Optical Coherence Tomography (OCT), in Figure 5.19j: the horizontal red dashed line overlaid in the tilted and zoomed in OCT image indicates that the centre of the ink dot and the surrounding support are at similar heights. Yet, the pulse reflected from the back surface paper - air is detected at a much later time when the ink is present on the front surface (24.69 ps ± 0.05 ps for pulse 3 against 24.34 ps ± 0.05 ps for pulse 5 in e, so a time difference of 0.35 ps): this significant difference in optical paths is therefore mostly due to the increase in refractive index caused by the presence of the ink, rather than to a local increase in thickness. Analysing the B-scan and waveforms in verso configuration (Figures 5.19h and i), with no prior knowledge of optical properties of the ink, the thickness of the ink layer could be easily overestimated by comparing the time delays of peak 3 (recorded at 24.13 ps ± 0.05 ps) and peak 5 (recorded at 23.89 ps ± 0.05 ps) in i: a difference in time delay of 0.24 ps between the pulses reflected from the back surface of the sheet would lead to a thickness of the ink of 36.5 µm, if the influence of the ink refractive index in the change in optical path is not taken into account. More care is therefore needed in the interpretation of ink inscriptions in verso configuration.

Similarly, the layer of graphite and clay left by a lead pencil on paper is too thin to induce any significant time delay between the pulses reflected from the front surface of the sheet, when imaging inscriptions written on the recto of the sheet. As a result, the terahertz image shown in Figure 5.20d, produced from the time delay of the pulse with the maximum amplitude (i.e. first reflected pulse) of the deconvolved waveforms, does not clearly show the inscriptions on the recto of the sheet. Only areas where the writing pencil has left an indentation on the paper surface (such as the pixel indicated with a solid grey arrow in 5.20e) would lead to a delay of the first reflected pulse (which is why the solid grey curve in a has the first reflected pulse delayed compared to the pulse from the blank paper in black). However, inscriptions written on the verso of the sheet always correspond to a delayed reflected pulse. Yet, any topographic change would
need to be caused by the writing tool rather than the thickness of the graphite layer.

As for the case of the iron gall ink dot described above, the optical path of the pulse reflected from verso inscriptions is increased by the large refractive index of the very thin layer of graphite it travels through before being reflected at the graphite - air interface. This is the reason why verso inscriptions are entirely legible in the image in Figure 5.20h, produced using the time delay of the pulse with the minimum amplitude (i.e. last reflected pulse) of the deconvolved waveforms.

The large refractive index of graphite causes the amplitude of the first reflected peak to be larger where the inscriptions are, compared to the pulse reflected from the surface of blank paper. As a result, Figure 5.20c showing the terahertz image produced using the maximum amplitude of the reflected waveforms very clearly displays the inscriptions written on the recto of the sheet. From the study of the iron gall ink dot above, it would be expected that the amplitude of the pulse reflected from the back surface of the sheet is larger in the presence of graphite inscriptions on the verso, due to the larger refractive index of graphite than paper, leading to a larger coefficient of reflection at the graphite - air interface than at the surrounding paper - air interface. However, graphite is much more absorbent than iron gall ink (as discussed above, graphite is a metal-like material, and therefore has both large refractive index and absorption coefficient), and the strong absorption of the pulse travelling through the thin layer of graphite before being reflected on the graphite - air interface causes the absolute amplitude of this pulse to be lower than the absolute amplitude of the pulse reflected at the surrounding paper - air interface. This is clearly observed when comparing the negative pulses from the dotted grey curve and the black curve in Figure 5.20a. As a result, Figure 5.20g produced using the amplitude of the pulse reflected from the back surface of the sheet, i.e. the minimum amplitude of the waveform, clearly shows the inscriptions from the verso of the sheet. Yet they appear brighter than paper, which is contrary to the centre of the iron gall ink dot in Figure 5.19f which is darker than paper. This is because the strong absorption from graphite decreases the minimum amplitude of the waveform, so with a less negative value than the pulse reflected from paper.
Figure 5.20: b) and f) Photographs of the front and back surfaces, respectively, of the same paper sheet written on the 24th of July 1946 with a lead pencil (i.e. containing graphite) by the same author as the letters discussed in Figures 5.18 and 5.17; c) and g) associated terahertz images acquired in reflection mode with RIM and a step size of 200 µm and obtained using the maximum and minimum amplitude of the deconvolved waveform; d) and h) terahertz images obtained using the time position of the maximum and minimum amplitude of the deconvolved waveform; e) B-scan along the dotted line indicated in b) and f); a) deconvolved waveforms extracted from pixel positions indicated with arrows in e): the solid grey line and arrow correspond to the presence of an inscription on the front surface of the paper sheet, the dotted grey line and arrow correspond to the presence of an inscription on the back surface of the paper sheet, and the black line and arrow correspond to the absence of inscription on either surfaces of the sheet.

It should be noted that, due to the same absorption phenomenon, a shadow of the *recto* inscription is observed in bright in this same Figure. Reflection and absorption from the ink are therefore competitive phenomena influencing the amplitude of the pulse reflected from inscriptions written on the *verso* of a sheet. It is therefore possible that the influence of one counterbalances the influence of the other, leading to a poor image contrast of the *verso* inscriptions.

Finally, it should be noted that pulses from the paper - ink or ink - paper interface are not observed in the deconvolved waveforms presented in Figure 5.20a. They start to be observed using a lower value for the HF cut-off of the double Gaussian filter in the
deconvolution process, due to the thinness of the graphite layer, but they are difficult to distinguish from the noise introduced with such a low HF value.

5.4.2 Inscriptions on verso of a parchment book cover

The Dutch National Archives hold bindings with covers produced from a parchment sheet which was previously used as an independent single sheet document, with maps or musical scores written on them. As illustrated in Figure 5.21a, such parchment sheets can be re-used such that the side with written information faces the board (made of cardboard or wood), and the corners of the parchment sheet are folded around the board and attached. A sheet of paper may be added between the parchment sheet and the board to increase the opacity of the parchment cover, as illustrated in Figure 5.21b. While some parchment covers detached from the board with time, revealing the inscriptions on the verso of the parchment sheet, others remain attached to the board, so that for a conservator, it is impossible to know whether or not archival information is present on the verso of the parchment cover, and to know the content of this archival information. From all studies discussed so far in this chapter, it seems that terahertz TD imaging in reflection mode might be useful in the case of such historical artefacts.

A mock-up was designed by using an historical parchment book cover from the reference collection of the ISH (Figure 5.21b), and by writing the inscription ”UCL” with a nib pen on the back of the parchment cover, using iron gall ink (IG1) for the letter ”U”, lamp black for the ”C” and minium for the letter ”L” (Figure 5.21c). The book cover was placed on the scanning stage of RIM attached to TPS Spectra 3000, with a metallic weight placed at the back of the cardboard to mimic the weight of the pages in a book, as illustrated in Figure 5.21d.

Analysing the B-scan shown in Figure 5.22c, three layers can be distinguished: the parchment is detected first, the thinner layer of paper is detected second, and the layer of cardboard is detected last (with only the front surface being displayed in Figure 5.22c). The raw waveform reflected from this stack shown in Figure 5.22a appears quite complex, while the deconvolved waveform shown in Figure 5.22b seems easier to interpret. Additionally, to separate time delays due to the presence of ink inscriptions on the verso from delays due to the non-flat parchment surface and further ease the interpretation of the data, it was decided to align all waveforms to their first re-
5.4. Depth pulse amplitude imaging

flected pulse (corresponding to the front surface of the parchment sheet), as shown in the B-scan in Figure 5.22d.

Figure 5.21: a) A book from the Dutch National Archives, the cover of which is made of a parchment sheet with a map drawn on the side attached to the cardboard. b) A historical book cover from the reference collection of the ISH. c) Schematic diagram explaining the succession of layers used for the model study: the inscriptions “UCL” were written on the back of the parchment cover (as shown by the photograph in the inset), the parchment cover was placed on top of a historical paper sheet with the side with inscriptions facing the paper sheet, and all were placed on top of the cardboard. d) The layout of the sample during a scan acquisition in reflection mode with RIM, with a metallic weight placed at the back of the cardboard, to mimic the weight of the pages in a book.
In above the studies, it was observed that the presence of ink inscriptions on the verso of a sheet delays the reflection of a pulse at the back surface of the sheet - air interface. Assuming that the layers of ink as left with the same nib pen by the same author are of similar thickness, the delay would increase with the refractive index of the ink. As a result, the terahertz image displayed in Figure 5.22f and produced using the time delay of the pulse reflected from the back surface of the parchment, shows a legible "C" brighter than parchment, as lamp black has an optical behaviour close to that of graphite, with a very large refractive index, and therefore a large time delay. "U" and "L" are only partially legible, due to their lower refractive index. Lamp black, like graphite, is also strongly absorbent. Minium, as shown in Figure 5.27, has spectral windows showing much stronger absorption than iron gall ink. As a result, both "C" and "L" appear brighter than parchment in Figure 5.22g produced from the amplitude of the pulse reflected from the back surface of parchment, as discussed in Figure 5.20. While from the study of the iron gall ink dot on paper imaged in verso configuration the letter "U" would be expected to appear darker than the surrounding parchment, it is not distinguishable here (although the top and bottom parts of this character are slightly darker than parchment). This might be caused by two phenomena. Firstly, it was observed that the larger refractive index of some parchment sheets compared to paper sheets could lead to a decrease in image contrast of ink inscriptions, in either recto or verso configuration (which would also explain why "L" is not so easily distinguishable). Secondly, the letter "U" is placed at a position where the paper sheet comes in contact with the parchment sheet, as observed in the B-scan in d): there is no longer an air gap between the parchment and paper sheets, so there is no longer a sharp change in refractive index which could be defined as a clear interface. Finally, the best image of these inscriptions is obtained by combining the influence of the decrease in absolute amplitude and the increase in time delay of the pulse reflected from the back surface of the parchment caused by the presence of the ink in the verso: the area under the curve between the two time positions indicated with vertical lines in b) was calculated and used to produce the image shown in Figure 5.22h.
5.4. Depth pulse amplitude imaging

Figure 5.22: a) Raw and b) deconvolved terahertz waveforms reflected from the parchment book cover shown in Figure 5.21, and collected from the position indicated with the red circle in f. B-scans along the dashed line indicated in f, and produced using c) the raw waveforms and d) the deconvolved waveforms aligned to the maximum pulse. e) Photograph of the inscriptions written on the back surface of the parchment book cover, and f - h) the associated terahertz images acquired with a step size of 100 µm. Images were produced using deconvolved waveforms, aligned to the first main pulse, selecting f) the time delay of the pulse reflected from the back surface of the parchment sheet - air interface, g) the amplitude of this same pulse and h) the area under the curve between 14 and 15 ps.
5.4.3 Overpainting: the case of an Ethiopian codex

An Ethiopian codex of an unknown date of production, from a private collection, contained text (presumably Ge’ez script) written with black-brown ink on most pages (Figure 5.23a), while some pages contained miniature paintings (Figure 5.23b). For some of these miniatures it could be clearly assessed that they had been painted over the original text, as the black-brown characters could be visually observed through the lightly painted and slightly transparent areas. However, for the painting on the verso of the first page of the codex shown in Figure 5.23b, no Ge’ez characters could clearly be visually distinguished. Still, darker lines could be noted through lightly painted areas: a study [231] explains that the artist would first sketch a preliminary drawing with charcoal, and would then delineate its main outlines in black ink. The purpose of the scan of this document with terahertz TD imaging was to assess whether or not text was present underneath the painting, and if any other design information could be revealed, such as changes between the final and preliminary drawings.

The first step was therefore to investigate whether the black-brown text would lead to a good image contrast with terahertz TD imaging in reflection, when present on the front surface of the document. The recto of the first page was investigated first (Figure 5.23), placing the open document on the scanning stage of RIM. Figure 5.23d shows the image obtained from the amplitude of the pulse reflected from the air - recto of the sheet interface: no inscriptions could be clearly distinguished, implying that the refractive index of the black-brown ink used for the text is similar to the refractive index of the parchment sheet in the terahertz region of interest. Studies of Ethiopian manuscripts indicate that iron gall ink was unknown or at least uncommon in the production of such codices [232, 233], and the bathophenanthroline test performed on this text did not indicate the presence of free iron(II). Some of these studies however report that the ink was produced from soot (from the combustion of various materials) left, and sometimes boiled, in water (or juice, or a solution of burnt plant products) [232–234]. Some of the preparations described in these studies are reminiscent of the preparation of bistre ink. The NIR spectra of the black-brown characters shown in Figure 5.24a are similar to the NIR spectrum collected from the model bistre ink dot on parchment in Figure 5.24b, while model bone black and lamp black ink dots on parchment show no reflectance across the full spectral range.
5.4. Depth pulse amplitude imaging

Figure 5.23: a) Photograph of an Ethiopian codex from a private collection; b) photograph of the painting on the verso of the first page. Terahertz images of the area corresponding to the red circle in a) and b) were acquired in reflection mode with RIM and a step size of 150 μm, using a round magnet to maintain the page flat. The side with inscriptions corresponded to the front surface during the scan, while the side with the painting corresponded to the back surface. Images were produced using deconvolved data, and c) the time delay of the first main pulse corresponding to the air - parchment interface, d) the amplitude of this same pulse, e) the time delay (after alignment of the waveforms to the first main pulse) of the last pulse corresponding to the paint-air interface, f) the amplitude of this same pulse, and g) the area under the positive pulse corresponding to the parchment - paint interface.
Further analysis with other spectroscopic tools such as FTIR spectroscopy might bring more conclusive results on the nature of this ink. This characterisation was however not performed as it could only be done non-destructively by using the ATR mode: the clamp of the ATR accessory would have risked leaving an indentation at the surface of the sheet, which would have increased the complexity of the interpretation of any subsequent terahertz scan. Still, the literature on Ethiopian manuscripts mentioned above and the results of NIR spectroscopy seem to indicate that the ink is not made of pure carbon, and would have a composition closer to bistre than to lamp black or bone black, which would explain its low contrast in terahertz images in reflection mode.

An interesting feature from the front surface could still be extracted from the terahertz scan: the surface of the parchment sheet used for Ethiopian codices was ruled, so that horizontal grooves would be indented on the surface and would facilitate writing of the text in equally spaced horizontal rows [232, 234]. These horizontal indentations can be clearly observed in Figure 5.23c, produced using the time delay of the pulse reflected at the air - front surface of the parchment interface.

The ink used for the black outlines of the drawing on the verso of the sheet is visually darker than the ink used for the text, and its NIR spectrum displayed in Figure 5.24a shows the same reflectance behaviour than model bone black or lamp black ink dots on parchment. A better image contrast would therefore be expected for these inscriptions. When opening the book at the page with the miniature and imaging this page in reflection mode with the gantry system (adapted to scanning wider areas on large objects), large areas of the painting could be investigated without the need to move the book between scans. The images displayed in Figure 5.25b and c produced from the amplitude of the air - painting interface, before and after deconvolution, clearly show the black ink outlines as brighter, which coincides with the presumed nature of the ink. It is however interesting to note that the features drawn with this ink do not lead to a difference in contrast when observing the image in Figure 5.23f with the miniature in verso configuration, produced using the amplitude of the pulse reflected at the painting - air interface: absorption through the paint layer and reflection at the surface of the paint seem to compete and lead to a poor contrast.
5.4. Depth pulse amplitude imaging

Figure 5.24: a) NIR spectra acquired from the Ethiopian codex discussed in Figure 5.23, using acquisition points from blank parchment, black characters in the text and black lines used as contours in the painting; b) NIR spectra acquired from model ink dots on parchment, made of bistre, bone black or lamp black mixed with Arabic gum.

In this same configuration, a positive pulse is observed before the large negative pulse corresponding to the paint-air interface: this pulse corresponds to the parchment-paint interface (corresponding to an increase in refractive index, which explains the sign of the pulse). This same pulse is observed with the painting in recto configuration as a negative pulse after the main positive pulse reflected corresponding to the air-paint interface. After mapping the area under the curve of this pulse with the drawing in either verso (Figure 5.23g) or recto (Figure 5.25d) configurations, subsurface writing could be observed but are not easily matched to the visually observed grey lines. Ruler lines under the paint can, however, be retrieved, as observed in Figure 5.25d. Results of other imaging techniques, such as NIR reflectance imaging to try and accessed subsurface features, could be used for comparison.

Images produced using parameters of the TD signal have so far been discussed, but further information about the nature of the ink or support as well as topographic features might be accessed as well using signal in the FD.
Figure 5.25: a) Photograph acquired using a LED bulb as the light source, and a Nikon-D300s DSLR full spectrum modified camera, with a UV filter: the image was processed with Photoshop to improve the contrast of subsurface features. Terahertz images of specific areas overlaid on a), acquired in reflection mode using the gantry system and a step size of 100 µm. Images were obtained using the maximum amplitude of b) the raw and c) the deconvolved reflected waveforms, and d) the area under the negative pulse corresponding to the paint-parchment interface.
5.5 Spectroscopic imaging

5.5.1 Dispersive behaviour of optical parameters

While the constant refractive index is a good indicator of the reflective properties and delaying effects of a material with respect to the terahertz pulse, which encompasses all frequencies, materials may exhibit a dispersive behaviour in the terahertz range, as discussed in Chapter 2. Using Equation 4.19 from Chapter 4, the frequency-dependent refractive index from pigments and dry inks was calculated and is displayed in Figure 5.26. Additionally, their frequency-dependent absorption coefficients were calculated using Equation 4.26 from Chapter 4 and are displayed in Figure 5.27. The first thing to observe is that the samples of vermilion, minium, verdigris and dry iron gall ink used in this study exhibit a reproducible spectral signature at room temperature between 0.15 and 3 THz.

Figure 5.27 shows that vermilion and minium exhibit intense and narrow absorption bands at the resonance frequencies 38, 42 and 89 cm$^{-1}$ (1.15, 1.27 and 2.66 THz) and 55, 62, 71 and 84 cm$^{-1}$ (1.64, 1.86, 2.14 and 2.51 THz), respectively. Iron gall ink shows three broad absorption bands at 38, 51 and 64 cm$^{-1}$ (1.14, 1.53 and 1.92 THz), and verdigris shows a sharp absorption band at 34 cm$^{-1}$ (1.02 THz) and a broad absorption band at 76 cm$^{-1}$ (2.27 THz). The specimens of minium, vermilion, verdigris and iron gall ink studied here are therefore crystalline materials with a resonant behaviour in the terahertz range 0.1 - 3 THz.

It should also be noted that the absorption features from vermilion (cinnabar) and minium are in agreement with the spectral signatures of the supplies from Zecchi, Florence, analysed with a different TD set-up in transmission and available on the online database thzdb.org [107]. The supply of verdigris from Zecchi, analysed using FTIR in the range 0.6 - 13 THz and available in the same online database also displays absorption bands at $\simeq$1 and 2.25 THz, but the SNR is poor below 5 THz. The absorption features from iron gall ink presented here are, however, the first ever reported.
Figure 5.26: Frequency-dependent refractive indices from different dry pigments, inks and binders, mixed with HDPE powder and pressed in duplicate or triplicate pellets as described in Chapter 4. Spectroscopic data acquired with TPS Spectra 3000, TerraView. The lines represent the average of data from triplicate or duplicate pellets, and the error bars represent standard deviations.

The resonant behaviour translates into an "anomalous" dispersion in refractive index values, as described in Chapter 2: the refractive index is larger than the resonant frequency at lower frequencies, and lower at larger frequencies.

If pigments and inks exhibit larger absorption or reflection properties in specific spectral regions, an improved image contrast of the inscriptions could be obtained using the amplitude of the FD signal in these specific spectral regions. Besides, these features are characteristic of a specific pigment or ink, so the frequency-dependent behaviour of the contrast in the FD image could reveal the nature of the ink or pigment.
In other words, using the mathematical tool of Fourier-transform to convert TD data into FD spectra, and benefiting from the strong resonant behaviour of some pigments and inks in the terahertz region, terahertz TD imaging could be used as a hyperspectral imaging tool. Yet, for the analysis of archival documents constituted of a layer of ink thinner than 100 \( \mu m \) (Appendix G) on a layer of support of \( \approx 70 \) to 300 \( \mu m \), two major concerns arise:

1. The overlapping signals reflected from the different interfaces spaced by less than the width of the incident terahertz pulse create a complex TD waveform
which can be thought of as a sum of periodic signals with different frequencies. The Fourier-transform of this waveform is therefore likely to exhibit spectral features, regardless of the spectral signature from different materials within a stack of layers. The position and width of these spectral features would depend on the degree of overlap (i.e., the thickness and composition of each layer) and on the number of layers. What role would these spectral features, arising from the morphology of the document, play if terahertz TD imaging is used as a hyperspectral imaging tool?

2. The layer of ink is smaller than the wavelength of the incident terahertz pulse (100 - 3000 µm): it is therefore fair to question whether the terahertz pulses transmitted through the air - ink interface and reflected from the ink - paper or paper - air interfaces are affected by the presence of the ink to the extent that they undergo frequency-dependent absorption, characteristic of the nature of the ink. If not, then terahertz TD imaging cannot be used as a hyperspectral imaging tool for the analysis of archival documents.

These two concerns will be discussed in Sections 5.5.2 and 5.5.3, respectively.

