Reconstruction tomographique et application en imagerie caméra Compton et en microscopie électronique

Voichita Maxim

▶ **To cite this version:**

Voichita Maxim. Reconstruction tomographique et application en imagerie caméra Compton et en microscopie électronique. Traitement du signal et de l'image [eess.SP]. Institut National des Sciences Appliquées de Lyon, 2019. tel-02946771

HAL Id: tel-02946771
https://hal.science/tel-02946771v1

Submitted on 23 Sep 2020

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

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

HABILITATION A DIRIGER DES RECHERCHES

présentée devant

l'Institut National des Sciences Appliquées de Lyon
et l'Université Claude Bernard LYON 1

Titre

Reconstruction tomographique et application en imagerie caméra Compton et en microscopie électronique

par

Voichiţa Maxim

Soutenue le 18 Mars 2019, devant la Commission d'examen :

Maï Nguyen-Verger	Professeur - Université de Cergy-Pontoise	Rapporteur
Claude Comtat	Ingénieur - CEA Orsay	Rapporteur
Jérôme Idier	DR CNRS - École Centrale de Nantes	Rapporteur
Denis Dauvergne	DR CNRS - Université Grenoble Alpes	Examinateur
Simon Masnou	Professeur - Université Lyon 1	Examinateur
Françoise Peyrin	DR INSERM, ESRF et Université de Lyon	Examinatrice

Laboratoire CREATIS

Table des matières

Introduction générale

Ce manuscrit présente essentiellement mes travaux de recherche menées au laboratoire CREATIS depuis ma nomination en tant que maître de conférences en 2005. Il contient une présentation succincte de mon parcours et de mes autres activités, au nombre desquelles mes activités d'enseignement. Les thématiques de recherche liées à ma thèse et au post-doctorat que j'ai effectué à l'université de Cambridge occupent une place restreinte dans ce document. En effet, les orientations que j'ai données à mes recherches au laboratoire CREATIS diffèrent sensiblement de mes activités antérieures.

Au début des années 2000 il était question de construire à Lyon un centre de traitement des tumeurs cancéreuses par hadron-thérapie. Cette technique d'irradiation des tumeurs avec des ions, proposée en 1946 par Robert Wilson, présente certains avantages comparé à la radio-thérapie par rayons X. Tout d'abord, il est possible en théorie de cibler plus précisément la tumeur et d'épargner les tissus sains environnants. L'ion qui se déplace dans la cible est ralenti de manière quasi-continue par les électrons du milieu et subit des collisions élastiques/inélastiques avec les électrons ainsi que des collisions élastiques avec les atomes. Ce ralentissement est caractérisé par le pouvoir d'arrêt $S(E) = -dE/dx$, dépendant de l'énergie de l'ion. Ce pouvoir d'arrêt augmente au cours du ralentissement. La section efficace d'interaction augmente lorsque l'énergie de la particule décroit et au final, le pouvoir d'arrêt chute. C'est le pic que l'on peut observer si l'on trace le pouvoir d'arrêt en fonction de la distance parcourue dans le matériau. Ce pic est appelé pic de Bragg d'après Sir William Henry Bragg qui a étudié le ralentissement de particules α dans l'air. L'énergie de l'ion est déposée dans le milieu par ionisation mais aussi par des réactions nucléaires qui produisent des fragments de particules voire la désintégration complète du projectile et/ou des noyaux de la cible. Des particules secondaires sont créées lors de ces interactions (fragments d'ions, neutrons, photons, électrons) et contribuent elles aussi à la dose reçue. En conclusion, les ions déposent une importante quantité d'énergie sur une très courte distance avant de s'arrêter complètement.

Ces propriétés balistiques sont très importantes en hadronthérapie car avec les ions, on obtient une meilleure précision sur le traitement des tumeurs en limitants les dommages aux organes à risques. Ainsi, les effets indésirables en hadronthérapie sont moindres pour le patient en comparaison de la radiothérapie conventionnelle.

En entrant dans le patient, les ions transfèrent leur énergie par des chocs successifs et provoquent une excitation et une ionisation des cellules cibles. En radiobiologie, on parle de transfert linéaire d'énergie ou TLE : c'est l'énergie moyenne déposée par unité de

longueur de la trajectoire de la radiation. Il faut distinguer deux catégories de radiations :

— Les radiations à faible TLE qui correspondent aux radiations photoniques X ou gamma,

— Les radiations à TLE élevé qui correspondent aux radiations particulaires tels que les hadrons.

Les particules chargées ont un TLE plus élevé que les photons car le dépôt d'énergie est plus important le long de la trajectoire. Ceci influe sur l'efficacité biologique relative (EBR), c'est-à-dire que lorsque le TLE croît le RBE croît également et les particules chargées ont donc un effet sur les cellules plus efficace que les photons X ou gamma. Cependant, lorsque le TLE atteint environ 100 keV/μm, l'EBR commence à décroître car l'énergie dépasse celle nécessaire pour détruire les cellules. Par conséquent, si on irradie des tissus avec un TLE trop élevé, les cellules ciblées meurent à partir de ce seuil et l'irradiation comporte des risques pour les cellules saines. Dans le cas de l'hadronthérapie par exemple où le TLE est élevé, les ionisations et les excitations des particules se produisent dans un espace assez petit, qui peut atteindre la dimension d'une biomolécule. Les dommages sont donc plus faciles à causer car l'irradiation provoque des lésions de l'ADN entrainant la mort des cellules. L'hadronthérapie s'avère donc plus efficace d'un point de vue biologique et est une alternative intéressante dans le cadre du traitement des cancers.

En contre-partie, le risque de focaliser par erreur le faisceau sur une région avoisinant la tumeur est grand, au vu de la difficulté d'estimer très précisément la composition chimique des tissus et par conséquent la fin de parcours du faisceau. A cela s'ajoutent les erreurs venant du positionnement du patient, des changements anatomiques entre l'imagerie pré-opératoire et le traitement ainsi que les mouvements physiologiques. Cela souligne l'importance d'un contrôle si possible en temps réel du traitement, par des mesures extérieures. Ces mesures peuvent se faire sur des particules secondaires émises au cours de l'irradiation, par exemple, les photons.

Avec l'essor des centres de proton et hadron-thérapie en Europe et dans le monde, la problématique du contrôle en ligne suscite beaucoup d'intérêt. Les chercheurs de l'Institut de Physique Nucléaire de Lyon se sont engagés dans cette voie. D'autres laboratoires dont CREATIS les ont rejoint dans une structure qui a pris le nom d'"Étoile". Cette structure a pris diverses formes, cependant la structure hospitalière n'a jamais vu le jour et le projet a été abandonné depuis. Il en reste cependant une activité de recherche autour de l'imagerie de contrôle en proton-thérapie et autour de la radio-biologie.

Lors de mon arrivée à CREATIS, le projet du centre hospitalier Étoile avait des chances pour aboutir et participait à un réseau Européen nommé ENLIGHT qui réunissait des groupes de divers pays promouvant ce type de traitement. Accompagnée et conseillée par Rémy Prost, qui dirigeait l'équipe dans laquelle j'ai été intégrée, j'ai cherché sous quelle forme nous pouvions participer à ce projet. Suite à des discussions avec les collègues de l'IPNL et à l'étude de la bibliographie, il nous a paru intéressant d'étudier le rayonnement γ-prompt. En 2007 nous avons obtenu un financement de thèse et nous avons démarré une nouvelle thématique dans le laboratoire, celle de l'imagerie caméra Compton. Le choix a été motivé en partie par le souhait de rester proches des thématiques IPNL et de

bénéficier de leur expertise. En effet, la même année commençait à l'IPNL une thèse sur la détection γ-prompt avec une caméra collimatée. Mais nous souhaitions aussi que les deux sujets soient suffisamment distincts. Notre laboratoire étant spécialisé en imagerie, il était naturel de s'intéresser à un détecteur capable de réaliser des images. L'idée de la caméra Compton nous est venue de l'astronomie et des travaux qu'Andreas Zoglauer a effectués au Max-Planck-Institut für extraterrestrische Physik de Garching, Munich. Nous avons beaucoup utilisé son manuscrit de thèse et son logiciel de simulation Monte Carlo capable de simuler l'acquisition caméra Compton, le pré-traitement de données et la reconstruction d'images.

Il nous est rapidement apparu que la tomographie caméra Compton était assez peu étudiée. Les algorithmes existants étaient en majorité itératifs, parfois en 2D, et le modèle d'acquisition, assez complexe, assez peu explicité dans les publications. Pour mieux comprendre les modèles et le processus de reconstruction, nous avons commencé par l'étude des méthodes analytiques. Nous avons proposé une méthode originale, qui a l'avantage d'être facile à implémenter et d'utiliser une grande partie de données. Il faut mentionner que les données provenant de caméra Compton sont redondantes, du moins en théorie, car en pratique ces redondances s'accompagnent de troncations des projections qui ne sont pas caractérisables avec le vocabulaire établi, celui des projections tronquées où de l'angle de vue limité. Il est probable que cette publication ait participé à faire naître un nouveau champ d'étude, celui de la transformée de Radon conique. Peu avant, les travaux de Mai Nguyen-Verger et Tuong T. Truong (laboratoire ETIS - ENSEA) sur la tomographie du diffusé Compton, avaient initié l'étude des projections coniques dans un cadre différent, où la diffusion a lieu dans l'objet et la mesure est la somme de projections coniques. Plusieurs articles sont parus ces dernières années sur l'étude et l'inversion de la transformée de Radon conique. Certains de ces travaux peuvent être reliés à l'imagerie par caméra Compton, d'autres à d'autres applications.

Les méthodes itératives, en particulier l'algorithme Maximum Likelihood Expectation Maximization (MLEM), sont assez bien connues. Néanmoins, leur particularisation au dispositif est assez délicate, car nécessite de définir le modèle adéquat pour les données. Les temps de calcul et les besoins en mémoire sont très élevés, ce qui impose des adaptations informatiques. Pour l'instant, nous avons effectué une parallélisation du code et nous envisageons une implantation sur carte graphique (GPU). À part la difficulté de trouver un modèle adapté, les données dont on dispose sont incomplètes, du fait de la géométrie de détection, et sont de plus fortement bruitées. Le bruit d'acquisition joue probablement plus dans cette modalité qu'en SPECT ou TEP, car les points d'émission sont situés sur un cône, et de faibles écarts sur la direction de l'axe ou sur l'angle d'ouverture produisent des écarts importants sur la retro-projection, ceci d'autant plus que l'on s'éloigne de la caméra. Il est important de prendre en compte ces erreurs et d'ajouter des connaissances a priori dans le processus de reconstruction.

Je fais à présent partie de l'équipe "Imagerie tomographique et thérapie par rayonnements", dans laquelle sont étudiées différentes techniques de tomographie avec des rayonnements de diverses natures. L'équipe est aussi très proche des applications, grâce à

la présence de chercheurs dans des centres de recherche (ESRF) ou de traitement (Centre de lutte contre le cancer, Léon Bérard). Le laboratoire fait partie du LabEx PRIMES (Physique, Radiobiologie, Imagerie Médicale et Simulation).

La structure du manuscrit est la suivante. La première partie contient une présentation générale de mon cursus, commençant avec mes études en mathématiques fondamentales complétés par des études en informatique, en passant par une thèse en mathématiques appliquées et ceci jusqu'à mes activités présentes en imagerie. Cette partie inclut la description de mes travaux de thèse et post-doctoraux, les enseignements que j'ai dispensés et la production scientifique. La deuxième partie passe en revue mes contributions en tomographie caméra Compton et en microscopie électronique de transmission. Ces contributions sont parfois personnelles, parfois issues de travaux que j'ai co-encadrés. La présentation est volontairement assez synthétique. Pour des précisions sur les aspects mathématiques et pour plus de détails, le lecteur intéressé pourra consulter les documents annexés ou cités. Enfin, la dernière partie contient les conclusions et perspectives.

Première partie

Présentation des activités

Chapitre 1

Curriculum Vitae

1.1 Synthèse de carrière et diplômes

Maître de conférences 2005-

INSA Lyon

- Recherche en imagerie médicale : tomographie mono-photonique avec une caméra Compton, application à l'imagerie γ-prompt pour le contrôle en ligne de la proton-thérapie. Depuis 2014 : tomographie en microscopie électronique de transmission.
- Co-auteur de 16 articles dans des revues internationales à comité de lecture.
- Enseignement : Mathématiques générales au département de Formation initiale aux métiers d'ingénieur (FIMI), Analyse numérique en 3ème année Génie Électrique, Représentation et analyse de données (FIMI), Imagerie TEP (Master Physique Médicale), Informatique (FIMI)
- Membre du Comité de Gestion et Animation pour la discipline mathématiques, premier cycle (FIMI) de l'INSA de Lyon en tant que responsable pour la discipline mathématiques de la deuxième année de la filière classique (2010-2017)
- Membre du conseil de l'école doctorale EEA depuis juillet 2018.
- Titulaire de la PEDR depuis octobre 2017.

ATER 2004-2005

INSA Toulouse

- Département Génie Mathématique et Modélisation, Laboratoire de Statistique et Probabilités
- Travaux dirigés de probabilités et statistique ; travaux dirigés et travaux pratiques (logiciel R) de qualité et fiabilité de mesures physiques ; travaux dirigés et travaux pratiques (langage C) d'analyse numérique.

Research Associate 2003-2004

Université de Cambridge

- Équipe Brain Mapping Unit du Département de Psychiatrie de l'Université de Cambridge,

dirigé par Pr. E. Bullmore. Projet financé par NIH "Wavelets, fractals and fMRI of brain adaptivity".
— Publications : Bullmore et al (NeuroImage 2005), Maxim et al (NeuroImage 2005), Sendur et al (IEEE TSP 2005)

ATER 2002-2003

Université Joseph Fourier, Grenoble 1

— Laboratoire de Modélisation et Calcul
— Travaux dirigés de statistique inférentielle en DEUG SV/STU 2ème année ; TD-cours de statistique descriptive en DEUG SV/STU 1ère année ; TD de mathématiques, IUP-MAI, 1ère année.

Doctorat de Mathématiques Appliquées 1999-2003

Université Joseph Fourier (Grenoble I)

— Laboratoire de Modélisation et Calcul (à présent Laboratoire Jean Kuntzmann)
— Titre : "Restauration de signaux bruités observés sur des plans d'expérience aléatoires".
— Dirigée par M.-L. Mazure et A. Antoniadis
— Jury de thèse : V. Perrier (INPG, présidente du jury), J.M. Poggi (Université Paris 5, rapporteur), J.L. Merrien (INSA Rennes, rapporteur), M.L. Mazure et A. Antoniadis (Université Joseph Fourier, Grenoble 1, directeurs de thèse), G. Grégoire (Université Pierre Mendès France, Grenoble 2, examinateur).
— Un article paru dans *Numerical Algorithms* en 2004, un article de conférence de 10 pages (Fifth AFA-SMAI conference on Curves and Surfaces, St. Malo, 2002)

Vacataire 1999-2002

Université Joseph Fourier, Grenoble 1

— travaux dirigés de mathématiques générales, travaux dirigés de statistique en DEUG

Enseignante en lycée 1997-1999

Lycée Mihai Eminescu, Satu Mare, Roumanie

— Poste de contractuel en Mathématiques pendant 1 an.
— Titularisation sur un poste en Informatique (concours similaire à l'agrégation française)
— Préparation des élèves aux Olympiades de Mathématiques

Master 2 de Mathématiques et Maîtrise d'Informatique 1996-1997

Université Babeş-Bolyai Cluj, Roumanie

— Spécialité Analyse Convexe et Approximation
— Bourse Tempus pour un séjour de trois mois au LMC-IMAG, Grenoble
— Stage de DEA effectué sous la direction de Pr. Pierre-Jean Laurent, sur le sujet "Espaces de Chebyshev et effets de forme en design géométrique"
— Une publication parue dans *Numerical Algorithms* en 1997
— Double cursus et double diplôme Mathématiques et Informatique

1.2 Activités d'encadrement

- **Thèses soutenues :**

1) Estelle Hilaire – 2012-2015 – 50% – bourse du département INSIS du CNRS, "Simulation et reconstruction 3D à partir de caméra Compton pour l'hadronthérapie : influence des paramètres d'acquisition," *INSA de Lyon*.
   — Co-direction avec F. Peyrin, DR INSERM.
   — Un article dans Phys. Med. Biol. en 2016 [A13] et deux conférences avec actes de 4 pages en premier auteur (IEEE NSS-MIC 2014 [C6], GRETSI 2013 [CN2]). Participation à un article paru dans Phys. Med. Biol. en 2016 [A12]. Participation à des conférences avec résumés courts (ex. ICTR-PHE 2014).
   — **Récompensée par un prix des Amis de l'Université de Lyon**.
   — Devenir : Co-fondatrice de la start-up Karaïb 3D en Martinique.
2) Xavier Lojacono – 2010-2013 – 50% – financé par le projet européen ENVISION, "Image reconstruction for Compton camera with application to hadrontherapy," *INSA de Lyon*.
   — Co-direction avec R. Prost, PU 61 INSA.
   — Un article paru dans IEEE TNS en 2013 [A9] et trois conférences avec actes de 4 pages en premier auteur (FULLY 3D [C3], GRETSI [C1], IEEE NSS-MIC [C4], en 2011). Participation à deux articles, parus respectivement dans IEEE TNS [A8, 2012] et Phys. Med. Biol. [A12, 2016]. Participations à des conférences avec résumés courts.
   — Devenir : ingénieur informaticien dans le secteur privé.
3) Mirela Frandes – 2007-2010 – 50% – financée par une allocation de recherche, "Gamma-ray detection and Compton camera image reconstruction with application to hadron therapy," *INSA de Lyon*.
   — Co-direction avec R. Prost, PU 61 INSA.
   — Deux articles parus dans IEEE TNS en 2010 [A7] et 2011. Deux conférences avec actes de 4 pages en premier auteur (IEEE NSS-MIC 2008 et 2009).
   — Devenir : enseignant-chercheur en Roumanie.

- **Thèse en cours :**

1) Yuemeng Feng – 2016-2019 – 50% – bourse CSC, "Dynamic detection design and acquisition modeling in tomographic reconstruction for Compton camera imaging," *Université de Lyon*.
   — Co-encadrement avec D. Sarrut, 50%.
   — Premiers résultats présentés à la conférence IEEE NSS-MIC 2018 [C8] ; participation à un article soumis [P3].

- **Encadrement de stage niveau Master**

1) Benjamin Ganguet (100%), Projet de Fin d'Études de 5 mois en 2018, Phelma, INP Grenoble. Titre "Tomographic reconstruction of contours with application in electron microscopy".
2) Yuemeng Feng (50%), stage de master-PFE de 6 mois en 2016, master EEAP, spécialité Signaux et Systèmes, INSA Lyon. Titre "Correction de flou dans des images de tomographie électronique".
   — Co-encadrement avec T. Grenier (MdC INSA Lyon).
   — Poursuite en thèse à CREATIS sous la direction de D. Sarrut et V. Maxim.
   — Travaux présentés à "The 16th European Microscopy Congress 2016" [C3], participation à un article dans Microsc. Microanal. 2016 [A14].
3) Khanh-Hung Tran (50%), Projet de Fin d'Études de 6 mois en 2016, département Génie Électrique de l'INSA de Lyon. Titre : "Automatisation de l'acquisition de séquences de projections en tomographie électronique de transmission".
   — Co-encadrement avec T. Grenier (MdC INSA Lyon).
   — Participation à un article dans Microsc. Microanal. 2016 [A14] et à un résumé de conférence [C3].
4) Tiphaine Meyet (100%), stage de 12 semaines en 4ème année Phelma, INP Grenoble. Titre "Réalisation en Python d'un logiciel d'analyse de données et de reconstruction d'images pour la caméra Compton". Transfert en Python du logiciel CORESI et ajout de nouvelles fonctionnalités pour l'analyse de données.
5) Hussein Banjak (100%), Projet de Fin d'Études de 6 mois en 2013, spécialité Signaux Multimédias et Technologie Réseaux, École Polytechnique Universitaire de l'Université de Nantes. Titre : "Fully three-dimensional analytic image reconstruction for Compton camera". *Poursuite en thèse au CEA, région parisienne puis post-doc à CREATIS.*
6) Estelle Hilaire (100%), Stage de Master de 5 mois en 2012, spécialité Physique Médicale, École Polytechnique Universitaire de l'Université de Lyon. Titre : "Développement d'un algorithme itératif de reconstruction d'images pour l'imagerie SPECT avec une caméra Compton". *Poursuite en thèse à CREATIS.*
7) Rohit Prasad (100%), stage de master 2 de 5 mois en 2012, master EEAP parcours Signal et Image à Lyon. Titre : "Comparative study of methods for calculating the transfer matrix of the Compton imaging system".
8) Xavier Lojacono (100%), stage de master de 5 mois en 2010, ENSEIRB. Titre : "Evaluation d'une méthode de reconstruction d'images pour une caméra Compton générique". *Poursuite en thèse à CREATIS.*

- **Encadrement de post-doctorat**

J'ai organisé le recrutement et je supervise le post-doctorat de Hussein Banjak, que nous avons recruté sur le projet ANR 3DCLEAN pour une durée de 2 ans. Son objectif était de développer un algorithme de reconstruction tomographique pour la microscopie environnementale rapide, algorithme implémenté sur GPU. Les données sont des projections incomplètes d'un objet susceptible de se déplacer et de se modifier au cours de l'acquisition. Son travail fait suite à deux stages de master : Yuemeng Feng a étudié la possibilité de déflouter en aveugle les images de projection (conférence The 16th European Microscopy Congress 2016 [3]) et Khanh Tran a étudié le réalignement d'images mais aussi la sélection automatique des projections valides. Le code de reconstruction, incluant une régularisation de type variation totale, a été validé sur des données réelles et des résultats ont été publiés dans Ultramicroscopy [A15]. A présent, Hussein Banjak étudie les méthodes deep learning dans le but d'accélérer encore la reconstruction ([C7]) et éventuellement de réduire les artefacts dus à l'angle manquant.

1.3 Responsabilités scientifiques

Responsable local du projet 3DCLEAN 2015-2018

ANR

Le projet 3D CataLytic Environmental lAb at the Nanoscale, sur la tomographie en microscopie environnementale rapide poursuit les travaux et la collaboration entamées dans le projet SPEE3D. Partenaires : MATEIS (INSA Lyon, coordinateur), IFPEN (Lyon), IPCMS (CNRS, Strasbourg), CREATIS (INSA Lyon). Montant de l'aide allouée à CREATIS : 160 k€, financement de deux ans de post-doctorat. Chercheurs impliqués : F. Peyrin (DR INSERM), I. Magnin (DR INSERM), T. Grenier (MdC INSA), V. Maxim, H. Banjak (post-doctorant). J'ai la responsabilité de signature pour l'utilisation des fonds ; j'ai organisé le recrutement du post-doctorant ; je supervise ses travaux.

Co-porteur du projet SPEE3D 2014-2016

BQR INSA

Suivi raPide et Environnemental de l'Évolution de nanomatériaux en 3D (15 k€) avec le laboratoire MATEIS. Tâche : déconvolution et réalignement des projections en tomographie électronique environnementale rapide. Financement de deux stages master–PFE, Y. Feng et K. Tran. Résultats : un article dans Microsc. Microanal. 2016 [A14] et deux actes de conférence "The 16th European Microscopy Congress 2016" [C3, C2]

Participant au projet européen ENVISION 2010-2014

Programme FP7-HEALTH-2009

Dans le projet European NoVel Imaging Systems for ION therapy notre tâche visait la reconstruction d'images pour la caméra Compton (financement d'une thèse de doctorat que j'ai co-dirigé et des missions). Collaboration avec l'IFIC (Université de Valence, Espagne) et OncoRay (TU Dresde, Allemagne). Publications : IEEE TNS 2013 [A9], IEEE TIP 2014 [A10], Phys. Med. Biol. 2016 [A12, A13]

Participant au projet régional de recherche en hadron-thérapie 2007-2013

CPER

Tâche : imagerie de contrôle en hadron-thérapie par détection des particules gamma prompt, reconstruction tomographique pour la caméra Compton (financement de missions et d'un stage master, E. Hilaire)

1.4 Autres activités et responsabilités

Conseil de l'École Doctorale 2018-

EEA, Lyon

Participation aux réunions et à des comités de suivi de thèse

Responsable pour la discipline mathématiques de la deuxième année de la filière classique 2010-2017

Premier cycle de l'INSA de Lyon

Responsable pour la discipline mathématiques de la deuxième année de la filière classique (15 groupes de TD regroupés en quatre "lanières" correspondant chacune à un amphi, 375 étudiants). Organisation de la progression, des contrôles de connaissances (5/année), des rattrapages, adaptation aux changements de maquette horaire et de programme intervenus suite aux changements de programme en lycée et à l'introduction des Parcours Pluridisciplinaires d'initiation à l'Ingénierie. Membre du Comité de Gestion et Animation pour la discipline mathématiques.

Membre expert du comité de suivi de thèse d'Amandine Verguet 2018

Institut Curie, Paris.

"Développement méthodologiques et informatiques pour la cryo-microscopie électronique en transmission à balayage des spécimens biologiques"

Travaux pratiques pour l'imagerie TEP 2015

École d'été Physique de l'Imagerie Médicale et Simulation

Participation à la réalisation d'un TP sur la simulation GATE pour la TEP dans le cadre de l'école d'été Physique de l'Imagerie Médicale et Simulation (Lyon 6-10 juillet 2015) avec la préparation d'un module pour la reconstruction tomographique MLEM en TEP.

Examinateur (expert en reconstruction d'images caméra Compton) pour la thèse de M. J.L. Ley 2015

Université de Lyon, spécialité Physique Médicale

"Mise en œuvre d'un démonstrateur de caméra Compton pour l'imagerie en médecine nucléaire et pour le contrôle en temps réel de l'hadronthérapie à l'aide des rayonnements gamma prompts". Composition du jury : S. Jan (rapporteur), G. Montarou (rapporteur), G. Chanfray (président), J.-P. Cussonneau (examinateur), D. Dauvergne (directeur de thèse), V. Maxim (examinateur), F. Peyrin (examinateur), P. Thirolf (examinateur).

Rapporteur : Agence Nationale de la Recherche, Université de Nantes, IEEE TIP, IEEE TNS, IEEE TMI, SIAM Journal on Mathematical Analysis, Inverse Problems, Applied Radiation and Isotopes, Japanese Journal of Applied Physics, Nuclear Inst. and Methods in Physics Research (A), EUSIPCO 2016.

Chapitre 2

Résumé des activités d'enseignement à l'INSA de Lyon

Cours de mathématiques 2010-

Deuxième année INSA (classes préparatoires intégrées)

Réalisation d'un polycopié pour le cours (63h puis 42h face-à-face), TDs (90h puis 86h). Ce polycopié est utilisé par plusieurs "lanières" de la filière classique et fait référence pour le contenu du programme de la deuxième année de classes préparatoire intégrées INSA Lyon.

Schémas numériques 2015-

Premier cycle INSA Lyon

Cours sur les schémas numériques pour les équations différentielles et aux dérivées partielles, dans le parcours pluridisciplinaire d'initiation à l'ingénierie appellé Modélisation numérique pour l'ingénieur mis en place à l'INSA de Lyon à la rentrée 2015 (cours+TD+TP, projet, 38h)

Analyse numérique 2015-

Département Génie Électrique, INSA Lyon

Interpolation polynomiale et spline, moindres carrés, intégration numérique, résolution de systèmes linéaires, conditionnement et erreur pour la résolution de système linéaire, résolution d'équations non linéaires, EDO et EDP. (Cours+TD+TP, 44h)

Représentation et analyse de données, outils statistiques et fréquentiels 2011-2015

Deuxième année INSA (classes préparatoires intégrées)

J'ai élaboré et dispensé la partie portant sur l'analyse temps-fréquence et temps-échelle (ondelettes) de l'Approfondissement Mathématique de la filière Eurinsa (cours+TD+TP,

21h).

Imagerie TEP, dans le module Imagerie Isotopique 2006-2009

Master 2 de Physique Médicale, Université Claude Bernard Lyon 1 (cours + TD + TP, 12h)

Travaux pratiques pour l'imagerie TEP 2015

École d'été Physique de l'Imagerie Médicale et Simulation

Participation à la réalisation d'un TP sur la simulation GATE pour la TEP dans le cadre de l'école d'été Physique de l'Imagerie Médicale et Simulation (Lyon 6-10 juillet 2015) avec la préparation d'un module pour la reconstruction tomographique MLEM en TEP.

Chapitre 3

Résumé des activités de recherche

3.1 Travaux effectués pendant la thèse

J'ai effectué ma thèse de doctorat au Laboratoire de Modélisation et Calcul de l'Université Joseph Fourier, Grenoble I, sous la direction conjointe des professeurs Marie-Laurence Mazure et Anestis Antoniadis. Cette thèse porte sur le développement d'algorithmes de subdivision et la statistique non paramétrique pour des problèmes d'estimation fonctionnelle.

Plus précisément, l'objectif principal du travail était l'estimation d'un signal contaminé par un bruit, ceci à l'aide de la décomposition en ondelettes. Les bases d'ondelettes permettent de bien approcher sur un petit nombre de coefficients les signaux réguliers par morceaux. Ces bonnes capacités d'approximation non linéaire conditionnent l'efficacité d'un débruitage du signal par seuillage des coefficients d'ondelettes. De plus, pour des signaux échantillonnés de manière discrète sur une grille régulièrement espacée, une transformée en ondelettes orthogonales s'implémente par un algorithme rapide à l'aide de cascades de filtres miroirs conjugués, permettant ainsi d'obtenir des algorithmes de débruitage dont la rapidité est suffisante pour les applications.

En restant dans le cadre de la régression non paramétrique, les informations a priori sur le signal f à restaurer s'expriment par une condition d'appartenance de f à un espace S de signaux vérifiant certaines conditions. Selon les propriétés géométriques de S, il est alors possible d'utiliser des méthodes d'estimation linéaires ou non linéaires permettant de minimiser l'erreur quadratique moyenne. Lorsque S est constitué de fonctions "lisses", le recours à des méthodes basées sur les décompositions d'ondelettes est moins justifié que dans le cas où les fonctions de S possèdent certaines formes de régularité par morceaux et sont spatialement peu homogènes. C'est dans ce contexte que les méthodes de seuillage d'ondelettes sont performantes grâce à leur pouvoir d'approximation adaptatif. C'est le cadre que nous avons choisi pour ce travail. Néanmoins l'optimalité de ce type de méthodes a été principalement étudiée dans le cadre d'un échantillonnage régulier. Nous avons cherché des méthodes d'ondelettes pour l'estimation non paramétrique de fonctions de régression à échantillonnage aléatoire.

Les méthodes classiques de reconstruction par ondelettes peuvent seulement traiter des données issues d'un échantillonnage équidistant, ce qui est une contrainte trop forte dans nombre d'applications, notamment en régression. Les recherches portent maintenant sur les façons de s'affranchir de cette contrainte.

Le problème étudié est :

Trouver la fonction f connaissant les valeurs

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

où X_i, $i = 1, \ldots, n$, est un échantillon de la variable aléatoire X de loi G (en bref la grille) et ε_i, $i = 1, \ldots, n$ est un bruit aléatoire.

Diverses méthodes ont été proposées pour remédier à cet inconvénient. L'idée commune d'une grande partie des méthodes qui ont été proposées est de se ramener à une estimation sur grille régulière. Nombre d'entre elles interpolent ou moyennent les données initiales aux points d'une grille équidistante. Appliquer un algorithme standard, pour échantillonnage régulier, à un échantillonnage soit non-équidistant donne des résultats insatisfaisants comme le montre l'exemple de la figure 3.1. Le plan d'expérience, représenté avec des astérisques, est déterministe : $X_i = G^{-1}(i/n)$, $i = 1, \ldots, n$. La fonction G est définie sur $[0, 1]$ par $G(x) = x + \frac{1}{10\pi}\sin(7\pi x + \pi)$. La ligne continue est la fonction f. A partir des valeurs (non bruitées) de cette fonction aux points d'échantillonnage, on construit deux approximations : en pointillé celle obtenue en considérant les points comme étant équidistants et en ligne brisée celle calculée avec une méthode adaptée que nous développerons ci-après. On remarquera les artefacts de la première méthode.

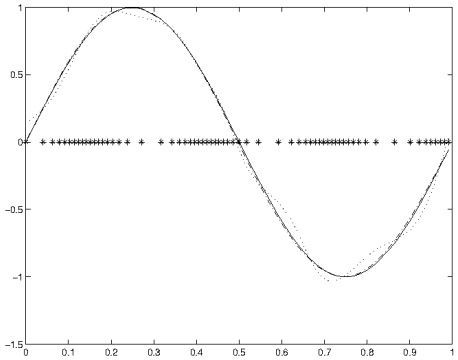


FIGURE 3.1 – Comparaison entre la méthode naïve, consistant à traiter l'échantillonnage comme déterministe (pointillés), et une méthode adaptée à l'échantillonnage (Brown et Cai, 1998) en ligne brisée. La fonction de réference, tracée avec une ligne continue, est connue aux points marqués par astérisque.

Une méthode intéressante est celle proposée par T. Cai [1]. En considérant les données comme étant équidistantes, il construit un approximant dans l'espace d'échelle le plus fin. Puis il compose cet approximant avec la fonction G, ce qui met en correspondance les

1. L. D. Brown, T. Cai. Wavelet shrinkage for nonequispaced samples. Annals of Statistics, 1998. Ce travail est issu du mémoire de thèse du second auteur, à Cornell University.

valeurs Y_i avec celles de la grille initiale, non équidistante. En décomposant le résultat dans la base d'ondelettes et en seuillant les coefficients, il obtient un estimateur dont le risque en moyenne quadratique est presque optimal pour les fonctions hölderiennes par morceaux ; ce risque est de même ordre de grandeur dans le cas des points équidistants.

Néanmoins, son résultat concerne les échantillonnages déterministes : les points de la grille sont de la forme $X_i = G^{-1}(t_i)$, pour $i = 1, \ldots, n$ et $t_i = i/n$. Quand l'échantillonnage est aléatoire et la fonction G inconnue on pourrait remplacer G par G_n, l'interpolation affine par morceaux de la fonction de répartition empirique, mais il n'est pas possible d'assurer la convergence de l'estimateur vers la fonction à reconstruire quand la taille de l'échantillon tend vers l'infini.

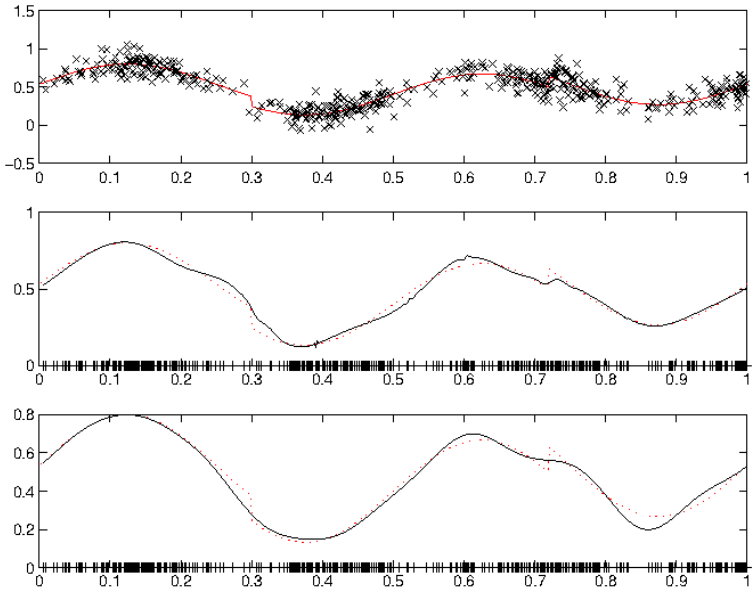


FIGURE 3.2 – Estimation de la fonction HeaviSine à partir d'un échantillon bruité de 512 points (représentés sur la première ligne par des croix), utilisant pour la grille d'abord l'estimateur log-spline $\widehat{G}_n$ (deuxième ligne), puis l'estimateur affine par morceaux G_n (dernière ligne). Le rapport signal-sur-bruit est ici SNR=2. Les discontinuités sont mieux réproduites lorsque l'estimation log-spline est utilisée pour la fonction définissant l'échantillonnage.

Nous proposons une modification de cette méthode qui permet d'obtenir la convergence même si les points sont aléatoires. L'idée est d'utiliser un estimateur $\widehat{G}_n$, par exemple de type log-spline, pour la fonction inconnue G. Cet estimateur, proposé par C. Stone [2], est défini sauf pour un événement dont la probabilité tend vers 0 quand la taille de l'échantillon tend vers l'infini. Ses propriétés nous permettent de montrer que, conditionnellement à la grille, le risque d'estimation relatif à des classes des fonctions hölderiennes reste presque optimal, en probabilité :

$$\sup_f E(\|f - \hat{f}_n\|_2^2 | X) = O_P\left(\left(\frac{\log n}{n}\right)^{\frac{2\alpha}{1+2\alpha}}\right),$$

2. C. J. Stone. Large sample inference for log-spline models. The Annals of Statistics, 1985.

α étant le régularité hölderienne de la fonction f. Un test avec une fonction présentant des discontinuités est montré dans la figure 3.2.

Ces résultats ont été présentés lors d'une communication à la Conférence Internationale "Curves and Surfaces" en juin 2002 à Saint Malo. Un article est paru dans les actes de cette conférence [C1].

Nous avons également travaillé sur une méthode de reconstruction des fonctions appartenant à des espaces de Besov. Ces espaces sont plus généraux que les espaces de Hölder et ils contiennent des fonctions moins lisses. Dans ce cas, nous utilisons la régression polynomiale locale pour estimer les valeurs de la fonction inconnue sur une grille équidistante, puis nous appliquons les algorithmes d'ondelettes à ces données, pour "lisser" le résultat. Cette méthode permet de traiter des données bruitées avec un bruit non-gaussien. Nous montrons que le risque L_2 d'estimation est presque optimal, pour des fonctions appartenant à des espaces de Besov B^s_{pq} avec $1 \leq p \leq 2$, $1 \leq q$ et $s > 1/p$, et avec certaines contraintes concernant les moments du bruit et la fonction grille.

A ce point, il est intéressant de comparer les deux méthodes. Nous avons montré qu'elles convergent à la vitesse presque optimale sur certaines classes de fonctions. Des résultats plus généraux montrent qu'aucun estimateur basé sur le seuillage des coefficients d'ondelettes ne peut dépasser cette vitesse. Cependant, on peut imputer à la première méthode le fait que le risque a été calculé conditionnellement à la grille et que les fonctions pour lesquelles cette vitesse est atteinte sont très lisses. Ces inconvénients viennent du fait que la fonction de grille G a été estimée, et le résultat est optimal à la lumière des résultats existant sur l'estimation d'une densité. Pour des fonctions aussi lisses il y a des estimateurs linéaires optimaux, mais qui ne sont pas adaptatifs. Autrement dit, nous avons besoin de connaître la régularité de la fonction afin de définir l'estimateur optimal, contrairement à notre estimateur qui s'adapte naturellement à la régularité de la fonction.

Le deuxième estimateur que nous avons défini atteint la vitesse de convergence optimale parmi les estimateurs non linéaires, pour des fonctions appartenant à des espaces de Besov B^s_{pq} avec $1 \leq p \leq 2$, $1 \leq q$ et $s > 1/p$. Ces espaces sont plus généraux que les espaces de Hölder. On sait que les estimateurs linéaires n'aboutissent pas au taux de convergence optimal sur ces espaces de fonctions peu régulières, et nous avons montré donc l'existence d'un estimateur basé sur des ondelettes qui converge plus vite. Par contre, celui-ci n'est pas adaptatif. De plus, nous avons remarqué empiriquement que l'algorithme nécessite un nombre minimal de points dans chaque sous-intervalle pour pouvoir appliquer correctement la régression polynomiale. Nous avons trouvé des conditions asymptotiques sur le nombre N de sous-intervalles en fonction du nombre n d'observations, qui nous ont confirmé le fait que le rapport n/N doit être grand. Dans l'implémentation algorithmique nous avons alors remplacé la régression polynomiale locale par la régression "ridge", qui est plus stable. Nous avons étudié empiriquement d'autres modifications susceptibles d'améliorer les performances de l'algorithme pour des échantillons de taille "raisonnable". Un exemple de débruitage est montré dans la figure 3.3.

L'avantage des méthodes basées sur l'interpolation ou la moyennisation réside dans le fait que l'on se ramène aux algorithmes d'ondelettes classiques et donc à des outils

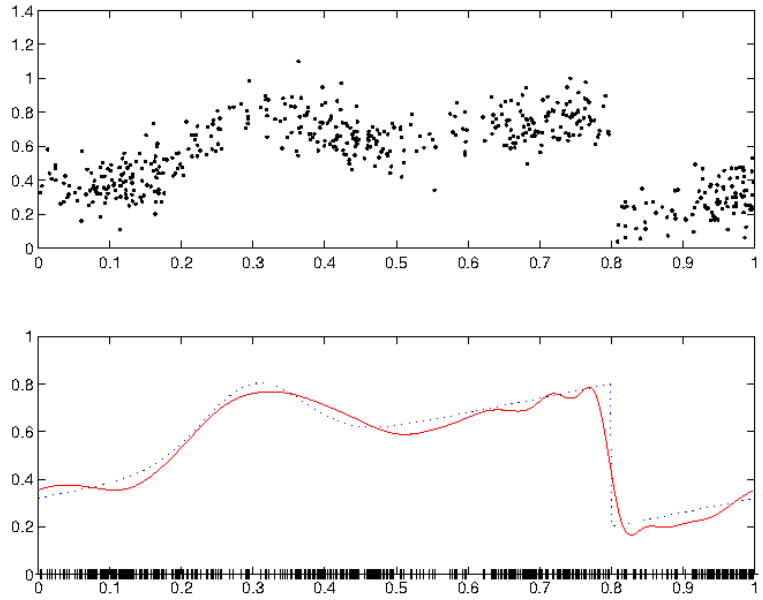
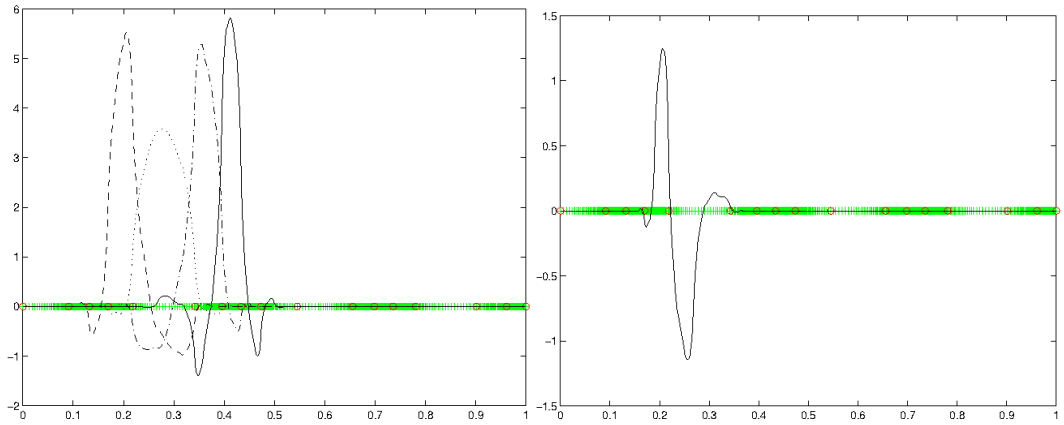


FIGURE 3.3 – Estimation de la fonction Blip à partir d'un échantillon bruité de 512 points (représentés sur la première ligne par des points), utilisant la régression "ridge" pour interpoler les valeurs de la fonction inconnue sur une grille équidistante avant d'appliquer un algorithme de débruitage par ondelettes standard. Le rapport signal-sur-bruit est ici SNR=2.

théoriques bien établis. Cependant, le principe n'est pas sans inconvénients. Une alternative sont les bases d'ondelettes adaptées au plan d'expérience. On est alors obligé d'étendre la définition des ondelettes et des analyses multirésolution (AMR), leur support théorique. En bref, une analyse multirésolution (AMR) est une décomposition de l'espace L_2 dans une suite de sous-espaces emboîtés (sous-espaces d'approximation). Si l'on renonce à la condition qui stipule que les ondelettes soient engendrées par une seule fonction, par translations et dilatations, on peut construire des AMR sur des intervalles compacts, ou adaptées aux échantillonnages non équidistants. Ce sont les AMR (et les ondelettes) dites de "seconde génération". Leur construction et la preuve des résultats théoriques sont néanmoins plus difficiles car la transformée de Fourier, systématiquement utilisée dans les preuves du cas classique, n'est plus adéquate.

Les fonctions de base de l'AMR sont construites par un algorithme itératif appelé schéma de subdivision. Les algorithmes de subdivision représentent, depuis longtemps, une méthode rapide et efficace pour construire des courbes et surfaces. Un tel algorithme reçoit en entrée un ensemble de points et des valeurs associées. A chaque étape, il insère des points entre les points de l'étape précédente et il associe à chaque point du nouvel ensemble une valeur qui est une combinaison affine des valeurs en les points voisins, de l'étape précédente.

Quand les coefficients de la combinaison affine changent avec le point (schéma non uniforme) et avec l'étape (schéma non stationnaire), il est assez difficile de décider si le schéma est convergent ou non, et de calculer la régularité des fonctions limites. Si la grille est du type $x_{j,k} = G^{-1}(2^{-j}k)$, la fonction G satisfaisant certaines propriétés relativement peu



(a) fonctions d'échelle (b) ondelette

FIGURE 3.4 – Système biorthogonal d'ondelettes construit à partir d'une analyse multi-résolution adaptée à la grille d'échantillonnage. Exemples de fonctions d'échelle en (a) et d'une fonction ondelette en (b).

contraignantes, nous montrons que les schémas de subdivision interpolateur de Lagrange et interpolateur en moyenne définis sur cette grille sont convergents. La démonstration se fait par comparaison du schéma non régulier avec le schéma régulier de même degré. Nous étudions les propriétés des fonctions fondamentales qui s'en déduisent ainsi que l'ordre d'approximation de ces schémas. Une approche un peu plus générale, applicable éventuellement à d'autres schémas de subdivision a fait ensuite l'objet d'une publication.

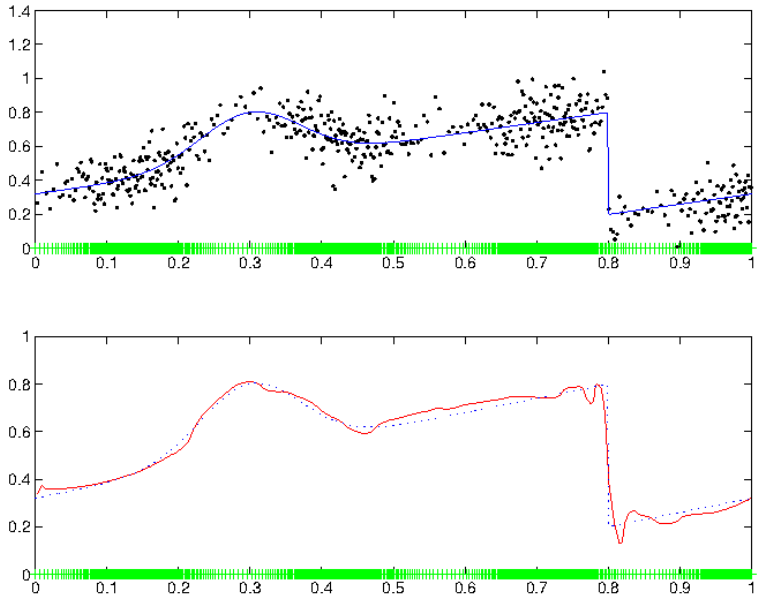


FIGURE 3.5 – Estimation de la fonction Blip à partir d'un échantillon bruité de 512 points (représentés sur la première ligne par des points), utilisant le seuillage dans une base d'ondelettes adaptées à la grille. Le rapport signal-sur-bruit est ici SNR=2.

Ces résultats ont été présentés lors d'une communication à la Semaine "Modélisation Géométrique et Approximation", Novembre 2001. Ils font également l'objet d'un article publié dans la revue Numerical Algorithms [A2].

Chacun des sous-espaces d'une AMR est engendré par une base de fonctions d'échelle. Nous montrons que les fonctions fondamentales du schéma interpolateur en moyenne, pondérées en accord avec la grille, sont les fonctions d'échelle d'une AMR. Pour cela, nous montrons la stabilité uniforme des bases à chaque niveau d'approximation. Nous construisons un système biorthogonal, formé des fonctions d'échelle primales et duales et des "ondelettes" primales et duales (figure3.4). La stabilité multi-échelle de ces bases, qui manque pour pouvoir réellement considérer ces fonctions ondelettes, reste un problème ouvert. Du point de vue numérique, les résultats concernant la reconstruction d'une fonction à partir des valeurs bruitées sont encourageants(voir figure 3.5).

3.2 Stage post-doctoral

D'octobre 2003 jusqu'en août 2004 j'ai effectué un stage post-doctoral dans l'équipe Brain Mapping Unit du Département de Psychiatrie de l'Université de Cambridge. J'ai travaillé dans le projet "Wavelets, fractals and fMRI of brain adaptivity", projet financé par le NIH. Mon travail était supervisé par le professeur Ed Bullmore (PUPH).

L'imagerie par résonance magnétique (IRM) fonctionnelle est une technique non invasive qui permet d'étudier la réponse du cerveau à des stimuli. L'activité neuronale a tendance à augmenter le flux sanguin dans les vaisseaux de la région corticale activée. Ceci se traduit par une augmentation du signal acquis par résonance magnétique. On peut ainsi détecter les régions du cerveau activées pendant l'accomplissement d'une certaine tache, l'influence d'une substance administrée au sujet, ou encore la connectivité entre diverses régions du cerveau. Les données se présentent sous la forme de séries temporelles d'images 3D. L'unité de volume est appelée voxel. La dimension spatiale peut être par exemple 64x64x16 et on analyse des séries temporelles de longueur de l'ordre 100-512 observations.

Dans la pratique, les difficultés que l'on rencontre dans l'analyse des données sont nombreuses. Le signal induit par l'activation est contaminé par du bruit provenant de différentes sources, avec un rapport signal-sur-bruit faible. Une première source de bruit est le mouvement du sujet durant l'acquisition des images. Les logiciels de correction de mouvement opèrent un réalignement des images, mais ils ne prennent pas en compte les différences de champ magnétique entre la position repère et la position réelle. Ceci fait que l'on retrouve dans la série temporelle uni-dimensionnelle de chaque voxel une tendance globale similaire aux séries temporelles qui enregistrent les translations et rotations du centre de gravité du cerveau. Cette dépendance peut être expliquée au niveau de chaque voxel par un modèle linéaire. Néanmoins, la modélisation n'est pas toujours parfaite (une des raisons étant que tous les voxels ne suivent pas la trajectoire du centre de gravité du cerveau) et on perd l'information spatiale.

La corrélation du bruit varie selon les voxels, indiquant par endroits de l'anti-persistance, du bruit faiblement corrélé, ou du bruit fortement corrélé. Le modèle que nous avons choisi est celui du bruit gaussien fractionnaire [A3], qui est un processus stationnaire de moyenne

nulle, caractérisé par deux paramètres : l'exposant de Hurst $H \in]0,1[$ et la variance σ^2. D'une part, nous avons cherché à savoir quelle est la distribution de ces deux paramètres dans le cerveau. D'autre part, nous nous sommes aussi intéressés à l'estimation du signal contaminé avec un bruit gaussien fractionnaire. Les filtres d'ondelettes orthogonales, même à support compact, permettent une très bonne décomposition en bandes de fréquence et décorrèlent asymptotiquement le signal. Cependant, les algorithmes d'ondelettes que nous avons trouvés dans la littérature souffrent d'au moins l'un des désavantages suivants :

— Ils approchent le spectre du signal par une formule du type

$$S(f) = constante(H, \sigma) * f^{1-2H}.$$

Cette formule est valable pour le processus continu ou encore dans un voisinage de l'origine pour les processus discrets, mais elle donne une approximation de faible qualité quand on s'éloigne de l'origine.

— Souvent l'estimation des deux paramètres se fait par moindres carrés à partir de l'approximation

$$\log S(f) \approx (1-2H)\log f + \log constante(H, \sigma).$$

L'erreur sur l'estimation de H induit une erreur exponentielle dans l'estimation de σ.

La taille réduite de nos échantillons nous empêchant d'utiliser uniquement les fréquences basses, nous avons construit un algorithme basé sur la formule exacte du spectre du bruit gaussien fractionnaire avec une résolution numérique de l'équation du maximum de vraisemblance. Il suffit alors de choisir une base d'ondelettes avec suffisamment de moments nuls et de ne pas prendre en compte les coefficients d'approximation et les coefficients des ondelettes proches des bords pour obtenir un estimateur robuste aux tendances de type polynomial.

Dans [A5] nous avons montré les résultats d'une étude comparative entre divers algorithmes d'estimation, en mettant en évidence leurs qualités et défauts vis-à-vis du problème à étudier. Les simulations que nous avons effectuées indiquent un comportement de l'algorithme basé sur le maximum de vraisemblance dans le domaine des ondelettes, comparable avec des algorithmes classiques en ce qui concerne l'estimation de H, mais il donne des résultats meilleurs en ce qui concerne l'estimation du paramètre σ.

Nous avons utilisé cet algorithme pour cartographier la corrélation temporelle du bruit dans le cerveau au repos (voir figure 3.6), et pour comparer un groupe de patients souffrant de la maladie d'Alzheimer à un groupe de volontaires sains. Dans le cerveau sain, nous avons observé de la corrélation négative dans les ventricules (due aux pulsations du liquide céphalo-rachidien) et de la longue mémoire surtout dans le cortex occipital. Cette observation est cohérente avec la conclusion de certaines études électro-physiologiques qui ont montré l'existence des oscillations structurées dans l'activité spontanée du cerveau au repos. En ce qui concerne la comparaison des groupes malades/volontaires sains, nous avons pu mettre en évidence des différences significatives dans des régions qui sont connues pour être affectées par la maladie d'Alzheimer. Pour le groupe de sujets malades nous avons observé des exposants de Hurst plus grands, probablement dus à la destruction

d'une partie des neurones ou des axones, impliquant une réduction de la dynamique du cerveau.

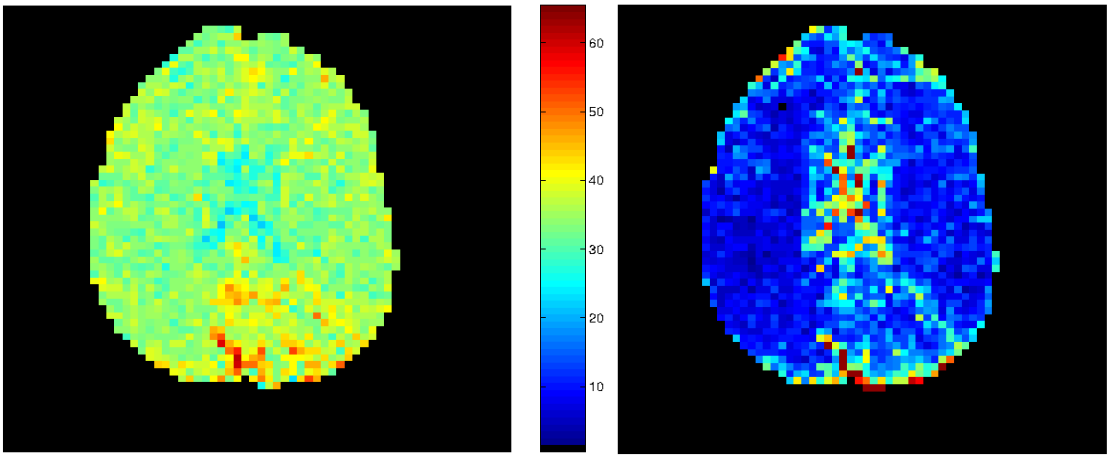


FIGURE 3.6 – Cartes de H et σ.

De la bonne estimation de la variance du bruit dépend la détection correcte des zones activées. Le niveau d'activation dans chaque voxel est calculé en utilisant des algorithmes paramétriques. Plus précisément, il s'agit de modèles linéaires, avec un bruit corrélé. Une bonne estimation de la variance du bruit est nécessaire en vue de la normalisation des cartes d'activation, avant d'appliquer les tests d'hypothèses. La corrélation du bruit fait que la variance est typiquement sous-estimée, ce qui induit une erreur de première espèce grande. L'algorithme que nous avons étudié est basé sur un Modèle Linéaire Généralisé, dans le domaine des ondelettes. Il estime simultanément les paramètres du bruit et le signal, par maximum de vraisemblance. Avec la variance estimée par cet algorithme, l'erreur de première espèce est proche du niveau de signification fixé, et ceci quelle que soit la valeur de H.

Les résultats ont été exposés en juin 2004 à la conférence OHBM de Budapest, Hongrie, et font l'objet d'un papier paru dans NeuroImage [A5].

Le grand nombre de tests à effectuer dans chaque expérience (généralement un test pour chaque voxel), induit une perte de puissance du test. La correction de Bonferroni assure le contrôle de l'erreur de premier type (FWER – Familywise Error Rate), mais elle est trop conservative. Nous avons utilisé la méthode FDR [3] (False Discovery Rate). Cette méthode contrôle l'espérance du taux de fausses positives parmi les hypothèses rejetées. Si toutes les hypothèses nulles sont vraies, elle contrôle le FWER ; on dit alors que le fait de contrôler le FDR implique le contrôle du FWER au sens faible. Par contre, comparé à la méthode de Bonferroni, elle se montre d'autant plus puissante que le nombre d'hypothèses alternatives est grand dans l'ensemble d'hypothèses à tester. D'où l'intérêt de l'utiliser dans des études ou l'on sait à l'avance que l'hypothèse nulle globale est fausse et l'on cherche à déterminer les points où l'alternative est vraie.

3. Y. Benjamini, Y. Hochberg. Controlling the False Discovery Rate : a practical and powerful approach to multiple testing. J. R. Statist. Soc. B, 1995.

En IRM fonctionnelle l'intérêt se porte en ce moment vers des méthodes capables de réduire le nombre d'hypothèses à tester et qui prennent en compte la dépendance spatiale entre les voxels voisins. Là encore les ondelettes se montrent utiles, car elles permettent de résumer l'information importante à un nombre réduit de coefficients significatifs. De plus, elles décorrèlent le bruit. En appliquant la méthode FDR (False Discovery Rate) à ce nombre réduit d'hypothèses on obtient un test bien plus puissant que les tests généralement utilisés, comme il a été montré [4] récemment. Dans ce papier, le choix du nombre des coefficients à garder se fait d'une manière très coûteuse. Pour cela nous avons étudié la possibilité d'utiliser des algorithmes de seuillage des coefficients d'ondelettes.

Les résultats ont été présentés en juin 2004 à la conférence OHBM de Budapest, Hongrie, et font l'objet d'un papier dans IEEE Transactions in Signal Processing [A4].

3.3 Activités exercées à l'INSA de Lyon

Au début des années 2000 était prévu la construction du centre Étoile de traitement du cancer par irradiation avec des faisceaux d'hadrons (ions lourds de type ^{12}C mais aussi protons) à Lyon. Le centre devait intégrer des facilités pour la recherche en hadronthérapie. Ce grand projet était principalement porté par des chercheurs de l'Institut de Physique Nucléaire de Lyon (IPNL) et fédérait nombre de scientifiques ainsi que des médecins de la région. Le centre Étoile n'a jamais vu le jour, mais le projet Étoile a suscité et facilité des activités autour de cette thématique. En parallèle des développements instrumentaux menés à l'IPNL et des recherches en radio-biologie, le laboratoire CREATIS a souhaité développer une thématique liée à l'imagerie pour l'hadron-thérapie. C'est ainsi qu'avec Rémy Prost (professeur INSA) et grâce à un financement doctoral (thèse de Mirela Frandes 2007-2010) nous avons démarré cette activité au sein du laboratoire. Les collègues de l'IPNL avaient identifié le potentiel de l'imagerie γ-prompt pour la vérification du traitement et démarraient l'investigation des caméras collimatées. Nous avons choisi la piste de l'imagerie 3D et nous avons découvert les travaux menés en astronomie sur l'imagerie des γ de moyenne énergie avec les travaux d'Andreas Zoglauer et sa thèse effectuée à Munich (Allemagne). Le logiciel qu'il avait développé, MEGAlib, ainsi que son expertise nous ont été d'une grande aide. Nous nous sommes vite aperçus qu'il y avait un vaste domaine de recherche encore très peu exploré lié à la reconstruction tomographique des images caméra Compton. Ce domaine correspondait bien aux thématiques du laboratoire et pouvait donner lieu à une collaboration avec les collègues physiciens. Le sujet n'était en rien lié à mes thèmes de recherche antérieurs, cependant il faisait appel à des compétences en analyse mathématique, géométrie, statistiques ainsi qu'en programmation.

Depuis, mes travaux s'articulent autour de la tomographie et du contrôle en ligne de la thérapie des tumeurs par faisceau de protons/ions. Dans la thèse de Mirela Frandes, co-encadrée avec R. Prost (actuellement professeur émérite INSA) nous avons été parmi

4. X. Shen, H.-C. Huang et N. Cressie. Nonparametric hypothesis testing for a spatial signal. Journal of the American Statistical Association, 2002

les premiers à proposer, lors de la conférence NSS-MIC de 2008 [C2]) puis dans un papier IEEE Trans. Nucl. Sci. de 2010 [C7]), l'utilisation de la caméra Compton pour l'imagerie SPECT des photons γ-prompt produits par le faisceau incident sur les tissus.

Nous nous sommes ensuite focalisés sur le problème de la reconstruction tomographique associé à ce dispositif d'imagerie, problème assez peu abordé dans la littérature et qui s'avère très mal conditionné pour les géométries de détection envisagées. En particulier, les méthodes analytiques étaient assez peu développées et nous avons proposé un algorithme original de type rétro-projection filtrée (Inverse Problems 2009 [A6]), IEEE TIP 2014 [A10]). Depuis, cette thématique connaît un certain succès dans la communauté des mathématiciens travaillant sur l'inversion des transformées de type Radon, car les données caméra Compton sont des projections sur des *surfaces* coniques. J'ai été invitée pour faire un exposé dans la conférence organisée en 2017 à Linz, Autriche, en honneur des 100 ans depuis la publication de l'article de Radon sur la reconstruction des fonctions à partir de leurs projections sur des lignes, puis à la conférence IPMS à Malte en 2018.

Les méthodes itératives (List-Mode MLEM) donnent des meilleurs résultats lorsque le problème direct est mal conditionné comme c'est notre cas. Nous avons commencé à les étudier dans le cadre du projet européen ENVISION finançant la thèse de Xavier Lojacono (co-dirigée par R. Prost). La principale difficulté dans ce cas consiste en l'estimation des paramètres de la reconstruction, qui sont la matrice de transfert du système et la sensibilité (IEEE TNS 2013 [A9], PMB 2016 [A12]).

Nous avons bénéficié de la proximité et des discussions avec des chercheurs de l'Institut de Physique Nucléaire de Lyon, UMR 5822, spécialistes en instrumentation et qui développent une caméra Compton (IEEE TNS 2012 et 2013 [A9], NIMA 2015 [A11], PMB 2016 [A12]), ainsi que du support d'un spécialiste des caméras Compton pour l'astronomie, Andreas Zoglauer travaillant au Space Sciences Laboratory, Berkeley, auteur du logiciel MEGAlib que nous utilisons pour la plupart des simulations (IEEE TNS 2010 [A7], conférences IEEE NSS-MIC 2008 [C2] et Fully 3D 2011 [C3]).

Le spectre énergétique des photons γ-prompt est très large et un dispositif d'imagerie doit convenir pour des énergies allant d'environ 1 MeV jusqu'à 8-10 MeV. Cette contrainte représente un défi pour la construction d'un détecteur, mais aussi pour l'exploitation des données afin de réaliser l'image. Dans la thèse d'Estelle Hilaire (co-dirigée par Françoise Peyrin, DR INSERM), nous avons été les premiers a commencer une étude de la distribution 3D par énergie visant à évaluer le potentiel de chaque image obtenue pour détecter une discordance avec le plan de traitement (PMB 2016 [A13]).

Les résultats que nous obtenons en terme de qualité d'images reconstruites étant assez mitigés, nous nous sommes orientés vers des modèles incluant de l'a priori et des algorithmes itératifs adaptés. Ceci est le sujet de la thèse de Yuemeng Feng qui a démarré en 2016. Avec Elie Bretin (MdC INSA Lyon, Institut Camille Jordan), nous avons trouvé la façon d'inclure un a priori de type variation totale (TV) dans un algorithme maximum-a-posteriori (MAP) avec maximisation de l'espérance (EM), qui permet de minimiser la log-vraisemblance Poissonienne sous contrainte TV. Nous avons déterminé un algorithme dual dont nous avons montré la convergence vers un minimum de la fonction coût. Dans

un article soumis nous comparons les performances de cette méthode aux résultats publiés dans la littérature et nous montrons que le fait de s'appuyer sur une approche de type EM et l'utilisation de la dualité lui permettent d'être plus rapide que les algorithmes d'optimisation habituels (de type primal-dual). Cependant l'application à la caméra Compton et au mode liste (LM-MLEM) présente quelques difficultés concernant les choix du paramètre de régularisation et du pas de l'algorithme dual. Des premiers résultats ont été présentés à la conférence NSS-MIC à Sydney en novembre 2018 [C8].

Nous avons démarré en 2014 une collaboration avec des chercheurs du laboratoire MATEIS (MATEriaux : Ingénierie et Science, UMR 5510) de l'INSA de Lyon travaillant dans la microscopie électronique en transmission. L'objectif que nous poursuivons consiste à réaliser la tomographie rapide en mode environnemental des échantillons en évolution ou sensibles sous faisceau électronique (Microsc. Microanal. 2016 [A14], conférence [D2]). Les images acquises - projections de l'image de l'échantillon - sont souvent floues, toujours mal alignées, et au cours de l'acquisition l'échantillon sort du champ de vue, étant emporté par le mouvement de rotation imposé au porte-échantillon. Nous avons étudié des pistes pour la correction de flou (stage de master de Yuemeng Feng, résultats présentés au Congrès Européen de Microscopie 2016 [D3]) et pour le réalignement de la séquence tomographique (stage PFE de Khanh Tran). Les deux stages ont été financés par le projet BQR INSA SPEE3D. Les travaux sur l'alignement sont poursuivis actuellement. Une présentation en a été faite au 19ème congrès mondial de microscopie (IMC 2018) [D6].

Un code pour la reconstruction tomographique itérative avec régularisation TV a été implémenté par Hussein Banjak, post-doctorant recruté dans le cadre du projet ANR 3DCLEAN. Ces travaux ont donné lieu à une publication dans Ultramicroscopy en 2018 [A15] et le code commence à être utilisé par MATEIS et pourrait être installé également au laboratoire IPCMS de Strasbourg, autre partenaire de l'ANR 3DCLEAN. Nous avons testé des méthodes pour diminuer l'artefact d'angle manquant dans le cadre du projet de fin d'études de Benjamin Ganguet (Phelma, 2018). Les résultats ont été assez décevants, excepté pour la régularisation TV, que ce soit employée avec hypothèse de bruit gaussien ou Poissonien. Ce stage a été l'occasion d'échanges avec le laboratoire IPCMS de Strasbourg. À présent, nous démarrons l'étude des méthodes de type apprentissage profond qui pourraient encore améliorer la reconstruction tomographique pour des données acquises avec un angle de vue très limité. Les premiers résultats ont été présentés au 19ème congrès mondial de microscopie (IMC 2018) [D7].

Chapitre 4

Production scientifique

4.1 Publications et exposés

Publications : Phys. Med. Biol. (deux), Inverse Problems (deux), IEEE TIP (une), IEEE TSP (une), IEEE TNS (trois), NeuroImage (deux), Human Brain Mapping (une), Ultramicroscopy (une), Numerical Algorithms (deux), etc.

- **Preprint**

{$U2$} **V. Maxim**, Y. Feng, H. Banjak, E. Bretin, "Tomographic reconstruction from Poisson distributed data: a fast and convergent EM-TV dual approach", *submitted* , (2018).

{$U1$} M. Haltmeier, **V. Maxim**, "Inversion of the conical Radon transform and Compton camera imaging", *preprint* , (2018).

- **Articles dans revues internationales à comité de lecture :**

{$A16$} **V. Maxim**, "Enhancement of Compton camera images reconstructed by inversion of a conical Radon transform", *Inverse Problems* **35**, no. 1, pages 1–21 (2019).

{$A15$} H. Banjak, T. Grenier, T. Epicier, S. Koneti, L. Roiban, A-S. Gay, I. Magnin, F. Peyrin, **V. Maxim**, "Evaluation of noise and blur effects with SIRT-FISTA-TV reconstruction algorithm: Application to fast environmental transmission electron tomography", *Ultramicroscopy* **189**, pages 109–123 (2018).

{$A14$} L. Roiban, S. Koneti, K. Tran, Y. Feng, T. Grenier, **V. Maxim** and T. Epicier, "Rapid Tomography in Environmental TEM: How Fast Can We Go to Follow the 3D Evolution of Nanomaterials in situ?", *Microsc. Microanal.* **22**, 5 (2016).

{$A13$} E. Hilaire, D. Sarrut, F. Peyrin and **V. Maxim**, "Proton therapy monitoring by Compton imaging: influence of the large energy spectrum of the prompt-γ radiation", *Phys. Med. Biol.* **61**, no. 8, pages 3127–3147 (2016).

{A12} **V. Maxim**, X. Lojacono, E. Hilaire, J. Krimmer, E. Testa, D. Dauvergne, I. Magnin and R. Prost, "Probabilistic models and numerical calculation of the system and sensitivity matrices in list-mode MLEM 3D reconstruction of Compton camera images", *Phys. Med. Biol.* **61**, no. 1, pages 243–264 (2016).

{A11} J. Krimmer, J.-L. Ley, C. Abellan, J.-P. Cachemiche, L. Caponetto, X. Chen, M. Dahoumane, D. Dauvergne, N. Freud, B. Joly, D. Lambert, L. Lestand, J.M. Létang, M. Magne, H. Mathez, **V. Maxim**, G. Montarou, C. Morel, M. Pinto, C. Ray, V. Reithinger, E. Testa and Y. Zoccarato, "Development of a Compton camera for medical applications based on silicon strip and scintillation detectors", *Nuclear Instruments and Methods in Physics Research Section A* **787**, 98–101 (2015).

{A10} **V. Maxim**, "Filtered Backprojection Reconstruction and Redundancy in Compton Camera Imaging", *IEEE Transactions on Image Processing* **23**, no. 1, pages 332–341 (2014).

{A9} X. Lojacono, M.-H. Richard, J.-L. Ley, E. Testa, C. Ray, N. Freud, J. M. Létang, D. Dauvergne, **V. Maxim** and R. Prost, "Low Statistics Reconstruction of the Compton Camera Point Spread Function in 3D Prompt-*gamma* Imaging of Ion Beam Therapy", *IEEE Trans. Nucl. Sci.* **60**, no 5, pages 3355–3363 (2013).

{A8} M-H. Richard, M. Dahoumane, D. Dauvergne, M. De Rydt, G. Dedes, N. Freud, J. Krimmer, J.M. Létang, X. Lojacono, **V. Maxim**, et al, "Design study of the absorber detector of a Compton camera for on-line control in ion beam therapy", *IEEE Trans. Nucl. Sci.* **59**, no 5, pages 1850–1855 (2012).

{A7} M. Frandes, A. Zoglauer, **V. Maxim**, R. Prost, "A tracking Compton-scattering imaging system for hadron therapy monitoring", *IEEE Transactions on Nuclear Science* **57**, Issue 1, pages 144–150 (2010).

{A6} **V. Maxim**, M. Frandes, R. Prost, "Analytical inversion of the Compton transform using the full set of available projections", *Inverse Problems* **25**, no 9, pages 1–21 (2009).

{A5} **V. Maxim**, L. Şendur, J. Fadili, J. Suckling, R. Gould, R. Howard, E. Bullmore, "Fractional Gaussian noise, functional MRI and Alzheimer's disease", *Neuroimage* **25**, no 1, pages 141–158 (2005).

{A4} L. Şendur, **V. Maxim**, B. Whitcher, E. Bullmore, "Multiple hypothesis mapping of functional MRI data in orthogonal and complex wavelet domains", *IEEE Transactions on Signal Processing* **53**, no 9, pages 3413–3426 (2005).

{A3} E. Bullmore, J. Fadili, **V. Maxim**, L. Şendur, B. Whitcher, J. Suckling, M. Brammer, M. Breakspear, "Wavelets and functional magnetic resonance imaging of the human brain", *Neuroimage* **23**, pages S234–S249 (2004).

{A2} **V. Maxim**, M-L. Mazure, "Subdivision schemes and irregular grids", *Numerical Algorithms* **35**, pages 1–28 (2004).

{A1} P-J. Laurent, M-L. Mazure, **V. Maxim**, "Chebyshev splines and shape parameters", *Numerical Algorithms* **15**, pages 373–383 (1997).

• Articles dans des revues d'audience nationale

{*AN*2} M. Fontana, J.-L. Ley, É Testa, J. M. Létang, **V. Maxim**, D. Dauvergne, "Versatile Compton Camera for High-energy Gamma Rays: Monte Carlo Comparison with Anger Camera for Medical Imaging", *Acta Physica Polonica B* **48**, no. 10 (2017).

{*AN*1} **V. Maxim**, "Redundancy and inversion of the Compton transform", *Annals of the University of Bucharest (mathematical series)* **4**, no. 1, pp. 141–154 (2013).

• Conférences internationales à comité de lecture et actes publiés :

{*C*8} Y. Feng, A. Etxebeste, J. M. Létang, D. Sarrut, **V. Maxim**, "Total variation regularization for list-mode MLEM reconstruction in Compton camera imaging", *IEEE NSS-MIC conference, Sydney, Australia* , (2018).

{*C*7} M. Fontana, D. Dauvergne, J. Krimmer, J. M. Létang, J.-L. Ley, **V. Maxim**, É Testa, "Comparison between Anger and Compton cameras for medical imaging : a Monte Carlo simulation study", *IEEE Nuclear Science Symposium and Medical Imaging Conference, Atlanta, USA, 21-28 October* , (2017).

{*C*6} E. Hilaire, D. Sarrut, F. Peyrin and **V. Maxim**, "Compton imaging of prompt-γ radiation in proton therapy : simulation of the influence of the detector response and of the patient attenuation", *IEEE NSS-MIC Conference, Seattle, USA* , (2014).

{*C*5} J. Krimmer, L. Balleyguier, J. Baudot, S. Brons, L. Caponetto, M. Chabot, X. Chen, M. Dahoumane, D. Dauvergne, M. De Rydt, et al., "Real-time online monitoring of the ion range by means of prompt secondary radiations", *3rd International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA)* , 1–8 (2013).

{*C*4} X. Lojacono, M.-H. Richard, C. Ray, D. Dauvergne, E. Testa, N. Freud, J. Létang, **V. Maxim**, R. Prost, "Image reconstruction for Compton camera applied to 3D prompt imaging during ion beam therapy", *IEEE Nuclear Science Symposium and Medical Imaging Conference, Valencia, Spain* , 3518 - 3521 (2011).

{*C*3} X. Lojacono, **V. Maxim**, R. Prost, F. Peyrin, A. Zoglauer, "A filtered backprojection reconstruction algorithm for Compton camera", *Fully 3D, Potsdam, Germany* , 1–4 (2011).

{*C*2} M. Frandes, A. Zoglauer, **V. Maxim**, R. Prost, "Compton-scattering based imaging system for hadron therapy monitoring", *IEEE NSS-MIC Conference Record* , pages 4853–4859 (2008).

{*C*1} **V. Maxim**, "Denoising signals observed on a random design", *Fifth AFA–SMAI Conference on Curves and Surfaces, Saint-Malo 2002, A. Cohen, J.L. Merrien, L. L. Schumaker eds.* , pages 303–312 (2002).

• Conférences nationales à comité de lecture et actes publiées :

{*CN*3} **V. Maxim**, "Redundancy and inversion of the Compton transform", *Annals of the University of Bucharest (mathematical series)* **4**, no. 1, pp. 141–154 (2013).

{$CN2$} E. Hilaire, C. Robert, X. Lojacono, C. Lartizien, I. Buvat and **V. Maxim**, "Imagerie Compton en protonthérapie : de la simulation GATE à la reconstruction d'images", *XXIVème Colloque GRETSI - Traitement du Signal et des Images (GRETSI'13), Brest, France* , 1–4 (2013).

{$CN1$} X. Lojacono, **V. Maxim**, R. Prost, "Calcul de la matrice des probabilités de transfert en imagerie Compton 3D", *XXIIIème Colloque GRETSI - Traitement du Signal et des Images (GRETSI'11), Brest, France* , 1–4 (2011).

- **Autres présentations récentes :**

{$P3$} **V. Maxim**, "Practical considerations for Compton camera analytic image reconstruction", *The 9th International Conference "Inverse Problems : Modeling and Simulation" (IPMS-2018) Malta, May 21 - 25* , (exposé invité) (2018).

{$P2$} **V. Maxim**, "Compton camera imaging and Radon transforms on the cone : From data to images", *100 Years of the Radon Transform, Linz, 27-31 March 2017* , (exposé invité) (2017).

{$P1$} **V. Maxim**, X. Lojacono, E. Hilaire, J. - L. Ley, D. Dauvergne, and E. Testa, "Data models for the Compton camera acquisition and their influence on the reconstructed image", *ICTR-PHE 2014, Geneva, Switzerland* , S66 (2014).

- **Résumés de conférence récents :**

{$D8$} A. Etxebeste, Y. Feng, J. M. Létang, **V. Maxim**, E. Testa, D. Sarrut, "An extension of the GATE Monte Carlo simulation toolkit to model Compton Camera systems", *The Third Geant4 International User Conference* , October 29-31, Bordeaux, France (2018).

{$D7$} H. Banjak, T. Grenier, T. Epicier, L. Roiban, **V. Maxim**, "Deep Neural Network for Iterative Image Reconstruction with Application to Fast Environmental Transmission Electron Tomography", *19th International Microscopy Congress (IMC19)* , Sydney, Australia, 9 - 14 September (2018).

{$D6$} T. Grenier, L. Roiban, H. Banjak, S. Koneti, **V. Maxim**, T. Epicier, "A robust method to acquire tilt series in a few seconds for Fast Operando Nano-Tomography in ETEM", *19th International Microscopy Congress (IMC19)* , Sydney, Australia, 9 - 14 September (2018).

{$D5$} T. Epicier, H. Banjak, A.-S. Gay, T. Grenier, S. Koneti, **V. Maxim**, L. Roiban, "Very Fast Tomography in the (E)TEM to Probe Dynamics in Materials during Operando and In Situ Experiments", *Invited conference at M&M* , August 5-9, Baltimore USA (2018).

{$D4$} H. Banjak, T. Grenier, **V. Maxim**, S. Koneti, L. Roiban, T. Epicier, "Towards one hertz electron tomography of dynamic processes under environmental conditions: expectations and limitations due to blur effects", *3rd International Conference on*

Tomography of Materials and Structures (ICTMS) , Lund, Sweden, 26th – 30th June (2017).

{*D*3} Y. Feng, K. Tran, S. Koneti, L. Roiban, A.-S. Gay, C. Langlois, T. Epicier, T. Grenier, **V. Maxim**, "Image deconvolution for fast Tomography in Environmental Transmission Electron Microscopy", *The 16th European Microscopy Congress 2016* , 6973 (2016).

{*D*2} S. Koneti, L. Roiban, **V. Maxim**, T. Grenier, P. Avenier, A. Cabiac, A.-S. Gay, F. Dalmas and T. Epicier, "Environmental Transmission Electron Tomography: fast 3D analysis of nano-materials", *The 16th European Microscopy Congress 2016* , 6438 (2016).

{*D*1} E. Hilaire, C. Robert, X. Lojacono, D. Sarrut, I. Buvat, F. Peyrin and **V. Maxim**, "Compton imaging in proton therapy: reconstructed image of the simulated prompt-γ distribution", *ICTR-PHE 2014, Geneva, Switzerland* , S43 (2014).

4.2 Logiciel

Le logiciel CORESI a été crée pour la reconstruction tomographique analytique et itérative LM-MLEM en imagerie caméra Compton. Il est utilisé par l'Institut de Physique Nucléaire de Lyon dans le cadre du développement d'une caméra Compton pour l'imagerie SPECT et dans l'équipe en complément de développement d'un module de simulation de caméra Compton dans le logiciel GATE.

Un logiciel de reconstruction tomographique pour la microscopie électronique a été réalisé dans le cadre du post-doctorat de Hussein Banjak, financé par le projet ANR 3DCLEAN. Ce logiciel est utilisé par les collègues du laboratoire Mateis de l'INSA de Lyon. Il fait l'objet d'un dépôt git en tant que logiciel libre.

Deuxième partie

Activités de recherche liées à la tomographie

Chapitre 5

L'imagerie caméra Compton et ses applications

Le contrôle de l'irradiation en hadron et proton thérapie nécessite de mesurer des particules secondaires émises pendant le traitement. Les premières approches ont été basées sur la mesure des photons gamma d'annihilation à l'aide de caméras TEP (Japon, Allemagne). Cette technique est toujours utilisée en hadron thérapie au Japon pour le contrôle a posteriori mais présente plusieurs inconvénients, dont la faible statistique, le wash-out physiologique qui modifie rapidement la distribution des fragments émetteurs produits par les collisions des ions avec des noyaux des tissus, les délais dus aux lois de décroissance radioactive, le bruit induit par les bien plus nombreux photons γ-prompt émis presque instantanément au cours de l'irradiation. La technologie de la TEP avec mesure du temps de vol pourrait améliorer les résultats mais le problème de la statistique, particulièrement sensible en proton thérapie, représente un verrou important ([49]). Plus récemment, l'imagerie SPECT des γ-prompts a été proposée. Nous en parlons dans la section §5.2.

La caméra Compton a été proposée pour imager des sources de rayonnement gamma dans diverses applications comme la médecine nucléaire, la sécurité intérieure et le démantèlement nucléaire. La caméra utilise la cinématique Compton pour reconstruire les chemins des rayons gamma qui interagissent dans ses détecteurs. Elle peut identifier et localiser des sources de rayonnement gamma ayant un spectre étendu, dans un champ de vision large, avec une efficacité et une spécificité améliorées par rapport aux dispositifs existants ([20, 95, 23]). Les avantages potentiels de cette technique d'imagerie, associés aux progrès technologiques, ont conduit au développement rapide d'un domaine de recherche visant l'optimisation des systèmes de détection d'une part et la reconstruction tomographique des images d'autre part. Dans la section §5.1 nous faisons un état de l'art sur la construction des caméras Compton et nous détaillerons les méthodes de reconstruction tomographique associées dans le chapitre 6.

Ces recherches reposent de manière significative sur les simulations Monte-Carlo. Au fil des différentes thèses que j'ai co-encadrées, nous avons développé des stratégies de

simulation pour pallier au manque d'outils adaptés pour l'imagerie médicale avec caméra Compton. Elles seront présentées dans la section §5.3.

5.1 Développement de la caméra Compton

Le fonctionnement de la caméra est basé sur la détection en coïncidence d'au moins deux interactions d'un photon γ dans des détecteurs. Les premières interactions sont des diffusions Compton sur des électrons d'un premier détecteur, électrons qui récupèrent une partie de l'énergie du photon. Dans les semi-conducteurs à pistes, des paires électron-trou sont formées et permettent de déterminer la position d'interaction ainsi que l'énergie déposée.

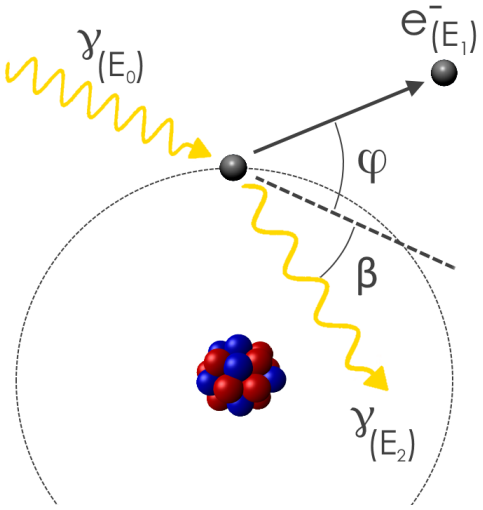


FIGURE 5.1 – Représentation schématique de la diffusion Compton d'un photon sur un électron.

Généralement on utilise aussi un deuxième détecteur ayant un haut potentiel d'absorption afin de déterminer l'énergie totale du photon γ. L'ensemble des données collectées lors de la détection des interactions d'un photon, en particulier des positions et des énergies, sera appelé ci-après évènement Compton. Connaissant l'énergie totale E_0 du photon et l'énergie E_1 déposée lors de la première interaction, on peut déterminer l'angle de diffusion d'après la formule

$$\cos\beta = 1 - m_e c^2 \frac{E_1}{E_0(E_0 - E_1)}, \tag{5.1}$$

où m_e est la masse de l'électron et c est la vitesse de la lumière. On peut ainsi associer à chaque évènement un cône Compton : il aura pour sommet le premier point d'interaction, l'axe dans la direction du photon diffusé et le demi-angle d'ouverture égal à l'angle de diffusion Compton. En faisant abstraction pour le moment des incertitudes, on peut dire que la source du photon se trouve sur la surface de ce cône.

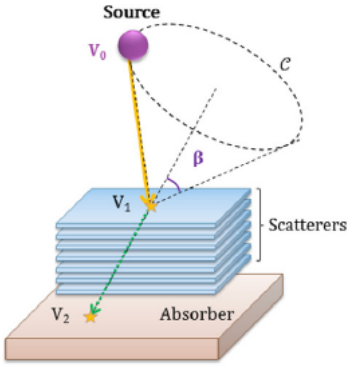


FIGURE 5.2 – Fonctionnement de la caméra Compton. À l'aide d'un algorithme de reconstruction tomographique on reconstruit l'image 3D de la source.

Les détecteurs ont fait des grands progrès depuis les premiers articles sur la caméra Compton, publiés au début des années 70 indépendamment par deux groupes, Schönfelder *et al* ([69]) et Todd *et al* ([88]) respectivement dans les domaines de l'astronomie et de la médecine nucléaire. Le grand champ de vision et le large spectre énergétique autorisé par la nouvelle technologie ont trouvé une première application réussie en astronomie avec COMPTEL (COMPton TELescope). COMPTEL a été embarqué en 1991 sur un satellite et est resté opérationnel pendant neuf ans. Le projet COMPTEL a été suivi par d'autres caméras embarquées sur des ballons stratosphériques. Certains exemples peuvent être trouvés à l'adresse http://andreaszoglauer.com/overview.html. Les premiers tests pour l'application de la médecine nucléaire ont été rapportés en 1983 par Singh et Doria ([76]).

A présent, les domaines d'application de la caméra incluent également la sécurité intérieure et le démantèlement de sites nucléaire. De petits prototypes ont été développés pour les applications de sécurité intérieure et de médecine nucléaire ([46, 82, 42]). Les développements instrumentaux ont été accompagnés par de développements logiciels dont le logiciel MEGAlib ([98]) basé sur la bibliothèque GEANT4 ([1]) et destiné aux simulations Monte-Carlo et à la reconstruction d'images. Ce logiciel a été développé en parallèle d'un prototype destiné à l'astronomie (MEGA) et est utilisé depuis par d'autres projets. Plus récemment, et favorisés par l'accident de Fukushima (Japon) initié par le tsunami lié au séisme du 11 mars 2011, des prototypes pour l'identification, la quantification et la localisation des sources de rayons gamma dans les centrales nucléaires et les déchets nucléaires ont été construits. On peut citer, par exemple, ASTROCAM7000HS de Mitsubishi Heavy Industries au Japon [83] et Polaris-H de H3Dgamma aux Etats-Unis, actuellement commercialisés. En France, au moins deux prototypes sont en construction, dans les projets ANR TEMPORAL ([27]) et ComptonCAM. L'imagerie tridimensionnelle n'est pas requise pour ces applications et des images bidimensionnelles sur des surfaces sphériques sont suffisantes. La position de la source peut alors être déterminée en déplaçant le détecteur portable. Les paramètres les plus importants d'un tel détecteur sont l'angle de vue, la résolution angulaire et le poids. Les détecteurs de petite taille et compacts sont préférables, bien qu'ils fournissent une résolution relativement médiocre pour l'identification des géométries complexes.

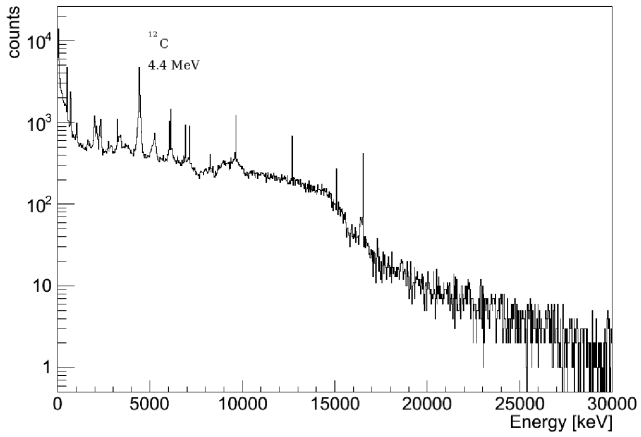


FIGURE 5.3 – Exemple de spectre γ-prompt en sortie d'un fantôme en PMMA irradié avec un faisceau de protons à 140 MeV. (source [A7])

La vérification de la profondeur du parcours des particules en proton et hadron thérapie est un défi majeur pour l'assurance qualité du traitement. Lors de l'irradiation des tissus par le faisceau ionique, des photons γ prompt sont émis quasi instantanément avec des énergies très diverses allant de quelques keV à plusieurs MeV. Le spectre, continu, possède quelques lignes proéminentes en dessous de 8 MeV (voir figure 5.3). Les gamma-caméras à fente sont relativement simples à produire et elles sont déjà commercialisées (entreprise IBA), mais souffrent d'une sensibilité faible. La caméra Compton présente l'avantage de pouvoir détecter des photons ayant des énergies dans une région de plusieurs MeV et de pouvoir produire des images 3D. Son utilisation pour la surveillance en protonthérapie est étudiée depuis dix ans [C2]. Les résultats d'un test avec un fantôme d'eau irradié avec un faisceau de protons et un petit prototype ont été publiés dans [40]. Des prototypes spécifiques sont en construction. Ils devraient être plus grands et plus complexes que ceux produits pour la localisation de sources de rayons γ dans les centrales nucléaires. Un petit prototype avec une surface d'environ 2×2 cm^2 a été testé récemment mais a donné des résultats mitigés concernant l'identification du fin de parcours du faisceau [85]. Cependant, de tels prototypes sont utiles pour explorer le comportement des détecteurs dans la gamme d'énergies pertinente pour l'imagerie γ prompt, pour mesurer les résolutions spatiale, énergétique et temporelle et pour évaluer l'influence du bruit ([26, 68]).

Pour les prototypes plus grands, différentes stratégies d'acquisition sont étudiées. En général, plusieurs diffuseurs sont montés pour améliorer l'efficacité de la diffusion aux hautes énergies (voir figure 5.2). Une bonne résolution énergétique des détecteurs est importante pour réduire autant que possible les erreurs sur la valeur estimée de l'angle de diffusion. Deux stratégies d'estimation peuvent être utilisées, avec deux ou trois sites d'interaction. Comme l'énergie totale des photons frappant le détecteur est inconnue, dans la configuration à deux sites l'absorption totale est nécessaire et l'absorbeur doit être choisi en conséquence. Dans la configuration à trois sites, l'énergie totale n'est pas nécessaire pour l'estimation de l'angle Compton, mais la résolution énergétique et l'ordonnancement des coups dans la séquence ont une influence plus grande sur le résultat. La première

configuration a été choisie à l'IPNL (voir [64]) et la seconde en Espagne (voir [54]). Les deux prototypes sont en construction et pour le moment seules certaines parties ont été testées.

Lorsque les diffuseurs sont suffisamment minces, l'électron extrait de l'atome avec une énergie suffisamment haute peut traverser les couches du diffuseur selon une trajectoire relativement linéaire. Avec cette information supplémentaire, la direction du photon incident peut être déterminée plus précisément et se limite idéalement à une ligne. De plus, en cas de création de paires, les chemins suivis par le positon et l'électron créés permettent de déterminer la ligne où le photon a été émis. Un prototype capable de suivre l'électron, appelé MEGA, a été construit pour l'imagerie en astrophysique ([97]). Une configuration similaire a été simulée pour la protonthérapie dans [A7] et un prototype est construit en Allemagne [2]. Une technologie différente pour le suivi de l'électron de recul consiste à utiliser des détecteurs à gaz ([90]).

Les isotopes actuellement employés en médecine nucléaire produisent des rayons γ dont l'énergie est inférieure à 200 keV. Ceci rend délicate l'utilisation de la caméra Compton puisqu'il a été montré qu'elle fonctionne bien au-dessus d'environ 300 keV. Une configuration en damier pouvant s'adapter mieux aux basses énergies a été proposée dans [59]. Les détecteurs pour énergies plus élevées peuvent être utiles pour de nouveaux isotopes ou pour les radionucléides utilisés dans la radiothérapie sélective interne. Les collimateurs étant adaptés à des énergies spécifiques, l'imagerie simultanée de plusieurs traceurs pourrait également constituer un marché pour les caméras Compton. De tels prototypes sont à l'étude et des essais sur petit animal ont été réalisés ([50, 84, 34]). Par exemple, dans [84] il est décrit un test avec un détecteur de petite taille qui a permis de distinguer l'accumulation de deux radio-isotopes dans une souris vivante. Dans [34], une image 3D d'une souris a été réalisée avec un dispositif ultra-compact mesurant $4.9 \times 5.6 \times 10.6$ cm^3 et pesant 580 g. L'image tridimensionnelle a été obtenue lors d'une acquisition à plusieurs angles de vue pour trois radio-isotopes : ^{131}I (364 keV), ^{85}Sr (415 keV) et ^{65}Zn (1116 keV).

Les premiers essais réussis d'acquisition caméra Compton ont donc été réalisés avec des prototypes de petite taille. Un avantage important qu'ils ont (mis à part le coût) est leur polyvalence et leur facilité d'utilisation. Ce sont souvent des adaptations de caméras développées pour d'autres applications, telles que le système SmartPET ([24]). L'efficacité des petits détecteurs est faible et nécessite de longues acquisitions. En outre, la résolution dans la direction orthogonale à la caméra est très médiocre, mais cela peut être résolu en déplaçant la caméra autour de l'objet. Une solution pour augmenter l'efficacité et la résolution spatiale consiste à disposer plusieurs caméras Compton dans un anneau. Il convient de mentionner que l'efficacité augmente plus que linéairement avec la surface de détection et avec le nombre de détecteurs. La construction de grandes caméras conçues pour des applications d'imagerie médicale prend plus de temps. La description d'un prototype non encore opérationnel peut être trouvée dans [38]. Une étude récente ([15]), comparant par simulation Monte-Carlo ce prototype à une caméra SPECT collimatée commerciale (la General Electric HealthCare Infinia), montre une efficacité accrue d'un

facteur supérieur à 10 et une résolution spatiale améliorée de la caméra Compton pour des énergies supérieures à 500 keV. Ces résultats font espérer à l'avenir une meilleure prise en charge des patients avec une dose et un temps d'acquisition réduits pour une qualité d'image comparable.

Un domaine de recherche très actif est celui des matériaux utilisés pour les détecteurs. Le diffuseur est souvent réalisé dans des semi-conducteurs, par exemple du silicium ([82, 84, 38, 2], ASTROCAM 7000HS), CdTe ou CdZnTe également appelé CZT ([26]). Le germanium est aussi un très bon candidat, ayant une meilleure résolution énergétique comparé au silicium lorsqu'il est refroidi. Ce dernier a par contre un élargissement Doppler plus faible à température ambiante. De nouveaux détecteurs en germanium avec refroidissement électronique ont été développés. Pour les absorbeurs, les cristaux scintillants sont préférés, comme par exemple le CsI ([82]), $Bi_4Ge_3O_{12}$ également appelé BGO ([38]), $LaBr_3$ ([2]), Lu_2SiO_5 également connu sous le nom de LSO ([26, 68]) mais les semi-conducteurs sont également employés (CdTe in [84] et ASTROCAM 7000HS). Certaines équipes préfèrent utiliser le même matériau à la fois pour le diffuseur et l'absorbeur, qui peut être un cristal scintillateur comme LYSO ou $CeBr_3$ ([27]), $Gd_3Al_2Ga_3O_{12}$ dopé Ce ([85, 34]), $LaBr_3$ ([54]), BGO ([25]) ou des semi-conducteurs comme le germanium ([24, 50]) et CZT (Caméra Polaris-H). Les détecteurs au gaz peuvent être utilisés pour la fabrication de diffuseurs. Certaines équipes étudient l'applicabilité des détecteurs au gaz pour l'astronomie ou le médical (par exemple [31] et autres publications de la même équipe). Malgré leur faible efficacité comparé aux semi-conducteurs, ils permettent de suivre l'électron de recul ce qui limite la reconstruction à une petite section de cône.

Les diffuseurs du prototype de caméra Compton à l'IPN de Lyon ([38]) sont en silicium et quatre matériaux avaient été considérés pour l'absorbeur : le germanate de bismuth (BGO), l'oxyorthosilicate de lutetium-yttrium (LYSO), l'iodure de sodium (NaI) et le bromure de lanthane ($LaBr_3$). Le LYSO et le BGO semblent les plus appropriés, en raison de leur grande section efficace photoélectrique [64], et le choix s'est finalement porté sur un absorbeur composé de cristaux de BGO.

5.2 Application à l'imagerie gamma-prompt en proton-thérapie

Dans un contexte où la protonthérapie et l'hadronthérapie se développent en Europe, le souci de sa justification médicale et celui de la qualité du traitement sont très présents. Les premiers travaux menés en Allemagne visent à adapter l'imagerie TEP et à imager les fragments d'atomes émetteurs β^+. Ce type d'imagerie post-traitement se pratiquait déjà au Japon où l'hadronthérapie s'est initialement développée (après sa découverte aux États-Unis). Elle nécessite que le patient soit sorti de la salle de traitement et conduit dans une salle d'imagerie. Cependant une part des photons produits par certains radio-isotopes de demi-vie très courte sont ainsi perdus et la distribution spatiale des autres radio-isotopes est modifiée par la circulation sanguine (wash-out physiologique). Comme en plus la durée

de l'acquisition ne peut pas être trop longue pour des raisons évidentes de confort du patient, le nombre de paires acquises est assez faible et la qualité de l'image insuffisante pour une évaluation fine de la zone endommagée. De plus, le résultat ne peut pas aider à prévenir les erreurs. Pour y remédier, au début des années 2000 l'équipe de W. Enghardt à Dresde a commencé à travailler sur le développement d'une caméra TEP en-ligne pour le nouveau centre DSI à Darmstadt. Certains succès ont été obtenus à commencer par ceux rapportés dans la thèse de K. Parodi (voir par exemple [57]), mais il a été observé que le signal TEP était noyé dans un signal γ-prompt bien plus fort. La TEP en-ligne semblait possible pour la thérapie par ions carbone mais pas pour la protonthérapie. A l'Institut de Physique Nucléaire de Lyon, qui est à l'origine du groupement de recherche Étoile ayant œuvré pour la construction, maintenant abandonnée, du premier centre de thérapie par ions carbone en France à Lyon, on commence alors à s'intéresser à l'imagerie γ-prompt à l'aide d'une caméra collimatée. En parallèle, dans sa thèse débutée en 2007, M. Frandes a proposé d'imager les photons γ-prompt avec une caméra Compton. Dans ses articles [C2] et [A7] elle montre des premiers résultats issus de la simulation Monte-Carlo d'un prototype adapté d'après le prototype MEGA construit à Munich en Allemagne pour l'astronomie. Ce prototype est assez complexe car constitué d'un nombre important de diffuseurs en silicium très fin. Il est capable de détecter les photons par diffusion Compton ou par création de paires. Aussi, à haute énergie il est possible de suivre la trace de l'électron de recul ce qui réduit l'incertitude sur la position de la source du photon incident à une ligne, au lieu d'un cône. Un algorithme assez sophistiqué est alors nécessaire pour traiter les données, disponible dans la suite logicielle MEGAlib que nous avons utilisé.

La particularité de la thérapie par ions est de pouvoir être très précise pour viser la tumeur et de permettre d'épargner les tissus situés derrière le pic de Bragg - région très réduite où un faisceau mono-énergétique dépose un maximum d'énergie. Cependant, pour bénéficier de cette caractéristique, il faut s'assurer que les pics de Bragg des faisceaux-crayon (pencil-beam en anglais) utilisés pour "peindre" la tumeur, se trouvent dans la tumeur et pas dans les tissus environnants. Premièrement, il faut s'interroger sur la précision intrinsèque de la balistique des ions due à des considerations physiques, y compris pour des matériaux simples de composition chimique connue comme c'est le cas de l'eau. Pour les tissus vivants, la composition chimique est habituellement estimée sur la base des unités Hounsfield et présente suffisamment de variabilité pour ajouter des incertitudes, même en l'absence de discrépances anatomiques avec le plan de traitement. Il a été montré dans [3] que ces erreurs peuvent atteindre facilement quelques millimètres. Viennent ensuite les erreurs de positionnement du patient et les modifications morphologiques entre l'imagerie pré-opératoire et le traitement. Deux alternatives se présentent alors, l'une étant de prévoir des marges suffisamment larges, l'autre étant d'imager les effets de l'irradiation. La question de la précision atteignable en radiothérapie par ions et celle de la possibilité de réaliser un contrôle per-opératoire par détection des particules secondaires sont encore ouvertes et suscitent l'intérêt de nombreuses équipes. D'après [55], la plupart des centres de protonthérapie utilisent une marge de 3.5% pour assurer la couverture de la tumeur, ce qui repousse la limite distale du faisceau jusqu'à 1 cm pour les tumeurs profondes. Aussi, les tirs dans la direction des organes à risque sont évités, par doute sur le lieu exact où

le faisceau s'arrêtera. Ces tirs pourraient pourtant être bénéfiques globalement, car ils éviteraient de toucher des organes importants.

Or, lors de l'irradiation des tissus avec des ions, de nombreuses particules secondaires sont produites et certaines peuvent être détectées à l'extérieur du patient. Parmi celles-ci, les photons γ-prompt, émis au maximum à quelques pico-secondes d'une interaction ion-noyau atomique du tissu. À l'aide de ces particules on pense pouvoir réaliser une mesure indirecte de la dose déposée, des tissus traversés par le faisceau, de la zone irradiée, ceci dans le but d'une vérification si possible en temps réel de la conformité du traitement réalisé avec le traitement planifié [48, 87, 70]. Plusieurs équipes ont alors pensé à l'imagerie SPECT et aux caméras collimatées. Relativement faciles à construire et à utiliser, ces caméras sont actuellement sur le point d'être implantées en clinique, un exemple étant la camera à fente (voir figure 5.4) de l'entreprise IBA [62].

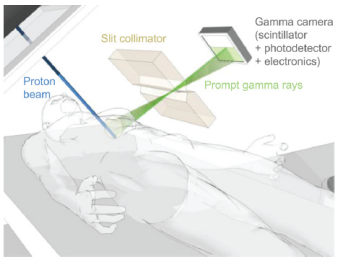


FIGURE 5.4 – Fonctionnement de la caméra avec collimateur à fente. Elle donne un apperçu 1D de la quantité de γ-prompts émis le long du faisceau ([60]).

Ces caméras nécessitent un collimateur de grande taille, capable d'arrêter des photons de 10 MeV, ce qui réduit leur champ de vue et leur sensibilité. Elles peuvent produire un signal 1D, le long du faisceau. Une autre idée intéressante consiste à utiliser un détecteur pour mesurer le temps de vol [18]. Moins onéreux et d'une complexité moindre comparés à la caméra Compton, tous ces dispositifs visent à localiser seulement la fin du parcours du faisceau dans les tissus. Une image 3D pourrait être bien plus utile car le faisceau traverse un milieu hétérogène. Plusieurs approches et dispositifs de contrôle en ligne ont été proposés et étudiés dans des projets du réseau européen ENLIGHT, dont le projet ENVISION auquel nous avons participé. L'étude bibliographique des méthodes de vérification du traitement en protonthérapie peut être trouvée dans [35] et [37].

Contrairement à la caméra collimatée qui est prévue pour fournir une image cumulée en 1D le long du faisceau, la caméra Compton nécessite aussi la reconstruction tomographique des images. Elle souffre d'un rendement assez faible, de coïncidences aléatoires, de bruit venant des autres particules secondaires dont les neutrons, et de la complexité algorithmique de la reconstruction d'images. En contrepartie, elle ne nécessite pas de collimation lourde qui de plus produit des particules secondaires par collision photon-collimateur, a un rendement supérieur aux caméras collimatées et peut fournir une image 3D. Tous ces arguments sont sujet à controverse car la complexité du problème et les difficultés techniques sont telles qu'à présent il n'y a pas vraiment une modalité capable d'assurer un contrôle efficace du dépôt de dose. Un prototype de caméra Compton est en construction

à l'Institut de Physique Nucléaire de Lyon. Des dispositifs expérimentaux sont étudiés par différentes équipes sur plusieurs continents, en particulier trois prototypes ont été étudiés dans le cadre du projet ENVISION, mais aucun développement industriel n'a encore été lancé en vue d'une application médicale.

5.3 Simulation de l'acquisition caméra Compton en conditions cliniques

Dans ce domaine les développements aussi bien instrumentaux que logiciels reposent en grande partie sur les simulations Monte-Carlo. La librairie Geant4 ([1]) est beaucoup utilisée, soit directement comme dans [64, 38, 16] soit avec une surcouche logicielle basée sur un interpréteur de commandes. Concernant les applications médicales on peut citer par exemple MEGAlib ([A7]), GAMOS ([24]) ou GATE ([CN2]).

Nous avons essentiellement utilisé le logiciel MEGAlib [98]. Celui-ci permet une prise en main facile, dispose d'une bonne documentation et a été validé sur plusieurs prototypes. Ce logiciel est composé de plusieurs modules, dont nous utilisons le module Geomega qui permet d'interpréter les commandes liées à la géométrie et la visualisation, le module Cosima qui sert à la simulation des particules et des interactions particules-matière, le module Sivan qui analyse les interactions, détecte les coïncidences diffuseur-absorbeur et produit en sortie les évènements utiles à la reconstruction d'images en mimant un détecteur idéal, et le module Revan qui réalise à peu près les mêmes étapes que Sivan, mais pour un détecteur réaliste. Ce dernier permet d'ajouter du bruit sur les positions et énergies mesurées et reconstruit les évènements sans information temporelle (figure 5.5).

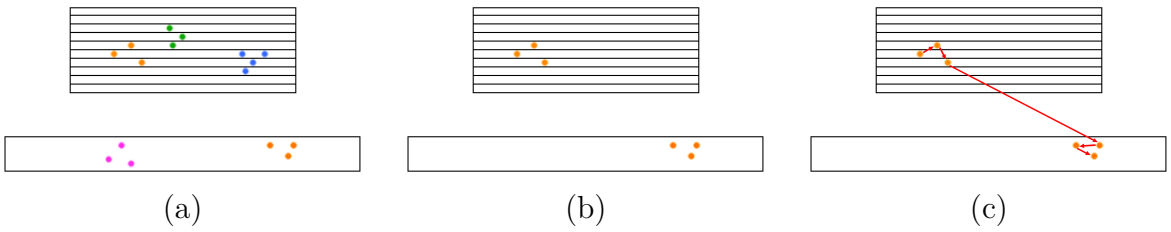


FIGURE 5.5 – Création des évènements à partir des interactions dans MEGAlib. Chaque photon est suivi individuellement (pas de coincidences aléatoires) et ses interactions sont représentées avec la même couleur (a). Les photons ayant touché seulement le diffuseur ou seulement l'absorbeur sont écartés. Pour un photon donné (b), l'ordre des interactions est inconnu pour Revan et déterminée par un algorithme (c). Dans la séquence ainsi ordonnée on retient les positions des deux premiers points ainsi que l'énergie déposée lors de la première diffusion et la somme des énergies déposées par la suite.

L'ordonnancement de la séquence des interactions a une importance cruciale, bien plus grande que celle de la résolution du détecteur. Une erreur à l'identification des deux premiers coups changera complétement le sommet et/ou l'axe du cône Compton. Quant

aux erreurs sur les mesures d'énergie, elles influent sur le demi-angle d'ouverture du cône.

Le logiciel MEGAlib n'est pas prévu à l'origine pour la simulation des sources voxelisées. Avec l'aide de l'auteur de MEGAlib, Estelle Hilaire a pu réaliser pendant son stage de master la simulation d'une source dans le fantôme voxelisé d'un patient, auquel elle a ajouté la définition des matériaux spécifiques à chaque organe. Ce travail était cependant assez fastidieux et les temps de calcul trop longs. Il nous a semblé plus pertinent d'utiliser un logiciel spécialement conçu pour l'imagerie médicale, le logiciel GATE ([30]). Cependant ce logiciel ne possédait ni les modules permettant la simulation d'une caméra Compton, ni un module de type Revan.

Dans sa thèse et en collaboration avec Charlotte Robert (qui travaillait alors au IMNC Paris), Estelle Hilaire a implémenté la caméra Compton dans GATE. Elle a réalisé une chaîne de simulation et analyse permettant de reproduire en simulation le plan de traitement avec un fantôme de patient. Alors que la simulation (étape (a) dans la figure 5.5) était faite par GATE, la sortie était transférée dans le format MEGAlib et analysée par Revan (étape (c)). Ces travaux ont donné lieu à une communication au colloque GRETSI en 2013 [CN2].

Cette solution ne nous a pas semblé satisfaisante, car trop complexe et par conséquent difficile à valider. Pour la suite de sa thèse et pour son article [A13] concernant la vérification du traitement en proton-thérapie, Estelle Hilaire a changé de stratégie et a utilisé le concept d'espace de phase de GATE (figure 5.6). La simulation de l'irradiation était effectuée dans GATE et les photons γ étaient capturés dans un espace de phase qui enregistrait leur position, direction et énergie. L'espace de phase était ensuite utilisé pour définir des sources dans MEGAlib. Cette approche permettait aussi d'accélérer les calculs, en faisant une seule irradiation proton, à partir de laquelle on déterminait la quantité approximative de photons produits par voxel du volume. Chaque voxel devient ensuite une source d'intensité connue, qui émet dans une direction aléatoire, des photons qui donneront des évènements Compton ou pas. On peut ainsi réaliser un grand nombre de tirages relativement indépendants, à condition que le nombre initial de protons soit suffisamment élevé.

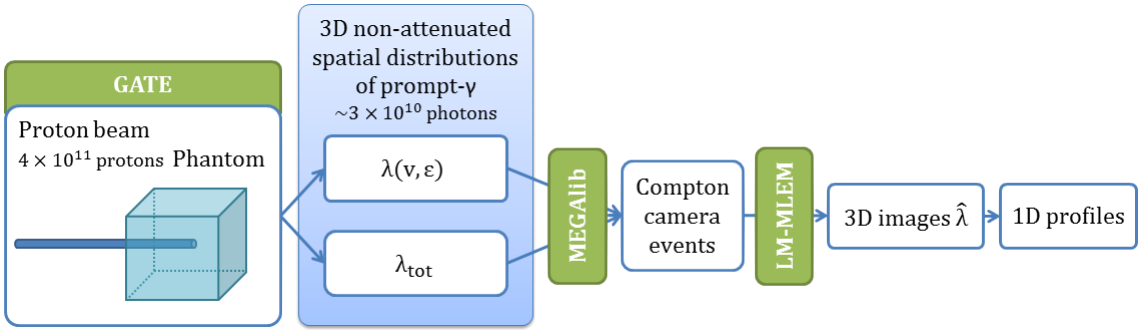


FIGURE 5.6 – Chaîne de simulation et analyse.

Un travail bien plus général est à présent réalisé par Ane Etxebeste (CREATIS, Centre Léon Bérard, Lyon), travail qui vise le développement dans Gate des macros spécifiques

pour la caméra Compton. Ceux-ci faciliteront la description des différents paramètres géométriques, d'acquisition, d'analyse de données. Une première validation a été réalisée pour des données idéales en comparant statistiquement les sorties GATE et MEGAlib. Une validation à l'aide des images reconstruites a été faite ensuite en collaboration avec Yuemeng Feng. La suite concerne les simulations réalistes et l'implémentation des analyses de type Revan. Ces travaux nous faciliteront la simulation pour fantômes de patient ainsi que la comparaison avec la SPECT classique par caméra collimatée, ces composantes étant déjà implémentées dans GATE.

Chapitre 6

Reconstruction d'images caméra Compton

Comme il a été expliqué dans le chapitre précédent et pour une acquisition parfaite des mesures, à chaque évènement détecté correspond une émission quelque part sur la surface d'un cône. Ce cône a pour sommet le point V_1 où le photon a été diffusé pour la première fois, un axe dont la direction $\Omega = \overrightarrow{V_2V_1}$ est opposée à la trajectoire du photon après cette première diffusion, et un demi-angle d'ouverture β calculé à partir de l'énergie E_1 transférée à l'électron et de l'énergie totale du photon, E_0 selon la formule (5.1). Supposant un temps d'acquisition infini pour une source éternelle, la fraction d'évènements acquis avec les mêmes caractéristiques (V_1, Ω, E_1, E_0) est proportionnelle à la moyenne de l'intensité f de la source sur la surface conique. Cette moyenne peut être complétée d'un poids. Nous obtenons ainsi une transformée de Radon généralisée de la fonction f de l'intensité de la source. Contrairement aux transformées de Radon sur des lignes, des cercles, des sphères ou même des coniques (intersections d'un cône avec un plan), celle-ci n'est étudiée que depuis peu et ceci grâce à son importance dans cette nouvelle technique d'imagerie.

On pourra noter une différence essentielle avec la cone-beam CT (CBCT) bien connue en imagerie médicale des rayons X. Même si dans les deux cas on parle des cônes, la CBCT mesure individuellement les intégrales d'un faisceau de demi-droites d'origine commune et remplissant un cône. On n'est donc pas très loin du modèle standard de l'intégration sur la ligne et de la transformée de Radon classique.

L'inverse analytique de la transformée de Radon conique est assez difficile à calculer mais des méthodes ont été proposées ces dernières années. De surcroît elle nécessite des hypothèses en termes de géométrie d'acquisition qui sont irréalisables en pratique ou très coûteuses. Il s'ensuit des acquisitions avec des projections manquantes et des artéfacts qui s'apparentent à la fois à l'angle manquant et aux projections tronquées. Les formules d'inversion existantes à ce jour sont parfois complexes et difficiles à reproduire dans un code informatique.

Quant aux méthodes itératives, qui résolvent le problème linéaire discret, la facilité de l'utilisation est relative. Premièrement, le modèle du problème direct est encore

insuffisamment maîtrisé et les simulations Monte Carlo ne peuvent pas vraiment aider vue la complexité du problème et la grande dimension de l'espace dans lequel vivent les données. Ensuite, l'implémentation de ce modèle direct est plus difficile que pour les modalités classiques car chaque projection, pour chaque évènement acquis, se fait selon une surface conique, voir une coque conique plus ou moins épaisse selon la confiance faite dans les mesures. Enfin, les temps de calcul sont très importants et les astuces algorithmiques qui s'appliquent par exemple en CBCT ne semblent pas pouvoir aider.

Parmi les méthodes itératives nous avons choisi d'utiliser la version list-mode (LM) de l'algorithme Maximum Likelihood Expectation Maximization (MLEM), bien connu en imagerie médicale. Cet algorithme est construit sur une hypothèse d'émission obéissant à une loi de Poisson qui semble ici naturelle. Cet algorithme converge en théorie vers le maximum de la fonction de vraisemblance, mais rien n'assure que ce maximum correspond à l'image de la source. En pratique, plus le nombre d'évènements est faible, plus l'image reconstruite est bruitée et ressemble à un ensemble de spots isolés. C'est pourquoi les itérations doivent être arrêtées selon un a priori que nous avons sur la régularité de la source. Une meilleure approche consiste d'ajouter cet a priori dans la fonction coût à minimiser. Le choix de l'arrêt des itérations est alors remplacé par un choix du poids de la régularisation dans la fonction coût.

Dans ce chapitre nous présentons les thématiques que nous avons abordées et les contributions apportées à la reconstruction d'images pour la caméra Compton.

6.1 La transformée de Radon conique

Le problème direct est modélisé à l'aide d'une transformation linéaire où les données sont des intégrales d'une fonction sur la surface d'un cône ou sur une coque conique – lorsque l'on modélise les incertitudes de mesure. L'inversion analytique est plus complexe que dans les modèles usuels rencontrés en médecine nucléaire, où les projections sont des intégrales sur des lignes. Elle se fait à l'aide des harmoniques sphériques, de la transformée de Fourier ou encore de la transformée de Hilbert. Dans tous les cas, pour des géométries d'acquisition usuelles les projections ne satisfont pas les conditions de complétude et les images présentent des artéfacts. Dans deux papiers [A6, A10], le premier publié en 2009 dans Inverse Problems et le second en 2014 dans IEEE TIP, nous avons proposé un théorème de type coupe-projection et un algorithme de type rétroprojection filtrée adaptés à la géométrie d'acquisition usuelle avec détecteurs plans et nous avons mis en évidence une façon de regrouper les projections dans des ensembles redondants.

Soit f une fonction réelle bornée et à support compact définie sur $\mathbb{R}^3$. La transformée de Radon conique consiste à calculer les intégrales pondérées de f sur des surfaces de cônes. Ces cônes sont définis par leur sommet $u \in \mathbb{R}^3$, leur axe $\Omega \in \mathbb{S}^2$ qui est la sphère unité de $\mathbb{R}^3$ et par le demi-angle d'ouverture $\beta \in]0, \pi[$. On note respectivement α et δ la colatitude et la longitude de l'axe en coordonnées sphériques (voir figure 6.1).

Le modèle généralement étudié dans la littérature est celui des transformées de Radon

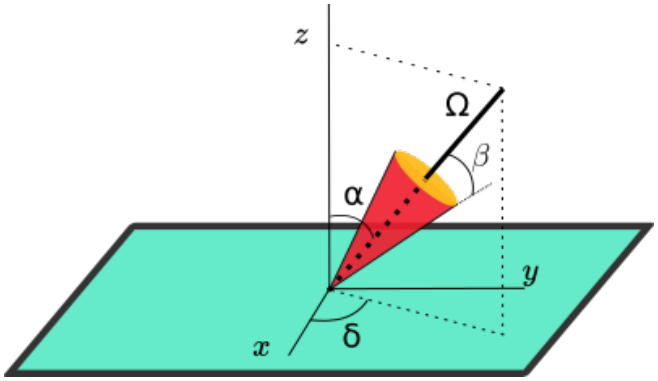


FIGURE 6.1 – Parametrisation du cône Compton.

coniques pondérées définies pour un $k \in \mathbb{Z}$ par

$$\mathcal{C}^k f(u, \alpha, \delta, \beta) = \int_{(v-u).\Omega = \|v-u\| \cos\beta} f(v) \frac{1}{\|v-u\|^k} dv, \tag{6.1}$$

où $\|\cdot\|$ est la norme euclidienne de $\mathbb{R}^3$. Dans cette présentation nous nous concentrerons sur le cas tridimensionnel, mais la transformation peut être définie plus généralement pour les fonctions sur $\mathbb{R}^n$. Pour $n = 2$, on parle plutôt de "V-line transform" ([7, 21]) ou "broken-ray transform" ([14]). Ce type de transformée peut servir en imagerie Compton pour des évènements à trois sites comme il est expliqué dans [7] ou encore en tomographie optique à diffusion unique. L'appellation "broken-ray transform" étant aussi utilisée par certains auteurs pour désigner l'intégration sur des lignes brisées composées de plus de deux segments, la terminologie "V-line transform" est privilégiée.

Les premiers travaux sur l'inversion de la transformée de Radon conique datent de 1994 avec l'article de Cree et Bones [11], suivis dans la periode avant 2005 de ceux de Basko *et al* [6], L. Parra [58], Tomitani et Hirasawa [89], B. Smith [78]. Dans la même période a été étudiée un problème proche, celui de la tomographie par rayonnement diffusé, consistant à utiliser le rayonnement diffusé dans l'objet, qui est à présent considéré comme bruit et écarté. Un exemple d'application possible en est la tomographie par émission de positons (TEP). Ce problème est plus compliqué dans le sens où le sommet du cône n'est pas connu et que l'on acquiert des projections coniques cumulées. On pourra consulter pour plus de détails les travaux de Nguyen *et al* , par exemple [51, 52]. Pour revenir à la caméra Compton, depuis les années 2010 plusieurs articles émanant de la communauté mathématique ont vu le jour et ont entériné la notion de transformée de Radon conique, notion qui portait le nom similaire de transformée de Radon sur des cônes dans [52]. Des références peuvent être trouvées dans Terzioglu *et al* , [86]. Des mini-symposiums spécifiques aux transformées de Radon généralisées sont proposés dans les conférences du domaine, par exemple "Inverse Problems : Modeling & Simulation" (Ölüdeniz-Fethiye, Turquie, puis Malte à partir de 2018) et "100 Years of the Radon Transform" organisée par le Radon Institute of Computational and Applied Mathematics de Linz, Autriche en 2017.

Ces transformées dépendent de six variables : les trois coordonnées du sommet, le demi-angle d'ouverture et les deux coordonnées angulaires de la direction de l'axe. La

transformation est redondante puisqu'elle passe d'un espace tridimensionnel à un espace de dimension six. Cet argument explique en partie la variété des techniques proposées pour son inversion. Plusieurs cas particuliers ont été traités, selon la restriction imposée à l'espace de données. Dans un article publié en 2009, nous avons étudié un modèle qui s'apparente à celui de l'équation (6.1) mais avec un poids légèrement différent. Pour faire un parallèle avec la transformée de Radon, nous l'avons nommé transformée Compton. Récemment nous avons montré que l'inverse de cette transformée s'adapte facilement au modèle (6.1) avec $k = 1$ [A16].

6.1.1 La transformée Compton

Pour des paramètres α, δ, β, avec $\alpha \in [0, \pi]$, $\beta \in]0, \pi[$ et $\delta \in]-\pi, \pi]$, on appellera projection Compton de f l'application $\mathscr{C}_{\alpha,\delta,\beta} f : \mathbb{R}^2 \times \{0\} \to \mathbb{R}$ définie par

$$\begin{aligned} \mathscr{C}_{\alpha,\delta,\beta} f(\mathbf{u}) &= K(\beta, E_0) \cos\alpha \int_{(\mathbf{v}-\mathbf{u}).\mathbf{\Omega}=\|\mathbf{v}-\mathbf{u}\|\cos\beta} f(\mathbf{v}) \cos\theta d\mathbf{v} \\ &= K(\beta, E_0) \cos\alpha \int_{(\mathbf{v}-\mathbf{u}).\mathbf{\Omega}=\|\mathbf{v}-\mathbf{u}\|\cos\beta} f(\mathbf{v}) \frac{z}{\|\mathbf{v}-\mathbf{u}\|} d\mathbf{v}, \end{aligned} \tag{6.2}$$

où $\mathbf{v} = (x, y, z)$ est un vecteur de $\mathbb{R}^3$ et θ est la colatitude du vecteur $\mathbf{v} - \mathbf{u}$. On appellera transformée Compton l'ensemble des projections de f. Le modèle est adapté pour une caméra dont le diffuseur est plan, de faible épaisseur mais de largeur infinie. On choisit le système de coordonnées de sorte que le diffuseur coïncide avec le plan $z = 0$. L'absorbeur est lui aussi considéré plan, mais cette hypothèse est seulement liée au facteur $\cos\alpha$ et est facilement modifiable. Ces facteurs $\cos\theta$ et $\cos\alpha$ de (6.2) sont introduits pour prendre en compte l'angle incident d'un faisceau de photons sur les détecteurs. Dans la figure 6.2 on peut voir le lien entre la formule (6.2) et les données. Une projection collectionne les évènements dont le cône Compton correspond à une forme et orientation données, mais dont le sommet est libre de se déplacer sur le diffuseur.

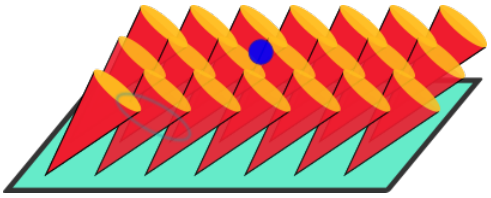
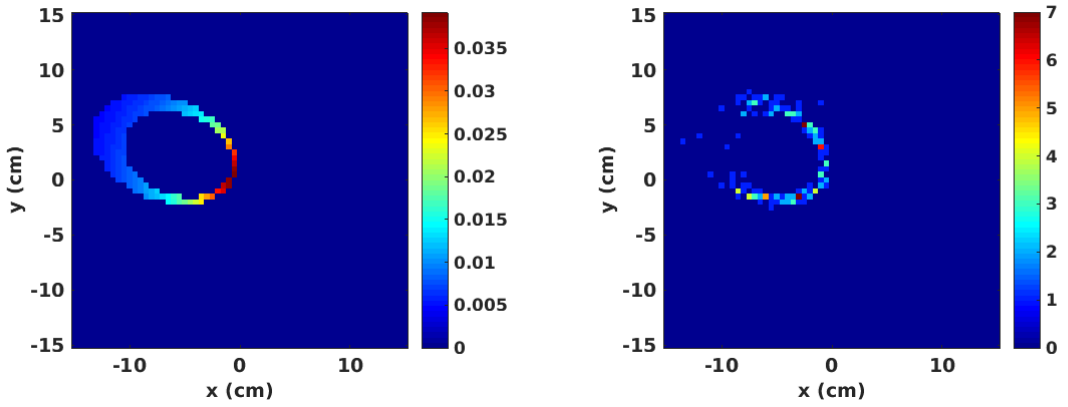


FIGURE 6.2 – L'image d'une projection $\mathscr{C}_{\alpha,\delta,\beta}$ est l'image dessinée sur la surface du diffuseur par les sommets des cônes de paramètres α, δ, β et qui intersectent la source. L'intensité de l'empreinte est proportionnelle à l'intégrale (6.2).

La projection d'une petite source sphérique est représentée dans la figure 6.3, projection calculée selon le modèle en (a) et à partir des données Monte Carlo en (b). On peut remarquer que les deux formes sont ressemblantes, cependant les intensités des pixels semblent différer légèrement au delà des variations statistiques.



(a) (b)

FIGURE 6.3 – Exemple de projection Compton d'une petite source sphérique. L'image (a) montre une projection deterministe calculée par intégration (6.2). Dans (b) on montre la même projection déterminée à partir de la simulation Monte Carlo. Le modèle utilisé pour calculer (a) décrit les données de manière imparfaite.

6.1.2 Théorème coupe-projection

Sur le modèle du théorème bien connu appelé coupe projection reliant les projections Radon à la transformée de Fourier de l'image, nous avons montré le théorème suivant. La figure 6.4 peut aider à le comprendre. Nour restreignons l'ensemble de données avec la contrainte $\alpha + \beta < \pi/2$ qui correspond à garder les cônes qui intersectent le diffuseur en un seul point, le sommet du cône. Avec cette restriction on a $\cos^2\alpha - \sin^2\beta > 0$. On note

$$b(\alpha,\beta) = \frac{\sin\beta}{\sqrt{\cos^2\alpha - \sin^2\beta}} \tag{6.3}$$

Dans un premier temps on définit la transformée Hankel-Fourier de f, qui est la transformée de Hankel en la coordonnée z de la transformée de Fourier de f par rapport aux coordonnées x et y :

$$(\mathscr{H}\mathscr{F}f)(\eta_1,\eta_2,\eta_3) := 2\pi \int_0^\infty z\mathscr{F}_2 f(\eta_1,\eta_2,z) J_0(2\pi z\eta_3) dz, \tag{6.4}$$

où $\mathscr{F}_2$ désigne la transformée de Fourier par rapport aux deux premières variables d'une fonction. On définit aussi dans le domaine fréquentiel le plan

$$\mathcal{P}_F(\delta) : \ \eta_1\cos\delta + \eta_2\sin\delta = 0 \tag{6.5}$$

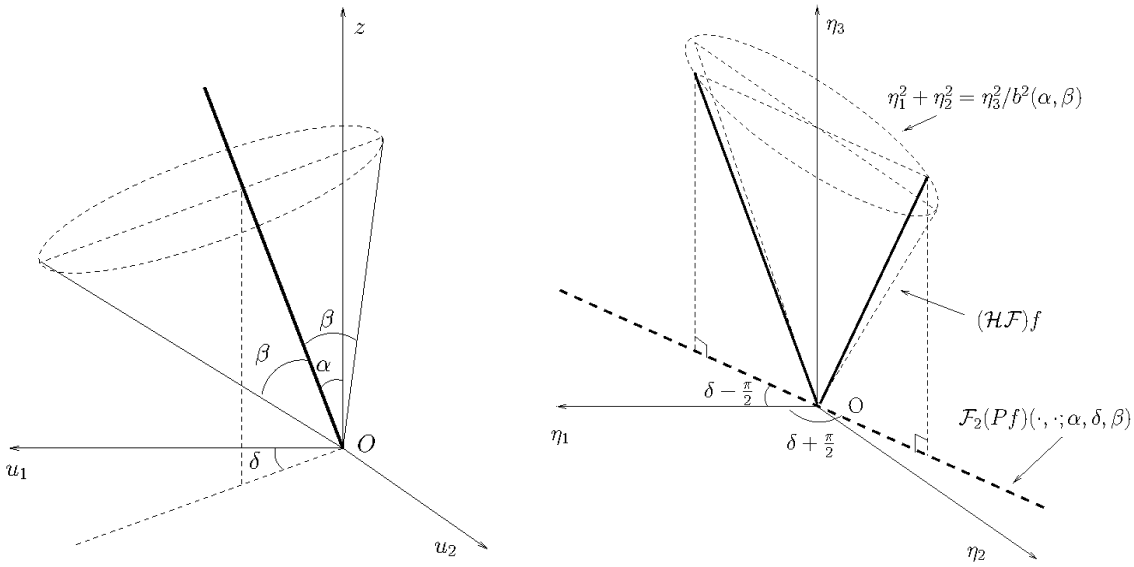
et le cône

$$\mathcal{C}_F(\alpha,\beta) : \ \eta_1^2 + \eta_2^2 = \frac{\eta_3^2}{b^2(\alpha,\beta)}. \tag{6.6}$$

On montre alors (voir [A6]) que pour $(\eta_1,\eta_2) \in \mathbb{R}^2$ tel que $\eta_1\cos\delta + \eta_2\sin\delta = 0$ on a

$$(\mathscr{H}\mathscr{F}f)(\eta_1,\eta_2,b(\alpha,\beta)\sqrt{\eta_1^2+\eta_2^2}) = \frac{1}{K(\beta,E)b(\alpha,\beta)\cos\alpha}\mathscr{F}_2(\mathscr{C}_{\alpha,\delta,\beta}f)(\eta_1,\eta_2), \tag{6.7}$$

d'où le théorème suivant (voir aussi la figure 6.4).



a) Domaine spatial, cône $C(0,0;\alpha,\delta,\beta)$ b) Domaine fréquentiel, cône $\mathcal{C}_F(\alpha,\beta)$.

FIGURE 6.4 – *a) Domaine spatial.* La projection Compton $\mathscr{C}_{\alpha,\delta,\beta}f$ est le résultat de l'intégration sur des surfaces de cônes d'axes parallèles et ayant le même demi-angle d'ouverture β. Chaque cône est une translation du cône $\mathcal{C}(0,0;\alpha,\delta,\beta)$. *b) Domaine fréquentiel.* Pointillés épais : ligne de la transformée de Fourier 2D de $\mathscr{C}_{\alpha,\delta,\beta}f$. Ligne continue épaisse : coupe dans la transformée Hankel-Fourier de f sur la surface du cône $\mathcal{C}_F(\alpha,\beta)$.

Théorème 6.1.1. *Soient $\alpha \in [0,\pi/2[$, $\delta \in [0,2\pi[$ et $\beta \in]0,\pi/2-\alpha[$ des angles donnés. En chaque point de la ligne brisée qui est l'ensemble $\mathcal{C}_F(\alpha,\beta) \cap \mathcal{P}_F(\delta)$, la valeur de la transformée Hankel-Fourier de f pondérée par le facteur $K(\beta,E)b(\alpha,\beta)\cos\alpha$ est égale à la valeur de la transformée de Fourier de la projection $\mathscr{C}_{\alpha,\delta,\beta}f$ au point projeté se trouvant à l'intersection du plan $\eta_3 = 0$ avec le plan $\mathcal{P}_F(\delta)$.*

Ce théorème montre qu'il est possible de remplir les valeurs de la transformée Hankel-Fourier de f sur la surface conique $\mathcal{C}_F(\alpha,\beta)$ à partir de l'ensemble des projections $\mathscr{C}_{\alpha,\delta,\beta}f$, avec $\delta \in [_\pi,\pi[$. Ensuite, lorsque α est constant et β parcourt l'intervalle $]0,\pi/2-\alpha[$, les cônes $\mathcal{C}_F(\alpha,\beta)$ remplissent le sous-espace $\eta_3 > 0$ (excepté la demi-droite ayant $\eta_1 = \eta_2 = 0$). La transformée Hankel-Fourier est inversible et il est donc possible à partir de ces valeurs de retrouver f. Pour ce faire, la meilleure solution d'un point de vue algorithmique consiste à utiliser une formule de retroprojection filtrée.

Sur les six paramètres spécifiant les projections, nous en avons fixé deux : l'altitude du sommet du cône et la valeur de δ. Cette dernière dépend du point où l'inverse sera calculée. La contrainte reliant les valeurs de α et β permet de décrémenter le nombre de degrés de liberté pour enfin atteindre trois, la dimension de l'espace image. L'inversion peut alors s'effectuer sans redondance dans les données.

D'autres méthodes d'inversion ont été proposées plus récemment. Une synthèse en est faite dans [U1], document non encore publié et donné en annexe.

6.1.3 Formule de retroprojection filtrée

Dans le cas $\alpha = 0$, qui correspond aux cônes d'axe vertical, une formule de retroprojection filtrée a été donnée par Cree et Bones dans [11]. Cette restriction est forte car elle correspond à une collimation entre diffuseur et absorbeur. La méthode que nous rappelons ci-après permet de s'affranchir de cette contrainte et d'utiliser tous les évènements dont $\alpha \in]0, \pi/2[$. Elle a été publiée pour la première fois dans les actes de la conférence Fully3D, suite au stage de master de Xavier Lojacono [C3].

En repartant du théorème coupe-projection pour projection coniques, on remplace la ligne de la transformée de Fourier d'une projection par la transformée de Fourier de la projection Radon 2D de la projection conique. On note $\mathscr{R}_\delta$ cette projection Radon 2D, indexée sur $\delta \in \mathbb{R}$. Elle associe à toute fonction g à support compact l'ensemble d'intégrales sur des lignes

$$\forall\, s \in \mathbb{R}, \quad \mathscr{R}_\delta g(s) = \int_{-\infty}^{\infty} g(s\cos\delta + t\sin\delta, s\sin\delta - t\cos\delta)dt. \tag{6.8}$$

On note $f_z : (x, y) \in \mathbb{R}^2 \mapsto f(x, y, z) \in \mathbb{R}$ la restriction de f à un plan horizontal. Pour $\tau \geq 0$ et $\delta \in]-\pi, \pi]$ la transformée intégrale définie pour tout $s \in \mathbb{R}$ par :

$$\mathscr{P}_{\tau,\delta} f(s) = \tau \int_0^{\infty} \int_{-\pi}^{\pi} \mathscr{R}_{\delta+\frac{\pi}{2}} f_z(s + z\tau\sin\varphi) z d\varphi dz, \tag{6.9}$$

vérifie

$$\mathscr{P}_{\tau,\delta} = \frac{1}{K(\beta, E_0)\cos\alpha} \mathscr{R}_{\delta+\frac{\pi}{2}} \mathscr{C}_{\alpha,\delta,\beta}, \tag{6.10}$$

lorsque $\tau = b(\alpha, \beta)$, où $b(\alpha, \beta)$ a été défini par la formule (6.3). Quand $\tau > 0$ est donné, les valeurs de α et β vérifiant $b(\alpha, \beta) = \tau$ peuvent être calculées l'une en fonction de l'autre d'après (6.3) qui peut aussi être écrite :

$$\tau\cos\alpha = \sqrt{1+\tau^2}\sin\beta, \tag{6.11}$$

car les applications $\beta \in \left(0, \frac{\pi}{2} - \alpha\right) \mapsto b(\alpha, \beta) \in \mathbb{R}_+^*$, pour $\alpha \in \left[0, \frac{\pi}{2}\right)$ donné, et $\alpha \in \left[0, \frac{\pi}{2} - \beta\right) \mapsto b(\alpha, \beta) \in \mathbb{R}_+^*$, pour $\beta \in \left(0, \frac{\pi}{2} - \alpha\right)$ donné, sont bijectives.

La transformée $\mathscr{P}$ fournit un ensemble de données inclus dans un espace tridimensionnel indexé sur $\tau > 0$, $\delta \in (-\pi, \pi]$ et $s \in \mathbb{R}$. l'ensemble des données peut être encore réduit, tout en gardant la dimension, compte tenu de la relation

$$\mathscr{R}_{\delta+\frac{\pi}{2}} \mathscr{C}_{\alpha,\delta+\pi,\beta} = \mathscr{R}_{\delta+\frac{\pi}{2}} \mathscr{C}_{\alpha,\delta,\beta}, \tag{6.12}$$

qui signifie que du point de vue de la transformée $\mathscr{P}$ les projections $\mathscr{C}_{\alpha,\delta+\pi,\beta}$ et $\mathscr{C}_{\alpha,\delta,\beta}$ partagent la même information.

Comme montré dans [A10], l'image de la source peut être reconstruite en tout point $\mathbf{v} = (x, y, z) \in \mathbb{R}^3$ en utilisant la formule

$$f(\mathbf{v}) = 2\pi \int_0^{\pi} \int_0^{\infty} \left(\int_{-\infty}^{\infty} \widehat{\mathscr{P}_{\tau,\delta} f}(\sigma) J_0(2\pi z\tau\sigma) |\sigma|^3 e^{2i\pi\sigma \mathbf{v}.\mathbf{d}_2} d\sigma \right) d\tau d\delta, \tag{6.13}$$

où $\mathbf{d}_2 = (-\sin\delta, \cos\delta, 0)$ et J_0 est la fonction de Bessel d'ordre zéro. La reconstruction se fait efficacement coupe par coupe suivant la direction $z > 0$ de la façon suivante. Les projections Radon des projections Compton sont filtrées avec un filtre dépendant de τ, paramètre dépendant de la forme des cônes ayant produit les projections et de z. Les projections filtrées sont intégrées selon τ puis retroprojetées sur des lignes à chaque altitude z.

Des variantes de la formule (6.13) peuvent être écrites et implémentées, selon que le filtrage se fait dans le domaine de Fourier (avec du zero-padding pour éviter le repliement spectral très visible du fait de la décroissance lente de la fonction de Bessel) ou dans le domaine réel. On notera aussi que la composante $|\rho|^3$ du filtre peut être vue comme un filtrage rampe couplé avec un laplacien. Une alternative consiste alors à utiliser la retroprojection filtrée de la transformée de Radon suivie d'un filtrage par le laplacien de l'image 2D d'une coupe. Ces variantes ainsi que l'importance d'un filtrage lissant (en occurrence un débruitage par minimisation de la variance totale) ont été discutées dans ([A16]).