5.5.2 Influence of topographic features and support thickness on the frequency-domain signal

To observe the influence of the morphology of the archival document on its spectral response when analysed with terahertz TD imaging, the same model inscriptions left by a clean nib pen on a sheet of print paper discussed in Figures 5.14 and 5.13, and the same sieve lines on a historical rag paper sheet discussed in Figure 5.15 are analysed in the FD. Figures 5.28b and d display the intensity of the FD spectrum for each frequency (y-axis) and for each pixel (x-axis) along the dashed line indicated in Figures 5.13 and 5.15. Figures 5.28b and d are therefore the FD equivalent of the TD B-scans displayed in Figures 5.13e and 5.15e. It can be observed that the spectral response of a paper sheet changes in the presence of indentations or sieve lines. As a result, there are frequencies (or spectral ranges) where the difference in amplitude between the FD signals from an area with no indentation or sieve line (indicated by arrows number 2 on Figures 5.28b and d) and an area with indentation or sieve line (arrows number 1) is maximum.
5.5. Spectroscopic imaging

Figure 5.28: a) Terahertz image from indentations left with a clean nib pen at the surface of a print paper, obtained using the amplitude of the FD signal at 0.96 THz. b) Associated FD B-scan along the same horizontal line as indicated in Figure 5.13c. c) Terahertz image from sieve lines on an historical paper sheet, obtained using the amplitude of the FD signal at 1.75 THz. d) Associated FD B-scan along the same horizontal line as indicated in Figure 5.15d. Arrow 1 indicates a pixel corresponding to the presence of an indentation or sieve line, while arrow 2 indicates a pixel with no specific topographic feature.

An example of such frequency is indicated by the horizontal dashed line. Figures 5.28a and c correspond to terahertz images produced at these frequencies. The inscription "UCL" can no longer be clearly distinguished in Figure 5.28a, compared to Figure 5.13b. However, sieve lines seem to alter the spectral response of rag paper more significantly than indentations do on print paper: the image of sieve lines produced from the difference in amplitude at 1.75 THz shows an improved contrast compared to the images presented in Figure 5.15, with an additional advantage that the resulting image also has a better spatial resolution, due to the use of a high frequency component of the signal.
5.5.3 Extraction of spectral information from an ink inscription

Figure 5.29 shows the raw TD B-scans and waveforms from ink dots on rag paper. We first note that the front and back surfaces of rag paper can be observed at around 23 ps and 24 ps, respectively. The back surface is delayed with the presence of an ink dot on the front surface, as the ink locally increases the optical path, due to both the added thickness of the ink layer, and the often larger refractive index of ink compared to the refractive index of paper. This delay modifies the overlap between the signal reflected from the air - front surface of the sheet and air - back surface of the sheet, and makes it possible to distinguish more or less clearly an additional peak as discussed in Figure 5.3, indicated with the star in Figure 5.29b. This creates an additional layer between the front and back surface of the sheet on the TD B-scans in Figure 5.29a. In the case of lamp black on rag paper only, neither this "additional" layer nor the back surface of the paper sheet are observed where ink is present on the front surface, simply because the lamp black ink dot has absorbed the entire incident pulse, and no pulse is transmitted through the dot.

An interesting phenomenon is observed in the TD B-scans from iron gall ink and minium ink dots on rag paper, and also in the TD B-scan from vermilion shown in Figure 5.30: a "ringing" effect is observed in the tail of the TD waveform reflected from an inked area, which translates into "ripples" observed after the back surface of the paper sheet on the TD B-scan, as if additional layers were present under the paper back surface. This phenomenon is systematically observed in all iron gall, minium and vermilion ink dots on rag paper and parchment sheets, although far less pronounced for iron gall ink dots, and is never observed for ink dots made of other pigments. These ink dots are the only pigments with characteristic spectral signatures in the terahertz range (with the exception of verdigris with a weak spectral feature at a low frequency, and a very broad feature at a high frequency). More specifically, the "ringing" and "rippling" phenomena are very clearly observed for the two pigments with very strong and narrow absorption bands, minium and vermilion.
This ringing effect is also clearly observed in the tail of the waveforms transmitted through pellets of minium or vermilion mixed with HDPE, as indicated in Figure 5.31. By calculating the time period of this "ringing" effect from the difference in time positions between 18 or 11 successive peaks, its frequency can be recovered. The frequency of \( \approx 1.7 \) THz was found for pellets of minium, and 1.2 THz for vermilion. These frequencies correspond to spectral regions of high absorption in the absorption coefficients of minium and vermilion shown in Figure 5.27. When analysing the frequency of the "ringing" effect from the tails of waveforms reflected from minium and vermilion ink dots on rag paper and parchment sheets (see Figure 5.30), similar values are obtained (1.7 - 1.8 THz for minium and 1 - 1.1 THz for vermilion).

Figure 5.29: a) Terahertz TD B-scans along a line across an ink dot on rag paper (sepia, malachite, lamp black, iron gall ink or minium mixed with Arabic gum), and b) associated raw TD waveforms extracted from pixel positions indicated with numbers. The inlet in b) shows the tail of the waveforms collected in positions 6 and 7 with a magnified scale. The star indicates the presence of an additional TD feature due to the presence of an ink dot at the surface of the sheet.
Figure 5.30: Tail of the TD waveforms reflected from a) mimbium and b) vermilion ink dots on rag paper with and without additional sizing, and on parchment, acquired from terahertz TD imaging scans using RIM, TeraView. "Backed" implies that a flat piece of cardboard was placed against the verso of the parchment sheet, between the two magnetic discs holding the sample, to flatten the surface of the parchment. The frequency of the observed sinusoidal "ringing" effect was calculated from the time-period between subsequent peaks.

It can therefore be concluded that there is a frequency-dependent absorption of the terahertz pulse transmitted through the thin layer of ink applied on paper and reflected at the ink - paper interface, and that this frequency-dependence is characteristic of the vibrational properties in the region 0.1 - 3 THz of the pigment. Unfortunately, this "ringing" effect has a weak amplitude and is significantly attenuated or lost when:

- Imaging the same ink dots in "verso" configuration, i.e. with the incident pulse impinging on the air - support interface, then support - ink, and finally ink - air: this is due to the fact that the reflected pulse that has been partially absorbed by the ink layer on the back surface travels through a layer of support before reaching the detector in reflection mode, and is therefore attenuated.

- When using a High Frequency (HF) cut-off parameter too high or too low for the double Gaussian filter used in the deconvolution process. An HF too high could cut-off this signal at 1.2 or 1.7 THz, and an HF too low would increase
the amplitude of noisy high frequency signals, which will obscure the signal of interest [76].

![Figure 5.31: Tail of the TD waveforms transmitted through a) minium and b) vermilion pellets, acquired in a nitrogen-purged environment using TPS Spectra 3000, Teraview.](image)

When analysing FD signals reflected from these ink dots with a strong resonant behaviour, it can be assumed from the results presented above that a different signal amplitude would be observed around 1.2 THz (38 cm$^{-1}$) for the vermilion ink dot, and around 1.7 THz (55 cm$^{-1}$) for the minium ink dot, compared to the amplitude of the signal reflected from the blank paper. Yet, the influence of topographic features and support thickness on the FD signal has also clearly been demonstrated above. Adding a layer of ink to a sheet of support would increase the complexity of the reflected TD waveform and would therefore change its spectral response, regardless of the resonant behaviour of the pigment within the ink layer.

Figure 5.32 compares the TD and FD B-scans along a line crossing the vermilion and minium ink dots on rag paper, and the influence of the sample morphology can be clearly observed. As discussed above, the presence of an ink layer introduces an increase in the optical path, which changes the time delay between the signals reflected from the front and back surfaces of the sheet. This phenomenon, and the presence
of a new interface, create the appearance of an additional layer between the front and back surface of the sheet, *i.e.* the grey patches observed between 23 and 24 ps in the TD B-scans for pixels within the inked area (Figures 5.32 *a* and *c*). Since pigment distribution within the ink dot, thickness of the ink layer and penetration of the ink into paper fibres are not perfectly uniform across an ink dot, two grey patches are observed for the minium ink dot and three for the vermilion ink dot, instead of a continuous grey layer. The exact same grey patches are observed in the FD B-scans within a frequency range where the rest of the pixels display no (or low) signal (Figures 5.32 *b* and *d*).

---

**Figure 5.32**: TD and FD terahertz B-scans along a line crossing *a* and *b*) a minium ink dot on rag paper and *c* and *d*) a vermilion ink dot on rag paper.
The spectral region below 1 THz is mostly influenced by the peak-to-peak amplitude of the low-frequency signal corresponding to reflection from the front surface: as a result, a larger constant refractive index of the ink layer leads to a larger amplitude of the FD signal below 1 THz. This is why the pixels corresponding to the inked area are very bright below 1 THz. With regards to resonant behaviour of the ink layer, it is difficult to observe a clear decrease in amplitude at either 1.7 or 1.2 THz for the pixels corresponding to minium and vermilion ink dots, respectively.

To summarize, when imaging a document with terahertz TD technology in reflection mode, the terahertz pulse can undergo dispersive absorption when transmitted through a thin layer of ink on paper or parchment before it is reflected on the ink-paper interface, but it seems that the change in the layered structure of the document with and without ink significantly modifies the shape of the reflected TD signal. These morphological modifications influence more the FD response of the document than the resonant behaviour of the materials present within the document.

However, in transmission mode, the signals transmitted through the single-sheet document with and without the model ink dot at the surface of the support have very similar shapes, as illustrated by the black and grey lines in Figures 5.33g, j and m. TD signals in reflection mode are also shown in Figures 5.33a and d for comparison. A vermilion ink dot on parchment was placed in the nitrogen-purged sample compartment of TPS Spectra 3000, TeraView, to be analysed with transmission spectroscopy. As discussed in the Chapter 4, the diameter of the spot size at the focal plane is approximately equal to 3.5 mm (iris aperture ”20”) at 100 cm$^{-1}$, 6.5 mm (iris aperture ”30”) at 40 cm$^{-1}$ and 9.5 mm (iris aperture ”40”) at 20 cm$^{-1}$.

To observe a spectral signature of an ink dot of $\approx 2$ mm of diameter, the iris diaphragm enables us to reduce the size of the spot around the dot, and therefore it reduces the influence of the surrounding support and enhances the influence of the ink dot on the spectral signature of the transmitted signal. Yet, vermilion absorbs strongly between 35 and 45 cm$^{-1}$, therefore the aperture should not be set lower than ”30”, in order to avoid strong scattering of the incident beam in the frequency range of interest. Figures 5.33g, h and i show the results of the single-point analysis (i.e. spectroscopy) in transmission mode of the vermilion ink dot on parchment. The signal is transmitted through a blank area of the parchment set as a reference (using the same aperture
size), and over 1800 scans are averaged for both the reference and the sample signals. Although the FD responses from blank parchment and parchment with ink dot are very similar, the absorbance spectrum indicates that stronger extinction of the incident pulse occurs around 35 cm\(^{-1}\). The experiment was repeated with an iris aperture of ”22.5” (dashed lines) and the same extinction band is observed, although the amplitude of the FD responses and the SNR of the absorbance spectrum are lower than for the aperture ”30”, as expected. This extinction band is within the range of large absorption and reflection for vermilion (see Figures 5.27 and 5.26): the spectral signature characteristic of vermilion can therefore be probed in the FD using the transmitted pulse.

Yet, when imaging a vermilion ink dot on parchment, or even on a less absorbing support like paper, with an imaging set-up set in transmission mode and with the sample analysed in the ambient environment (T-Ray 4000, Picometrix, Figures 5.33 to o), two major issues arise, preventing the observation of the characteristic dispersive behaviour of vermilion:

- A ringing effect is observed in the tail of both the reference and sample waveforms, and arise from the dispersive absorption behaviour of ambient water vapour. As a result, the absorption bands characteristic of water vapour are observed in the reference and sample FD responses (19, 25, 33, 37, 39, 40, 41 and 47 cm\(^{-1}\); the same spectral bands as measured with TPS Spectra 3000, using a sample compartment filled with ambient water vapour - data not shown here). Yet, as the ambient relative humidity can change during the acquisition of an imaging scan, these same bands are still observed in the absorbance spectrum (either as positive or negative bands) calculated using signals from a pixel within a blank parchment or paper area and from a pixel within the ink dot. The band of water vapour at 33 cm\(^{-1}\) would unfortunately mask the extinction band due to the presence of vermilion.

- Absorption from ambient water vapour lowers the SNR and reduces the possibility to observe the spectral signature of vermilion. Plus, when acquiring an image, the number of scans acquired per pixel is often much lower than when performing a point measurement for spectroscopy, in order to reduce the image acquisition time. In the experiments performed with T-Ray 4000, Picometrix,
10 scans were acquired and averaged, compared to 1800 scans for spectroscopic measurement with TPS Spectra 3000, TeraView. This also leads to a poorer SNR and dynamic range, and decreases the likelihood to observe resonance behaviour of the vermilion ink dot.

In light of the above, the results published by Fukunaga et al. [38], showing the use of a terahertz TD imaging set-up in transmission mode as a hyperspectral imaging tool for the analysis of a historical document in the terahertz range, can be much better appreciated. While the number of scans performed to image the document is not specified, the authors do specify that the system used was developed by the RIKEN research institute [235]. This is a highly sensitive terahertz imaging system using a surface plasmon resonant terahertz sensor. The detection sensitivity and the imaging contrast for small sample amounts are reported to be considerably better than those of conventional terahertz transmission imaging without such sensor. This superior sensitivity would explain why the spectral signature of vermilion (cinnabar) can be very clearly identified in the spectral response from the signal transmitted through the red ink, and would explain the dispersive behaviour of image contrast of the ink, optimized at 1.05 THz, i.e. 35 cm\(^{-1}\). In conclusion, terahertz TD imaging can be used as a hyperspectral imaging tool for analysis of single-sheets documents with inks showing a strong resonance behaviour in the terahertz range only if the system is set in transmission mode, and if using experimental conditions optimizing the SNR for ink: a highly sensitive sensor, a purged environment, a significant number of scans acquired per pixel, a small spot size etc.
Figure 5.33: a, d, g, j and m) TD signals, b, e, h, k and n) FD signals and c, f, i, l and o) absorbance spectra, collected from a single sheet of support with (black line) and without (grey line) a vermilion ink dot on its surface. Absorbance spectra were calculated by taking the logarithmic ratio of the two FD responses. a to f) were extracted from a set of imaging data produced in reflection mode with RIM, Teraview, for a vermilion ink dot on a - c) rag paper, and d - f) parchment. g to i) correspond to single-point analysis of a vermilion ink dot on parchment placed in the nitrogen-purged sample compartment of TPS Spectra 3000, Teraview. An iris diaphragm was placed around the ink dot, with an aperture of "30" (6.5 mm, solid line) and a smaller aperture of "22.5" (dashed line). j to o) were extracted from a set of imaging data produced in transmission mode and in the ambient environment with T-Ray 4000, Picometrix, for a vermilion ink dot on j - l) parchment, and m - o) rag paper.
5.6 Future work and trends

5.6.1 Image processing tools to improve contrast

Once a terahertz image is produced using a specific parameter of the terahertz signal, the first step to improve its contrast is, of course, to adapt its grey scale / brightness scale. Image processing algorithms can further be used to improve the image contrast of inscriptions and for character recognition. Image segmentation consists of grouping pixels into regions according to pre-defined criteria. The simplest image processing based on segmentation consists of choosing a threshold value within the image brightness scale, and making the image binary: every pixel value above the threshold value is set to 1, and the rest is set to zero (see Figure 5.34b). Image segmentation can also be based on a region growing algorithm: an initial seed pixel is selected and the brightness value of its neighbouring pixels are compared to both its brightness value and a user-defined threshold value to determine whether or not these pixels should be integrated in the grouping region (see Figures 5.34c and d). Pixel values can also be statistically analysed to define groups, such as with the use of K-means clustering algorithms (see Figure 5.34e), assigning a pixel to a grouping region with a mean brightness value closest to the brightness value of that pixel. Region growing and K-means clustering algorithms are supervised, since the user pre-defines a seed pixel or a few pixels to be considered as belonging to different groups, also called initial cluster centres. While such algorithms are freely available online in Matlab, or are already implemented in commonly used software tools such as ImageJ or Origin Lab, the difficulty lies in finding the right seed pixel or initial cluster centres, and the right threshold value to improve legibility of the inscriptions. Computer science expertise would be required to progress this line of research.

Images produced using different parameters of the terahertz signal (amplitude of the FD signal at a specific frequency and maximum amplitude of the TD signal, for example) and revealing different features or characters can be merged, or superimposed with different opacity ratios, to produce a holistic and contrasted image of the document, using commercially available image processing tools such as Adobe Photoshop, or Inkscape. This idea was proposed by Bianca Jackson in 2014 to reveal the shape of a bird in a terahertz image from wall etchings in a Paleolithic cave [237].
Figure 5.34: a) Original terahertz image obtained from the maximum amplitude of the waveform reflected from a typewritten inscriptions on the letter dated 13.07.1946. b) Image processed using a threshold of 210 over a brightness scale of 0 - 255 and a conversion to a binary image with the ImageJ software tool. c) Image processed using a Matlab region growing algorithm [236], with the pixel labelled "4" in a as a seed, and a threshold of 0.15. d) Image processed using the same region growing algorithm, with the pixel labelled "2" in a as a seed, and a threshold of 0.0833. e) Image processed using a K-means clustering algorithm implemented in the Origin Lab software, with the 5 pixels indicated with red rectangles in a as initial cluster centres.

5.6.2 Continuous Wave imaging

In Section 5.3.2.3, it was discussed that Prussian blue pigment was detected within the ink used for the historical typewritten letters, and that the constant refractive indices from Prussian blue and Antwerp blue pigments acquired from art suppliers were similar to the constant refractive indices from blank paper samples cut from the historical letters. This explained why no ink inscriptions could be observed in the terahertz images of the letters produced using the maximum amplitude of the deconvolved reflected waveforms (see Figures 5.17b, h and n). When investigating the dispersive behaviour of those same samples, it can be observed that while the refractive indices of Prussian and Antwerp blues are relatively frequency-independent (flat curve), the absorption coefficients show a low and very large slope before and after ≃60 cm\(^{-1}\), until the detection limit of the set-up is reached at ≃71 cm\(^{-1}\) (≃79 cm\(^{-1}\) for Antwerp blue, which additionally contains barium sulfate, as discussed in Chapter 3, and therefore has a lower effective pigment to HDPE mass concentration in the associated pellets). The results comparing pellets with pigment concentrations of 10% and 1% shown in Chapter 4 indicate that this region of strong absorption extends to ≃100 cm\(^{-1}\).
5.6. Future work and trends

The spectral region between $\simeq 70$ and $100\ \text{cm}^{-1}$ therefore corresponds to a region of high absorption for inks containing Prussian blue, in comparison to the absorption from paper in the same spectral region. It is therefore expected that the historical letters typewritten with Prussian blue studied in this chapter would lead to terahertz images with a very good contrast between the ink inscriptions and the surrounding support, if studied in transmission mode in the spectral range $70 - 100\ \text{cm}^{-1}$.

Yet, when imaging with a terahertz TD set-up, the SNR is often poor in this spectral region, particularly if the number of acquired and averaged signals per pixel is reduced in order to reduce the acquisition time. In collaboration with David Giovannacci from the Laboratoire de Recherche des Monuments Historiques (LRMH), France, and Bianca Jackson, University of Reading, UK, T-Ray 4000, Picometrix, was set in transmission mode and the signature "RANDALL", typewritten on the letter from 13.07.46, was imaged in the ambient environment with 500 averaged scans per pixel (compared to 10 averaged scans per pixel in the study of ink dots with the same set-up). The result of the scan was still unsuccessful: the amplitude of the transmitted signals reached the noise floor from $\simeq 40\ \text{cm}^{-1}$ (1.2 THz) onwards and the inscriptions were not legible on terahertz images produced using parameters from either TD or FD. However, using a CW set-up in transmission mode with an incident beam with a frequency within the range $70 - 100\ \text{cm}^{-1}$, a better SNR could be obtained and the Prussian blue inscriptions could lead to well-contrasted terahertz images. Unfortunately, by using a CW set-up,
depth information would be lost if using fixed sensor heads and sample: the image collected from a scan in transmission mode would be the superposition of information from all layers and it would be very difficult to reconstruct textual information originating from one specific layer. However, as with X-ray absorption imaging, CT could be used to recover depth information. Such experiment with a CW set-up would be worth trying, but was unfortunately not possible within the time frame of this project.

5.6.3 Algorithm to remove occlusions in complex artefacts
In the case study discussed in Figure 5.2, removal of the shadows of the inscriptions on the image from the second and third fold was attempted manually, by selecting a specific pixel within the shadowed area on the image of the second (or third fold), and adding the weighted value of the same pixel from the image of the fold(s) above: to no longer observe occlusions, the result of the sum should equal the result of the same sum for pixels corresponding to blank paper on all folds. Researchers are currently developing algorithms to automatically remove accumulating occlusions in even more complex systems with up to 9 overlaid sheets with inscriptions [229].

5.6.4 Algorithm to extract the pulse reflected from the layer of interest in complex layered systems
In the case study discussed in Figure 5.2, the presence of air gaps between folds (low refractive index), their thickness and the thickness of the paper layers in relation to the terahertz pulse width and duration, ensure the observation of intense pulses corresponding to reflections at the air - surface of the fold interfaces (this was true for both the raw and the deconvolved signals). Only a very simple algorithm was needed to extract the amplitude of the pulses corresponding to the subsequent air - surface of the first fold, air - surface of the second fold and air - surface of the third fold interfaces, as discussed above, and reconstitute the image from the inscriptions at the surface of each fold.

However, in complex multi-layered systems with a varying number of layers across the scanned area, with some thin layers in relation to the pulse width and duration, with layers showing non-flat surfaces, and layers with close refractive indices, the pulse reflected at the surface of the same material may either show a large or a low amplitude, depending on the material and on the thickness of the adjacent layer. Such a case can be found when analysing overlaid painted scenes in an easel painting
with terahertz TD imaging in reflection [120]. In such a case, a custom algorithm, with user-defined windowing functions, threshold values and number of interfaces, must be developed and applied to the deconvolved signals, in order to extract the signal reflected from the surface of the material of interest (e.g. the ground layer) and reveal a hidden painted layer. The development of such an algorithm can be enabled by knowing the number of layers to be examined, using, for example, physical paint cross-sections sampled from the painting [120].

When imaging 9 overlaid sheets with inscriptions, Redo-Sanchez et al. observed that deconvolution methods using frequency-based filters (a double-Gaussian filter such as the one used in this study, or a Wiener filter as used in the study of the easel painting [120]) do not work well when the SNR during image acquisition is poorer than 10 dB. The authors recommend the use of a probabilistic pulse extraction method to extract a pulse from the layer of interest in densely layered samples imaged with a terahertz TD system in low SNR conditions. They used this method to reveal textual content from individual sheets of paper [229].