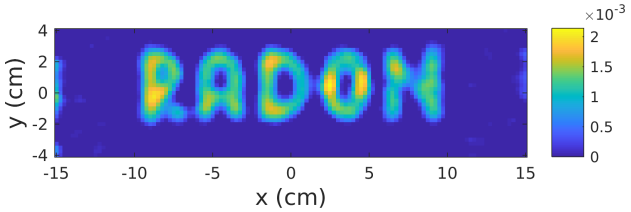


FIGURE 6.5 – Source parallele au détecteur et données Monte-Carlo idéales simulées avec MEGAlib et analysées avec SIVAN. Un filtrage TV a été appliqué à l'image reconstruite. (source [A16])

6.2 Modélisation du problème direct

Un évènement est défini par deux positions, celle de la première interaction V_1 et celle de la seconde V_2 dans le détecteur, et par deux énergies, E_1 et E_2, où E_2 correspond à la somme des énergies déposées par le photon après la diffusion en V_1. On peut aussi définir l'évènement par les deux positions et par l'angle de diffusion Compton, calculé à partir des énergies déposées d'après (5.1). D'autres paramètres peuvent s'y ajouter, dont la section efficace de diffusion Compton donnée par la formule de Klein-Nishina et les incertitudes sur les mesures. Les évènements peuvent être regroupés selon leur caractéristiques dans des tableaux qui rappellent les sinogrammes, ou bien être traités un par un, généralement selon l'ordre d'arrivé. Dans ce dernier cas on parle d'algorithme en mode liste. Dans tous les cas, il est possible d'associer à chaque évènement ou classe d'évènements un index qu'on notera généralement i. Le problème direct consiste à exprimer le nombre n_i d'évènements par classe (en mode liste la classe contient zéro ou un élément) en fonction des nombres f_j de photons émis par les voxels. On obtiendra un système linéaire dont l'équation type est :

$$n_i = \sum_{j=1}^{J} a_{ij} f_j, \tag{6.14}$$

où a_{ij} est la proportion de photons émis par le voxel j et détectés dans la classe i. Cependant la matrice A de ce système n'étant pas connue, on la remplace par la matrice système T dont l'élément t_{ij} est défini comme la probabilité d'acquérir un évènement de caractéristiques indexées par i, sachant qu'il a été émis par le voxel d'indice j. Lorsque le taux d'émission est faible, l'écart entre les deux matrices est très important, et requiert une modélisation probabiliste qui remplace les valeurs réelles n_i et f_j par des valeurs moyennes μ_i et λ_j reliées par

$$\mu_i = \sum_{j=1}^{J} t_{ij}\lambda_j. \tag{6.15}$$

On cherche alors les valeurs du vecteur $\boldsymbol{\lambda}$ en fonction non pas des valeurs μ_i qui sont inconnues, mais en fonction des observations n_i, sous l'hypothèse supplémentaire selon laquelle la loi suivie par les différentes variables est une loi de Poisson.

6.2.1 Modélisation mathématique de l'acquisition

Les premiers travaux publiés sur le calcul des probabilités t_{ij} sont probablement ceux de Wilderman dans une série de papiers introduisant la reconstruction par l'algorithme List-Mode Maximum Likelihood Expectation Maximization (LM-MLEM) pour la reconstruction tomographique d'images de caméra Compton, en 2D puis en 3D (dont [94]). Les phénomènes physiques qui entrent en jeu étant assez compliqués et difficiles à décrire dans le modèle, des simplifications sont nécessaires. La simplification extrême consiste à associer un poids de 1 aux voxels intersectant la surface conique et zéro aux autres. De manière plus réaliste, on peut considérer que la probabilité t_{ij} d'acquérir un évènement de caractéristiques indexées par i, sachant qu'il a été émis par un voxel d'indice j est proportionnelle à la surface de l'intersection entre le cône Compton associé à i avec le voxel j. Ce modèle simple peut être affiné en y ajoutant l'influence d'autres phénomènes physiques. L'angle solide couvert par l'élément détecteur au point émetteur est un de ces facteurs car les sources émettent de manière isotrope dans 4π. Cet angle solide dépend de la distance source-point de détection ainsi que de l'angle incident sur la surface. S'ajoutent ensuite les incertitudes de mesure, sur la position de l'interaction et l'énergie déposée. Toute erreur de mesure sur l'énergie induisant une erreur dans le calcul de l'angle de diffusion Compton, les incertitudes sur la mesure de l'énergie conduisent naturellement à des incertitudes sur l'angle d'ouverture du cône Compton. Par conséquent l'intégrale de surface doit être remplacée par une intégrale volumique avec un poids modélisant ces incertitudes.

Le modèle proposé par Sauve *et al* dans [66] est défini comme le produit entre la probabilité qu'un photon γ émis au point M atteigne le détecteur au point V_1 (calculé à partir de l'angle solide), la section efficace de diffusion Compton donnée par la formule de Klein-Nishina, des facteurs qui dépendent seulement de l'évènement et qui se simplifient dans la formule de reconstruction MLEM, le facteur $1/\sin\beta$ et une densité probabiliste de type von Mises pour modéliser les incertitudes de mesure. Le facteur $1/\sin\beta$ favorise les voxels situés vers l'intérieur du cône épaissi pour prendre en compte les incertitudes.

Pour des distributions de faible variance, la loi de von Mises peut être approchée par une gaussienne, dont les paramètres sont plus aisés à estimer.

Il reste que la modélisation des probabilités t_{ij} n'est pas évidente. Pour les algorithmes MLEM et LM-MLEM il est aussi nécessaire d'estimer le vecteur des sensibilités

$$s_j = \sum_i t_{ij}, \tag{6.16}$$

où l'index i couvre tous les évènements possibles, et pas seulement ceux réalisés. La signification de ce terme est liée à la probabilité de détection d'un photon émis dans le voxel j. Cette probabilité peut être calculée sans trop de difficultés par simulation Monte Carlo, mais cette approche semble difficilement applicable pour la matrice système T.

Une analyse de quelques configurations pour le modèle direct a été faite dans [A12]. Une étude plus approfondie est en cours, dans la thèse de Yuemeng Feng dont la soutenance est prévue pour l'automne 2019.

6.2.2 Calcul numérique de la matrice système

Supposant le modèle connu, le calcul numérique de la matrice système peut se faire de différentes façons. Une méthode basée sur le calcul de la distance des voxels de l'image à la surface conique a été proposée et implémentée par Estelle Hilaire au cours de son stage de master 2 de physique médicale que j'ai encadré et a été utilisée dans [A13]. Le stage a été financé par le projet régional de recherche en hadron-thérapie. Une autre méthode, basée sur la modélisation du cône avec des génératrices a été développée dans sa thèse par Xavier Lojacono, dans le cadre du projet européen ENVISION. Elle a été présentée au colloque GRETSI'11 [CN1]

Plus précise, la seconde méthode est plus difficilement applicable lorsqu'on veut prendre en compte les incertitudes de mesure. Cela reste néanmoins possible indirectement, en introduisant dans le modèle de détection (6.14) une convolution avec la fonction d'étalement du point (PSF) comme suit. Cette réécriture du modèle permet aussi de déplacer la difficulté du calcul de la matrice système vers l'estimation de la fonction d'étalement du point dans tout point de l'image.

6.2.3 Apport de la PSF image pour la reconstruction

Pour faciliter l'exposé, considérons le cas simplifié de la transformée de Radon 2D, altérée par une PSF du détecteur de type gaussienne d'écart-type σ_1. Ce cas de figure est rencontré en SPECT ou en TEP collimaté. Nous noterons G la version discrétisée de l'opérateur de convolution associé. Les observations faites suivent (en ignorant les variations statistiques) le modèle :

$$\tilde{\mu} = GT\boldsymbol{\lambda}, \tag{6.17}$$

où l'opérateur T correspond à la projection Radon, donc à une intégration sur des lignes. On sait dans ce cas que l'image reconstruite d'une source ponctuelle, qui est donc la

fonction d'étalement d'un point de l'image, est également une gaussienne de même écart-type (on pourra consulter par exemple [43]). La différence consiste dans la dimension : alors que la première est une fonction définie sur $\mathbb{R}$, la seconde est une fonction bi-dimensionnelle. Nous noterons H l'opérateur linéaire de convolution dans le domaine image. Formellement, on a la permutation $GT = TH$ et les données viennent d'un modèle

$$\tilde{\mu} = TH\boldsymbol{\lambda}. \tag{6.18}$$

L'opérateur $\widetilde{T} = GT = TH$ correspond à une projection Radon sur des "lignes épaissies" par la gaussienne 1D. Il peut être implémenté comme tel, ou bien comme une composition d'applications linéaires. Dans tous les cas, la précision de cet "épaississement" est cruciale, bien que la valeur de l'écart-type σ_1 ne soit pas connue et puisse seulement être estimée, soit à partir des caractéristiques physiques du détecteur, soit, le plus souvent, à partir de la fonction d'étalement du point dans l'image. Cette dernière est calculée en résolvant (6.18) pour $H\boldsymbol{\lambda}$, à partir de données d'une source ponctuelle. Résoudre (6.17), ou de manière équivalente (6.18), revient en théorie à reconstruire d'abord l'image $H\boldsymbol{\lambda}$ de la source convoluée avec la PSF puis à déconvoluer pour in fine déterminer $\boldsymbol{\lambda}$. Lorsque l'estimation σ_2 est inférieure à la vraie valeur, la déconvolution sera incomplète et la résolution sera recouverte de manière partielle. Inversement, lorsque $\sigma_2 > \sigma_1$, la déconvolution n'est pas possible car elle conduirait à une image dont les hautes fréquences tendent vers l'infini. Numériquement on observe, proche des bords francs, des artéfacts assimilables au phénomène de Gibbs. Ces artefacts ont été observés en imagerie TEP et SPECT et étudiés pour des données de simulation par exemple par Snyder *et al* dans [79], puis plus en détail par Stute *et al* dans [81]. Nous y reviendrons au moment de la présentation de l'algorithme MLEM.

Deux autres difficultés s'ajoutent dans le cas des données réelles ou issues de simulations réalistes, incluant du bruit de mesure. D'abord l'existence de la matrice H permettant la permutation $GT = TH$ n'est pas assurée dans tous les cas. En particulier son existence semble difficile à justifier de manière théorique dans le cas de la caméra Compton. On peut alors se questionner sur la possibilité réelle d'augmenter la résolution des images par déconvolution, comme initialement proposé en TEP dans les travaux de A.J. Reader ([63]). L'autre difficulté vient du fait que la PSF image est adaptative, sa forme varie d'un pixel à l'autre. L'approche est cependant souvent utilisée en imagerie nucléaire et a donné des bons résultats pour la caméra Compton dans [33] et [29]. Nous reviendrons sur sa mise en œuvre dans la section dédiée à l'algorithme MLEM.

6.3 Reconstruction tomographique itérative par l'algorithme LM-MLEM

Les algorithmes itératifs laissent plus de souplesse dans la prise en compte du modèle physique d'acquisition et donnent actuellement des meilleurs résultats. Plusieurs équipes s'y sont intéressées, dont celles de M. Rafecas (Espagne, puis Allemagne), W. Enghart

(Allemagne), J. Fessler, A. Zoglauer, J. Polf (États Unis) ainsi que des équipes de Corée et Japon.

Deux classes d'algorithmes suscitent de l'intérêt actuellement. La première classe est celle de l'algorithme Maximum Likelihood Expectation Maximization (MLEM) qui cherche à déterminer le vecteur $\boldsymbol{\lambda}$ le plus à même d'avoir produit les données n_i définies dans l'équation (6.14), sous une hypothèse d'émission qui suit une loi de Poisson. L'autre, est celle de l'algorithme nommé Stochastic Origin Ensemble (SOE) qui a été proposé plus récemment par A. Sitek [77] et adapté à la caméra Compton par Andreyev *et al* dans [4]. Appartenant à la famille des méthodes bayésiennes, SOE fournit une estimation Monte-Carlo de la distribution a posteriori du nombre de photons émis $\mathbf{f}$, à la différence de MLEM qui permet de calculer seulement sa moyenne $\boldsymbol{\lambda}$. D'après les auteurs, l'algorithme assure aussi une convergence plus rapide que MLEM.

6.3.1 Le maximum de vraisemblance en imagerie d'émission

La fonction de log-vraisemblance sous une hypothèse d'émission suivant une loi de Poisson est :

$$\ell(\boldsymbol{\lambda}|\mathbf{n}) = -\sum_{i=1}^{I}\sum_{j=1}^{J} t_{ij}\lambda_j + \sum_{i=1}^{I} n_i \ln\left(\sum_{j=1}^{J} t_{ij}\lambda_j\right) - \sum_{i=1}^{I} \ln(n_i!). \tag{6.19}$$

Ses dérivées partielles sont

$$\frac{\partial \ell}{\partial \lambda_k}(\boldsymbol{\lambda}|\mathbf{n}) = -s_k + \sum_{i=1}^{I} t_{ik} \frac{n_i}{\sum\limits_{j=1}^{J} t_{ij}\lambda_j} \tag{6.20}$$

et son gradient

$$\nabla \ell(\boldsymbol{\lambda}|\mathbf{n}) = -s + T^* \left[\frac{\mathbf{n}}{T\boldsymbol{\lambda}}\right]. \tag{6.21}$$

Les éléments de la matrice hessienne $\mathscr{H}_\ell(\boldsymbol{\lambda}|\mathbf{y})$ de ℓ sont données par

$$\frac{\partial^2 \ell}{\partial \lambda_j \partial \lambda_k}(\boldsymbol{\lambda}|\mathbf{n}) = -\sum_{i=1}^{I} \frac{t_{ij}t_{ik}n_i}{\mu_i^2} \tag{6.22}$$

et pour tout $\mathbf{u} \in \mathbb{R}^J$,

$$\mathbf{u}^T \mathscr{H}_\ell(\boldsymbol{\lambda}|\mathbf{n})\mathbf{u} = -\sum_{i=1}^{I} \left(\frac{\sqrt{n_i}}{\mu_i}\sum_{j=1}^{J} t_{ij}u_j\right)^2 \leq 0, \tag{6.23}$$

qui montre que ℓ est concave. Son opposée est convexe et coercive, ce qui implique l'existence d'un maximum de la log-vraisemblance (et par monotonie du logarithme, d'un maximum de vraisemblance). Quand $\operatorname{rang}(T) = J$, l'opérateur T est injectif et la matrice hessienne de ℓ est définie négative. L'argument du maximum est alors unique. Quand $\operatorname{rang}(T) < J$, la hessienne est semi-définie négative et le maximum peut être atteint en plusieurs points. Le maximum de vraisemblance

$$\arg\max_{\boldsymbol{\lambda}\in\mathbb{R}^J} \ell(\boldsymbol{\lambda}|\mathbf{n}), \tag{6.24}$$

peut alors avoir de multiples solutions.

On parle de mode liste lorsque le regroupement des données se fait dans des éléments si petits que chacun contient au plus un évènement. On peut pour cela échantillonner selon le temps en plus de l'échantillonnage induit par les détecteurs. Dans ce cas on a toujours $n_i = 1$ car on retient seulement les éléments non nuls.

6.3.2 Convergence de l'algorithme MLEM

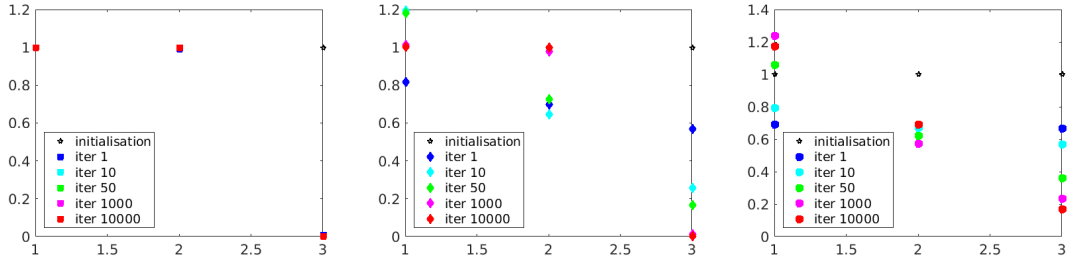
L'usage d'un algorithme itératif s'impose pour déterminer une solution. L'algorithme Expectation Maximization (EM) dont les bases ont été posées par Dempster *et al* dans [12] est souvent préféré. Il présente l'avantage de conserver la positivité des estimations au cours des itérations et de converger relativement vite. L'adaptation au modèle Poissonien des données d'imagerie γ et TEP a été publiée par Lange et Carson [41] et par Vardi *et al* [91]. La preuve de sa convergence a été donnée pour ce cas particulier par A.N. Iusem [28], la preuve initialement publiée dans le cas général étant erronée. Cette preuve concerne le cas des solutions multiples ou situées sur le bord du domaine de définition. D'autres tentatives ont été publiées auparavant, mais à notre connaissance soit elles restent incomplètes soit elles sont plus compliquées (voir en ce sens la discussion dans l'article de A.N. Iusem [28]).

L'algorithme consiste à choisir une valeur initiale $\lambda^{(0)} \in (\mathbb{R}_+^*)^J$ puis, pour $n \in \mathbb{N}$, de calculer la $(n+1)^{\text{ième}}$ estimation du maximum de vraisemblance selon la formule

$$\lambda_j^{(n+1)} = \lambda_j^{(n)} \frac{1}{s_j} \sum_{i=1}^{I} t_{ij} \frac{n_i}{\sum_{k=1}^{J} t_{ik} \lambda_k^{(n)}}. \tag{6.25}$$

En mode liste on remplace $n_i = 1$. D'après [28], en cas de solutions multiples, l'algorithme converge vers une solution parmi celles qui possèdent le plus petit nombre de composantes nulles, correspondant à l'absence d'activité radioactive. Notons que (6.25) définit un algorithme de point fixe, dont l'analyse pourrait se faire à l'aide des outils élémentaires à partir de la matrice de Jacobi. Des tentatives ont été faites en ce sens dans la littérature, mais ne semblent pas avoir été remarquées et la preuve que nous avons pu trouver était incomplète donc difficile à analyser.

Dans la pratique, l'algorithme peut converger numériquement vers une fausse solution. Considérons l'exemple simple suivant. Nous avons pris $\boldsymbol{\lambda} = (1, 1, 0)$ et nous avons défini la matrice système à partir de la convolution par une gaussienne de moyenne nulle et variance $\sigma \in \{0.3, 1, 2\}$. Nous sortons ainsi du cadre de la tomographie, cependant l'algorithme (6.25) est aussi appliqué à la déconvolution d'images sous le nom de Richardson-Lucy. Pour toutes les valeurs de σ la matrice carrée T était inversible. Les données ont été calculées selon la formule $\mathbf{n} = T\boldsymbol{\lambda}$, ce qui fait que mathématiquement la solution exacte existe et est unique, égale à $\boldsymbol{\lambda}$. Nous sommes donc dans un cas où l'algorithme devrait converger vers cette solution. Nous avons fait tourner l'algorithme sur 10000 itérations à partir de la valeur initiale $(1, 1, 1)$. Les résultats pour quelques itérations sont montrés



(a) $\sigma = 0.3$ (b) $\sigma = 1$ (c) $\sigma = 2$

FIGURE 6.6 – Illustration de la convergence lente de l'algorithme MLEM pour des matrices système mal conditionnées. Lorsque σ augmente de (a) à (c), le conditionnement de T augmente de 1.09 à 9.30 puis 138.85. Dans le dernier cas (c), après 40 et jusqu'à 10^4 iterations l'algorithme MLEM semble converger numériquement vers une solution "lisse" qui ne correspond pas à la solution (unique dans ce cas) du maximum de vraisemblance.

dans la figure 6.6. On peut observer la convergence très rapide pour la plus petite valeur de l'écart-type, et la convergence numérique vers une fausse solution pour la plus grande valeur.

Nous avons tenté d'expliquer ce comportement par le conditionnement du système, qui prend les valeurs $c(T) \in \{1.09,\ 9.30,\ 136.85\}$ et se dégrade lorsque σ augmente. Cependant cette façon de mesurer la qualité de la solution numérique ne semble pas convenir : dédoubler une ligne du système dégrade le conditionnement, mais ne devrait pas changer le résultat théorique. Dans le même sens, nous avons constaté qu'un élargissement de la surface de la caméra Compton, qui conduit à une meilleure estimation de la source, produit aussi une dégradation du conditionnement. Par contre, la distance de Kullback-Leibler souvent utilisée de manière équivalente à la fonction de log-vraisemblance et qui devrait décroître vers zéro est un meilleur indicateur. Ses valeurs respectives selon les valeurs croissantes de σ sont 10^{-11}, 3×10^{-8} et 3×10^{-6}.

L'algorithme MLEM déconvolue par étapes, en ajoutant petit à petit les hautes fréquences. Plus le filtre est lissant, plus l'évolution de la solution est lente. On peut tenter une interprétation à partir de la formule (6.25) que nous réécrivons sous la forme

$$\boldsymbol{\lambda}^{(n+1)} = \frac{\boldsymbol{\lambda}^{(n)}}{\mathbf{s}} \left(T^* \left[\frac{\mathbf{n}}{T\boldsymbol{\lambda}^{(n)}} \right] \right). \tag{6.26}$$

Dans le cas de la convolution par la gaussienne on a $T^* = T$. A chaque itération, les erreurs relatives $\mathbf{n}/(T\boldsymbol{\lambda}^{(n)})$ sont lissées avant d'être utilisées pour la mise à jour de la solution. Plus l'opérateur T^* est lissant, plus la convergence sera lente.

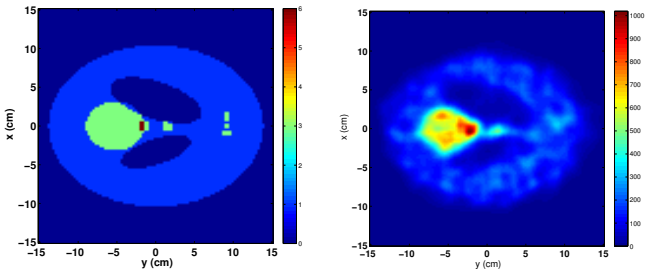
6.3.3 Implémentation de l'approche PSF dans l'algorithme MLEM

Lorsque les données sont bruités et que le modèle de bruit est connu comme dans l'équation (6.17), et lorsque ce modèle admet une représentation équivalente de la forme

(6.18) avec H la fonction d'étalement du point dans l'image, l'algorithme MLEM devient :

$$\boldsymbol{\lambda}^{(n+1)} = \frac{\boldsymbol{\lambda}^{(n)}}{\mathbf{s}} \left(H^* T^* \left[\frac{\mathbf{n}}{T(H\boldsymbol{\lambda}^{(n)})} \right] \right). \tag{6.27}$$

Quand la PSF est à symétrie radiale, on a $H^* = H$. Dans ce cas, deux convolutions s'ajoutent à chaque itération MLEM. Dans le cas contraire, on a une convolution et une corrélation. Une itération consiste en une convolution entre la PSF et la solution actuelle $\boldsymbol{\lambda}^{(n)}$, une retroprojection des erreurs, une convolution/correlation de l'image des erreurs et une multiplication du résultat par $\boldsymbol{\lambda}^{(n)}/\mathbf{s}$. Le coût de ces opérations est faible comparé au coût d'une itération pour la version list-mode de MLEM. Cette approche est très répandue en TEP et SPECT, et plus de détails peuvent être trouvés dans les travaux de A.J. Reader (voir par exemple [63]) et pour la caméra Compton dans [33, 29]. La difficulté de l'estimation de la matrice T est alors transférée à l'estimation de la PSF image. Comme nous l'avons expliqué précédement, cette étape est importante car l'utilisation d'une PSF mal adaptée conduit à des oscillations autour des discontinuités ([81]). Yuemeng Feng, étudiante dont je co-dirige la thèse avec David Sarrut, étudie actuellement cette piste par comparaison avec la méthode où les incertitudes sont modélisées dans la matrice système que nous avions déjà programmée [A12, A13]. Un exemple d'image reconstruite extrait de [A12] est donné dans la figure 6.7.



(a) Référence (b) Reconstruite

FIGURE 6.7 – Fantôme Shepp-Logan reconstruite avec l'algorithme LM-MLEM [A12].

En résumé, l'algorithme (LM)-MLEM converge toujours vers une solution du maximum de vraisemblance mais cette solution pourrait ne pas être unique. L'algorithme converge lentement quand l'opérateur de projection agit comme un filtre passe-bas et est connu pour donner des résultats bruités dans les autres cas surtout lorsque le nombre de particules détectées est faible. Dans les régions d'émission uniforme, on peut observer des spots d'activité plus intense qui se forment au fil des itérations. Ils sont dus à la non-linéarité de l'algorithme, au fait que les voxels sont liés non pas par des relations de voisinage mais par leur supposée contribution aux mêmes évènements acquis, et aussi par le mauvais conditionnement de la déconvolution suite à un lissage. La solution la plus simple consiste à arrêter les itérations selon des a priori utilisateur ou selon des protocoles établis. Un post-lissage peut aussi être appliqué. Cependant il semble plus judicieux d'ajouter de l'information a priori sous la forme d'une pénalité dans la fonction coût à minimiser (au sens où maximiser la log-vraisemblance équivaut à minimiser son opposée). Nous en reparlerons dans la section §6.5.

6.4 Illustration en imagerie γ-prompt

L'imagerie γ-prompt pour le contrôle de l'irradiation en protonthérapie a été le fil directeur des thèses de M. Frandes, X. Lojacono et E. Hilaire. Au cours de la première nous avons commencé a étudier la transformée de Radon conique (dont une forme particulière que nous avons appelé transformée Compton) et les algorithmes d'inversion analytiques. M. Frandes a aussi étudié l'apport d'une régularisation de type L_1 par seuillage des coefficients d'ondelettes. La thèse de X. Lojacono a été l'occasion d'étudier un modèle mathématique de l'acquisition de photons et de chercher des pistes pour la discrétisation et l'implémentation numérique de ces modèles dans la matrice système. Ayant une formation en génie biomédical, E. Hilaire a choisi d'aller le plus loin possible vers l'application médicale. La première étape en ce sens a été de modifier notre façon de réaliser les simulation Monte Carlo comme nous l'avons brièvement présenté dans la section 5.3. Elle a ensuite cherché à évaluer les performances atteignables par l'imagerie gamma-prompt avec une caméra Compton ressemblant au prototype envisagé par l'IPNL et à identifier des pistes d'amélioration. L'idée de départ venait d'une séries d'articles de J. Verburg (Harvard Medical School and Massachusetts General Hospital, Boston) concernant les propriétés du spectre gamma-prompt. Nous en faisons ci-après un bref résumé.

Le spectre gamma-prompt est perçu comme continu, avec quelques lignes proéminentes. L'observation intéressante de ces travaux est que le spectre évolue spatialement en fonction de la profondeur du faisceau et qu'il a une certaine signature en fin de parcours. La fin de parcours (end-of-range) noté R_{80} est définie comme le point où la courbe de la dose déposée descend en dessous de 80% de son maximum. Par exemple, lorsqu'un volume d'eau est irradié avec un faisceau de protons, les lignes les plus fortes dans le spectre expérimental (voir figure 6.8) sont à 6.13 MeV, produite par la transition $^{16}\mathrm{O}(\mathrm{p},\mathrm{p}')\,^{16}\mathrm{O}^*$, à 4.44 MeV, produite par la transition $^{16}\mathrm{O}(\mathrm{p},\mathrm{p}'\alpha)\,^{12}\mathrm{C}^*$ et à environ 5.2 MeV, combination probable des lignes dues aux transitions $^{16}\mathrm{O}(\mathrm{p},\mathrm{pp})\,^{15}\mathrm{N}^*$ (5.27 MeV) et $^{16}\mathrm{O}(\mathrm{p},\mathrm{x})\,^{15}\mathrm{O}^*$ (5.24 MeV). Le spectre change ensuite radicalement derrière la fin de parcours et ces lignes ne sont plus visibles. D'où l'idée d'identifier la fin du parcours grâce au spectre mesuré avec un détecteur collimaté. De plus, les lignes du spectre sont révélatrices de la composition chimique du tissu traversé, autre information qui peut être utile pour le contrôle.

Un faisceau émis par un cyclotron est typiquement fractionné en paquets d'ions émis à des intervalles réguliers. La figure 6.9 montre que ces lignes sont visibles dans le spectre pendant le cycle d'émission en dépit du fond dû aux neutrons.

La figure 6.10 montre la coordination dans l'évolution en profondeur de ces émissions spécifiques avec la dose. Cette dernière est caractérisée en hadron-thérapie par un plateau où les tissus sont relativement peu endommagés suivie d'une augmentation assez rapide jusqu'à un maximum et une chute brutale d'intensité. Le pic d'intensité porte le nom de pic de Bragg et les tissus situés derrière celui-ci sont presque entièrement épargnés. L'émission gamma-prompt suit les évolutions de la dose et la figure montre que l'émission totale caractérise moins bien la fin de parcours que les émissions spécifiques aux lignes majeures. On parle généralement de corrélation entre le profil gamma-prompt et le profil

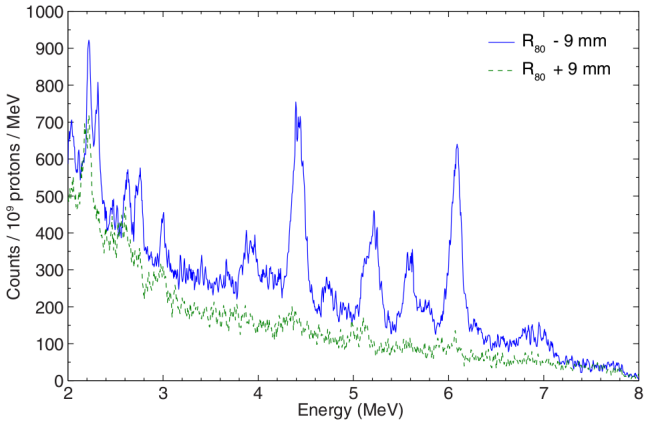


FIGURE 6.8 – Spectre énergétique du rayonnement gamma-prompt intégré sur une fenêtre temporelle de 2 ns, mesuré 9 mm avant et 9 mm derrière la fin du parcours notée R_{80} (figure extraite de [92]).

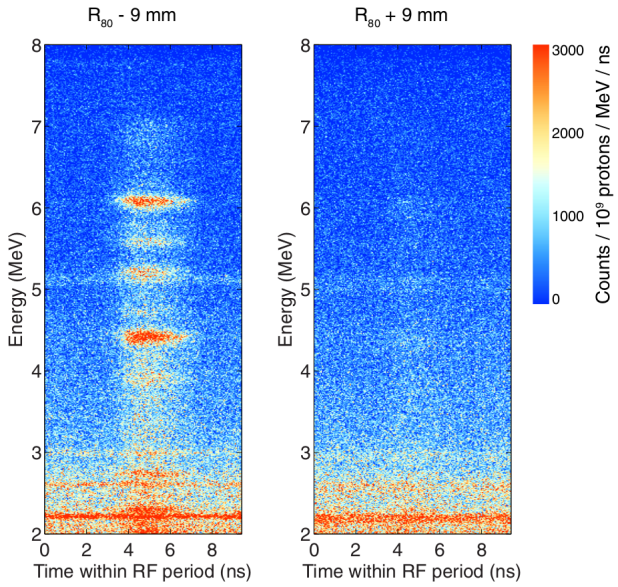


FIGURE 6.9 – Histogrammes temps-énergie, measurées 9 mm avant R_{80} et 9 mm derrière R_{80} pour un faisceau proton et un volume d'eau (figure extraite de [92]).

de dose, même si ce terme n'est pas très adapté et que la relation exacte entre les deux reste pour le moment méconnue.

En identifiant avec précision le spectre énergétique de l'émission gamma-prompt il serait aussi possible d'identifier la composition chimique et par conséquent la nature des tissus traversés par un faisceau. La figure 6.11 montre le profil en profondeur pour deux matériaux : eau et polyéthylène, contenant les deux isotopes les plus répandus dans les tissus humains, l'oxygène et le carbone.

Dans sa thèse, Estelle Hilaire a cherché à reproduire par simulation certaines des observations de Verburg et al, en 3D, et d'analyser la réponse de la caméra Compton à ces

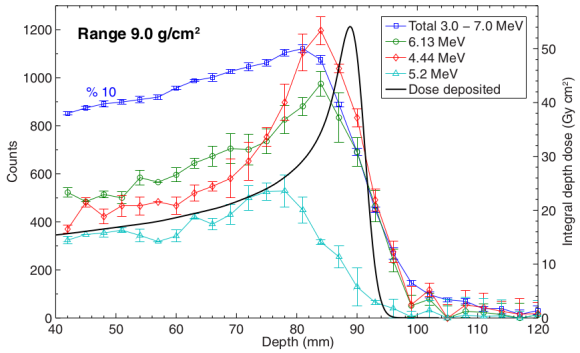


FIGURE 6.10 – Distribution en profondeur des émissions gamma-prompt cumulé et discrètes comparées au profil de dose (ligne noire) (figure extraite de [92]).

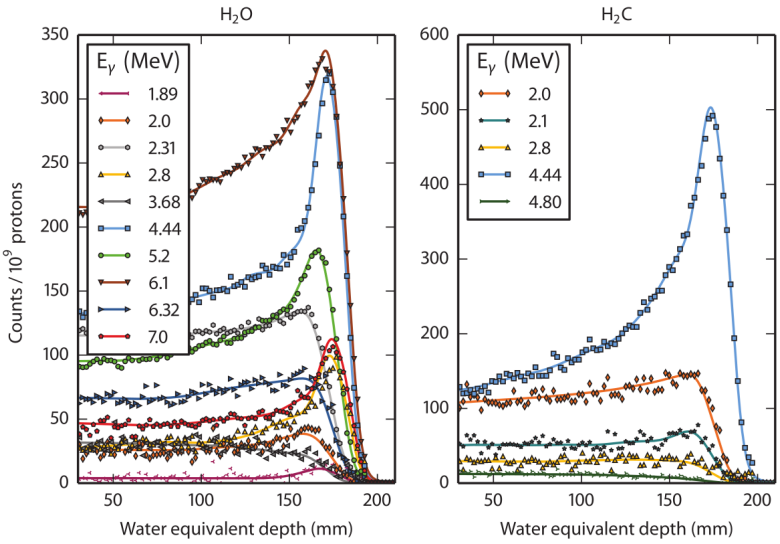


FIGURE 6.11 – Profils mesurés d'émission gamma-prompt à différentes énergies pour un faisceau de protons irradiant un volume d'eau (à gauche) et un volume de polyéthylène (à droite) (figure extraite de [93]).

énergies spécifiques. La résolution énergétique de la caméra étant attendue inférieure au détecteur collimaté utilisé par Verburg et al, elle a choisi d'étudier des gammes d'énergie de largeur 0.5 MeV, entre 0.1 et 10 MeV. La figure 6.12 confirme que la distribution spatiale pour certaines de ces gammes présente un pic vers la fin du parcours ce qui est favorable pour la reconstruction et pour l'identification du R_{80}.

Les distributions simulées ont été utilisées ensuite pour générer les données caméra Compton. Pour chaque distribution, un volume de $20 \times 8 \times 2.5$ cm^3 divisé en $100 \times 41 \times 13$ voxels et centré sur la ligne du faisceau, a été reconstruit avec l'algorithme itératif LM-MLEM. Des coupes centrales extraites des images 3D reconstruites avec 15 itérations sont représentées en figure 6.13. Les échelles de couleurs ne sont pas affichées et diffèrent d'une coupe à l'autre. Cette image est obtenue en sélectionnant les photons dont l'énergie a été presque entièrement déposée dans le détecteur. Une telle sélection n'est pas possible

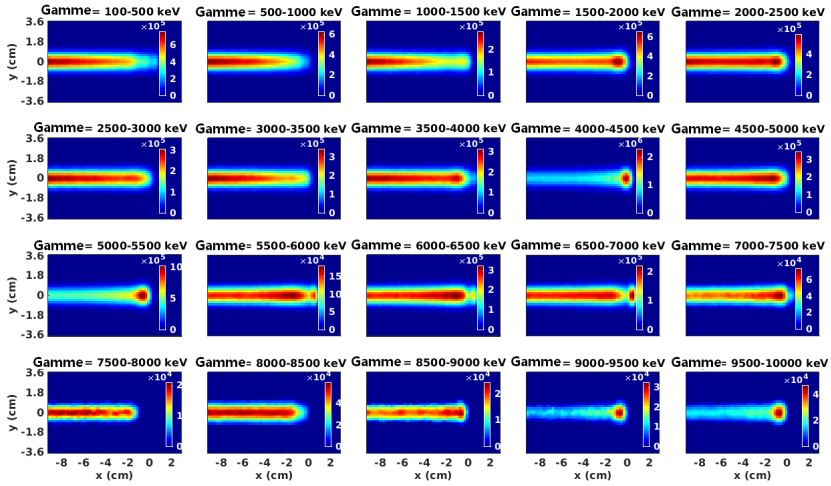


FIGURE 6.12 – Distributions gamma-prompt $\lambda(\cdot, \varepsilon_k)$ obtenues à partir de l'irradiation d'un fantôme d'eau avec un faisceau de protons d'énergie 120 MeV. Fenêtres d'énergie entre 100 keV et 10 MeV échantillonnées avec un pas de 500 keV. Coupes centrales de 2 mm d'épaisseur centrées en $z = 0$ cm.

pour une vraie acquisition, d'où la nécessité de disposer d'un détecteur ayant une bonne capacité d'absorption et très précis. Les images sans cette sélection sont montrées dans la thèse et sont bien moins précises.

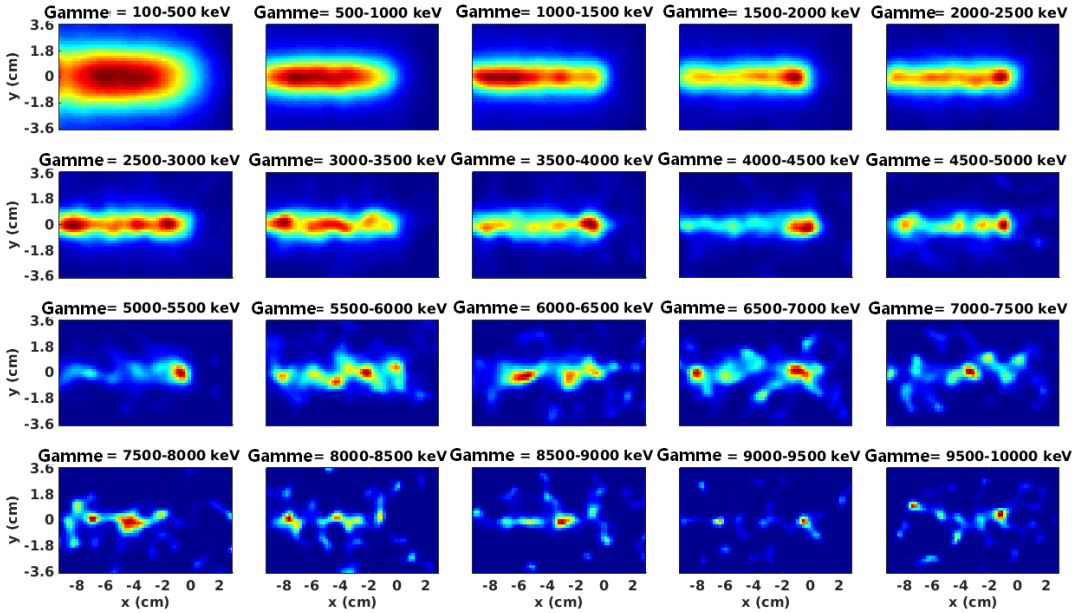
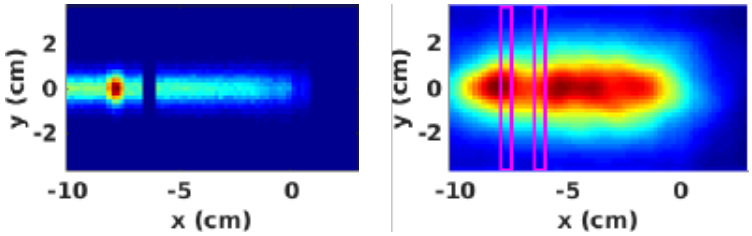


FIGURE 6.13 – Images reconstruites des distributions gamma-prompt $\widehat{\lambda}(\cdot, \varepsilon_k)$ après sélection des évènements ayant déposé presque toute leur énergie dans le détecteur. Voir figure 6.12 en comparaison.

L'analyse de la structure spatio-énergétique de l'émission γ-prompt, son lien avec les performances de l'algorithme de reconstruction et l'influence sur la détection du pic de Bragg ont été présentées à la conférence IEEE NSS/MIC en 2014 [C6] et ont été



(a) Référence (b) Reconstruite

FIGURE 6.14 – Image simulée et image reconstruite de l'émission γ-prompt produite par un faisceau de protons traversant un volume d'eau avec un insert d'os et un insert d'air (images extraites de la thèse d'E. Hilaire).

publiés dans [A13]. En dépit de quelques résultats prometteurs concernant la possibilité d'identifier des écarts avec le plan de traitement, la qualité des images reste en dessous des attentes, surtout pour un passage au patient et à une composition hétérogène du corps irradié, comme nous avons pu le vérifier par simulation dans la thèse d'Estelle Hilaire. Nous avons cependant identifié quelques uns des points pouvant être améliorés. À part la reconstruction, les propriétés de détection de la caméra dont l'absorption totale sont très importantes. Aussi, il serait utile d'étudier quelle pré-sélection réalisable en pratique sur les données pourrait améliorer les images.

6.5 Techniques de régularisation

L'algorithme MLEM est un algorithme de point fixe dérivé d'un modèle statistique poissonien sur les données. Il intègre naturellement la contrainte de positivité sur les voxels de l'image, contrairement aux autres méthodes de reconstruction pour lesquelles cette contrainte s'obtient par projection de la solution ou de l'itérée sur l'ensemble d'images positives. Lorsqu'on le compare à d'autres algorithmes algébriques, on observe une convergence plus rapide et une meilleure robustesse aux données aberrantes. D'après [32], ceci s'explique par le fait que la formule (6.25) peut être vue comme une méthode de gradient pré-conditionnée,

$$\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n)} + \boldsymbol{\alpha} \nabla \ell(\boldsymbol{\lambda}^{(n)} | \mathbf{n}), \tag{6.28}$$

avec $\boldsymbol{\alpha} = \boldsymbol{\lambda}^{(n)}/\mathbf{s}$. En contre-partie, il a tendance à produire des zones d'accumulations d'intensité dans des régions où l'intensité devrait être uniforme. En effet, la variance de la loi de Poisson étant égale à sa moyenne, la variance dans l'image reconstruite n'est pas uniforme et augmente avec la concentration d'émetteurs. Ce résultat n'est pas tellement lié à l'algorithme, mais plutôt à la solution du maximum de vraisemblance elle-même [75]. Le signal se retrouve ainsi noyé dans le bruit et au minimum il est nécessaire d'interrompre les itérations bien avant la convergence et de lisser le résultat. L'inconvénient de cette approche est d'abord qu'elle nécessite de choisir a priori le nombre d'itérations, puis d'induire un écart supplémentaire avec les données lors de l'étape de lissage.

Au lieu de chercher une estimation de $\boldsymbol{\lambda}$ sous la forme d'un maximum de la vraisemblance ℓ, il est préférable de calculer le maximum a posteriori (MAP) associé à une densité a priori dont la log-vraisemblance est une fonction $-G$. De manière équivalente, on cherche une solution du minimum de

$$F = -\log \ell(\cdot|\mathbf{n}) + G, \tag{6.29}$$

où G est supposée continue et convexe, mais pas nécessairement lisse. Numériquement, cet ajout d'a priori peut se faire parfois en modifiant la formule du MLEM, parfois en insérant une étape de lissage entre deux pas MLEM. Il a été montré dans [22] qu'une modification mineure de la formule (6.25) permet de calculer l'estimateur MAP pour G lisse :

$$\lambda_j^{(n+1)} = \lambda_j^{(n)} \frac{1}{s_j + \alpha \partial G / \partial \lambda_j(\boldsymbol{\lambda})} \sum_{i=1}^{I} t_{ij} \frac{n_i}{\sum_{k=1}^{J} t_{ik} \lambda_k^{(n)}}, \tag{6.30}$$

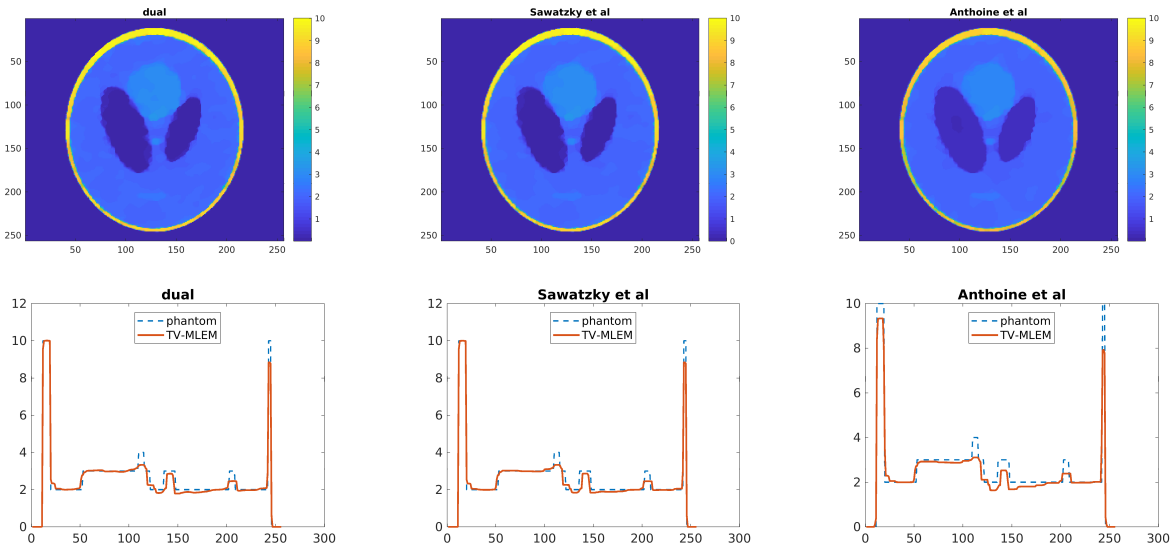
où α est un paramètre de relaxation. Cette formule a été utilisée dans le cas non-lisse aussi (a priori de type variation totale) avec une approximation lisse du critère [56, 61]. Ce schéma est cependant de type explicite et donc instable, nécessitant un choix de la valeur de α très faible. La deuxième approche a été discutée par Silverman et al dans [75] pour un lissage quadratique. Les auteurs sont parvenus a relier l'étape intermédiaire de lissage à un critère pénalisé modulo quelques adaptations. Ces adaptations consistent à abandonner l'invariance spatiale habituelle des critères car les images MLEM n'ont pas des propriétés de biais et variance uniformes.

Nuyts et Fessler [53] comparent les résultats de deux méthodes : reconstruction MLEM suivie d'un lissage adaptatif et critère pénalisé par un a priori quadratique pondéré par les données. Ils concluent en trouvant des avantages aux deux méthodes. La première est la plus simple à appliquer en clinique et la plus rapide. Elle donne aussi des résultats légèrement meilleurs que la seconde à condition que le filtre soit bien choisi - et la PSF est assez éloignée du modèle classique de la gaussienne et très dépendante des conditions d'acquisition. Elle nécessite aussi la convergence du MLEM avant d'appliquer le lissage. Quand l'acquisition introduit un floutage des données, le nombre d'itérations acquises peut être de l'ordre de quelques centaines. L'utilisation d'une régularisation quadratique pondérée par les données est numériquement plus complexe mais peut donner des bons résultats. La pondération est imposée par le fait que la fonction de log-vraisemblance n'est pas invariante par translation. Une pénalité invariante par translation, c'est-à-dire qui ne tient pas compte du contenu de l'image, va conduire à un biais et à une variance variables spatialement.

Comme le soulignent Stayman et Fessler [80], les approches pénalisées peuvent être utilisées aussi comme technique d'accélération de la convergence car la pénalité améliore le conditionnement du problème. De plus, le résultat d'un tel algorithme est à la fois en bonne adéquation avec les données (selon le premier critère, la fonction de vraisemblance) et compatible avec les connaissances a priori dont nous disposons.

L'inconvénient des méthodes basées sur une pénalité quadratique est qu'elles lissent trop les images aux frontières des objets. C'est pourquoi nous avons choisi d'utiliser un a

priori de type variation totale, et ce choix n'est pas nouveau dans la littérature. L'originalité de notre approche consiste dans le choix de la méthode numérique de résolution, basée sur une décomposition EM (Expectation-Maximization, comme dans MLEM) et un algorithme dual pour l'étape M. L'étape E reste la même que pour MLEM, ce qui au final donne un algorithme où une solution intermédiaire est calculée à chaque pas par la formule MLEM et cette solution intermédiaire est ensuite lissée pour satisfaire à la condition de variation totale. À noter que le lissage ne peut pas se faire avec les algorithmes classiques, construits pour du débruitage ([9]) ou pour un critère d'attache aux données $\|Ax - y\|$ quadratique ([8]), car on se retrouverait de nouveau avec une non-uniformité du biais et de la variance dans l'image. Intuitivement, il faudrait plutôt utiliser une approche avec des poids variables comme dans [32] et (6.28). C'est en résumé et de façon intuitive l'approche choisie par Sawatzky et al [67]. Nous avons proposé un nouvel algorithme, spécialement conçu pour le critère de log-vraisemblance (qui correspond aussi à une distance de Kullback-Leibler) et nous avons démontré mathématiquement la convergence globale de l'algorithme. Comparé aux autres travaux de la littérature, notre méthode est plus rapide qu'une méthode standard en optimisation convexe non-lisse (par exemple l'algorithme Chambolle-Pock [10] utilisé dans [5]) de par l'inclusion naturelle de la contrainte de positivité et de l'approche EM, stable (contrairement à des approches implicites [56, 61] ou semi-explicites [96, 47]) et convergente.



(a) Méthode duale (b) Méthode Sawatzky et al. (c) Méthode Anthoine et al.

FIGURE 6.15 – Comparaison entre la méthode duale proposée, celle de Sawatzky et al. et celle d'Anthoine et al. après 1000 itérations MLEM.

Dans la figure 6.15 on peut voir le résultat de la reconstruction du fantôme de Shepp-Logan à partir de 36 projections (une projection tous les 5°) avec la méthode duale, la méthode Sawatzky et al, et la méthode Anthoine et al. Après 1000 itérations les images reconstruites et les profils sont presque identiques. Dans la figure 6.16 on montre la décroissance de la fonction-coût et de l'erreur quadratique moyenne pour les trois méthodes. On remarque la décroissance plus rapide des algorithmes qui utilisent les étapes

EM. La description de la méthode et d'autres exemples sur des données de type TEP 2D simulées à partir de la fantôme de Shepp-Logan sont donnés en annexe (article soumis).

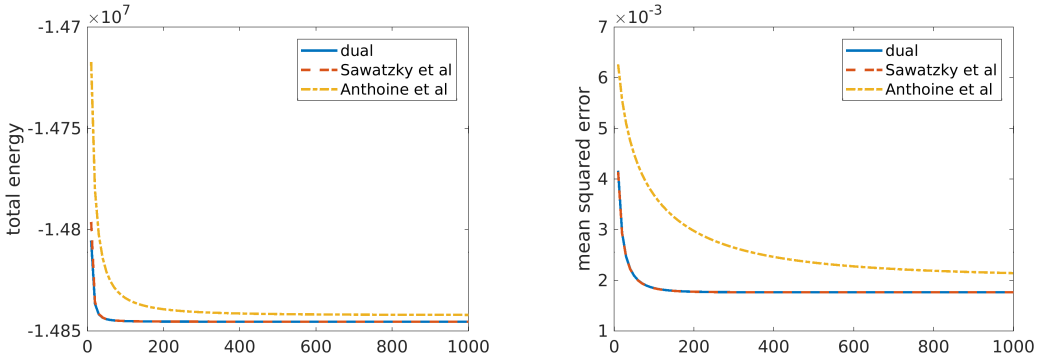


FIGURE 6.16 – Comparaison en termes de minimisation du critère et de l'erreur quadratique moyenne entre la méthode duale proposée (ligne bleue), la méthode de Swatzky et al [67] (ligne rouge) et celle de Anthoine et al. [5] (ligne jaune).

L'application de cette méthode à la reconstruction d'images caméra Compton est en cours d'étude. Des difficultés sont apparues liées à l'indisponibilité d'une estimation précise du vecteur sensibilité et à la difficulté du choix de la constante multipliant la fonction pénalité. Ces difficultés sont amplifiées par les temps de calcul importants requis par l'algorithme MLEM dans le cas de la caméra Compton. Un résumé a été présenté à la conférence IEEE NSS-MIC 2018 [C8].

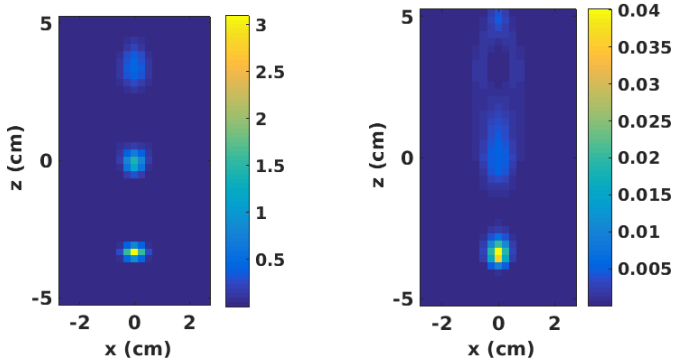
6.6 Comparaison analytique-itératif

Une comparaison de la méthode analytique et de la méthode itérative LM-MLEM a été faite à partir de simulations pour une source ponctuelle et pour le prototype développé à l'IPNL. L'étude a été publiée dans [A9]. Cette étude montre que la méthode itérative est plus robuste vis-à-vis du bruit statistique venant d'acquisitions avec un faible nombre d'évènements. Des travaux que nous avons menés depuis nous permettent d'aller plus loin dans cette analyse.

Nous avons ainsi observé que l'algorithme itératif donne de meilleurs résultats pour des données bruitées et des faibles statistiques, mais cela peut venir en partie d'un lissage dû à un arrêt prématuré de l'algorithme. Le lissage des images analytiques fait apparaître les sources qui auparavant étaient complétement noyées dans le bruit.

L'algorithme itératif peut fonctionner en mode liste contrairement à l'analytique, ce qui lui évite d'ajouter des erreurs par regroupement de données dans des classes, sur des caractéristiques moyennes. Les erreurs sur les valeurs des paramètres angulaires (inclinaison de l'axe, α, et demi-angle d'ouverture du cône, β) ont des conséquences importantes sur l'image reconstruite et ceci indépendamment de l'algorithme considéré. Dans la figure 6.17 on montre des images reconstruites avec l'algorithme MLEM (standard, non pas en mode

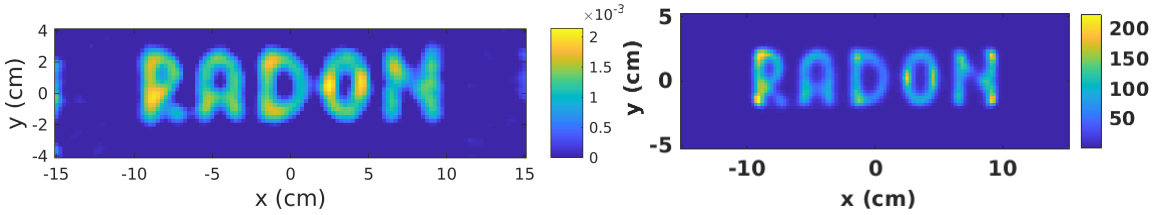
liste), à partir de projections déterministes non bruitées de trois sources ponctuelles dans un volume d'air. Dans (a) les projections sont simulées avec $\alpha = 2.5°$ et utilisées comme si elles avaient été mesurées à $\alpha = 0°$. Dans (b), les projections avec $\beta = 10°$ et $\beta = 15°$ sont sommées puis l'image est reconstruite comme si l'angle de diffusion avait été $\beta = 12.5°$. Dans les deux cas, les intensités relatives des sources sont différentes et le flou augmente lorsqu'on s'éloigne du détecteur. L'erreur sur β semble la plus pénalisante.



(a) (b)

FIGURE 6.17 – Image reconstruite avec l'algorithme MLEM à partir de projections avec (a) une erreur de 2.5° sur l'inclinaison de l'axe α et (b) une erreur de 2.5° sur l'angle de diffusion Compton β. Dans les deux cas, ving itérations ont été effectuées.

Un autre argument en faveur des algorithmes itératifs c'est qu'ils permettent de prendre en compte les incertitudes de mesure et le bruit statistique. Ils donnent l'image la plus probable compte tenu des données, ce qui fait que les voxels sont liés entre eux non pas par leur relation de voisinage mais par leur rapport aux mêmes évènements. En contre-partie, on perd la linéarité : imager simultanément deux sources n'est pas équivalent à sommer les images acquises séparément.



(a) (b)

FIGURE 6.18 – Comparaison d'images reconstruites avec (a) l'algorithme analytique et (b) l'algorithme MLEM.

En contre-partie, une méthode analytique sera toujours largement plus rapide, fournira une image dans laquelle les propriétés du bruit seront plus uniformes, et une telle méthode est linéaire. Pour diminuer le bruit on pourra soit introduire un filtre passe-bas en complément du filtre de la reconstruction, soit post-lisser l'image [A16]. La figure 6.18 montre des coupes de volumes reconstruits avec les deux méthodes. Le nombre d'évènements est 1.7×10^7 pour le premier et 8×10^5 pour le second. Le temps de calcul est de l'ordre de la

minute pour le premier et d'un peu moins de 4 heures pour le second, avec parallélisation sur 32 cœurs (80 heures sans parallélisation).

En conclusion, les méthodes analytiques restent probablement le meilleur choix pour des acquisitions abondantes en données et avec des détecteurs performants. Les méthodes itératives avec critère pénalisé par des a priori sur l'image sont a privilégier dans les autres cas.

6.7 Développement logiciel

Le logiciel CORESI inclut plusieurs fonctionnalités liées à l'imagerie caméra Compton. Il implémente les algorithmes analytique et itératifs MLEM et LM-MLEM mais aussi des procédures pour la simulation de projections déterministes. Le développement de ce logiciel se poursuit avec la thèse de Yuemeng Feng qui modélise le problème de reconstruction sous la forme d'un problème inverse contraint afin de pouvoir ajouter de l'information a priori sur l'objet imagé. Plus précisément, des modules pour la reconstruction basée PSF (point spread function) et pour la régularisation par variation totale ont été ajoutés et sont testés. De cette façon nous espérons réduire l'influence de la géométrie d'acquisition incomplète et du faible nombre de photons détectés. Depuis peu, après intervention des informaticiens du laboratoire, nous disposons d'une version parallélisée du code LM-MLEM. Le logiciel CORESI est mis à disposition de nos collègues de CREATIS et de l'Institut de Physique Nucléaire de Lyon.

Chapitre 7

Imagerie tomographique en microscopie électronique de transmission

Mon intérêt pour cette thématique est relativement récent. Suite à des discussions avec des collègues du laboratoire des Sciences des Matériaux MATEIS de l'INSA de Lyon, nous avons soumis un premier projet dans le cadre de l'appel BQR INSA. Notre tâche dans ce projet consistait à proposer des méthodes basées sur le traitement des images visant d'une part à faciliter leur analyse de données (par la sélection automatique des images nettes d'une séquence tomographique) et, d'autre part, à contribuer à l'amélioration des images reconstruites (en dé-floutant des images moyennement floues). Un projet ANR a été obtenu qui a permis l'embauche d'un post-doctorant et de deux stagiaires de niveau Master.

7.1 Contexte scientifique

L'étude 3D à l'échelle du nanomètre d'échantillons soumis à des sollicitations extérieures, comme la contrainte, la température ou des stimuli électriques, ouvre la voie d'une meilleure compréhension des phénomènes dynamiques au sein de la matière condensée. Pour atteindre cette résolution, la microscopie électronique en transmission (MET) est une technique incontournable. Grâce aux développements récents du mode dit "environnemental", on peut aujourd'hui suivre l'évolution des matériaux sous l'influence de la température et de l'environnement gazeux, donc appréhender des phénomènes physico-chimiques, tels des transformations de phase ou des réactions chimiques en milieu réactif.

Le microscope électronique en transmission environnemental du CLYM (Consortium Lyon - St-Etienne de Microscopie) permet d'observer l'échantillon dans un milieu gazeux tout en réalisant une acquisition tomographique sur une plage angulaire d'environ 140°. Pour d'autres microscopes, qui fonctionnent avec une enceinte sous vide, il est possible d'utiliser une cellule environnementale et de faire circuler un flux gazeux entre l'extérieur et

l'échantillon. La cellule environnementale étant plus encombrante qu'un porte-échantillon classique, elle conduit à une limitation plus importante de l'angle d'inclinaison au cours de l'acquisition tomographique.

L'acquisition environnementale impose des contraintes de temps plus strictes liées à l'évolution de l'échantillon suite aux réactions chimiques. Pour cette raison l'acquisition doit se faire le plus rapidement possible. L'idée naturelle consiste alors à remplacer l'acquisition pas-à-pas par une acquisition en rotation continue. Les problèmes typiques de la tomographie sont alors amplifiés :

— Au vu de la vitesse de rotation, l'intervention manuelle de l'utilisateur pour replacer le faisceau sur la région d'intérêt n'est plus envisageable. Afin d'automatiser la procédure, un code informatique capable de suivre en temps réel l'échantillon et de prévoir ses déplacements est nécessaire.
— Les différents déplacements de l'échantillon et du faisceau, en rotation et en translation, ainsi que les instabilités mécaniques font que certaines images sont floues. Ces images doivent être retirées de la séquence et éventuellement dé-floutées.
— L'acquisition avec une caméra rapide, qui par ailleurs réduit l'influence du flou, conduit à des séries de centaines voire de milliers d'images. Celles-ci doivent être alignées par rapport à l'axe de rotation commun. Cette opération est à présent très fastidieuse malgré l'aide logicielle disponible.
— Plus l'échantillon évolue vite en conditions environnementales, ou même sous l'effet du faisceau, plus l'acquisition devra être rapide et éventuellement à dose réduite. Le bruit dans les projections et l'angle manquant dégradent les volumes reconstruits et imposent l'ajout d'a priori dans l'algorithme de reconstruction.

7.2 Sélection automatique des images floues et correction de flou

Dans son projet de fin d'études d'ingénieur INSA, Khanh Tran a étudié des critères de classification d'images selon leur netteté et a proposé un algorithme permettant de retirer les images floues de la séquence par une méthode d'apprentissage. L'utilisateur doit classer quelques images. Des caractéristiques comme : le contraste, l'information mutuelle, la moyenne, l'écart-type, le nombre de points-clé identifiés par l'algorithme SIFT, des paramètres de texture, sont utilisées pour séparer les classes et pour évaluer les images restantes. Un premier tri des caractéristiques a été fait sur la base des courbes ROC (Receiver Operator Curve). Plusieurs classifieurs ont ensuite été testés et des combinaisons prometteuses de caractéristiques ont été mises en évidence.

La correction de flou a été le sujet du stage de master de Yuemeng Feng. Elle a testé des méthodes analytiques directes (filtrage de Wiener, filtrage Constrained Least Squares) et des méthodes itératives dont FISTA. Cependant dans notre cas le filtre de convolution n'est pas connu. Les méthodes testées (la méthode de Matlab, la méthode de Shan et al [71]) n'ont pas donné de bons résultats. Son choix s'est porté sur une déconvolution

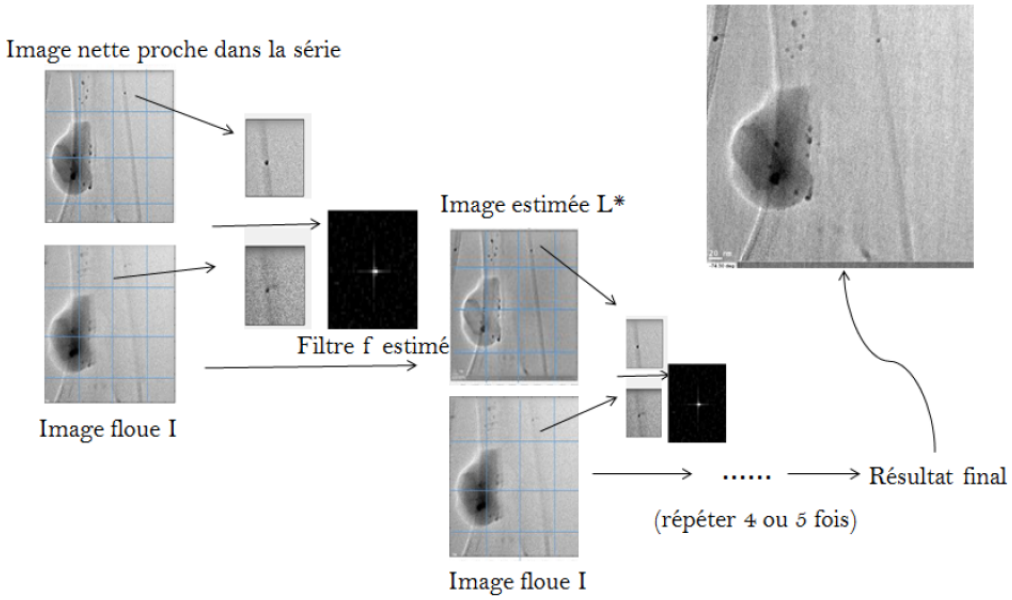


FIGURE 7.1 – Procédure de défloutage semi-aveugle d'images d'une série tomographique en microscopie électronique (Master Y. Feng).

semi-aveugle, où le filtre de convolution était estimé sur une partie de l'image floue à l'aide d'une image nette proche dans la séquence (voir figure 7.1). Ces résultats ont été présentés sous forme d'affiche au Congrès Européen de Microscopie de 2016 [D3].

7.3 Alignement d'images et correction automatique de déplacement

La projection de l'objet au cours de la rotation se déplace dans l'image acquise par la caméra. On observe une translation dans le plan xy du goniomètre mais il ne semble pas y avoir de rotation notable autour d'un axe parallèle au faisceau. L'estimation du déplacement entre deux images consécutives peut fournir des paramètres pour régler le microscope : en mode pas-à-pas et après chaque acquisition, on pourrait appliquer au porte-objet le même déplacement mais dans le sens inverse. De plus, comme le parcours du porte-objet semble reproductible (voir figure 7.2), on peut envisager pour le mode continu, le calcul prévisionnel de la dérive puis le déplacement du faisceau en parallèle de l'acquisition. Ce travail est poursuivi par Thomas Grenier (MdC INSA, CREATIS) en collaboration avec nos collègues Thierry Epicier et Lucian Roiban du laboratoire Mateis et une présentation à ce sujet a été faite au 19ème congrès international de microscopie IMC'19 [D6] en 2018.

L'alignement automatique des projections est un enjeu important en tomographie électronique. Des algorithmes très perfectionnés existent, et pourtant cette opération reste largement manuelle et demande un temps et un effort considérables. Pour comprendre, prenons l'exemple de la figure 7.3. Lorsqu'on utilise la méthode la plus répandue de mise

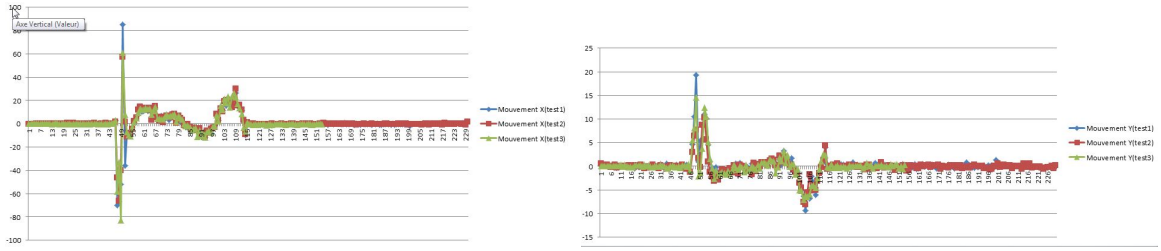


FIGURE 7.2 – Déplacements de l'échantillon lors de trois acquisitions successives identiques.

en correspondance, celle basée sur le calcul de la corrélation, on trouve dans l'image de droite plusieurs rectangles qui correspondent au rectangle sélectionné à gauche.

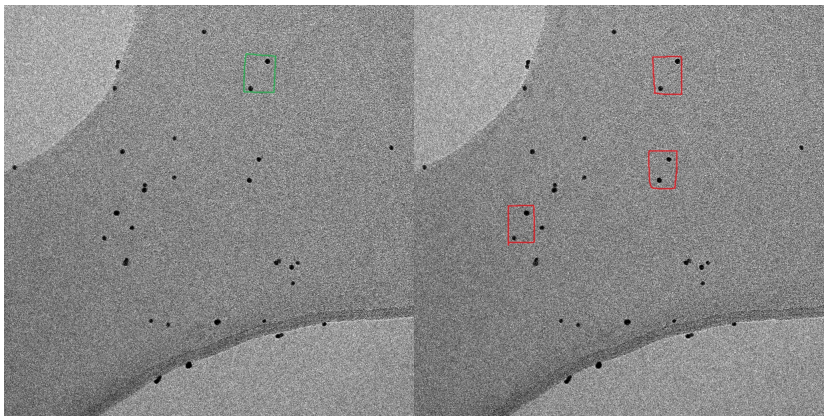


FIGURE 7.3 – Mise en correspondance par correlation de deux images de microscopie électronique. (Projet de fin d'études de K. Tran).

Dans cette image la mise en correspondance est basée sur des amers placés dans l'échantillon, en occurrence des billes d'or. L'utilisation des ces amers sphériques facilite la tâche de mise en correspondance, comme le montre la figure 7.4 où la méthode ORB basée sur des descripteurs et points-clés a été utilisée. Cependant ils ne peuvent pas être utilisés sur tous les échantillons et dans certains cas ils nuisent à l'image quand l'objet a un contraste bien plus faible.

La séquence tomographique nécessite un pré-alignement très précis avant la reconstruction de l'image. Cet alignement peut se faire par plusieurs méthodes. La méthode classique nécessite le calcul de la corrélation croisée. Khanh Tran a aussi testé des méthodes plus récentes : SIFT, SURF et ORB. Les résultats sont encourageants et montrent que l'association des points clés a été faite correctement. Mais ce travail n'a pas abouti à une prise en compte de la particularité de la séquence, celle d'être une séquence d'images tomographiques.

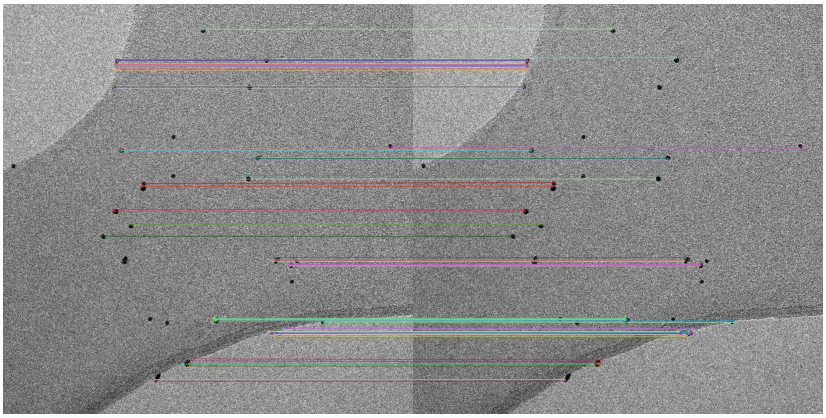


FIGURE 7.4 – Mise en correspondance par descripteurs et points-clé de deux images de microscopie électronique. (Projet de fin d'études de K. Tran).

7.4 Reconstruction tomographique sous contrainte de variation totale

L'ajout de régularité améliore le conditionnement du problème de reconstruction tomographique en microscopie électronique aussi, et permet de réduire les artefacts d'angle manquant. Dans son article [A15], Hussein Banjak, post-doctorant dans le projet 3DCLEAN, a proposé l'algorithme SIRT-FISTA-TV qui est une technique de reconstruction algébrique régularisée. Les méthodes algébriques représentent le problème de reconstruction sous la forme d'un système linéaire $\mathbf{p} = A\mathbf{f}$, où $\mathbf{p}$ est le vecteur de données de projection, A est la matrice système et $\mathbf{f}$ est le vecteur contenant le volume à reconstruire. Ces méthodes supposent implicitement un modèle de bruit gaussien et cherchent à résoudre le système en minimisant l'erreur ℓ_2. Autrement dit, on cherche une solution de

$$\arg\min_{\mathbf{f}} \|A\mathbf{f} - \mathbf{p}\|_2, \tag{7.1}$$

avec A généralement de grande taille. Comme illustré dans la figure 7.5, le processus SIRT-FISTA-TV consiste en trois grandes étapes : i) itération SIRT (Simultaneous Iterative Reconstruction Technique), ii) débruitage TV, iii) mise à jour du volume selon la technique d'accélération FISTA.

L'algorithme SIRT [17] est souvent utilisé en reconstruction tomographique et a tendance à être plus stable au bruit comparé à ART [19]. En partant d'une image initiale $f^{(0)}$, qui peut être égale à zéro partout, on itère selon la formule :

$$\mathbf{f}^{(k)} = \mathbf{f}^{(k-1)} - \lambda \cdot \frac{1}{A^*\mathbf{1}} A^* \left[\frac{A\mathbf{f}^{(k-1)} - \mathbf{p}}{A\mathbf{1}} \right]. \tag{7.2}$$

Ici λ est un paramètre de relaxation. Comme illustré dans la figure 7.5, chaque itération SIRT consiste en trois phases : projection de l'estimation pour produire les projections synthétisées, calcul de la différence entre les projections acquises et les projections synthétisées, retro-projection pondérée de l'erreur pour mettre à jour le volume. SIRT est donc

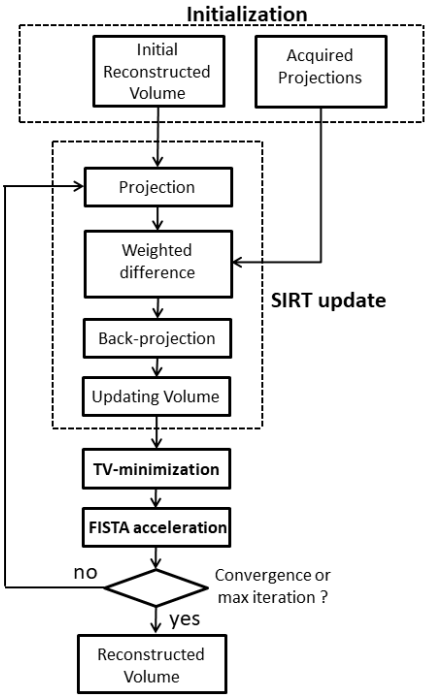


FIGURE 7.5 – Schéma de l'algorithme SIRT-FISTA-TV.

une version pondérée de l'algorithme de descente du gradient pour la minimisation de $\|A\mathbf{f} - \mathbf{p}\|_2$ dont l'expression est :

$$\mathbf{f}^{(k)} = \mathbf{f}^{(k-1)} - \lambda A^*(A\mathbf{f}^{(k-1)} - \mathbf{p}). \tag{7.3}$$

L'ajout d'une régularisation TV permet de réduire le bruit tout en conservant les contours. Cette technique, initialement proposée pour le débruitage par Rudin *et al* [65], est très utilisée en tomographie et permet de réduire de manière significative les artefacts quand le nombre de projections est faible et quand l'angle de vue est réduit [73]. Le problème régularisé est :

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}\in\mathbb{R}^N} \left\{ \|A\mathbf{f} - \mathbf{p}\|_2^2 + \mu\|\mathbf{f}\|_{TV} \right\}, \tag{7.4}$$

où le paramètre $\mu \geq 0$ contrôle le degré de lissage et $\|\mathbf{f}\|_{TV}$ est la norme l_1 des magnitudes des gradients de l'image $|\nabla f|_{i,j,k}$. La pénalité est exprimée par :

$$\|\mathbf{f}\|_{TV} = \sum_{i,j,k} \sqrt{(f_{i,j,k} - f_{i-1,j,k})^2 + (f_{i,j,k} - f_{i,j-1,k})^2 + (f_{i,j,k} - f_{i,j,k-1})^2} \tag{7.5}$$

où $f_{i,j,k}$ est la valeur du voxel d'indice (i, j, k).

Nous avons utilisé le schéma proximal forward-backward nommé FISTA [8] permettant de résoudre

$$\min_{\mathbf{f}\in\mathbb{R}^N} \{\phi(\mathbf{f}) + \psi(\mathbf{f})\}, \tag{7.6}$$

avec ψ une fonction continue et convexe pas nécessairement lisse et ϕ une fonction continument différentiable ayant un gradient Lipschitz de constante $L(\phi)$. Dans notre

cas, $\phi(\mathbf{f}) = \|A\mathbf{f} - \mathbf{p}\|_{l_2}^2$ et $\psi(\mathbf{f}) = \mu\|\mathbf{f}\|_{TV}$. Le nom FISTA signifie Fast ISTA : FISTA est une accélération de ISTA. ISTA résout (7.6) par une technique de splitting. À chaque itération, on effectue un pas de gradient sur la fonction lisse ϕ,

$$\mathbf{g}^{(k)} = \mathbf{f}^{(k-1)} - \frac{1}{L}\nabla\phi(\mathbf{f}^{(k-1)}), \tag{7.7}$$

où $L > 0$ est le pas de gradient, puis on cherche dans le voisinage de ce nouveau point une meilleure estimation de (7.6) qui réduit aussi la valeur de ψ. Cette nouvelle estimation est donnée par :

$$\mathbf{f}^{(k)} = \underset{\mathbf{f}\in\mathbb{R}^N}{\arg\min} \left\{ \psi(\mathbf{f}) + \frac{L}{2}\|\mathbf{f} - \mathbf{g}^{(k)}\|_2^2 \right\}. \tag{7.8}$$

Avec les notations de l'analyse convexe, $\mathbf{f}^{(k)} = \text{prox}_{\psi/L}(\mathbf{g}^{(k)})$. Dans SIRT-FISTA-TV le pas de gradient (7.7) est remplacé par une itération SIRT (7.2) avec $\lambda = 1/L$ et $L \geq L(\phi)$, la norme spectrale de A^*A. La version publiée utilise une approximation lisse de la norme TV et un algorithme de descente de gradient pour résoudre (7.8). Une meilleure approche consiste à utiliser l'algorithme de Chambolle [9], cependant les résultats numériques sont très proches dans les deux cas.

L'accélération FISTA s'obtient en remplaçant $\mathbf{f}^{(k-1)}$ dans (7.7) par une combinaison linéaire des deux dernières estimations $\bar{\mathbf{f}}^{(k-1)}$ et $\bar{\mathbf{f}}^{(k-2)}$, où par $\bar{\mathbf{f}}^{(k)}$ on note maintenant la solution de (7.8). La nouvelle valeur de $\mathbf{f}^{(k)}$ est :

$$\mathbf{f}^{(k)} = \bar{\mathbf{f}}^{(k)} + \left(\frac{t^{(k-1)} - 1}{t^{(k)}} \right) \left(\bar{\mathbf{f}}^{(k)} - \bar{\mathbf{f}}^{(k-1)} \right), \tag{7.9}$$

où $\bar{\mathbf{f}}^{(k)}$ est l'image obtenue après la minimisation TV dans l'itération (k). Le paramètre $t^{(k)}$ est mis à jour dans chaque itération selon la formule

$$t^{(k)} = \frac{1 + \sqrt{1 + 4\left(t^{(k-1)}\right)^2}}{2}, \tag{7.10}$$

avec $t^{(0)} = 1$. Avec FISTA, le nombre total d'itérations nécessaires pour produire un résultat de qualité similaire est réduit comparé à ISTA et donc à SIRT-TV. Comme il a été montré dans [8], la complexité de FISTA est de l'ordre $O(1/k^2)$ (avec k le nombre d'itérations) comparé à la complexité $O(1/k)$ que ISTA peut avoir dans les cas les plus défavorables.

Les réductions de bruit et d'artefacts rendues possibles par cet algorithme permettent de mieux identifier des nano-particules dans des images de microscopie et de mieux évaluer leur taille y compris lorsqu'on acquiert peu d'images et que certaines sont floues. Les images débruitées sont plus faciles à analyser, segmenter et quantifier. Un exemple en est donné sur la figure 7.6.

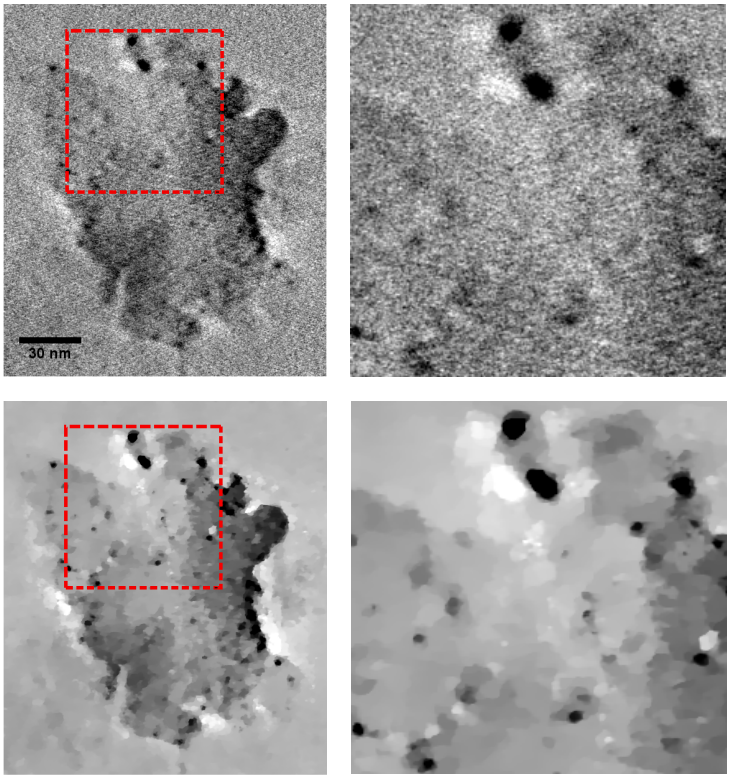


FIGURE 7.6 – Reconstruction d'un échantillon Pd/Al_2O_3 avec SIRT (ligne en haut) et SIRT-FISTA-TV (ligne en bas). Une coupe XY de chaque volume reconstruit est montrée dans la première colonne et un agrandissement de la région marquée par un rectangle rouge est montré dans la seconde colonne.

Troisième partie

Conclusions et perspectives

Chapitre 8

Conclusions et perspectives

8.1 Imagerie camera Compton

Depuis mon recrutement à CREATIS, mes recherches se sont articulées autour de la reconstruction tomographique. Pour la modalité caméra Compton, assez peu répandue et non commercialisée, je m'intéresse à la modélisation du problème direct, à la simulation de données, aux méthodes de reconstruction directes et itératives ainsi qu'à leur mise en place numérique. Mon travail dans cette thématique a été motivé par la perspective de la construction d'un centre d'hadron-thérapie à Lyon, puis par les développements instrumentaux pour la construction d'une caméra à l'Institut de Physique Nucléaire de Lyon. J'ai co-encadré quatre thèses dans cette thématique, financées par une allocation de thèse, un projet européen, une bourse INSIS et une bourse CSC du gouvernement chinois. Cette thématique pourrait se poursuivre par des études concernant l'imagerie nucléaire thérapeutique, qui intéresse le Centre Léon Bérard et la partie de l'équipe construite autour de David Sarrut, hébergée dans le centre.

La caméra Compton peine à convaincre le milieu médical de l'imagerie nucléaire SPECT, en dépit d'une sensibilité estimée à près de 100 fois supérieure aux caméras Anger classiquement utilisées. Les arguments avancés sont les conditions d'utilisation de ces détecteurs (par exemple, le germanium nécessite un refroidissement très important et d'être mis sous vide, ce qui conduit à un encombrement important), le coût, la gamme d'énergie optimale qui commence à environ 3-400 keV alors que les isotopes actuellement utilisés émettent des photons γ d'énergies autour de 100 keV.

À partir des années 2010, de grands espoirs ont été mis dans l'imagerie caméra Compton qui pouvait en théorie donner accès à une carte 3D de la zone irradiée pendant la proton ou hadron-thérapie. Ces deux modalités de stérilisation des tumeurs avec des faisceaux d'ions sont d'une précision supérieure aux radio-thérapies classiques et présentent une meilleure efficacité biologique pour des tumeurs radio-résistantes. Plusieurs équipes ont mené des investigations et ont montré que cette modalité pouvait donner des renseignements sur le traitement, du moins dans des conditions un peu idéalisées, avec des cibles homogènes et des temps de traitement assez longs. Mais les résultats qui sont publiés à ce jour ont fait

quelque peu retomber l'enthousiasme initial. En dépit des efforts et des simplifications rendues possibles par les simulations, les images reconstruites présentent des artefacts, une résolution spatiale très faible, assez peu de ressemblance à l'image de référence [54, 68]. Les espoirs des entreprises se tournent vers d'autres modalités, moins exigeantes, moins coûteuses mais conduisant néanmoins à des signaux 1D. La caméra Compton peut fournir cette information 1D et permet d'observer des écarts dans la direction du faisceau [A13],[25]. Mais elle peut faire davantage, car elle offre des images en 3D.

En 2008 nous avons commencé à travailler sur l'imagerie Compton et à évaluer son potentiel pour l'imagerie γ-prompt. Nous utilisions alors pour nos simulations et pour la reconstruction d'images un logiciel initialement développé pour des applications en astronomie. Nous avons présenté les premiers résultats à la conférence NSS-MIC de 2009. En 2010 nous avons publié un article. Grâce à ces travaux nous avons pu participer ensuite au projet européen ENVISION (2010-2014). Nous avons montré que la caméra Compton présentait un intérêt certain, mais nous avons aussi constaté qu'assez peu de résultats existaient concernant la reconstruction d'images. Nous avons depuis mieux compris la nature des données, développé un nouvel algorithme de reconstruction, développé des codes pour la reconstruction analytique et pour la reconstruction itérative MLEM (Maximum Likelihood Expectation Maximization), mis en place des outils pour la simulation de l'acquisition caméra Compton d'un traitement. Les premiers travaux que nous avons publiés sur la reconstruction analytique dans la revue Inverse Problems ont probablement participé à l'initiation de travaux théoriques sur ce qu'on appelle maintenant la transformée de Radon conique.

Je continue à croire que tout le potentiel de la caméra Compton n'a pas été exploré et exploité. Des nouveaux détecteurs apparaissent sur le marché, comme les détecteurs en germanium de très petite taille, qui ont un refroidissement électronique et offrent une très bonne résolution spatiale et énergétique [45]. Des configurations originales sont imaginées pour la SPECT clinique, comme par exemple le détecteur plan configuré comme un plateau d'échec ([59], présentation invitée à la conférence NSS-MIC 2016 à Strasbourg). Nous nous sommes aussi aperçus [A13], et ces résultats viennent d'être confirmés par d'autres équipes [68, 25] que la capacité d'absorption totale des photons en imagerie γ-prompt est un facteur essentiel. On observe clairement également les limites de la géométrie d'acquisition. Au départ, l'objectif était de pouvoir réaliser l'image à l'aide d'une seule caméra, de faible étendue (2-9 cm). Il apparaît clairement, au vue des études menées, qu'aucun algorithme actuellement implémenté ne parvient à donner des résultats convenables face au mauvais conditionnement du problème inverse à résoudre, mauvais conditionnement induit par la troncation très sévère des projections Compton. L'ajout d'a priori sous forme de régularisation peut améliorer les résultats. Il faut cependant avoir à l'esprit qu'une image plus belle n'est pas toujours synonyme d'un meilleur diagnostique. Il se peut que les améliorations purement algorithmiques ne soient pas suffisantes. Une autre approche consiste à chercher des géométries d'acquisition plus proches des conditions de complétude. Très peu d'études ont été faites pour le moment en ce sens [78, 39]. On peut mentionner également les enjeux du temps réel, la prise en compte des paramètres expérimentaux réalistes (filtrage rapide des données brutes). L'ajout d'un hodoscope faisceau ([13])

pourrait améliorer les résultats, en participant à la sélection des coïncidences vraies.

Une utilisation de la caméra Compton en *médecine nucléaire* pourrait permettre de réduire la dose radioactive administrée au patient ou encore une meilleure exploitation des radionucléides émettant à la fois des positons et des rayons γ. Pour le patient, ce serait une avancée comparable à l'apparition des caméras TEP 3D, non collimatées. Bien que plusieurs modalités d'imagerie soient à l'étude pour la vérification du traitement *en proton-thérapie*, aucune n'a été pour l'instant validée sur patient. La vision 3D de la caméra Compton reste un atout très important pour observer les effets de l'irradiation sur des cibles hétérogènes. Elle peut en théorie observer et identifier des sources d'énergies distinctes, ce qui permettrait d'identifier aussi la nature des tissus traversés pour la comparer avec l'image de l'examen radiologique.

8.2 Tomographie rapide en microscopie électronique de transmission

La nano-tomographie électronique connait aujourd'hui un fort engouement en matière d'amélioration de la résolution spatiale dans la perspective d'une vision 3D à l'échelle atomique. Pour autant, peu d'efforts sont faits actuellement sur l'amélioration de la résolution temporelle, un des verrous limitant le suivi d'objets en conditions dynamiques, qui est naturellement la voie pour étudier non plus des objets inertes, mais des objets en évolution dans des conditions proches de leur utilisation réelle. C'est notamment le cas des applications en catalyse, où des expériences quasiment en vraie grandeur peuvent être reproduites à l'intérieur du microscope.

Mon travail sur cette thématique a débuté assez récemment et les premiers résultats que nous avons obtenus datent de l'été 2016, grâce au financement INSA à travers le BQR SPEE3D et au travail des deux stagiaires que j'ai co-encadrés avec mon collègue T. Grenier. Nous collaborons étroitement avec nos collègues du laboratoire MATEIS de l'INSA. La proximité des deux laboratoires et du microscope facilitent les discussions et l'avancement du travail. Le code de reconstruction tomographique développé dans le cadre du post-doctorat de Hussein Banjak commence à être utilisé par nos collègues et devrait être transféré également au IPCMS de Strasbourg, autre partenaire du projet.

Les modalités de poursuite de ce projet ne sont pas définies à présent et dépendent des opportunités et des possibilités de financement. Plusieurs thématiques de collaboration ont été identifiées, certaines directement liées à la tomographie rapide, d'autres indépendantes. On peut par exemple mentionner les travaux liés au suivi en temps réel de l'échantillon et à la reproductibilité du mouvement ([D6]). Concernant la tomographie, il me semble intéressant de travailler sur l'alignement des projections, dans une approche intégrante reconstruction tomographique - alignement. Cet alignement peut être vu comme une simple translation ou plus généralement comme un problème d'identification d'angle de projection et translation, dans le cas des objets se déplaçant librement dans un volume.

8.3 Perspectives générales

L'essentiel de mes travaux de recherche à l'INSA Lyon ont porté sur la reconstruction caméra Compton, qui est un dispositif d'imagerie pour la SPECT, non collimaté et acceptant des énergies supérieures aux caméras SPECT cliniques (caméras Anger). L'application en vue était initialement l'imagerie gamma-prompt pour le contrôle en ligne de la proton-thérapie. À présent, nous investiguons aussi son applicabilité en imagerie nucléaire thérapeutique. Depuis peu, j'ai commencé à m'intéresser à des modalités d'imagerie tomographique standard. De nombreux problèmes intéressants apparaissent dans ces domaines du fait des avancées technologiques qui élargissent le champ des expériences possibles. On peut mentionner les détecteurs rapides et les problèmes d'alignement, la possibilité de réaliser des acquisitions parcimonieuses pour réduire la dose absorbée par l'échantillon ou les tissus, ainsi que l'imagerie spectrale.

L'ajout d'a priori semble d'une importance cruciale en reconstruction tomographique, dès lors que l'on considère des projections tronquées et des faibles doses. Cet a priori vient d'une expérience précédente, et s'exprime souvent sous la forme d'une contrainte d'uniformité dans l'image (image lisse, formée d'objets distincts à l'intérieur desquels l'intensité varie peu). Mathématiquement, on pénalisera les différences entre les voxels voisins, avec des fonctions-pénalité assez simples pour permettre leur étude et l'implémentation numérique. On utilise souvent un critère quadratique ou de type variation totale. La tendance actuelle est de transférer ce processus d'apprentissage aux machines, ce qui permettrait d'élargir la classe de contraintes vers des fonctions plus complexes sans expression analytique. Ce nouveau degré de liberté est à double tranchant car à défaut d'un ensemble d'apprentissage suffisamment exhaustif, les résultats risquent d'être trop proches des entrées de l'apprentissage.

Les approches variationnelles sont à présent très utilisées pour résoudre des problèmes inverses avec des contraintes de régularité. Nous les avons utilisées pour imposer une contrainte de type variation totale au problème tomographique avec des données poissoniennes ([U2]). D'autres problèmes liés à la tomographie pourraient être traités avec ce point de vue. On peut mentionner le problème de reconstruction tomographique dynamique ([44]), la segmentation ([72]), l'imagerie spectrale et l'acquisition parcimonieuse pour une réduction de la dose ([74, 36]).

Pour qu'elle soit utile, une méthode doit être relativement facile à implémenter. Par exemple, les algorithmes en tomographie, retro-projection filtrée, ART, SIRT, MLEM, sont tous basés sur des équations très simples. L'algorithme analytique que nous avons proposé pour la reconstruction caméra Compton bénéficie lui aussi de la simplicité de l'approche retro-projection filtrée. L'algorithme itératif, de type MLEM, est plus long en temps de calcul mais a été récemment parallélisé. Ceci nous permettra de tester plus facilement des modèles ainsi que l'ajout de la régularisation sur des données plus représentatives. Les codes que nous développons ont été jusqu'à présent distribués uniquement à des collaborateurs et continuent à évoluer. Afin de les distribuer plus largement, un travail spécifique concernant la facilité d'utilisation et la documentation doit être effectué.

Bibliographie

[1] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, and al. Geant4-A simulation toolkit. Nucl. Instrum. Methods A, 506 :205–303, 2003.

[2] S Aldawood, PG Thirolf, A Miani, M Böhmer, G Dedes, R Gernhäuser, C Lang, S Liprandi, L Maier, T Marinšek, et al. Development of a compton camera for prompt-gamma medical imaging. Radiation Physics and Chemistry, 140 :190–197, 2017.

[3] Pedro Andreo. On the clinical spatial resolution achievable with protons and heavier charged particle radiotherapy beams. Physics in Medicine & Biology, 54(11) :N205, 2009.

[4] A. Andreyev, A. Sitek, and A. Celler. Fast image reconstruction for Compton camera using stochastic origin ensemble approach. Medical Physics, 38 :429, 2011.

[5] Sandrine Anthoine, Jean-François Aujol, Yannick Boursier, and Clothilde Melot. Some proximal methods for Poisson intensity CBCT and PET. Inverse problems and Imaging, 6(4) :565–598, 2012.

[6] R. Basko, G. L. Zeng, and G. T. Gullberg. Application of spherical harmonics to image reconstruction for the Compton camera. Phys. Med. Biol., 43 :887–894, 1998.

[7] R. Basko, G.L. Zeng, and G.T. Gullberg. Fully three dimensional image reconstruction from "V"-projections acquired by Compton camera with three vertex electronic collimation. In IEEE Nuclear Science Symposium, volume 2, pages 1077–1081. IEEE, 1997.

[8] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences, 2(1) :183–202, 2009.

[9] A. Chambolle. An algorithm for total variation minimization and applications. Journal of Mathematical imaging and vision, 20(1-2) :89–97, 2004.

[10] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision, 40(1) :120–145, 2011.

[11] M. J. Cree and P. J. Bones. Towards direct reconstruction from a gamma camera based on Compton scattering. IEEE Transactions on Medical Imaging, 13(2) :398–407, 1994.

[12] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 39 :1–38, 1977.

[13] S Deng, D Dauvergne, GN Lu, H Mathez, and Y Zoccarato. Very fast front end asic associated with multi-anode pmts for a scintillating-fibre beam hodoscope. Journal of Instrumentation, 8(01) :C01047, 2013.

[14] Lucia Florescu, Vadim A Markel, and John C Schotland. Inversion formulas for the broken-ray radon transform. Inverse Problems, 27(2) :025002, 2011.

[15] M. Fontana, D. Dauvergne, J.M. Letang, J.L. Ley, and É. Testa. Compton camera study for high efficiency SPECT and benchmark with Anger system. Physics in Medicine & Biology, 62(23) :8794, 2017.

[16] Mattia Fontana, JL Ley, É Testa, Jean Michel Létang, V Maxim, and D Dauvergne. Versatile Compton camera for high-energy gamma rays : Monte Carlo comparison with Anger camera for medical imaging. Acta Physica Polonica B, 48(10), 2017.

[17] P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. Journal of theoretical biology, 36(1) :105–117, 1972.

[18] C. Golnik, F. Hueso-González, A. Müller, P. Dendooven, W. Enghardt, F. Fiedler, T. Kormoll, K. Roemer, J. Petzoldt, A. Wagner, and G. Pausch. Range assessment in particle therapy based on prompt γ-ray timing measurements. Phys. Med. Biol., 59 :5399–5422, 2014.

[19] Richard Gordon, Robert Bender, and Gabor T Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. Journal of theoretical Biology, 29(3) :471IN1477–476IN2481, 1970.

[20] Jerome Edward Gormley, WL Rogers, NH Clinthorne, DK Wehe, and GF Knoll. Experimental comparison of mechanical and electronic gamma-ray collimation. Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment, 397(2-3) :440–447, 1997.

[21] R. Gouia-Zarrad and G. Ambartsoumian. Exact inversion of the conical Radon transform with a fixed opening angle. Inverse Problems, 30(4) :045007, 2014.

[22] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. IEEE Transactions on Medical Imaging, 9(1) :84–93, Mar 1990.

[23] Li Han, W Leslie Rogers, Sam S Huh, and Neal Clinthorne. Statistical performance evaluation and comparison of a compton medical imaging system and a collimated anger camera for higher energy photon imaging. Physics in Medicine & Biology, 53(24) :7029, 2008.

[24] LJ Harkness, P Arce, DS Judson, AJ Boston, HC Boston, JR Cresswell, J Dormand, M Jones, PJ Nolan, JA Sampson, DP Scraggs, A Sweeney, I Lazarus, and J Simpson. A compton camera application for the gamos geant4-based framework. Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment, 671 :29–39, 2012.

[25] F. Hueso-Gonzalez, G. Pausch, J. Petzoldt, K. E. Roemer, and W. Enghardt. Prompt gamma rays detected with a BGO block Compton camera reveal range deviations of therapeutic proton beams. IEEE Trans. on Nuclear Science, 2017.

[26] F. Hueso-Gonzàlez, C. Golnik, C. M. Berthel, A. Dreyer, W. Enghardt, F. Fiedler, K. Heidel, T. Kormoll, H. Rohling, S. Schone, R. Schwengner, A. Wagner, and G. Pausch. Test of Compton camera components for prompt gamma imaging at the ELBE bremsstrahlung beam. Journal of Instrumentation, 9 :P05002, 2014.

[27] A Iltis, H Snoussi, L Rodrigues de Magalhaes, MZ Hmissi, C Tata Zafiarifety, G Zeufack Tadonkeng, and C Morel. Temporal imaging CeBr3 Compton camera : A new concept for nuclear decommissioning and nuclear waste management. In EPJ Web of Conferences, volume 170, page 06003. EDP Sciences, 2018.

[28] Alfredo N. Iusem. A short convergence proof of the EM algorithm for a specific Poisson model. Brazilian Journal of Probability and Statistics, 6(1) :57–67, 1992.

[29] Meei-Ling Jan, Ming-Wei Lee, and Hsuan-Ming Huang. PSF reconstruction for Compton-based prompt gamma imaging. Physics in Medicine & Biology, 63(3) :035015, 2018.

[30] Sébastien Jan, G Santin, D Strul, Steven Staelens, K Assie, D Autret, S Avner, R Barbier, M Bardies, PM Bloomfield, et al. GATE : a simulation toolkit for PET and SPECT. Physics in Medicine & Biology, 49(19) :4543, 2004.

[31] Shigeto Kabuki, Kaori Hattori, Ryota Kohara, Etsuo Kunieda, Atsushi Kubo, Hidetoshi Kubo, Kentaro Miuchi, Tadaki Nakahara, Tsutomu Nagayoshi, Hironobu Nishimura, et al. Development of electron tracking Compton camera using micro pixel gas chamber for medical imaging. Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment, 580(2) :1031–1035, 2007.

[32] L. Kaufman. Implementing and accelerating the EM algorithm for positron emission tomography. IEEE transactions on medical imaging, 6(1) :37–51, 1987.

[33] Soo Mee Kim, Hee Seo, Jin Hyung Park, Chan Hyeong Kim, Chun Sik Lee, Soo-Jin Lee, Dong Soo Lee, and Jae Sung Lee. Resolution recovery reconstruction for a compton camera. Physics in Medicine & Biology, 58(9) :2823, 2013.

[34] A. Kishimoto, J. Kataoka, T. Taya, L. Tagawa, S. Mochizuki, S. Ohsuka, Y. Nagao, K. Kurita, M. Yamaguchi, N. Kawachi, K. Matsunaga, H. Ikeda, E. Shimosegawa, and J. Hatazawa. First demonstration of multi-color 3-D in vivo imaging using ultra-compact Compton camera. Scientific Reports, 7, 2017.

[35] Antje-Christin Knopf and Antony Lomax. In vivo proton range verification : a review. Physics in Medicine & Biology, 58(15) :R131, 2013.

[36] Ville Kolehmainen, Samuli Siltanen, Seppo Järvenpää, Jari P Kaipio, P Koistinen, M Lassas, J Pirttilä, and E Somersalo. Statistical inversion for medical x-ray tomography with few radiographs : II. Application to dental radiology. Physics in Medicine & Biology, 48(10) :1465, 2003.

[37] J Krimmer, D Dauvergne, JM Létang, and É Testa. Prompt-gamma monitoring in hadrontherapy : A review. Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment, 878 :58–73, 2018.

[38] J. Krimmer, J.-L. Ley, C. Abellan, J.-P. Cachemiche, L. Caponetto, X. Chen, M. Dahoumane, D. Dauvergne, N. Freud, B. Joly, D. Lambert, L. Lestand, J.M. Létang, M. Magne, H. Mathez, V. Maxim, G. Montarou, C. Morel, M. Pinto, C. Ray, V. Reithinger, E. Testa, and Y. Zoccarato. Development of a Compton camera for medical applications based on silicon strip and scintillation detectors. Nucl. Instrum. Methods A, 787(0) :98 – 101, 2015.

[39] P. Kuchment and F. Terzioglu. Three-dimensional image reconstruction from Compton camera data. SIAM Journal on Imaging Sciences, 9(4) :1708—-1725, 2016.

[40] S. Kurosawa, H. Kubo, K. Ueno, S. Kabuki, S. Iwaki, M. Takahashi, K. Taniue, N. Higashi, K. Miuchi, T. Tanimori, et al. Prompt gamma detection for range verification in proton therapy. Current Applied Physics, 12(2) :364–368, 2012.

[41] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. J Comput Assist Tomogr, 8(2) :306–316, 1984.

[42] G. Llosa, J. Bernabeu, D. Burdette, E. Chesi, N.H. Clinthorne, K. Honscheid, H. Kagan, C. Lacasta, M. Mikuz, P. Modesto, et al. Last results of a first Compton probe demonstrator. IEEE Transactions on Nuclear Science, 55(3) :936–941, 2008.

[43] Alfred K Louis. Feature reconstruction in inverse problems. Inverse Problems, 27(6) :065010, 2011.

[44] Felix Lucka, Nam Huynh, Marta Betcke, Edward Zhang, Paul Beard, Ben Cox, and Simon Arridge. Enhancing compressed sensing photoacoustic tomography by simultaneous motion estimation. SIAM J. Imaging Sciences, 11(4) :2224–2253, 2018.

[45] V. Marian, J. Clauss, R. Abou-Khalil, J.-O. Beau, B. Pirard, J. Flamanc, and M.-O. Lampert. μGe : a miniature Germanium detector for immediate operation. In IEEE NSS-MIC conference, 2016.

[46] L. Mihailescu, K.M. Vetter, M.T. Burks, E.L. Hull, and W.W. Craig. SPEIR : a Ge Compton camera. Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment, 570(1) :89–100, 2007.

[47] Arthur Mikhno, Elsa D Angelini, Bing Bai, and Andrew F Laine. Locally weighted total variation denoising for ringing artifact suppression in PET reconstruction using PSF modeling. In Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, pages 1252–1255. IEEE, 2013.

[48] Chul-Hee Min, Chan Hyeong Kim, Min-Young Youn, and Jong-Won Kim. Prompt gamma measurements for locating the dose falloff region in the proton therapy. Applied physics letters, 89(18) :183517, 2006.

[49] M Moteabbed, Samuel España, and H Paganetti. Monte Carlo patient study on the comparison of prompt gamma and PET imaging for range verification in proton therapy. Physics in Medicine & Biology, 56(4) :1063, 2011.

[50] S. Motomura, Y. Kanayama, M. Hiromura, T. Fukuchi, T. Ida, H. Haba, Y. Watanabe, and S. Enomoto. Improved imaging performance of a semiconductor Compton camera GREI makes for a new methodology to integrate bio-metal analysis and molecular imaging technology in living organisms. J. Anal. At. Spectrom., 28 :934–939, 2013.

[51] M.K. Nguyen and T.T. Truong. On an integral transform and its inverse in nuclear imaging. Inverse Problems, 18(1) :265, 2002.

[52] M.K. Nguyen, T.T. Truong, and P. Grangeat. Radon transforms on a class of cones with fixed axis direction. Journal of Physics A : Mathematical and General, 38(37) :8003, 2005.

[53] Johan Nuyts and Jeffrey A Fessler. A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution. IEEE transactions on medical imaging, 22(9) :1042–1052, 2003.

[54] P. G. Ortega, I. Torres-Espallardo, F. Cerutti, A. Ferrari, J. E. Gillam, C. Lacasta, G. Llosà, J. F. Oliver, P. R. Sala, P. Solevi, and M. Rafecas. Noise evaluation of Compton camera imaging for proton therapy. Phys. Med. Biol., 60 :1845–63, 2015.

[55] Harald Paganetti. Range uncertainties in proton therapy and the role of monte carlo simulations. Physics in Medicine & Biology, 57(11) :R99, 2012.

[56] V. Y. Panin, G. L. Zeng, and G. T. Gullberg. Total variation regulated EM algorithm. IEEE Transactions on Nuclear Science, 46(6) :2202–2210, Dec 1999.

[57] K Parodi, Wolfgang Enghardt, and T Haberer. In-beam pet measurements of β+ radioactivity induced by proton beams. Physics in Medicine & Biology, 47(1) :21, 2001.

[58] L.C. Parra. Reconstruction of cone-beam projections from Compton scattered data. IEEE Transactions on Nuclear Science, 47(4) :1543–1550, 2000.

[59] G. Pausch, A. Schulz, and W. Enghard. A novel scheme of compton imaging for nuclear medicine. In IEEE NSS-MIC conference, 2016.

[60] I Perali, A Celani, L Bombelli, C Fiorini, F Camera, E Clementel, S Henrotin, G Janssens, D Prieels, F Roellinghoff, et al. Prompt gamma imaging of proton pencil beams at clinical dose rate. Physics in Medicine & Biology, 59(19) :5849, 2014.

[61] M Persson, D Bone, and H Elmqvist. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. Physics in Medicine & Biology, 46(3) :853, 2001.

[62] M Priegnitz, S Barczyk, L Nenoff, C Golnik, I Keitz, T Werner, S Mein, J Smeets, F Vander Stappen, G Janssens, L Hotoiu, F Fiedler, D Prieels, W Enghardt, G Pausch, and C Richter. Towards clinical application : prompt gamma imaging of passively scattered proton fields with a knife-edge slit camera. Physics in Medicine and Biology, 61(22) :7881, 2016.

[63] Andrew J Reader, Peter J Julyan, Heather Williams, David L Hastings, and Jamal Zweit. Em algorithm system modeling by image-space techniques for pet reconstruction. IEEE Transactions on Nuclear Science, 50(5) :1392–1397, 2003.

[64] M-H Richard, M Dahoumane, D Dauvergne, M De Rydt, G Dedes, N Freud, J Krimmer, JM Létang, X Lojacono, V Maxim, et al. Design study of the absorber detector of a compton camera for on-line control in ion beam therapy. IEEE Transactions on Nuclear Science, 59(5) :1850–1855, 2012.

[65] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Physica D : nonlinear phenomena, 60(1-4) :259–268, 1992.

[66] A. C. Sauve, A. O. Hero, W. L. Rogers, S. J. Wilderman, and N. H. Clinthorne. 3D image reconstruction for a Compton SPECT camera model. IEEE Transactions in Nuclear Science, 46(6) :2075–2084, 1999.

[67] A. Sawatzky, C. Brune, F. Wubbeling, T. Kosters, K. Schafers, and M. Burger. Accurate EM-TV algorithm in PET with low SNR. In Nuclear Science Symposium Conference Record, 2008. NSS'08. IEEE, pages 5133–5137. IEEE, 2008.

[68] Sebastian Schöne, Wolfgang Enghardt, Fine Fiedler, Christian Golnik, Guntram Pausch, Heide Rohling, and Thomas Kormoll. An image reconstruction framework and camera prototype aimed for Compton imaging for in-vivo dosimetry of therapeutic ion beams. IEEE Trans. on Nuclear Science, 2017.

[69] M. Schönfelder, A. Hirner, and K Schneider. A telescope for soft gamma ray astronomy. Nuclear Instruments and Methods, 107(2) :385–394, 1973.

[70] A Schumann, M Priegnitz, S Schoene, W Enghardt, H Rohling, and F Fiedler. From prompt gamma distribution to dose : a novel approach combining an evolutionary algorithm and filtering based on gaussian-powerlaw convolutions. Physics in Medicine and Biology, 61(19) :6919, 2016.

[71] Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. ACM Trans. Graph., 27(3) :73 :1–73 :10, August 2008.

[72] Li Shen, Eric Todd Quinto, Shiqiang Wang, and Ming Jiang. Simultaneous reconstruction and segmentation with the mumford-shah functional for electron tomography. Inverse Problems & Imaging, 12(6) :1343–1364, 2018.

[73] Emil Y Sidky, Chien-Min Kao, and Xiaochuan Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. Journal of X-ray Science and Technology, 14(2) :119–139, 2006.

[74] Samuli Siltanen, Ville Kolehmainen, Seppo Järvenpää, JP Kaipio, P Koistinen, M Lassas, J Pirttilä, and E Somersalo. Statistical inversion for medical x-ray tomography with few radiographs : I. general theory. Physics in Medicine & Biology, 48(10) :1437, 2003.

[75] BW Silverman, MC Jones, JD Wilson, and DW Nychka. A smoothed EM approach to indirect estimation problems, with particular, reference to stereology and emission tomography. Journal of the Royal Statistical Society. Series B (Methodological), 52(2) :271–324, 1990.

[76] M. Singh and D. Doria. An electronically collimated gamma camera for single photon emission computed tomography. Part II : Image reconstruction and preliminary experimental measurements. Medical Physics, 10(4) :428–435, 1983.

[77] A Sitek. Representation of photon limited data in emission tomography using origin ensembles. Physics in Medicine & Biology, 53(12) :3201, 2008.

[78] B. Smith. Reconstruction methods and completeness conditions for two Compton data models. J. Opt. Soc. Am. A, 22 :445–459, March 2005.

[79] Donald L Snyder, Michael I Miller, Lewis J Thomas, and David G Politte. Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. IEEE Transactions on Medical Imaging, 6(3) :228–238, 1987.

[80] J Webster Stayman and Jeffrey A Fessler. Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems. IEEE transactions on medical imaging, 23(3) :269–284, 2004.

[81] Simon Stute and Claude Comtat. Practical considerations for image-based PSF and blobs reconstruction in PET. Physics in medicine and biology, 58(11) :3849, 2013.

[82] J.P. Sullivan, S.R. Tornga, and M.W. Rawool-Sullivan. Extended radiation source imaging with a prototype Compton imager. Applied Radiation and Isotopes, 67 :617–624, 2009.

[83] S. Takeda, A. Harayama, Y. Ichinohe, H. Odaka, S. Watanabe, T. Takahashi, H. Tajima, K. Genba, D. Matsuura, H. Ikebuchi, Y. Kuroda, et al. A portable Si/CdTe Compton camera and its applications to the visualization of radioactive substances. Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment, 787 :207–211, 2015.

[84] S. Takeda, H. Odaka, S. Ishikawa, S. Watanabe, H. Aono, T. Takahashi, Y. Kanayama, M. Hiromura, and S. Enomoto. Demonstration of in-vivo multi-probe tracker based on a Si/CdTe semiconductor Compton camera. IEEE Transactions on Nuclear Science, 59(1) :70–76, 2012.

[85] T. Taya, J. Kataoka, A. Kishimoto, Y. Iwamoto, A. Koide, T. Nishio, S. Kabuki, and T. Inaniwa. First demonstration of real-time gamma imaging by using a handheld Compton camera for particle therapy. Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment, 831 :355–361, 2016.

[86] Fatma Terzioglu, Peter Kuchment, and Leonid Kunyansky. Compton camera imaging and the cone transform : a brief overview. Inverse Problems, 34(5) :054002, 2018.

[87] Etienne Testa, M Bajard, M Chevallier, Denis Dauvergne, Fabrice Le Foulher, Nicolas Freud, JM Létang, JC Poizat, Cédric Ray, and Mauro Testa. Dose profile monitoring with carbon ions by means of prompt-gamma measurements. Nuclear Instruments and Methods in Physics Research Section B : Beam Interactions with Materials and Atoms, 267(6) :993–996, 2009.

[88] R. Todd, J. Nightingale, and D. Everett. A proposed Gamma camera. Nature, 251 :132–134, 1974.

[89] T. Tomitani and M. Hirasawa. Image reconstruction from limited angle Compton camera data. Phys. Med. Biol., 47 :2129–2145, 2002.

[90] D. Tomono, T. Mizumoto, A. Takada, S. Komura, Y. Matsuoka, Y. Mizumura, M. Oda, and T. Tanimori. First on-site true gamma-ray imaging-spectroscopy of contamination near Fukushima plant. Scientific Reports, 7 :41972, 2017.

[91] Y. Vardi, L.A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. Journal of the American Statistical Association, 80(389) :8–20, 1985.

[92] Joost M Verburg, Kent Riley, Thomas Bortfeld, and Joao Seco. Energy-and time-resolved detection of prompt gamma-rays for proton range verification. Physics in Medicine & Biology, 58(20) :L37, 2013.

[93] Joost M Verburg and Joao Seco. Proton range verification through prompt gamma-ray spectroscopy. Physics in Medicine & Biology, 59(23) :7089, 2014.

[94] S.J. Wilderman, W.L. Rogers, G.F. Knoll, and J.C. Engdahl. Fast algorithm for list mode back-projection of Compton scatter camera images. IEEE Transactions on Nuclear Science, 45 :957–962, June 1998.

[95] D. Xu and Z. He. Gamma-ray energy-imaging integrated deconvolution. Nucl Instr Meth, A574(1) :98–109, 2007.

[96] Ming Yan, Alex AT Bui, Jason Cong, and Luminita A Vese. General convergent expectation maximization (em)-type algorithms for image reconstruction. Inverse Problems & Imaging, 7(3) :1007–1029, 2013.

[97] A. Zoglauer, R. Andritschke, P.F. Bloser, and G. Kanbach. Imaging properties of the mega prototype. In Nucl. Sci. Symp. Conf. Rec., number 3, pages 1694–98, Oct. 2003.

[98] A. Zoglauer, R. Andritschke, and F. Schopper. MEGAlib–The medium energy gamma-ray astronomy library. New Astron. Rev., 50 :629–632, October 2006.

Quatrième partie

Articles annexés

A Tracking Compton-Scattering Imaging System for Hadron Therapy Monitoring

Mirela Frandes, Andreas Zoglauer, Voichita Maxim, and Rémy Prost, *Member, IEEE*

***Abstract*—Hadron therapy for, e.g., cancer treatment requires an accurate dose deposition (total amount and location). As a consequence, monitoring is crucial for the success of the treatment. Currently employed PET imaging systems are not able to provide information about the deposed dose fast enough to allow stopping the therapy in case of a discordance with the treatment plan. We are currently investigating an imaging system based on a combined Compton scattering and pair creation camera capable of imaging gamma rays up to 50 MeV. The camera would be able to measure the complete spectrum of emitted gamma rays during the therapy session. We have performed Monte Carlo simulations for three different proton beam energies in a typical hadron therapy scenario. They show that the location of the gamma-ray distribution decay and the falloff region of the deposed dose are related. The reconstructed images prove that the proposed system could provide the required imaging and dose location capabilities.**

***Index Terms*—Compton imaging, Compton scattering, hadron therapy.**

I. Introduction

HADRON therapy is an effective treatment using charged particle beams of protons or carbon ions. It proved its efficiency in case of tumors which are hard to treat by conventional radiotherapy because they are either radio-resistant, deep-seated, or close to vital organs. The hadron beam enables a high precision in the dose deposition due to the existence of a Bragg peak at the end of the beam range. In order to exploit this benefit a strict control of the location of the delivered radiation dose is extremely important, thus millimeters precision is mandatory. The measurement of gamma rays originating from nuclear reactions of the hadrons within the body is a way to fulfill this requirement. Current systems, such as the in-beam PET system proposed by [1], exploit the coincident 511 keV gamma rays from annihilation of positrons emitted during the beta decay of generated radioactive isotopes. However, the low number of generated isotopes which undergo beta decay, and their decay time require quite long data acquisition times. Therefore it only gives a posttherapy information about the location of the deposed dose. Unlike conventional PET scanners, in-beam PET has limited angle of coverage due to the gantry, reducing the quantitative precision. Moreover, positron activity created by nuclear interactions undergoes metabolic processes and is washed out via the blood flow. The ultimate goal, however, is to monitor location and deposed dose of the beam in real-time in 3D. One way to achieve this goal is to measure the complete spectrum of the emitted gamma rays during the interaction of the hadron beams with the target. This includes nuclear gamma rays which are emitted by the relaxation of generated nuclei. This emission is isotropic, and its energy spectrum ranges from roughly 100 keV up to 20 MeV, accompanied by a more energetic background of neutrons. A relation between the spatial distribution of gamma rays following nuclear reactions and the dose falloff region was reported by [2] with an experimental proton beam. There, the nuclear gamma rays are observed by a collimated system counting only those gamma rays which are emitted with an 90° angle with respect to the beam direction.

In general, in nuclear medicine, the detection systems (PET, SPECT, etc.) are optimized for energies below 1 MeV. Thus the measurement of energetic gamma rays from nuclear interactions exceeds the capabilities of all existing medical systems. However, those gamma rays are a prime target for astronomical imaging systems, such as tracking Compton and pair telescopes [3], detectors using microtime projection chambers [4], etc. Those systems do not require collimators and therefore can achieve a high efficiency and large field-of-view. Furthermore, they can use gamma ray and electron tracking to identify and thus reject a large portion of the background. Therefore, the adaptation of such a system to hadron therapy monitoring might prove advantageous.

An applicability investigation of an astronomical system for medical imaging, an electron tracking Compton camera based on a micro pixel gas chamber, was reported by [5] using low energy, e.g., 364 keV, gamma-ray sources. In the following, we propose an instrument operating in the required energy regime, called Medium-Energy Gamma-ray Astronomy (MEGA) prototype. MEGA was the first fully calibrated and successfully operating combined Compton and pair telescope, capable of measuring gamma rays in the energy range from roughly 400 keV up to 50 MeV. Based on the MEGA prototype measurement principle, an advanced Hadron Therapy Imaging HTI system model was designed and optimized for detection of gamma rays in the Compton energy regime. It is capable of tracking a gamma ray through several Compton interactions until it is stopped via photoelectric effect, and also of tracking recoil electrons. The Compton scatter angle of the first interaction and the positions of the first two interactions allow to restrict the direction of the

Manuscript received January 14, 2009; revised June 09, 2009. Current version published February 10, 2010.

M. Frandes, V. Maxim, and R. Prost are with the CREATIS-LRMN; CNRS UMR5220; Inserm U630; INSA-Lyon, University of Lyon, 69100 Villeurbanne, France (e-mail: frandes@creatis.insa-lyon.fr; maxim@creatis.insa-lyon.fr; prost@creatis.insa-lyon.fr).

A. Zoglauer is with the Space Sciences Laboratory, University of California, Berkeley, CA 94720 USA (e-mail: zog@ssl.berkeley.edu).

Digital Object Identifier 10.1109/TNS.2009.2031679

0018-9499/$26.00 © 2010 IEEE

Authorized licensed use limited to: INSA. Downloaded on March 01,2010 at 09:08:33 EST from IEEE Xplore. Restrictions apply.

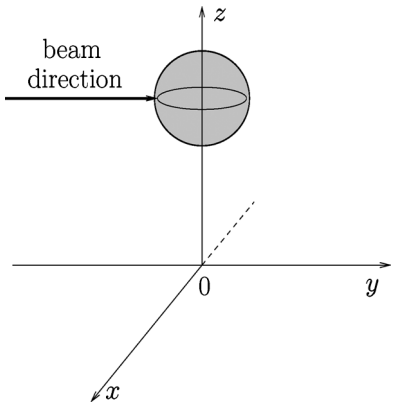


Fig. 1. Schematic representation of the setup used for gamma ray emission simulations. The center of phantom represented by a PMMA sphere is situated at $z = 41.5$ cm.

incident gamma ray to a cone surface. The reconstruction of the recoil electron direction enables to further restrict the origin distribution to a cone segment.

The paper is organized as follows. In Section II, we compare the simulated dose deposition and the gamma-ray emission distribution in a PMMA phantom irradiated by a proton beam. Section III introduces the proposed detection technique based on Compton gamma-ray tracking and describes the design of the HTI as well as its simulated performance. Section IV presents the reconstructed images of Compton scatter events simulated with the HTI. Section V discusses the impact of the used imaging algorithm on the presented results. We conclude in Section VI.

II. Gamma Ray Emission

When charged particles such as protons traverse matter they lose their energy primarily through ionization and inelastic scattering. The latter generates, beside others, excited nuclei which emit gamma rays during their deexcitation. If some isotopes are unstable, further gamma rays may be generated during their radioactive decay. In the following, we concentrate on the detection of gamma rays emitted during the proton irradiation, since only those could enable real-time hadron therapy diagnostics.

The first step is to determine their emission patterns and spectra via simulations. The simulation configuration consists of a PMMA ($C_5H_8O_2$) phantom, which was irradiated by three monoenergetic pencil proton beams with an intensity fixed at 10^6 protons/second. The start point of beam is along y axis at -7.5 cm. The phantom is a sphere with 7.5 cm radius centered at 41.5 cm on z axis (Fig. 1). The beam energies were selected from the potential treatment range, 70, 100, and 140 MeV.

Simulations were performed using the Geant4 package [6] (version 9.1) with its front end Cosima [7]. The Geant4 Livermore package including the Doppler-broadening extension G4LECS [8] was used for electromagnetic interactions and the standard QGSP-BIC-HP physics list for hadronic interactions.

Fig. 2 shows the dose deposition as well as the simulated distribution of originated nuclear gamma rays along the beam path for the first second of irradiation. Just before the proton comes to rest, a pronounced peak in dose deposition can be observed, called Bragg peak.

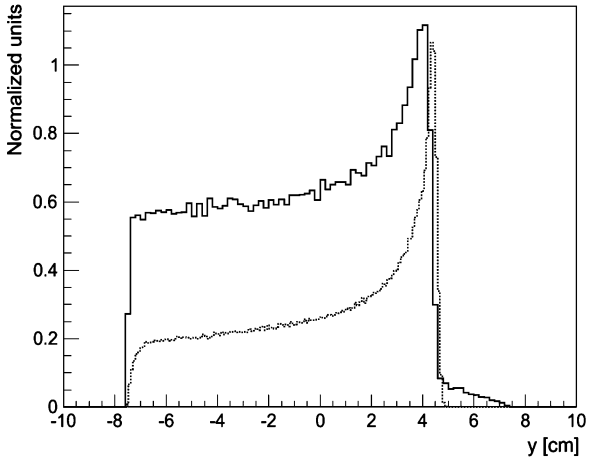


Fig. 2. Simulated depth dose deposition profile (dotted line) and simulated depth distribution profile of originated nuclear gamma rays (solid line) in PMMA phantom for proton beam irradiation at 140 MeV. The start point of beam in PMMA is $y = -7.5$ cm.

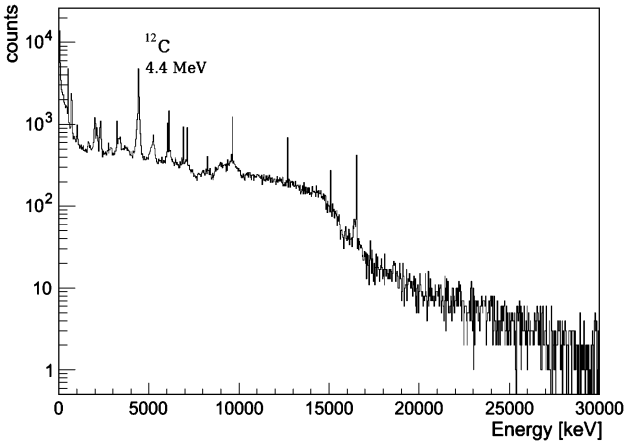


Fig. 3. Simulated energy spectrum of gamma rays with energies below 30 MeV leaving the PMMA phantom during irradiation by a proton beam at 140 MeV.

The simulated energy spectrum of emitted gamma rays (i.e. those leaving the phantom) during the 140 MeV proton beam irradiation can be found in Fig. 3. It shows a continuum ranging up to roughly 20 MeV and some nuclear lines. The most prominent lines are the 511 MeV line from positron annihilation and the 4.4 MeV excitation line from carbon. The same nuclear lines can be observed for the other considered proton beam energies, 70 and 100 MeV.

Fig. 4 shows the origin position of the gamma rays leaving the phantom during the first second of irradiation. In this case, origin means either the creation location, if the gamma ray left the phantom unscattered, or the last scatter position (last position of direction change), if the gamma ray was subject to, e.g., a Compton scatter.

Simulations indicate that during the first second of irradiation, roughly 10% of the protons at 70 MeV lead to gamma rays, which escape the PMMA phantom isotropically. When the beam energy is 140 MeV, the number of gamma photons increases to roughly 30% of the delivered protons.

The distribution of emitted particles is a function of their origin position. Obviously, close to the entry position of the

Authorized licensed use limited to: INSA. Downloaded on March 01,2010 at 09:08:33 EST from IEEE Xplore. Restrictions apply.

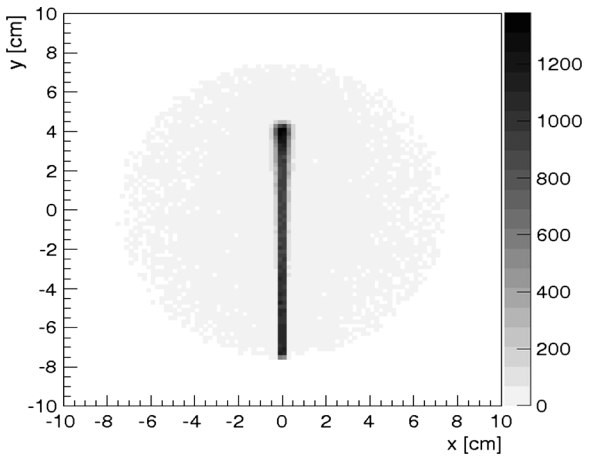


Fig. 4. Simulated 2D image projection along $x-y$ axis of the emission location of gamma rays leaving the PMMA phantom during irradiation by the proton beam at 140 MeV.

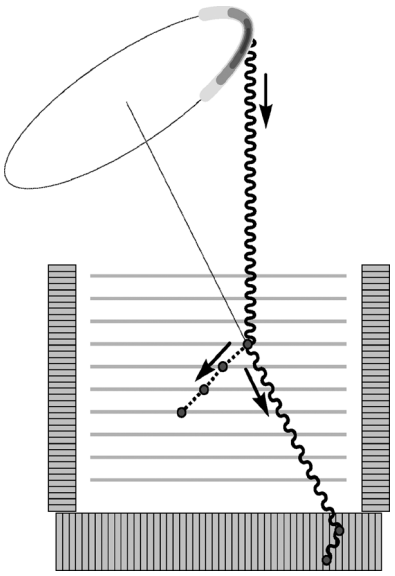


Fig. 5. Compton imaging technique based on gamma-ray and electron tracking. Illustration of the basic detection sub-systems: a central tracker (the scattering sub-system) is surrounded by a calorimeter (the absorption sub-system).

beam into the phantom lower energy gamma rays can more easily escape the phantom. In addition, simulations show an increased concentration of escaping gamma rays with energies above 1 MeV close to the Bragg peak. Gamma rays with energies between 1 and 10 MeV represent the dominant emission during the irradiation, with an intensity peak at the location of the Bragg maximum. This is therefore the target energy range for proton therapy monitoring, and includes the Compton regime and the low-energy pair regime.

III. A Compton-Based Imaging System for Hadron Therapy

A. A Tracking Compton Imaging System

The proposed gamma-ray detector HTI (Hadron Therapy Imaging) consists of two detector subsystems (see Fig. 5). In the tracker, composed of several layers made of thin silicon wafers, the initial interaction happens (either Compton scattering or pair creation), and the electrons and positrons are tracked. A calorimeter surrounds the lower hemisphere of the tracker and stops all secondary particles. It measures energy and interaction positions.

The simulation model of the tracker comprises 36 layers of double-sided Silicon-strip detectors. Each layer consists of 2 by 2 wafers with a length of 10 cm, thickness of 0.5 mm and a strip pitch of 0.5 mm. A uniform 1-sigma energy resolution of 1 keV and a 10 keV threshold were assumed in the electron tracker. The calorimeter consists of $LaBr_3$ bars (surface area: 5×5 mm, length: 8 cm bottom, 4 cm side). An energy resolution of 4.3% FWHM at 662 keV, and a threshold of 28 keV was assumed. For an event to be triggered at least one hit in the tracker and one hit in the calorimeter is required.

In the tracker, gamma rays undergo either Compton scattering or pair creation. If the energy of the produced electrons (and positrons) is above $\sim$500 keV then they pass through several layers and the direction of the recoil electron or electron-positron-pair can be determined during data analysis.

B. Data Analysis Tools

Simulations of the HTI system have been performed using the same tool as the simulations of the gamma-ray emission patterns, Cosima, a Monte-Carlo tool based on Geant4, which is part of the MEGAlib package [9].

For Compton telescopes the first step of data analysis is the event reconstruction. The simulation (like real measurements) results in a set of positions and energies. To determine the origin of the gamma rays, the interaction sequence has to be determined by analyzing all possible paths of the gamma ray in the tracker and the calorimeter as well as of all secondaries in the tracker. For this step three different approaches are available, the classic approach (Compton sequence reconstruction) [3], a Bayesian approach [10], and a neural network approach [11]. Due to the large amount of generated simulation data, and the limited amount of available computing resources, the fastest approach, the classic approach has been chosen for the event reconstruction. Since the HTI allows for gamma ray as well as electron tracking, sufficient redundant measurements about the event is recorded to enable background rejection reliably. For the hadron therapy application, the most important background sources are random coincidences, other particles originating from the phantom (e.g., neutrons), and incompletely absorbed events.

The next and most challenging step during data analysis is image reconstruction. The knowledge of the first and second interaction position of the gamma ray along with the measured energies allows restricting the origin of the gamma ray to a surface of a cone. Its opening angle can be calculated via the standard Compton equation. Moreover, if an electron track is present, reconstructing the direction of the recoil electron enables to restrict the incident gamma-ray direction further to a small segment of the cone. The applied list-mode maximum-likelihood expectation-maximization image reconstruction algorithm (LM-ML-EM) is an event-by-event approach,

Authorized licensed use limited to: INSA. Downloaded on March 01,2010 at 09:08:33 EST from IEEE Xplore. Restrictions apply.

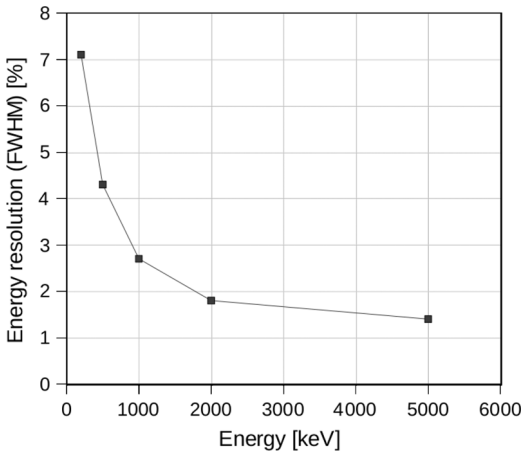


Fig. 6. Energy resolution of the HTI simulation model.

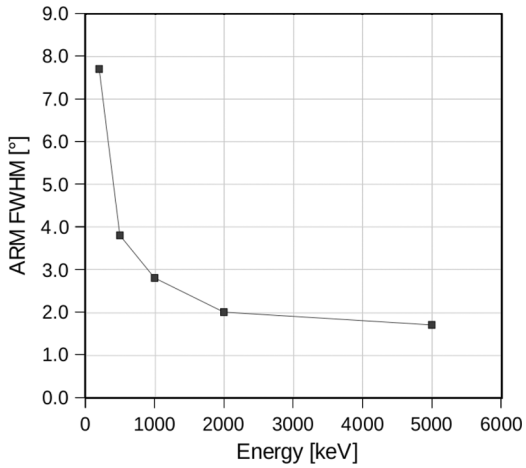


Fig. 7. Angular resolution of the HTI simulation model as FWHM of the ARM.

which avoids large data-space matrices. This imaging algorithm was developed based on the 2D near-field imaging algorithm proposed by [12] for medical imaging, being extended to also include tracked Compton events. The system response matrix is computed similar to [13]. Each event response is determined from the calculated emission and interaction probabilities, which are described by the Compton cones and arcs profiles. The profiles represent the distributions of possible true event cones for the measured one, and the distribution of possible true scatter planes for the measured scatter plane. The simplest form to determine the shape of the profile, which is currently used, is by an 1D Gaussian approximation of the widths corresponding to the values derived from the Angular Resolution Measure (ARM) and the Scatter Plane Deviation (SPD). The ARM is a distribution given by the smallest distance of the known origin of the gamma ray to the Compton cone. The SPD is defined as the angular distance on the Compton cone between the known origin of the photons and the calculated one. It is a measure of the length for the Compton scatter arc, while the ARM is a measure for the width of the Compton scatter cone or arc.

C. Estimated Performance

Determining the origin of a gamma ray with a Compton camera utilizes energy and position measurements. Therefore the accuracy of the reconstruction is limited by the uncertainties associated to those measurements. In addition, the angular resolution of Compton telescopes is limited by Doppler-broadening resulting from the unknown momentum of the electron bound to its nucleus. Finally the efficiency of the telescope is limited by the amount of available Silicon layers, the number of completely absorbed events, and the detection thresholds.

The final energy resolution is defined by the energy resolution in the tracker and in the calorimeter. It improves with increasing energy of the initial gamma ray (Fig. 6).

The angular resolution of a telescope can be defined by the FWHM of the ARM of the telescope. The FWHM of the ARM of the HTI simulation model is, at lowest energies, limited by Doppler-broadening, at medium energies by energy and position resolution, and at higher energy by position resolution alone (Fig. 7).

The on-axis effective area is defined as the efficiency in producing a detected event for each incident gamma-ray photon.

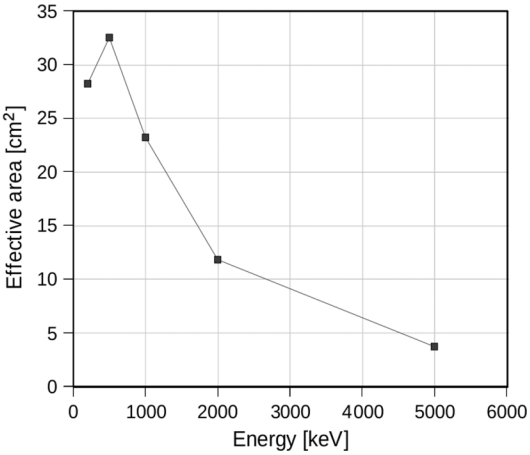


Fig. 8. On-axis photopeak effective area of the HTI simulation model after event with no energy cut selection.

The on-axis photopeak effective area of the HTI simulation model can be found in Fig. 8. The peak is situated around a few hundred keV. At low energies, the probability for a coincidence in tracker and calorimeter decreases due to a small cross section for Compton scattering. At high energies the probability for a completely absorbed event decreases due to escapes and due to more interaction sites in the tracker, which generally increases the risk of interactions in passive material.

IV. Results

In order to estimate the imaging performance of the detector, a PMMA sphere (radius 7.5 cm) was placed at 8 cm distance from the imaging system (Fig. 9). The phantom was irradiated by a proton beam with different energies selected from the expected treatment range, 70, 100, and 140 MeV and an intensity fixed at 10^6 protons/s. The total simulation time is 20 s.

The simulated data was reconstructed as described in Section III-B.

Fig. 10 shows reconstructed 2D images, where the reconstruction plane is the beam plane (at $z = 41.5$ cm) for the three different beam energies. Only Compton events with energies above 450 keV were reconstructed. As expected, elevating

Authorized licensed use limited to: INSA. Downloaded on March 01,2010 at 09:08:33 EST from IEEE Xplore. Restrictions apply.

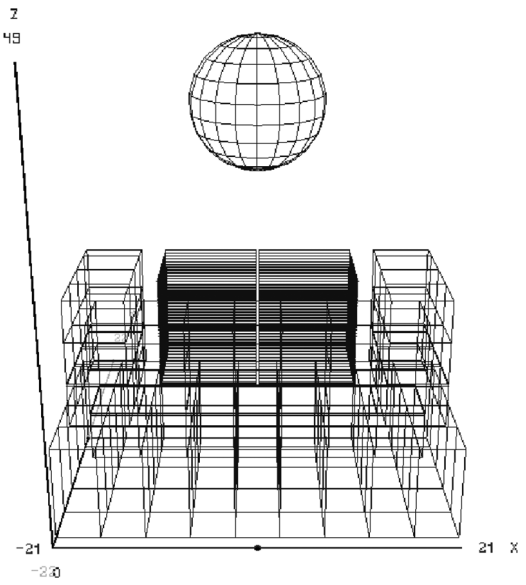


Fig. 9. Wire-frame representation of geometry setup as used for hadron therapy simulations. A cut-view by a back-side plane was used.

the proton beam energy from 70 to 140 MeV results in increasingly long, line-shaped reconstructed beams path.

The intensity profiles of the reconstructed images along the central vertical line are illustrated in Fig. 11. For the proton beam at 70 MeV, the falloff region of the deposed dose is after 3.5 cm from the entry point into the phantom (at -7.5 cm on the y axis). When increasing the beam energy to 100 and 140 MeV, the Bragg peak is translated to 6.5 and 12 cm, respectively (Fig. 11, dashed lines).

The length of the reconstructed gamma-ray source depends on the beam energy being related to the proton range in the phantom. The decay of the reconstructed gamma ray activity allows retrieving information about the location of the Bragg peak. However, assuming ideal events, i.e. no measurement uncertainties (energy, position), and no escaping events or deposits in passive material, the intensity profile of Compton reconstructed image obtained at 140 MeV shows a roughly abrupt decay at the Bragg peak position (Fig. 11, solid line). Moreover, with the current imaging algorithm the Bragg peak cannot be recovered precisely. This fact is expected since the reconstruction algorithm estimates the source distribution by a maximum-likelihood function using a crude approximation for the model of data acquisition process. The application of maximum-likelihood estimation based algorithms in reconstruction of radioactive sources in PET reported the existence of both noise and edge artifacts [14]. Secondly, the detector response is not adapted for events with high-incident energy. In addition, the algorithm does not include corrections for absorptions in the phantom, and does not include the prior knowledge that the source is line shaped. A more detailed discussion about the reconstruction process is presented in the next section.