5.6.5 Expanding the library of terahertz optical properties to modern archival materials

As discussed in Chapter 3, the selection of pigments and inks used in the above study was based on the frequency of their use in archival documents, their predisposition to degrade the carrier, the availability of their components, and the feasibility of their production in a laboratory. Yet, archival collections also contain documents from the last two centuries, which may have been written with synthetic dyes and pigments such as aniline, nigrosine, methylene blue and Prussian blue, amongst others, as mentioned in Chapter 3 and as illustrated by the letters from 1946 discussed above.

In the historical reference collection of Islamic paper from the ISH, a single-sheet document from the 19th century written in Ottoman Turkish script clearly displays inscriptions written with two different types of ink (see Figure 5.36a). The ink used for the main text (ink b, second line) shows stains typical of degradation products from iron gall ink accumulating on the other side of the page (brownish stains), while the ink used for the header inscriptions (ink a, first line) does not. The bathophenantroline test indicates the presence of free iron(II) ions in ink b, and not for ink a.
Figure 5.36: a) Photograph of an Islamic manuscript from the historical reference collection of the ISH (sample ID: AP50). The circled area was scanned in reflection mode with TPS Spectra 3000 with RIM, with a scanning step size of 150 µm. b) Terahertz image obtained from the maximum amplitude of the main pulse from the raw TD signal. c) Terahertz image obtained from the maximum amplitude of the main pulse from the TD signal after deconvolution. d) and e) Terahertz image obtained from the amplitude at 13.57 ps and 13.78 ps, respectively, of the TD signal after deconvolution and time alignment of the main pulse from all pixels. f) Terahertz image obtained from the amplitude at 2.2 THz of the FD signal. g) TD signals from different pixels with either ink a on the recto, ink b on the recto, ink b on the verso or blank paper. Signals are displayed after deconvolution and time alignment of the main pulse from all pixels, to facilitate the interpretation.
5.6. Future work and trends

It can be concluded that the inscriptions surrounded by brownish stains are written with iron gall ink containing free iron(II), while the other black inscriptions have a different composition.

Figure 5.36 shows the results of the terahertz TD imaging in reflection mode of the document, using TPS Spectra 3000 with RIM, and placing the document within a round magnet to maintain it flat. Using the maximum amplitude of the main reflected pulse, from either the raw or the deconvolved waveforms (Figure 5.36b and c respectively), only inscriptions written with ink b appear very well-contrasted, while inscriptions written with ink a are not legible. Using the amplitude of deconvolved waveforms at the time position corresponding to the ink - paper interface represented by a negative peak (13.57 ps in Figure 5.36g), a very good contrast is still observed for inscriptions with ink b, but inscriptions with ink a show a better contrast and start to be legible (see Figure 5.36d): a very faint negative peak is observed in the waveforms from ink a at this time position (black line in Figure 5.36g), while no negative pulse is observed for the waveform reflected from paper (blue line in Figure 5.36g).

The presence of ink b on the verso of the sheet introduces a paper - ink interface, which would lead to a positive reflected pulse (due to the change from low to high refractive index). This positive pulse is observed at 13.78 ps in the waveforms reflected from the paper with ink inscription on the verso (green line in Figure 5.36g) and the associated image in Figure 5.36e indicates the location of ink b on the verso of the sheet in white, matching the position of the brownish stain in the photograph in Figure 5.36a.

In the above study of sieve lines, it was observed that images in the FD at a high frequency are likely to lead to an improved image contrast and resolution. As a result, Figure 5.36f was produced from the amplitude of the FD signal at 2.2 THz and very clearly displays the lead and chain lines of paper. It is interesting to note that although Figure 5.36c shows a better contrasted ink b than Figure 5.36b, the presence of ink on the verso of the sheet and of the top chain line are discernible in dark in Figure 5.36b while they are less discernible or not at all in Figure 5.36c. This demonstrates once more both the gain and loss of information when performing deconvolution.
Figure 5.37: a) Near-infrared reflectance spectra from triplicate point measurements on blank paper (black lines), ink lines of ink a (red lines) and ink b (blue lines). A reflectance value of 1 was set as the reference with a spectralon white standard. b) Average ATR-FTIR spectra from triplicate point measurements on blank paper (black line), ink lines of ink a (red line) and ink b (blue line). Spectra were normalized at 1027.58 cm$^{-1}$.

To understand the poor contrast of the black ink a in terahertz images in reflection, ATR-FTIR and NIR spectroscopic analysis were performed on both ink a, ink b and blank paper to try and withdraw information on the chemical composition of the two inks. Results are displayed in Figure 5.37. Both inks absorb strongly from 500 to 2500 nm (reflectance close to zero). Comparing these spectra with the spectra from model and historical carbon-based and iron gall inks available in Appendix H, it would seem that this near-infrared absorption behaviour is characteristic of the presence of lamp black or ivory black (and the chemically similar pigments carbon black and bone black). Mixed inks, containing both iron gall ink and carbon-based pigments such as lamp black, bone black or ivory black, have traditionally been used in Islamic calligraphy [238]. Although the ATR-FTIR spectra from both inks in Figure 5.37b are strongly dominated by the spectral signature from the paper underneath, no significant absorption is observed at 560, 1020 or 2010 cm$^{-1}$ (markers of the signature of ivory black and bone black as discussed in Chapter 3). It can therefore be assumed that ink b is a mixed ink made of iron gall ink and lamp black (or any black made from other burnt resin or oil). The presence of the added lamp black pigment in ink b would explain the very good contrast observed in terahertz images from inscriptions written with this ink. The stronger absorbance at $\approx 1317$ cm$^{-1}$ and in the region 1620 - 1650 cm$^{-1}$ in the ATR-FTIR spectrum from ink b may indicate the presence of the carboxylic, ketone and aldehyde functional groups, which could be linked to oxidation of cellulose in
paper caused by the iron gall ink.

Due to the very poor contrast observed in terahertz images in reflection mode from ink \(a\), it seems unlikely that this ink contains lamp black or carbon black, as it was demonstrated above that such pigments show very good image contrast in the terahertz region. Instead, the following hypothesis can be formulated. The inscriptions written with ink \(a\) correspond to numbers "40" (on the verso here) and "41" (on the recto) written in Ottoman Turkish script (the same script as the main text, taught before 1928 [239]) and are likely to indicate the page number of the folio. Some of the dyes used in fountain pen inks during the 20\(^{th}\) century are near-infrared absorbent, such as, for example, copper phthalocyanine and nigrosin [240–244]. Yet, such dyes were introduced later than the date of production of the document as disclaimed by the supplier (19\(^{th}\) century) [240]. Comparing this document with other documents from the historical reference collection of Islamic paper from the ISH and from documents available online [245–247], this page numbering practice seems uncommon. Therefore, it is possible that the document was produced in the 19\(^{th}\) century, but that the page numbering was added in the 20\(^{th}\) century by a person familiar with the Ottoman Turkish script, using an ink of a more modern composition. To confirm this hypothesis, chemical characterisation of the document with complementary analytical techniques would be necessary, but the study of the optical properties in the terahertz region of dyes and pigments from the past and present centuries would be equally as important.

5.6.6 X-ray attenuation and phase contrast imaging

The X-ray imaging study published in 2009 of the carbonized scroll from the Herculaenum site used the attenuation of the irradiance (energy) of the X-rays transmitted through the scroll as the contrasting parameter, and did not provide a contrasted image of inscriptions within the scroll [248]. It was argued that this was because carbon-based ink was likely used for the production of the document. Such inks have the same X-ray absorption properties as the papyrus support. Yet, as discussed in this chapter and in Chapter 3, black carbon-based pigments exist with different chemical compositions, and could therefore display different optical properties in the X-ray spectral range (0.01 - 10 nm, \(3.10^{16}\) - \(3.10^{19}\) Hz, 100 eV - 100 keV): it is possible that some carbon based ink inscriptions with a larger metallic content absorb X-rays more strongly than the sur-
rounding support, and can be detected with X-ray attenuation imaging. Furthermore, the most recently published X-ray imaging study of a carbonized scroll from the Herculaeum site used the changes in phase as a contrasting parameter, and did provide contrasted images of inscriptions within the scroll [47]. Yet, phase-contrast imaging is sensitive to phase change, i.e. changes in the optical path, either arising from changes in refractive index across the path of the beam, or changes in thickness / topography, much like what is observed with terahertz TD imaging in reflection mode. To better understand the origin of contrast in X-ray attenuation and phase-contrast imaging of documents, a similar study to the one performed above with model ink dots and various historical documents would be worth performing in the X-ray spectral range with different set-ups. A preliminary study was performed in collaboration with Jennifer Griffiths and the X-ray phase contrast imaging group at UCL Institute of Biomedical Engineering, on carbon-based ink dots applied on historical parchment. The results are displayed in Figure 5.38. They indicate that inks containing a larger metallic content, such as sepia ink $S$ containing sodium in NaCl salt and bone black ink $IB$ containing calcium, can be detected with X-ray attenuation imaging using a mammographic x-ray source, which is a mono energetic source with a peak energy between 17 and 18 keV, while they could not be detected with a micro-CT system with a source operating at 40 kVp (maximum beam energy of 40 keV). Additionally, a larger mass ratio of pigment to binder leads to a better image contrast. With X-ray phase contrast imaging, none of the inks gave a different response to the surrounding parchment. The only sources of contrast are the raised edges of the ink dots leading to circular shapes in the X-ray phase-contrast image, particularly for lamp black $LB1$ and sepia $S$ on the bottom line. The discrepancies between the results shown in Figure 5.38c and the published X-ray phase contrast images of the Herculaneum scroll might be explained by the presence of lead in the ink, as revealed by a recently published study [249].
Figure 5.38: 

- **a)** Photograph of carbon-based ink dots on parchment: $LB1$ represents lamp black pigment mixed with rabbit skin glue, $LB2$, $IB$ and $B$ represent lamp black, bone black and bistre pigments mixed with Arabic gum, $S$ represents sepia ink. The first, second and third lines correspond to a pigment to binder mass ratio of 1:6, 1:3 and 1:2, respectively; 
- **b)** X-ray absorption image; 
- **c)** X-ray phase-contrast image.
5.6.7 Diversity of supplies and round robin tests

As shown in Figure 5.35, Prussian blue powders acquired from different suppliers but with the same composition (as assessed by their mid-infrared spectra) have the same optical properties in the terahertz range. This approach of diversifying the supplies therefore strengthens the reliability of the terahertz optical properties from Prussian blue presented in this thesis. While it was not possible to adopt this approach for all materials investigated during this project, due to its limited time-frame, the spectral signatures of vermilion, minium, lapis, malachite and verdigris acquired from Cornelissen and Son, London, are in agreement with the spectral signatures from the same pigments supplied by Zecchi, Florence, provided elsewhere [107], which strengthens their reliability. However, to assess if the refractive index and absorption coefficient of a pigment supply presented in Figures 5.26 and 5.27 faithfully renders the optical properties of the pigment in question, it would be recommended to perform further analysis of the pigment using different supplies, but also using different set-ups (round robin test).

In this Chapter, it was discussed that iron gall ink, vermilion and minium have a characteristic spectral signature in the terahertz range 0.15 - 3 THz, and that this signature could be probed in an ink inscription on a support. The results suggest that THz-TDS or imaging could be used as a non-invasive tool to detect the presence of corrosive iron gall ink in documents, differentiating it from inert carbon-based black ink, and could become a valuable alternative to the use of the bathophenanthroline indicator, which requires contact with the surface of the document. Yet, the spectral signature and optical properties of iron gall ink have only been shown for one recipe of iron gall ink. A large variety of recipes has historically been used, which might lead to different crystallisation processes upon drying and therefore different spectral responses in the terahertz region. Similarly, the spectrum and optical constants of only one parchment sheet have been displayed and, contrary to paper, there exist no published sources for comparison: the parchment sheet studied here has larger reflective and absorption properties than the other studied paper sheets and than the values for paper found in the literature. Do other iron gall inks and parchment sheets have the same optical properties and spectral responses as those studied in this Chapter? If not, what parameters participate in the change of optical properties and spectral responses? These research questions will be explored in the next Chapters.
Chapter 6

Terahertz time-domain spectroscopy of supports and iron gall inks

6.1 Terahertz time-domain spectroscopy of supports

Samples of $\approx 2 \text{ cm} \times 2 \text{ cm}$ were cut from larger historic and contemporary parchment and paper sheets, maintained flat with a round magnet and placed in the nitrogen-purged compartment of TPS Spectra 3000. Each sample was analysed in triplicate while changing the orientation of the sample between measurements, to account for any scattering that might depend on the orientation of paper or parchment fibres. Absorbance spectra of 32 historic parchment sheets from the reference collection of the ISH are displayed in Figure 6.1.a. While no spectral feature is observed above 35 cm$^{-1}$, apart from features from water vapour at $\approx 56$, $88$ and $93$ cm$^{-1}$, broad bands are observed for all parchment sheets, but at different frequencies, below 35 cm$^{-1}$. As discussed in Chapter 4, these bands are not absorption bands caused by vibrational properties of the material, but are caused by the delayed pulse(s) arising from internal reflections within the sheet and overlapping with the main pulse reflected at the air - front surface of the sheet interface (the so-called "etalon effect"). Since the time delay of the subsequent pulse(s) is dependent on the optical path within the sheet, it varies with the thickness of the parchment sheet. As a result, in the FD, the number and position of the broad bands observed below 35 cm$^{-1}$ depend on the thickness of parchment. This is particularly evident when comparing the TD waveforms and absorbance spectra from parchment sheets sorted according to their thickness, as shown in Figure 6.2. It should be noted that the thickness of a parchment sheet is not systematically homoge-
neous across a sheet. Consequently, when analysing three different samples cut from the same parchment document (used as a support for the imaging study of ink dots discussed in Chapter 5) but with a different thickness, the bands due to the etalon effect are observed at different frequencies, as illustrated in Figure 6.3.

Figure 6.1: Absorbance spectra from a) 32 parchment sheets of ≃2 cm x 2 cm sampled from documents dating from the mid-17th to the early 20th century, and b) a pellet of cellulose (grey) and 11 paper sheets of ≃2 cm x 2 cm sampled from contemporary printing paper (blue) and filter paper (brown), 19th rag paper sheets (black and red) and letters from 1946 discussed in Chapter 5.

Figure 6.2: a) Absorbance spectra from 32 parchment sheets, also presented in Figure 6.1a, but sorted according to the thickness of each sheet, with dashed lines representing borderline samples (samples which can belong to two categories); b) Associated TD waveforms.

The etalon effect is the main reason why we observe the very broad bands below 35 cm\(^{-1}\) (≃1 THz) in absorbance spectra from paper sheets displayed in Figure
6.1. Terahertz time-domain spectroscopy of supports

6.1.b. Yet, the degree of overlap between the pulse reflected on the air-front surface of the sheet interface and the pulse(s) arising from multiple internal reflections within the sheet also depends on the time width of the incident pulse: it follows that the number, spectral width and position of the bands observed in the FD and caused by the etalon effect depend on the performances of the set-up used for spectroscopy. It can therefore be argued that the narrow bands observed below 1 THz for different types of paper by Naftaly et al. [129], and defined by the authors as "characteristic absorption peaks", are caused by the etalon effect. These bands are therefore not characteristic of a specific vibrational motion in paper, but are characteristic of the optical path of the paper sheet, i.e. its thickness as well as its frequency-dependent optical parameters. This would explain why a correlation is found between the weight per area of the paper sheets and the number and frequency position of the spectral bands below 1 THz: the weight per area of paper would increase with paper thickness, leading to a different overlap between the main pulse and the subsequent pulses arising from multiple internal reflections. Unfortunately, the authors do not provide data about the thickness of investigated paper sheets (nor the TD waveforms associated with the absorbance spectra), so the link between the etalon effect and the narrow spectral bands below 1 THz cannot be confirmed. Still, Trafela et al. [128] observe a broad band below 1 THz for a single sheet of Whatman filter paper (Grade 1, 180 µm), and one to two similar narrow bands as observed by Naftaly et al. when analysing an increased number of overlaid filter paper sheets: this strengthens our argument. Finally, Mechelen et al. [132] used what could be considered as the most accurate approach to extract the frequency-dependent optical properties from paper sheets, by fitting the TD waveform reflected from a sheet of paper on a copper reflector with a function accounting for multiple internal reflections. This function also uses the Lorentz model to describe the frequency-dependence of the complex refractive index of the paper sheet, and the Drude model for a copper reflector. Using this procedure, neither the real nor the imaginary part of the complex refractive index (associated with what is commonly referred to as refractive index and absorption coefficient, respectively) show any spectral feature below 1 THz: this study confirms that any spectral feature observed below 1 THz for paper is highly unlikely to be associated with any characteristic vibrational behaviour of paper.
What is certain is that paper can display spectral features arising from the vibrational behaviour of cellulose at 73 cm\(^{-1}\) (2.2 THz) and 100 cm\(^{-1}\) (3 THz), as shown in Figure 6.1 \(b\) and also mentioned by Trafela et al [128]. This discussion leads to a reconsideration of the Partial Least Square (PLS) calibrations presented by Trafela et al. [128] for determination of the lignin content in historic paper sheets. The authors observed spectral bands below 1 THz in the transmittance spectra of lignin, which they attributed to the characteristic vibrational behaviour of this molecule. Yet, another study [250] concluded that there are no signature absorption peaks for lignin in the frequency range of 0.3 - 2.0 THz: the bands observed by Trafela et al. are therefore likely due to the etalon effect. In reality, the PLS calibrations presented by Trafela et al. reflect the cellulose content of paper sheets, rather than their lignin content, which is why the loading weight associated with this calibration shows that the spectral regions which predominantly contributed to the separation of paper samples are the spectral bands located at 2.2 and 3 THz, i.e. the spectral features characteristic of vibrational properties of cellulose. The reason why this statistical model was judged to be of reasonable quality to determine the lignin content of historic paper is because a higher lignin content could be associated with a lower cellulose content in paper. The authors attributed the difference of signature between the spectrum of lignin and the loading weights of PLS calibrations to the possible presence of intermolecular bonds forming between lignin and cellulose, and modifying the vibrational behaviour of lignin. In the light of the above, this hypothesis can be rejected.
6.1. Terahertz time-domain spectroscopy of supports

Absorption coefficients and refractive indices from paper and parchment sheets calculated by TPS Spectra software might be inaccurate, as spectral bands at low frequencies caused by the etalon effect may affect the phase unwrapping procedure (see Section 4.1.4.2 in Chapter 4). The model described by Mechelen et al. [132] to extract optical parameters from four different paper sheets enables us to estimate the thickness of sheets with a precision and accuracy comparable to values measured with a calliper, which means that the refractive index and absorption coefficient calculated with this model are highly accurate. Yet, such a procedure requires some prior knowledge on the nature of light-matter interactions within a sheet. These interactions can change from one sheet to another, as different sheets have different porosity, fibrous structure, coating and filler content. For example, Mechelen et al. chose to describe the vibrational behaviour of a paper sheet with only one Lorentzian oscillator with a resonant frequency above 20 THz, accounting for the fibrous structure. Yet, for copy paper, the authors observed that an additional oscillator below 10 THz would be needed to accurately describe the vibrational behaviour of this paper sheet and extract accurate frequency-dependent refractive index and absorption coefficient. To objectively compare the optical properties from 32 historic parchment sheets and 7 paper sheets, this model would therefore be difficult to implement.

The alternative we suggest is to use constant values calculated using waveforms after deconvolution, i.e. values independent from the overlap between the main reflected pulse and the pulses arising from multiple internal reflections. Considering that paper and parchment sheets show little dispersion, refractive indices extracted from their deconvolved waveforms could be considered a good estimate of the average refractive index over the spectral range 0.15 - 3 THz. The constant refractive index was therefore calculated using equation 4.20 in Chapter 4, with $\Delta t$ the time delay between the main pulses of the sample and reference deconvolved waveforms. The absorption coefficient was calculated using equation 4.26 in Chapter 4, with the constant refractive index instead of the frequency-dependent refractive index, and the ratio between the amplitude of the main pulses of the sample and reference deconvolved waveforms, instead of the frequency-dependent transmittance ratio.

$$n = 1 + \frac{2\Delta t}{d}$$

\hspace{1cm} (6.1)
\[ \alpha = \frac{2}{d} \left[ \ln \left( \frac{4n}{(n+1)^2} \right) - \ln \left( \frac{A_{\text{sample, deconvolved waveform}}}{A_{\text{reference, deconvolved waveform}}} \right) \right] \tag{6.2} \]

Figure 6.4 lists the constant refractive index and absorption coefficient values extracted from historic and contemporary paper (grey squares) and parchment (black squares) sheets, together with some published values of refractive index of paper for comparison\(^1\) (grey stars).

While some paper and parchment sheets have similar refractive indices (between 1.4 and 1.5), the trend is for parchment sheets to be more reflective and absorbent than paper sheets. Yet, it can be observed in Figure 6.4 that the parchment sheet used for the imaging study of ink dots (represented by three black triangles, for the three different areas of the sheet used for the application of inks) and the parchment book cover used for the case study in Chapter 5 (represented by a black circle) have a refractive index and an absorption coefficient larger than the average value for the set of investigated parchment sheets. As a result, the findings of the imaging study with ink dots presented in Chapter 5 should be reviewed: the contrast in terahertz images of ink dots applied on the selected parchment sheet was observed to be particularly poor because the refractive index of the sheet is particularly large. The difference in image contrast in reflection mode between inks on paper and inks on parchment is, in most cases, not as significant as the difference reported in Chapter 5, since the difference between the average refractive index values of paper and parchment is \(\simeq 0.15\), as opposed to 0.3 - 0.4 for the difference between the parchment and paper sheets used in the imaging study.