V. Discussion

In this section, the capability of the proposed imaging system using the LM-ML-EM algorithm is studied. To achieve the reconstruction precision required by our application, an accurate

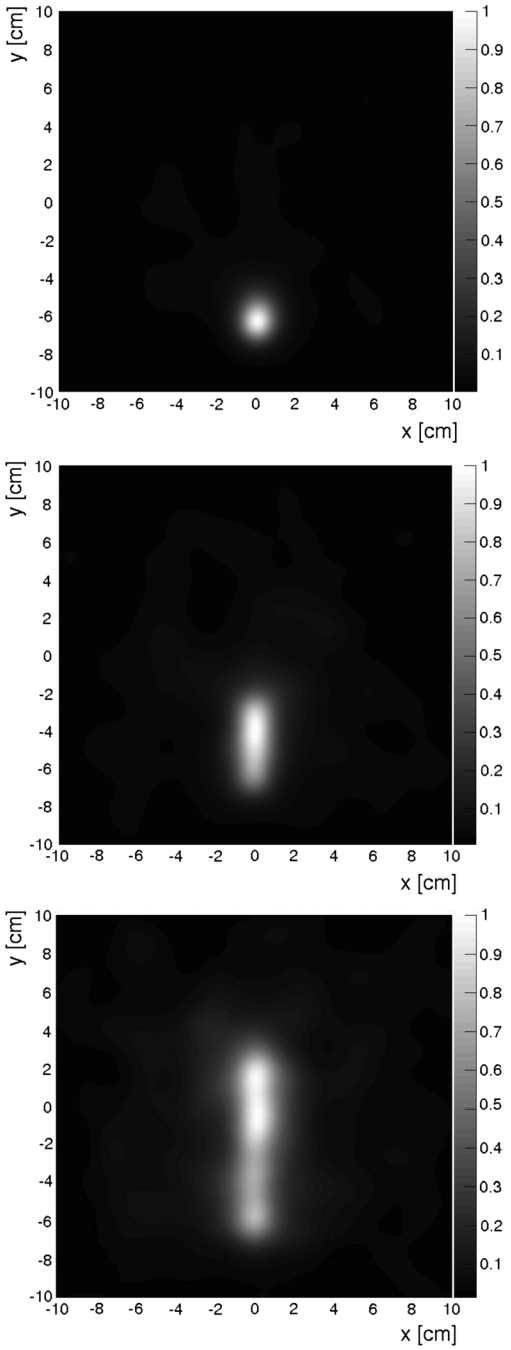


Fig. 10. Compton reconstructed images of measured gamma rays during simulated irradiation of PMMA phantom by a proton beam at 70 MeV (top), 100 MeV (middle), and 140 MeV (bottom). The reconstruction plane is the beam plane.

modeling of the detection system is necessary in order to correctly calculate the detection sensitivity and the response of individual events, which determine the estimated quality of the image reconstruction.

In the present algorithm, the response of each event is approximated via transition probabilities represented by the Compton cones and arcs profiles. For Compton events, where the energy transferred to the recoil electron is not sufficient to produce a track, i.e., incident gamma rays with energies below 2 MeV, the width and the shift of the cone is mainly determined by the

Authorized licensed use limited to: INSA. Downloaded on March 01,2010 at 09:08:33 EST from IEEE Xplore. Restrictions apply.

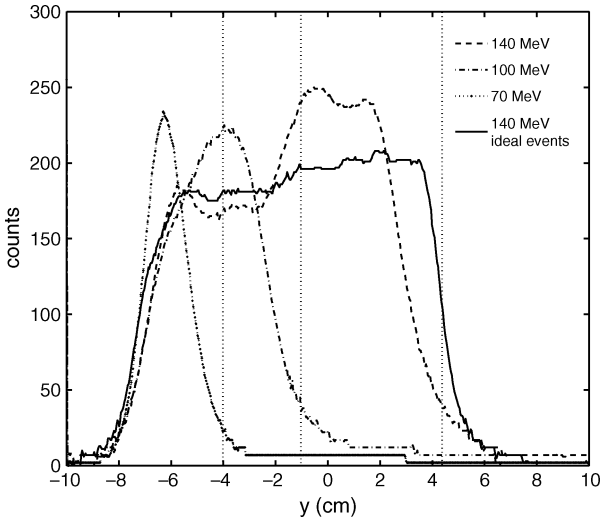


Fig. 11. Intensity profiles of the Compton reconstructed images assuming ideal events (solid line) and assuming measurement uncertainties (discontinuous lines). Bragg peak positions of proton beams at 70, 100, and 140 MeV are indicated by dotted vertical lines.

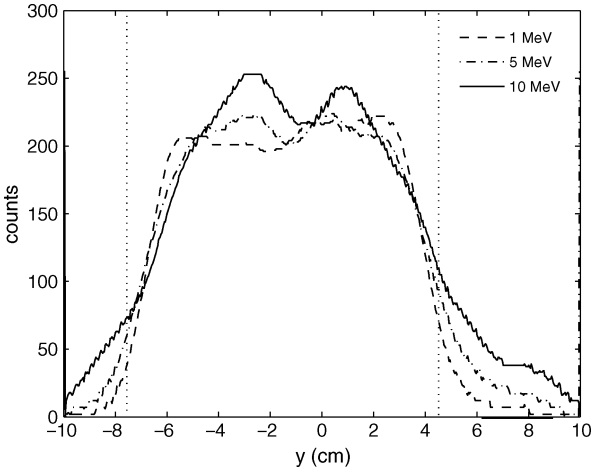


Fig. 12. Intensity profiles of the Compton reconstructed images of gamma-ray line sources. The edge of line sources are represented by dotted vertical lines.

energy and position measurement. With increasing energy, the events have an electron track and the origin is restricted to an arc of the cone-section whose length is determined by Molière Scattering, but the profile becomes broader due to energy leakage. Also, using a list-mode algorithm, a difficult problem to resolve is the absolute normalization, i.e., reconstructing intensity/flux during the image reconstruction process. Since a correct normalization for each event is extremely time consuming, this operation has not been done for the presented images and the sensitivity coefficient is assumed to be one.

In order to evaluate the reconstruction performance, a gamma-ray line source was placed in the PMMA phantom at the same distance from the detector as the proton beam. The length of the sources is equal the 140 MeV proton beam path in the phantom, i.e., 12 cm. A monoenergetic spectrum was assumed for all sources. The number of reconstructed events is the same as for the images of emitted gamma rays activity during the 140 MeV proton beam irradiation. An equivalent response of the imaging algorithm can be observed, i.e., the edge artifact is present in all reconstructed images being more pronounced by increasing the source energy (Fig. 12). This is due to usual artifacts generated by the reconstruction algorithm, the modelling of the system response, and the low number of photons. Moreover, the imaging algorithm does not take into account the different absorption probabilities which the photons encounter when they are emitted at different depths in the phantom, i.e. it is less likely to measure a photon from the center of the phantom than from the edge of the phantom.

The present imaging algorithm includes the ordered-subsets acceleration technique proposed by [15]. One iteration took 2 s on an Intel Core2 Duo CPU, T7700@2.40 GHz, and 2 GB of RAM, while the initial back-projection consumed nearly 120 s. However, more sophisticated acceleration techniques could be envisioned, e.g., space-alternating generalized EM, strategies of row-action mode with dynamic relaxation, etc. Acceleration could also be achieved by using an architecture exploiting parallel processing of data (e.g., using multiple cores or the GPU). The latter method will allow storing a greater number of Compton events, and an extension into 3D of the algorithm.

A camera prototype will be implemented under the grant ENVISION of the European Union Frame work seventh Program (FP7).

VI. Conclusion

Gamma rays in the Compton scattering energy regime are predominantly emitted during proton beam irradiation of a PMMA phantom. Thus they are the primary target for an on-line monitoring system, which measures the dose deposition and the Bragg peak location during hadron therapy.

Simulations revealed that using a tracking Compton camera to detect gamma rays emitted during irradiation of a PMMA phantom by a typical hadron therapy beam, allows to recover the beam path within the phantom. The obtained intensity profiles show a correlation between the decay of the gamma rays spatial distribution and the Bragg peak. However, further improvements of the image reconstruction algorithm are necessary to clearly extract the Bragg peak position with millimeters precision from the reconstructed images.

References

[1] K. Parodi, W. Enghardt, and T. Haberer, "In-beam PET measurements of $beta^+$-radioactivity induced by proton beams," *Phys. Med. Biol.*, vol. 47, pp. 21–36, 2002.

[2] C.-H. Min, C.-H. Kim, M.-Y. Youn, and J.-W. Kim, "Prompt gamma measurements for locating the dose falloff region in the proton therapy," *Appl. Phys. Lett.*, vol. 89, 2006.

[3] A. Zoglauer, "First light for the next generation of Compton and pair telescopes," Doctoral thesis, Technische Univ. München, Munich, Germany, Dec. 2005.

[4] T. Tanimori *et al.*, "MeV gamma-ray imaging detector with micro-TPC," *New Astron. Rev.*, vol. 48, pp. 263–268, 2004.

[5] S. Kabuki *et al.*, "Development of electron tracking Compton camera using micro pixel gas chamber for medical imaging," *Nuncl. Instrum. Methods A*, vol. 580, pp. 1031–1035, 2007.

[6] S. Agostinelli *et al.*, "Geant4-A simulation toolkit," *Nucl. Instrum. Methods A*, vol. 506, pp. 205–303, 2003.

[7] A. Zoglauer, R. Andritschke, S. E. Boggs, F. Schopper, G. Weidenspointner, and C. B. Wunderer, "MEGAlib: Simulation and data analysis for low-to-medium-energy gamma-ray telescopes," in *Proc. SPIE*, July 2008, vol. 7011, p. 70113F.

Authorized licensed use limited to: INSA. Downloaded on March 01,2010 at 09:08:33 EST from IEEE Xplore. Restrictions apply.

[8] R. M. Kippen, The GEANT low energy Compton scattering (GLECS) package for use in simulating advanced Compton telescopes vol. 48, pp. 221–226, 2004.
[9] A. Zoglauer, R. Andritschke, and F. Schopper, "MEGAlib—The medium energy gamma-ray astronomy library," *New Astron. Rev.*, vol. 50, pp. 629–632, October 2006.
[10] A. Zoglauer, R. Andritschke, G. Kanbach, and S. E. Boggs, "A Bayesian-based method for particle track identification in low-energy pair-creation telescopes," in *AIP Conf. Proc.: The First GLAST Symp.*, S. Ritz, P. Michelson, and C. Meegan, Eds., 2007, vol. 921, pp. 584–585.
[11] A. Zoglauer and S. E. Boggs, "Application of neural networks to the identification of the Compton interaction sequence in Compton imagers," in *Proc. IEEE Nuclear Sci. Symp. Conf.*, 2007, vol. 6, pp. 4436–4441.
[12] S. J. Wilderman, N. H. Clinthorne, J. A. Fessler, and W. Les Rogers, "List-mode maximum likelihood reconstruction of Compton scatter camera images in nuclear medicine," in *Proc. 1998 IEEE Nucl. Sci. Symp.*, 1998, vol. 3, pp. 1716–1720.
[13] S. J. Wilderman, J. A. Fessler, N. H. Clinthorne, J. W. LeBlanc, and W. L. Rogers, "Improved modeling of system response in list-mode EM reconstruction of Compton scatter camera images," *IEEE Trans. Nucl. Sci.*, vol. 48, pp. 111–116, 2001.
[14] D. L. Snyder, M. I. Miller, L. J. Thomas, and D. G. Politte, "Noise and edge artifacts in maximum-likelihood reconstrcutions for emission tomography," *IEEE Trans. Med. Imag.*, vol. 3, pp. 228–238, 1987.
[15] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, pp. 601–609, 1994.

Authorized licensed use limited to: INSA. Downloaded on March 01,2010 at 09:08:33 EST from IEEE Xplore. Restrictions apply.

Markus Haltmeier and Voichița Maxim

Inversion of the conical Radon transform and Compton camera imaging

Abstract: The conical Radon transform maps a function defined on the three-dimensional Euclidian space to its integrals over circular cones. Among others, this transform is relevant for image reconstruction in emission tomography using Compton cameras. Recently, there has been major progress toward its theoretical understanding and inversion. The family of all cones depends of six variables. Therefore, the full conical Radon transform is highly overdetermined and a large variety of inversion methods have been proposed and implemented. In this paper, we provide a selected review on the inversion the conical Radon transform. We focus on analytic inversion techniques applicable to data from Compton camera imaging and present them in a systematic and unified way.

Keywords: Cone transforms, Compton camera imaging, Radon transform, Compton transforms, emission tomography, SPECT.

Classification: 44A12, 45D05, 92C55.

1 Introduction

The conical Radon transform, also known under the name cone transform, maps a function $f\colon \mathbb{R}^3 \to \mathbb{R}$ to its integrals over conical surfaces (so-called conical projections). The integrals may be with respect to the standard surface measure or may include some additional weight. We aim for recovering the original function from all conical projections or probably a subset of them. Inversion of conical Radon transforms became popular recently in the mathematical community and significant progress has been made (see, for example, [28, 25, 13, 44, 32, 11, 21, 53, 8] and further references given below).

Among others, inversion of the conical Radon transform arises in Compton camera imaging where the function f represents the density distribution of sources of γ-ray photons (see Section 3). The idea of using Compton scattering of photons as a mean to detect their incoming direction and to exploit this information for

Markus Haltmeier, University of Innsbruck, Department of Mathematics, Technikerstrasse 13, A-6020 Innsbruck, Austria, e-mail: markus.haltmeier@uibk.ac.at

Voichița Maxim, Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, Villeurbanne, France, e-mail: voichita.maxim@creatis.insa-lyon.fr

imaging was published in the early seventies [45, 54] in the context of gamma-ray astronomy and nuclear medicine, respectively. Its actual application to astronomy with the COMPTEL telescope [46] came shortly afterwards. Other applications have been proposed since then and promoted a rich research work on the instrument development on the one and on image reconstruction algorithms on the other hand.

The full conical Radon transform depends on six variables: the three coordinates of the apex, the half-opening angle, and the two angular coordinates of the axis direction. The transform is thus highly over-determined as it maps a function defined on a three-dimensional space to a function defined on a six-dimensional space. This argument partly explains the variety of techniques that have been proposed for its inversion. Several particular cases have been treated, where the dimension of the data space is reduced. Although in this presentation we will focus on the three-dimensional case, the conical Radon transform may be generally defined for functions on $\mathbb{R}^n$. For example, the interesting special case $n = 2$ has applications in imaging with one-dimensional Compton cameras [4, 32].

The main aim of this paper is to provide a unifying overview on analytic inversion techniques for the conical Radon transform, and to make a link with existing Compton camera prototypes and possible detection geometries.

Outline

This article is organized as follows. In Section 2 we define the conical Radon transform and discuss some elementary properties. In Section 3 we describe imaging with Compton cameras, report recent progress on Compton camera design and discuss the relation to the conical Radon transform. The main part of this article is Section 4, where we give a systematic overview on analytic inversion approaches. The paper ends with a discussion and outlook presented in Section 5 where we in particular discuss iterative reconstruction approaches in Compton camera imaging.

2 The conical Radon transform

Let f be some bounded and compactly supported function defined on $\mathbb{R}^3$ and consider the circular cones

$$C(u, \beta, \omega) \triangleq \{v \in \mathbb{R}^3 \mid (v - u) \bullet \beta = \|v - u\| \cos(\omega)\} \tag{1}$$

having apex $u \in \mathbb{R}^3$, axis β on the unit sphere $\mathbb{S}^2 \subseteq \mathbb{R}^3$, and half-opening angle $\omega \in (0, \pi)$. Here and in the following, $\|\cdot\|$ denotes the standard Euclidian norm on $\mathbb{R}^3$ and $\bullet$ is the corresponding inner product. For any $k \in \mathbb{Z}$, we write

$$\mathbf{C}^{(k)} f(u, \beta, \omega) \triangleq \int_{C(u,\beta,\omega)} f(v) \frac{1}{\|v-u\|^k} dv \tag{2}$$

for the integral of f over the cone $C(u, \beta, \omega)$ with weight $\|v-u\|^{-k}$, where dv denotes the standard surface measure. We are in particular interested in the inversion of f from $\mathbf{C}^{(k)} f$ known on certain subsets of all conical surfaces.

Definition 1. *Let $f\colon \mathbb{R}^3 \to \mathbb{R}$ be integrable with respect to the measure $\|v-u\|^{-k} dv$ over almost every cone in a manifold $\mathcal{M} \subseteq \mathbb{R}^3 \times \mathbb{S}^2 \times (0, \pi)$. We call*

$$\mathbf{C}^{(k)} f\colon \mathcal{M} \to \mathbb{R}\colon (u, \beta, \omega) \mapsto \mathbf{C}^{(k)} f(u, \beta, \omega) \tag{3}$$

the conical Radon transform or simply the cone transform of the function f (restricted to the class $\mathcal{M}$ with weight $\|v-u\|^{-k}$).

We will always suppose that f verifies properties ensuring that $\mathbf{C}^{(k)} f$ is well-defined. Typical instances of the cone transform defined by (2), (3) are $\mathbf{C}^{(0)} f$, which represents the integrals of f over the conical surfaces, $\mathbf{C}^{(1)} f$, which is the integral of the line integrals of f over the generatrices of the cone, and $\mathbf{C}^{(2)} f$ in which the values of f are weighted by the solid angle subtended at v by the detector element containing u. It is still unclear which weight should be taken in formula (2) to best model the Compton camera acquisition process. In the absence of measurement uncertainties and if attenuation and additional interactions are neglected, the model that best fits the data seems to be obtained for $k = 2$. This model was originally proposed by Wilderman *et al* in a series of papers (see *e.g.*, [41]) and was afterwards validated against Monte-Carlo simulations in [29] with a small modification related to the solid angle formula.

A simple and natural idea for inversion is to use the conical Radon transform to recover some other Radon transform of f for which an efficient inversion procedure is known. An interesting particular case is obtained when the half-opening angle ω takes the value $\pi/2$, where the conical projection with axis $\beta \in \mathbb{S}^2$ reduces to the classical planar Radon projection

$$\mathbf{R}_\beta f\colon \mathbb{R} \to \mathbb{R}\colon s \mapsto \int_{\beta^\perp} f(s\beta + y) dy\,, \tag{4}$$

where $\beta^\perp \triangleq \{y \in \mathbb{R}^3 \mid \beta \bullet y = 0\}$. More precisely, we have $\mathbf{C}^{(0)} f(u, \beta, \pi/2) = \mathbf{R} f(\beta, u \bullet \beta) \triangleq \mathbf{R}_\beta f(u \bullet \beta)$. Several explicit inversion formulas for the Radon

transform exist and are well known (see, *e.g.*, [34]). A minimal manifold for the stable inversion of the conical Radon transform is the set of conical projections corresponding to $\mathcal{M} = \{(u, \beta, \pi/2) \mid \beta \in \mathbb{S}^2 \wedge u \in \mathbb{R}\beta\}$. Such a restriction has limited interest in practice, but the idea of transforming a set of cone projections into a set of three-dimensional Radon projections has been successfully investigated in the literature and we will return to it later in our presentation.

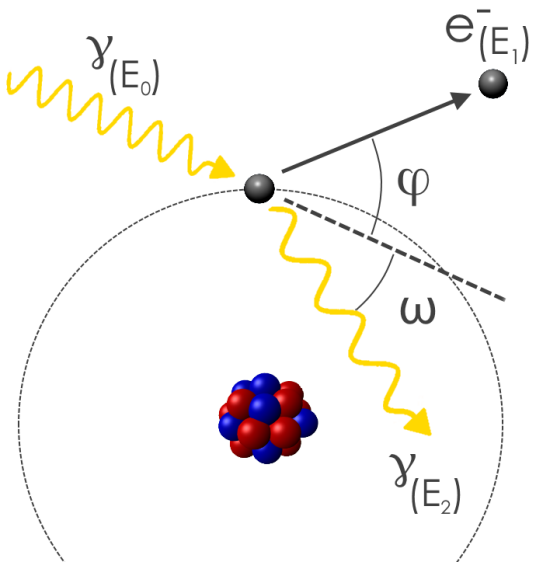


Fig. 1: Schematic representation of Compton scattering of a γ-photon on an electron.

3 Link to Compton camera imaging

3.1 Compton cameras

At sufficiently high energy, photons exhibit particle behavior and are elastically scattered by free electrons in the medium. This property allows to detect γ-photons and to obtain information on their energy E_0 and their emission position. Once the photons hits an electron, it is deflected by angle ω and goes off with an energy E_2. The recoil electron leaves the nucleus with a part E_1 of the γ-ray energy in a direction making an angle φ with the incoming ray direction (see Figure 1). This type of interaction is called Compton scattering, and the distribution of different values of ω is governed by the so-called Klein-Nishina differential cross-section. The energy is conserved in the interaction, thus $E_0 = E_1 + E_2$. If the

energies E_1 and E_2 are measured, the angles ω and φ can be calculated via explicit formulas.

For Compton camera imaging we are interested in the Compton scattering angle ω, which is given by the following equation:

$$\cos(\omega) = 1 - m_e c^2 \frac{E_1}{E_2(E_1 + E_2)} \,. \tag{5}$$

Here m_e is the mass of the electron and c is the speed of the light. If, additionally, the scattering position u and the direction β of the scattered photon are known, then the opposite direction of the incident photon defines a generatrix of the cone with vertex u, axis β and half-opening angle ω. Supposing that this photon, generated by a source with intensity distribution f, did not underwent other interactions, its emission point is known to lie on the cone $C(u, \beta, \omega)$. The expected number of photons detected with the same parameters u, β and ω can be modeled with the conical Radon transform $\mathbf{C}^{(k)} f(u, \beta, \omega)$. Note that when the direction of the recoil electron would be known additionally, then the source of the incoming photon could be assigned the line making the intersection between two cones of vertex u and half-opening angles ω and φ. Whilst the inverse problem would become much simpler, the complexity of the detector design and of the data processing step increases accordingly. Detectors that observe γ-particles undergoing Compton scattering are called Compton cameras or telescopes (see Figure 2).

3.2 Compton camera prototypes

The principle of Compton cameras was first proposed early 1970s independently by two groups, Schönfelder *et al* [45] and Todd *et al* [54] in the fields of gamma-ray astronomy and nuclear medicine, respectively. The large acceptance angle and the large energy spectrum allowed by the new technology found a first successful application in astronomy with COMPTEL (COMPton TELescope). COMPTEL was born in 1991 on a satellite Compton Gamma-Ray Observatory for nine years. The COMPTEL project was then followed by other projects of cameras boarded on stratospheric balloon flights. The first tests for nuclear medicine application were reported in 1983 by Singh and Doria [47]. Nowadays, the application domains of Compton cameras also include homeland security and nuclear decommissioning.

Small prototypes were developed for homeland security and nuclear medicine applications ([30, 51, 23]). The instrumentation developments were accompanied by software developments and one of them is the MEGAlib software [59] based

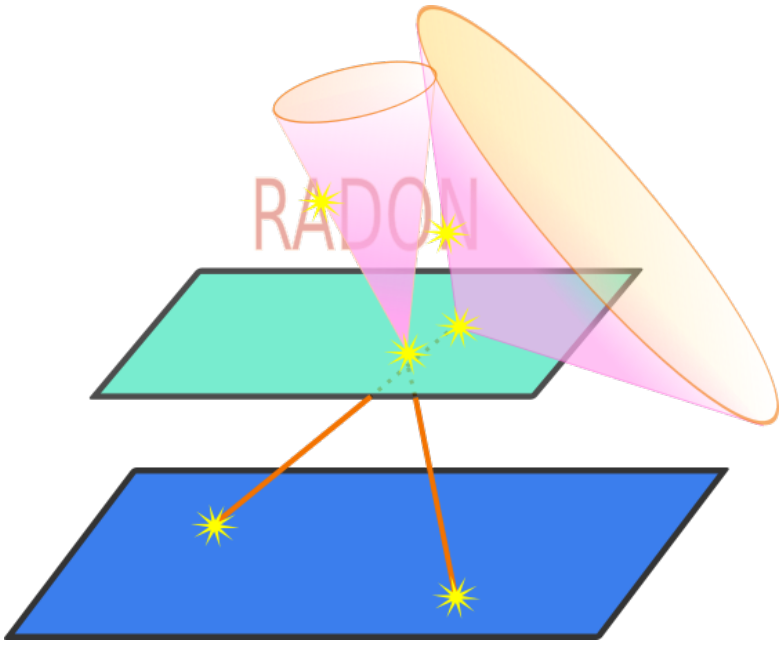


Fig. 2: Schematic representation of a Compton camera composed of a scatterer (green) and an absorber (blue). The source is represented as the word RADON. Photons emitted from the source are first scattered then absorbed in the detector, and the position where they were emitted can only be determined at this stage as belonging to a conical surface.

on the GEANT4 [1] library and intended for Monte-Carlo simulations and image reconstruction. More recently, and boosted by the Fukushima accident initiated by the tsunami following the earthquake on March 11th 2011, prototypes for the identification, quantification, and localization of gamma-ray sources in nuclear power plant and nuclear waste were constructed. We may cite for instance ASTROCAM7000HS from Mitsubishi Heavy Industries in Japan [16] and Polaris-H from H3Dgamma in USA, currently commercialized. In these applications three-dimensional imaging is not compulsory and two-dimensional images on spherical surfaces without depth resolution are sufficient. The position of the source may then be determined by moving the hand-held detector and based on parallax. The most important parameters of these cameras are the view angle, the angular resolution and the weight. Small and compact detectors are preferred although they would provide a relatively poor resolution for 3D imaging of complicated shapes.

In nuclear medicine, multi-hole or pinhole collimators are used to image radioisotope distributions. However, collimators are adapted to specific energies and imaging of the metabolism for the characterization of cancers may benefit from the diagnosis by means of multiple probes. Detectors for higher energies may be useful for new isotopes or for radionuclides employed in Selective Internal Radiation Therapy. Prototypes for small animal tests are under study [33, 52, 19].

In [52] is described a test with a small surface detector which allowed to distinguish the accumulation of two radio-pharmaceuticals injected in a living mouse. In [19] the first 3D image of a mouse with a ultra-compact camera measuring $4.9 \times 5.6 \times 10.6$ cm^3 and weighting 580g is reported. The three-dimensional image was obtained in a multi-angle data acquisition for three radioisotopes: ^{131}I (364 keV), ^{85}Sr (415 keV) and ^{65}Zn (1116 keV).

Gamma rays of higher energy can penetrate through septa of the collimator or produce secondary particles that are afterwards unintentionally detected. As the gamma ray energy increases, thicker septa have to be used, which degrades the collimator aperture ratio. Particle range verification in proton and hadron therapy is a major challenge for SPECT imaging. During the irradiation of the tissues by the ion beam, prompt gamma photons are emitted almost instantaneously with energies in a wide range going from a few keV to several MeV. Slit cameras are relatively simple to produce and they are already commercialized (IBA company). Compton camera should be able to outweigh knife-edge and slit cameras as it can detect photons with energies that span over a region of several MeV and can produce 3D images. Their use in proton therapy treatment monitoring has been investigated from a decade ago [9, 10]. A test with a water phantom irradiated with a proton beam was reported in [22].

3.3 Practical aspects

The main limiting factor in Compton camera imaging is the angular resolution. In a real-life detector system the electrons are neither free nor at rest, but bound to a nucleus. The energy distribution between recoil electron and scattered gamma ray changes, while the sum of the interaction energies is still exactly equal to the total gamma-ray energy. In 1929, Du Mond [7] interpreted a measured broadening of Compton spectra as Doppler broadening induced by the velocity of the electrons. As a consequence of the broadening, the real scatter angle and the one calculated using the standard Compton equation (5) differ. The uncertainty due to Doppler broadening is dominated by the distribution of the Compton profile, which approximately follows a normal distribution. It is thus natural to assume that the energy uncertainty also follows the normal distribution and its standard deviation was calculated by Matscheko *et al* [24]. An approximate formula for the broadening $\Delta\omega$ due to energy uncertainty was derived by Solomon and Ott [50]. Doppler broadening adds blur to reconstructed images. Hirasawa et Tomitani [17] extended the reconstruction method they proposed in [55] to account and correct for Gaussian uncertainties on the Compton scattering angle.

Another main source of degradation is attenuation in the object, which comes under the form of blur and intensity sliding from the center to the boundary of the object. At the moment, very few studies tackle this problem. One of them is [12] in the three-dimensional case with constant attenuation coefficient, and [14] in the two-dimensional case. Detectors may have large planar surface, or they may be small but authorized to move on more or less complicated paths. Their scatterer may have a small thickness or it may be a relatively large block with continuous detection volume as the gaseous detectors [56]. Depending on the type of detector, the choice of the best adapted inversion method for three-dimensional reconstructions may vary.

4 Analytic inversion methods

In this section, we give an overview over analytic inversion methods for inverting the conical Radon transform. Interestingly, often the resulting formulas can be rewritten to bring up known transforms from the Radon family. Selected examples will be presented in more detail.

4.1 Inversion for fixed half-opening angle

Let us first consider inversion methods which apply to fixed opening angle $\omega \in (0, \pi)$ of the cones of integration. If the full conical Radon transform is available, then the reconstructions of f obtained for different ω can subsequently be averaged to reduce the noise of the final reconstruction.

Methods bases on spherical harmonics decomposition

As we mentioned above, the three-dimensional Radon transform (4) of f is obtained immediately when $\omega = \pi/2$. Quite surprisingly, the same can be computed from conical Radon projections with an arbitrary fixed value of $\omega \in (0, \pi)$. This has first been demonstrated by R. Basko *et al* in [5], using orthogonal decomposition in spherical harmonics and linking the coefficients of $\mathbf{C}^{(0)} f(u, \cdot, \omega)$ with the coefficients of the weighted divergent beam transform

$$\forall \eta \in \mathbb{S}^2 : \quad \mathbf{P}_\eta^{(1)} f(u) \triangleq \int_0^\infty f(u + s\eta) s \, ds \, . \tag{6}$$

The main idea is to determine the three-dimensional Radon projections of f on all planes containing a given point u.

The first step consist in calculating the coefficients of the spherical harmonics expansion of the cone projections $\mathbf{C}^{(0)}f(u,\,\cdot\,,\omega)$ considered as function of the axis of the cone and with vertex u and half-opening angle ω taken as constants. From them, the spherical harmonics coefficients of the weighted divergent beam transform $\mathbf{P}_u^{(1)}f$ can be computed. The coefficients of $\mathbf{P}_u^{(1)}f$ are then further used to estimate the coefficients and subsequently the values of $\mathbf{C}^{(0)}f(u,\cdot,\pi/2)$. As noted above, this yields the three-dimensional Radon of f on planes containing the vertex u. To recover f, the above steps have to be repeated for a set of points u that is sufficiently large to guarantee that a complete set of Radon projections is obtained. A plane could be such a set, although it is not a minimal one.

Filtered backprojection type procedure

A different inversion method avoiding costly spherical harmonics decomposition exists for vertices moving freely in a plane P, arbitrary fixed half-opening angle $\omega \in (0, \pi/2)$ and a slightly different conical Radon transform,

$$\mathbf{C}_{\cos}f(u,\beta,\omega) \triangleq \int\limits_{C(u,\beta,\omega)} f(v)\frac{(v-u)\bullet\nu_P}{\|v-u\|}dv, \tag{7}$$

where ν_P is the normal to the plane P. In the following we suppose that the coordinate system is chosen such that $P = \mathbb{R}^2 \times \{0\}$ is the horizontal plane at $z = 0$. A central-slice theorem that links the Hankel transform of the Fourier transform of the conical Radon projections, to the Fourier transform of slices of f parallel to the plane, was given by Maxim *et al* in [28]. Later a filtered back-projection formula has been derived by Maxim in [25], that we recall hereafter.

Let θ and φ be the polar angle and the azimuth of the cone axis $\beta = \beta(\theta,\varphi)$, and we restrict the values of θ to the interval $(0, \pi/2 - \omega)$ depending on the given half-opening angle value ω. The cone projections $\mathbf{C}_{\cos}f(\cdot,\beta,\omega)\colon P \to \mathbb{R}$ with β and ω fixed define a real-valued function on the plane. The two-dimensional Radon transform in the plane $P = \mathbb{R}^2 \times \{0\}$ may be applied to define a new transform,

$$\mathbf{T}_\beta f \triangleq \mathbf{R}_{\varphi+\pi/2}(\mathbf{C}_{\cos}f(\,\cdot\,,\beta,\omega))\,. \tag{8}$$

Here $\mathbf{R}_{\varphi+\pi/2}$ is the linear Radon projection in the plane P on the line spanned by the vector $(\cos(\varphi+\pi/2), \sin(\varphi+\pi/2))$. The transform $\mathbf{T}_\beta f$ give thus the Radon projection of the two-variable function $\mathbf{C}_{\cos}f(\,\cdot\,,\beta,\omega)$ on lines parallel to

the projection of the axis β on P. Incidentally we may note that the number of degrees of freedom of the cone transform was reduced by one and this reduction lead to three variables, two for the vertex $u \in P$ and one for the polar angle θ.

We thus obtained some two-dimensional Radon projections $\mathbf{T}_\beta f$. The values of f are calculated from them by a specific filtered back-projection formula that acts in two-dimensions, on slices of f parallel to P. Each projection $\mathbf{T}_\beta f$ has to be convolved with the filter $H_{z,\theta}$ having frequency response $\widehat{H}_{z,\theta}(\rho) = |\rho|^3 J_0(2\pi z b(\theta)\rho)$, $\rho \in \mathbb{R}$, where we put $b(\theta) = \sin\omega/\sqrt{\cos^2\theta - \sin^2\omega}$ for each $\theta \in (0, \pi/2 - \omega)$. Finally, the filtered projections are back-projected at each altitude z, giving the inversion formula:

$$f(x,y,z) = 2\pi \int_0^\pi \int_0^{\pi/2-\omega} (\mathbf{T}_\beta f * H_{z,\theta})(-x\sin\varphi + y\cos\varphi) b'(\theta) d\theta d\varphi. \qquad (9)$$

The computation of this filtered back-projection formula can be efficiently done numerically for each slice parallel to the plane P and may be seen as a two-dimensional Radon inverse from synthetic Radon projections calculated from the conical Radon projections. We denote by $K_{z,\theta}$ the filter having frequency response $\rho \in \mathbb{R} \mapsto J_0(2\pi z b(\theta)\rho)$ and set

$$p_{z,\varphi} f(s) \triangleq \int_0^{\pi/2-\omega} (\mathbf{T}_\beta f * K_{z,\theta})(s) b'(\theta) d\theta \,.$$

Then the reconstruction formula (9) becomes:

$$f(x,y,z) = 2\pi \mathbf{R}^\sharp [(p_{z,\varphi} f)''](x,y)\,, \qquad (10)$$

where $\mathbf{R}^\sharp$ denotes any left inverse of two-dimensional Radon transform (see [34]). Formula (10) shows that each slice of f can be calculated by Radon inversion from the second derivatives of the synthetic projections $p_{z,\varphi} f$.

As shown in [26], formula (9) can be adapted to invert the weighted conical Radon transform $\mathbf{C}^{(1)} f$. When the projections $\mathbf{T}_\beta f$ are calculated from $\mathbf{C}^{(1)} f$ instead of $\mathbf{C}_{\cos} f$, the inversion formula becomes

$$f(x,y,z) = 2\pi z \int_0^\pi \int_0^{\pi/2-\omega} (\mathbf{T}_\beta f * H_{z,\theta})(-x\sin\varphi + y\cos\varphi) b'(\theta) d\theta d\varphi, \qquad (11)$$

where each slice of f is further multiplied by the altitude z. If the full set of conical Radon projections is available, then the reconstructions of f calculated from different values of ω can be averaged to reduce the noise. The example shown in Figure 3 was produced with such a procedure.

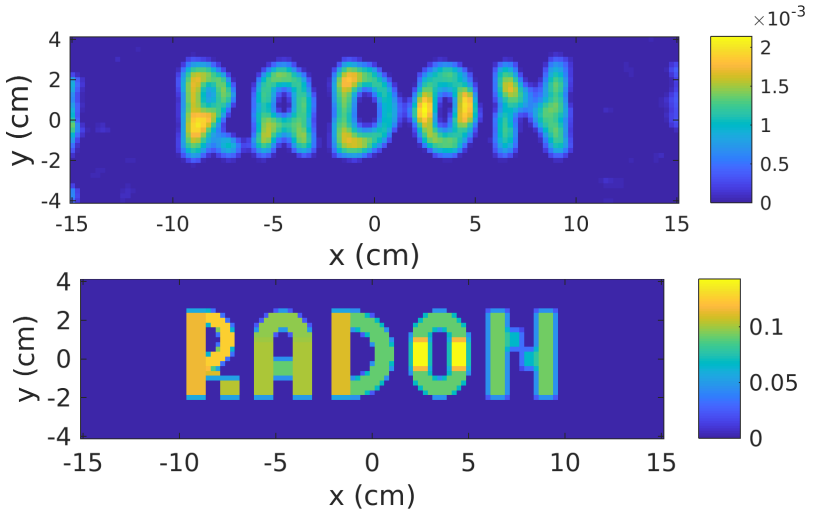


Fig. 3: Example of reconstructed image (top) and reference image (bottom). A three-dimensional source with thickness 1 cm was reconstructed using formula (9) from Monte-Carlo simulated data reproducing the acquisition of 17×10^6 emitted photons with a finite extent Compton camera. Doppler broadening was simulated; the precision of the detectors is supposed infinite. A total variation filtering is applied to the reconstructed image.

4.2 Methods using spherical deconvolution

A circular cone is the union of a set of half-lines, and the integrals over half-lines define the divergent beam (also known as cone-beam) transform having known inverse. As a consequence, if the cone-beam integrals can be determined from the conical Radon transform, the values of f can be recovered, too. Note that recovering cone-beam integrals is of interest on its own for applications where the view in depth of the source is not necessary or impossible to obtain due to the large distance between the source and the detector as it is the case *e.g.*, in astronomy and nuclear plant survey (see Section 3.2). In this case, the position and intensity of a source should be searched for on a sphere surrounding a point-like detector, and then the cone-beam projections yield the cumulative intensity of the gamma-ray source as a function of the incident direction.

The spherical deconvolution methods presented below recover the cone-beam integrals from the conical projections with the same vertex [40, 55]. Parra [40] proposed to calculate, for any fixed apex $u \in \mathbb{R}^3$, the cone-beam projections $\mathbf{P}_\eta^{(0)} f(u) = \int_0^\infty f(u + s\eta) ds$ with $\eta \in \mathbb{S}^2$ from the restricted cone projections

$$\mathbf{C}^{(1)} f(u, \beta, \omega) = \int_{\eta \bullet \beta = \cos(\omega)} \mathbf{P}_\eta^{(0)} f(u) d\eta \,, \tag{12}$$

where $d\eta$ is the arc-length measure on the circle $\{\eta \in \mathbb{S}^2 \mid \eta \bullet \beta = \cos(\omega)\}$. He proposed to first eliminate one degree of freedom by integrating on the

half-opening angle ω,

$$\int_0^{\pi} \mathbf{C}^{(1)}(u,\beta,\omega)h(\cos(\omega))d\omega = \int_{\mathbb{S}^2} \mathbf{P}_{\eta}^{(0)} f(u)h(\eta \bullet \beta)d\eta\,, \tag{13}$$

where h is a real-valued kernel function defined on $(-1,1)$ that may be taken as the Klein-Nishina differential cross-section of Compton scattering.

The right-hand side of equation (13) is a convolution in spherical coordinates and is rotation invariant. Adequate series decomposition allows to transform convolution in a product then to determine a deconvolution formula. As for Cartesian coordinates deconvolution, division by zero and noise enhancement may occur. To address this issue, L. Parra derives a filtered back-projection algorithm which combines back-projection and deconvolution in a single step,

$$\forall \eta \in \mathbb{S}^2 : \quad \mathbf{P}_{\eta}^{(0)} f(u) = \int_{-1}^{1} \int_{\mathbb{S}^2} \mathbf{C}^{(1)}(u,\beta,\omega)H(\eta \bullet \beta,\omega)d\beta d\omega\,, \tag{14}$$

where the established inverse filter is given by

$$H(\cos(\omega),\cos(\omega)') \triangleq \sum_{n=0}^{+\infty} \frac{2n+1}{4\pi} \frac{1}{H_n} P_n(\cos(\omega))P_n(\cos(\omega)')\,. \tag{15}$$

This formula makes use of the Legendre polynomials P_n and of the coefficients related to the back-projected image in the orthogonal decomposition, recalculated in [55] as $H_n = \int_{-1}^{1} h(t)P_n^2(t)dt$. The inverse filter H can be computed prior to the reconstruction and stored.

This inversion procedure (14), (15) was further generalized by Tomitani and Hirasawa in [55]. Instead of integrating ω over the full interval $(0,\pi)$, they showed that a smaller fixed range can be considered. Their motivation comes from practice, where the Compton camera can only measure projections for a set of scattering angles limited bellow by the threshold imposed by the electronic noise present in the first detector, and above by the typical camera geometry made of two parallel detectors. It should be noted that the extreme case of the method from [55] obtained as the measure of the integration range goes to zero, can be compared with the configuration considered by Basko *et al* in [5]. At least numerically, in this case a single value of ω (and a single vertex) will be considered, and from the associated cone projections the divergent beam projections are evaluated. The main differences are the inclusion in [40, 55] of the weighting function h, that may be chosen to have a physical interpretation in Compton camera imaging, and the outcome, which is a weighted divergent projection in [5].

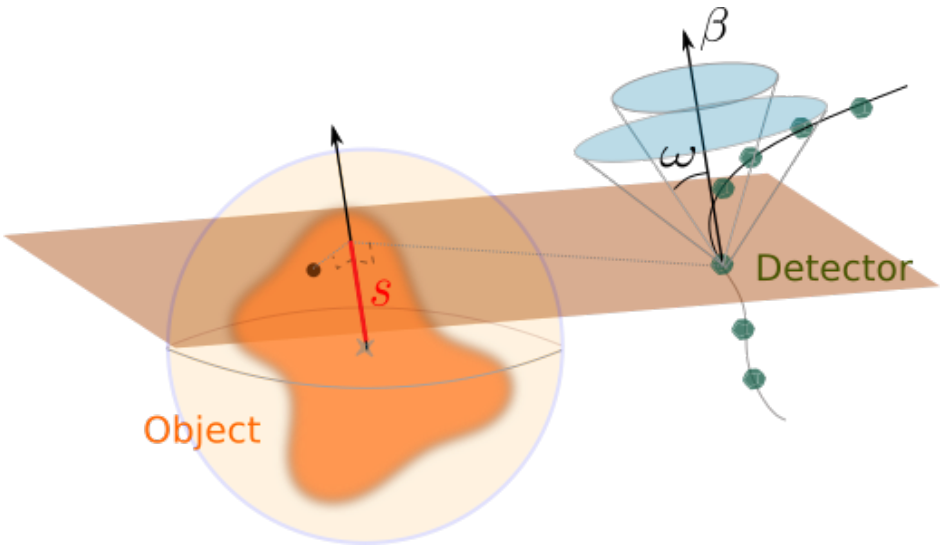


Fig. 4: Tuy's condition for complete data acquisition. A set $\mathcal{D}$ of detection positions should be chosen such that each plane intersecting the source also intersects $\mathcal{D}$ at some point with coordinates u. For β the normal to the plane, all the projections $\mathbf{C}(u, \beta, \cdot)$ should be measured.

4.3 Vertex restriction and Tuy's condition

A different class of analytic inversion methods that exploit the redundancy in the axis and the half-opening angle have been chosen in the works of Smith [49], Kuchment and Terzioglu [20, 21] and Palamodov [39]. Interestingly, although very different in philosophy, these methods share a common starting point with the method introduced by Parra (see Section 4.2). The first step consists to integrate over the opening angle of the conical projections multiplied by some weight function as in (13), and the second step calculates some Radon projection of the data, which may be the three-dimensional Radon transform, the x-ray transform or a divergent beam transform. The main conceptual difference is that the vertex of the cone is not fixed in the integration procedure. Using such a methodology, explicit inversion formulas are derived for $k = 0$ in [49, 20, 21], $k = 1$ in [49, 21], and $k = 2$ in [21, 39].

For instance, in [21] the values

$$G_k(\beta, s) = \int_0^\pi \mathbf{C}^{(k)}(u, \beta, \theta) h_k(\cos\theta) d\theta \,, \tag{16}$$

are computed as an intermediate step for all axis directions β and well-chosen vertices u with $s = u \bullet \beta$ and a certain kernel h_k. The function f then can be

determined at each reconstruction point v through the formula [21]

$$f(v) = c_k \int_{\mathbb{S}^2} \partial_s^{k+1} G_k(\beta, v \bullet \beta) d\beta \,, \tag{17}$$

for some constants c_k and ∂_s^{k+1} denoting the $(k+1)$-th derivative with respect to the second component. For each v in the support of f and for each direction $\beta \in \mathbb{S}^2$, values of $G_k(\beta, s)$ corresponding to an arbitrary small neighbourhood of $v \bullet \beta$ are sufficient for evaluation for (17). A vertex should therefore lie a plane orthogonal to β at distance $s = v \bullet \beta$. Thus a *Tuy's condition* ([57]) can be formulated that ensures that a set $\mathcal{D}$ of detection points is sufficient for the inversion. Roughly speaking, for each point from the support of f, any hyperplane containing this point should intersect $\mathcal{D}$ at least once, and the detector placed in the point of intersection can measure the conical Radon projection for all half-opening angles in $(0, \pi)$, see Figure 4. This condition was further refined by different authors and the differences are related to the specific formulation they have chosen [49, 21, 39].

4.4 Inversion from vertices on a manifold and restricted axis

Axes perpendicular to a surface

In one of the first published studies on the inversion of the conical Radon transform, Cree and Bones [6] found an inversion formula for $\mathbf{C}^{(0)} f$ by restricting u to a plane P and considering cone axes perpendicular to that plane. In this case the number of parameters is three: two for the apex and one for the half-opening angle.

Inversion formulas for vertices on certain hyper-surfaces and axes perpendicular to it were given by Haltmeier *et al* for general weight $k \in \mathbb{Z}$ and general spatial dimension. The case where the vertices lie on the plane $P = \mathbb{R}^2 \times \{0\}$ for arbitrary values of k has been considered in [13], where the compact inversion formula

$$f(x, y, z) = -\frac{z^k}{4\pi^2} \int_0^{\pi/2} \frac{1}{\cos(\omega)^{1+k}} \int_{\mathbb{S}^1} (\Delta g)((x, y) + z \tan(\omega)\eta, \omega) \, d\eta \, d\omega \,, \tag{18}$$

from data $g(x, y, \omega) \triangleq \mathbf{C}^{(k)} f((x, y, 0), \nu_P, \omega)$ has been derived. Here Δ is the Laplacian in the plane $P = \mathbb{R}^2 \times \{0\}$ and $\nu_P = (1, 0, 0)$ the unit normal to P. Further results for this transform have been obtained by Moon [31].

For the case where the vertex moves on a sphere and the vertex is orthogonal to the sphere, [44] Schiefeneder and Haltmeier derive an analytic inversion approach based on spherical harmonics decompositions. In [14], Haltmeier *et al* derive a similar inversion procedure for the conical Radon transform including attenuation $e^{-\lambda\|u-v\|}$ in dimension $n = 2$. Extensions of these results allowing attenuation in arbitrary dimension are established in [15].

Non-perpendicular axis

The formula of [6] for the planar geometry and orthogonal axis was generalized by Maxim *et al* ([28, 25]) for axes making an arbitrary angle $\theta \in [0, \pi/2)$ with the unitary normal ν_P to the plane, for the transform $\mathbf{C}_{\cos} f$ defined in equation (7). The inversion process is similar to the one described in Section 4.1 and can be adapted to invert $\mathbf{C}^{(1)} f$. For $\omega \in (0, \pi/2 - \theta)$, let $b(\omega) = \sin\omega / \sqrt{\cos^2\theta - \sin^2\omega}$. Let φ be the azimuthal angle of the cone axis β and $\mathbf{T}_{\omega,\varphi} f = \mathbf{R}_{\varphi+\pi/2}(\mathbf{C}^{(1)} f(\cdot, \beta, \omega))$. The filter $H_{z,\omega}$ for the filtered back-projection formula has frequency response $\widehat{H}_{z,\omega}(\rho) = |\rho|^3 J_0(2\pi z b(\omega)\rho)$ for $\rho \in \mathbb{R}$. The function f is calculated as

$$f(x, y, z) = 2\pi z \int_0^{\pi} \int_0^{\pi/2-\theta} (\mathbf{T}_{\omega,\varphi} f * H_{z,\omega})(-x\sin\varphi + y\cos\varphi) b'(\omega) d\omega d\varphi, \quad (19)$$

and obviously, the values of f calculated from different choices of θ can then be averaged to reduce the noise.

Several inversion formulas have been derived by Moon and Haltmeier in [32] where the apex moves on the surface of a circular cylinder and the axis points to the central axis of the cylinder. In [18], Jung and Moon present an explicit inversion method using a three-dimensional family of cones, with vertices on a line, variable opening angle and axis restricted to a circle. Their approach is based on computing the cone-beam projections from the conical Radon transform and a new inversion formula from cone-beam projections on a line.

5 Discussion

In this paper we presented a unifying review on analytic inversion approaches for the conical Radon transform with vertices restricted to a set of detectors as it arises in Compton camera imaging. Note we have not discussed cases where the values of $\mathbf{C}^{(k)} f$ are known only for single ω and β but for all vertices $u \in \mathbb{R}^3$. In

this case, the mapping $u \mapsto \mathbf{C}^{(k)} f(\cdot, \beta, \omega)$ can be seen as the convolution of f with a distribution supported on the cone. Deconvolution formulas were given by Florescu *et al* in [8], Gouia-Zarrad and Ambartsoumian [11] for $k = 0$ and by V. Palamodov [39] for $k \in \{0, 1\}$. Also, we did not study inversion of cumulative (or compound) conical Radon transforms that have been studied by Nguyen *et al* in [35, 36] in the context of SPECT imaging of Compton scattered photons.

Historically, in Compton camera development, the first images were produced with iterative methods. Nowadays these methods are still widespread in the user community. Statistical iterative reconstruction is the most widely used class of tomographic reconstruction methods in emission tomography. These methods are based on a probabilistic description of the data, mainly the Poisson distribution, and allow to include accurate models for data corrections and image regularization. Initially adapted from positron emission tomography (PET) and single particle emission computed tomography (SPECT) by Wilderman *et al* (see *e.g.*, [58]), the Maximum Likelihood Expectation Maximization (MLEM) with its list-mode (LM) version is the most known such algorithm. Some recent developments can be fond in [29] and references therein. LM-MLEM aims to find the average emissions per image voxel that maximizes the likelihood of the measured data assuming that this emission follows a Poisson probability law. More recently, the stochastic origin ensemble (SOE) algorithm was proposed by Sitek *et al* ([48, 2]). Belonging to the Bayesian methods family, this algorithm provides a Monte-Carlo estimate of the posterior distribution of the emitted data and claim to converge faster than LM-MLEM algorithm. Similarly to MLEM, SOE also requires a system response model represented in formula (2) by the integration on the cone and the weight. The model can also accommodate object attenuation.

An important feature of these algorithms is that they allow for individual exploitation of the events, and thus avoid binning uncertainties and the blur and artifacts they introduce in the images. Some numerical tests have shown that the most harmful for reconstruction are errors on the half-opening angle of the cone ([26]). With a binning step of five degrees, we are already at the level of the full width at half maximum for Doppler broadening. While iterative algorithms tend to give better results than direct inversion in the most difficult situations where the acquisition noise on the measures is high, the number of acquired events is low, or the acquisition geometry leads to severely ill-conditioned problems, it also makes extensive use of the computer resources and requires very long computation time. However, recent development of parallel computing and also graphics processing units allows to push the limit of data sets that may be treated with such methods (see *e.g.*, [37]). Direct inversion methods might be preferred however in low-noise experiments. Also, they can be used for efficient pre-conditioning. The reconstruction of the cumulated contribution of two sources

coincides with the addition of individual reconstructions. Iterative methods do not verify this property, making quantitative evaluation of their outcome challenging. When the images are noisy, they may be post-processed as we have done with the image shown in figure 3. But iterative methods more easily account for a priori information on the source through penalized-likelihood reconstruction. Quadratic priors (see *e.g.*, [38]) favor smooth images. Total variation priors are preferred for low-counts experiments and for piecewise smooth images (see *e.g.*, [43, 42, 3, 27]). Regularization has similar effect to frequency damping in filtered back-projection but is more respectful of data fidelity criteria and may better preserve the contours than frequency cut.

In the recent years, the Compton camera imaging modality gave birth to a new generalized Radon transform. Remarkable progress has been made in its analysis and in the development of inversion methods. However, various effects on the data make this problem challenging. It is likely that the advent of new methods and new paradigms will help to enhance spatial resolution of the tomographic images and promote Compton cameras in key applications as *e.g.*, medical imaging.

Bibliography

[1] S. Agostinelli et al. Geant4-A simulation toolkit. *Nucl. Instrum. Methods A*, 506:205–303, 2003.

[2] A. Andreyev, A. Sitek, and A. Celler. Fast image reconstruction for Compton camera using stochastic origin ensemble approach. *Med. Phys.*, 38:429, 2011.

[3] S. Anthoine, J. Aujol, Y. Boursier, and C. Melot. Some proximal methods for Poisson intensity CBCT and PET. *Inverse Probl. Imaging*, 6(4):565–598, 2012.

[4] R. Basko, G. L. Zeng, and G. T. Gullberg. Analytical reconstruction formula for one-dimensional Compton camera. *IEEE Trans. Nucl. Sci.*, 44(3):1342–1346, 1997.

[5] R. Basko, G. L. Zeng, and G. T. Gullberg. Application of spherical harmonics to image reconstruction for the Compton camera. *Phys. Med. Biol.*, 43:887–894, 1998.

[6] M. J. Cree and P. J. Bones. Towards direct reconstruction from a gamma camera based on Compton scattering. *IEEE Trans. Med. Imag.*, 13(2):398–407, 1994.

[7] J. W. M. Du Mond. Compton modified line structure and its relation to the electron theory of solid bodies. *Phys. Rev.*, 33(5):643, 1929.

[8] L. Florescu, V. A. Markel, and J. C. Schotland. Inversion formulas for the broken-ray Radon transform. *Inverse Problems*, 27(2):025002, 2011.

[9] M. Frandes, A. Zoglauer, V. Maxim, and R. Prost. Compton-scattering based imaging system for hadron therapy monitoring. In *Nuclear Science Symposium Conference Record, 2008. NSS'08. IEEE*, pages 4853–4859. IEEE, 2008.

[10] M. Frandes, A. Zoglauer, V. Maxim, and R. Prost. A tracking Compton-scattering imaging system for hadron therapy monitoring. *IEEE Trans. Nucl. Sci.*, 57:144–150, 2010.

[11] R. Gouia-Zarrad and G. Ambartsoumian. Exact inversion of the conical Radon transform with a fixed opening angle. *Inverse Problems*, 30(4):045007, 2014.

[12] R. Gouia-Zarrad and S. Moon. Inversion of the attenuated conical Radon transform with a fixed opening angle. *Math. Method. Appl. Sci.*, (in press).

[13] M. Haltmeier. Exact reconstruction formulas for a Radon transform over cones. *Inverse Problems*, 30(3):035001, 2014.

[14] M. Haltmeier, S. Moon, and D. Schiefeneder. Inversion of the attenuated V-line transform with vertices on the circle. *IEEE Trans. Comput. Imaging*, 3(4):853–863, 2017.

[15] M. Haltmeier and D. Schiefeneder. Variational regularization of the conical Radon transform with vertex on the sphere, 2018. to be submitted.

[16] A. Harayama et al. A portable Si/CdTe Compton camera and its applications to the visualization of radioactive substances. *Nucl. Instrum. Methods Phys. Res. A*, 787:207–211, 2015.

[17] M. Hirasawa and T. Tomitani. An analytical image reconstruction algorithm to compensate for scattering angle broadening in Compton cameras. *Phys. Med. Biol.*, 48:1009–1026, 2003.

[18] C. Jung and S. Moon. Exact inversion of the cone transform arising in an application of a Compton camera consisting of line detectors. *SIAM J. Imaging Sci.*, 9(2):520–536, 2016.

[19] A. Kishimoto et al. First demonstration of multi-color 3-D in vivo imaging using ultra-compact Compton camera. *Sci. Rep.*, 7, 2017.

[20] P. Kuchment and F. Terzioglu. Three-dimensional image reconstruction from Compton camera data. *SIAM J. Imaging Sci.*, 9(4):1708—1725, 2016.

[21] P. Kuchment and F. Terzioglu. Inversion of weighted divergent beam and cone transforms. *Inverse Probl. Imaging*, 11(6):1071–1090, 2017.

[22] S. Kurosawa, H. Kubo, K. Ueno, S. Kabuki, S. Iwaki, M. Takahashi, K. Taniue, N. Higashi, K. Miuchi, T. Tanimori, et al. Prompt gamma detection for range verification in proton therapy. *Current Applied Physics*, 12(2):364–368, 2012.

[23] G. Llosa et al. Last results of a first Compton probe demonstrator. *IEEE Trans. Nucl. Sci.*, 55(3):936–941, 2008.

[24] G. Matscheko, G. A. Carlsson, and R. Ribberfors. Compton spectroscopy in the diagnostic x-ray energy range: II. Effects of scattering material and energy resolution. *Phys. Med. Biol.*, 34(2):199, 1989.

[25] V. Maxim. Filtered backprojection reconstruction and redundancy in Compton camera imaging. *IEEE Trans. Image Process.*, 23(1):332–341, 2014.

[26] V. Maxim. Enhancement of Compton camera images reconstructed by inversion of a conical Radon transform, 2018. submitted.

[27] V. Maxim, Y. Feng, H. Banjak, and E. Bretin. Tomographic reconstruction from Poisson distributed data: a fast and convergent EM-TV dual approach, 2018. submitted.

[28] V. Maxim, M. Frandes, and R. Prost. Analytical inversion of the Compton transform using the full set of available projections. *Inverse Problems*, 25(9):1–21, 2009.

[29] V. Maxim, X. Lojacono, E. Hilaire, J. Krimmer, E. Testa, D. Dauvergne, I. Magnin, and R. Prost. Probabilistic models and numerical calculation of system matrix and sensitivity in list-mode MLEM 3D reconstruction of Compton camera images. *Phys. Med. Biol.*, 61(1):243–264, 2016.

[30] L. Mihailescu, K. M. Vetter, M. T. Burks, E. L. Hull, and W. W. Craig. SPEIR: a Ge Compton camera. *Nucl. Instrum. Methods Phys. Res. A*, 570(1):89–100, 2007.

[31] S. Moon. On the determination of a function from its conical Radon transform with a fixed central axis. *SIAM Journal on Mathematical Analysis*, 48(3):1833–1847, 2016.

[32] S. Moon and M. Haltmeier. Analytic inversion of a conical Radon transform arising in application of Compton cameras on the cylinder. *SIAM J. Imaging Sci.*, 10(2):535–557, 2017.

[33] S. Motomura, Y. Kanayama, M. Hiromura, T. Fukuchi, T. Ida, H. Haba, Y. Watanabe, and S. Enomoto. Improved imaging performance of a semiconductor Compton camera GREI makes for a new methodology to integrate bio-metal analysis and molecular imaging technology in living organisms. *J. Anal. At. Spectrom.*, 28:934–939, 2013.

[34] F. Natterer. *The Mathematics of Computerized Tomography*, volume 32 of *Classics in Applied Mathematics*. SIAM, Philadelphia, 2001.

[35] M. K. Nguyen and T. T. Truong. On an integral transform and its inverse in nuclear imaging. *Inverse Problems*, 18(1):265, 2002.

[36] M. K. Nguyen, T. T. Truong, and P. Grangeat. Radon transforms on a class of cones with fixed axis direction. *J. Phys. A*, 38(37):8003, 2005.

[37] V. Nguyen, S. Lee, and M. N. Lee. GPU-accelerated 3D Bayesian image reconstruction from Compton scattered data. *Physics in Medicine and Biology*, 56:2817–2836, 2011.

[38] J. Nuyts and J. A. Fessler. A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution. *IEEE Trans. Med. Imag.*, 22(9):1042–1052, 2003.

[39] V. Palamodov. Reconstruction from cone integral transforms. *Inverse Problems*, 33(10):104001, 2017.

[40] L. C. Parra. Reconstruction of cone-beam projections from Compton scattered data. *IEEE Trans. Nucl. Sci.*, 47(4):1543–1550, 2000.

[41] A. C. Sauve, A. O. Hero, W. L. Rogers, S. J. Wilderman, and N. H. Clinthorne. 3D image reconstruction for a Compton SPECT camera model. *IEEE Trans. Nucl. Sci.*, 46(6):2075–2084, 1999.

[42] A. Sawatzky, C. Brune, F. Wubbeling, T. Kosters, K. Schafers, and M. Burger. Accurate EM-TV algorithm in PET with low SNR. In *IEEE Nuclear Science Symposium Conference Record, 2008*, pages 5133–5137, 2008.

[43] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen.

Variational methods in imaging, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.

[44] D. Schiefeneder and M. Haltmeier. The Radon transform over cones with vertices on the sphere and orthogonal axes. *SIAM J. Appl. Math.*, 77(4):1335–1351, 2017.

[45] M. Schönfelder, A. Hirner, and K Schneider. A telescope for soft gamma ray astronomy. *Nucl. Instrum. Meth.*, 107(2):385–394, 1973.

[46] V. Schönfelder, H. Aarts, K. Bennett, et al. Instrument description and performance of the imaging gamma-ray telescope COMPTEL aboard the Compton Gamma-Ray Observatory. *Astrophysical Journal Supplement Series*, 86:657, 1993.

[47] M. Singh and D. Doria. An electronically collimated gamma camera for single photon emission computed tomography. Part II: Image reconstruction and preliminary experimental measurements. *Med. Phys.*, 10(4):428–435, 1983.

[48] A. Sitek. Representation of photon limited data in emission tomography using origin ensembles. *Phys. Med. Biol.*, 53(12):3201, 2008.

[49] B. Smith. Reconstruction methods and completeness conditions for two Compton data models. *J. Opt. Soc. Am. A*, 22:445–459, March 2005.

[50] C. J. Solomon and R. J. Ott. Gamma ray imaging with silicon detectors – a Compton camera for radionuclide imaging in medicine. *Nucl. Instrum. Methods Phys. Res. A*, 273(2-3):787–792, 1988.

[51] J. P. Sullivan, S. R. Tornga, and M. W. Rawool-Sullivan. Extended radiation source imaging with a prototype Compton imager. *Appl. Radiat. Isot.*, 67:617–624, 2009.

[52] S. Takeda, H. Odaka, S. Ishikawa, S. Watanabe, H. Aono, T. Takahashi, Y. Kanayama, M. Hiromura, and S. Enomoto. Demonstration of in-vivo multi-probe tracker based on a Si/CdTe semiconductor Compton camera. *IEEE Trans. Nucl. Sci.*, 59(1):70–76, 2012.

[53] F. Terzioglu, P. Kuchment, and L. Kunyansky. Compton camera imaging and the cone transform: a brief overview. *Inverse Problems*, 34(5):054002, 2018.

[54] R. Todd, J. Nightingale, and D. Everett. A proposed Gamma camera. *Nature*, 251:132–134, 1974.

[55] T. Tomitani and M. Hirasawa. Image reconstruction from limited angle Compton camera data. *Phys. Med. Biol.*, 47:2129–2145, 2002.

[56] D. Tomono, T. Mizumoto, A. Takada, S. Komura, Y. Matsuoka, Y. Mizumura, M. Oda, and T. Tanimori. First on-site true gamma-ray imaging-spectroscopy of contamination near Fukushima plant. *Sci. Rep.*, 7:41972, 2017.

[57] H. K. Tuy. An inversion formula for cone-beam reconstruction. *SIAM J. Appl. Math.*, 43(3):546–552, 1983.

[58] S. J. Wilderman, W. L. Rogers, G. F. Knoll, and J. C. Engdahl. Fast algorithm for list mode back-projection of Compton scatter camera images. *IEEE Trans. Nucl. Sci.*, 45:957–962, June 1998.

[59] A. Zoglauer, R. Andritschke, and F. Schopper. MEGAlib–The medium energy gamma-ray astronomy library. *New Astron. Rev.*, 50:629–632, October 2006.

Enhancement of Compton camera images reconstructed by inversion of a conical Radon transform

Voichiţa Maxim
Université de Lyon, CNRS, Inserm
INSA-Lyon, CREATIS, UMR5220, U630, F-69621, Villeurbanne, France
E-mail: voichita.maxim@creatis.insa-lyon.fr

Abstract. We present a new inversion formula for a weighted conical Radon transform modelling Compton camera data. The formula exploits a large proportion of the acquired events and is easy to implement into fast algorithms. We give for it two equivalent formulations relying on known properties of the two-dimensional Radon transform and we test a semi-iterative algorithm for one of them. From a practical point of view, methods robust to measurement noise and to low number of events are required. We show that adding a constraint on the total variation of the final image strongly improves the results. We illustrate our arguments with Monte-Carlo simulated data in both low and realistic noise configurations.

Keywords: image reconstruction, conical Radon transform, Compton transform, analytic algorithms, Compton camera, total variation

Submitted to: *Inverse Problems*

1. Introduction

The recent interest in the conical Radon transforms was driven by the development of Compton cameras for medical and homeland security imaging applications. The idea of using Compton scattering of γ-rays as a mean to detect their incoming direction and to exploit this information for imaging was first proposed in the early 1970s independently by two groups, Schönfelder *et al* [1] and Todd *et al* [2] in the fields of γ-ray astronomy and nuclear medicine, respectively. This new technology was successfully implemented in the COMPton TELescope COMPTEL, born in 1991 on a satellite Compton Gamma-Ray Observatory for nine years. Other projects of cameras boarded on stratospheric balloon flights were realized since then. The first tests for nuclear medicine application were reported in 1983 by Singh and Doria [3]. With the technological progress, medical applications could in a near future benefit from the large acceptance angle and the large energy spectrum allowed by the Compton imaging devices. Nowadays, the application

fields of Compton cameras were extended to imaging in homeland security and nuclear decommissioning ([4, 5, 6, 7]).

The guiding thread of this study is to keep as close as possible to real applications, where the incompleteness of the data set and the acquisition noise reduce the quality of the images. Section 2 is devoted to the introduction of the main concepts employed in the paper. It contains a brief presentation of the Compton camera followed by a short state-of-the-art on conical Radon transforms, an introduction to total variation denoising ([8]) and to the Simultaneous Iterative Reconstruction Technique (SIRT, [9]). In section 3 we derive an inversion formula for a class of weighted conical Radon transforms, for cones with vertices in a plane (the scatterer) and arbitrary axis directions and half-opening angles. The filtered backprojection inversion formula we propose is closely related to the one derived in [10]. We propose three methods for the numerical computation of the inverse, two of them being analytic and the third semi-iterative. As the noise and the low number of acquired events lead to poor reconstructed images, the three of them are associated with total variation denoising or regularization. Finally we evaluate the methods with Monte-Carlo simulated data reproducing the acquisition of a complex shaped source with a finite-size detector. Two levels of uncertainties on the position and energy measurements are considered in the simulation. Section 4 describes the data, generated with a Monte-Carlo software from nuclear physics. In section 5 we show some numerical examples. We demonstrate that the methods can reconstruct the images from statistically distributed counts, but with a resolution depending on the geometry of acquisition and on the noise level.

2. Introduction to Compton camera imaging and to total variation denoising

In this section we introduce the principle of Compton camera imaging and we give a concise state-of-the-art on conical Radon transforms and their inverses. We also recall some widely used techniques that will be used in the next sections. They are respectively the Simultaneous Iterative Reconstruction Technique (SIRT, [9]) that is largely employed for tomographic reconstruction and the total variation (TV) denoising methodology originally introduced by Rudin *et al* in [8].

2.1. Compton scattering and conical projections

In nuclear medicine imaging Compton cameras might eventually surpass the collimated Anger cameras and allow either reduction of the radio-tracer dose injected to the patient or reduction of the acquisition time ([11]). In other medical applications where γ imaging serves to the monitoring of the treatment, the number of emitted photons is too small to allow three dimensional imaging with collimated cameras. Additionally, the collimators hardly cope with a large energy spectrum. For this reasons, applications as treatment monitoring in proton therapy or selective internal radiation therapy are even of more

interest but also more challenging (see *e.g.*, [12, 13, 14, 15, 16]).

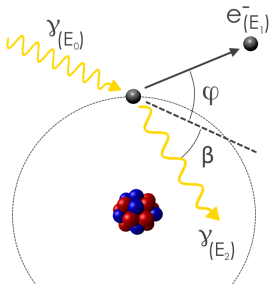


Figure 1: Inelastic scattering of a photon on an electron. A part of the energy of the photon is transferred to the electron and the last one is ejected from the atom's shell. In the simplistic model of the electron at rest the two angles, β and φ, are uniquelly determined by the energies.

Compton cameras detect γ rays that interact with the detector at least twice. The first interaction should be a Compton scattering, where a part of the energy of the ray is transferred to the recoiling electron (see figure 1). The last one deposits its energy in the detector, making possible to identify its position V_1 and to measure the energy E_1 lost by the photon. The γ particle should further interact at least once at some point V_2. When the energies E_1 of the recoiled electron and E_0 of the γ ray are known, the scattering angle may be deduced through the formula

$$\cos\beta = 1 - \frac{m_e c^2 E_1}{(E_0 - E_1)E_0}, \tag{1}$$

with m_e the mass of the electron at rest and c the speed of the light. In the absence of noise, the point where the gamma ray was emitted lies on the surface of the cone having vertex at V_1, axis directed by the unitary vector $\mathbf{\Omega} = \overrightarrow{V_2V_1}/\|\overrightarrow{V_2V_1}\|$ and half-opening angle β. Here $\|\ \|$ denotes the usual Euclidean norm of a vector in the three-dimensional space. As a consequence, data acquired with a Compton camera having ideal position and energy resolutions use to be modelled as integrals of the weighted intensity of the source on conical surfaces. These models are gathered under the name of weighted conical Radon transforms, cone transforms or Compton transforms.