Parchment sheets investigated in this spectroscopic study have various states of degradation, as assessed by their shrinkage temperature [68] (ranging from 37.1 to 61.6 °C). Yet, their absorbance spectra reported in Figure 6.1 a are featureless. This implies that no intermolecular vibrations related to the ordered structure of fibres of collagen or resulting from conformational changes upon denaturation of collagen can be probed with THz-TDS in the range 0.15 - 3 THz. Yet it is possible that the swelling, glassiness and conformational changes in fibres of collagen resulting from

---

\(^1\)Constant refractive index values for paper reported by Naftaly \textit{et al.} [129] are the average of the refractive index values over the spectral range 0.1 - 1.6 THz, and values reported by Mousavi \textit{et al.} [131] were calculated using a fitting procedure similar to the one presented by Mechelen \textit{et al.} and were observed to be real and constant below 3 THz.
6.1. Terahertz time-domain spectroscopy of supports

their denaturation alter the optical path and overall absorption properties of the parchment sheet in the terahertz range under study. However, like the ash content in paper sheets [128, 132], calcium salts used in the process of some of the parchment sheets and detected with FTIR spectroscopy ($\approx 710, 875$ and $1410 \text{ cm}^{-1}$ for calcium carbonate and $\approx 670, 1110$ and $1620 \text{ cm}^{-1}$ for calcium sulfate [170]) could also alter the optical properties of the parchment sheets. Finally, the compactness of the fibrous structure of the sheet, originating from the provenance and process of the skin, could also have an influence.

![Figure 6.4](image)

**Figure 6.4:** a) Constant refractive index and b) absorption coefficient from contemporary and historic paper sheets (grey) and parchment sheets (black). Error bars indicate standard deviations of triplicate measurements. Additional refractive index values from paper reported in the literature [129, 131] are displayed as stars. Values displayed as triangles represent rag paper and parchment sheets used as supports for ink dots in the imaging study discussed in Chapter 5. The value displayed as a circle represents the value from the parchment book cover discussed in Chapter 5. Average values for paper and parchment are displayed as grey and black horizontal lines, respectively.

Appendix I reports the results of principal component analysis performed on the dataset of 32 parchment samples discussed in Figure 6.1 a: they reveal that the shrinkage temperature shows no or little correlation with the constant refractive index or absorption coefficient of the parchment sheet, compared to the influence of the presence of calcium salts on the surface of a parchment sheet or the presence of a dense network of fibres, both leading to a stronger reflection and absorption from the parchment. It can therefore be concluded that the state of degradation of a parchment sheet cannot be assessed using its optical constants in the terahertz range 0.15 - 3 THz. Appendix
I also demonstrates that the bands at low frequencies arising from the etalon effect do not reflect the state of degradation of parchment.

### 6.2 Terahertz time-domain spectroscopy of iron gall inks

Chapter 5 shows that the model iron gall ink IG1 has a characteristic spectral signature in the terahertz range 0.15 - 3 THz, leading to a ringing effect observed in the terahertz TD B-scan from an ink dot on a single sheet of support. This dispersive behaviour differentiates it from inert carbon-based inks. Yet, IG1 shows the same image contrast as bone black in terahertz TD C-scans from an ink dot on a single sheet of support. To determine if these optical properties are specific to the composition of this model ink, or if most iron gall inks would show the same optical behaviour, iron gall inks with different compositions were analysed with THz-TDS.

Terahertz absorbance spectra from the individual components of iron gall ink are presented in Figure 6.5a. Sharp spectral features can be observed for iron(II) sulfate heptahydrate at 37, 50 and 64 cm\(^{-1}\), and a broad band centred around 84 cm\(^{-1}\). For anhydrous copper(II) sulfate, sharp features are observed at 78, 83 and 95 cm\(^{-1}\), and for gallic acid at 35, 50, 69 and 85 cm\(^{-1}\). Arabic gum and tannic acid do not exhibit any distinct spectral feature. The spectrum of iron(III) sulfate is also shown in Figure 6.5a as oxidation of iron(II) into iron(III) may take place during preparation or drying of the ink in contact with ambient air [251, 252]; consequently, spectral features of iron(III) sulfate might be observed in the spectra of ink samples. Yet, the only spectral feature observable in all triplicate spectra from iron(III) sulfate is at 56 cm\(^{-1}\) and is not very pronounced: this band is likely due to the spectral signature of water remaining in a small concentration in the purged chamber during the acquisition of the spectra, since Ruggiero et al. [251] observe the absence of spectral signatures for iron(III) sulfate hydroxide dihydrate.

It should be noted that the spectral feature at \(\approx 37\) cm\(^{-1}\) for iron(II) sulfate is not systematically observed for all iron(II) sulfate supplies. Ruggiero et al. [251] explain that it is unlikely for this band to originate in the absorption of an optical phonon, as it was not reproducible in solid-state DFT simulations. Instead, it is believed to be caused
by a polariton absorption (see Chapter 2) from iron(II) sulfate heptahydrate. Iron(II) sulfate tetrahydrate, however, does not show such a band, either in the experimental or the simulated spectra, but still shows the bands at 51 and 64 cm\(^{-1}\). PXRD results from iron(II) sulfate shown in Appendix A and discussed in Chapter 3 revealed the presence of the two hydrates. Therefore, depending on the ratio of iron(II) sulfate tetra- and hepta-hydrate (which may evolve over time with oxidation into iron(III) sulfate), the band at \(\simeq 37\) cm\(^{-1}\) is either observed or absent.

![Figure 6.5: Terahertz absorbance spectra of individual constituents of iron gall ink: gallic acid, tannic acid, copper sulfate, iron(II) sulfate and Arabic gum, and a possible reaction product: iron(III) sulfate; b) average absorbance spectra of historically informed iron gall inks IG1, IG2, IG3 and IG5.](image)

Figure 6.5: a) Terahertz absorbance spectra of individual constituents of iron gall ink: gallic acid, tannic acid, copper sulfate, iron(II) sulfate and Arabic gum, and a possible reaction product: iron(III) sulfate; b) average absorbance spectra of historically informed iron gall inks IG1, IG2, IG3 and IG5.
The model iron gall ink $IG1$ clearly displays the spectral signature of iron(II) sulfate as shown in Figure 6.5. It therefore contains water-soluble iron(II) ions unbound to gallic acid. Yet, as discussed in Chapter 2, free, i.e. uncomplexed, iron(II) ions catalyse oxidative degradation of cellulose in paper. Using THz-TDS, it can therefore be concluded that $IG1$ is a corrosive ink. When decreasing the amount of iron(II) sulfate incorporated in the preparation of the ink, it can be observed in Figure 6.6 a that the spectrum of the ink becomes featureless. Yet it cannot be asserted that this ink is not oxidative. Although it has been shown in Chapter 4 that TPS Spectra 3000 could detect up to 7.2 mg of iron(II) sulfate in a pellet prepared with a 10% mass concentration of sample to HDPE, it has also been shown that the intensity of the absorption peaks and absorbance baseline characteristic of iron(II) sulfate decreases with decreas-
ing amount of iron(II) sulfate in the pellet. Therefore when the mass ratio of tannic acid to iron(II) sulfate introduced in the preparation is 4x ($IG0.25$) to 10x ($IG0.1$) more than in $IG1$, the strong rising baseline in the absorbance spectrum, caused by featureless tannic acid and Arabic gum, likely masks the weak absorption peaks of iron(II) sulfate, if present. Besides, the absence of spectral features in the absorbance spectrum of iron gall inks prepared with large amounts of tannic acid makes them spectrally similar to inert carbon-based inks. However, when using gallic acid in the preparation of ink, the characteristic absorption peaks of this acid are observed in the spectral signature of the ink from a molar ratio of iron(II) sulfate to gallic acid equal to 0.9 onwards, as illustrated in Figure 6.6 b. Therefore, although the oxidative nature of the ink is still not possible to assess for the same reason as for tannic acid, non-degraded iron gall inks prepared with gallic acid and iron(II) sulfate would always display a spectral signature, and could therefore be distinguished from inert carbon-based inks.

To summarize, if the spectral signature probed from an ink inscription is similar to the spectral signature of iron(II) sulfate, it can be concluded that this ink is a corrosive iron gall ink. The reverse is however not true. If the spectral signature probed from an ink inscription is similar to the signature of gallic acid, it can only be concluded that this ink is iron gall ink (corrosive or not). If the spectral signature probed from a black ink inscription is featureless, no information on the nature of the ink can be deduced, as it can either be an iron gall ink with an excess of tannic acid, or a carbon-based ink.

When using terahertz TD imaging, rather than spectroscopy, to withdraw information from the nature of the ink, no information on the corrosive nature of the ink can be deduced: if a clear ringing effect is observed in the B-scan of a black ink inscription, then it can only be concluded that the ink is iron gall ink, as gallic acid and iron(II) sulfate have similar spectral signatures and might lead to a ringing effect with a similar frequency at the tail of TD waveforms. However, if no ringing effect is observed, then the nature of the ink cannot be deduced.
Figure 6.7: Average constant refractive indices and absorption coefficients of historically informed iron gall inks prepared following different recipes, measured in transmission from 400-mg pellets with a mass ratio of pigment to polyethylene of 10%. Error bars indicate standard deviations of triplicate measurements.

While the different chemical compositions of iron gall inks presented above lead to different spectral signatures, it is interesting to note that they have similar refractive indices and absorption coefficients, regardless of the nature and amount of acid, nor the purity of the iron(II) sulfate supply (i.e. the amount of copper(II) sulfate introduced during the preparation) as shown in Figures 6.7 and 6.8. Both frequency-dependent and independent refractive indices and absorption coefficients are shown here as, in theory, there exist no frequency-independent $n$ and $\alpha$, unless the material shows no dispersion, which is a reasonable assumption for parchment and paper, based on their (quasi) featureless absorbance spectra and based on the results presented by Mechelen et al. [132]. This explains why the frequency-independent absorption coefficients of featureless iron gall inks ($IG1 / r_{\text{molar}} = 0.1$ and 0.25) are different to the constant absorption coefficients from the other inks in Figure 6.6, while all inks show baselines with similar absorption values in Figure 6.8. All these model iron gall inks could therefore be categorized as likely to give a good terahertz image contrast in reflection mode, based on the study performed in Chapter 5, since their constant refractive index is comprised between 1.8 and 2.15.
The optical properties of freshly prepared iron gall inks have so far been discussed, and might be different from the optical properties of historic iron gall inks, as these may have undergone chemical changes such as hydrolysis and oxidation. Determination of refractive indices and absorption coefficients from historic iron gall inks would be strenuous as:

- It would require a sample of ink from the surface of the support big enough to prepare a spectroscopic pellet thick enough to observe no etalon effect, and not too diluted to easily measure the difference in thickness between the reference and sample pellets, and obtain a reproducible refractive index and absorption coefficient. As discussed in Chapter 4, 76 mg would be required to make a pellet with a 5% mass ratio, a diameter of 13 mm, and a measurable difference in thickness with the reference pellet. Assuming a density of iron gall ink close to the density of iron(II) sulfate (≃2.8 g/cm³) and an ink thickness of ≃50 µm, 76 mg of ink would correspond to a sampling area of ≃5.5 cm x 5.5 cm. This is a very large quantity, practically impossible to extract from a historic document.

- Penetration of the ink into fibres of the support leads to a gradual change in refractive index from the ink to the support: there is no longer a sharp ink-support interface, which makes it difficult to accurately determine the refractive index of the ink layer when comparing the time of collection of the pulse reflected from
the air - ink interface and the subsequent pulse, in the deconvolved waveform acquired in reflection mode.

Thermally degrading a model ink in a ventilated oven at 70 °C prior to pressing it into a pellet would mostly lead to dehydration of the ink, as shown in Appendix J [253], which is not a likely natural degradation path.

While it is not possible to obtain refractive indices and absorption coefficient from historic inks, or model inks which have undergone an artificial degradation process mimicking a natural degradation path, it might still be possible to extract the spectral signature from historic iron gall inks, and assess if their corrosive nature could be determined with THz-TDS. For the reasons detailed in Chapter 3, the ATR acquisition mode was preferred over the transmission mode to measure the spectral signature from ink lines on historic documents. Yet, spectral information is collected from the entire area in contact with the ATR crystal, the diameter of which is ≃6 mm: the width of the investigated ink line being ≃0.5 - 2 mm, spectral information from the support is also collected and might mask the spectral information from ink. As a result, ink was also scraped off from the surface of some documents with a scalpel, and the powder (2 - 5 mg, i.e. ≃1.5 - 3.5 mm²) was spread over the entire ATR crystal, to increase the possibility to observe a spectral signature. Spectral signatures from several inks on historic documents from the reference collection of the ISH are listed in Figure 6.9: all inks showed a positive reaction to the iron(II) indicator strip, confirming the presence of free iron(II) ions and therefore the corrosive nature of the ink. Figure 6.9 shows that spectral signatures extracted from ink powders are not always reproducible: this is likely due to the quality of the contact between the ATR crystal and the ink powder. Appendix J further shows that the extraction of the spectral signature from ink lines or ink dots largely depends on the quantity of ink deposited on the surface, and its penetration into the fibres of the support. The ATR mode, although less invasive than cutting a sample from a document and placing it in the nitrogen-purged sample compartment for analysis in transmission mode, does not lead to a systematically detectable spectral ink signature. Yet, the spectral signature of iron(II) sulfate is clearly observed for the ink powder sampled from the document OR37: the use of THz-TDS in the ATR mode was therefore successful in assessing the corrosive nature of this ink.
Figure 6.9: Absorbance spectra from historic iron gall inks, acquired in ATR mode. Measurements were either performed directly on the ink inscriptions, or on 2 - 5 mg of ink powder sampled from the surface of the parchment document with a scalpel. The scale of the absorbance spectrum of the ink powder sampled from document OR37 has been enlarged in the bottom graph, to observe the weak and broad spectral features at 36 and 49 cm\(^{-1}\).
6.3 Conclusions

- Historic parchments have a featureless terahertz spectrum, while historic papers may display the weak spectral band of cellulose. Bands observed at low frequency are likely spectral artefacts caused by multiple internal reflections of the terahertz pulse within the sheet.

- Historic parchment is on average more reflective than historic paper, and would therefore tend to lead to a poorer image contrast of ink inscriptions in reflection mode, although the images of ink dots on parchment shown in Chapter 5 represent an extreme scenario because the refractive index of parchment is exceptionally high.

- All model iron gall inks presented in this chapter belong to the category that is likely to give a good terahertz image contrast in reflection mode, regardless of their composition. A preliminary conclusion could therefore be that most iron gall inks are likely to give a good terahertz image contrast in reflection mode, although a systematic imaging study of more historic iron gall inks would be needed to confirm this hypothesis beyond doubt. This would require access to a large number of historic documents written with iron gall inks.

- Differentiation between iron gall and carbon-based inks using terahertz TD spectroscopy is possible if the spectral features associated with iron(II) sulfate or gallic acid are observed. In particular, if the spectral signature of iron(II) sulfate is observed (peaks at 38 - not always observed, 51 and 64 cm$^{-1}$), it can be concluded that such an ink is a corrosive iron gall ink. However, this differentiation is not systematic, as the absence of spectral features in the spectrum extracted from a black ink can be associated with the amorphicity of carbon-based inks, as well as the preparation of the iron gall ink or the low SNR of the technology used to extract this spectral information.
Chapter 7

Conclusions

7.1 Contribution to new knowledge

The project dealt with metrology of terahertz time-domain spectroscopy and imaging in relation to inks and pigments as used in the context of historic libraries and archives. Specifically, I have contributed to the body of knowledge by:

- Providing a database of precise frequency-dependent refractive indices and absorption coefficients from pigments and supports historically used in archival documents, and explaining how to use this database to better understand and interpret contrast in terahertz images of historical documents. In particular, inks have been separated into three categories with respect to image contrast in reflection mode. Lamp black, carbon black and graphite are very likely to appear contrasted in terahertz images. Vermilion, minium, iron gall ink, bone black, ivory black, lapis-lazuli and malachite are likely to appear contrasted in terahertz images, though this depends on the optical properties of the actual support. Bistre, verdigris, sepia and Prussian blue are unlikely to appear contrasted in terahertz images. For documents written with an ink belonging to this latter category, the use of terahertz TD imaging in reflection mode as an alternative technology for digitisation is not recommended as the textual content of the document would not be legible in terahertz images.

- Showing that iron gall inks can contain spectral signatures of iron(II) sulfate or gallic acid in the terahertz region 0.15 - 3 THz [253]. Iron gall inks can be corrosive and degrade the writing support (paper or parchment). Since free iron(II)
ions catalyze this degradation, our results imply that it is possible to detect corrosive iron gall inks using terahertz time-domain spectroscopy.

- Proving that, when analysing a sample diluted with HDPE and pressed into a pellet, moderately grinding the sample powder in fine HDPE powder with a pestle and mortar before pressing it improves the precision and accuracy of the optical properties extracted from this pellet [209]. In particular, the standard deviation of the refractive index of Prussian blue has been observed to reduce from 0.14 to 0.02 following this procedure. This will significantly contribute to terahertz metrology.

- Showing that deconvolution of the terahertz TD signal influences the contrast in images based on a peak amplitude. Specifically, sieve lines, ruling lines and indentations left by writing tools (nib pen or typewriter) at the surface of paper or parchment can be observed in images produced from peak amplitude, if no deconvolution of the signal is used. While deconvolution of the signal has become routine when analysing layered structures, since it facilitates the assignment of different layers and since it is now automated in commercially available software tools, it is important to remain aware that for documents, some archival information might be best observed using the raw signal.

- Showing that it is possible to reveal the textual content on both recto and verso sides of a folded letter within a closed envelope with terahertz time-domain imaging.

- Showing that the spectral signature of iron gall inks prepared in excess of iron(II) sulfate progressively decreases in amplitude when exposed to a temperature of 70 °C in dry conditions, and linking this spectral change to the dehydration of iron(II) sulfate heptahydrate into iron(II) sulfate monohydrate [253]. This information is of use to heritage scientists as thermal degradation of iron gall inks is widely used to identify and understand chemical degradation paths in these inks.

- Demonstrating the influence of multiple internal reflections of the pulse transmitted through paper and parchment sheets on their frequency-dependent refractive index and absorption coefficient. Specifically, it is argued that any spectral
7.2 Impact: current and future

The results of this project were disseminated to heritage scientists through poster and oral presentations at international conferences [254–260]. Through publications in peer-reviewed journals [209, 253], terahertz spectral signatures from carbon-based pigments, iron gall inks and their components were disseminated to analytical scientists and best practice on the preparation of pellets to extract accurate and precise refractive index and absorption coefficients reached the readership of terahertz scientists. The published spectral signatures of iron(II) sulfate, copper sulfate and gallic acid attracted the attention of computational chemists from the University of Syracuse, USA. While the research interests of this group are focused on Density Functional Theory to understand the vibrational behaviours leading to spectral signatures of compounds in the terahertz range, they found that this knowledge could be directly applied to the understanding of spectral signatures from historical iron gall inks, thanks to the collaboration with the ISH. This collaboration further led to publications in peer-reviewed journals with a readership of chemists and analytical scientists [191, 251]. Feedback on and discussions of the project took place during seminars with staff members of the Dutch National Archives and the National Library of the Netherlands in October 2012 and August 2015. Finally, the aim and research approach of this project were shared with the general public through a presentation at The Bartlett, UCL Faculty of the Built Environment, which was video-recorded and disseminated online on YouTube [261].

The knowledge and research approach discussed in this thesis serve as a guide to provide the best terahertz technological answer to curatorial research questions. What follows is a list of questions raised by conservators and researchers during this project, that I will attempt to answer based on the current state-of-the-art:
1. I want to access textual information beneath layers of ink and/or support in a document, what can I do?

Terahertz TD imaging is a cutting-edge, expensive and often time-consuming technique requiring a certain level of expertise for data interpretation, particularly since much research is still being undertaken to understand the interaction between terahertz radiation and matter and on metrology. As a result, rather than just scanning a document with this technology and spending resources to obtain a legible image from inscriptions within a stack of layers, the first step should be to ensure that the inscriptions which need to be studied are likely going to lead to a contrasted image with the available terahertz set-up and imaging configuration. It should be established whether there are other imaging configurations, terahertz set-ups or imaging technologies which might be better adapted to reveal such inscriptions. For example, at BAM Federal Institute for Materials Research and Testing, Germany, multispectral imaging in combination with laser cleaning was used to attempt to recover blackened (i.e. censored) text passages on a postcard sent from the ghetto Piaski in the German-occupied zone of Poland during WWII [262]. Since this research was unsuccessful, terahertz TD technology could potentially help to recover parts of the original text. The research approach which can be suggested, based on the results of this doctoral project, is the following:

• The first step would be to try and recover as much information about the nature of the inks used for the text and for the censorship, whether there is any historical information available (any literature about what material or process was used for censorship at that time) or analytical information (e.g. results of a spectroscopic study).

• The second step would be to compare the frequency-dependent refractive index and absorption coefficient from the inks and from the support. If the optical properties of the inks, pigments or dyes used in the postcard are not reported in the database provided in our study, or are not published elsewhere, then there are two options: i) if there is a significant amount of metal in one ink but not in the other, an X-ray based imaging technique
would first need to be considered, as it would lead to a significant contrast and better image resolution than terahertz-based imaging techniques, ii) if not, a THz-TDS study of the component(s) from the inks would need to be performed to determine $n(\nu)$ and $\alpha(\nu)$. If optical properties from both inks used for the censorship and for the text OR optical properties from the ink used for the text and the support are too similar, terahertz TD imaging is unlikely going to give a contrasted image of the censored text.

- The third step would be to determine the terahertz imaging set-up and configuration which would optimize the image contrast. If a large difference is observed between the baseline of $n(\nu)$ from the inks and the support, then a terahertz TD imaging set-up in reflection mode should be used, provided deconvolution can be performed on the reflected signal (in order to separate the pulses reflected from the air - censorship ink interface from the pulse reflected from the censorship ink - text ink or text ink - paper interface). If a large difference is observed between the baselines of $\alpha(\nu)$ from the inks and the support, then either a terahertz TD or a CW imaging set-up in transmission mode should be used, without the need for deconvolution of the signal if using a terahertz TD, since the signal would be transmitted through all layers, and the contrast of the image produced with the first and main transmitted pulse would only depend on the difference in absorption between the text ink and the censorship ink. A CW imaging set-up at a high frequency would be more appropriate, as the difference in absorption coefficients is larger and the spatial resolution is improved at high frequency. Finally, if the baselines of $n(\nu)$ and $\alpha(\nu)$ are similar for all materials in the stack, but the dispersive behaviours are different (e.g. one material has a clear spectral signature), then it would be worth using a CW imaging set-up tuned to the frequency where the optical properties from the different materials show the highest difference.

- The last step would be to scan a small area of the document and progressively reduce the scanning step size until the inscriptions are just legible, and since a reduction of the step size does not improve the image contrast.
This step optimizes the scanning acquisition time and the file size.

2. **Can this technology be used for mass digitisation of an entire archival collection, or can it only be used to digitize a few items in a collection?**

   Provided that all items in the collection are similar (e.g. only postcards written with the same ink, or only folded letters in an envelope written with the same ink on the same kind of support), once the suitable imaging approach has been determined for one item in the collection, it can be similarly applied to all other items. However, long acquisition times, file size and cost of the technology would deter from using terahertz imaging for mass scanning.

3. **Could we read a whole book?**

   Younus Ayesha states in her doctoral thesis [103] that a terahertz pulse can propagate through 200 pages of regular printing paper (80 µm thick, 80 g/m², $n_{mean\ over\ 0.1–2THz} = 1.54 \pm 0.05$) before its amplitude reaches the limit of detection of their terahertz TD set-up in transmission mode. In reflection mode, since the signal travels twice through the material before it reaches the detector, the limit of detection would be reached for 100 pages. Yet it was discussed in Chapter 4 that a poor image contrast (or signal to noise ratio) is obtained for a target imaged already under two sheets of rag paper, particularly if using a high frequency to produce the terahertz image, so particularly at frequencies leading to a good image resolution. Inscriptions written with a material leading to a very good image contrast, such as graphite, carbon black or lamp black, could still be differentiated from the support at larger depths with a low signal to noise ratio, while inscriptions written with a material from the categories "likely" and "unlikely to give a good image contrast" as defined in Chapter 5 would not be legible if located beneath a few pages. Additionally, each page should be within the focal region of the terahertz beam to ensure that the signal from the page is optimal and that the inscriptions can be spatially resolved. This would require a lens which could ensure that 100 pages of 80 µm are simultaneously in focus (*i.e.* a lens with a focal depth of 8 mm or more), that inscriptions less than 1-mm wide are (laterally) resolved and that each page of 80 µm is (axially) resolved. The latest developments in the design of terahertz lenses enable us to reach a
focal depth larger than 120 mm with a terahertz TD imaging set-up and a spatial resolution similar to a commercial terahertz TD imaging set-up [263]. An alternative solution using commercially available lenses could also be to perform several subsequent terahertz TD scans of the same book area while re-adjusting the focus to the next set of pages between scans. Finally, if we consider a book constituted of 100 printing paper pages with a text printed with carbon black ink, provided shadowing effects or occlusions could be automatically removed using algorithms such as the one under development by Redo-Sanchez [229], air gaps between pages with a thickness lower than the axial resolution of the terahertz TD set-up would limit the possibility to reveal text from all pages within the book. While a layer of ink between two paper pages that are stuck together (i.e. with no air gap) could still be detected due to the difference in refractive index between the layer of ink and the adjacent layers of paper, if the layer of ink is too close to another layer of ink (i.e. two regions with text facing each other with a very small air gap in between), then there is no change in optical properties perceived by the terahertz pulse and the two layers cannot be differentiated. The presence of an air gap between two layers of ink, thicker than the axial resolution of the imaging set-up, ensures that the text at the surface of each page can be (axially) resolved. A terahertz TD set-up developed by Takayanagi et al. in 2009 used a 17 fs wide pulse (almost twenty times narrower than the terahertz pulse of the TPS Spectra 3000) and was able to reach an axial resolution of 2 µm [140], revealing very fine depth information in closely spaced low-absorbing Teflon sheets. While such technology seems able to resolve small air gaps within a closed book, two issues arise from the use of very narrow terahertz pulses:

- Small oscillations (similar to the previously described “ringing effect”) superimpose on the main pulse [140]. This is likely due to the long life-time resonance of the emitting antenna, coherently re-emitting terahertz pulses at a frequency corresponding to its resonance frequency. It follows that small peaks resulting from reflection at the air-paper interface from a layer deep within the stack of pages would be difficult to differentiate from the oscillations due to the ”ringing effect” arising from the resonating antenna.
• The energy is distributed throughout a broader spectral range for a sharp pulse, compared to a short pulse (Takayanagi [140] shows a reference spectrum extending beyond 20 THz, with a maximum amplitude at \( \approx 2 \) THz, compared to the TPS Spectra 3000 extending to 3 THz and peaking below 0.6 THz - 20 cm\(^{-1}\)). However, this broad spectral distribution of the energy also means that for the same emitting power, the maximum amplitude in the frequency-domain reachable with a short pulse is lower than the one reachable with a broader pulse. It follows that for absorbing materials like paper, with a rising absorption baseline, high frequency components are lost faster for a sharp pulse than for a broad pulse when travelling through multiple sheets. In other words, pulse-broadening increases and therefore the axial (and lateral) resolution becomes poorer for a small stack of papers when using a sharp pulse than when using a broad pulse. So while the study by Takayanagi shows promising results for low absorbing materials like Teflon, the same performances are unlikely to be obtained for absorbing materials like paper, unless a higher emitting power is used. A threshold for this emitting power would then need to be defined so that the desired axial and lateral resolutions are reached, without heating and damaging the document.

While the latest technological developments pave the way for the use of terahertz technologies in scanning of a whole book, sophisticated signal and image processing, as well as inverse problem solving, would be needed to overcome the complexity of the signal due to the use of sharp pulses. If there is prior knowledge about the document, such as an estimate of the number and thickness range of the investigated pages, simulating the signal that has propagated through the document would help in interpreting the complex experimental signal. However, when it comes to historical documents, such prior knowledge rarely exists.

4. Can this research inspire the design of bespoke technology for use by conservators in their practice?

A hand-held probe to detect the presence of iron(II) sulfate within ink inscriptions at the surface of a document could be designed. A conservator could
7.2. Impact: current and future

rapidly and non-invasively assess the presence of iron(II) sulfate in excess in the ink by searching for peaks at 38, 50 and 64 cm\(^{-1}\) in the spectrum. A similar outcome than for the bathophenanthroline test is expected: if these peaks are observed, the ink can be considered as corrosive. However, if these peaks are not observed, nothing can be concluded on the nature of the ink. Yet, it is unlikely that such technology could ever replace the use of the iron(II) indicator strips for the bathophenanthroline test, due to their very low price (\(\approx £20\) for a pack of 100).

5. **Is this methodology ready to be routinely used in museum and conservation laboratories?**

This methodology is designed to help scientists ensure that they are using the most adapted technology, image acquisition, signal parameter and image processing technique to extract the archival information of interest to conservators. A heritage scientist who identifies a document for which he or she believes that terahertz technology can be of use would be able to define the specifications that the terahertz set-up should meet to increase the likelihood of extraction of the archival information of interest, using the results and methodology described in this thesis. This scientist would then need to contact the terahertz community to inquire which company or institution is equipped with a terahertz set-up meeting these specifications. If the analysis performed externally is conclusive and if the information from several other documents in the collection with similar structure and composition has yet to be uncovered, the set-up can be acquired by the conservation laboratory and routinely used for the analysis of these documents. Terahertz science is still under development and although software interfaces of terahertz equipments can be very user-friendly, terahertz time-domain data sets remain quite complex to interpret. Besides, terahertz technology is still expensive. Yet with the increasing demand, more affordable, compact and lighter instruments are to come. I therefore believe that in the next twenty years, terahertz technology would meet the budget and requirements of museum and conservation laboratories, with a broad database of instructions on how to best acquire and interpret terahertz data, and would therefore become an instrument routinely
used by heritage scientists in museum and conservation laboratories, without the need to collaborate with the terahertz community.

Finally, the knowledge and research approach developed in this thesis could be transferred to inspection of post. In particular, if performing a rapid TD spectroscopic point-analysis of a letter to detect the presence of drugs or hazardous chemical substances concealed in an envelope, this extra layer needs to have different optical properties than the layers of ink and support from the letter. Moreover, to safely identify the nature of the compound on the basis of its spectral signature, it should be ensured that no ink resonates at similar frequencies. Mail screening for security purposes is a target market for companies designing terahertz instruments such as TeraView Ltd [264].
Appendices
Appendix A

Complementary characterization and imaging

A.1 Fourier-transform infrared spectroscopy

Spectroscopic data in the mid-infrared region was acquired to give complementary information to terahertz spectroscopy on the chemical composition of pigment and ink powders, as well as to determine the nature of ink in some historical documents. Measurements were performed using the Alpha-P Bruker FTIR spectrometer with OPUS software, version 6.5. The ATR module equipped with a diamond crystal was preferred for analysis owing to the absence of sample preparation and small sample volumes required. The ATR mode further enabled to analyse independently the recto and verso of parchment sheets, due to the low penetration depth of the evanescent wave (≈2 µm at 1000 cm\(^{-1}\) in paper or parchment). It should be noted that while a good contact between the surface of the paper or parchment sheet and the crystal should be maintained during a measurement to improve the signal to noise ratio, the clamp has been observed to leave indentations on the surface of the sheet on some occasions. As the conditions for total internal reflection are not met for samples with a refractive index larger than the refractive index of diamond (≈2.4), some pigment powders were pressed in KBr pellets and analysed with the transmission module. Spectroscopic data was acquired within the wavenumber range 375 - 4000 cm\(^{-1}\), where characteristic absorption peaks from both organic and inorganic compounds can be found. 64 scans were performed at 4 cm\(^{-1}\) spectral resolution.
Figure A.1: FTIR spectrum of carbon black, graphite and lamp black powders acquired in transmission mode, together with the FTIR spectrum of a blank KBr pellet: 375 - 650 (broad), 714, 876, 1032 - 1270, 1385, 1431, 1630, 2854, 2923, 3100 - 3700 cm$^{-1}$ (broad).

Figure A.2: ATR-FTIR spectrum of bone black: 469, 560, 599, 642, 697, 873, 960, 1022, 1085, 1409, 1452, 1575 (broad), 2012 cm$^{-1}$.

Figure A.3: ATR-FTIR spectrum of ivory black: 473, 560, 599, 874, 960, 1019, 1085 (shoulder), 1416, 1453, 1545 (broad), 1601 (broad) cm$^{-1}$. 
A.1. Fourier-transform infrared spectroscopy 257

Figure A.4: FTIR spectrum of bistre in transmission: 427, 483, 505, 523, 572, 711, 747, 780, 820, 870, 1046, 1097, 1388, 1419, 1511, 1598, 2924, 3403 cm\(^{-1}\) (broad).

Figure A.5: FTIR spectrum of malachite in transmission: 426, 500, 520, 570, 707, 742, 777, 817, 864, 1040, 1096, 1379, 1489, 3302, 3397 cm\(^{-1}\).

Figure A.6: ATR-FTIR spectrum of lapis lazuli: 388, 429, 569, 643, 692, 960, 1100 cm\(^{-1}\) (shoulder).
Figure A.7: FTIR spectrum of vermilion in transmission.

Figure A.8: FTIR spectrum of minium in transmission: 387, 453, 489 (weak and broad), 511 (weak and broad), 532, 685 and 1410 cm\(^{-1}\) (broad).

Figure A.9: ATR-FTIR spectrum of verdigris: 507, 519, 626, 686, 803, 946, 1032, 1049, 1102, 1263, 1354, 1417, 1440, 1595, 1650, 3269, 3367 and 3463 cm\(^{-1}\).
A.1. Fourier-transform infrared spectroscopy

Figure A.10: ATR-FTIR spectra of pigments in the Prussian blue family, supplied by Cornelissen and Son, London (Prussian blue and Antwerp blue), Zecchi, Florence (Prussian blue) and Kremer Pigmente, Munich (Milori blue): 494, 605, 633, (982, 1066, 1171), 1414, 1458, 1611, 2080, (2855, 2925, 2955), 3256 and 3596 cm\(^{-1}\).

Figure A.11: ATR-FTIR spectrum of Arabic gum: 507, 519, 626, 686, 803, 946, 1032, 1049, 1102, 1263, 1354, 1417, 1440, 1595, 1650, 3269, 3367 and 3463 cm\(^{-1}\).

Figure A.12: ATR-FTIR spectrum of rabbit skin glue: 483, 868, 973, 1030, 1081, 1162, 1200, 1235, 1331, 1403, 1447, 1528, 1628, 1743, 2853, 2922, 3074 and 3275 cm\(^{-1}\).
A.2 Powder X-ray diffraction

PXRD was used to demonstrate the presence of polymorphs and their crystalline nature in compounds used in the preparation of iron gall inks, since polymorphs are likely to lead to different spectral signatures in the terahertz region. Measurements were performed by Martin Vickers and Jeremy Karl Cockcroft at the UCL Department of Chemistry using a Stoe Stadi-P PXRD system. The powder was held in a fine capillary thread, to reduce any orientation-led artefact. A CuKα1 source set at 40 kV and 30 mA was used, sending an incident beam at 1.54060 Å. A 2θ scan was performed from 5° to 64.895° for copper sulfate and from 2° to 45.065° for gallic acid, tannic acid, and gum Arabic, with a step of 0.495° every 20 s. As iron strongly fluoresces when using a CuKα1 source, a MoKα1 set at 50 kV and 40 mA was used to analyse iron(II) sulfate, sending an incident beam at 0.70930 Å. A 2θ scan was performed from 2° to 24.77°, with a step of 0.495° every 20 s. Diffractograms were compared to those available in the Powder Diffraction File database developed by The International Centre for Diffraction Data [265].

Figure A.13: Powder X-ray diffractogram of Arabic gum.
A.2. Powder X-ray diffraction

Figure A.14: Powder X-ray diffractogram of gallic acid.
Figure A.15: Powder X-ray diffractogram of iron(II) sulfate, analyzed in a disc sample holder (green line), and in a capillary thread (black line).
Figure A.16: Powder X-ray diffractogram of copper sulfate.
A.3 X-ray fluorescence (XRF) spectroscopy

PXRD had revealed the presence of crystalline hydrated calcium sulfate (gypsum) in a model iron gall ink, while no calcium was deliberately introduced in the preparation of the ink. It was therefore suspected that calcium was present in Arabic gum, in an amorphous form not revealed by PXRD. XRF spectroscopy was used as a rapid characterization method to qualitatively assess the presence of calcium in the gum.

Measurements were performed by Jacqueline Moon (UCL Institute for Sustainable Heritage) using the handheld Delta Dynamix air path X-ray Fluorescence spectrometer from Innov-X, Canada, available at the UCL Institute of Archaeology. The XRF was held in a clamp with the detector facing up, and the powder Arabic gum deposited on top. This allowed the X-rays to pass through the sample into the air and avoided interference from a supporting material. The handheld XRF was used in 'Soil mode'. This is a multi beam analysis method improving the lowest limits of detection for all analysed elements, including heavy metals, transition metals and light elements. 'Powershot' was used to analyse the full elemental range. The analysis time was 30 s; 10 s for each beam.
A.4 Scanning electron microscopy with energy dispersive X-ray spectroscopy

To determine the amount of metallic compounds such as calcium and phosphorus in carbon-based pigments and assess the likelihood to observe carbon-based inks on parchment when imaged with X-ray imaging systems, SEM-EDS was used, with the help of Dafydd Griffiths from the UCL Institute of Archaeology. Measurements were performed on pure bone black and ivory black powders, as well as carbon-based ink dots on parchment. Each sample was mounted on an aluminium stub and was first examined with a S-3400N Hitachi scanning electron microscope at an operating voltage of 20 kV, using secondary electrons emitted by atoms near the surface to produce a well-contrasted image of the sample, mostly rendering surface variations. To avoid accumulation of electrostatic charges at the surface of the low-conducting parchment, air was introduced in the chamber at a pressure of 70 Pa to conduct away the charges. A user-defined area of the sample was then analysed with EDS, using a INCA X-sight X-ray detector, Oxford Instruments, to identify the elements present in the ink dot or parchment surface, or in the pigment powder. The resulting spectrum is the average of all spectra acquired for each pixel within the user-defined area, and the spectrum energy range was set between 0 - 20 keV.

<table>
<thead>
<tr>
<th>Element</th>
<th>Ivory black</th>
<th>Bone black</th>
</tr>
</thead>
<tbody>
<tr>
<td>C (K)</td>
<td>(17.03, 17.58, 18.64) = 17.75 ± 0.82</td>
<td>(25.23, 20.66) = 22.95 ± 3.23</td>
</tr>
<tr>
<td>O (K)</td>
<td>(56.78, 55.46, 55.83) = 56.02 ± 0.68</td>
<td>(50.97, 50.84) = 50.91 ± 0.09</td>
</tr>
<tr>
<td>Mg (K)</td>
<td>(1.91, 1.87, 1.91) = 1.90 ± 0.02</td>
<td>(0.58, 0.51) = 0.55 ± 0.05</td>
</tr>
<tr>
<td>P (K)</td>
<td>(9.93, 9.82, 9.63) = 9.79 ± 0.15</td>
<td>(7.76, 8.48) = 8.12 ± 0.51</td>
</tr>
<tr>
<td>Ca (K)</td>
<td>(14.34, 15.27, 13.99) = 14.53 ± 0.66</td>
<td>(13.08, 18.34) = 15.71 ± 3.72</td>
</tr>
<tr>
<td>Ratio Ca/C</td>
<td>(0.84, 0.87, 0.75) = 0.82 ± 0.06</td>
<td>(0.52, 0.89) = 0.68 ± 0.26</td>
</tr>
<tr>
<td>Ratio P/C</td>
<td>(0.58, 0.56, 0.52) = 0.55 ± 0.03</td>
<td>(0.31, 0.41) = 0.35 ± 0.07</td>
</tr>
</tbody>
</table>

Table A.1: Atomic percentage ratios of elements present in ivory black and bone black pigment powders (Cornelissen and Son, London) obtained by energy dispersive X-ray spectroscopy.
A.5 Scanning electron microscopy

Due to its excellent spatial resolution, SEM was used to observe particle sizes \( (a \text{ priori} < 80 \, \mu m) \) and pigment distribution within a powder mixture prepared from Prussian blue pigment and HDPE powders ground with a pestle and mortar. The powder mixture was dispersed onto a conductive carbon adhesive tape and imaged with a Hitachi TM 3030 scanning electron microscope. Images were obtained in 'COMPO' mode using back scattered electrons, as this mode renders compositional differences: Prussian blue containing iron ions is heavier than HDPE and would appear as bright on the images, while HDPE would appear dark. Images were obtained with a 800x magnification and a 5 kV voltage.

A.6 3D microscopy

3D microscope imaging was further used to observe particle sizes and pigment distribution within the powder mixture prepared from Prussian blue pigment and HDPE powders moderately ground with a pestle and mortar. Due to the spherical shape of the particles, 3D microscope was preferred over regular optical microscopy to obtain a focused image of the full particle or cluster of particles. The powder mixture was dispersed onto a conductive carbon adhesive tape was imaged with a 3D VHX-5000 microscope (Keyence) using a 1000x magnification and the 'fine depth composition' mode allowing to reconstruct a 3D image from successive 2D images acquired at focal planes spaced by 0.5 \( \mu m \).

The surface of Prussian blue pellets prepared for terahertz spectroscopy was also imaged with the 3D microscope at a 200x magnification to observe pigment distribution within the HDPE matrix, as the porosity and surface irregularities of the pellet would prevent the surface from being investigated with regular optical microscopy.

Finally, the surface of some ink dots on paper or parchment was also investigated with 3D microscope imaging, using a 100x magnification and the 'fine depth composition' mode with a height step of 0.5 \( \mu m \). This enabled investigation of the surface height of ink dots, informing on the thickness of the ink layer and/or the surface irregularities of the support.
A.7 Optical coherence tomography

Optical coherence tomography of black ink dots on paper and parchment was performed in order to evaluate the advantages and drawbacks of this depth imaging technique, compared to terahertz time-domain imaging. Images were performed at the Photonics Centre, University of Kent, with Adrian Bradu, using a swept-source system operating at a central wavelength of 1050 nm and a sweeping rate of 100 kHz for fast data acquisition. A diagram of the set-up and further details about the instrument can be found elsewhere [266].

A.8 Profilometry

To observe indentations left at the surface of the support by a clean nib pen, the portable TRACEiT profilometer, Artec, was used. This system analyses an area of 5 x 5 mm, with a lateral and axial resolution of 2.5 μm.

A.9 X-ray imaging

X-ray absorption and phase-contrast imaging were performed on carbon-based ink dots on parchment, to assess whether well-contrasted images of inks with no or low metal content and low density on organic supports with similar properties could be obtained with X-ray based techniques. This was to objectively inform conservators of the relevance of the use of terahertz-based technologies compared to other existing technologies for depth imaging of archives.

An X-ray micro-CT imaging system from X-TeK Systems Ltd., UK (now Nikon) operating at 40 kVp and ≃100 mA.s⁻¹, and equipped with an intensified CCD camera, was first used in transmission mode by Jennifer Griffiths at the UCL Institute of Biomedical Engineering to obtain a planar image of the ink dots on parchment, based on their X-ray absorption properties: no ink dot could be discerned from the parchment support. It should be noted that 40 kVp was the minimum kV peak reachable with this system (a low kVp leading to a good image contrast).