2.2. The weighted conical Radon transform

More precisely, the weighted conical Radon transform use to be defined as follows. For some given cone, let us denote $\alpha \in [0, \pi]$ and $\delta \in (-\pi, \pi]$ the polar and azimuthal angles of the axis direction $\mathbf{\Omega}$, $\mathbf{u} \in \mathbb{R}^3$ the vector of coordinates of its vertex and $\beta \in (0, \pi)$ its half-opening angle. The cone may then be defined as the set

$$C(\mathbf{u}, \alpha, \delta, \beta) = \{\mathbf{v} \in \mathbb{R}^3 :\ (\mathbf{v} - \mathbf{u}).\mathbf{\Omega} = \|\mathbf{v} - \mathbf{u}\| \cos\beta\}, \tag{2}$$

and the projection associated to it as

$$\mathscr{C}^d_{\alpha,\delta,\beta} f(\mathbf{u}) = \int_{C(\mathbf{u},\alpha,\delta,\beta)} f(\mathbf{v}) \frac{1}{\|\mathbf{v} - \mathbf{u}\|^d} d\mathbf{v}, \tag{3}$$

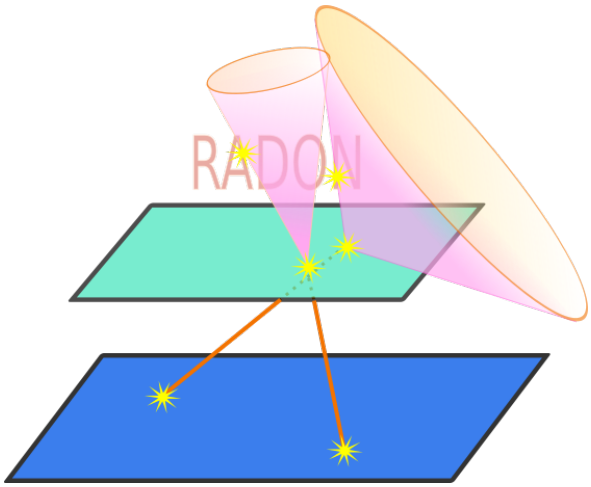


Figure 2: Schematic representation of a Compton camera composed of a scatterer (green) and an absorber (blue). The source is represented as the word RADON. Photons emitted from the source are first scattered then absorbed in the detector, and the position where they were emitted can only be determined at this stage as belonging to a conical surface.

where f is the intensity function and d use to be an integer. The weighted conical Radon transform of a function f integrable with respect to the measure $\|\mathbf{v}-\mathbf{u}\|^{-d}d\mathbf{v}$ over almost all cones with parameters in some manifold $\mathscr{M} \in \mathbb{R}^3\times[0,\pi]\times(-\pi,\pi]\times(0,\pi)$, is the application

$$(\mathbf{u},\alpha,\delta,\beta) \in \mathscr{M} \mapsto \mathscr{C}^d_{\alpha,\delta,\beta} f(\mathbf{u}). \tag{4}$$

A first restriction to some $\mathscr{M}$ may be imposed by the integrability condition. A second restriction comes from the particularity of the conical Radon transform to be redundant, since the dimension of the data space is larger than the dimension of the image space. Consequently, its inversion requires to select a sub-set from the data and, when possible, to average several inverses in order to reduce the noise.

In the last years some inversion methods accompanied by adapted selection techniques were developed. In [17], Basko *et al* have shown that the three-dimensional Radon projections on planes containing a given point can be calculated from the set of projections on cones having the vertex at that point, arbitrary axis direction and given constant half-opening angle. Cone-beam projections (also called divergent-beam projections) can be evaluated from conical projections having the same vertex (see *e.g.*, the works of L. Parra [18] and Tomitani and Hirasawa [19]). Detectors placed on a line were considered by Jung and Moon in [20].

The projections on cones with axes perpendicular to some surface provide a complete set. Inversion formulas in such configurations were given by Cree and Bones [21] and M. Haltmeier [22] for the plane, Schiefeneder and Haltmeier [23] for the sphere, S. Moon [24] for surfaces of revolution and Moon and Haltmeier [25] for the cylinder. Inversion methods that restrict the useful set of projections to cones with axes perpendicular on a manifold will generally neglect a large part of the acquired data and

thus provide very noisy images. A particular case where they may be of interest is the one of small detectors that move around the patient (see *e.g.*, [26]). Such detectors can anyway measure projections only for small values of α.

When the values of the projections are known for single axis direction and half-opening angle but for all vertices $u \in \mathbb{R}^3$, the mapping (4) can be seen as a convolution of f with a distribution supported on the cone. Deconvolution formulas were given by Gouia-Zarrad and Ambartsoumian ([27]), F. Terzioglu ([28]) and V. Palamodov ([29]). The inconvenient of this approach for the application to Compton camera is that measurements are required in all points of the space, including inside the source. This is obviously impossible to realize since the detectors are always placed outside the body. The methods proposed by B. Smith ([30]), Kuchment and Terzioglu ([31, 32]), Terzioglu *et al* ([33]) and one of the methods given by V. Palamodov in [29] allow to select the points of some manifold $\mathscr{M} \subset \mathbb{R}^3$ that do not intersect the source. The manifolds $\mathscr{M}$ are characterised by some admissibility (see [31, 32]) or completness (see [30]) conditions that are very similar to each other, with specificities related to the inversion formula.

Connected to the topic we discuss here is the inversion of the cumulative (or compound) conical Radon transform addressed *e.g.*, in [34] by Nguyen and Truong. This transform arises in nuclear medicine imaging when one wants to make an image with the photons scattered inside the patient. The differences in the inversion methods are however important and the problem is more complicated in the sense that the projections are not measured individually.

2.3. Conical Radon transform for cones with vertices in a plane

A slightly different class of weighted conical transforms was analysed in [35] under the name of Compton transforms. Their specificity is to employ a weight different from (3), weight that accounts for the incident angle of the incoming particle, and to consider the set of cones with vertices in a plane. Two transforms from the class differ either by the polar angle α or by the half-opening angle β and can be averaged together for noise variance reduction as shown in [10]. The main result of [35] is the central-slice theorem from section 4.1, proving the invertibility of the transform. Derived afterwards in [10], the associated filtered back-projection formula benefits from a relative simplicity and also from the fact that conical projections with axis directions not orthogonal to the detector are allowed.

To ensure uniqueness of the inverse, it is assumed that all the projections on cones with vertices in some given plane are measured. This completeness condition is obviously not verified in practice. The method can still give reasonably good results for Monte-Carlo simulated data, providing addition of some smoothing. In [10], this was obtained by frequency damping. In this paper we propose to use total variation either for denoising the solution or as a regularization step associated with an iterative algorithm.

2.4. Total variation regularization

Originally proposed by Rudin *et al* in [8], total variation penalty function is recommended for denoising piecewise constant images, as it promotes smoothness while still conserving sharp edges. Total variation may also improve tomographic reconstruction of images from low-dose acquisitions (see *e.g.*, [36]), or from limited-angle data (see *e.g.*, [37, 38]).

We suppose without loose of generality that the volume is a cube divided in N^3 voxels, rearranged in a vector $\mathbf{g}$, and we note $X = \mathbb{R}^{3N}$. The volume is transformed through a linear map of matrix R and the measurements $\mathbf{p}$ are noisy realizations of $R\mathbf{g}$. We search for $\mathbf{g}$ solution of the regularized problem:

$$\arg\min_{\mathbf{g}\in X} \left\{ \frac{1}{2}\|R\mathbf{g} - \mathbf{p}\|_2^2 + \mu TV(\mathbf{g}) \right\}, \tag{5}$$

with $\mu \geq 0$ the regularization parameter and TV the total variation norm. For Ω an open subset of $\mathbb{R}^3$, the total variation of a function $g \in L_1(\Omega)$ is defined by:

$$TV(g) = \sup \left\{ \int_\Omega g(x) \mathrm{div}\,(\varphi(x)) dx : \ \varphi \in C_c^1(\Omega; \mathbb{R}^3),\ \|\varphi(x)\| \leq 1 \ \forall x \in \Omega \right\}. \tag{6}$$

If g has gradient $\nabla g \in L_1(\Omega)$, then $TV(g) = \int_\Omega \|\nabla g(x)\| dx$. The discrete form of the total variation uses to be written as:

$$TV(\mathbf{g}) = \sum_{1\leq i,j,k\leq N} \|(\nabla \mathbf{g})_{i,j,k}\|, \tag{7}$$

and is not differentiable in zero. Let us put $Y = (\mathbb{R}^3)^{3N}$. An equivalent expression for the discrete total variation is:

$$TV(\mathbf{g}) = \sup \left\{ \langle \boldsymbol{\varphi}, \nabla \mathbf{g} \rangle_Y : \ \boldsymbol{\varphi} \in Y \text{ such that } \|\varphi_{i,j,k}\| \leq 1 \text{ for all } 1 \leq i,j,k \leq N \right\}, \tag{8}$$

where $\langle \boldsymbol{\varphi}, \boldsymbol{\psi} \rangle_Y = \sum_{i,j,k} (\sum_{\ell=1}^3 \varphi_{i,j,k}^\ell \psi_{i,j,k}^\ell)$ is the Euclidean dot product in Y. With the appropriate definition of the discrete divergence, $\langle \boldsymbol{\varphi}, \nabla \mathbf{g} \rangle_Y = -\langle \mathrm{div}\, \boldsymbol{\varphi}, \mathbf{g} \rangle_X$. In this work we rely on the notations from [39], and we apply the Chambolle's algorithm proposed therein to solve the denoising problem when R is the identity matrix.

When R is not the identity matrix a splitting approach can be used, with a gradient descent step that decreases the value of the data fidelity term $\|R\mathbf{g} - \mathbf{p}\|_2$ and a TV-denoising step that smooths the solution. This approach was previously applied by Beck and Teboulle in [40] for deconvolution problems and the gradient descent algorithm was used for the first task. When R stands for the discrete Radon transform, the SIRT algorithm may be applied. SIRT may be considered as a weighted gradient descent algorithm with more stable results.

2.5. The SIRT algorithm for the inversion of the Radon transform

In this paper we make use of the iterative inversion of the Radon transform. The Simultaneous Iterative Reconstruction Technique (SIRT, [9]) is one of the algorithms widely employed for this task. Let R be the system matrix describing the discrete Radon

transform in 2D or the X-ray transform in 3D, $\mathbf{p}$ the vector of data and $\mathbf{g}$ the vector containing all the voxels of the volume we wish to reconstruct. The SIRT algorithm starts from an initial guess $\mathbf{g}^{(0)}$ that may be taken as the null vector or the filtered back-projection solution then iterates following the formula:

$$\mathbf{g}^{(k+1)} = \mathbf{g}^{(k)} - \lambda \left(R^* \left[\frac{R\mathbf{g}^{(k)} - \mathbf{p}}{R\mathbf{1}} \right] \right) / \left(R^* \mathbf{1} \right), \tag{9}$$

where λ is a descent step, $\mathbf{1}$ is a vector of ones and R^* is the Hermitian conjugate of R. The divisions of vectors is done component-wise. In all our numerical tests we take $\lambda = 0.2$. The SIRT algorithm may give reconstructions more precise than the filtered back-projection, but when the data is very noisy a regularization of the solution is necessary. The best way to introduce this regularization is to add it as a penalty term in the minimization problem, as explained in section 2.4.

3. Theoretical and numerical inversion of the conical Radon transform

In this section we give an analytic inversion formula for the transform defined in equation (3) for $d = 1$, where the integral of the intensity of the source is weighted by the inverse of the distance from the source point to the vertex of the cone. We propose three equivalent formulations for the inverse and discuss them from a numerical and discrete point of view. One of the methods we propose is semi-iterative, in the sense that a part of the computation, aiming to transform the Compton data in two-dimensional Radon data, is done analytically. Then an iterative algorithm alternating SIRT ([9]) iterations and TV denoising is applied in order to calculate a regularized inverse of the Radon transform. The other methods are analytical and based on the commutativity of the Radon transform with differentiation.

3.1. Discussion on the Compton camera data modelling

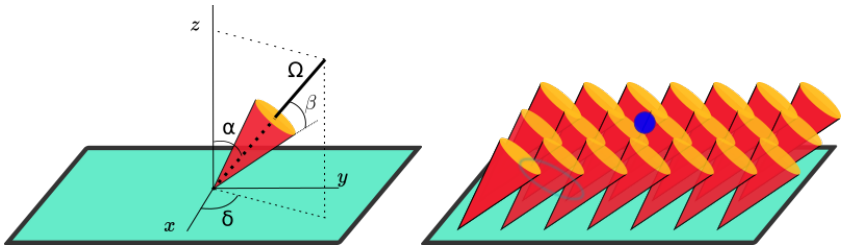
We consider in this work cameras with planar detectors (see figure 2). In the scope of a simpler presentation and without losing generality, hereafter we consider a camera containing a single scatterer layer placed at altitude $z_s = 0$ and an absorber placed at some altitude $z_a < 0$. When the scatterer has several layers, a conical transform should be defined for each layer. The intensity f is then calculated as the mean of their inverses. We suppose that the probabilities for an incident γ ray to be Compton scattered in the first detector and then completely absorbed in the second detector are both equal to one, so that these probabilities do not appear in the model. The differential cross section of a γ photon having initial energy E_0 and scattered by an angle β, is given by the Klein-Nishina formula ([41]), and is denoted by $K(\beta, E_0)$. For a given cone, let $\mathbf{u} \in \mathbb{R}^2 \times \{0\}$ be the position vector of the vertex and $\mathbf{\Omega} = (\sin\alpha\cos\delta, \sin\alpha\sin\delta, \cos\alpha)$ the direction vector of its axis, where α and δ are its polar and azimuthal angles (see figure 3 (a)).

When a source of γ particles having intensity distribution represented by a function $f : \mathbb{R}^3 \to \mathbb{R}$ is observed by a Compton camera, the fraction of events recorded with

parameters $(\mathbf{u}, \alpha, \delta, \beta)$ was considered in [35] proportional to

$$\begin{aligned} \mathscr{C}_{\alpha,\delta,\beta} f(\mathbf{u}) &= K(\beta, E_0) \cos\alpha \int_{C(\mathbf{u},\alpha,\delta,\beta)} f(\mathbf{v}) \cos\theta d\mathbf{v} \\ &= K(\beta, E_0) \cos\alpha \int_{C(\mathbf{u},\alpha,\delta,\beta)} f(\mathbf{v}) \frac{z}{\|\mathbf{v}-\mathbf{u}\|} d\mathbf{v}, \end{aligned} \tag{10}$$

where $\mathbf{v} = (x, y, z)$ is a vector from $\mathbb{R}^3$ and θ is the polar angle of the vector $\mathbf{v} - \mathbf{u}$. The factors $\cos\theta$ and $\cos\alpha$ from (10) account for the arbitrary incidence angle of the γ ray on the scatterer and absorber, respectively. For any given parameters α, δ, β, with $\alpha \in [0, \pi]$, $\beta \in (0, \pi)$ and $\delta \in (-\pi, \pi]$, the function $\mathscr{C}_{\alpha,\delta,\beta} f : \mathbb{R}^2 \times \{0\} \to \mathbb{R}$ was called in [35] a Compton projection of f (see figure 3 (b)). In order to keep the apelation from [35] and to distinguish it from other conical Radon transforms from the literature, the integral transform $\mathscr{C}$ operating on functions $f : \mathbb{R}^3 \to \mathbb{R}$ is hereafter referred to as the Compton transform.



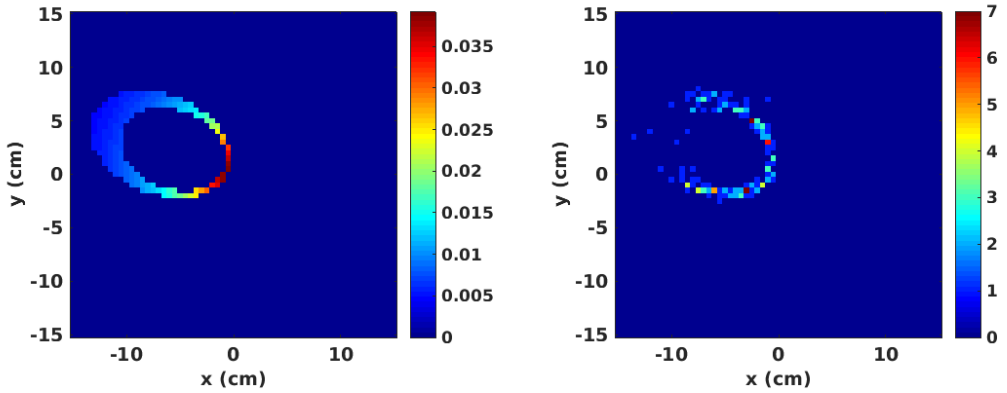
(a) (b)

Figure 3: (a) Parameterization of the Compton cone. (b) Construction of the Compton projection $\mathscr{C}_{\alpha,\delta,\beta}$. The Compton projection is the image drawn on the surface of the scatterer by the vertices of the cones with parameters α, δ, β that intersect the source. The intensity of the print is the integral from (10).

In figure 4 are represented a deterministic Compton projection and its counterpart from a Monte-Carlo simulation. The source is a small sphere. The deterministic projection from (a) was calculated in each pixel of the scatterer following equation (10). In (b) a Compton camera acquisition with a small spherical source emitting γ photons was simulated with the software MEGAlib (The Medium-Energy Gamma-ray Astronomy library, [42]), software based on the nuclear physics simulation library Geant4 [43]. The simulation in figure 4 accounts for Doppler broadening *i.e.*, small errors on the measured energies produced when the photon is scattered on an electron which is not at rest. In this example, the positions of the hits and the sum of energies are exact.

In (10) the weight $\cos\theta$ only accounts for the incidence angle of the γ ray on the scatterer. In a series of papers (see *e.g.*, [44]), Wilderman *et al* suggested that the weight should also depend on the distance $\|\mathbf{v} - \mathbf{u}\|$ through the solid angle subtended by the element of detector at the source. The model,

$$\mathscr{C}^*{}_{\alpha,\delta,\beta} f(\mathbf{u}) = K(\beta, E_0) \int_{C(\mathbf{u},\alpha,\delta,\beta)} f(\mathbf{v}) \frac{\cos\theta}{\|\mathbf{v}-\mathbf{u}\|^2} d\mathbf{v}, \tag{11}$$



(a) (b)

Figure 4: Example of Compton projections of a small spherical source. In (a) the projection is calculated deterministically following (10). In (b) a Monte Carlo simulation of the same detector-source configuration is shown. The model serving for the deterministic projection (a) only approximatelly describes the data in (b).

was recently validated against Monte-Carlo simulations ([45]). As the source is supposed to lie in the half-space $z > 0$, the function f is zero for values of $\mathbf{v}$ close to $\mathbf{u}$. Thus there is no singularity in the expression under the integral (11). As we shall show in section 3.2, the inversion formula given in [10] for (10) can be adapted easily to $d = 1$ and the transform

$$\mathscr{C}^1_{\alpha,\delta,\beta} f(\mathbf{u}) = K(\beta, E_0) \int_{C(\mathbf{u},\alpha,\delta,\beta)} f(\mathbf{v}) \frac{1}{\|\mathbf{v}-\mathbf{u}\|} d\mathbf{v}, \tag{12}$$

which is closer to the model suggested by the Monte Carlo simulations and in practice seems to give better results.

3.2. Inversion of the conical Radon transform

In this section we derive a formula allowing to calculate the source intensity from the conical Radon projections (12). To avoid technical difficulties throughout the following, we assume f to be Schwartz function supported in the upper half space.

Let $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ be the standard basis of $\mathbb{R}^3$. The vectors $\mathbf{d}_1 = (\cos\delta, \sin\delta, 0)$, $\mathbf{d}_2 = (-\sin\delta, \cos\delta, 0)$ and $\mathbf{e}_3$ also form an orthonormal basis of $\mathbb{R}^3$, defining a reference frame $Otsz$ obtained by rotation of $Oxyz$ about the Oz axis. Let us consider the cone $C(\mathbf{u}_0, \alpha, \delta, \beta)$ with $\mathbf{u}_0 = t_0\mathbf{d}_1 + s_0\mathbf{d}_2$. Since the source belongs to the half-space $z > 0$, for a single-layer scatterer one should always have $\alpha \in \left[0, \frac{\pi}{2}\right)$. No restriction apply on δ, which is taken in the interval $(-\pi, \pi]$. Physically, the Compton angle β can take any value in $(0, \pi)$. However the parametrization we choose hereafter for the cone constrains to keep only the Compton projections with $\alpha+\beta < \frac{\pi}{2}$, leading to a selection of the cones that intersect the scatterer in a single point, the vertex. For such admissible values of

α and β one has $\cos^2\alpha - \sin^2\beta > 0$. We define:

$$a(\alpha,\beta) = \frac{\sin\beta\cos\beta}{\cos^2\alpha - \sin^2\beta}, \tag{13}$$

$$b(\alpha,\beta) = \frac{\sin\beta}{\sqrt{\cos^2\alpha - \sin^2\beta}}, \tag{14}$$

$$c(\alpha,\beta) = \frac{\sin\alpha\cos\alpha}{\cos^2\alpha - \sin^2\beta}. \tag{15}$$

The intersection of the cone with a horizontal plane at some given altitude $z > 0$ is an ellipse that may be described by the parametric equation:

$$\begin{cases} t = & t_0 + zc(\alpha,\beta) & +za(\alpha,\beta)\cos\varphi \\ s = & s_0 & +zb(\alpha,\beta)\sin\varphi \end{cases}, \quad \varphi \in (-\pi,\pi]. \tag{16}$$

With the cone thought of as a stack of ellipses, its parametric equation may then be readily obtained. The surface integral (12) at a point from the scatterer having coordinates $\mathbf{u}_0 = t_0\mathbf{d}_1 + s_0\mathbf{d}_2$, may then be expressed as

$$\begin{aligned} &\mathscr{C}^1_{\alpha,\delta,\beta} f(t_0\mathbf{d}_1 + s_0\mathbf{d}_2) = K(\beta,E_0)b(\alpha,\beta) \\ &\times \int_0^\infty \int_0^{2\pi} f((t_0 + zc(\alpha,\beta) + za(\alpha,\beta)\cos\varphi)\mathbf{d}_1 + (s_0 + zb(\alpha,\beta)\sin\varphi)\mathbf{d}_2 + z\mathbf{e}_3)d\varphi dz. \end{aligned} \tag{17}$$

Hereafter we make use of the two-dimensional Radon transform, that associates to a function $g:\ \mathbb{R}^2 \to \mathbb{R}$ with bounded support a set indexed on the parameter $\delta \in (-\pi,\pi]$ of line integrals $\mathscr{R}_\delta$, defined as:

$$\forall\, s \in \mathbb{R}, \quad \mathscr{R}_\delta g(s) = \int_{-\infty}^{\infty} g(s\cos\delta + t\sin\delta, s\sin\delta - t\cos\delta)dt. \tag{18}$$

We note $f_z : (x,y) \in \mathbb{R}^2 \mapsto f(x,y,z) \in \mathbb{R}$ the restriction of f to some horizontal plane. For $\tau \geq 0$ and $\delta \in (-\pi,\pi]$, we define the integral transform given for all $s \in \mathbb{R}$ by:

$$\mathscr{P}_{\tau,\delta} f(s) = \tau \int_0^\infty \int_{-\pi}^{\pi} \mathscr{R}_{\delta+\frac{\pi}{2}} f_z(s + z\tau\sin\varphi)d\varphi dz. \tag{19}$$

Simple calculations starting from (17) and (18) allow to show that the transforms $\mathscr{P}$ and $\mathscr{C}^1$ are related to each other by

$$\mathscr{P}_{\tau,\delta} = \frac{1}{K(\beta,E_0)} \mathscr{R}_{\delta+\frac{\pi}{2}} \mathscr{C}^1_{\alpha,\delta,\beta}, \tag{20}$$

whenever $\tau = b(\alpha,\beta)$. The transform $\mathscr{P}$ produces a data set included in a three dimensional space, indexed by $\tau > 0$, $\delta \in (-\pi,\pi]$ and $s \in \mathbb{R}$. The required data set may be further reduced by taking into account the identity

$$\mathscr{R}_{\delta+\frac{\pi}{2}} \mathscr{C}^1_{\alpha,\delta+\pi,\beta} = \mathscr{R}_{\delta+\frac{\pi}{2}} \mathscr{C}^1_{\alpha,\delta,\beta}, \tag{21}$$

meaning that from the point of view of the transform $\mathscr{P}$, the projections $\mathscr{C}^1_{\alpha,\delta+\pi,\beta}$ and $\mathscr{C}^1_{\alpha,\delta,\beta}$ share the same information.

Taking the Fourier transform in (19) leads to:

$$\widehat{\mathscr{P}_{\tau,\delta} f}(\sigma) = 2\pi\tau \int_0^\infty \widehat{\mathscr{R}_{\delta+\frac{\pi}{2}} f_z}(\sigma) J_0(2\pi z\tau\sigma)dz, \tag{22}$$

where J_0 is the zero-order Bessel function of the first kind. Then, by the projection-slice theorem applied to the Radon projection and for $d_2 = (-\sin\delta, \cos\delta, 0)$ we get:

$$\widehat{\mathscr{P}_{\tau,\delta} f}(\sigma) = 2\pi\tau \int_0^\infty \frac{1}{z} \widehat{f_z}(\sigma d_2) J_0(2\pi z\tau\sigma) z dz. \tag{23}$$

In (23), $f_z = 0$ for z close to the origin. Formulas (22), (23) are valid for all Schwartz functions f supported in the upper half space and both appearing integrals are absolutely convergent. We recognize in (23) the Hankel transform of $\widehat{f_z}/z$. Finally the inversion of the Hankel transform leads to

$$\widehat{f_z}(\sigma d_2) = 2\pi z\sigma^2 \int_0^\infty \widehat{\mathscr{P}_{\tau,\delta} f}(\sigma) J_0(2\pi z\tau\sigma) d\tau, \tag{24}$$

and to the reconstruction formula:

$$f(x,y,z) = 2\pi z \int_0^\pi \int_0^\infty \left(\int_{-\infty}^\infty \widehat{\mathscr{P}_{\tau,\delta} f}(\sigma) J_0(2\pi z\tau\sigma) |\sigma|^3 e^{2i\pi\sigma(-x\sin\delta + y\cos\delta)} d\sigma \right) d\tau d\delta. \tag{25}$$

Note that when $b(\alpha,\beta) = \tau > 0$, either α or β may be calculated from (14) as function of the other parameter through the relation:

$$\tau\cos\alpha = \sqrt{1+\tau^2}\sin\beta, \tag{26}$$

since both applications $\beta \in \left(0, \frac{\pi}{2} - \alpha\right) \mapsto b(\alpha,\beta) \in \mathbb{R}_+^*$, for some given $\alpha \in \left[0, \frac{\pi}{2}\right)$, and $\alpha \in \left[0, \frac{\pi}{2} - \beta\right) \mapsto b(\alpha,\beta) \in \mathbb{R}_+^*$, for some given $\beta \in \left(0, \frac{\pi}{2} - \alpha\right)$, are one-to-one. This means that in (25), the projections $\mathscr{P}_{\tau,\delta}$ can be calculated as the mean of an infinity of conical projections, one for each polar angle α or one for each Compton angle β.

Numerically, the reconstruction can be done slice by slice (following the direction orthogonal to the camera) as follows. The Radon projections of the Compton projections are filtered in the Fourier domain with a filter depending on the parameter τ, related to the cone pattern generating the projection, and on the altitude z. The filtered projections are then integrated on τ then back-projected on lines in each horizontal slice. The filtering in the Fourier domain relies on the evaluation of the Bessel function J_0 on a grid sufficiently fine to avoid aliasing as much as possible. Consequently, zero-padding of the projections is necessary implying intensive use of the memory when all projections are processed simultaneously.

The evaluation of the reconstruction formula may produce very noisy results as the reconstruction filter $|\sigma|^3$ amplifies the noise in the data while emphasizing high frequencies. A practical yet insufficient solution is the damping of the higher frequencies during the reconstruction. Some post-filtering remains necessary and may be obtained by total variation denoising.

3.3. Alternative formulations of the inverse Compton transform

In a tempered distribution sense the Bessel function $\sigma \mapsto J_0(2\pi z\tau\sigma)$ is the Fourier transform of

$$k_{z,\tau}(s) = \frac{1}{2\pi} \frac{2\mathrm{rect}\left(\frac{s}{2z\tau}\right)}{\sqrt{(z\tau)^2 - s^2}}, \tag{27}$$

where rect denotes the rectangular function. With ramp the impulse response of the ramp filter $\sigma \mapsto |\sigma|$, the filtered projections may be rewritten as:

$$\int_{-\infty}^{\infty} \widehat{\mathscr{P}_{\tau,\delta} f}(\sigma) J_0(2\pi z\tau\sigma)|\sigma|^3 e^{2i\pi\sigma \mathbf{v}.\mathbf{d}_2} d\sigma = -\frac{1}{(2\pi)^2} \left(\mathscr{P}_{\tau,\delta} \star k_{z,\tau} \star \text{ramp}\right)''(\mathbf{v}.\mathbf{d}_2). \tag{28}$$

With the notation $g_{z,\tau,\delta} = \mathscr{P}_{\tau,\delta} \star k_{z,\tau}$ we then have

$$\int_{-\infty}^{\infty} \widehat{\mathscr{P}_{\tau,\delta} f}(\sigma) J_0(2\pi z\tau\sigma)|\sigma|^2 e^{2i\pi\sigma(\mathbf{v}.\mathbf{d}_2)} d\delta = -\frac{1}{(2\pi)^2} (g_{z,\tau,\delta} \star \text{ramp})''(\mathbf{v}.\mathbf{d}_2). \tag{29}$$

The function $g_{z,\tau,\delta}$ may be calculated either in the temporal or the frequency domain. The first alternative seems preferable, because the highly oscillatory Bessel function requires a high sampling rate to limit aliasing as much as possible (see figures 5 and 6). This frequency domain sampling may be obtained by zero-padding the projections $\mathscr{P}_{\tau,\delta}$.

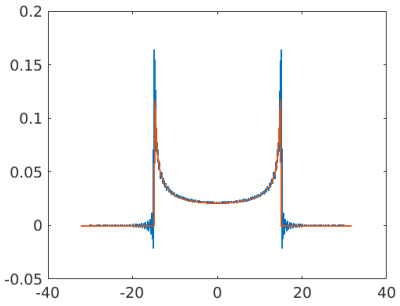


Figure 5: Inverse Fourier transform of the discretized Bessel function (blue line) compared to $k_{z,\tau}$ (red line).

Let us now put for each $s \in \mathbb{R}$,

$$p_z(s,\delta) = z \int_0^{\infty} g_{z,\tau,\delta}(s) d\tau. \tag{30}$$

From (29) we then deduce that

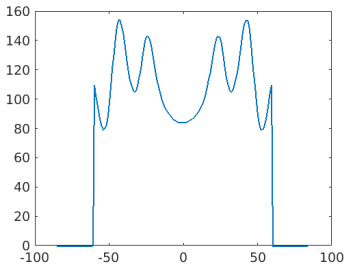
$$\int_0^{\infty} \left(\int_{-\infty}^{\infty} \widehat{\mathscr{P}_{\tau,\delta} f}(\sigma) J_0(2\pi z\tau\sigma)|\sigma|^3 e^{2i\pi\sigma(\mathbf{v}.\mathbf{d}_2)} d\delta \right) d\tau = -\frac{1}{(2\pi)^2 z} (p_z(\cdot,\delta) \star \text{ramp})''(\mathbf{v}.\mathbf{d}_2), \tag{31}$$

and finally from (25) we obtain

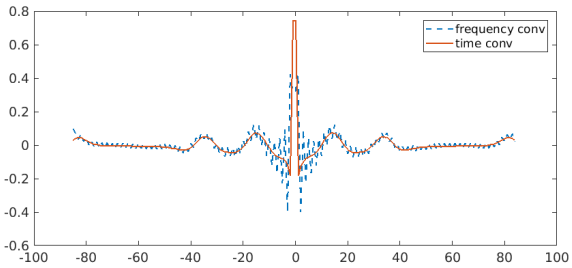
$$f(x,y,z) = -\frac{1}{2\pi} \left(\mathscr{R}^{-1} \left[\frac{\partial^2 p_z}{\partial s^2} \right] \right)(x,y), \tag{32}$$

where $\mathscr{R}^{-1}$ is the two-dimensional inverse Radon transform. We have thus shown that the image of the source may be reconstructed slice by slice as the inverse of the two-dimensional Radon transform, from synthetic Radon projections calculated by post-processing the Compton projections. From a numerical point of view, formula (32) has several advantages:

(i) It gives the possibility to use iterative methods combined with advanced filtering for the inversion of the Radon transform, still conserving a fast method compared to direct use of iterative methods on the raw Compton data;



(a) raw projection $\mathscr{P}_{\tau,\delta}$



(b) second derivative of filtered projection $g_{z,\tau,\delta}$

Figure 6: (a) Example of projection $\mathscr{P}_{\tau,\delta}$, *i.e.*, Radon projection of the Compton projection, for a source composed of three points. (b) Its filtered versions $g''_{z,\tau,\delta}$ for some arbitrary value z, with the convolution calculated in the frequency domain (blue doted line) and in the time domain (red continuous line). The oscillations due to aliasing that may be observed in the right-hand image are important, but this is not always the case. Aliasing is more visible for large values of $z\tau$. Since they are attenuated by the usual smoothing of projections in the reconstruction process, the influence of the filtering domain may finally have little impact on the images.

(ii) It has lower memory requirements when the convolution with the filter k is done in the time domain compared to Fourier domain convolution with the highly-oscillating Bessel function;

(iii) It makes use of a more precise calculation of the filtered projections $g_{z,\tau,\delta}$ which are no more subject to aliasing. An example is shown in figure 6.

In our experiments, only the first point appeared as important. The third point had little impact on the results, but this observation may be biased by the small distance we put between the source and the detector.

3.3.1. Semi-iterative inversion of the Compton transform

The images obtained from (32) may be denoised with a total variation filtering. However, post-filtering increases the projection errors and a better approach is to apply an iterative algorithm to the regularized problem as explained at the end of section 2.4. In this case we choose to solve a problem of type (5). Following the procedure described in section 2.4, SIRT steps are alternated with total variation denoising steps completed with the Chambolle's algorithm from [39]. If λ is the descent step of SIRT and μ is the regularization parameter multiplying the TV norm in (5), the parameter of the total variation denoising should be taken equal to the product $\lambda\mu$.

3.3.2. Analytic solution written as a second derivative

In (32) the back-projection is preceded by the differentiation of the projections. The derivative is calculated by finite differences and this operation induces some circular blurring in the reconstructed images.

The commutativity of the Radon transform with derivatives is a classical result (see *e.g.*, [46]). For any $g : \mathbb{R}^2 \mapsto \mathbb{R}$ sufficiently smooth,

$$\mathscr{R}_\delta\left(\partial_1 g\right) = (\cos\delta)(\mathscr{R}_\delta g)' \quad \text{and} \quad \mathscr{R}_\delta\left(\partial_2 g\right) = (\sin\delta)(\mathscr{R}_\delta g)'.$$

and when $\Delta = \partial_{11} + \partial_{22}$ is the Laplace operator,

$$\mathscr{R}_\delta\left(\Delta g\right) = (\mathscr{R}_\delta g)''.$$

We thus may rewrite (32) as

$$f(x,y,z) = -\frac{1}{2\pi}\Delta(\mathscr{R}^{-1}p_z)(x,y). \tag{33}$$

In the continuous case this formula is equivalent to (32) and thus to (25). In the discrete case the equivalence do not hold any more and formula (33) may produce less blurry images. Numerically, the Laplacian can be calculated by finite differences. In this work we used the del2 function from MATLAB.

4. Description of the data

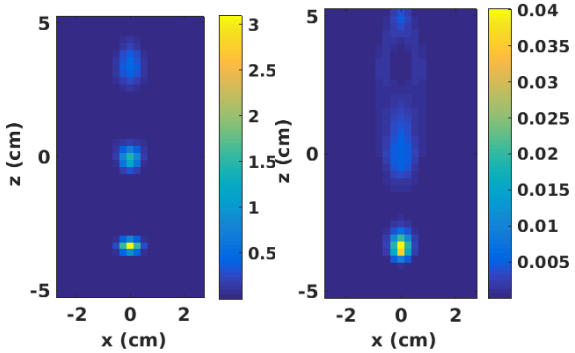
We generated Compton camera data with the Monte-Carlo simulation software MEGAlib (The Medium-Energy Gamma-ray Astronomy library) described in [42]. This software is composed of several modules, and is based on the Geant4 library [43]. The COSIMA module allows to simulate the interactions between the particles and the matter. The output consists in a list of interactions for each individual photon and for the secondary particles produced during these interactions. The ordering of the hits in the sequence is currently available only in simulations, as the time resolution of the detectors is too poor. In a real acquisition, the camera would provide a set of (not ordered) interactions and specific algorithms are required to reconstruct the sequence of hits.

The Monte-Carlo data may then be analysed with one of the two modules SIVAN and REVAN. The first one retrieves the ideal positions of the hits and the exact energies of the secondary particles. Only Doppler broadening may affect the measurements. Additionally the order of the interactions is known in this case and the vertex and axis of the cone are thus the true ones. When the data are processed with REVAN, errors are added both on positions and energies. These errors reverberate on the value of the scattering angle, on the position of the vertex and on the direction of the axis of the cone. The order of the interactions is determined based on kinematics considerations and may be erroneous. The first hit is eventually wrongly identified and leads to a miss-positioning of the vertex of the cone. The same may hold for the second hit, with consequences on the axis direction.

4.1. Influence of the errors on the angular parameters

The errors on the angular parameters of the Compton cone have important consequences on the reconstructed images regardless the algorithm. For instance in figure 7 we show

iterative Maximum Likelihood Expectation-Maximization (MLEM) reconstructions from deterministic projections of three identical point-like sources. The algorithm was originally adapted to Compton camera imaging by Wilderman *et al* in [44] and here we use the implementation from [14]. In (a) the projections are simulated with $\alpha = 2.5°$ and reconstructed as they were acquired at $\alpha = 0°$. In (b), the projections for $\beta = 10°$ and $\beta = 15°$ are summed together then reconstructed as they were taken at $\beta = 12.5°$. In both cases the intensities of the sources are different after reconstruction and the blur increases as we move further from the detector. Whilst the errors on α produced in this experiment only an elongation of the source, the influence of the errors on β is more critical.



(a) (b)

Figure 7: Image reconstructed with the Maximum Likelihood Expectation-Maximization algorithm (MLEM) from projections with (a) an error of 2.5° on the axis inclination α and (b) an error of 2.5° on the Compton scattering angle β (see text for details). Twenty iterations of the MLEM algorithm were runned.

Figure 7 shows that the artefacts in the vertical direction are not specific to analytic reconstruction. They are not necessarily related to an erroneous modelling of the data since in this example the model from which the data is generated is perfectly known. Elongation of the source in the direction orthogonal to the camera is usual in Compton camera imaging and is often related to the small solid angle subtended by the detector at a point from the source. As it is produced by a loss of projections, it could be assimilated to a "limited angle" artefact. In this example the camera was designed to be relatively large and close to the sources. We shall thus conclude that data binning strongly reinforces the artefacts in the vertical direction.

4.2. Compton camera simulated data

For the purpose of evaluation of the analytic methods we simulated a Compton camera composed of six Silicon scatterer layers and one absorber composed of $LaBr_3$ crystals. Each layer has dimensions $30 \times 30 \times 0.2$ cm^3 and the absorber is composed of 128^2 crystals arranged in a 73 cm-side square. Each crystal has dimensions $0.5 \times 0.5 \times 4$ cm^3.

The scatterers are placed with the centres at altitudes $z \in \{-5, -6, \ldots, -10\}$ and the absorber at $z = -22$.

For the noisy data, the spatial resolution in the scatterer is about 1 mm. In the absorber, the hit is always considered at the center of a crystal. The energy errors are randomly drawn from Gaussian distributions. In the scatterer its standard deviation is 1 keV. In the absorber, the standard deviation depends on the measured energy. For instance its value is 11.4 keV at 511 keV.

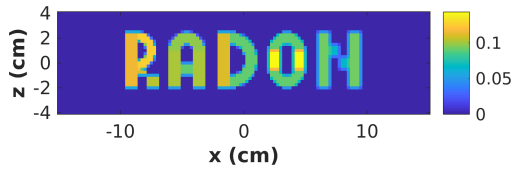


Figure 8: The reference image - one slice from a 1 cm thick source representing the word RADON.

The non-uniform source intensity has the form of the word RADON, with a thickness of 1 cm in the third direction (see figure 8). A total number of 10^9 γ-particles was emitted with energies 511 keV. After discarding not Compton and not reconstructible events (see conditions on the angles α and β in section §3.2), about 17×10^6 reconstructible events were selected in the ideal case and about 8.1×10^6 in the realistic case.

The data were grouped in a five-dimensional array. The dimensions correspond to the layer, the azimuthal angle α, the polar angle δ, the parameter τ and the coordinate s in the Radon projection of the Compton projection $\mathscr{P}_{\tau,\delta}$. The resulting array had $6 \times 45 \times 90 \times 45 \times 170$ elements. The higher the sampling density for the angular parameters, the better the reconstructed image should be as the sampling errors produce blur and artefacts that are much more severe than in imaging modalities based on line projections (see figure 7).

5. Results and discussions

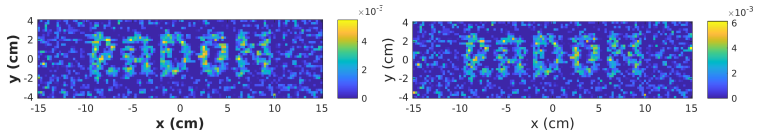
Volumes having $120 \times 120 \times 33$ cubic voxels with side of 2.5 mm were reconstructed from ideal and noisy data. The resolution in the direction perpendicular to the camera is known to be worse than in slices parallel to it. We thus considered two situations where the word was placed either parallel to the camera or orthogonal to it. We also tested a reduction of the number of counts by a factor ten.

We compare three reconstruction methods: *analytic* stands for equation (32), *Laplacian* stands for (33) and *semi-iterative* stands for the algorithm described in paragraph 3.3.1.

5.1. Resolution in slices parallel to the detector

In this test we place the word-shaped source parallel to the detector with the center at altitude $z = 0$. The central slice is thus at 5 cm from the detector.

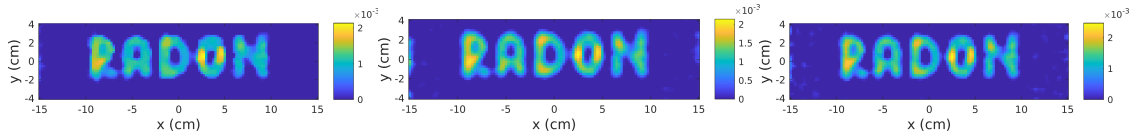
The first example is made with ideal data. We recall that ideal data include however some errors on energies induced by Doppler broadening and also statistical errors coming from the random generation of particles in the source. Even in this case, the noise in the reconstructed images is consequent as it may be seen in figure 9. Some denoising is necessary and we suggest that TV denoising would be a good choice in order to preserve the sharp edges of the source.



(a) *analytic* (b) *Laplacian*

Figure 9: Central slices from the word-shaped source placed parallel to the detector. The data are processed with SIVAN, meaning that only mild noise is considered (see text for details). The *analytic* and *Laplacian* methods give similar and noisy results.

Figure 10 shows on the left and center the central cuts of the same volumes as in figure 9 but after TV denoising. On the right is shown the TV-regularized SIRT reconstruction. The results for the three methods are very similar although the *analytic* reconstruction is slightly blurry (see letters R and O).



(a) *analytic* ($\mu = 0.2$) (b) *Laplacian* ($\mu = 0.05$) (c) *semi-iterative* ($\mu = 0.01$)

Figure 10: The source is placed parallel to the detector and the collected events are ideal except the Doppler broadening. Central slices from the word-shaped source after TV denoising. For the semi-iterative method the result is shown after 30 iterations.

The evolution of the errors for the semi-iterative method is shown in figure 11 for the central slice. The figures of merit are the cost function from equation (5) displayed in blue and the root mean square error displayed in red. The iterations are started from the *analytic* solution.

When a realistic noise is added to the data, in practice by treating the output of the simulation with the module REVAN, the results are largely degraded but the shape of the letters is almost conserved (see figure 12). It should be noted however that the size of the voxel is here of $(2.5 \text{ mm})^3$ which is very low when we know that the spatial

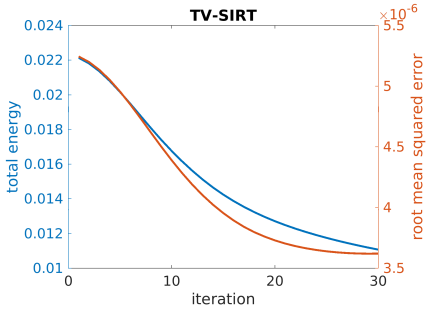
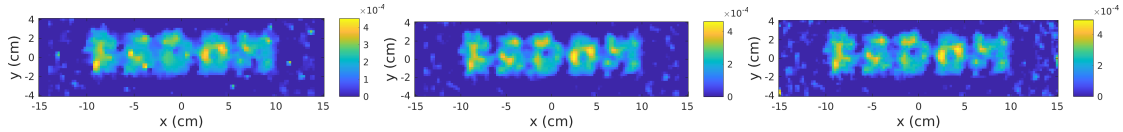


Figure 11: Evolution of the errors with the number of iterations for the regularized semi-iterative method. Ideal data, as produced in the output of the SIVAN module.

resolution currently obtained with collimated detectors in medical SPECT imaging is close to 5 mm. Some degradation factors as attenuation in the patient and random coincidences were not considered in our experiment.



(a) *analytic* ($\mu = 0.12$) (b) *Laplacian* ($\mu = 0.05$) (c) *semi-iterative* ($\mu = 0.01$)

Figure 12: Realistic data and TV denoising for the word-shaped source placed parallel to the detector. For the semi-iterative method the result is shown after 30 iterations.

Figure 13 shows the evolution of the errors for the semi-iterative method with realistic data.

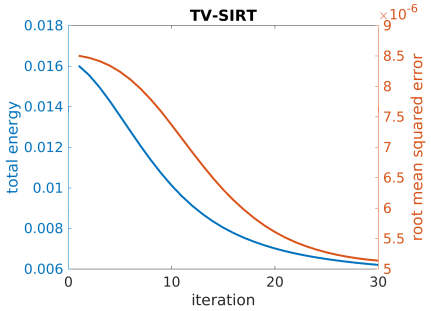


Figure 13: Evolution of the errors with the number of iterations for the regularized semi-iterative method. Realistic data, as produced in the output of the REVAN module.

5.2. Low statistics acquisition

The quantity of available data has a strong influence on the noise as it is shown in figure 14. The image is reconstructed from a sub-set of $1.6{\times}10^6$ ideal events taken from the

previous simulation (thus about one tenth of the sample).

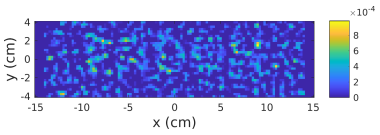
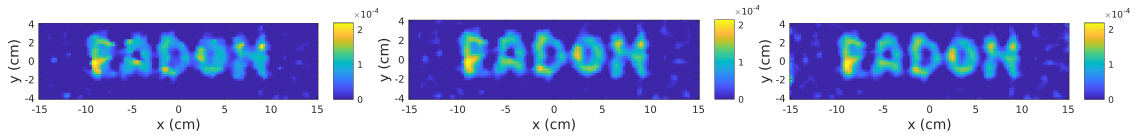


Figure 14: Low statistics reconstruction from ideal data, with the *analytic* formula.

Again, imposing low total variation to the reconstructed image helps to strongly improve the result. It may be noted in figure 15 that in this case the methods *Laplacian* and *semi-iterative* seem to outweigh the *analytic* method and to give sharper results.



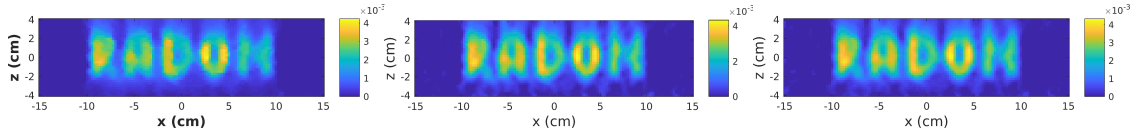
(a) *analytic* ($\mu = 0.15$) (b) *Laplacian* ($\mu = 0.07$) (c) *semi-iterative* ($\mu = 0.02$)

Figure 15: Low statistics acquisition. Central slices from the word-shaped source placed parallel to the detector after TV denoising. The data are processed with SIVAN. For the semi-iterative method the result is shown after 20 iterations.

5.3. Resolution in slices perpendicular to the detector

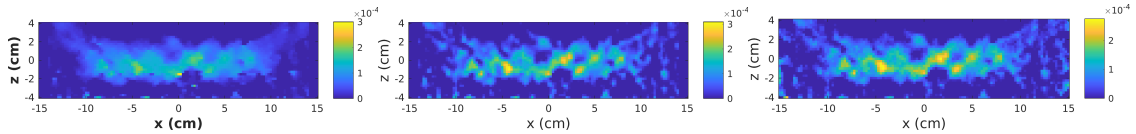
In this test we place the word-shaped source perpendicular to the detector. The bottom of the source is at 3 cm from the up-most scatterer. The central slice refers in this case to $y = 0$ and this slice is thus perpendicular to the detector, too. A blurring in the vertical direction can be observed. This blurring is usual in Compton camera imaging regardless the reconstruction method and uses to be attributed to projections loss due to the finite size of the detector. At our knowledge, the only theoretical study on the artifact produced by the data incompletness is the one from [29], where a "plume"-shaped artifact is mentionned but under the convolutional conical Radon transform model. In section 4.1 we showed that the measurement errors and the binning are additional degradation factors. The relative influence the geometry and the errors on the angular parameters may have is an important question not yet elucidated.

The first test is done with ideal data and the results are displayed in figure 16. Let us mention that even ideal data are affected by the Doppler broadening that induce errors on the Compton angle. Those errors have a bell-shaped distribution with full width at half maximum of a few degrees, the exact value depending on the material and on the temperature. None of the methods presented here could retrieve the shape of the source when realistic data was used. The result is in all cases blurry and noisy and only the rough location of the word can be established. Again, the results from figure 17 show that the *analytic* method give the blurriest results.



(a) *analytic* ($\mu = 0.15$) (b) *Laplacian* ($\mu = 0.04$) (c) *semi-iterative* ($\mu = 0.01$)

Figure 16: Word perpendicular to the camera and ideal data from SIVAN output. Central slices after TV denoising of the volume. For the semi-iterative method the result is shown after 20 iterations.



(a) *analytic* ($\mu = 0.15$) (b) *Laplacian* ($\mu = 0.04$) (c) *semi-iterative* ($\mu = 0.01$)

Figure 17: Word perpendicular to the camera and realistic data from REVAN output. Central slices after TV denoising of the volume. For the semi-iterative method the result is shown after 30 iterations.

6. Conclusions

The purpose of this work is to push the limit of the attainable resolution in Compton camera imaging. We developed an inverse for a conical Radon transform and we tested three implementations for its numerical calculation. We demonstrated that imposing low total variation to the reconstructed volume strongly improves the result even when the level of the noise in the image is high. The results given by the three methods were very similar for ideal high statistics data. When the quality of the data drops, the *Laplacian* method, where the derivation is applied to the volume instead of the projections, and the semi-iterative method gave slightly better results in terms of sharpness. The best solution both in time spent in the reconstruction and quality of the image seems to be the *Laplacian* method. In the less favourable situation where the resolution in the vertical direction is tested with noisy data we did not succeeded to reconstruct an acceptable image of the source. The errors on the axes directions, on the half-opening angle and on the sequence of interactions, cumulated with the approximations induced by the binning, prevent reconstructing a source with small details as the one we chose for our tests. It should be noted however that we did not applied here clever data processing, consisting *e.g.*, to select only events that have deposited in the detector a total energy close to the known initial energy of the photon. Our idea was to explore the limits of the methods, and in some situations the initial energy is unknown.

Acknowledgements

The author acknowledges the financial support of the French National Research Agency through the ANR project 3DCLEAN (ANR-15-CE09-0009) and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the ANR. The author would like to thank the anonymous reviewers for careful reading of the manuscript and thoughtful comments.

References

[1] M. Schönfelder, A. Hirner, and K Schneider. A telescope for soft gamma ray astronomy. Nuclear Instruments and Methods, 107(2):385–394, 1973.

[2] R. Todd, J. Nightingale, and D. Everett. A proposed Gamma camera. Nature, 251:132–134, 1974.

[3] M. Singh and D. Doria. An electronically collimated gamma camera for single photon emission computed tomography. Part II: Image reconstruction and preliminary experimental measurements. Medical Physics, 10(4):428–435, 1983.

[4] E.A. Wulf, B.F. Phlips, W.N. Johnson, J.D. Kurfess, E.I. Novikova, P. O'Connor, and G. De Geronimo. Compton imager for detection of special nuclear material. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 579(1):371–374, 2007.

[5] W. Lee and T. Lee. A compact Compton camera using scintillators for the investigation of nuclear materials. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 624(1):118–124, 2010.

[6] S. Takeda, A. Harayama, Y. Ichinohe, H. Odaka, S. Watanabe, T. Takahashi, H. Tajima, K. Genba, D. Matsuura, H. Ikebuchi, Y. Kuroda, et al. A portable Si/CdTe Compton camera and its applications to the visualization of radioactive substances. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 787:207–211, 2015.

[7] C.G. Wahl, W.R. Kaye, W. Wang, F. Zhang, J.M. Jaworski, A. King, Y.A. Boucher, and Z. He. The Polaris-H imaging spectrometer. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 784:377–381, 2015.

[8] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena, 60(1-4):259–268, 1992.

[9] P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. Journal of theoretical biology, 36(1):105–117, 1972.

[10] V. Maxim. Filtered backprojection reconstruction and redundancy in Compton camera imaging. IEEE Transactions on Image Processing, 23(1):332–341, 2014.

[11] M. Fontana, D. Dauvergne, J.M. Letang, J.L. Ley, and É. Testa. Compton camera study for high efficiency SPECT and benchmark with Anger system. Physics in Medicine & Biology, 62(23):8794, 2017.

[12] M. Frandes, A. Zoglauer, V. Maxim, and R. Prost. A tracking Compton-scattering imaging system for hadron therapy monitoring. IEEE Transactions on Nuclear Science, 57:144–150, 2010.

[13] S. Kurosawa, H. Kubo, K. Ueno, S. Kabuki, S. Iwaki, M. Takahashi, K. Taniue, N. Higashi, K. Miuchi, T. Tanimori, et al. Prompt gamma detection for range verification in proton therapy. Current Applied Physics, 12(2):364–368, 2012.

[14] E. Hilaire, D. Sarrut, F. Peyrin, and V. Maxim. Proton therapy monitoring by Compton imaging: influence of the large energy spectrum of the prompt-γ radiation. Physics in Medicine & Biology, 61(8):3127, 2016.

[15] S Aldawood, PG Thirolf, A Miani, M Böhmer, G Dedes, R Gernhäuser, C Lang, S Liprandi,

L Maier, T Marinšek, et al. Development of a compton camera for prompt-gamma medical imaging. Radiation Physics and Chemistry, 140:190–197, 2017.

[16] E Muñoz, J Barrio, D Bemmerer, A Etxebeste, F Fiedler, F Hueso-González, C Lacasta, JF Oliver, K Römer, C Solaz, et al. Tests of macaco compton telescope with 4.44 mev gamma rays. Journal of Instrumentation, 13(05):P05007, 2018.

[17] R. Basko, G. L. Zeng, and G. T. Gullberg. Application of spherical harmonics to image reconstruction for the Compton camera. Phys. Med. Biol., 43:887–894, 1998.

[18] L.C. Parra. Reconstruction of cone-beam projections from Compton scattered data. IEEE Transactions on Nuclear Science, 47(4):1543–1550, 2000.

[19] T. Tomitani and M. Hirasawa. Analytical image reconstruction of cone-beam projections from limited-angle Compton camera data. IEEE Trans. on Nuclear Science, 50(5):1602–08, October 2003.

[20] C-Y Jung and S. Moon. Exact inversion of the cone transform arising in an application of a Compton camera consisting of line detectors. SIAM Journal on Imaging Sciences, 9(2):520–536, 2016.

[21] M. J. Cree and P. J. Bones. Towards direct reconstruction from a gamma camera based on Compton scattering. IEEE Transactions on Medical Imaging, 13(2):398–407, 1994.

[22] M. Haltmeier. Exact reconstruction formulas for a Radon transform over cones. Inverse Problems, 30(3):035001, 2014.

[23] D. Schiefeneder and M. Haltmeier. The Radon transform over cones with vertices on the sphere and orthogonal axes. SIAM Journal on Applied Mathematics, 77(4):1335–1351, 2017.

[24] S. Moon. Inversion of the conical radon transform with vertices on a surface of revolution arising in an application of a Compton camera. Inverse Problems, 33(6):065002, 2017.

[25] S. Moon and M. Haltmeier. Analytic inversion of a conical Radon transform arising in application of Compton cameras on the cylinder. SIAM Journal on Imaging Sciences, 10(2):535–557, 2017.

[26] A. Kishimoto, J. Kataoka, T. Taya, L. Tagawa, S. Mochizuki, S. Ohsuka, Y. Nagao, K. Kurita, M. Yamaguchi, N. Kawachi, K. Matsunaga, H. Ikeda, E. Shimosegawa, and J. Hatazawa. First demonstration of multi-color 3-D in vivo imaging using ultra-compact Compton camera. Scientific Reports, 7, 2017.

[27] R. Gouia-Zarrad and G. Ambartsoumian. Exact inversion of the conical Radon transform with a fixed opening angle. Inverse Problems, 30(4):045007, 2014.

[28] F. Terzioglu. Some inversion formulas for the cone transform. Inverse Problems, 31(11):115010, 2015.

[29] Victor Palamodov. Reconstruction from cone integral transforms. Inverse Problems, 33(10):104001, 2017.

[30] B. Smith. Reconstruction methods and completeness conditions for two Compton data models. J. Opt. Soc. Am. A, 22:445–459, March 2005.

[31] P. Kuchment and F. Terzioglu. Three-dimensional image reconstruction from Compton camera data. SIAM Journal on Imaging Sciences, 9(4):1708—-1725, 2016.

[32] P. Kuchment and F. Terzioglu. Inversion of weighted divergent beam and cone transforms. Inverse Problems & Imaging, 11(6):1071–1090, 2017.

[33] Fatma Terzioglu, Peter Kuchment, and Leonid Kunyansky. Compton camera imaging and the cone transform: a brief overview. Inverse Problems, 34(5):054002, 2018.

[34] M.K. Nguyen and T.T. Truong. On an integral transform and its inverse in nuclear imaging. Inverse Problems, 18(1):265, 2002.

[35] V. Maxim, M. Frandes, and R. Prost. Analytical inversion of the Compton transform using the full set of available projections. Inverse Problems, 25(9):1–21, 2009.

[36] A. Sawatzky, C. Brune, F. Wubbeling, T. Kosters, K. Schafers, and M. Burger. Accurate EM-TV algorithm in PET with low SNR. In Nuclear Science Symposium Conference Record, 2008. NSS'08. IEEE, pages 5133–5137. IEEE, 2008.

[37] E.Y. Sidky, C.M. Kao, and X. Pan. Accurate image reconstruction from few-views and limited-

angle data in divergent-beam CT. Journal of X-ray Science and Technology, 14(2):119–139, 2006.

[38] M Persson, D Bone, and H Elmqvist. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. Physics in Medicine & Biology, 46(3):853, 2001.

[39] A. Chambolle. An algorithm for total variation minimization and applications. Journal of Mathematical imaging and vision, 20(1-2):89–97, 2004.

[40] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences, 2(1):183–202, 2009.

[41] O. Klein and Y. Nishina. Über die streuung von strahlung durch freie elektronen nach der neuen relativistischen quantendynamik von dirac. Zeitschrift für Physik, 52(11-12):853–868, 1929.

[42] A. Zoglauer, R. Andritschke, and F. Schopper. MEGAlib–The medium energy gamma-ray astronomy library. New Astron. Rev., 50:629–632, October 2006.

[43] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, and al. Geant4-A simulation toolkit. Nucl. Instrum. Methods A, 506:205–303, 2003.

[44] S.J. Wilderman, W.L. Rogers, G.F. Knoll, and J.C. Engdahl. Fast algorithm for list mode back-projection of Compton scatter camera images. IEEE Transactions on Nuclear Science, 45:957–962, June 1998.

[45] V. Maxim, X. Lojacono, E. Hilaire, J. Krimmer, E. Testa, D. Dauvergne, I. Magnin, and R. Prost. Probabilistic models and numerical calculation of system matrix and sensitivity in list-mode MLEM 3D reconstruction of Compton camera images. Phys. Med. Biol., 61(1):243–264, 2016.

[46] F. Natterer. The mathematics of computerized tomography, volume 32. Siam, 1986.

IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology

Phys. Med. Biol. **61** (2016) 243–264 doi:10.1088/0031-9155/61/1/243

Probabilistic models and numerical calculation of system matrix and sensitivity in list-mode MLEM 3D reconstruction of Compton camera images

Voichita Maxim[1], Xavier Lojacono[1], Estelle Hilaire[1], Jochen Krimmer[2], Etienne Testa[2], Denis Dauvergne[2], Isabelle Magnin[1] and Rémy Prost[1]

[1] Université de Lyon, CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon, Université Lyon 1, France
[2] Université de Lyon, CNRS/IN2P3, UMR 5822, IPNL, F-69622 Villeurbanne, France

E-mail: voichita.maxim@creatis.insa-lyon.fr

Received 4 June 2015, revised 28 September 2015
Accepted for publication 10 November 2015
Published 4 December 2015

Abstract

This paper addresses the problem of evaluating the system matrix and the sensitivity for iterative reconstruction in Compton camera imaging. Proposed models and numerical calculation strategies are compared through the influence they have on the three-dimensional reconstructed images. The study attempts to address four questions. First, it proposes an analytic model for the system matrix. Second, it suggests a method for its numerical validation with Monte Carlo simulated data. Third, it compares analytical models of the sensitivity factors with Monte Carlo simulated values. Finally, it shows how the system matrix and the sensitivity calculation strategies influence the quality of the reconstructed images.

Keywords: tomographic reconstruction, Compton camera imaging, SPECT, MLEM, Monte Carlo simulation

(Some figures may appear in colour only in the online journal)

1. Introduction

The Compton camera, which is generally employed in astronomy for two-dimensional imaging of distant γ sources, may also be of interest for SPECT imaging in nuclear medicine and for the monitoring of proton or hadron therapy (e.g. Frandes *et al* 2010, Peterson *et al* 2010, Richard *et al* 2012, Hueso-González *et al* 2014). Being collimator free, an ideal camera could

 © 2016 Institute of Physics and Engineering in Medicine Printed in the UK

acquire all incident photons regardless of their direction and allow the source image to be reconstructed in three dimensions with a stationary and relatively small detector.

The Compton camera detects two or more interactions of a γ photon and measures the positions and transferred energies. The first interactions are Compton scatterings of the photon on electrons from the atoms of the detector and the last interaction is thought to be an absorption through a photo-electric effect. The most important acquisition parameters are the positions of the first and second hits, the energy deposited during the first hit, and also the total energy deposited on the detector when the initial energy of the particle is unknown. The direction of the incident photon is then confined to the surface of a cone with its vertex placed at the first interaction point, the axis in the direction of the scattered photon, and the opening angle calculated from the Compton scattering formula that in absence of Doppler broadening is:

$$\cos\beta = 1 - \frac{m_e c^2 E_1}{E_0(E_0 - E_1)}, \tag{1}$$

where $m_e c^2$ is the mass energy of an electron, E_0 is the total energy of the incident photon and E_1 is the energy transferred to the electron in the first Compton scattering. The energy E_1 is the actual value measurable in the detector.

The image of the distribution of the γ-photons source is then obtained by tomographic reconstruction techniques from a large set of overlapping Compton cones.

Originally adapted for the Compton camera by Wilderman *et al* (1998a), the list-mode maximum likelihood expectation maximization (LM-MLEM) algorithm is currently the most widely used for the reconstruction of Compton camera images. The algorithm allows us to approximate the image λ of the source by sharpening iteratively the coarse approximation, which is the intersection set of the Compton cones related to the measured events. At each step a vector $\widehat{\lambda}^{(\ell)}$, containing the intensities $\widehat{\lambda}_j^{(\ell)}$ for all of the M voxels from the image, is evaluated. The initialization step consists in calculating $\widehat{\lambda}^{(0)}$, generally as the back-projection in the image volume of all the N events. The sequence $\widehat{\lambda}^{(\ell)}$ is then evaluated iteratively through the equation

$$\widehat{\lambda}_j^{(\ell+1)} = \frac{\widehat{\lambda}_j^{(\ell)}}{s_j} \sum_{i=1}^{N} t_{ij} \frac{1}{p_i^{(\ell)}}, \quad \text{with } p_i^{(\ell)} = \sum_{k=1}^{M} t_{ik} \widehat{\lambda}_k^{(\ell)}, \tag{2}$$

where t_{ij} are elements of the system matrix T, indexed on the events i and on the voxels j. The value t_{ij} corresponds to the probability of a photon emitted by the jth voxel to be detected with the parameters of the ith event. The sensitivity correction factor $s_j = \sum_i t_{ij} > \sum_{i=1}^{N} t_{ij}$ is a summation over all possible measurable cones of response, not only those that are measured, and corresponds to the probability of a photon emitted by the jth voxel to be detected by the camera. Derived from the MLEM algorithm developped by Dempster *et al* (1977), the list-mode (LM) version initially proposed for tomography in Lange and Carson (1984) and Shepp and Vardi (1982) is a limit case where the extremely fine sampling of the space of acquisition parameters (e.g. positions, energies, time) makes that at most one event is measured for each parameter. As a consequence, only terms corresponding to measured events appear in (2). As stated in previous studies, (Dempster *et al* 1977, Vardi *et al* 1985), under some mild hypotheses, although not always verified in practice, the sequence $\widehat{\lambda}_j^{(\ell)}$ of MLEM estimators converges to the maximum likelihood estimator of λ. Note that λ is the parameter (mean and variance) of the Poisson distribution of the source.

In theory, both the system matrix and the sensitivity can be calculated by Monte Carlo simulations. Patient attenuation, camera acquisition parameters and other random uncertainties and noise can then be included in the calculation. The weakness of the Monte Carlo approach stems from the dimension of the space where the Compton camera data lie. The device has to acquire for each incident γ photon at least two positions of interactions and the scattering angle. Composed of at least two detector layers, the minimum dimension of the data space is then five, when the depth of interaction within the layer is unavailable. The Monte Carlo approach is thus impractical for calculation of the system matrix but can still be employed for evaluation of the sensitivity matrix.

Deterministic evaluation of the system matrix is a problem more generally related to the formulation of the direct problem in Compton imaging, and constitutes a link between iterative and analytic reconstruction methods. There is a broad choice of models that have been proposed in the literature, most of them being associated with direct inversion by analytic methods. In this context, they are additionally constrained by the difficulty to find an integral operator solving the inverse problem. In Cree and Bones (1994), Basko *et al* (1998) and Smith (2005), the projections are modeled as integrals of the intensity distribution of the source on the surface of Compton cones. Other authors define the projections as sums of the cone-beam projections on the generatrices of the cone (e.g. Parra 2000, Tomitani and Hirasawa 2002, Smith 2005). The disadvantage of this approach lies in the over-estimation of the contribution of points located close to the apex of the cone. Both models ignore the incidence angle of the photons on the camera and the related variability of the efficiency of the detector, introduced in Maxim *et al* (2009) and Maxim (2014). The decrease with the distance in the solid angle covered by detector element at the source may produce parallax artifacts when extended sources are imaged in 3D. A method correcting for the distance from the source to the detector and also for the uncertainties in the Compton angle measurement was proposed by Hirasawa and Tomitani (2003).

The description of direct projections is less extensive in the literature on iterative methods. One of the advantages of these methods is that they can handle complex models and account for various acquisition parameters, for noise and for attenuation sources. A relatively extensive probabilistic formulation for 2D image reconstruction was given in Wilderman *et al* (1998a). The sensitivity was taken to be constant for all the voxels in the image, perfect energy and spatial resolutions in the detectors were assumed, and Doppler effect was ignored. These limitations were removed in a later work (Wilderman *et al* 2001), although the model mainly remains 2D driven.

The numerical calculation of the system matrix is also challenging. Each of its elements corresponds to the integral of a function on a piece of cone surface or cone shell. A 2D method of calculation was proposed by Wilderman *et al* (1998b), where the pixels of a slice that are intersected by the conical section are identified by a fast marching algorithm. The algorithm may be extended to 2D reconstruction by repeating the calculation for all the slices of the volume. Later referred to as the ellipse-stacking method, the algorithm was compared with a ray-tracing method by Kim *et al* (2007). According to the aforementioned study, the performances of the method in terms of quality of the reconstructed images are equivalent but the ray-tracing method seems faster. A method based on the sampling of the cone surface with non-uniformly distributed rays was employed by Lojacono *et al* (2013). Uncertainties in the calculation of the Compton angle were included in the tool Mimrec (Zoglauer *et al* 2011), part of the MEGAlib software (Zoglauer *et al* 2006) that we used in our study for simulations and data analysis.

Reducing the calculation time and the memory burden is another important challenge in Compton imaging. Approximated calculation of the system matrix on GPU was proposed by Nguyen *et al* (2011). Arguing that the ray-tracing method is not adapted to fast implementation

on GPU, the authors propose to approximate the cord length resulting from the intersection of a ray with a voxel by the orthogonal distance from the center of the voxel to the ray. In a second method, they propose to replace each voxel by a point. An approach based on matrix decomposition and adapted to a particular Compton camera prototype was proposed by Sauve *et al* (1999). The calculation of the matrix exploits symmetries that follow from a specific sampling pattern in the object domain.

Some alternatives to the MLEM algorithm were suggested in the literature. In particular, the stochastic ensemble origin (SOE) algorithm (Mackin *et al* 2012) was shown to give very good results for perfect spatial, energy, and time resolution Compton cameras. The projection model is recalculated at each iteration as a function of both the data and the activity distribution at the previous iteration. The advantage of the MLEM approach is to rely on a rigorous mathematical formalism. Moreover, when the data statistics is too low to avoid convergence, dimensionality reduction techniques may be applied. E.g. Lee *et al* (2008) considered the maximum a posteriori (MAP) extension together with the row-action maximum likelihood algorithm (RAMLA) and the block sequential regularized maximum likelihood (BSREM) algorithms.

The first issue we address in this paper is the fully 3D formulation of the transition probabilities defining the system matrix. The model we propose accounts for the limited resolutions of the detectors, for the size of the voxels, and for the inclination of both the axis and the generatrices of the cone with respect to the detector surface.