A Molybdenum mammographic x-ray source with a peak energy between 17 and 18 keV (nearly mono energetic) was further used, as this system is more sensitive to small differences in X-ray absorption. It was combined with an Edge Illumination X-ray Phase Contrast Imaging system developed at the UCL Institute for Biomedical
Engineering to produce an image of the dots based on x-ray refraction and interference effects rather than absorption. A Rigaku M007 source, using a molybdenum target and a focal spot of 70 µm, was used at 40 kVp and 25 mA. To generate optimal image contrast and reduce radiation dose, a 30 µm-thick molybdenum filter was placed in front of the X-ray source. The detector is the ANRAD SMAM amorphous selenium flat panel, with a pixel size of 85 µm. Edge-illumination used masks made of gold on graphite before the sample and the detector [267]. The source-to-sample and sample-to-detector distances were 1.6 m and 0.4 m, respectively.
Appendix B

Phase unwraping algorithm

Bold text indicates sections added from the original algorithm provided by TeraView Ltd.

% CALCULATION OF THE FREQUENCY-DEPENDENT REFRACTIVE INDEX & ABSORPTION COEFFICIENT USING A CORRECT PHASE UNWRAPPING
% APPROACH AS DESCRIBED BY Withayachumnankul (2014) - NO APODISATION
% INPUT ARGUMENTS: tr — TIME-DOMAIN COORDINATES OF REFERENCE WAVEFORM (IN UNITS OF mm’s)
% ts — TIME-DOMAIN COORDINATES OF SAMPLE WAVEFORM (IN UNITS OF mm’s)
% r — TIME-DOMAIN REFERENCE WAVEFORM
% s — TIME-DOMAIN SAMPLE WAVEFORM
% d_um — THICKNESS IN MICROMETERS

function [ref_ind, abs_coeff]=RIcorrectphaseunwrap(tr,ts,r,s,d_um)

%...CALCULATE COMPLEX-VALUED FREQUENCY-DOMAIN SPECTRA
nr = length(tr);
Nr = 2 * nr;
R=fft(r,Nr);
ns = length(ts);
Ns = 2 * ns;
S=fft(s,Ns);

%...CALCULATE FREQUENCY-DOMAIN COORDINATES

tmin = min(tr);
tmax = max(tr);
trng = tmax - tmin;
c_ms = 2.99792458e8;
dt_mm = trng/[nr-1];
dt_m = dt_mm/1000;
dt_s = dt_m/c_ms;

vmax_Hz = 1/[2*dt_s];
dv_Hz = vmax_Hz/[Nr-1];
v_Hz = dv_Hz *[0:Nr-1]’;
v_THz = v_Hz / 1e12;%...FREQUENCY IN THz
v_cm = v_Hz / [100*c_ms];%...CONVERT Hz TO cm-1

dv_cm = dv_Hz / [100*c_ms];%...CONVERT Hz TO cm-1

%...CALCULATE UNWRAPPED TRANSMITTANCE PHASE

Nr = length(R);
Ns = length(S);
if Ns < Nr

N = Nr/2;%...USE ONLY FIRST HALF OF SPECTRUM (AS SECOND HALF OF ARRAY RETURNED FROM fft CONTAINS NEGATIVE FREQUENCIES)
else

N = Ns/2;
end

T = S ./ R;%...COMPLEX-VALUED TRANSMITTANCE

PHI = angle(T(1:N));%...TRANSMITTANCE PHASE (OVER FIRST N POINTS ONLY)

%...UNWRAP TRANSMITTANCE PHASE

for i = 33 : N
dPHI = PHI(i-1) - PHI(i); %...PHASE DIFFERENCE BETWEEN NEIGHBOURING FREQUENCIES
irem = ((i):N)'; %...INDICES OF ALL SUBSEQUENT (REMAINING) FREQUENCIES
if dPHI <= pi
    PHI(irem) = PHI(irem) + 2*pi;
elseif dPHI >= -pi
    PHI(irem) = PHI(irem) - 2*pi;
end
end

%...GENERATE ARRAY OF FREQUENCIES (WAVENUMBERS) IN UNITS OF cm⁻¹
v = dv_cm * (0:N)';

%...IF REFERENCE AND SAMPLE WAVEFORMS ARE NOT ALIGNED ON TIME-DOMAIN AXIS, CALCULATE AND ADJUST FOR PHASE OFFSET
trMin = min(tr);
tsMin = min(ts);
if trMin ~= tsMin
    dtMin_mm = tsMin - trMin;
dtMin_cm = dtMin_mm / 10;
    alpha = 2*pi * dtMin_cm * 2; %...MULTIPLY DIFFERENCE IN REFERENCE AND SAMPLE POSITIONS BY 2 BECAUSE LIGHT PATH IS TWICE THE SCAN DELAY LENGTH
    if alpha ~= 0
        for i = 33 : N
            PHI(i) = PHI(i) - alpha * v(i);
        end
    end
end
end
% LINEAR FIT OF PHI BETWEEN 14.1 - 21.6 CM-1

P = polyfit(v(33:50),PHI(33:50),1);
for i=1:32
    PHI(i)=P(1)*v(i)+P(2);
end
PHI=PHI-P(2); % FORCING INTERCEPTION AT ZERO

d_mm = d_um / 1000;
d_cm = d_mm / 10; % SAMPLE THICKNESS IN SAME UNITS AS WAVELENGTH (cm)

% CALCULATION OF THE REFRACTIVE INDEX & ABSORPTION COEFFICIENT FROM THE EXTRAPOLATED PHASE
ref_ind = zeros(N,1);
abs_coeff = zeros(N,1);
for i=1:N
    ref_ind(i) = 1 - PHI(i) / (2*pi * v(i)*d_cm);
    abs_coeff(i) = 2/(d_um*10^(-4))*(log(abs(4.*ref_ind(i)/(ref_ind(i)+1)^2*R(i)/S(i))));
end
Appendix C

Deconvolution algorithm

%DOUBLE GAUSSIAN FILTER FUNCTION TO BE USED FOR DECONVOLUTION
% t: time (of the reference waveform)
% r: reference waveform (baseline corrected)
% s: sample waveform (baseline corrected)
% HF: High Frequency filter
% LF: Low Frequency filter

function [h]=deconvolution(t,s,r,HF,LF)
f=DoubleGaussianFilter(t,r,HF,LF);
n=length(t);
S=fft(s,n);
R=fft(r,n);
F=fft(f,n);
quotient=S./R;
mult=quotient.*F;
h=ifft(mult,n);
end
Appendix D

Algorithm to extract peak amplitude from a layer of interest

% Load the imaging file saved as a .tvl file with the TPI Imaging Suite
% software, TeraView Ltd., and converted into a .coating file readable with
% Matlab

    % Load '.coating' file
load('-mat',file)

    TPIdata = TPIcs;
clear TPIcs
thzData = TPIdata.myData;
clear TPIdata

    % Load coordinates x and y
xyz = thzData.XYZ’;
x = xyz(:,1);
y = xyz(:,2);

    %...LOAD TIME-DOMAIN WAVEFORMS
sRaw = thzData.rawWaveform;

    I=length(sRaw(:,1)); %Lines of the matrix = time position
J=length(sRaw(1,:)); %Columns of the matrix = pixel position

n=4; %insert here the number of main peaks observed on the waveform
P=zeros(J,n); %initialization of matrix listing the amplitude of the n main peaks for each pixel
T=zeros(J,n); %initialization of matrix listing the time position of the n main peaks for each pixel

i=1; % initialization of the time position
j=1; % initialization of the pixel position

locsn=zeros(n);
LOCSn=zeros(n);
z=zeros(J,n);

for j=1:J
[pks,locs]=findpeaks(sRaw(:,j), 'SORTSTR','descend'); %find all the peak amplitudes
and positions of each waveform and sort them from larger to lower amplitude
P(j,:)=pks(1:n); %Select the nth main peaks
locsn=locs(1:n)';
LOCSn=sort(locsn,'ascend'); %find the time position of the nth main peaks and sort
them by ascending order

T(j,:)=LOCSn(:);
z(j,1:n)=sRaw(T(j,1:n),j)'; %extract the amplitude of the nth main reflected peak on the
time axis
end;

N=1; %Select the position on the time axis of the peak to be virtually flattened (1
= first main reflected peak, 2 = second main reflected peak...)
zN=z(:,N);
% Display image of the Nth peak
tri = delaunay(x,y); % Delaunay triangulation
trisurf(tri,x,y,zN);
shading interp
Appendix E

Algorithm for image normalisation

%LOADING THE .dat OR THE .THzImage FILES
fname='Path';
M=load('-ASCII',fname);

%DEFINING THE IMAGE SIZE AND THE X AND Y COORDINATES
m=size(M,1);
n=size(M,2);
x=1:m;
y=1:n;

%DEFINING REGION OF INTEREST = INK DOT
%USER-DEFINED LIMITS OF INK DOT, EXCLUDING EDGES
xmin=16;
xmax=22;
ymin=23;
ymax=28;

rROI_x=(xmax-xmin)/2;
rROI_y=(ymax-ymin)/2;

%CENTRE OF INK DOT
xc_ROI=xmin+rROI_x;
yc_ROI=ymin+rROI_y;
% RADIUS OF INK DOT (CHOOSE THE SMALLEST)
rROI = min(rROI_x, rROI_y);

% PIXELS INSIDE THE INK DOT
for i = 1:m
for j = 1:n
rad(i,j) = sqrt((i-xc_ROI).^2 + (j-yc_ROI).^2); % RADIUS FROM CENTRE OF INK DOT
end;
end;
i_center_ink_dot = find(rad <= rROI); % INDICES OF PIXELS WITHIN INK DOT
M_center_ink_dot = M(i_center_ink_dot); % PIXEL VALUES CORRESPONDING TO THE INK DOT

%iROI_x = [find(xc_ROI-rROI & x1;xc_ROI+rROI)];
%iROI_y = [find(yc_ROI-rROI & y1;yc_ROI+rROI)];
%M_center_ink_dot = M(iROI_x, iROI_y);

% DEFINING SUB REGION = SUPPORT

% USER-DEFINED LIMITS OF SUPPORT, EXCLUDING THE INK DOT EDGES
x_SUB_min = 13;
x_SUB_max = 24;
y_SUB_min = 20;
y_SUB_max = 31;

rSUB_x = (x_SUB_max-x_SUB_min)/2;
rSUB_y = (y_SUB_max-y_SUB_min)/2;

% CENTRE OF SUPPORT
\[ xc_{SUB} = x_{SUB\_min} + r_{SUB\_x}; \]
\[ yc_{SUB} = y_{SUB\_min} + r_{SUB\_y}; \]

% RADIUS OF SUPPORT (CHOOSE LARGEST)
\[ r_{SUB} = \max(r_{SUB\_x}, r_{SUB\_y}); \]

% USER-DEFINED REDUCED WINDOW OF 3.6 * 3.6 MM², REGARDLESS OF STEP SIZE
\[ x_{reduced\_min} = 12; \]
\[ x_{reduced\_max} = 25; \]
\[ y_{reduced\_min} = 19; \]
\[ y_{reduced\_max} = 32; \]
\[ M_{reduced} = M(x_{reduced\_min};x_{reduced\_max}, y_{reduced\_min};y_{reduced\_max}); \]

% PIXELS OUTSIDE SPECIFIC RADIUS FROM CENTRE OF SUPPORT
\[ \text{rad}_{SUB} = \text{zeros}(m,n); \]
for \( i = x_{reduced\_min}; x_{reduced\_max} \)
for \( j = y_{reduced\_min}; y_{reduced\_max} \)
\[ \text{rad}_{SUB}(i,j) = \sqrt{(i-xc_{SUB})^2 + (j-yc_{SUB})^2}; \]
end;
end;
i_{SUB} = \text{find}(\text{rad}_{SUB} \leq r_{SUB} \& \text{rad}_{SUB} \geq 0); \]
\[ M_{support} = M(i_{SUB}); \]

% STANDARDIZATION OF THE IMAGE IN THE REDUCED WINDOW:
\[ s_{SUB\_avg} = \text{mean}(M_{support}(:)); \]
\[ s_{ROI\_avg} = \text{mean}(M_{center\_ink\_dot}(:)); \]

%INITIALISATION
M_stand=zeros(x_reduced_max-x_reduced_min+1,y_reduced_max-y_reduced_min+1);

i=1;
j=1;
while i<x_reduced_max-x_reduced_min+1
while j<y_reduced_max-y_reduced_min+1
M_stand(i,j)=M_reduced(i,j)/sSUBavg;
j=j+1;
end
j=1;
i=i+1;
end

imagesc(M_stand); figure(gcf)
axis off
colormap('gray')
caxis([1 1.5])
colorbar
axis image

%CALCULATION OF THE CONTRAST PERCENTAGE VALUE, FROM THE AVERAGE SIGNAL REFLECTED FROM
%THE INK DOT AND THE AVERAGE SIGNAL REFLECTED FROM THE SUPPORT
Contrast=sROIavg/sSUBavg;
Appendix F

Raw terahertz images from ink dots

Figure F.1: Photographs of ink dots on rag paper and associated terahertz images acquired in reflection mode using RIM attached to TPS Spectra 3000, with step size between 75 and 150 µm. Terahertz images were produced using maximum amplitude of the reflected waveforms. All images were normalized and are displayed on the same grey scale.
Appendix F. Raw terahertz images from ink dots

Figure F.2: Photographs of ink dots on additionally sized rag paper and associated terahertz images acquired in reflection mode using RIM attached to TPS Spectra 3000, with step size between 75 and 150 μm. Terahertz images were produced using maximum amplitude of the reflected waveforms. All images were normalized and are displayed on the same grey scale.
Figure F.3: Photographs of ink dots on parchment and associated terahertz images acquired in reflection mode using RIM attached to TPS Spectra 3000, with step size between 75 and 150 μm. Terahertz images were produced using maximum amplitude of the reflected waveforms. All images were normalized and are displayed on the same grey scale.

Figure F.4: Photographs of three different bistre and vermilion ink dots applied on rag paper, additionally sized rag paper and parchment, and associated terahertz images acquired in reflection mode using RIM attached to TPS Spectra 3000, with step size between 75 and 150 μm. Terahertz images were produced using maximum amplitude of the reflected waveforms. All images were normalized and are displayed on the same grey scale.
Appendix G

OCT and 3D images of model ink dots

Figure G.1: OCT images of a) bistre on rag paper, b) bistre on parchment, c) iron gall ink (IG1) on rag paper and d) iron gall ink (IG1) on parchment; 3D images of e) vermilion on rag paper and f) vermilion on parchment. Experimental details are available in Appendix A.
Appendix H

Near-infrared spectra from model and historical ink inscriptions

![Near-infrared spectra](image)

**Figure H.1:** a - c) NIR spectra of iron gall ink inscriptions on three historic documents (sample ID: a) Eva36, b) Eva47 and c) Eva37), all responding positively to the bathophenanthroline test. d) NIR spectra of bistre, bone black, lamp black and sepia mixed with Arabic gum and applied on rag paper. The black curve is the average NIR spectrum from the support, with error bars representing the standard deviation from triplicate measurements.
Thirty two parchment sheets of approximately 2 cm x 2 cm were sampled from documents dating from the mid-17th to the early 20th century, with known thickness and shrinkage temperature [68], and were analyzed with terahertz time-domain spectroscopy using TPS Spectra 3000 in transmission mode. Each parchment sample was analyzed in triplicate while changing the orientation of the sample between measurements to account for scattering.

Additionally, both flesh and grain sides of the 32 parchment samples were analyzed using attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy, using the Alpha-P Bruker at 4 cm\(^{-1}\) resolution. The presence of calcium compounds on either sides of the sheets was confirmed by the presence of spectral features at \(\approx 710, 875\) and \(1410\) cm\(^{-1}\) for calcium carbonate and \(\approx 670, 1110\) and \(1620\) cm\(^{-1}\) for calcium sulfate dihydrate [170]. These salts were introduced during the preparation of the skin to produce a parchment sheet to be used as a writing support, as discussed in Chapter 2. The introduction of calcium salts was used to remove the excess of grease and moisture.

The surface on both flesh and grain sides was also observed with an optical microscope in reflection mode (SP-400, Brunel Microscope), and it either showed a clear network of parchment fibres, or a very dense and uniform structure in which barely any fibre could be distinguished even at the highest magnitude (400x).

Principal Component Analysis (PCA) was performed using Origin Pro 8.6 software on a dataset containing, for each of the 32 samples (i.e. observations), a) the
Appendix I. Principal Component Analysis of historical parchment optical constants

constant refractive index or $b$) the constant absorption coefficient, the shrinkage temperature of fibres sampled from its surface ($ST$), the absence (0) or presence (1) of calcium salts on either of its surfaces ($Ca$) and the absence (0) or presence (1) of a clear and loose fibre network on either of its surfaces ($FS$, for Fibrous Structure). Figures I.1 and I.2 illustrate the set of variables and observations used for PCA. The correlation matrix of the set of data was calculated and used to extract principal components. The results are presented in Figure I.3 as biplots, i.e. 2D scatter plots with the first two principal components as $x$ and $y$ axis and the 32 samples as points, with superimposed loading vectors for each variable ($n$ or $\alpha$, $Ca$, $ST$, $FS$) indicating the direction of variance separation. The percentage of variance associated with the first two principal components cumulatively is $a)$ 70.50% using $n$ and $b)$ 67.65% using $\alpha$.

Figure I.1: 

- **a)** Photograph of the document OR81 from the reference collection of the ISH.
- **b)** FTIR spectra acquired from both sides of the parchment sheet, normalized at 1540 cm$^{-1}$, corresponding to the amide II band of the peptide linkage in the collagen protein [174].
- **c)** and **d)** Micro-photographs from both sides of the parchment sheet, obtained using an optical microscope in reflection mode, with a magnification 400x.
- **e)** Values of the variables used to perform PCA, for the observation OR81.
The state of degradation, i.e. the shrinkage temperature, does not seem to contribute to the refractive index nor to the absorption properties of parchment, as the angle between $ST$ and $n$ (or $\alpha$) vectors is close to 90°. However, it appears that the presence of calcium salts is positively correlated with $n$ (or $\alpha$), i.e. the angle between $Ca$ and $n$ vectors is lower than 90°. It can be concluded that the presence of calcium salts on the surface of a parchment sheet leads to a stronger reflection and absorption of parchment in the terahertz region below 100 cm$^{-1}$. The fibrous structure ($FS$) is negatively correlated with $n$ (or $\alpha$), i.e. the angle between $FS$ and $n$ vectors is greater than 90°: this negative correlation implies that the lower value (0), associated with a very dense and homogeneous structure, corresponds to a larger refractive index and absorption coefficient of the parchment sheet.

**Figure I.2:** a) Photograph of the document OR94 from the reference collection of the ISH. b) FTIR spectra acquired from both sides of the parchment sheet, normalized at 1540 cm$^{-1}$, corresponding to the amide II band of the peptide linkage in the collagen protein [174]. c) and d) Micro-photographs from both sides of the parchment sheet, obtained using an optical microscope in reflection mode, with a magnification 400x. e) Values of the variables used to perform PCA, for the observation OR94.
Figure I.3: Biplots displaying the results of the PCA on the set of parchment properties and showing the contribution of the shrinkage temperature (ST) of the parchment fibres, the presence of calcium salts (Ca) and the presence of a clear fibre network (FS) on either side of the parchment sheet, to a) the constant refractive index, and b) the constant absorption coefficient.

Therefore a parchment sheet with a dense fibrous structure on both flesh and grain sides tends to absorb and reflect more terahertz radiation than a parchment sheet with a clear and loose network of fibres on either side, which allows for more terahertz radiation to be transmitted through.

**Conclusion 1**

There is no or little correlation between the shrinkage temperature of fibres sampled from a parchment sheet and the constant refractive index or absorption coefficient from this same sheet. This therefore implies that the state of degradation of a parchment sheet cannot be assessed using its optical constants in the terahertz range.

However, given the correlation between the shrinkage temperature and the fibrous structure (i.e. acute angle between FS and ST vectors), it is possible that the fibrous structure parameter accounts for all of the following: the state of degradation of parchment, the animal species used for the skin and the process of skin manufacture. Besides, while the above study was based on optical constants, the state of degradation of parchment may affect the dispersive behaviour of a parchment sheet (i.e. the frequency-dependence of the refractive index and absorption coefficient). A degraded parchment with swollen fibres collapsing in disordered arrangement would have a different fibrous structure than a freshly prepared parchment sheet, and might therefore lead to scatter-
ing, absorption and speed of propagation with a different frequency-dependence. If this is the case, for a parchment sheet with a given thickness, the pulse transmitted through a degraded area would not only have a different amplitude and time-delay than the pulse transmitted through a non-degraded area, but it would also have a different FWHM. As a result, the overlap between the main pulse and the pulse originating from multiple internal reflections would be modified. In the FD domain, this modification of the etalon effect would lead to a different number, position and/or shape of broad bands at low frequencies. Five samples of historic parchments were analysed using THz-TDS before and after their thermal degradation for 2 h in a ventilated oven at 150 °C. During this treatment, all samples had undergone obvious degradation, as shown in Figure I.4: their colour had changed and they had become stiffer and had curved.

**Figure I.4:** Photographs from five historic parchment samples (≈2 cm x 2 cm) before (top) and after (bottom) thermal degradation for 2 h in a ventilated oven at 150 °C.

Figure I.5 shows absorbance spectra and the time-domain waveforms from these parchment samples before and after thermal degradation, and Table I.1 lists the associated optical constants. No significant change is observed in the broad spectral bands at low frequencies, which means that the dispersive behaviour of the sheet has not been altered by thermal degradation. However, the slope of the absorbance spectrum is systematically reduced after thermal degradation. This change of slope is due to the decrease in amplitude of the main transmitted pulse. Dehydration of the sheets upon
Appendix I. Principal Component Analysis of historical parchment optical constants
degradation in the oven is a phenomenon that is highly likely going to participate in this
decrease in pulse amplitude. Dehydration would also explain why weak water vapour
bands are still observed around $56 \text{ cm}^{-1}$ for the sample before degradation, while they
are not for the same sample after degradation. Thermal degradation also leads to a
decrease in optical constants, as observed in Table I.1.

**Conclusion 2**
The dispersive behaviour of a parchment sheet is not altered by the state of degradation
of this sheet. As a result, the bands at low frequencies arising from the etalon effect do
not reflect the state of degradation of parchment. On the other hand, the hydration state
of a parchment sheet affects its optical constants. As a result, THz-TDS cannot be used
as a tool to monitor the state of degradation of a parchment sheet, as changes in the
optical constants of parchment are influenced by the environment it has been exposed
to (relative humidity and temperature).

<table>
<thead>
<tr>
<th>Sample name in Figure I.5</th>
<th>Original</th>
<th>Thermally degraded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>a</td>
<td>1.65 ± 0.04</td>
<td>40.85 ± 1.80</td>
</tr>
<tr>
<td>b</td>
<td>1.58 ± 0.02</td>
<td>46.21 ± 1.23</td>
</tr>
<tr>
<td>c</td>
<td>1.54 ± 0.02</td>
<td>44.29 ± 1.99</td>
</tr>
<tr>
<td>d</td>
<td>1.51 ± 0.06</td>
<td>38.53 ± 3.33</td>
</tr>
<tr>
<td>e</td>
<td>1.68 ± 0.00</td>
<td>49.51 ± 0.85</td>
</tr>
</tbody>
</table>

**Table I.1:** Optical constants from five historic parchment samples ($\approx 2 \times 2 \text{ cm}$) before and
after thermal degradation for 2 h in a ventilated oven at 150 °C. Constants were
calculated using Equations 6.1 and 6.2 in Chapter 6.
Figure I.5: Terahertz absorbance spectra and associated time-domain waveforms from five historic parchment samples (≈2 cm x 2 cm) before (black curve) and after (grey curve) thermal degradation for 2 h in a ventilated oven at 150 °C.
Appendix J

Terahertz spectral analysis of model inks in ATR mode

To investigate the spectral signature from iron gall ink inscriptions on historical documents in the least invasive way, ATR mode was preferred over the transmission mode. The sensitivity of such configuration was tested using model iron gall ink dots. 2 and 6 µL of IG1, IG2 and IG3 model inks were applied on historical rag paper, contemporary filter paper and historical parchment sheets, and their spectral signatures were acquired with the ATR configuration.

Figure J.1 shows that, on parchment, the IG3 ink dot (ink with gallic acid) has more intense spectral features than IG1 and IG2 ink dots (prepared with tannic acid). It should be noted that the spectral signature of IG3 is reproducible as it was observed both on 06/07/12 and 29/11/12 (see Figure J.3). As the inks had been applied on parchment 4 months before acquisition of the spectra shown in Figure J.1, it is possible that the ink with gallic acid (IG3c) has degraded / dehydrated at a different rate than the inks prepared with tannic acid (IG1c and IG2c), leading to spectral features with different intensities.

Figure J.2 shows that the spectral signature of the IG3 ink is more pronounced when IG3 is applied on parchment than when applied on rag paper or filter paper: in the latter case, the spectral signature cannot be distinguished. This may be due to larger penetration of the ink in paper fibres than in the denser fibre network of parchment: as the evanescent terahertz wave escaping from the crystal only penetrates a dozen of microns into the sample, if the ink has low penetration into parchment, the signal collected from the evanescent wave is mostly the one of the ink layer, not the parchment
Figure J.3 shows that with a larger amount of ink applied on parchment, the more distinct the spectral signature of the ink dot becomes. This phenomenon, however, does not apply to ink on rag paper: the spectra are not reproducible and none exhibit any distinct spectral features.

In conclusion, while the ATR mode is less invasive than cutting a sample from a document and placing it in the nitrogen-purged sample compartment for analysis in transmission mode, it does not systematically detect spectral signatures of an ink on the surface of a document. This might be due to the difficulty to ensure a good contact between the ATR crystal and the surface of the ink, and/or to the lower SNR of the ATR mode compared to the transmission mode (58 dB at 0.3 THz for the ATR mode against 70 dB at 0.3 THz for the transmission mode [268]).

Figure J.1: Triplicate ATR-THz spectra from three different ink dots made of 6 µL of iron gall ink models applied on parchment.
Figure J.2: Triplicate ATR-THz spectra from an ink dot made of 2 µL of IG3 ink applied on rag paper, filter paper and parchment.

Figure J.3: Top: triplicate ATR-THz spectra from ink dots made of 2µL (IG3a) and 6 µL (IG3c) of IG3 ink applied on parchment; bottom: ATR-THz spectra from an ink dot made of 6 µL of IG3 ink applied on rag paper.
Systematic study of terahertz time-domain spectra of historically informed black inks

Tiphaine Bardon, a Robert K. May, b Philip F. Taday b and Mattija Strlic a,
a
The potential of terahertz-time domain spectroscopy (THz-TDS) as a diagnostic tool for studies of inks in historical documents is investigated in this paper. Transmission mode THz-TDS was performed on historically informed model writing and drawing inks. Carbon black, bistre and sepia inks show featureless spectra between 5 and 75 cm⁻¹ (0.15–2.25 THz); however, their analysis still provided useful information on the interaction of terahertz radiation with amorphous materials. On the other hand, THz-TDS can be used to distinguish different iron gall inks with respect to the amount of iron(II) sulfate contained, as sharp spectral features are observed for inks containing different ratios of iron(II) sulfate to tannic or gallic acid. Additionally, copper sulfate was found to modify the structure of iron(II) precipitate. Furthermore, Principal Component Analysis (PCA) applied to THz-TDS spectra, highlights changes in iron gall inks during thermal degradation, during which a decrease in the sharp spectral bands associated with iron(II) sulfate is observed. ATR-FTIR spectroscopy combined with THz-TDS of dynamically heated ink samples indicate that this phenomenon is due to dehydration of iron(II) sulfate heptahydrate. While this research demonstrates the potential of THz-TDS to improve monitoring of the chemical state of historical documents, the outcomes go beyond the heritage field, as it also helps to develop the theoretical knowledge on interactions between terahertz radiation and matter, particularly in studies of long-range symmetry (polymorphism) in complex molecular structures and the role played by the surrounding matrix, and also indicates the potential of THz-TDS for the optimization of contrast in terahertz imaging.

Introduction

The twofold purpose of archives and libraries is to guarantee that their collections of historical records are preserved and that the information remains accessible and available to the public. Research leading to a better understanding of degradation of archival collections and to improvement of the readability of documents that may be difficult to open (such as scrolls) or handle (degraded manuscripts) is therefore of major interest to curators and conservators. For this purpose, X-ray spectroscopic techniques are increasingly used, while X-ray imaging techniques offer depth-resolved images of documents in a non-invasive way, thus preserving the integrity of the document. Still, the set-ups are substantial in size and immovable, and to date, it has only been possible to analyze or reveal inscriptions containing inorganic components (e.g. in iron gall inks, inorganic pigments). However, throughout history many manuscripts have been written using carbon inks produced from the soot of organic materials. To complement the results obtainable with X-ray techniques and enable characterization and imaging of a broader range of archival objects, other non-invasive spectroscopic and imaging techniques with greater penetration depth are needed.

Terahertz time-domain spectroscopy (THz-TDS) and imaging has emerged in the last decade, with the development of ultra-short pulsed lasers. As terahertz radiation induces changes of amplitude in inter-molecular vibrations between neighbouring molecules, materials with different chemical composition or crystal structure have different absorption and dispersion responses. Therefore, terahertz spectroscopy can be used for the characterization of both inorganic and organic material and consequently, terahertz imaging can produce a contrasted image of a layered sample based upon differences in composition. This allows for non-invasive analysis of objects to depths of up to a few millimetres, depending on sample material and experimental conditions. This is particularly useful for the analysis of historical objects requiring minimal handling.

To examine the potential of THz-TDS as a diagnostic and imaging tool for studies of manuscripts, it is crucial to first investigate and understand the interaction of the terahertz pulse with individual materials used in the production of documents such as pigments, inks, parchment or paper. Recent work has made use of THz-TDS analysis of ink by itself to assist...
in the interpretation and improvement of terahertz images of model or historic manuscripts: Labaune et al.\textsuperscript{7} compared terahertz spectra acquired in reflection and transmission modes of carbon black and red ochre inks (frequently used for the production of papyrus documents) to select the model ink which provided the highest contrast when imaging inscriptions written on overlaid papyri. Similarly, Fukunaga et al.\textsuperscript{8} investigated terahertz spectra of a red ink inscription, a stain and the parchment base of a medieval manuscript to select the frequencies at which terahertz images of the document would show enhanced readability of the text. Additionally, terahertz spectra of a few black inks ("tannin ink", lamp black and ivory black) as well as various pigments have been made available in an online terahertz spectral library.\textsuperscript{9}

In this database, the measurements have largely been made using conventional Fourier transform spectroscopy in the range 0.6–13 THz which offers a lower signal-to-noise ratio below 3 THz than THz-TDS.\textsuperscript{10} Terahertz sources used for imaging at room temperature mainly operate below 4 THz.\textsuperscript{11,12} Any time-domain spectroscopic data of materials acquired below 3 THz would therefore complement the online terahertz spectral library and help improving and interpreting terahertz images of manuscripts. Finally, many documents written with iron gall inks are extensively damaged, with dark-coloured degradation products, or even loss of degraded support in and around the inscriptions, rendering the documents difficult to handle and read.\textsuperscript{13} However, no terahertz spectroscopic information on the chemical processes involved in the degradation of iron gall ink has, to date, ever been published.

This research presents the first systematic comparative study of the optical properties in the terahertz region 5–100 cm\textsuperscript{-1} of a representative range of black inks commonly found in historical documents, and also presents information about the chemical structure and the degradation processes of black inks. As this article represents the early stage of the work on terahertz technology applied to historical documents, it aims to provide essential background knowledge on the interaction of historical black inks with terahertz pulses based on the analysis of clear and accurate quantitative data. Therefore, THz-TDS in transmission mode of large amounts of inks models with known provenance and composition was used.

**Materials and methods**

**Samples**

**Ink models.** The selection of inks used in this study was based on their occurrence in archival documents, their predisposition to degrade the carrier, the availability of their components, and the feasibility of their production in a laboratory.\textsuperscript{7,13–15} Two main types of black inks were investigated: carbon black inks, where the black colour comes from the dispersion of soot particles from burnt organic materials in a mixture of water and binder; and iron gall inks, produced from the reaction of iron(III) with gallic acid moieties resulting in a dark blue complex upon oxidation, which degrades into brown-black over time. Bistre and sepia inks were also prepared since as dark brown inks widely used in manuscripts or drawings they are difficult to discern from degraded iron gall ink: bistre is brownish due to phenolic compounds from soot, whilst sepia is produced by cuttlefish and owes its colour to melanin.\textsuperscript{16}

The preparation of iron gall inks (IG) was mainly driven by the instructions based on nine historical recipes.\textsuperscript{17} Although originally made from crushed gall nuts, IG inks were prepared in this study using pure tannic or gallic acids (from Chinese gall nuts, Sigma Aldrich) to increase control over the experiments. Gall nuts contain almost 70% tannic acid,\textsuperscript{18} therefore the ratio of ~70% of the mass of nuts indicated in the recipe was used to determine the mass of pure acid needed for the four different recipes of iron gall ink examined: IG1, IG2, IG3 and IG5. Historic recipes suggest that crushed nuts can either be decanted or boiled in water,\textsuperscript{19} thus increasing the hydrolysis of tannic acid into gallic acid. Both pure acids were used for the preparation of inks in this research, tannic acid from Chinese gall nuts (T0200, Sigma Aldrich) and gallic acid (G7384, Sigma Aldrich). The purity and amount of iron sulfate contained in the ‘vitrail’ used by ink makers may have varied a lot, therefore, different ratios of iron and copper sulfates (p.a., Sigma Aldrich) were investigated. Similarly, the ratio of iron sulfate and acid in the ink also varied significantly, and additional inks with different mass or molar ratios than indicated in the reference recipe\textsuperscript{19} were prepared. Table 1 presents the mass of the different ingredients for each iron gall ink, including Arabic gum (Kremer, Munich), mixed with 50 ml of deionised water.

Soot for carbon black inks mainly came from burnt oil (lamp black pigment) and burnt bones or ivory (bone black or ivory black pigment). Lamp black (Cornelissen and Son, London), ivory black (Cornelissen and Son, London) and bistre (12100, Kremer, Munich) pigments, as well as sepia powder (acquired from a local fishmonger) were used for this study.

| Table 1 Compositions of iron gall inks, for preparation with 50 ml of deionised water\textsuperscript{a} |
|-----------------------------------------------|------------------|-----------------|-----------------|-----------------|
| Sample name | Tannic acid | Gallic acid | Iron(s) sulfate | Copper(s) sulfate | Arabic gum |
| IG1 | 1700 mg | 2500 mg | 25 mg | 1250 mg |
| IG2 | 1700 mg | 2500 mg | 250 mg | 1250 mg |
| IG3 | 1700 mg | 2500 mg | 25 mg | 1250 mg |
| IG5 | 1700 mg | 2500 mg | 1250 mg |
| IG1/\textsubscript{r=0.25} | 1700 mg | 425 mg | 4.25 mg | 1250 mg |
| IG1/\textsubscript{r=0.1} | 1700 mg | 170 mg | 1.70 mg | 1250 mg |
| IG2/\textsubscript{r=0.1} | 340 mg | 111 mg | 250 mg |
| IG2/\textsubscript{r=0.2} | 340 mg | 278 mg | 250 mg |
| IG3/\textsubscript{r=0.9} | 340 mg | 500 mg | 250 mg |
| IG3/\textsubscript{r=2} | 340 mg | 1112 mg | 250 mg |
| IG3/\textsubscript{r=5} | 340 mg | 2780 mg | 250 mg |

\textsuperscript{a} IG1 composition is modified for the preparation of IG2 by increasing the amount of copper(s) sulfate, for IG3 by using gallic instead of tannic acid; for IG5 by omitting copper(s) sulfate; for IG1/\textsubscript{r=0.25} and IG1/\textsubscript{r=0.1} by reducing the mass ratio of iron(s) sulfate to tannic acid in IG1. IG1/\textsubscript{r=0.1} = 1, IG1/\textsubscript{r=0.25} = 0.5, and IG3/\textsubscript{r=0.9} = 0.9 has the same molar ratio of iron(s) sulfate to acid as IG3.
Thermal degradation. To improve the understanding of the chemical processes involved in the degradation of iron gall ink, independently from the interaction of the ink with the support (paper or parchment), the different IG ink samples were dried, and then degraded in a ventilated oven (Carbolite, Eurotherm) at 70 °C for 1, 3, 8 and 22 h in dry conditions to accelerate degradation. It should be noted that this accelerated degradation was not meant to mimic natural degradation. Dried and degraded inks were then ground into a fine powder to make pellets for THz-TDS.

Terahertz time-domain spectroscopy

For terahertz spectroscopic measurements in transmission, dried iron gall inks and individual constituents of the iron gall and carbon black inks were ground into powder in an agate mortar, mixed with polyethylene powder and pressed into duplicate or triplicate pellets of 400 mg each with a 10% mass concentration of sample to polyethylene in the compressed matrix.

Spectroscopic data was acquired using a TPS spectra 3000 (TeraView Ltd., Cambridge UK) in transmission mode within the range 5–100 cm\(^{-1}\) (~0.15–3 THz). 500 scans were collected at 1.2 cm\(^{-1}\) spectral resolution. The sample compartment was purged with nitrogen to minimize terahertz absorption by atmospheric water vapour. This methodology was preferred over THz-TDS using the Attenuated Total Reflection (ATR) mode, as it leads to a signal more intense and easier to interpret, with reduced influence of the ambient environment.

Blackman-Harris 3-term apodization was used for the Fourier transformation of the measured transmitted pulse. Sample spectra were background-corrected against a reference spectrum from a 360 mg polyethylene pellet. Absorbance spectra, refractive indices and absorption coefficients were calculated (using the amplitude and phase of the terahertz pulse transmitted through sample pellets as well as measured pellet thickness) and displayed using the TPS spectra software (TeraView Ltd.).

The spectra from duplicate or triplicate measurements were averaged and the resulting spectrum for each sample was displayed in the range where the standard deviation was the lowest (usually 5–75 cm\(^{-1}\) or 5–100 cm\(^{-1}\)).

Multivariate data analysis

Principal Component Analysis (PCA) was performed using Origin Pro 8.6 software. An increase with frequency of the absorption baseline was observed for all raw THz spectra, and may be interpreted as Mie scattering;\(^{27}\) the spherical particles of ground samples could scatter the incident terahertz pulse if their diameters are similar to the wavelengths of the radiation (from 5–0.1 mm; i.e. 2–100 cm\(^{-1}\)). The scattering effect, and thus the absorption baseline, generally increases with shorter wavelengths, i.e. at higher frequencies. To ensure that only spectral features associated with the chemical structure of a sample contribute to the variance, rather than changes associated with particle size, the baseline was removed by calculating the second derivative of the absorbance spectra. The second derivative was preferred over the first derivative, as it removes the baseline while maintaining the intensity and positions of peaks in a spectrum.

The second derivative of the terahertz spectra was calculated using the Stavitzky–Golay function with a polynomial order of 3, over a window of 25 points.

PCA was performed on the second derivative of the spectroscopic data. Both correlation and covariance matrix of the set of data were calculated and used to extract the first 3 principal components, but only the scores for the covariance matrix were plotted as this allows for better separation of the data.

Results and discussion

Carbon black, bistre and sepia pigments

Fig. 1 shows that sepia, bistre, lamp black and ivory black pigments do not exhibit any distinct spectral features in the range below 2.25 THz (75 cm\(^{-1}\)), but rather a difference in overall absorbance from one pigment to another. Lamp black pigment with a mass ratio of 10% in a 400 mg pellet absorbs and/or scatters terahertz radiation so strongly that no transmitted pulse could be detected. At 1%, a transmitted pulse was detected, which also showed a featureless terahertz spectrum for lamp black pigment. For the same concentration of pigment in the pellet (10%), a decrease in absorbance in the terahertz region 5–75 cm\(^{-1}\) is observed from lamp black (no transmitted pulse detected at all) to ivory black, and then to sepia and bistre. The latter two have very similar absorbance. It should be noted that lamp black and ivory black come from the combustion of oil or tar and bone or ivory and consist of highly amorphous carbon, whereas bistre contains phenolic compounds\(^{28}\) and sepia is a mixture of organic macromolecules.\(^{29}\) Besides, lamp black is known to be nearly pure amorphous carbon while ivory black also contains calcium phosphate and calcium carbonate from the original bone or ivory.\(^{30}\)

A similar spectral comparison of black inks containing carbon has been performed by Vahur et al.\(^{31}\) in the far-infrared region 230–550 cm\(^{-1}\) and leads to a similar observation: charcoal and vine black, both pure amorphous carbon, show featureless spectra (or with a few faint bands which might be attributed to additives or impurities), while bone black and
Ivory black show the same spectral features thought to be due to the presence of calcium phosphate.

To summarize, Fig. 1 shows that totally or partially amorphous organic compounds do not exhibit distinct spectral features in the measured terahertz region and that larger carbon amorphousness within the material seems to induce a higher overall absorption of terahertz radiation. Additionally, the increase in absorbance with frequency observed in all organic materials, as other phenomena may occur, such as scattering; however, more research is required to confirm this.

The frequency dependence of the refractive indices and absorption coefficients of these amorphous samples is valuable in the interpretation of terahertz images. For example inks with higher refractive indices are expected to more strongly reflect the incident terahertz pulse, thereby providing high contrast with the surrounding support material when imaging in reflection mode. From the results shown in Fig. 1, ivory black ink should give a stronger reflected signal than that from bistre ink. This further indicates that the refractive index of ivory black pigments is very high but rapidly decreases below 13 cm\textsuperscript{-1}, and then reaches a quite steady value: this implies that there is an optimum frequency range (5–13 cm\textsuperscript{-1}) to obtain a good image of ivory black ink inscriptions in reflection imaging. Similarly, the frequency dependence of the absorption coefficient of sepia powder shown in Fig. 1 indicates that sepia inscriptions are expected to be best observed at higher frequencies in transmission mode.

\begin{align*}
\alpha(\omega) &= C(\omega)(\gamma(\omega) + A + Br^2)
\end{align*}

where \(C\) is a measure of the charge fluctuations caused by structural irregularities in the material \(C\) reflects the degree of “disorder” and \(\gamma\) is the Debye vibrational density of states, which can only be calculated independently by inelastic neutron scattering. Without the results of the neutron scattering experiment, it is unfortunately impossible to confirm the validity of this law for the amorphous samples of black pigments. Since the above mentioned theories were developed to compare crystallinity and amorphousness in glass, it may not be justified to apply them to powders derived from ground organic materials, as other phenomena may occur, such as scattering; however, more research is required to confirm this.

\textit{The frequency dependence of the refractive indices and absorption coefficients of amorphous materials put forward by Taraskin et al.}\textsuperscript{24} suggests a different power-law frequency dependency of the coefficient of absorption \(\alpha\):
In conclusion, valuable information about the optical properties of carbon black, bistre and sepia pigments can be extracted with THz-TDS, even if featureless absorbance spectra are observed in transmission. The values of optical coefficients can enable the interpretation of terahertz images of layered structures such as ink on support and aid in the differentiation of inks based on a comparison of the frequency dependence of image contrast from such samples. Yet, in historical documents, the optical coefficients of the ink binders and possible additives, together with the optical coefficients of the support (paper, parchment, papyrus) may also be frequency dependent and influence the contrast of the terahertz image. Further investigation of the optical properties of these materials would also aid in the interpretation and improvement of terahertz images of documents. Finally, terahertz absorption of powders of amorphous organic materials presented here may be of interest in other research fields (e.g. drug research).