The second issue is the validity of the model and how it compares to Monte Carlo simulated data. Direct comparison seems inappropriate with regards to the large dimension of the data space. Instead, we test the model on an aggregated set resulting from the data.

Several analytic methods for the calculation of the sensitivity factors were previously described in the literature. Among them we can mention those by Wilderman *et al* (2001), for a straight-forward and easy-to-implement approach, and Sullivan *et al* (2009), for a more complete formulation. In this work we compare the values obtained by Monte Carlo simulation with the values obtained by deterministic calculation. The calculation methods, their influence on the reconstructed images, and the relation of the sensitivity to the spectrum of the γ source constitutes the third issue of the paper.

The probabilistic model we propose for the calculation of the system matrix accounts for the solid angle covered by an elementary surface containing the first scattering point, at a point from the volume to be reconstructed. As a consequence, among the points that can contribute to a given event, the points located close to the detector are favored. On the contrary, the sensitivity factor, which decreases with the distance to the detector, tends to favor at each iteration the furthest points. Consequently, using realistic values for the sensitivity factors is mandatory. The joint influence of the transition probabilities formulation and of the sensitivity factors on the reconstructed images is the fourth issue of the current study.

In this work we concentrate on mono-energetic sources, but the initial energy of the photons was generally not taken into account in the reconstruction algorithm. For poly-energetic sources, methods that specifically model energy leakage from high to low energies are necessary. Indeed, incomplete absorptions from the higher-energy sources contaminate the acquisition of the lower-energy sources. Some fraction of the intensity distribution of the high-energy sources is thus transferred toward low-energy sources as a background. In addition, the sensitivity is energy-dependent. A probabilistic model for the calculation of the system matrix in four dimensions aiming to integrate the spectral deconvolution step in the image reconstruction was proposed by Xu and He (2007) and demonstrated on a system composed of four mono-energetic point sources. In a study by Gillam *et al* (2011), the spectral deconvolution

algorithm was demonstrated for the large spectrum of γ photons emitted during a proton therapy treatment.

The remainder of this paper is organized as follows. Section 2 gives a probabilistic description of the processes involved in Compton detection, leading to an analytic model of the system matrix elements. Simplified analytic descriptions of the sensitivity are also presented in this paper. Section 3 describes the configuration of the simulations and the methodology for the evaluation of the results from section 2. Section 4 shows the results concerning the evaluation of the models, a set of reconstructed images, and the influence that the choice of the parameters has on the reconstructed images. Finally, section 5 presents a discussion and some conclusions.

2. Deterministic models

2.1. Compton camera acquisition

Let $\mathfrak{p}$ be a γ photon emitted at a point M. A physical event is produced when $\mathfrak{p}$ activates at least two detectors. An additional constraint is given, stipulating that the last interaction should be in the absorber, in order to strengthen the probability of total absorption of the photon in the camera. A schematic representation of the camera acquisition principle is given in figure 1.

The two first interactions, as measured by the camera, are denoted V_1 and V_2. They correspond to real interactions at points $\widetilde{V}_1$ and $\widetilde{V}_2$, respectively. The measured loss of energy in the first hit, E_1 corresponds to the energy $\widetilde{E}_1$ transferred to the electron in the first scattering process. The sum of energies deposited by the scattered photon in the camera, denoted $\widetilde{E}_2$, produces a measurement E_2. Ideally, $\widetilde{E}_1 + \widetilde{E}_2$ should be equal to E_0, the initial energy of the detected photon. The unobservable Compton scattering angle $\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}$, theoretically given as

$$\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M} = \arccos\left(\frac{\overrightarrow{\widetilde{V}_2\widetilde{V}_1}\cdot\overrightarrow{\widetilde{V}_1M}}{\|\overrightarrow{\widetilde{V}_2\widetilde{V}_1}\|_2\,\|\overrightarrow{\widetilde{V}_1M}\|_2}\right), \tag{3}$$

may be then estimated from (1) and the measured energies by a value β. When the initial energy of the photon is known (mono-energetic source), it is preferable to replace E_2 by the difference $E_0 - E_1$ in order to avoid the influence of incomplete absorptions. Let $y = (V_1, V_2, \beta)$ be the vector of measurements, realization of the random variable $\mathscr{Y}$ containing the measured positions of interaction and the Compton angle as derived from (1). The real physical event that has produced the measure y is $\tilde{y} = (\widetilde{V}_1, \widetilde{V}_2, \widetilde{\beta})$, realization of the random variable $\widetilde{\mathscr{Y}}$ (see figure 1).

2.2. Probabilistic model for the system matrix elements

In the following, italic capital letters designate random variables. They mainly serve to distinguish between the probability distribution functions defined in the section.

Let x be the vector of coordinates of a point M from the source. The probability of occurrence of the measured event y given that $\mathfrak{p}$ was emitted at x is then

$$p_{\mathscr{Y}|\mathscr{X}}(y|x,E_0) = \int p_{\mathscr{Y}|\widetilde{\mathscr{Y}}}(y|\,\tilde{y},E_0)\,p_{\widetilde{\mathscr{Y}}|\mathscr{X}}(\tilde{y}\,|x,E_0)\mathrm{d}\tilde{y}. \tag{4}$$

Since, e.g. the measured value of V_1 does not depend on $\widetilde{V}_2$ or on $\widetilde{\beta}$ but only on $\widetilde{V}_1$, we have

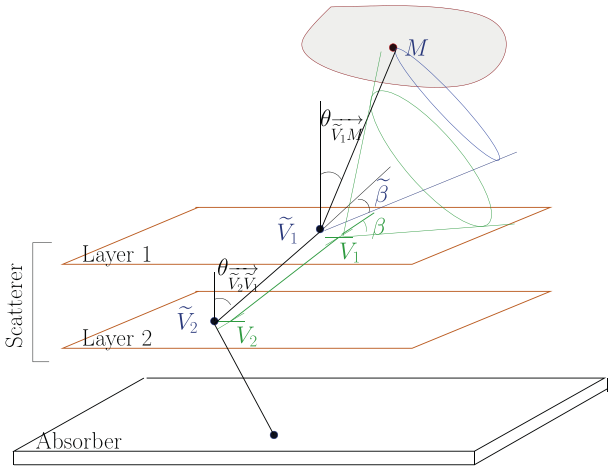


Figure 1. Compton camera acquisition principle. In this example, the camera is composed of two scatterers and one absorber layers. The two first positions of interaction of the photon emitted at M are $\widetilde{V}_1$ and $\widetilde{V}_2$. The real scattering angle is $\widetilde{\beta}$. The parameters measured by the camera are the locations V_1 and V_2 and the energies, from which the Compton angle β is calculated. The point M where the photon is emitted lies on the cone having the apex in $\widetilde{V}_1$, the axis collinear with the scattered ray direction and the half-opening angle $\widetilde{\beta}$.

$$p_{\mathscr{Y}|\widetilde{\mathscr{Y}}}(y|\tilde{y},E_0) = p_{\mathscr{Y}_1|\widetilde{\mathscr{Y}}_1}(V_1|\widetilde{V}_1,E_0)p_{\mathscr{Y}_2|\widetilde{\mathscr{Y}}_2}(V_2|\widetilde{V}_2,E_0)p_{\beta|\tilde{\beta}}(\beta|\widetilde{\beta},E_0). \tag{5}$$

The terms $p_{\mathscr{Y}_1|\widetilde{\mathscr{Y}}_1}(V_1|\widetilde{V}_1,E_0)$ and $p_{\mathscr{Y}_2|\widetilde{\mathscr{Y}}_2}(V_2|\widetilde{V}_2,E_0)$ depend on the spatial resolutions of the detectors. For spatially sensitive detector elements, they may be taken as uniform or Gaussian distributions. E.g. the Gaussian distribution was chosen by Xu and He (2007). The term $p_{\beta|\tilde{\beta}}(\beta|\widetilde{\beta},E_0)$ depends on the energy resolution of the detectors and on the probability $p_{\mathscr{E}_1,\mathscr{E}_2|E_0}(E_1,E_2|E_0)$.

The probability $p_{\widetilde{\mathscr{Y}}|\mathscr{X}}(\tilde{y}|x,E_0)$ is the product of several probabilities: the probability of Compton interactions in the scatterer, the probability of the Compton scattering process (including Doppler-broadening), absorption probabilities (in the different environments such as the object, air, absorption of the scattered photon in the scatter detector), and the absorption probability of the photon in the absorber. It also depends on the solid angle subtended at the origin of the particle M by the detector element containing the first hit, which should be considered in 3D since the emission is isotropic, and on the solid angle subtended at the position of the first hit by the detector element where the second hit takes place, which should be considered in 2D since it is related to a cone surface uncertainty through the scattering angle $\widetilde{\beta}$. Let $\theta_{\overrightarrow{\widetilde{V}_1M}}$ and $\theta_{\overrightarrow{\widetilde{V}_2\widetilde{V}_1}}$ be the angles of incidence of $\mathfrak{p}$ on the first and second activated detectors, respectively (see figure 1).

The absorption probabilities have been discussed previously (e.g. Wilderman *et al* 1998a, Parra 2000). As they depend on the path followed by the photon in the camera, they are difficult to calculate and we chose to neglect them hereafter. We argue that, for M on the surface of the cone defined by the parameters $\tilde{y}$, the density $p_{\widetilde{\mathscr{Y}}|\mathscr{X}}(\tilde{y}|x,E_0)$ may then be considered as proportional to

$$K(\widetilde{\beta}\,|E_0)\frac{\left|\cos(\theta_{\overrightarrow{\widetilde{V}_1M}})\right|}{\|\overrightarrow{\widetilde{V}_1M}\|_2^2}\frac{\left|\cos(\theta_{\overrightarrow{\widetilde{V}_2\widetilde{V}_1}})\right|}{\|\overrightarrow{\widetilde{V}_1\widetilde{V}_2}\|_2}, \tag{6}$$

where $K(\widetilde{\beta}\,|E_0)$ is the Compton scattering cross-section at the energy E_0. Since the proportionality factors get cancelled in the MLEM algorithm, in the following we put:

$$p_{\widetilde{\mathscr{Y}}\,|\,\mathscr{X}}(\tilde{y}\,|x,E_0)\propto K(\widetilde{\beta}\,|E_0)\frac{\left|\cos(\theta_{\overrightarrow{\widetilde{V}_1M}})\right|}{\|\overrightarrow{\widetilde{V}_1M}\|_2^2}\frac{\left|\cos(\theta_{\overrightarrow{\widetilde{V}_2\widetilde{V}_1}})\right|}{\|\overrightarrow{\widetilde{V}_1\widetilde{V}_2}\|_2}\delta(\widetilde{\beta}-\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}). \tag{7}$$

An expression for the probability $p_{\mathscr{Y}\,|\,\mathscr{X}}(y|x,E_0)$ may now be derived by incorporating (5) and (7) into (4):

$$\begin{aligned} p_{\mathscr{Y}\,|\,\mathscr{X}}(y|x,E_0)=&\int_{\mathbf{R}^3}\frac{\left|\cos(\theta_{\overrightarrow{\widetilde{V}_1M}})\right|}{\|\overrightarrow{\widetilde{V}_1M}\|_2^2}p_{\mathscr{Y}_1\,|\,\widetilde{\mathscr{Y}}_1}(V_1|\widetilde{V}_1,E_0)\\ &\int_{\mathbf{R}^3}\frac{\left|\cos(\theta_{\overrightarrow{\widetilde{V}_2\widetilde{V}_1}})\right|}{\|\overrightarrow{\widetilde{V}_1\widetilde{V}_2}\|_2}p_{\mathscr{Y}_2\,|\,\widetilde{\mathscr{Y}}_2}(V_2|\widetilde{V}_2,E_0)\\ &\int_0^{\pi}K(\widetilde{\beta}\,|E_0)p_{\beta\,|\,\widetilde{\beta}}(\beta|\widetilde{\beta}\,,E_0)\delta(\widetilde{\beta}-\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M})\mathrm{d}\widetilde{\beta}\,\mathrm{d}\widetilde{V}_2\mathrm{d}\widetilde{V}_1. \end{aligned} \tag{8}$$

After integration for $\widetilde{\beta}$ we finally get:

$$\begin{aligned} p_{\mathscr{Y}\,|\,\mathscr{X}}(y|x,E_0)=&\int_{\mathbf{R}^3}\frac{\left|\cos(\theta_{\overrightarrow{\widetilde{V}_1M}})\right|}{\|\overrightarrow{\widetilde{V}_1M}\|_2^2}p_{\mathscr{Y}_1\,|\,\widetilde{\mathscr{Y}}_1}(V_1|\widetilde{V}_1,E_0)\\ &\int_{\mathbf{R}^3}\frac{\left|\cos(\theta_{\overrightarrow{\widetilde{V}_2\widetilde{V}_1}})\right|}{\|\overrightarrow{\widetilde{V}_1\widetilde{V}_2}\|_2}p_{\mathscr{Y}_2\,|\,\widetilde{\mathscr{Y}}_2}(V_2|\widetilde{V}_2,E_0)\\ &K(\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}|E_0)p_{\beta\,|\,\widetilde{\beta}}(\beta|\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M},E_0)\mathrm{d}\widetilde{V}_2\mathrm{d}\widetilde{V}_1. \end{aligned} \tag{9}$$

With the spatial resolution of semi-conductor stripped detectors, the uncertainty on the direction of the Compton cone axis may be considered as negligible compared to the uncertainty of the measured Compton scattering angle. In our experiments its modeling did not improved the reconstructed images. Consequently, the real and measured points where $\mathfrak{p}$ hits the detector may be merged, leading to the simplified formulation:

$$\begin{aligned} p_{\mathscr{Y}\,|\,\mathscr{X}}(y|x,E_0)=&\frac{\left|\cos(\theta_{\overrightarrow{V_1M}})\right|}{\|\overrightarrow{V_1M}\|_2^2}\frac{\left|\cos(\theta_{\overrightarrow{V_2V_1}})\right|}{\|\overrightarrow{V_1V_2}\|_2}\\ &K(\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}|E_0)p_{\beta\,|\,\widetilde{\beta}}(\beta|\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M},E_0). \end{aligned} \tag{10}$$

For a completely absorbed photon in a detector with ideal energy resolution it may be considered that

$$p_{\beta|\widetilde{\beta}}(\beta|\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M},E_0)=\delta(\beta-\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}). \tag{11}$$

For a finite resolution detector, a convenient choice may be the Gaussian distribution with mean $\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}$ and standard deviation $\sigma_{\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}}$,

$$p_{\beta|\widetilde{\beta}}(\beta|\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M},E_0)=\frac{1}{\sigma_{\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}}\sqrt{2\pi}}\exp\left(-\frac{(\beta-\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M})^2}{2\sigma^2_{\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}}}\right). \tag{12}$$

The unknown mean $\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}$ may be replaced by the measured value β. Different estimations of the standard deviation $\sigma_{\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}}$ were proposed by Ordonez *et al* (1997), depending on the type of acquisition system. In our case, the standard deviation may be calculated from the known detector uncertainties $\sigma_{\widetilde{E}_1}$ and $\sigma_{\widetilde{E}_2}$ as:

$$\sigma_{\widetilde{\beta}}=\frac{m_ec^2}{E_0^2\sin(\widetilde{\beta})}\sqrt{\sigma^2_{\widetilde{E}_1}+\frac{\widetilde{E}_1^2(\widetilde{E}_2+E_0)^2}{\widetilde{E}_2^4}\sigma^2_{\widetilde{E}_2}}. \tag{13}$$

The unknown real values are to be replaced in (13) by the measured values. The Gaussian distribution was previously used (e.g. Hirasawa and Tomitani 2003, Ortega *et al* 2015). A mixture of Gaussians that better accounts for large tails induced by Doppler broadening was proposed by Wilderman *et al* (2001) and a Fisher Von-Mises model adapted to the positive values of the Compton angle was suggested by Sauve *et al* (1999).

The element t_{ij} of the system matrix T is the probability of observing the physical event y_i when a photon is emitted by the voxel v_j. Assuming that a voxel is sufficiently small to allow a constant probability inside the voxel we get:

$$\begin{aligned} t_{ij}&=\int_{v_j}p_{\mathscr{Y}_i|\mathscr{X}}(y_i|x,E_0)p(x|M\in v_j)\mathrm{d}x\\ &=\frac{1}{\mathrm{vol}(v_j)}\int_{v_j}p_{\mathscr{Y}_i|\mathscr{X}}(y_i|x,E_0)\mathrm{d}x, \end{aligned} \tag{14}$$

where $\mathrm{vol}(v_j)$ denotes the volume of the voxel and does not depend on the index j when all the voxels have the same size. From the proposed formula (10) and (14), the elements of the matrix may be expressed as:

$$\begin{aligned} t_{ij}=&\frac{1}{\mathrm{vol}(v_j)}\frac{\left|\cos(\theta_{\overrightarrow{V_2V_1}})\right|}{\|\overrightarrow{V_1V_2}\|_2}\\ &\int_{v_j}\frac{\left|\cos(\theta_{\overrightarrow{V_1M}})\right|}{\|\overrightarrow{V_1M}\|_2^2}K(\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}|E_0)p_{\beta|\widetilde{\beta}}(\beta|\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M},E_0)\mathrm{d}x. \end{aligned} \tag{15}$$

The rest of this work is partly dedicated to the numerical validation of this proposed model.

It may be seen from (2) that the factor multiplying the integral in (15) gets canceled in the iterative reconstruction process and may thus be ignored.

Derived from (15), the model:

$$t^0_{ij}=\int_{v_j}K(\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}|E_0)p_{\beta|\widetilde{\beta}}(\beta|\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M},E_0)\mathrm{d}x, \tag{16}$$

is better adapted when the sensitivity is uniformly set to one. Its effect on the images is compared in section 4 with model (15) coupled to a non-constant sensitivity.

2.3. Deterministic calculation of the sensitivity

A Compton camera is often composed of a relatively small scatterer and a larger absorber. The absorber may be located at a distance from the scatterer that is smaller than the distance from the source to the scatterer. Also, it may happen that the first two hits take place in two layers of the scatterer. For an extended source, the solid angle covered by the scatterer at a point of the source is highly dependent on the position of the point. In contrast, since the scatterer's extent is generally smaller that the extent of the absorber, the variations of the solid angle of the absorber at the scatterer's points are small. If we also choose to neglect the variations of the Compton scattering differential cross-section and the size of a voxel, confounded with its center O_j, we may consider that s_j is approximately equal to the solid angle subtended by the scatterer at O_j. For a thin, single-layer planar scatterer composed of L detector elements (e.g. individual crystals or surfaces delimited by strips) having centers Q_ℓ, the solid angle subtended by the detector at O_j may be approximated by the sum of the solid angles covered by each elementary detector surface at O_j. The following approximation of the sensitivity may be thus driven, except for some constant factor:

$$s_j = \sum_{\ell=1}^{L} \frac{\left|\cos(\theta_{\overrightarrow{Q_\ell O_j}})\right|}{\|\overrightarrow{Q_\ell O_j}\|_2^2}. \tag{17}$$

Equation (17) is derived from a similar expression given by Wilderman *et al* (2001), where the probabilities of the different interactions in the detector were also included.

For a multi-layer scatterer system, an approximate value can be calculated as the sum of the sensitivities calculated in (17) for each individual layer. We will further refer to it as the multi-layer estimation of the sensitivity matrix.

For a small detector, the terms from the sum (17) are approximately the same. Another rough estimate of the sensitivity for multi-layer systems is to consider that when seen from the source's voxel v_j, the scatterer block is small. The sensitivity factor s_j may then be approximated by the solid angle covered at O_j by a surface element placed at the center Q of the block, through the formula:

$$s_j = \frac{\left|\cos(\theta_{\overrightarrow{QO_j}})\right|}{\|\overrightarrow{QO_j}\|_2^2}. \tag{18}$$

We will further refer to this method as the volume-based estimation of the sensitivity matrix.

3. Simulations and evaluation methodology

3.1. Two experimental setups

Two Compton cameras were modeled for the purpose of this work. The first one, described hereafter as set-up (A), is a non-realistic extended camera dedicated only to the verification of the probabilistic model from section 2.2. The second one, described hereafter as set-up (B), is inspired by a prototype under construction. This second camera was employed in all the imaging experiments presented in section 4.

(A) *Large camera for the validation of the system matrix model.* The dimensions were chosen to avoid as much as possible the artifacts due to the boundaries. The scatterer is composed of three silicon scatter detectors, each having dimensions of $32 \times 32 \times 0.5\,\text{cm}^3$. One layer

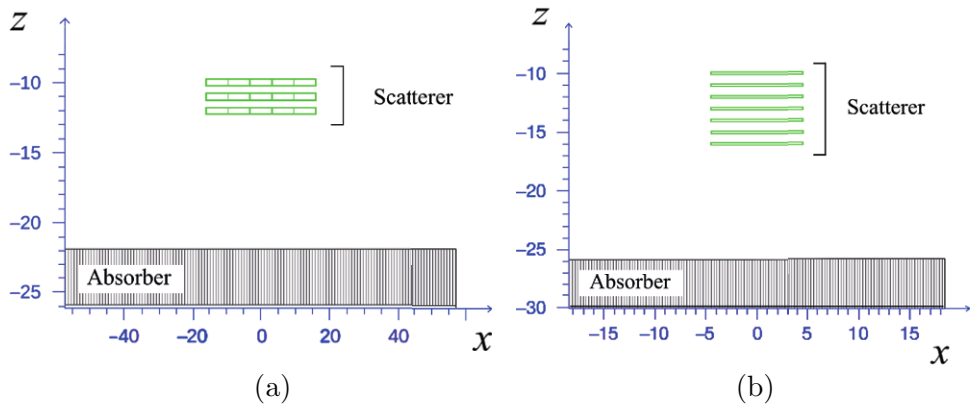


Figure 2. Two geometries of Compton camera. (a) Extended camera acquiring photons emitted by a point source with coordinates (0, 0,−5). This camera is used for the verification of the system matrix model. (b) Camera from the examples. The dimensions are in centimeters.

is composed of 5×5 wafers of $6.3 \times 6.3 \times 0.5\,\text{cm}^3$, each of them containing 128 strips on each side of the layer. The scatterers are centered at $(0, 0, -10)$, $(0, 0, -11)$ and $(0, 0, -12)$, respectively. The absorber detector centered at $(0, 0, -24)$ is composed of 200×200 LSO blocks, each of them measuring $0.5 \times 0.5 \times 4\,\text{cm}^3$, for a total size of $114 \times 114 \times 4.2\,\text{cm}^3$ (2 mm of non-sensitive material representing the electronics was added at the bottom of the absorber). Each individual LSO block is encapsulated in a millipore separator of 0.35 mm thick. The geometry of the experiment is depicted in figure 2(a).

(B) *Multi-layer Compton camera.* The camera is represented in figure 2(b) and has characteristics inspired from Krimmer *et al* (2015). The camera was used for the Monte Carlo calculation of the sensitivity matrices and for the illustrations given with different test sources. Its scatterer is composed of 7 silicon layers measuring $9 \times 9 \times 0.2$ cm^3 each with 1 cm separating their central planes. Each layer contains 2×128 strips. The LSO absorber is made of 64^2 crystal blocks, each of them measuring $0.5 \times 0.5 \times 4$ cm^3. The crystals are separated by 0.7 mm of millipore. The absorber has a total dimension of $37 \times 37 \times 4.2\,\text{cm}^3$.

For both cameras, the spatial resolution of the absorber is related to the size of the individual crystals, since they are not position sensitive. The *xy* coordinates of a hit in the scatterer are set as the coordinates of the two closest orthogonal strips. The *z* coordinates are taken as the center of the scatterer layer. Energy deposits that are closer than 2 mm for the scatterer and 5 mm for the absorber are binned in a single hit.

The energy resolution in the absorber is set as a Gaussian function of the energy deposited by the photon. An uniform standard deviation of 1 keV is set for the Gaussian energy resolution of the scatterers. A hit is recorded when the deposited energy exceeds 10 keV for the scatterer and 20 keV for the absorber.

3.2. Simulation and data analysis

The Monte Carlo simulations are realized with the tool Cosima (Zoglauer *et al* 2009) of the MEGAlib software (Zoglauer *et al* 2006), version 2.28.01. Based on Geant4 (Agostinelli *et al* 2003), version 09-06-patch-03, the Cosima tool has been validated on several Compton cameras, e.g. MEGA, NCT, COSI and HEMI.

During the simulation, the hits of each emitted photon are recorded separately from the hits of the other photons. The simulated data are then analyzed either with Sivan or with Revan, both tools being parts of MEGAlib. Sivan recovers ideal measurements and is employed only for the validation of the system matrix model. Revan identifies the Compton events and reconstructs the unknown paths followed by the γ photon in the camera. The detector effects engine ensures that the input used for Revan exactly resembles measured data, and that no prior knowledge from the simulation is retained. More details on the strategies allowing to select the Compton events from other types of events and on the path reconstruction methods are given elsewhere (Zoglauer 2005, Zoglauer *et al* 2006).

We consider an event in the detector with at least one hit in the scatterer and at least one hit in the absorber, irrespective of what really happened in the detector. For instance, the hit in the scatterer could also originate from a 511 keV photon originating from the annihilation of a pair-creation produced positron. When more than two hits are recorded for the same emitted photon, the most likely path of the photon in the camera is calculated. The first hit was hypothesized to always take place in the scatterer. The photon is not analyzed and no event is created when two or more hits take place in the same detector (scatterer wafer or absorber ctystal), since in real experiments these type of events cannot be resolved. No cut on the total energy deposited by the photon is made, except when this is explicitly stated in the text. As a consequence, the reconstructed Compton events are affected by spatial and energy uncertainties, by incomplete absorptions, and possibly by the wrong sequencing of the hits in the calculated path.

3.3. Monte Carlo estimation of the sensitivity matrix

A mono-energetic point source is placed in a $30 \times 30 \times 10$ cm^3 volume centered at (0, 0, 0). The source is then moved inside the volume with steps of 1 cm and let to emit 10^7 gamma photons at each position. The number of Compton events at each position of the source is recorded to produce the Monte Carlo estimation of the sensitivity vector. A linear interpolation is then applied to the 3D vector moving from the coarse grid of simulation to the finer grid necessary in the image reconstruction algorithm. The calculation is repeated for two energies, namely, 662 keV and 5 MeV.

3.4. Methodology for the evaluation of the transition probabilities

This section describes a method for the experimental validation of the results presented in section 2.2 in the case of an ideal acquisition system. The elements of the system matrix are then defined by the density $p_{\mathscr{Y}|\mathscr{X}}(\tilde{y}|x, E_0)$ given in (7). In order to limit the influence of the boundaries, the camera described as set-up (A) is simulated along with a close point source placed at coordinates (0, 0, −5).

The analysis of the data set gives the real sequence of interactions. The energies are exact up to the Doppler broadening and the equality $E_2 = E_0 - E_1$ is imposed. The coordinates of the hits are known exactly. For the sake of simplification, the z coordinate of each hit is replaced with the z coordinate of the center of the activated detector.

Owing to the dimension of the data space, equal to $3 \times 3 \times 1$ when the variables are V_1, V_2 and β, an extremely large simulation would be necessary for a statistically reliable estimation of the system matrix. We will instead verify a consequence of (7), related to a particular binning of the data described hereafter. In order to simplify the presentation, let us focus on γ first scattered in the layer centered at $z = -10$ cm (the one closer to the source) then interacted in the layer centered at $z = -11$ cm. The source is located at $M(0, 0, -5)$.

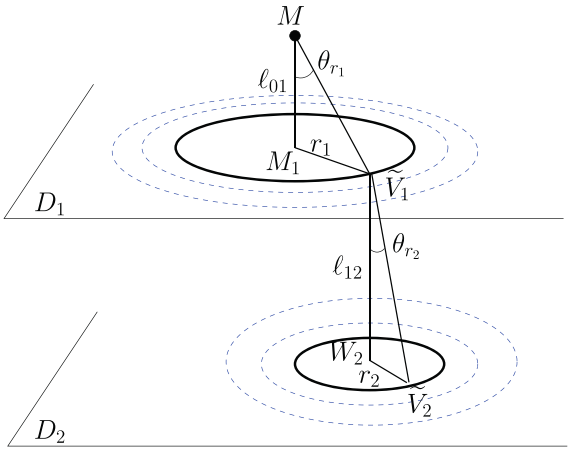


Figure 3. The data acquired from the point source M are binned on sets of concentric circles. The first set of concentric circles is contained in a first layer, the second set in the second layer or in the absorber. All detectors are reduced to their central plane and the vertical coordinates of the hits are ignored.

Let M_1 be the perpendicular projection of M on the central plane of the first scatterer (see figure 3). Let $r_1 \in (0, 16)$ be the radius in centimeters of a circle $C(M_1, r_1)$ centered at M_1. Let $\widetilde{V}_1$ be a point on this circle, representing a possible location for the first hit of a detected event. Let W_2 be the projection of $\widetilde{V}_1$ on the central plane of the second layer, the one placed at altitude $z = -11$ cm. For some $r_2 \in (0, 8)$, let $C(W_2, r_2)$ be the circle of radius r_2 cm, centered at W_2, and $\widetilde{V}_2$ a running point of the circle. Let $\widetilde{\beta}$ stand for some scattering angle in the range $(0, \pi)$. Note that for each given pair $(\widetilde{V}_1, \widetilde{V}_2)$, there is a single value $\widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M}$ such that the conical surface $\mathscr{C}(\widetilde{V}_1, \widetilde{V}_2, \widetilde{\beta}_{\widetilde{V}_1,\widetilde{V}_2,M})$ contains the point M.

Let us denote $\ell_{01} = \|\overrightarrow{MM_1}\|_2$ the distance from the source to the first hit layer and ℓ_{12} the distance between the two first hit detectors. In our example, $\ell_{01} = 5$ cm and $\ell_{12} = 1$ cm. Following equation (7), the probability of observing photons that successively interact first on the circle $C(M_1, r_1)$ and secondly on the circle $C(W_2, r_2)$, corrected for variations in the Compton scattering differential cross-section $K(\widetilde{\beta}\,|E_0)$,

$$P(r_1, r_2) = \frac{1}{r_1 r_2} \int_{C(M_1,r_1)} \int_{C(W_2,r_2)} \int_0^{\pi} \frac{1}{K(\widetilde{\beta}\,|E_0)} p_{\widetilde{\mathscr{Y}}\,|\,\mathscr{X}}(\tilde{y}\,|x, E_0) \mathrm{d}\widetilde{\beta}\, \mathrm{d}\widetilde{V}_2 \mathrm{d}\widetilde{V}_1 \tag{19}$$

satisfies the proportionality relation

$$P(r_1, r_2) \propto \varphi_1(r_1)\varphi_2(r_2), \tag{20}$$

with

$$\varphi_1(r_1) = \frac{\ell_{01}}{(r_1^2 + \ell_{01}^2)^{3/2}} \quad \text{and} \quad \varphi_2(r_2) = \frac{\ell_{12}}{r_2^2 + \ell_{12}^2}, \tag{21}$$

comming from the solid angles in (7). In practice, discrete values are given to the radii r_1 and r_2 and the circles are replaced by disjoint annuli. The probability $P(r_1, r_2)$ may be estimated from the data and the linear dependence with respect to the two solid angles may be verified with a linear least squares fit.

Table 1. Joint distributions of the first two hits for real (top table) and measured (bottom table) events.

		Second hit			
		D_1	D_2	D_3	*Absorber*
First hit	D_1	122	104	78	371
	D_2	29	109	90	350
	D_3	130	30	99	350

		Second hit			
		D_1	D_2	D_3	*Absorber*
First hit	D_1	0	81	57	313
	D_2	28	0	73	303
	D_3	14	28	0	308

Note: Top table: several interactions may take place in the same detector. Bottom table: only one interaction is allowed in each detector (scatterer layer or absorber crystal). The numbers represent thousands of observed events.

4. Results

In this section we first present results supporting the theoretical developments from section 3.4. Then we compare analytically calculated to Monte Carlo calculated sensitivity. We finally discuss the previous observations in light of three examples of reconstructed images, for a point-like source, a line-shaped source, and a 2D Shepp–Logan brain phantom.

4.1. Validity of the system matrix model

A point source emitting γ photons with energy 662 keV was simulated. From $4 \cdot 10^7$ emitted particles, about $2.7 \cdot 10^6$ Compton events were triggered by the camera. Table 1 shows the distribution of the two first interactions. The left-hand panel shows the counts (in thousands of events) for real coordinates, the right hand panel shows the counts (in thousands of events) for measured coordinates. The first two interactions may both be located in the same detector (left-hand table). We chose to discard these events when we work in realistic conditions, thus they do not appear in the right-hand table. The scatterer's layers are denoted D_1, D_2, D_3, with D_1 the one closest to the source.

The real events were then binned as described in section 3.4, by replacing the circles with annuli of 2 mm width. We chose a sampling of 2° for the Compton angle. As already stated in section 3.4, the dependency of the function $P(r_1, r_2)$ with respect to $\varphi_2(r_2)$ should be linear. For $r_1 = 9, 11, \ldots, 79$ mm, a line was fitted to the data $\{(\varphi_2(r_2), P(r_1, r_2)) \ : \ r_2 = 3, 5, \ldots, 79 \text{ mm}\}$. Figure 4 shows the slopes obtained in panel (a) and the coefficient of determination denoted R^2 measuring the goodness-of-fit in panel (b). The R^2 statistics indicates how successful the fit is in explaining the variation of the data and is defined as the ratio of the sum of squares of the regression (*SSR*) and the total sum of squares (*SST*). For response values denoted $(y_i)_{i=1,n}$ having the mean $\bar{y}$, and for the estimated response vales $(\hat{y}_i)_{i=1,n}$, the expressions of *SSR* and *SST* are:

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \quad \text{and} \quad SST = \sum_{i=1}^{n} (y_i - \bar{y})^2. \tag{22}$$

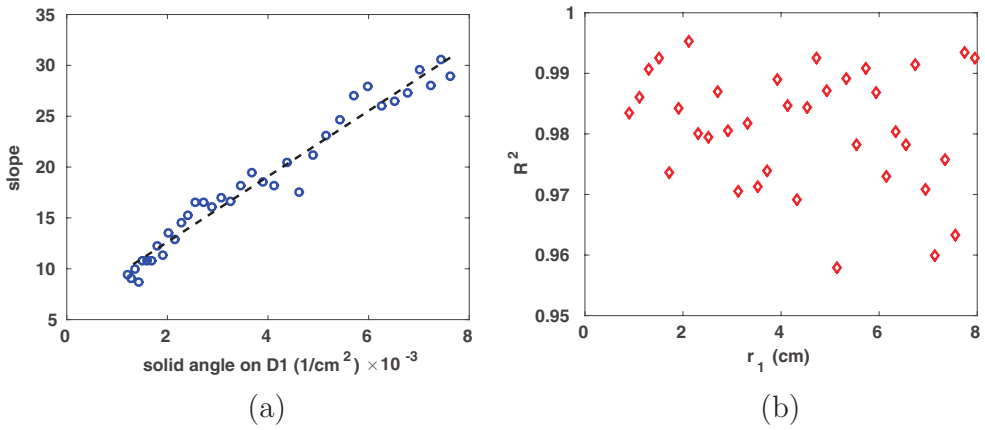


Figure 4. Mono-energetic point source emitting at 662 keV. (a) Slope from the linear fit of $P(r_1, r_2)$ *w.r.* to $\varphi_2(r_2)$ represented as function of $\varphi_1(r_1)$. (b) Quality of the fit expressed by the R^2 statistic.

The slopes are represented as a function of the solid angle $\varphi_1(r_1)$ and the R^2 values as a function of r_1.

The values of R^2, close to 1, suggest that the expected linear dependency on φ_2 is verified. As can be seen in figure 4(a), the linear dependency on φ_1 was also verified. Note that a few of the smallest values and the half largest values of r_i, $i = 1, 2$, were excluded from the study. Indeed, for values of r_i close to 0 the area of the annuli is relatively small and yelds poor statistics. For the largest values, the annuli are not completely included in the detectors. The plots from figure 4 were created for a first hit in D_1 and a second hit in D_2.

4.2. Sensitivity matrix

Slices from the sensitivity matrices calculated for a volume of $30 \times 30 \times 10\,\text{cm}^3$ and different models are shown in figure 5.

We denote S_{vol} and S_{layer} the volume-based estimation and the multi-layer estimation, respectively. The Monte Carlo sensitivities are indexed on the energy of the source. Figures 6(a) and (b) compares the values of $S_{662\,\text{keV}}$ with the values of S_{vol} and S_{layer} after division of all matrices by their respective mean. The observed element-by-element quotient over the volume is in both cases in the range (0.8, 1.8). Relatively to the mean value, the elements of the analytic matrices are under-estimated in the furthest regions of the lowest slices and they are over-estimated in the axis of the detector. Some variability may also be observed between neighboring voxels, due to statistical fluctuations in the calculation of $S_{662\,\text{keV}}$. Figure 6(c) compares the values of the two Monte Carlo calculated sensitivities, $S_{662\,\text{keV}}$ and $S_{5\,\text{MeV}}$. The observed element-by-element quotient over the volume is in the range (0.85, 1.13).

For the source emitting at 662 keV, we also evaluated the variability on 100 runs of the Monte Carlo estimated sensitivity for the central position, i.e. the point having coordinates (0, 0, 0), and for the point at (1, 0, 0) cm. The observed counts are reported in table 2.

Table 2 shows that for the chosen statistics, the difference between the mean sensitivities at two points 1 cm away from each other is of the order of the standard deviation. Therefore, our Monte Carlo sensitivity estimator is relatively poor and a more precise estimator may lead to improvement of the reconstructed images.

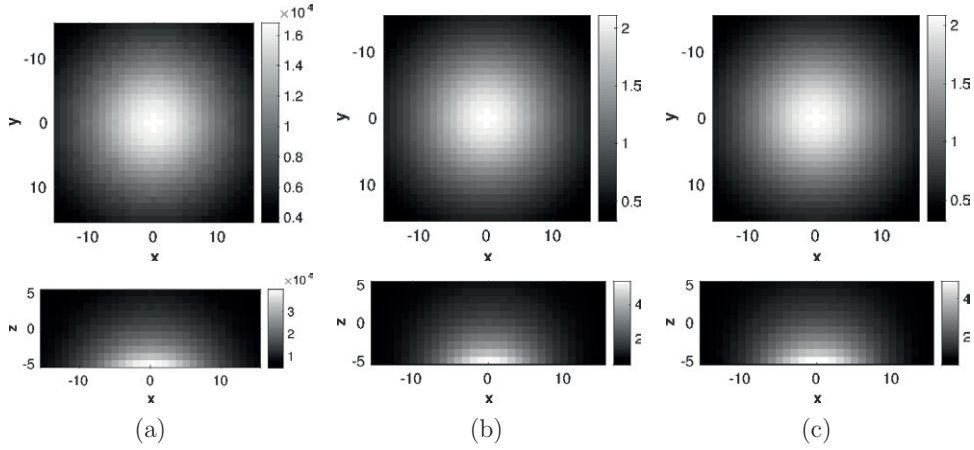


Figure 5. Variation of the sensitivity in the volume centered at (0, 0, 0) having dimensions $30 \times 30 \times 10$ cm^3. In the first row the slices at $z = 0$ cm and in the second row the slices at $y = 0$ are shown. (a) 662 keV (Monte Carlo), (b) analytic (volume), (c) analytic (multi-layer).

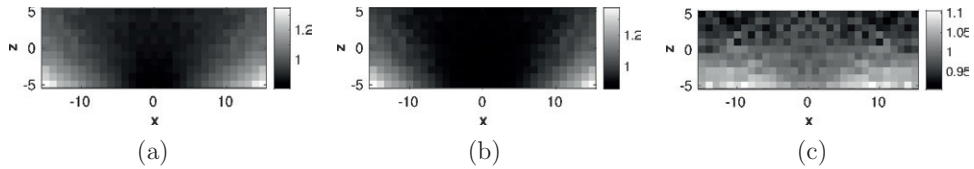


Figure 6. Comparison between sensitivities. (a) Monte Carlo calculated $S_{662\,\mathrm{keV}}$ divided by analytic S_{vol} sensitivities. (b) Monte Carlo calculated $S_{662\,\mathrm{keV}}$ divided by analytic S_{layer}. (c) Monte Carlo calculated $S_{662\,\mathrm{keV}}$ divided by Monte Carlo calculated $S_{5\,\mathrm{MeV}}$ sensitivities. Only one slice orthogonal to the detector in the plane $y = 0$ is shown. All the matrices were divided by their respective mean beforehand.

Table 2. Variability of the Monte Carlo calculation for two elements of the sensitivity matrix $S_{662\,\mathrm{keV}}$.

	Mean	Std	Minimum value	Maximum value
(0, 0, 0)	16 804	132	16 519	17 207
(1, 0, 0)	16 665	125	16 405	17 015

Note: All the values represent counts.

4.3. Influence of the two parameters on the reconstructed images

As can be seen from equation (2), the system matrix T and the sensitivity S have an opposite influence at each iteration of the MLEM algorithm. The factor t_{ij} represents the relative weight of the j-th voxel between the voxels that contribute to the i-th event. It aims to reinforce the intensity of voxels having a high relative weight. In particular, by the mean of the solid angle subtended by the detector at the center of the voxel, it heightens the voxel intensities close to the detector. In contrast the factor s_j reinforces the voxels that are located far from the detector. An alternative solution is from then to set the sensitivity to 1 for all the voxels and to withdraw the solid-angle factors from the calculation of the system matrix as

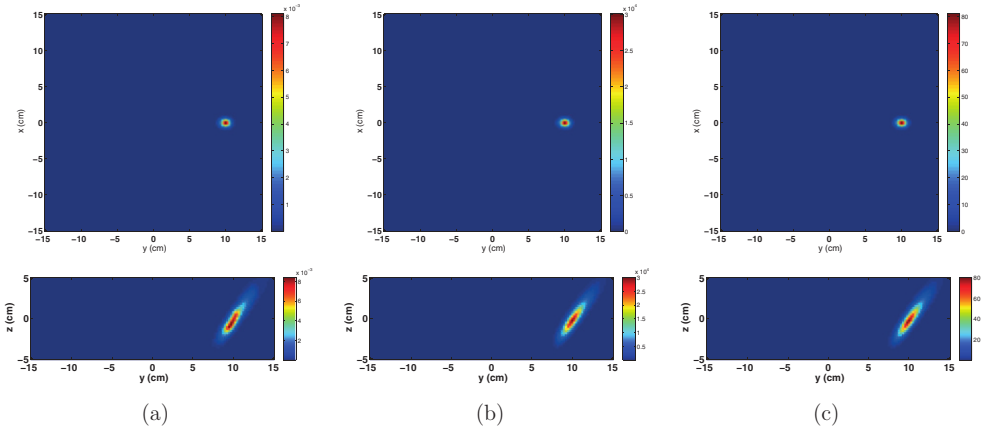


Figure 7. Point source placed in (0, 10, 0) cm emitting γ photons with energy of 662 keV, reconstructed with different models of the sensitivity matrix. (a) MC sensitivity, (b) analytic sensitivity, (c) sensitivity set to 1.

in equation (16). In the experiments to follow, this variant will be considered together with the proposed model.

The two analytic estimations of the sensitivity gave reconstructed images that cannot be differentiated from each other by the human eye. For this reason, only the images obtained with the volume-based sensitivity are shown.

In all the given examples, a volume with dimensions $30 \times 30 \times 10\,\text{cm}^3$ centered at (0, 0, 0) was reconstructed. The volume was divided into $120 \times 120 \times 40$ cubic voxels of 2.5 mm for each side.

The iterations were stopped after 20 steps for the tests with simple sources. For the Shepp–Logan phantom, 40 iterations were runned in order to extract the smallest features. Augmenting the number of iterations leads to mild improvement for point sources and to slow divergence for the other images reflected by the formation of small clusters in homogeneous regions.

4.3.1. Points. A point-like source was placed in (0, 10, 0) cm, i.e. 5.5 cm away from the edge of the scatterer in the y direction. In two distinct runs, the source was let to emit isotropically 4×10^7 γ particles with energies 662 keV and 5 MeV, respectively.

About 3.3×10^4 valid detections were obtained for the source emitting at 662 keV. Three images of the volume were reconstructed that differ by the sensitivity model. Slices containing the source are shown in figure 7, parallel to the detector in the top row and orthogonal to it in the bottom row.

About 2×10^4 valid events were acquired from the source emitting γ photons with an energy of 5 MeV. As with the previous source, three images of the volume were reconstructed. Horizontal and vertical slices containing the source are shown in figure 8.

4.3.2. Lines. The images of a linear source having endpoints (0, −2, 0) cm and (0, 10, 0) cm were reconstructed for the energies 662 keV and 5 MeV. The source was let to emit isotropically 10^8 gamma photons.

For the source emitting at 662 keV, about 1.3×10^5 valid events were acquired. From these, about 77% had total measured energies above 90% of the initial energy of the photons. Slices of the reconstructed images in the three scenarii (Monte Carlo calculated sensitivity, analytic

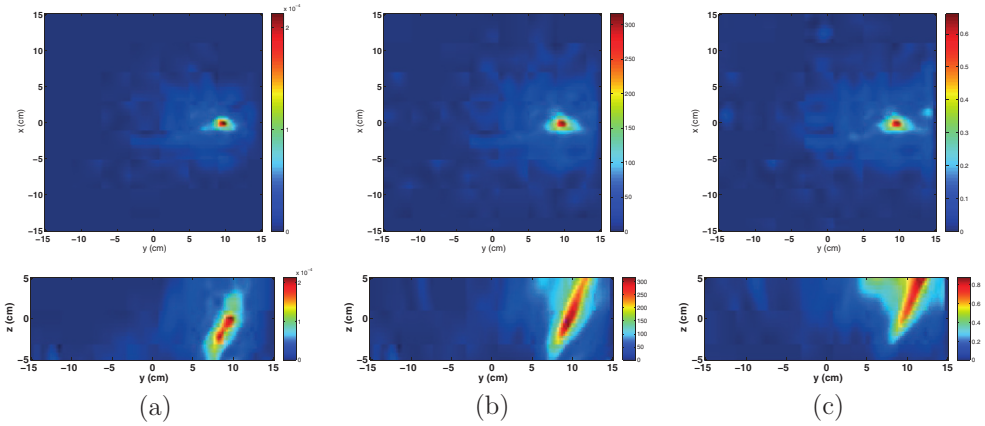


Figure 8. Point source placed in (0, 10, 0) cm emitting γ photons with energy of 5 MeV, reconstructed with different models of the sensitivity matrix. (a) MC sensitivity, (b) analytic sensitivity, (c) sensitivity set to 1.

sensitivity, and sensitivity set to 1) are shown in figure 9. It can be observed that the sensitivity set to 1 is the worst option between the three analyzed.

For the source emitting at 5 MeV, 7.4×10^4 valid events were acquired. Only about 10% had a total energy above 90% of the initial energy. Slices of the reconstructed images in the three scenarii (Monte Carlo calculated sensitivity, analytic sensitivity, and sensitivity set to 1) are shown in figure 10.

Reconstructed profiles of the lines were extracted from the images and are shown in figure 11. The images were previously divided by their respective means. An additional scaling factor was then applied, the same for all profiles in a panel. The evolution of the measured length of the line as a function of the number of iterations is plotted in figure 12.

4.3.3. Shepp–Logan. The last experiment was done with a 2D Shepp–Logan brain phantom emitting γ photons with an energy of 662 keV.

The reference distribution is shown in figure 13(a). Central slices from the reconstructed distribution are shown in figures 13(b) and (c). The image from panel (b) was produced from 2×10^5 events. A selection based on the total deposited energy in the detector was then operated and gave the image from panel (c). A total number of 1.5×10^5 events were selected, with total energy in the range of $\pm 10\%$ around the known initial energy of the photons. Analytic sensitivity was used in both cases. The selection based on the total energy slightly improves the results, although some small structures are still hidden.

5. Discussions and conclusion

Compton camera allows us to increase the proportion of detected photons compared with traditional collimated detectors used in SPECT imaging and may be successful for imaging relatively high energy sources. It can also produce a 3D image of the source from a still detector. However, because of the truncation of the projections induced by the loss of all photons not directed towards the camera, of the errors made in the estimation of the total energy of the photons, and of the wrong reconstruction of the path of the photon in the camera, this image is

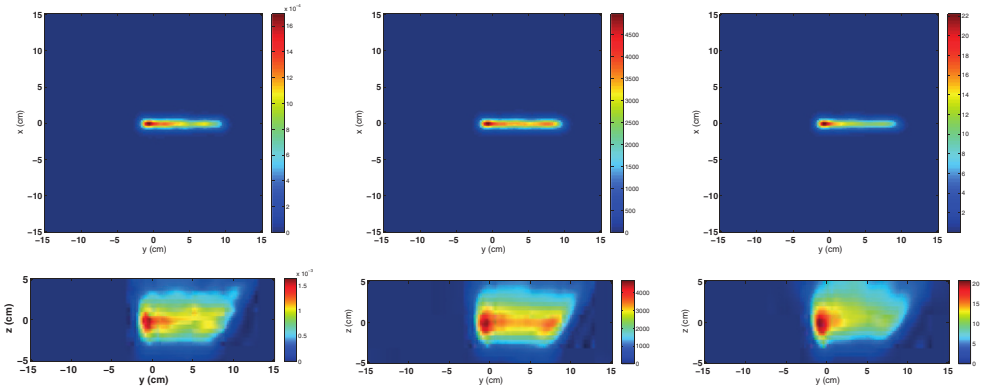


Figure 9. Images of a linear source with extremities (0, −2, 0) cm and (0, 10, 0) cm, emitting γ photons with energy of 662 keV, reconstructed with different models of the sensitivity matrix. (a) MC sensitivity, (b) analytic sensitivity, (c) sensitivity set to 1.

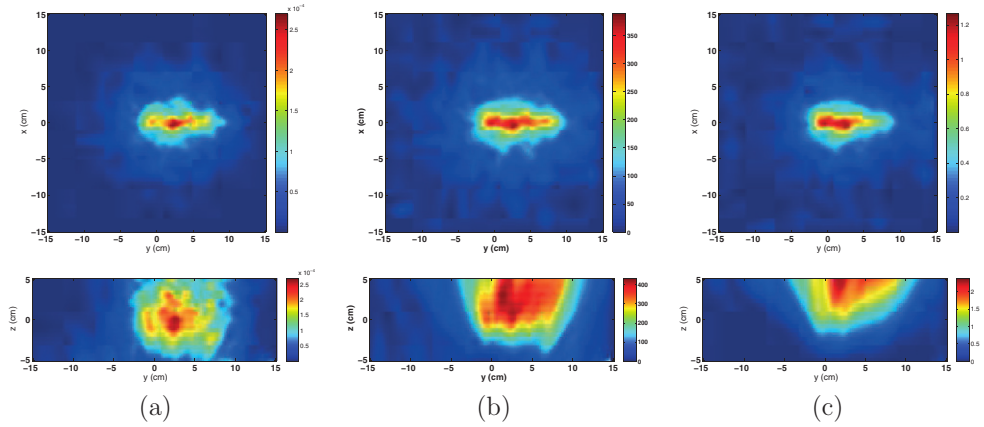


Figure 10. Images of a linear source with extremities (0, −2, 0) cm and (0, 10, 0) cm, emitting γ photons with energy of 5 MeV, reconstructed with different models of the sensitivity matrix. (a) MC sensitivity, (b) analytic sensitivity, (c) sensitivity set to 1.

affected by artifacts. The models used for the definition of the system and sensitivity matrices have important effects on the results.

We proposed a probabilistic model of the system matrix and a method for its numerical evaluation. We conclude that the solid angles covered by the detectors at the source points have to be accounted for in the calculation of the elements of the system matrix. The variant where only the distances of the points to the surface of the Compton cone is considered, together with the sensitivity set to 1, was also evaluated.

Figure 6(c) shows that the values of Monte Carlo calculated sensitivities for two energies are close to each other except for a proportionality factor previously eliminated via division by their respective means.

Monte Carlo calculation of the sensitivity was compared with a simple and fast analytic calculation. We observed that in some cases the analytic sensitivity gave better results than its Monte Carlo calculated counterpart. The reason is that the variance of our Monte Carlo

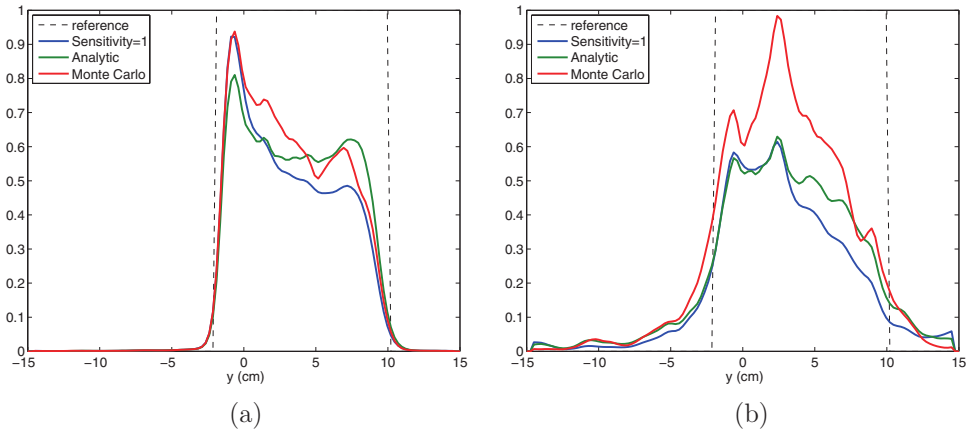


Figure 11. Central profiles from the reconstructed distributions for the line having extremities (0, −2, 0) cm and (0, 10, 0) cm. The source emitting at 662 kev is presented in (a) and the source emitting at 5 MeV is presented in (b). All the 3D distributions were scaled to a unitary mean.

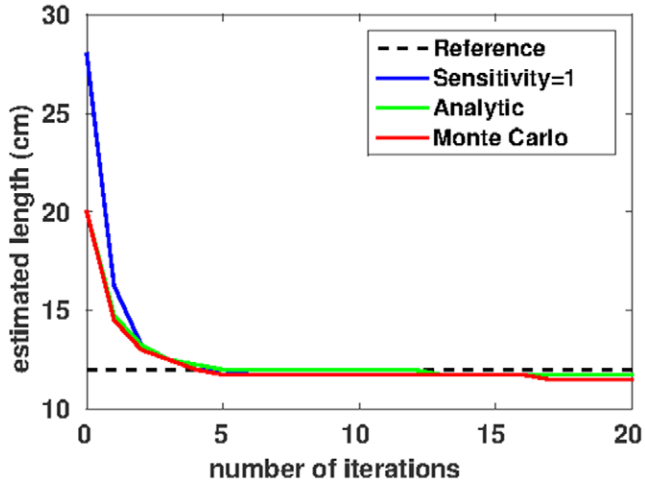


Figure 12. Length of the linear source emitting at 662 keV, as function of the number of iterations. The length is calculated as the full width at 20% of the mean intensity over the line.

sensitivity estimator is high with respect to the difference between its values in nearby voxels (see table 2). We conclude that an extensive Monte Carlo calculation is certainly better and could account for anatomical information, but the analytic estimate gives reasonably good results at a lower cost. We also conclude that the analytic method outperforms the short-run and thus biased Monte Carlo calculation and is an effective choice for the imaging of poly-energetic sources.

For a relatively low-energy point source, we observed that the differences between the three models of sensitivity (analytic, Monte Carlo, and identically set to 1) are minor. Some differences appear at the higher considered energy. The image reconstructed with the sensitivity set to 1 and the system matrix modeled as in (16) yelded the worst results, characterized

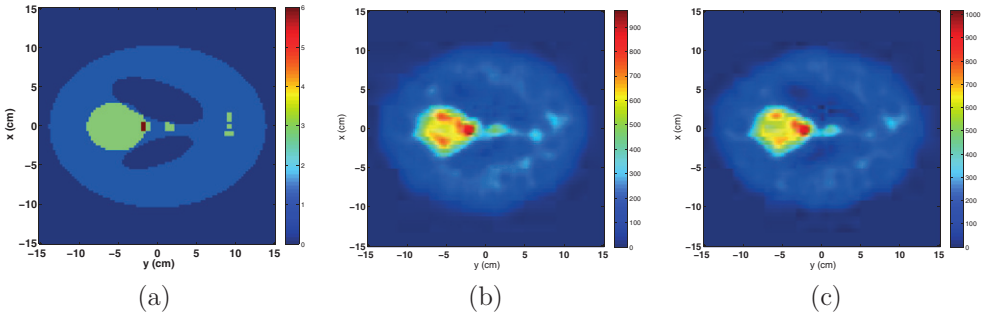


Figure 13. Two-dimensional Shepp–Logan phantom emitting at 662 keV, reconstructed in 3D with 40 iterations. The reference distribution is shown in (a). The reconstructed images are slices at $z = 0$ cm, corresponding to the position of the phantom. Realistic events without selection were considered for the reconstructed distribution shown in (b). A selection of the total energy deposited in the camera was operated for the image shown in (c).

by the largest point spread function and by a wrong localization of the maximum intensity in the slice orthogonal to the detector. Favored by incomplete absorptions and by a wrong recovery of the sequence of interactions of the photons in the camera, the noise increases with the energy. The spatially varying sensitivity improves the reconstructed images, especially in the direction orthogonal to the camera.

The images produced for a line source confirm the observations made for point sources. The variant model corresponding to the sensitivity set to 1 shows poor results especially for the points located outwith the range of the camera. The large variance in the Monte Carlo sensitivity leads to heterogeneities in the intensity of the source. In all cases, the vertical distribution of the source is poorly reproduced. The finite size of the detector in horizontal directions leads to severe truncation of the projections, corresponding to photons that miss the detector, and to specific vertical elongation artifacts. This problem is not specific to Compton imaging. Considerable work was done in classical tomography aiming to determine wether local tomography is possible in specific detection geometries and for which region of the object. As an example, Zhang and Zeng (2007) present an analysis of the truncation projection matrix and of its Moore-Penrose inverse matrix that helps to determine the recoverable region-of-interest in 2D MLEM image reconstruction from line-driven projections.

Finally, the performances of the algorithm were demonstrated on a Shepp–Logan brain phantom. It is shown that applying an energy selection window aiming to select almost completely absorbed events improves the reconstructed images. At least for the considered camera that has a relatively thin absorber, the benefits of the selection become more clear as the energy increases.

Acknowledgments

This research was supported by the ENVISION project (co-funded by the European Commission under the FP7 Collaborative Projects Grant Agreement Nr. 241851FP7), by the ETOILE Research Program (PRRH/UCBL, under CPER 2007-13 funding) and by the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon. The authors would like to thank the anonymous reviewers for careful reading of the manuscript and thoughtful comments.

References

Agostinelli S *et al* 2003 Geant4—a simulation toolkit *Nucl. Instrum. Methods* A **506** 205–303

Basko R, Zeng G L and Gullberg G T 1998 Application of spherical harmonics to image reconstruction for the Compton camera *Phys. Med. Biol.* **43** 887–94

Cree M J and Bones P J 1994 Towards direct reconstruction from a gamma camera based on Compton scattering *IEEE Trans. Med. Imaging* **13** 398–407

Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc.* B **39** 1–38

Frandes M, Zoglauer A, Maxim V and Prost R 2010 A tracking Compton-scattering imaging system for hadron therapy monitoring *IEEE Trans. Nucl. Sci.* **57** 144–50

Gillam J E, Lacasta C, Torres-Espallardo I, Juan C C, Llosá G, Solevi P, Barrio J and Rafecas M 2011 A Compton imaging algorithm for on-line monitoring in hadron therapy *Proc. SPIE Medical Imaging* **7961** 796110–8

Hirasawa M and Tomitani T 2003 An analytical image reconstruction algorithm to compensate for scattering angle broadening in Compton cameras *Phys. Med. Biol.* **48** 1009–26

Hueso-González F *et al* 2014 Test of Compton camera components for prompt gamma imaging at the ELBE bremsstrahlung beam *J. Instrum.* **9** P05002

Kim S, Lee J, Lee M, Lee J, Lee C, Kim C-H, Lee D and Lee S 2007 Two approaches to implementing projector-backprojector pairs for 3D reconstruction from Compton scattered data *Nucl. Instrum. Methods Phys. Res.* A **571** 255–8

Krimmer J *et al* 2015 Development of a Compton camera for medical applications based on silicon strip and scintillation detectors *Nucl. Instrum. Methods* A **787** 98–101

Lange K and Carson R 1984 EM reconstruction algorithms for emission and transmission tomography *J. Comput. Assist. Tomogr.* **8** 306–16

Lee S-J, Lee M N, Nguyen V-G, Kim S and Lee J S 2008 Three dimensional edge-preserving regularization for Compton camera reconstruction *IEEE Nuclear Science Symp. and Medical Imaging Conf.* pp 1095–7863

Lojacono X, Richard M-H, Ley J-L, Testa E, Ray C, Freud N, Létang J M, Dauvergne D, Maxim V and Prost R 2013 Low statistics reconstruction of the Compton camera point spread function in 3D prompt-γ imaging of ion beam therapy *IEEE Trans. Nucl. Sci.* **60** 3355–63

Mackin D, Peterson S, Beddar S and Polf J 2012 Evaluation of a stochastic reconstruction algorithm for use in compton camera imaging and beam range verification from secondary emission during proton therapy *Phys. Med. Biol.* **57** 3537–53

Maxim V, Frandes M and Prost R 2009 Analytical inversion of the Compton transform using the full set of available projections *Inverse Problems* **25** 1–21

Maxim V 2014 Filtered backprojection reconstruction and redundancy in Compton camera imaging *IEEE Trans. Image Process.* **23** 332–41

Nguyen V-G, Lee S-J and Lee M 2011 GPU-accelerated 3D Bayesian image reconstruction from Compton scattered data *Phys. Med. Biol.* **56** 2817–36

Ordonez C E, Bolozdynya A and Chang W 1997 Dependence of angular uncertainties on the energy resolution of Compton cameras *IEEE Nuclear Science Symp.* vol 2

Ortega P G *et al* 2015 Noise evaluation of Compton camera imaging for proton therapy *Phys. Med. Biol.* **60** 1845–63

Parra L C 2000 Reconstruction of cone-beam projections from Compton scattered data *IEEE Trans. Nucl. Sci.* **47** 1543–50

Peterson S, Robertson D and Polf J 2010 Optimizing a three-stage Compton camera for measuring prompt gamma rays emitted during proton radiotherapy *Phys. Med. Biol.* **55** 6841–56

Richard M-H *et al* 2012 Design study of the absorber detector of a compton camera for on-line control in ion beam therapy *IEEE Trans. Nucl. Sci.* **59** 1850–5

Sauve A C, Hero A O, Rogers W L, Wilderman S J and Clinthorne N H 1999 3D image reconstruction for a Compton SPECT camera model *IEEE Trans. Nucl. Sci.* **46** 2075–84

Shepp L and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imaging* **1** 113–22

Smith B 2005 Reconstruction methods and completeness conditions for two Compton data models *J. Opt. Soc. Am.* A **22** 445–59

Sullivan J, Tornga S and Rawool-Sullivan W 2009 Extended radiation source imaging with a prototype Compton imager *Appl. Radiat. Isot.* **67** 617–24

Tomitani T and Hirasawa M 2002 Image reconstruction from limited angle Compton camera data *Phys. Med. Biol.* **47** 2129–45

Vardi Y, Shepp L and Kaufman L 1985 A statistical model for positron emission tomography *J. Am. Stat. Assoc.* **80** 8–20

Wilderman S J, Clinthorne N, Fessler J A and Rogers W L 1998a List-mode maximum likelihood reconstruction of Compton scatter camera images in nuclear medicine *Nuclear Science Symp. Conf. Rec.* **3** 1716–20

Wilderman S J, Fessler J, Clinthorne N, LeBlanc J and Rogers W 2001 Improved modeling of system response in list mode EM reconstruction of Compton scatter camera images *IEEE Trans. Nucl. Sci.* **48** 111–6

Wilderman S J, Rogers W L, Knoll G F and Engdahl J C 1998b Fast algorithm for list mode back-projection of Compton scatter camera images *IEEE Trans. Nucl. Sci.* **45** 957–62

Xu D and He Z 2007 Gamma-ray energy-imaging integrated deconvolution *Nucl. Instrum. Methods* A **574** 98–109

Zhang B and Zeng G L 2007 Two-dimensional iterative region-of-interest (ROI) reconstruction from truncated projection data *Med. Phys.* **34** 935–44

Zoglauer A 2005 First light for the next generation of Compton and pair telescopes *PhD Thesis* Munich Germany

Zoglauer A, Andritschke R and Schopper F 2006 MEGAlib—the medium energy gamma-ray astronomy library *New Astron. Rev.* **50** 629–32

Zoglauer A, Boggs S, Galloway M, Amman M, Luke P and Kippen R 2011 Design, implementation, and optimization of MEGAlib's image reconstruction tool Mimrec *Nucl. Instrum. Methods Phys. Res.* A **652** 568–71

Zoglauer A, Weidenspointner G, Galloway M, Boggs S E and Wunderer C 2009 Cosima—the Cosmic simulator of MEGAlib *Nuclear Science Symp. Conf. Rec.* pp 2053–9

IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology

Phys. Med. Biol. **61** (2016) 3127–3147 doi:10.1088/0031-9155/61/8/3127

Proton therapy monitoring by Compton imaging: influence of the large energy spectrum of the prompt-γ radiation

Estelle Hilaire, David Sarrut, Françoise Peyrin and Voichiţa Maxim

Université de Lyon, CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon, Université Lyon 1, France

E-mail: maxim@creatis.insa-lyon.fr

Received 8 June 2015, revised 24 February 2016
Accepted for publication 2 March 2016
Published 23 March 2016

Abstract

In proton therapy, the prompt-γ (PG) radiation produced by the interactions between protons and matter is related to the range of the beam in the patient. Tomographic Compton imaging is currently studied to establish a PG image and verify the treatment. However the quality of the reconstructed images depends on a number of factors such as the volume attenuation, the spatial and energy resolutions of the detectors, incomplete absorptions of high energy photons and noise from other particles reaching the camera. The impact of all these factors was not assessed in details. In this paper we investigate the influence of the PG energy spectrum on the reconstructed images. To this aim, we describe the process from the Monte Carlo simulation of the proton irradiation, through the Compton imaging of the PG distribution, up to the image reconstruction with a statistical MLEM method. We identify specific PG energy windows that are more relevant to detect discrepancies with the treatment plan. We find that for the simulated Compton device, the incomplete absorption of the photons with energy above about 2 MeV prevents the observation of the PG distributions at specific energies. It also leads to blurred images and smooths the distal slope of the 1D PG profiles obtained as projections on the central beam axis. We show that a selection of the events produced by γ photons having deposited almost all their energy in the camera allows to largely improve the images, a result that emphasizes the importance of the choice of the detector. However, this initial-energy-based selection is not accessible in practice. We then propose a method to estimate the range of the PG profile both for specific deposited-energy windows and for the full spectrum emission. The method relies on two parameters. We use a learning approach for their estimation and we show that it allows to detect few millimeter shifts of the PG profiles.

0031-9155/16/083127+21$33.00 © 2016 Institute of Physics and Engineering in Medicine Printed in the UK

Keywords: proton therapy, monitoring, tomographic reconstruction, Compton camera, MLEM, Monte Carlo simulation

(Some figures may appear in colour only in the online journal)

1. Introduction

In proton therapy cancer treatments, interactions between the beam and the patient's body lead to nuclear reactions that produce secondary PG radiation. It was shown that some correlation exists between the emission point of those photons and the Bragg peak position (e.g. Kurosawa *et al* (2012), Min *et al* (2006), Stichelbaut and Jongen (2003) and Testa *et al* (2009)). Ongoing works aim to design imaging systems able to monitor the proton range by exploiting PG radiation. Among the proposed designs, we can mention SPECT systems such as knife-edge-slit, pinhole and parallel-slit collimator cameras (e.g. Bom *et al* (2012), Kim *et al* (2009), Lee *et al* (2012), Cambraia Lopes *et al* (2012), Perali *et al* (2014), Smeets *et al* (2012), Verburg and Seco (2014)), as well as Compton cameras (e.g. Frandes *et al* (2010), Kormoll *et al* (2011), Kurosawa *et al* (2012) and Richard *et al* (2012)). Both mechanically collimated SPECT systems and Compton cameras aim to determine the spatial distribution of the PG emission. Alternative approaches for the verification of the treatment are currently investigated. For instance, Verburg *et al* (2013) suggest to exploit the spatial variability of the gamma-ray spectrum and Golnik *et al* (2014) propose a method relying on timing measurements.

As mentioned by Polf *et al* (2009), the fall-off of the real PG profile is shifted from the Bragg peak because of the decrease in production of gamma rays for proton energies below of about 25 MeV. The profile reconstructed from the data shows additional offsets due to attenuation in the patient, acquisition factors and sequential errors introduced during data processing and image reconstruction steps.

During the treatment, secondary particles other than γ photons are produced. In the literature related to collimated cameras it is shown that those sources of noise may be partly avoided by adequate shielding of the detector (Min *et al* 2006), energy selection (Testa *et al* 2009, Smeets *et al* 2012) and time-of-flight selection (Verburg *et al* 2013, Pinto *et al* 2014). In a recent work, Ortega *et al* (2015) studied the influence of neutrons on the images obtained from Compton camera data. They also studied the noise induced by random coincidences that may occur when the detector becomes saturated. An other issue related to real-time quality control is the ability of the detection system to produce reliable outcome from a low number of data. In a simulation study, Lojacono *et al* (2013) evaluated from the reconstructed images the properties of the point spread function of a Compton camera as function of the number of detected events.

As already shown in Styczynski *et al* (2009) and more recently by Verburg *et al* (2013), the energy spectrum of the PG emission is spatially varying. At each given depth it depends on the residual energy of the beam. When a water phantom is irradiated with a proton beam, discrete gamma lines at 4.44, 5.2 and 6.13 MeV may be observed in the spectrum. Moreover, these lines are salient just before the end-of-range of the beam and disappear behind it. It is also known that the energy spectrum depends strongly on the irradiated tissue. Observation of the spectrum with a collimated camera may thus give important insight on the range of the beam (Polf *et al* 2009, Verburg *et al* 2013, Perali *et al* 2014) and on the chemical composition of the irradiated tissues (Moteabbed *et al* 2011, Polf *et al* 2009, 2013, Verburg and Seco 2014).