**Historically informed model iron gall inks**

Solutions of iron gall inks IG1, IG2, IG3 and IG5 prepared with the same ratio of iron(II) sulfate to tannic or gallic acid as described in the literature\(^{13}\) have a similar pH of 2.0 ± 0.1. It could be assumed that all inks have a similarly acidic environment which would lead to similar chemical reactions during their preparation, and it is therefore valid to compare the chemical changes occurring during degradation. These inks exhibit a purple-black colour which may come from both the black ferro–gallic complex in solution and the formation of black precipitate dispersed in the solution.\(^{13}\)

Terahertz absorbance spectra of the individual components of iron gall ink are presented in Fig. 2a. Sharp spectral features can be observed for iron(II) sulfate heptahydrate at 37, 50 and 64 cm\(^{-1}\), for copper(II) sulfate anhydrous at 78, 83 and 95 cm\(^{-1}\), and for gallic acid at 35, 50, 69 and 85 cm\(^{-1}\). Arabic gum and tannic acid, do not exhibit any distinct spectral feature.

The spectrum of iron(III) sulfate is also shown in Fig. 2a as oxidation of iron(II) into iron(III) may take place during the preparation or the drying of the ink in contact with the ambient air; consequently, the spectral features of iron(II) sulfate might be observed in the spectra of the ink samples. Yet, the only spectral feature observable in all triplicate spectra from iron(III) sulfate is at ~56 cm\(^{-1}\) and is not very pronounced: this band could be due to the spectral signature of water remaining in a small concentration in the purged chamber during the acquisition of the spectra (THz pulses being very sensitive to water absorption\(^{25}\)).

It should be noted that the spectrum of gallic acid presented in Fig. 2a is in good agreement with the spectrum published by Wu et al.,\(^{26}\) while the spectrum of anhydrous copper sulfate only partially agrees with the results of Fu et al.,\(^{27}\) presented as a table listing THz absorption bands for different copper sulfate hydrates. No spectrum of anhydrous copper sulfate is presented in their study, which could allow for an objective comparison of the spectra.

Terahertz absorbance spectra of IG1, IG2, IG3 and IG5 are shown in Fig. 2b and all four ink compositions clearly exhibit three sharp spectral features at 38, 51 and 64 cm\(^{-1}\). An increase with frequency of the absorption baseline can also be observed and is attributed to Mie scattering. As sharp spectroscopic
features only arise from the coordinated vibrations of the same inter-molecular interactions in long-range symmetrical structures, this result shows that inks prepared with the same mass ratios of iron(n) sulfate to acid but with different copper sulfate contents or different acids share a similar long-range symmetry in their molecular structure. These spectral features coincide with those of pure, dried iron(n) sulfate heptahydrate (observed at 37, 50 and 64 cm\(^{-1}\), meaning that the same long-range molecular symmetry can be found in historically informed iron gall inks and pure iron(n) sulfate. As little is known about the vibrational behaviour of hydrated iron sulfates in the terahertz region, it is yet unclear whether these spectral features are due to the hydration state of iron(n) ions, a crystalline structure of iron(n) or sulfate ions, or to another ordered molecular structure.

Other pure constituents of iron gall inks such as gallic acid and copper(n) sulfate exhibit different distinct spectral features below 100 cm\(^{-1}\) but these features are not observed in the spectra of IG1, IG2, IG3 and IG5. Nevertheless, it seems that the relative intensity of the absorbance peaks at 38, 51 and 64 cm\(^{-1}\) varies from one ink composition to the other. Still, Fig. 2e indicates a larger standard deviation of the absorbance data at these frequencies for some inks. To ensure that the differences of relative peak magnitudes observed between the averaged spectra from IG1, IG2, IG3 and IG5 presented in Fig. 2a are due to a difference in ink composition rather than to the influence of one of the triplicate spectra on the averaged data, the second derivative of all triplicate spectra from all inks was further analyzed.

When comparing the area under the curve at frequencies 33–42 cm\(^{-1}\) for the second derivative of the triplicate spectra from IG1, IG2, IG3 and IG5, it was observed that triplicates from each separate ink have similar values and that IG2 triplicates show much larger values than all other inks (IG5 triplicates showing the lowest values). Similarly, when studying the area under the curve at frequencies 60–70 cm\(^{-1}\), triplicates from IG2 show much lower values than the other inks. IG2 has a copper sulfate content ten times larger than IG1 and IG3 (IG5 having no copper sulfate at all). Yet neither of the peaks at 38 cm\(^{-1}\) and 64 cm\(^{-1}\) was observed in the spectra of pure copper sulfate. This result suggests that copper sulfate modifies the actual structure of iron gall inks, which is a significant finding.

Additionally, the comparison of the area under the curve at 45–54 cm\(^{-1}\) for the second derivative of the triplicate spectra of all inks clearly distinguishes IG3 triplicates from the other inks as IG3 triplicates show much larger values. Gallic acid was used for the preparation of IG3, while all other inks have been prepared with tannic acid, and gallic acid also shows a characteristic spectral feature around 50 cm\(^{-1}\). Yet, none of the other spectral features of gallic acid are observed at 69 cm\(^{-1}\) or 85 cm\(^{-1}\) in the spectrum of IG3.

It is thus unclear whether the larger intensity of the band at 51 cm\(^{-1}\) for IG3 is due to an enhanced vibration of the molecular structure of iron(n) sulfate in presence of gallic acid, or to a synergistic vibration at the same frequency of gallic acid and iron(n) sulfate; this needs further research.

In conclusion, although inks with the same mass ratio of iron sulfate to acid contain spectral features similar to those associated with iron(n) sulfate, the findings indicate that copper sulfate plays a role in the vibrations of the molecular structure of the inks and that the presence of gallic acid leads to an increase band at 51 cm\(^{-1}\). These small spectral differences could potentially be used to distinguish between different ink compositions.

Refractive indices and absorption coefficients of IG1, IG2, IG3 and IG5 are presented in Fig. 2d and e, and show some differences: the refractive index decreases by ~0.2–0.4 over the range 3–100 cm\(^{-1}\), while the absorption coefficient increases with frequency. The sharp peaks at 38, 51 and 64 cm\(^{-1}\) observed in the absorbance spectra are still observed in the refractive index and absorption coefficient. It should be noted that the values of refractive index shown in Fig. 2d are averages of the values obtained for each triplicate pellet. The associated standard deviation can be as large as 0.4, without taking into account the errors that may arise from the measurement of the pellets thickness. Therefore, the values presented here can only inform on the order of magnitude and the frequency dependence of the refractive indices of iron gall inks. As a result of the large variance of the refractive index data, it cannot be asserted that IG3 is less reflective than the other ink samples, nor can it be asserted that inscriptions written with different iron gall inks can be distinguished with THz imaging.

Influence of iron(n) sulfate and gallic acid

The spectral features common to the absorbance spectra of IG1, IG2, IG3 and IG5 are thought to be due to the high concentration of iron(n) sulfate heptahydrate relative to the amount of acid. The iron(n) sulfate reacts in solution to form the coloured complex, however it is believed that a proportion of the iron(n) sulfate that does not react with the acid is responsible for the appearance of its spectral features in spectra from the IG ink, the intensity of which may be influenced by copper sulfate or gallic acid as discussed above. It is therefore expected that a lower amount of iron(n) sulfate heptahydrate added during the preparation would, once the stoichiometric proportion of acid and iron(n) sulfate was reached, lead to all the iron(n) sulfate reacting with the acid to form the ferro–gallic complex. Consequently, the iron(n) sulfate heptahydrate spectral features should no longer be observed for iron gall inks containing lower concentrations of iron(n) sulfate.

To confirm this hypothesis, iron gall inks with different ratios of iron(n) sulfate to acid were prepared and analyzed with THz–TDS. While gallic acid has a defined molecular structure, tannic acid consists of a glucose core linked to either gallic, di-gallic or tri-gallic acid moieties, which makes the calculation of its molar mass difficult. This is why mass ratios were calculated for inks containing tannic acid and molar ratios for inks with gallic acid.

Fig. 3 shows a decrease of the intensity of the peaks at 51 and 64 cm\(^{-1}\) with decreasing amounts of iron(n) sulfate in IG3, until the ink composition reaches the ratios of approximately one iron(n) sulfate entity for one gallic acid entity, i.e. for IG3/molar = 0.9, and one iron(n) sulfate entity for two gallic acid entities, i.e. for IG3/molar = 0.5.
At these ratios, the spectral signatures of both iron(II) sulfate and gallic acid are observed. These ratios indeed correspond to stoichiometric ratios of the iron gall complex most commonly found in the literature.28,29 Below these molar ratios, the spectral feature at 64 cm\(^{-1}\) is no longer observed and peaks around 35, 69 and 85 cm\(^{-1}\) appear, with increasing intensities due to decreasing iron to acid ratios. As the baseline slightly obscures these peaks, the second derivative of the absorbance spectra was further analyzed and confirms the presence, position and change in intensity of these peaks. Fig. 2a reveals that the peaks at approximately 35, 50, 69 and 85 cm\(^{-1}\) correspond exactly to the spectral features of gallic acid. However, the peaks observed around 69 cm\(^{-1}\) and 85 cm\(^{-1}\) in the spectrum of IG3/\(r\) molar \(\equiv 0.1\) are broader and less prominent than the peaks distinguished in pure gallic acid.

Similarly, Fig. 3 shows that the spectral features of iron(II) sulfate are no longer observed for IG1 inks with low mass ratios of iron(II) sulfate to acid (IG1/\(r\) mass \(\equiv 0.25\) and IG1/\(r\) mass \(\equiv 0.1\)), and that their spectra are featureless, similar to the spectrum of tannic acid.

Nevertheless, lower concentration of a component in the spectroscopic pellet would also lead to decreased absorbance intensity, as illustrated in Fig. 1 with lamp black showing an absorbance signal for a pellet prepared with 1% of pigment to polyethylene, while no absorbance signal could be detected for a pellet with 10% of the pigment. When preparing ink solutions with decreased iron sulfate to acid molar ratios, the proportion of iron sulfate to acid also decreases in the 40 mg of powder ink used to prepare the pellet, which leads to a decreased ratio of iron sulfate to polyethylene in the pellet. Thus, similar differences in the intensity of peaks with decreasing amounts of iron sulfate added in solution might as well occur if using pure and dry iron sulfate and gallic acid for the preparation of the pellet, with no reaction of complexation involved.

However, it is certain that a reaction of complexation takes place when adding iron sulfate and gallic or tannic acid in solution, since a rapid colour change of the solution is observed immediately after the preparation of the ink. This reaction is known to be fast, and thermodynamically favourable. Finally, the large differences between the spectra of pure gallic acid and IG3/\(r\) molar \(\equiv 0.1\), compared to the spectral differences between IG3/\(r\) molar \(\equiv 0.1\) and \(\equiv 0.2\), seem to indicate a reaction of the acid with the iron sulfate, rather than solely a reduced concentration of acid in the pellet. The results shown in Fig. 3 therefore suggest that with THz-TDS it is possible to observe differences in the composition of the ferro–gallic complex, although further studies are still required.

Thermally degraded iron gall inks

Since historically informed model iron gall inks show sharp terahertz spectral features associated with the molecular structure of iron(n) sulfate heptahydrate, or gallic acid, and since minor differences in composition of the ink result in minor differences in the spectral signatures, this leads us to believe that chemical changes, which might occur during ink degradation, could give rise to changes in terahertz spectral features.

Fig. 4 indicates some differences in the intensity of the sharp spectral bands at \(~38, 50\) and \(64\) cm\(^{-1}\) between samples of the same ink composition degraded for different time periods at 70 °C. No additional spectral features are observed after
degradation, apart from a spectral band at ~56 cm⁻¹ present in the spectra of some samples of IG3, which is the only ink produced from gallic acid rather than tannic acid (Table 1). Although all triplicate pellets of degraded IG3 exhibit this band, and although no band at 56 cm⁻¹ could be observed in the spectrum of the reference pellet of polyethylene, it cannot yet be asserted that the presence of this feature is not due to water vapour: the sample compartment might have been ideally purged only for the measurement of the reference pellet. The increase in absorbance at higher frequencies due to Mie scattering and the overlap of the spectra of ink samples degraded to different extents make any association of the evolution of the peak intensity with degradation time difficult.

PCA performed on the second derivative of the THz spectra of IG1, IG2, IG3 and IG5 between 10 and 90 cm⁻¹ is used to reveal, in an unsupervised way, if the differences in peak intensities are linked to the period of degradation. Fig. 5 shows that PC1 is related to the intensity of the peaks at 38, 51, 56 and 64 cm⁻¹ and that its value is decreasing for samples degraded for a longer period of time. This result clearly means that the longer the degradation time, the less intense the peak intensities, even when including some variability in the composition of the ink.

Such a good correlation between degradation times and peak intensities also confirms that the spectral differences between samples originating from the same ink composition are not linked to any sudden chemical or physical change occurring mainly during the first hour of heating of the sample, or other chemical changes induced by other factors (e.g. drying or grinding processes, exposure to different environmental conditions).

In conclusion, THz-TDS combined with PCA clearly shows that thermal degradation of iron gall inks leads to chemical reactions and/or structural rearrangements within the ink, resulting in a decrease in the intensity of sharp spectral features. In the spectroscopic study of carbon black, sepia and bistre inks, featureless spectra in the terahertz region were associated with amorphous materials, therefore, it might be concluded that thermal degradation of iron gall inks causes a loss of long-range symmetry or crystallinity within the ink structure.

Additionally, the area under the curve of the second derivative of absorbance spectra from IG3 samples at 54–58 cm⁻¹ was calculated. It confirms that the magnitude of the peak at 56 cm⁻¹ increases with longer thermal degradation. This confirms that this feature is not random, which would be expected if it was due to varying amounts of water vapour remaining in the sample compartment. Therefore, it is due to thermal degradation.

Iron(n) sulfate heptahydrate can either be oxidized or dehydrated at increased temperatures. While iron oxides and iron(n) sulfate are crystalline, iron(n) sulfate monohydrate can be amorphous. Therefore, during heating of inks, formation of crystalline oxidation products is expected to lead to additional spectral features, while dehydration could lead to featureless spectra. Consequently, it is believed that the decrease of bands at 51 and 64 cm⁻¹ for all inks with longer degradation times is due to a dehydration phenomenon, while the increase of the band at 56 cm⁻¹ for IG3 with longer degradation time is due to an oxidation phenomenon.

To confirm this hypothesis, ATR-FTIR spectroscopy was additionally performed on IG1, IG2, IG3 and IG5. Fig. 6a shows a decrease in intensity of the spectral band located at ~1645 cm⁻¹ relatively to the band at ~1610 cm⁻¹ for the sample of IG1 degraded 22 h at 70 °C compared to the non-degraded IG1 sample. The band at ~1645 cm⁻¹ is present in the spectrum of pure iron(n) sulfate heptahydrate and this frequency range corresponds to the O–H bending vibration of water molecules. The same phenomenon is observed for IG2 and IG5. This would confirm that dehydration is taking place during thermal degradation. On the contrary, increased intensity, broadening and a frequency shift of the maximum is observed for IG3 with longer degradation times. Additionally, a band located at ~1700 cm⁻¹ present in the spectrum of non-degraded IG3 is no longer observed for IG3 degraded for 22 h.

To further confirm that the origin of the decrease of the bands at 51 and 64 cm⁻¹ in Fig. 4 is due to dehydration, pellets of iron sulfate heptahydrate and model iron gall ink IG1 were placed in a nitrogen-purged chamber and heated from room...
temperature to above 125 °C (at which iron sulfate monohydrate should form) at 5 °C min⁻¹. Details about pellet preparation for this experiment can be found elsewhere.³¹ The purged environment prevents any oxidation phenomenon of the iron sulfate (or any other ink compounds): only dehydration can be observed. Fig. 7 shows that with increasing temperature, the decrease in intensity and the broadening of the peaks at 51, 64 and around 80 cm⁻¹ can be observed for both iron sulfate and iron gall ink, until a completely featureless spectrum is detected. It proves that the terahertz spectrum of iron sulfate monohydrate is featureless, and further confirms the results of ATR-FTIR spectroscopy: the decrease in intensity of the bands at 51 and 64 cm⁻¹ observed in the spectra of the thermally degraded iron gall inks is due to dehydration, rather than oxidation.

Yet, the different spectral behaviour of IG3 both in terahertz and ATR-FTIR analysis could indicate that different iron gall inks, prepared with the exact same recipe and proportions, do not undergo the exact same degradation processes depending on the nature of the acid used (gallic acid, tannic acid, or oak gall extracts). Considering the poor spectral feature of iron(II) sulfate at 56 cm⁻¹ in Fig. 2a, we cannot yet assert that the oxidation of iron(II) into iron(III) is responsible for the increase of the band at 56 cm⁻¹ in the spectra of degraded IG3 samples. Further research on the terahertz spectral signature of iron(III) oxides is needed to assess if the different spectral behaviour of IG3 is really due to an oxidation phenomenon of the iron. Additionally, similar heating experiments in a purged environment will be performed for IG3 and inks prepared with oak gall extracts, to further understand the role of the acid on the dehydration and oxidation phenomenon.

The possible oxidation of gallic or tannic acid into a quinoid form³² does not seem to be observable in the THz spectra of Fig. 4, since no additional spectral feature appears during degradation of the inks, apart from the band at 56 cm⁻¹. This does not mean there is no oxidation of the acid, but rather that if there is, the chemical compounds formed in this reaction do not exhibit sharp spectral bands in the low THz region. Previous work has shown that para-quinones show distinct spectral features in the low THz region³³ originating from the vibration of the hydrogen bonds holding the dimers of para-quinones together, but no further work has been published yet on the THz spectral signature of other quinoid compounds.
Conclusions

The potential of THz-TDS for the study of historical black inks has been explored and important results reported, shedding significant new light on the interpretation of terahertz spectra in general.

THz-TDS was shown to give valuable information on the macro-molecular structure of a material such as its long-range macromolecular symmetry, based on the presence or absence of sharp spectral features, or its degree of amorphousness, from the overall absorbance of featureless spectra. This was used to compare carbon black, bistre and sepia pigments (made from the soot of different organic materials, and from cuttlefish). This was also used to differentiate models of iron gall inks (issued from the reaction of iron ions with gallic acid moieties) depending on the ratios of iron(II) sulfate and acid, the nature of the acid and the amount of copper sulfate used for their preparation.

Principal Component Analysis was combined with THz-TDS to enable the gradual thermal degradation of iron gall ink models to be followed, via the progressive decrease in the intensity of sharp spectral features. This has shown, for the first time, that changes in molecular configuration occur during thermal degradation of ink. ATR-FTIR spectroscopy combined with THz-TDS of samples dynamically heated in a purged environment confirms that these spectral changes arise from a change in the hydration state of the iron ion.

This research also discusses the frequency-dependence of the refractive index and absorption coefficient acquired from THz-TDS, and suggests the exploitation of this information for the optimization of experimental conditions in THz imaging of layered structures (such as inks on supports).

In conclusion, this research contributes to the understanding of the interactions between terahertz radiation and matter, and pioneers terahertz spectroscopic studies of iron compounds, and their reaction with tannins. It shows that valuable chemical information can be obtained from iron gall inks using THz-TDS in transmission mode. The ATR mode could be investigated to assess if this chemical information can be extracted from smaller amounts of ink in an ambient environment, or even directly from the surface of an ink inscription. This chemical information could possibly also be extracted from the complex data acquired during contactless THz-TDS imaging of a document.

This study also opens the way to further research on the absorption of terahertz pulse by powders of amorphous organic materials, on THz-TDS study of metallic ions in different environments (different oxidation and hydration states), as well as on THz-TDS study of quinoid compounds possibly formed by oxidation of gallic acid.

Acknowledgements

The authors would like to thank the partners TeraView, Ltd and the Nationaal Archief. Professor Sir Michael Pepper, UCL Department of Electronic and Electrical Engineering and Professor Tom Fearn, UCL Department of Statistical Science, are gratefully acknowledged for their advice. UK Engineering and Physical Research Council Doctoral Training Centre in Virtual Environments, Imaging and Visualisation is acknowledged for the financial support. The authors are particularly grateful to Dr Bianca Jackson and the C2RMF, Paris, for sharing their knowledge on THz technology and for enabling the acquisition of the first images of the ink samples applied on paper and parchment.

Notes and references

3 R. Baumann, D. C. Porter and W. B. Seales, 9th International Conference on NDT of Art, Jerusalem, Israel, 2008.
5 W. B. Seales, J. Griffioen, R. Baumann and M. Field, 10th International Conference on non-destructive investigations and microanalysis for the diagnostics and conservation of cultural and environmental heritage, Florence, Italy, 2011.
Bibliography


laser using vanadium oxide microbolometer focal plane arrays. In B. F. And- 
drensen, G. F. Fulop, and P. R. Norton, editors, Proceedings of SPIE, Infrared 
Technology and Applications XXXIV, volume 6940, pages 69402Y1–69402Y12, 
2008.


[116] Cultural Heritage Advanced Research Infrastructures Synergy for a Multi-
disciplinary Approach to conservation/restoration (CHARISMA), Project co-
 funded by the European Commission within the action ’Research Infrastruc-
tures’ of the ’Capacities’ Programme GA No. FP7- 228330. [available online]. 
http://www.charismaproject.eu/.


Polytechnique, 2011.

thesis of a seventeenth-century panel painting by reflection terahertz time-domain 
imaging (THz-TDI): contribution of ultrafast optics to museum collections in-

Reflection terahertz time-domain imaging for analysis of an 18th century neo-

[121] P. Bartolini, A. Taschin, J. Tasseva, R. Torre, M. Barucci, C. Riminesi, and 
J. Striova. Drawing media as probed by THz time - domain and infrared spec-
troscopy. In THz-ARTE, Terahertz Advanced Research Techniques For Non-
Inavisve Analysis in Art Conservation, Frascati, Italy, 2014.


330 Bibliography


[177] R. T. Downs. The RRUFF Project: an integrated study of the chemistry, crystallography, Raman and infrared spectroscopy of minerals. In Program and Ab-


[246] Digital Archives and Special Collections. [available online]. http://digital.info.soas.ac.uk/.


Bibliography


[261] Undercover investigation of documents with terahertz. [available online], April 2014. https://www.youtube.com/watch?v=nVDpZs1-P7Q.