While the development of PG collimated cameras is advanced enough to allow tests in clinical conditions, they are still limited to provide 1D information. Next generation PG systems

should allow 2D or 3D imaging. The capability of Compton cameras to provide such images may be an important step forward for the monitoring of the treatment for beams going through heterogeneous tissues. In theory, the Compton camera efficiency is superior to Anger cameras, but it is also disadvantaged by the necessity to measure coincident events. Their capabilities struggle to convince and the performances of the available tomographic reconstruction algorithms is currently an important bottleneck. Some reconstruction algorithms were proposed in the literature, either based on analytic solutions (e.g. Maxim (2014), Smith (2005) and Tomitani and Hirasawa (2002)) or on iterative methods (e.g. Gillam *et al* (2011), Lojacono *et al* (2013), Mackin *et al* (2012) and Wilderman *et al* (1998)).

The main purpose of this paper is to investigate through simulations the capability of a proposed reconstruction algorithm to produce 3D maps of the PG distributions at given energies from the spectrum, to identify the limitations, and to extract from the images estimators of the fall-off position. To our knowledge, this is the first study of the 4D spatial and spectral PG distribution, with a Compton camera. As we focus on the analysis of the influence of the energy on the reconstructed images, number of noise sources were neglected and the number of emitted protons was set in the simulation to provide sufficient PG outcome for the reconstruction algorithm. The conclusions of our study were drawn for $4 \cdot 10^{11}$ emitted protons, although some illustrations concerning the total distribution are given for 10^9 primaries. As a comparison, the results from the measurements reported in Verburg *et al* (2013) for a collimated camera are means over 5 runs of 10^{10} primaries each. In Verburg and Seco (2014) a sub-millimeter precision on the proton range measurement was attained with $5 \cdot 10^8$ protons. When the number of primaries was lowered to 10^8 protons, the precision dropped to about 3 mm in water and about 4 mm in solid water. In Perali *et al* (2014) is shown that a slit camera allows to reach precision of 4 mm on range retrieval, with as few as about 10^8 protons. The PG profiles used in the study are samples drawn from a rescaled low-noise profile, by addition of Poisson noise. The reference low-noise profile was obtained by measuring the PG emission of a PMMA phantom irradiated with 10^{11} protons.

We simulated a two-stage Compton camera with limited footprint and design similar to the one described in Krimmer *et al* (2014). The camera is composed of seven layers of scatterer made in silicon and a 4 cm-thick absorber made of LYSO crystals. Because of incomplete absorptions, the information on the initial energy of the particles may be lost. An important issue from then is to determine the PG measured energy windows that provide relatively well reconstructed images and are thus best adapted to detect discrepancies with the treatment plan. PG profiles are extracted from the images and an estimator of the distal fall-off position is proposed and employed to characterize the ability of the simulated Compton camera to provide relevant information on the quality of the treatment by exploiting radiation in the selected energy window. Although our results are strongly related to the properties of the simulated camera, the proposed image reconstruction and range calculation methods could be applied to other designs.

2. Materials and methods

In this section, we present the different steps leading to the reconstruction of the Compton camera image of the PG distribution and to the estimation of the PG profile fall-off position. The proposed work-flow is summarized in figure 1. The irradiation of a water phantom with a proton pencil beam is simulated with the software GATE (Jan *et al* 2011, Sarrut *et al* 2014). A joint space-energy distribution $(v, \varepsilon) \in \mathbb{R}^3 \times \mathbb{R}_+^* \mapsto \lambda(v, \varepsilon)$ is obtained and the marginal distribution $\lambda_{\text{tot}} = \sum_\varepsilon \lambda(\cdot, \varepsilon)$ is deduced. The simulation of the Compton imaging of

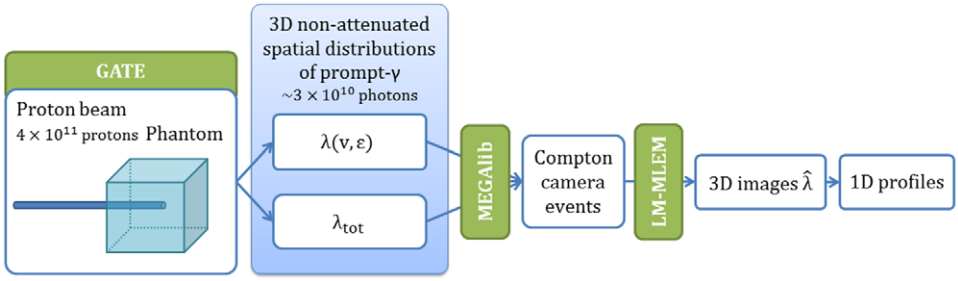


Figure 1. Simulation of the irradiation of a water phantom by a proton pencil beam with GATE. The distribution λ of the PG emission is extracted. The interaction of the PG photons in the Compton camera is simulated with MEGAlib. The 3D distribution $\widehat{\lambda}$ is reconstructed with a LM-MLEM algorithm and projected on the central beam axis to obtain a 1D profile.

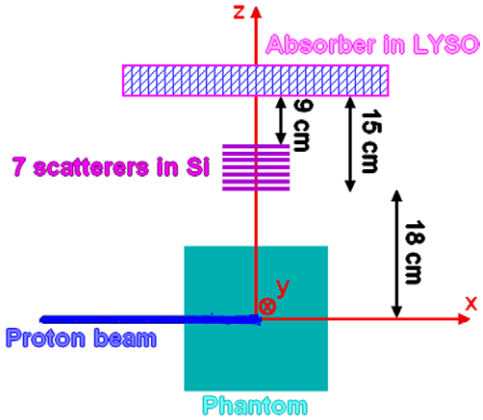


Figure 2. Architecture of the simulation.

the PG distributions is then realized with the software MEGAlib (Zoglauer *et al* 2006). We introduce the method used to reconstruct the images from data recorded in the camera. Finally we present a method for the calculation of the depths from 1D profiles obtained from the reconstructed images.

2.1. Simulation process

2.1.1. Simulation of the prompt-γ emission. We simulated the irradiation of a water phantom of dimensions $(20\,\text{cm})^3$ with a proton pencil beam of energy 120 MeV having 2D Gaussian profile with $(\sigma_y, \sigma_z) = (4, 2)$ mm (see figure 2). The pencil beam emits 4×10^{11} protons. Note that in clinical conditions, about 10^8 protons are emitted per spot, 10^{10} protons are generally used for the irradiation of the distant layer of the tumor and about 10^{11} protons are emitted per fraction of irradiation. Our set-up corresponds thus to the dose delivered to the entire treatment plan, concentrated in one single spot, which is probably above what could be achieved in clinic. However, taking into account the simulated efficiency of the modeled camera, which is of about 10^{-3}, the chosen intensity allows us to study the performances of the proposed reconstruction algorithm for different slices of the gamma energy spectrum.

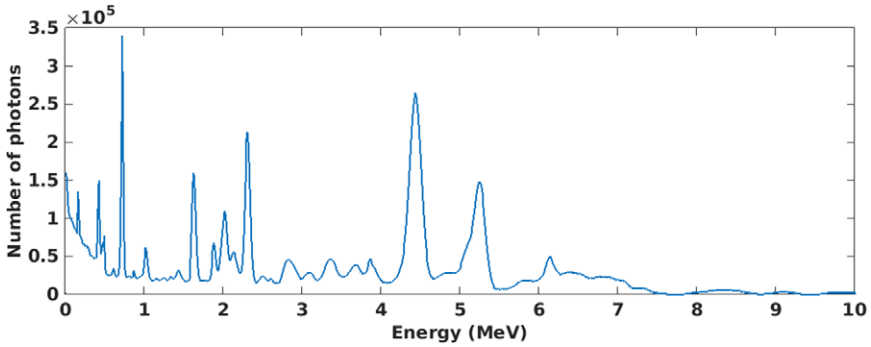


Figure 3. Energy spectrum $\sum_v \lambda(v, \cdot)$ of the PG radiation produced by a proton pencil beam of energy 120 MeV in water (bins of 10 keV).

The simulation of the irradiation was performed with GATE, a Geant4-based Monte-Carlo software. We used GATE version 7 with the QGSP_BIC_HP_EMY physics list, and Geant4 version 9.6.3. To limit the computation time, at this stage only 4×10^8 protons were simulated, the remaining 10^3 factor being applied to the PG distribution as mentioned bellow.

The non attenuated 4D joint energy-spatial frequency distribution λ of the PG emission was then recorded. The marginal distribution of the energy spectrum of the PG, cut at the interval (100 keV, 10 MeV), is shown in figure 3. Some dominant discrete γ lines may be observed, in particular those at 4.4 MeV and 5.2 MeV. They are produced by proton-induced reactions on ^{16}O at energies up to 50 MeV and were shown to characterize the end of the range of the beam (Verburg *et al* 2013). Note that all the secondary particles except PG were removed as soon as they were created.

Bins of $\Delta\varepsilon = 500$ keV were taken for the energy and the phantom was sampled into 100^3 voxels of 2 mm side. For each given energy bin having center ε_k and width $\Delta\varepsilon$, a three dimensional spatial distribution $\lambda(\cdot, \varepsilon_k)$ was obtained. It associates to each voxel v_j from the phantom volume the number $N_{k,j}$ of proton-induced photons emitted in the voxel v_j with an energy in the interval $[\varepsilon_k - \Delta\varepsilon/2, \varepsilon_k + \Delta\varepsilon/2)$. To decrease the variance induced by the limited statistics of the Monte Carlo simulation, the spatial distributions were smoothed with a 2D Gaussian filter having a FWHM of 2 pixels, for each slice parallel to the detector. The resulting distribution γ was then multiplied with 10^3, leading to the targeted statistics that corresponds to 4×10^{11} emitted protons. Each voxel v_j of the volume was then considered as an isotropic mono-energetic source of photons. Note that this procedure leads to the loss of some physical information related to the PG particles, namely the exact emission positions and energies, and the direction of the particles. In turn, it allows to lower the computational cost of the simulation and to reduce the influence of statistical fluctuations. Fast simulation of PG emission is a topic of high importance for proton-therapy monitoring. A method based on track length estimator was recently proposed in Kanawati *et al* (2015).

2.1.2. Simulation of the Compton camera acquisition. The simulated camera consists of a scatterer made of seven silicon layers of dimensions $9.6 \times 9.6 \times 0.2$ cm^3 with 1 cm gap between two successive layers, placed at 9 cm distance from an absorber made of LYSO measuring $37 \times 37 \times 4.2$ cm^3. Each silicon scatterer contains 2×128 strips with energy resolution of 2.35 keV FWHM. The LYSO crystals of the segmented absorber measure $0.5 \times 0.5 \times 4$ cm^3 each and have an energy resolution depending upon a Gaussian distribution on the energy of

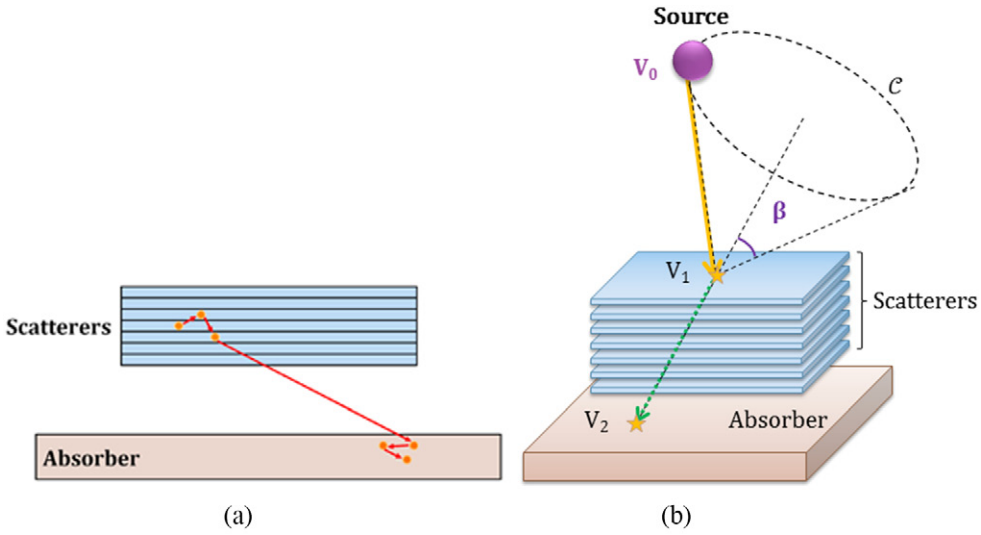


Figure 4. (a) A γ photon undergoes Compton scatterings in a first detector called scatterer, that might be composed of several layers, and is finally absorbed in a second detector called absorber. The order in the sequence of interactions cannot be established but only estimated. (b) From position and energy measures, the point that generated the photon is located on the surface of the cone. (a) Sequence of Compton interactions. (b) Compton camera principle.

the incoming photon. The standard deviation σ of the Gaussian varies from 3.5% to 1.2% in the range of 0.1 MeV to 10 MeV. The assumed energy resolution is too optimistic in view of new data (Roemer *et al* 2015), but has no major impact on the results presented here. The camera and its positioning with respect to the phantom are shown in figure 2. The position of a hit in the absorber is calculated with a barycentric rule and the z coordinate is set as the middle of a crystal (the depth of interaction is supposed unknown).

The Compton camera acquisition of the γ photons was simulated with the software Cosima (Zoglauer *et al* 2009), part of MEGAlib version 2.28. Built on Geant4 too, this software was designed for Compton camera simulated and real experiments.

The individual recording of the photons avoids random coincidences. An event is triggered when at least one interaction is detected in each detector of the camera. Spatial and energy resolutions of the detectors and Doppler broadening are accounted for. The order of the hits of a given photon is supposed unknown and the most likely path is calculated from interaction positions and deposited energies with the software Revan, also part of the MEGAlib library.

The Compton camera detects a photon in two steps. The photon, having initial energy E_0, is scattered in a first detector (scatterer) at some position V_1 and deposits an energy denoted E_1. The scattered photon may further interact in the first detector and, ideally, is finally absorbed in a second detector (absorber) by photoelectric effect (see figure 4). The energy deposited by the photon after the first Compton scattering will be denoted hereafter E_2. Ideally, the remaining energy $E_0 - E_1$ is entirely deposited in the camera and we should have $E_0 = E_1 + E_2$. If this is the case, the conservation law of energy allows to recover the scattering angle of the photon (Compton 1923) through the equation:

$$\cos\beta = 1 - \frac{m_e c^2 E_1}{E_2(E_1 + E_2)}, \quad \text{with } m_e c^2 = 511 \text{ keV}. \tag{1}$$

Generally a part of the initial energy is deposited in non sensitive material or is taken away by the photon or its derivatives escaping from the device. Thus, the measured total energy $E_{\text{tot}} = E_1 + E_2$ does not always match the initial energy E_0, leading to an overestimation of the Compton angle.

In case of multiple interactions in the camera (the most probable case in case of high energies), the Revan software decides which of the interaction locations is taken as the first and second ones defining the cone axis. The energy associated to the hit identified to be the first is considered from then as E_1 and E_2 comprises the sum of all energies deposited in the other hits. The calculated positions of the first and second hit, denoted V_1 and V_2, and the energies E_1 and E_2 define an event $e = (V_1, E_1, V_2, E_2)$. The incoming path of the initial ray is then supposed to lie on the surface of the cone defined by the apex V_1, the half-opening angle β (i.e. the scattering angle of the photon) and the axis direction $\overrightarrow{V_2V_1}$ (see figure 4(b)).

2.2. Reconstruction method

Let $(e_i)_{i=1,\ldots,N_\gamma}$ be the list of recorded events and $(\lambda_j)_{j=1,\ldots,N_v}$ be the discretized 3D image of the source, with N_v the total number of voxels. The 3D image is reconstructed using the LM-MLEM algorithm (Shepp and Vardi 1982, Wilderman *et al* 1998). From an initial estimate $\lambda^{(0)}$, the algorithm calculates successive approximations $\lambda^{(\ell)}$ of $\widehat{\lambda}$, the maximum likelihood estimator of the image. The sequence $\lambda^{(\ell)}$ is supposed to converge towards the vector λ representing the mean emission intensities in the voxels of the volume. The iterations are based on the equation:

$$\lambda_j^{(\ell+1)} = \frac{\lambda_j^{(\ell)}}{s_j} \sum_{i=1}^{N_\gamma} t_{ij} \frac{1}{p_i^{(\ell)}}, \text{ with } p_i^{(\ell)} = \sum_{k=1}^{N_v} t_{ik} \lambda_k^{(\ell)}, \tag{2}$$

where t_{ij} is the probability for a photon emitted by the voxel v_j to be detected as event e_i and s_j is the probability for a photon emitted by the voxel v_j to be detected. The matrix $T = (t_{ij})_{i=1,\ldots,N_\gamma, j=1,\ldots,N_v}$, generally called the system matrix, and the sensitivity $S = (s_j)_{j=1,\ldots,N_v}$ are the key parameters of the algorithm.

2.2.1. Calculation of the system matrix. The system matrix T accounts for the specific conical shape of the set of source points that could produce an event and for the known energy uncertainties of the detector through a spatial kernel h_i. For a given event $e_i = (V_1, E_1, V_2, E_2)$ let us denote β_i the estimated Compton angle. We define h_i as a Gaussian of mean β_i and standard deviation σ_{β_i} calculated, as suggested by Ordonez *et al* (1997), from the deposited energies and the resolution of the detectors as:

$$\sigma_{\beta_i} = \frac{m_e c^2}{\sin \beta_i} \sqrt{\left(\frac{1}{E_1^2} - \frac{1}{E_{\text{tot}}^2}\right)^2 \mathrm{d}E_1^2 + \frac{\mathrm{d}E_2^2}{E_{\text{tot}}^4}}. \tag{3}$$

Let M be a point from the volume to reconstruct. For δ_M the angular distance between $\overrightarrow{V_2V_1}$ and $\overrightarrow{V_1M}$, the expression of the spatial kernel is then:

$$h_i(M) = \frac{1}{\sqrt{2\pi}\,\sigma_{\beta_i}} \exp\left(-\frac{1}{2}\left(\frac{\delta_M - \beta_i}{\sigma_{\beta_i}}\right)^2\right). \tag{4}$$

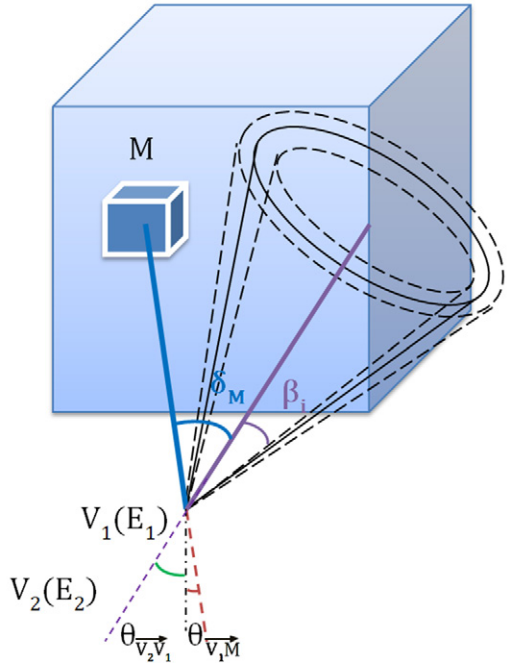


Figure 5. Parameters for the calculation of the probabilities t_{ij}.

For a given vector $\vec{w}$, we denote $\theta_{\vec{w}}$ the angle between $\vec{w}$ and the normal to the camera. Figure 5 shows the parameters entering in the definition of the system matrix. According to Maxim *et al* (2016), we model the elements of the matrix T as:

$$t_{ij} = \frac{|\cos(\theta_{\overrightarrow{V_2V_1}})|}{\|\overrightarrow{V_2V_1}\|} \int_{M\in v_j} K(\delta_M, E_{\text{tot}}) \frac{|\cos(\theta_{\overrightarrow{V_1M}})|}{\|\overrightarrow{V_1M}\|^2} h_i(M)\mathrm{d}v, \tag{5}$$

where $K(\delta_M, E_{\text{tot}})$ is the Compton scattering cross section with parameters calculated from the data. The spatial kernel h_i allows to account for measurement uncertainties by considering that every point of the image may produce the given event, with a probability depending on its angular distance to the surface of the associated Compton cone.

For the numerical evaluation of (5), we denote O_j the center of the voxel v_j and we propose to use the approximation:

$$t_{ij} = K(\delta_{O_j}, E_{\text{tot}}) \frac{|\cos(\theta_{\overrightarrow{V_2V_1}})|}{\|\overrightarrow{V_2V_1}\|} \frac{|\cos(\theta_{\overrightarrow{V_1O_j}})|}{\|\overrightarrow{V_1O_j}\|^2} h_i(O_j), \text{ if } |\delta_{O_j} - \beta_i| \leqslant 3\sigma_{\beta_i}, \tag{6}$$

and $t_{ij} = 0$ otherwise. Proportionality to the volume of the voxel was ignored in (6).

2.2.2. Calculation of the sensitivity. The sensitivity $S = (s_j)_{j=1,\ldots,N_v}$, that allows in (2) to compensate for the loss of photons emitted in distant voxels, can be assessed either by Monte Carlo simulation or approximated analytically as in Wilderman *et al* (2001). In the reconstructions performed in this work, the calculation method did not have a significant influence on the results. The analytical method was hereafter chosen for its satisfactory performances at negligible computational cost.

The analytical computation consists in considering some simplifications in the detector's geometry. Assuming that the absorber is wide compared to the scatterer and placed close to it, that the Compton cross sections $K(\beta, E_{\text{tot}})$ have similar influence for all voxels and that the size of a voxel is small, the variations in s_j are essentially due to variations of the solid angle subtended by the scatterer volume at the centre O_j of the voxel. Assuming that the side of the scatterer is small compared to the distance from the source to the detector, we put:

$$s_j \propto \frac{|\cos(\theta_{\overrightarrow{QO_j}})|}{\|QO_j\|^2}, \tag{7}$$

where Q is the center of the scatterer block.

2.3. Calculation of the depth of the distal fall-off of the prompt-γ profile

Compared to the dose fall-off which corresponds to a limit on the irradiated region, the PG profile fall-off has less meaning for the treatment. Its main role is to provide a quantitative tool for the measure of range shifts. In the literature, its estimation vary as function of the detection system.

Let us call f the 1D profile obtained by projecting the 3D PG distribution λ on the proton central beam axis. The fall-off of the simulated PG profile was taken as it is usually done for the Bragg peak at 80% of the maximum value M, i.e. $d = f^{-1}(0.8\ M)$. For the reconstructed profiles, the abscissa of the point where the profile drops under 80% of the maximum value is not always suitable and presents some variability. Depending on the shape of the curve, a specific percentage or a mean over a range of percentages may be more convenient.

For high-variability curves, we suggest to calculate the fall-off as the mean over a range of percentages $[p_{\min}, p_{\max}]$ according to:

$$\widehat{d}_{p_{\min},p_{\max}} = \frac{1}{(p_{\max} - p_{\min})M} \int_{p_{\min}M}^{p_{\max}M} f^{-1}(y)\mathrm{d}y \tag{8}$$

An example is illustrated in figure 6. The range $[p_{\min}M, p_{\max}M]$ is displayed in blue and the mean value $\widehat{d}_{p_{\min},p_{\max}}$ is represented by the green line. It may be shown that the two areas to the left and right of the mean, highlighted in blue, are equal.

For profiles showing low variability, one may take $p_{\min} = p_{\max}$, leading to the estimated fall-off position $\widehat{d}_{p_{\min},p_{\max}} = f^{-1}(p_{\min}M)$.

Some other methods have been proposed in the literature. In Ortega *et al* (2015), the value of M is the height of the deepest peak and $\widehat{d} = f^{-1}(0.8M)$ or $\widehat{d} = f^{-1}(0.5M)$. For profiles presenting important variability, a prior fit of the data with some smooth function is recommended (Smeets *et al* 2012, Gueth *et al* 2013, Janssen *et al* 2014).

Obviously, it exists an infinity of couples of parameters $(p_{\min}, p_{\max})$ leading to $\widehat{d}_{p_{\min},p_{\max}} = d$. An important issue from then is the choice of the parameters and it seems reasonable to learn the values from some previous experience. Two examples will be given in section 3.4.

2.4. Calculation of the range shift

One of the challenges for a proton therapy monitoring system is the range shifts detection. By analogy with the beam range, we define the PG deduced proton range as the distance from the phantom entry to the fall-off position of the PG profile.

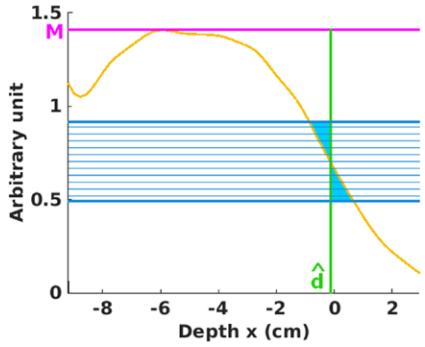


Figure 6. Depth calculation.

A shift of 5 mm is produced by increasing the beam energy from 120 MeV to 123 MeV. The same result is obtained when the energy increases from 123 MeV to 126 MeV. At 120 MeV, the simulated Bragg peak is located at $x = 6$ mm, which corresponds to a range of the beam in the phantom of 106 mm. As expected, the PG deduced proton range is shorter and measures 100 mm.

We define the estimated range shift induced by the increase of the energy of the beam as the difference between the corresponding estimated fall-off positions.

3. Results

3.1. Prompt-γ distribution as function of the energy

Figure 7 shows 2D slices from the 3D distributions $\lambda(\cdot, \varepsilon_k)$ of the PG rays produced in the phantom, for the energy bins described in section 2.1.1. It should be noticed that the color scale differs from one slice to another.

It may be seen that the energies up to 1.5 MeV are more represented at the beginning of the path of the beam. For some other energies, the distribution presents a peak close to the end-of-range. This is the case for instance for the windows [4–4.5] MeV and [5–5.5] MeV which include two of the remarkable gamma lines mentioned by, e.g. Verburg *et al* (2013).

3.2. Influence of the prompt-γ energy on the reconstructed images

The distributions simulated as described in section 2.1.1 and plotted in figure 7 were then used to generate Compton camera data. For each distribution, a volume of $20 \times 8 \times 2.5\,\text{cm}^3$ divided in $100 \times 41 \times 13$ voxels centered on the axis of the beam was reconstructed with the iterative LM-MLEM algorithm described in section 2.2. Central horizontal slices from the results at iteration 15 are shown in figure 8. Given the considered geometry, the acquired Compton projections are truncated. Despite correction with the sensitivity factor in (2), the reconstructed images are still not quantitative. Therefore, the color scales are not displayed in figures 8 and 9. Less studied for Compton tomography, the topic of image reconstruction from incomplete projections is better understood in classical tomography and some analytical conditions leading to uniqueness of the solution of the region of interest imaging problem were established. In Zhang and Zeng (2007), the properties of the MLEM algorithm are investigated and compared to analytical solutions. The MLEM algorithm is shown to give in practice better results than analytical methods when the uniqueness conditions are not verified.

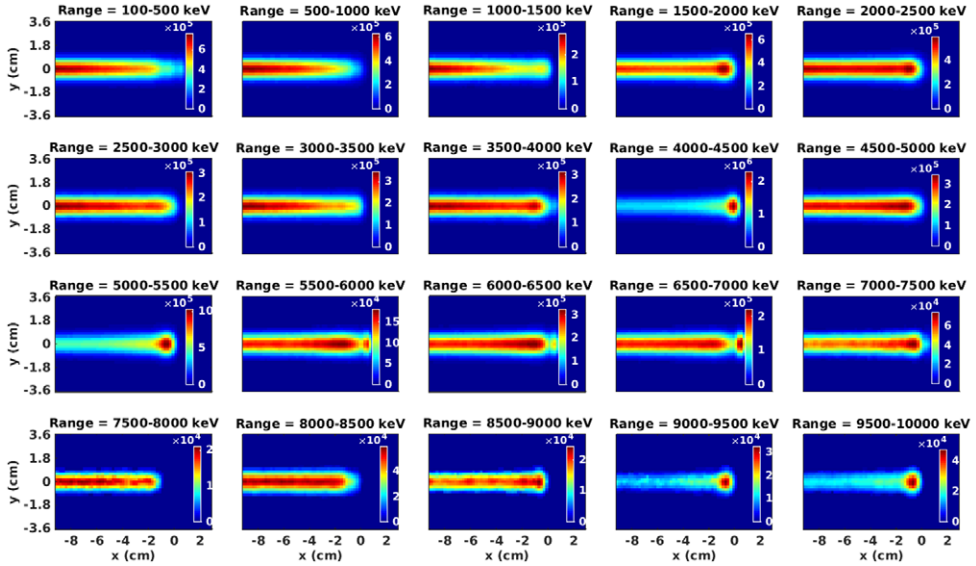


Figure 7. 2D slices from the 3D distributions $\lambda(\cdot, \varepsilon_k)$ of PG rays produced by the irradiation of a water phantom with a proton pencil beam of energy 120 MeV. Energy windows between 100 keV and 10 MeV sampled with a 500 keV step. Central slices at $z = 0$ cm of width 2 mm.

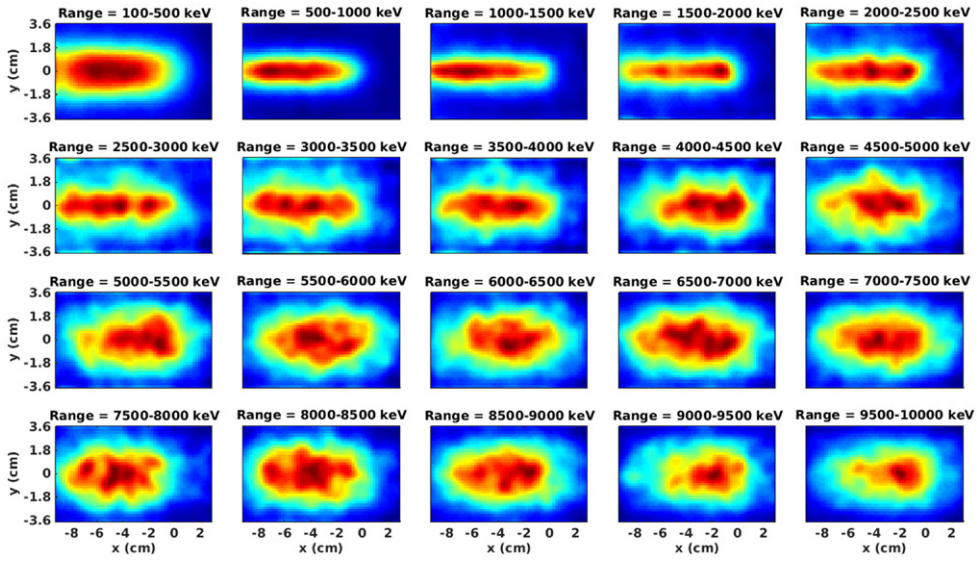


Figure 8. Reconstructed images of the PG distribution $\widehat{\lambda}(\cdot, \varepsilon_k)$. See figure 7 for the reference distributions.

It may be observed that the accuracy of the reconstructed images decreases as the energy increases. An explanation is that at relatively high energies, the photons undergo numerous interactions in the camera and also are more likely to be incompletely absorbed (see figure 10). The errors on the estimated Compton angle and on the reconstructed sequence of

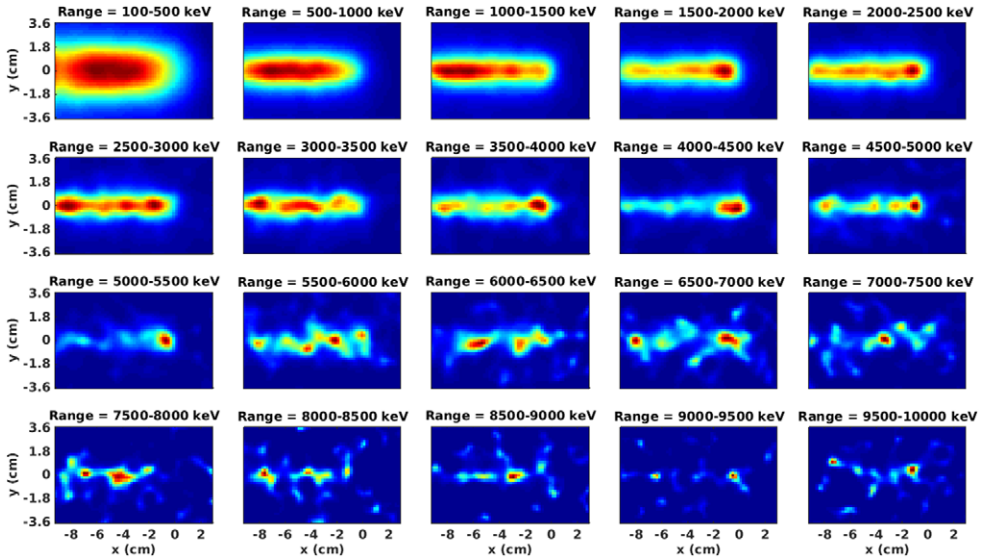


Figure 9. Reconstructed images of the PG distribution $\widehat{\lambda}(\cdot, \varepsilon_k)$ after selection of the events having deposited a total energy E_{tot} verifying $|E_{\text{tot}} - E_0| \leqslant 250$ keV. See figures 7 and 8 for a comparison.

hits lead to the blurring of the reconstructed distribution. Although the range of the beam is still roughly identifiable, the sharpness of the PG distribution is lost at the higher energies.

Some improvement may be obtained by imposing an energy selection to the detected events, i.e. selecting events whose total deposited energy belongs to a given window. Figure 9 shows slices reconstructed from only events for which the absolute error on the total energy $|E_{\text{tot}} - E_0|$ is below $\Delta\varepsilon/2 = 250$ keV.

It may be seen from figures 8 and 9 that in the window [0.1, 2] MeV, the improvement produced by the energy selection is incremental. The differences are more striking above 2.5 MeV and the energy selection largely benefits to the peaked distributions from the bins [4, 4.5] MeV and [5, 5.5] MeV.

The evolution of the quality of the images with the energy of the γ photons may be more easily observed on the profiles plotted in figure 11. The profiles are calculated as projections from the central slice on 1 cm around the line $y = 0$ cm. This lateral integration does not cover the full 2D distribution but the chosen interval includes the relevant information around the axis of the beam and partly rule out the blur. The influence of the energy selection (dashed blue lines) is obvious above about 2 MeV. Up to 6 MeV the profiles reconstructed after energy selection are very close to the PG profile, especially near to the fall-off where they are almost confounded. Without energy selection, the reconstructed profiles are close to the simulated profile only up to about 2 MeV.

For energies above 6 MeV, the number of available events drops under $1.5 \cdot 10^3$ and the results are hardly exploitable whatever the energy selection parameter. Indeed, the modeled Compton camera does not correctly detect high energy photons. As shown in figure 12(a), the number of events selected as valid by the reconstruction algorithm decreases with the energy. Also, at the higher energies, the number of hits per incident photon increases. E.g. at 4.5 MeV, the photons interact with energy deposits above 20 keV, in average 2.3 times in the absorber and 5.8 times in the scatterer. In addition, when looking at figure 12(b),

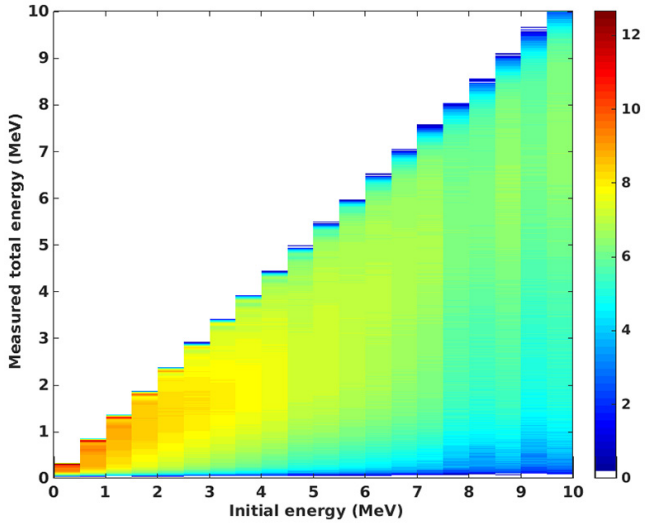


Figure 10. Distribution of the total measured energy as a function of the initial energy (logarithmic scale). The distribution of the initial energy is the one plotted in figure 3.

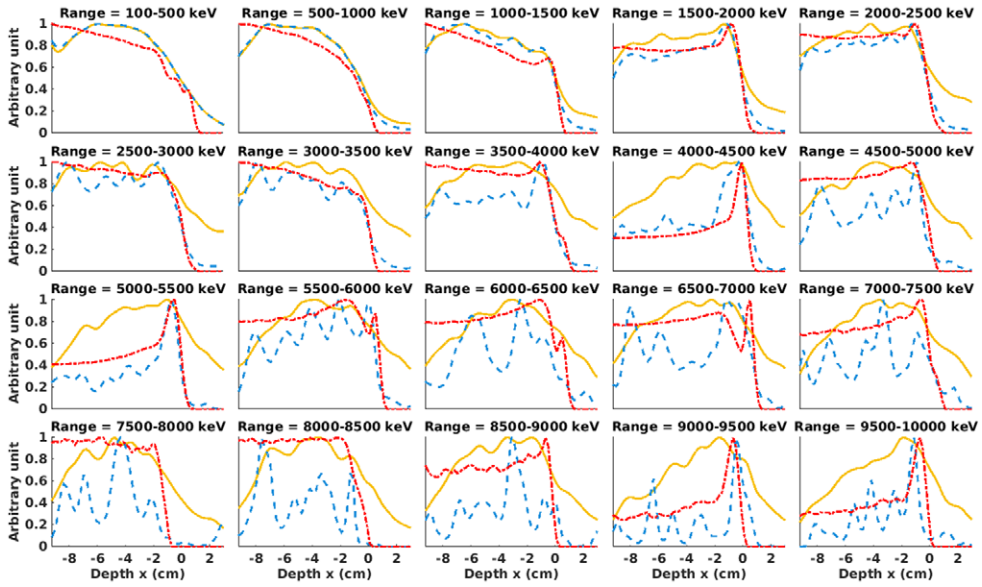


Figure 11. Overlay of three PG profiles: reference (dotted red line), reconstruction from events with total energy E_{tot} verifying $|E_{\text{tot}} - E_0| \leqslant 250$ keV (dashed blue line) and reconstruction from all the events (solid yellow line), for energy windows between 100 keV and 10 MeV with a 500 keV step. Projections calculated from the central slice on 1 cm around the line $y = 0$ cm.

it appears that the proportion of completely absorbed events (with a tolerance of ± 250 keV) among all the valid events decreases quickly with the energy. This means that more the energy increases, greater the number of noisy events becomes. The reconstruction of a photon path in the camera becomes complex, the Compton angle is overestimated and the

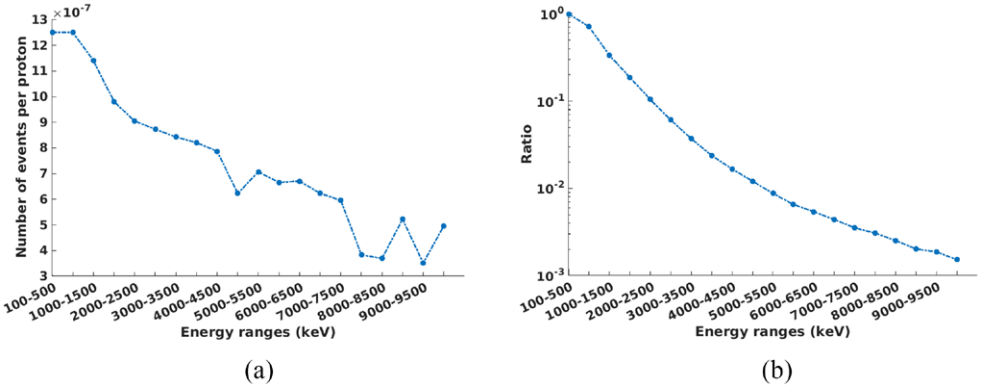


Figure 12. (a) Number of events selected as valid by the reconstruction algorithm without energy selection, per emitted proton. (b) Proportion of events selected as completely absorbed with $|E_{\text{tot}} - E_0| \leqslant 250$ keV, among all valid events.

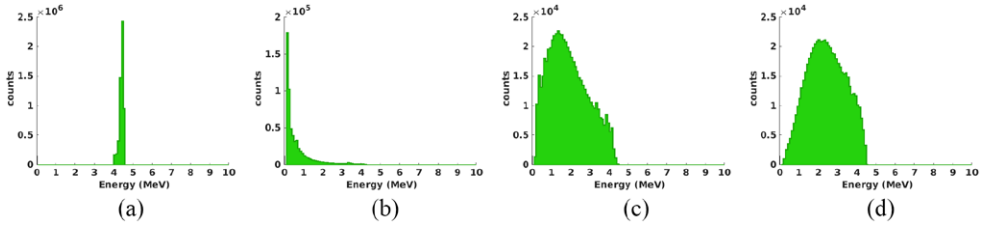


Figure 13. Energy spectra for events detected with initial energy between 4 MeV and 4.5 MeV. Counts obtained for 4×10^{11} emitted protons. (a) Initial energy spectrum. (b) Distribution of energies transferred to electrons (first hit). (c) Energy spectrum of the scattered photons. (d) Energy spectrum of coincident events.

images are less accurate. Random coincidences and detector load were not accounted for in the study. It may be mentioned that the observed mean number of hits per emitted proton was $4.65 \cdot 10^{-5}$ for the scatterer block and $2.33 \cdot 10^{-5}$ for the entire segmented absorber block.

3.3. Influence of the measured total energy selection on the reconstructed images

The limited capability of the camera to provide the initial energy of the incoming photons makes impossible to observe the spatial distribution at some given initial energy. As an example, in figures 13 and 14, we observe the spectra of initial and measured energies from events with energy in the windows [4, 4.5] MeV and [0.1, 10] MeV. The distribution from figure 14(a) is the same as in figure 3 and the figure 13(d) is the plot of a column from figure 3. In each figure, panel (a) represents the initial energy spectrum. Panel (b) display the distribution of the energy transferred to the electron at the first hit. Panel (c) shows the distribution of E_2, the energy deposited in the camera by the scattered photon. The last panel (d) represents the measured sum energy spectrum of coincident events in the camera.

The images shown on the upper row of figure 15 represent slices of the simulated PG spatial distribution for some selected initial energy windows. The windows are R_{tot}, which contains all the events, $R_1 = [0.5, 2.5]$ MeV, $R_2 = [4, 4.7]$ MeV and $R_3 = [4.7, 5.5]$ MeV.

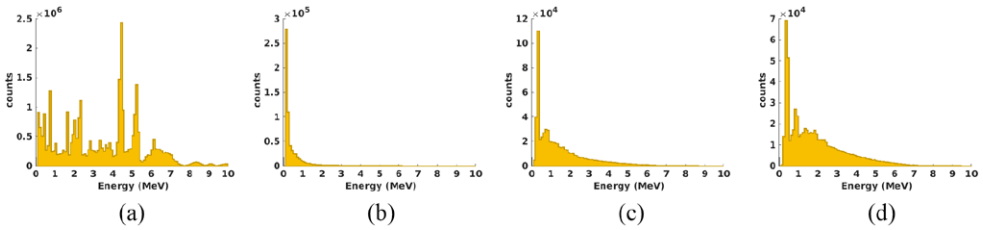


Figure 14. Energy spectra for events detected with initial energy between 100 keV and 10 MeV. Counts obtained for 4×10^{11} emitted protons. (a) Initial energy spectrum. (b) Distribution of energies transferred to electrons (first hit). (c) Energy spectrum of the scattered photons. (d) Energy spectrum of coincident events.

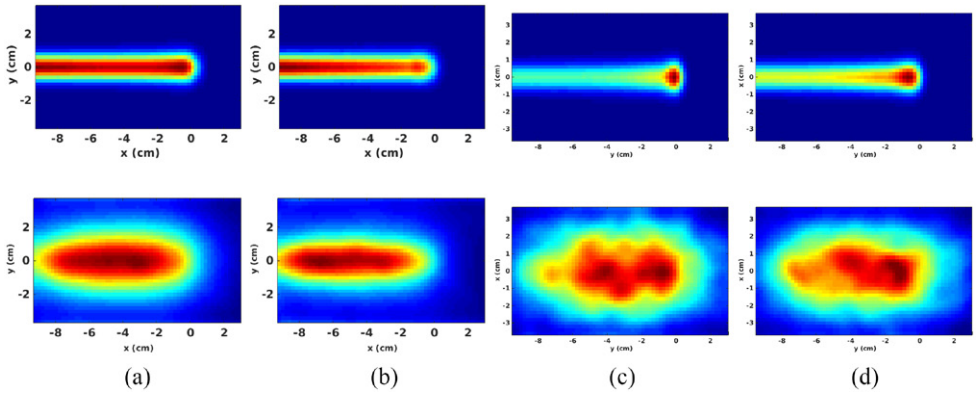
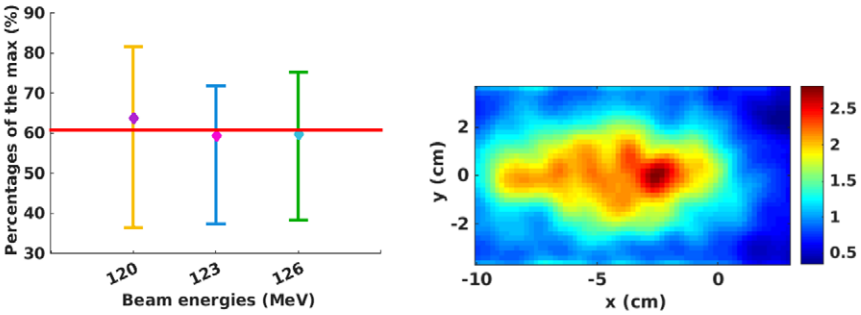


Figure 15. PG distributions for different energy windows. Upper line: simulated reference distributions. Bottom line: reconstructed distributions. (a) All events. (b) Events with total energy in $R_1 = [0.5, 2.5]$ MeV. (c) Events with total energy in $R_2 = [4, 4.7]$ MeV. (d) Events with total energy in $R_3 = [4.7, 5.5]$ MeV.

These reference distributions obtained from simulations serve for comparison for the reconstructed images shown on the bottom row. The reconstructed images are produced by selecting from the data only the photons having deposited in the camera an energy E_{tot} belonging to the specified range. The slices are obtained from volumes whose dimensions and number of voxels are the same as in section 3.2. On the first column, the distributions λ_{tot} and $\widehat{\lambda}_{\text{tot}}$ include PG of all energies and the reconstructed image is calculated from 4.85×10^6 events in 15 iterations. The number of events available for windows R_1, R_2 and R_3 was respectively 2×10^6, 1.58×10^5 and 1.1×10^5 events. The first energy window R_1 was chosen as it contains energies well suited for the modeled camera. Moreover, the resulting image and profile show low variance. The two other windows were chosen as corresponding to salient peaks in the energy spectrum of the PG emission, and to spatial distributions relatively well reproduced in figures 8 and 11.

3.4. Detection of shifts

The aim of this section is to evaluate the ability of the Compton camera to identify range shifts. As explained in sections 2.3 and 2.4, we first need to calculate fall-off positions, that were defined as functions of two parameters p_{min} and p_{max}.



(a) (b)

Figure 16. (a) Percentages of the maximum value of the profile, corresponding to the simulated PG position, calculated over 100 samples (10^4 events each and 15 iterations) for each of the 3 beam energies. Diamonds represent median values and the vertical lines connect the minimum and maximum values. (b) Exemple of image reconstructed from 10^4 events and 15 iterations, without energy selection. Figure 15(a) shows the image of the same distribution reconstructed from 4.85×10^6 events.

3.4.1. Variability of the best percentage value. For each beam energy, we reconstructed 100 profiles from 10^4 events each, without any energy selection. For each profile f_k obtained from the samples, we evaluated the p_k percentage of the maximum M_k of the curve, corresponding to the simulated fall-off d, as:

$$p_k = \frac{f_k(d)}{M_k} \times 100\%. \tag{9}$$

Then, we calculate the maximum, the minimum and the median of the values p_k for $k = 1, \ldots, 100$ and the 3 beam energies. The results are plotted in figure 16(a). The median values over the samples are pictured by the diamonds on the line segments. The extremities of the line segments correspond to the extreme values. The central red line $y = 60.81\%$ represents the median over all the values whatever the beam energy.

It may be seen from figure 16(a) that the median values are close one to another for the three considered energies. This suggest that the median percentage calculated for one given energy may be extrapolated for other energies.

Note that for the efficiency of the considered camera, 10^4 detected photons correspond to about 10^9 emitted protons, i.e. to 10 of the stronger spots. The mean values over the 100 samples of the errors $|\widehat{d}_{p_{\min},p_{\max}} - d|$ with $p_{\min} = p_{\max} = 0.61$, are for the three considered energies of 2.2, 2.9 and 3.3 mm. The standard deviations are 3.1, 5 and 5.8 mm. In clinical conditions, we may thus expect an average error of 3 mm on the estimated fall-off, but the results present a large variance.

3.4.2. Estimation of the shift for the statistically learned best percentage. The statistical study from the previous section 3.4.1 gives a median best percentage value. We applied the method described in section 2.3 to estimate the PG deduced proton range for the total reconstructed profiles, i.e. profiles reconstructed without any energy selection, for each of the three beam energies. The profiles are shown in figure 17. We choose $p_{\max} = p_{\min} = 60.81\%$, the median over the values p_k calculated for the 100 profiles and three energies. We can see from table 1 the PG deduced proton range is very well estimated with a sub-millimetric error.

Table 1. Error on the estimated PG deduced proton range, in millimeters.

Beam energy	120 MeV	123 MeV	126 MeV
PG deduced proton range (mm)	100	105	110
Error (mm)	0.2	0.5	0.9

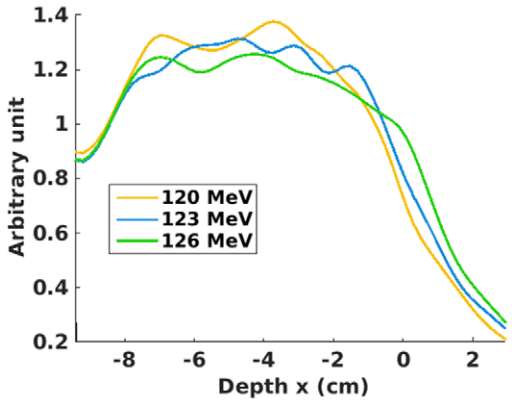


Figure 17. Overlay of profiles reconstructed from all events.

However, we can apply this method only to profiles from a set-up identical to the one for which the learning was done. The best percentage calculated for profiles reconstructed from all data cannot be extrapolated to profiles calculated following energy selection.

3.4.3. A posteriori error. For the three energies of the proton beam, overlays of reference and reconstructed profiles are shown in figure 18. The slope of the reconstructed profiles near the Bragg peak is less sharp than for the profiles obtained by energy selection and shown in figure 11. For all beam energies, the total reconstructed profile is closer to the R_1 profile and likewise, the R_2 and R_3 curves are very similar.

Figure 18 shows that the 80% criterion used to calculate the Bragg peak location can not be applied to the reconstructed profiles. To estimate the PG deduced proton range from the curves in figure 18, we calculate the estimated fall-off as a mean over a range of percentages with the method described in section 2.3.

In order to account for the observed variability, the values for $p_{\min}$ and $p_{\max}$ have to be set individually for each type of profile (namely total, R_1, R_2, R_3). For each given profile we define the error function F of variables $(p_{\min}, p_{\max}) \in [0.2, 1]^2$, $p_{\min} \leqslant p_{\max}$ by the relation:

$$F(p_{\min}, p_{\max}) = |\widehat{d}_{p_{\min}, p_{\max}} - d|, \tag{10}$$

where $\widehat{d}_{p_{\min}, p_{\max}}$ is the estimated position of the fall-off given by (8) and d is the reference value obtained from simulations. The mean over the three beam energies of the error functions is shown in figure 19 for the four considered PG energy window.

The minimum value of each function F gives the combinations $(p_{\min}, p_{\max})$ that minimize the error on the estimated PG deduced proton range. However, a best combination is curve-specific. For each PG energy window we dispose of three curves, corresponding to the three beam energies. In what follows we apply a leave-one-out approach to find a best combination $(p_{\min}, p_{\max})$ and then to calculate the errors on the estimated PG deduced proton ranges shown in table 2. E.g. when the reference values of the fall-off position are considered as known

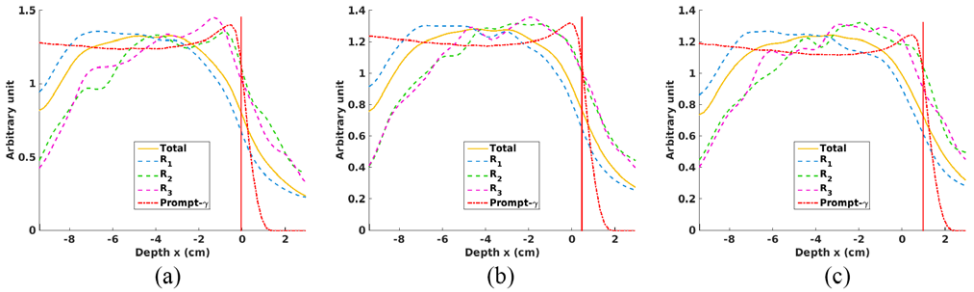


Figure 18. Overlay of the PG profile (dotted red line) and the reconstructed profiles calculated as projection of the central slice on 1 cm around the line $y = 0$ cm. The vertical red line stands for the PG fall-off. (a) [120 MeV] (b) [123 MeV] (c) [126 MeV].

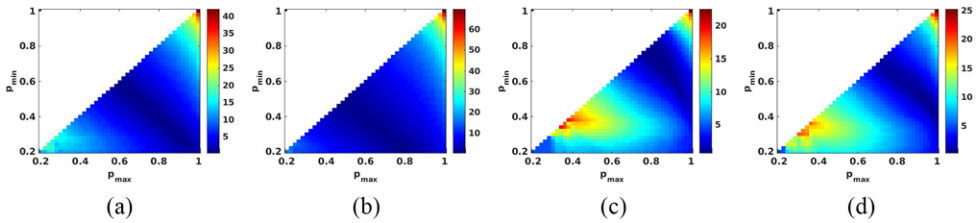


Figure 19. Mean over the three beam energies of the error functions F for four energy windows. (a) Total. (b) R_1. (c) R_2. (d) R_3.

Table 2. Error on the estimated PG deduced proton range, in millimeters.

Beam energy	120 MeV				123 MeV				126 MeV			
PG range (mm)	100				105				110			
Energy window	Total	R_1	R_2	R_3	Total	R_1	R_2	R_3	Total	R_1	R_2	R_3
Error (mm)	0.9	0.6	2.8	1.6	1	0.7	−0.7	1.5	−0.6	−0.6	1.3	0

Note: The error is calculated as the difference $\widehat{d}_{p_{\min},p_{\max}} - d$.

for the beam energies 120 and 123 MeV, for each energy window, two error functions can be calculated. The argument of the minimum of their mean is set as the best combination of percentages. The errors on the estimated position of the fall-off are then calculated for the test energy 126 MeV and are shown in the last four columns of table 2.

The three beam energies we chose, namely 120, 123 and 126 MeV, provide about 5 mm shifts in the position of the Bragg peak. The shifts in the PG profiles observed from the reconstructed images are compared to the theoretical values in table 3. The shifts were calculated for the test samples, i.e. for each beam energy the coordinates of the PG profile fall-off is evaluated with the percentages estimated from the two other energies.

4. Discussions and conclusions

The distribution of the prompt-gamma emission during irradiation in proton therapy is closely related to the deposited dose. The goal of this work was double. On one hand, we investigate the ability of a specific Compton camera and a proposed reconstruction algorithm to provide an image of the distribution of the PG emission with and without energy selection. On the

Table 3. Shifts measured from the reconstructed profiles for different beam energies.

Beam energies increase	120–123 MeV				123–126 MeV			
Reference PG shift (mm)	5				5			
Energy window	Total	R_1	R_2	R_3	Total	R_1	R_2	R_3
Estimated shift (mm)	5.3	5.2	1.6	5	3.5	3.9	7	3.6

other hand, we proposed a method for the estimation of the PG fall-off and we showed that it allows to detect shifts with millimetre accuracy when some prior experience is available.

The spatial distribution of the photons varies with the energy, and some energies were previously shown to characterize the end-of-range of the beam. Imaging of the PG emission for specific energy windows could thus give important information on the treatment. We show that when the initial energy of the photons is known, the reconstructed images relatively well reproduce the simulated distributions up to 5.5 MeV. However, as the PG initial energy can not be sorted, a selection window can be applied only to the total energy deposited by each individual photon. As a consequence, for each selection, the reconstructed image is contaminated with noise caused by incomplete absorption of higher energy photons.

The performances of the Compton camera simulated for this work are better for the window $R_1 = [0.5, 2.5]$ MeV. Compared to other windows, the R_1 window is more advantageous since it benefits from a larger statistics and presents a lower variance. Moreover, the selection on energies below 2.5 MeV leads to an image similar to the one of the total distribution. The available number of events is still in the same order of magnitude. Nevertheless, this energy window is in practice the most affected by noise from neutrons.

The depth of the distal edge of the PG distribution was calculated from the reconstructed images as the coordinate of some percentage of the maximum of the projected profile. For the total distribution, the percentage was previously learned on a set of 100 samples of low-statistics reconstructed images. We observed mild influence of the beam energy on the median percentage. With this approach, we observed less than 1 mm offset between the depths of the fall-off for the reconstructed and simulated profiles. We also evaluated the errors, for four energy windows, when only two samples are available in the learning set. In this case the maximum error we found was of about 3 mm. We also noticed that the smoother the distribution, better the estimation of the fall-off is. With the considered camera, the best values were found for the total and R_1 energy windows. Theoretical shifts of 5 mm in the range of the beam were then estimated with an error lower than 1 mm for the R_1 window. For the higher energy windows the error attained at its maximum 3.4 mm. In the case of a real treatment, the PG profile would present more variability. The machine learning approach proposed in Gueth *et al* (2013) would be then better suited and could be carried on data simulated from the treatment plan.

All the typical noise sources in PG imaging were not taken into account in this work. The attenuation in the emitting volume was not discussed here because in the described context it had no significant influence on the reconstructed images. Random coincidences were also avoided in our simulations by recording the interactions of each particle separately. The noise coming from neutrons was ignored here, but it was quantified as very important at low energies in other works. Consequently, the Compton camera would also detect neutrons in the energy window R_1 and the results we obtained for this window are likely to be affected. Exploitation of high energy photons may be a solution but requires an enhancement of the detection efficiency compared to our set-up. In clinical practice, importance should be accorded to the ability of a detector to accurately measure the initial energy of the photons.

Discrepancies with the treatment plan also arise from volume composition changes due to misplacements of the patient, anatomical evolution or wrong estimation of the stopping power

of the tissues. Since the emission spectrum is tissue dependent, the precise estimation of the joint spatial and energy distribution is an important issue in protontherapy monitoring.

Acknowledgment

This work was partly supported by the ENVISION project (co-funded by the European commission under the FP7 Collaborative Projects Grant Agreement Nr. 241851FP7), by ETOILE's Research Program (PRRH/UCBL, under CPER 2007-13 funding) and by the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon. The authors would like to thank the anonymous reviewers for careful reading of the manuscript and thoughtful comments.

References

Bom V, Joulaeizadeh L and Beekman F 2012 Real-time prompt gamma monitoring in spot-scanning proton therapy using imaging through a knife-edge-shaped slit *Phys. Med. Biol.* **57** 297

Cambraia Lopes P C, Pinto M, Simões H, Biegun A, Dendooven P, Oxley D, Parodi K, Schaart D and Crespo P 2012 Optimization of collimator designs for real-time proton range verification by measuring prompt gamma rays *IEEE Nuclear Science Symp. and Medical Imaging Conf.* pp 3864–70

Compton A H 1923 A quantum theory of the scattering of x-rays by light elements *Phys. Rev.* **21** 483

Frandes M, Zoglauer A, Maxim V and Prost R 2010 A tracking Compton-scattering imaging system for hadrontherapy monitoring *IEEE Trans. Nucl. Sci.* **57** 144–50

Gillam J E, Lacasta C, Torres-Espallardo I, Juan C C, Llosá G, Solevi P, Barrio J and Rafecas M 2011 A Compton imaging algorithm for on-line monitoring in hadrontherapy *SPIE Medical Imaging 2011: Physics of Medical Imaging Conf.* vol 7961

Golnik C *et al* 2014 Range assessment in particle therapy based on prompt γ-ray timing measurements *Phys. Med. Biol.* **59** 5399

Gueth P, Dauvergne D, Freud N, Létang J, Ray C, Testa E and Sarrut D 2013 Machine learning-based patient specific prompt-gamma dose monitoring in proton therapy *Phys. Med. Biol.* **58** 4563

Janssen F, Landry G, Lopes P C, Dedes G, Smeets J, Schaart D, Parodi K and Verhaegen F 2014 Factors influencing the accuracy of beam range estimation in proton therapy using prompt gamma emission *Phys. Med. Biol.* **59** 4427–41

Jan S *et al* 2011 GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy *Phys. Med. Biol.* **56** 811–901

Kanawati W E, Létang J, Dauvergne D, Pinto M, Sarrut D, Testa E and Freud N 2015 Monte Carlo simulation of prompt γ-ray emission in proton therapy using a specific track length estimator *Phys. Med. Biol.* **60** 8067–86

Kim D, Yim H and Kim J W 2009 Pinhole camera measurements of prompt γ-rays for detection of beam range variation in proton therapy *J. Korean Phys. Soc.* **55** 1673–6

Kormoll T, Fiedler F, Schöne S, Wüstemann J, Zuber K and Enghardt W 2011 A Compton imager for *in vivo* dosimetry of proton beams—a design study *Nucl. Instrum. Methods Phys. Res.* A **626** 114–9

Krimmer J *et al* 2014 Development of a Compton camera for medical applications based on silicon strip and scintillation detectors *Nucl. Instrum. Methods Phys. Res.* A **787** 98–101

Kurosawa S *et al* 2012 Prompt gamma detection for range verification in proton therapy *Curr. Appl. Phys.* **12** 364–8

Lee H R *et al* 2012 Design optimization of a 2D prompt-gamma measurement system for proton dose verification *J. Korean Phys. Soc.* **61** 239–42

Lojacono X, Richard M H, Ley J L, Testa E, Ray C, Freud N, Létang J M, Dauvergne D, Maxim V and Prost R 2013 Low statistics reconstruction of the Compton camera point spread function in 3D prompt-γ imaging of ion beam therapy *IEEE Trans. Nucl. Sci.* **60** 3355–63

Mackin D, Peterson S, Bedar S and Polf J 2012 Evaluation of a stochastic reconstruction algorithm for use in Compton camera imaging and beam range verification from secondary gamma emission during proton therapy *Phys. Med. Biol.* **57** 3537–53

Maxim V, Lojacono X, Hilaire E, Krimmer J, Testa E, Dauvergne D, Magnin I and Prost R 2016 Probabilistic models and numerical calculation of system matrix and sensitivity in list-mode MLEM 3D reconstruction of Compton camera images *Phys. Med. Biol.* **61** 243–64

Maxim V 2014 Filtered backprojection reconstruction and redundancy in Compton camera imaging *IEEE Trans. Image Process.* **23** 332–41
Min C H, Kim C H, Youn M Y and Kim J W 2006 Prompt gamma measurements for locating the dose falloff region in the proton therapy *Appl. Phys. Lett.* **89** 183517
Moteabbed M, España S and Paganetti H 2011 Monte Carlo patient study on the comparison of prompt gamma and PET imaging for range verification in proton therapy *Phys. Med. Biol.* **56** 1063–1082
Ordonez C E, Bolozdynya A and Chang W 1997 Dependence of angular uncertainties on the energy resolution of Compton cameras *IEEE Nuclear Science Symp.* vol 2
Ortega P G *et al* 2015 Noise evaluation of Compton camera imaging for proton therapy *Phys. Med. Biol.* **60** 1845–63
Perali I *et al* 2014 Prompt gamma imaging of proton pencil beams at clinical dose rate *Phys. Med. Biol.* **59** 5849–71
Pinto M, Dauvergne D, Freud N, Krimmer J, Létang J M, Ray C, Roellinghoff F and Testa E 2014 Design optimization of a TOF-based collimated camera prototype for online hadrontherapy monitoring *Phys. Med. Biol.* **59** 7653–74
Polf J, Panthi R, Mackin D, McCleskey M, Saastamoinen A, Roeder B and Beddar S 2013 Measurement of characteristic prompt gamma rays emitted from oxygen and carbon in tissue-equivalent samples during proton beam irradiation *Phys. Med. Biol.* **58** 5821–31
Polf J, Peterson S, Ciangaru G, Gillin M and Beddar S 2009 Prompt gamma-ray emission from biological tissues during proton irradiation: a preliminary study *Phys. Med. Biol.* **54** 731–43
Richard M H *et al* 2012 Design study of the absorber detector of a Compton camera for online control in ion beam therapy *IEEE Trans. Nucl. Sci.* **59** 3496–500
Roemer K *et al* 2015 Characterization of scintillator crystals for usage as prompt gamma monitors in particle therapy *J. Instrum.* **10** P10033
Sarrut D *et al* 2014 A review of the use and potential of the GATE Monte Carlo simulation code for radiation therapy and dosimetry applications *Med. Phys.* **41** 064301
Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imaging* **1** 113–22
Smeets J *et al* 2012 Prompt gamma imaging with a slit camera for real-time range control in proton therapy *Phys. Med. Biol.* **57** 3371
Smith B 2005 Reconstruction methods and completeness conditions for two Compton data models *J. Opt. Soc. Am.* A **22** 445–59
Stichelbaut F and Jongen Y 2003 Verification of the proton beam position in the patient by the detection of prompt gamma-rays emission *39th PTCOG Meeting (San Francisco)*
Styczynski J, Riley K, Binns P, Bortfeld T and Paganetti H 2009 Su-dd-a3-03: can prompt gamma emission during proton therapy provide *in situ* range verification? *Med. Phys.* **36** 2425
Testa E, Bajard M, Chevallier M, Dauvergne D, Foulher F L, Freud N, Létang J M, Poizat J, Ray C and Testa M 2009 Dose profile monitoring with carbon ions by means of prompt-gamma measurements *Nucl. Instrum. Methods Phys. Res.* B **267** 993–6
Tomitani T and Hirasawa M 2002 Image reconstruction from limited angle Compton camera data *Phys. Med. Biol.* **47** 2129–45
Verburg J M, Riley K, Bortfeld T and Seco J 2013 Energy- and time-resolved detection of prompt gamma-rays for proton range verification *Phys. Med. Biol.* **58** L37–49
Verburg J M and Seco J 2014 Proton range verification through prompt gamma-ray spectroscopy *Phys. Med. Biol.* **59** 7089
Wilderman S, Fessler J, Clinthorne N, LeBlanc J and Rogers W 2001 Improved modeling of system response in list mode EM reconstruction of Compton scatter camera images *IEEE Trans. Nucl. Sci.* **48** 111–6
Wilderman S J, Clinthorne N H, Fessler J A and Rogers W L 1998 List-mode maximum likelihood reconstruction of Compton scatter camera images in nuclear medicine *IEEE Nuclear Science Symp. Conf. Record* vol 3 pp 1716–20
Zhang B and Zeng G L 2007 Two-dimensional iterative region-of-interest (ROI) reconstruction from truncated projection data *Med. Phys.* **34** 935–44
Zoglauer A, Andritschke R and Schopper F 2006 MEGAlib—the medium energy gamma-ray astronomy library *New Astron. Rev.* **50** 629–32
Zoglauer A, Weidenspointner G, Galloway M, Boggs S and Wunderer C 2009 Cosima—the cosmic simulator of MEGAlib *Nuclear Science Symp. Conf. Record* pp 2053–9

TOMOGRAPHIC RECONSTRUCTION FROM POISSON DISTRIBUTED DATA: A FAST AND CONVERGENT EM-TV DUAL APPROACH

Voichiţa Maxim*, Yuemeng Feng, Hussein Banjak

Univ Lyon, INSALyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F69621, Villeurbanne, France

Elie Bretin

Univ Lyon, INSA de Lyon, CNRS UMR 5208, Institut Camille Jordan, 20 avenue Albert Einstein, F-69621 Villeurbanne Cedex, France

Abstract. This paper focuses on tomographic reconstruction for nuclear medicine imaging, where the classical approach consists to maximize the likelihood of Poisson distributed data using the iterative Expectation Maximization algorithm. In this context and when the quantity of acquired data is low and produces low signal-to-noise ratio in the images, a step forward consists to incorporate a total variation prior on the solution into a MAP-EM formulation. The novelty of this paper is to propose a convergent and efficient numerical scheme to compute the MAP-EM optimizer, based on a splitting approach which alternates an EM step and a dual-TV-minimization step. The main theoretical result is the proof of stability and convergence of this scheme. We also present some numerical experiments where we compare the proposed algorithm with some other algorithms from the literature.

1. **Introduction.** Emission tomography is based on detecting radiation emitted from the patient, enabling clinicians to identify the position and size of tumours, to control the quality of a treatment or to conduct diagnostic procedures such as coronary perfusion. Important issues relating to the invasive character of this type of imaging is the reduction of the tracer dosage and acquisition time. As a consequence, the clinical emission data uses to be strongly affected by Poisson noise.

Tomographic reconstruction is an inverse problem that may be solved by analytic or iterative algorithms. Analytic algorithms, based on the inversion of the Radon transform, are fast but produce noisy images. Iterative ones may integrate knowledge on the detection process, on the data noise and on the expected image. For instance, the positivity constraint is naturally ensured by some of these algorithms. The most widespread iterative algorithm for nuclear medicine tomography is the Maximum Likelihood Expectation Maximization (MLEM) algorithm ([15, 32]) with its variants from which we may cite for instance the ordered subsets EM (OSEM [12]) and the row action maximum likelihood algorithm (RAMLA [5]). This class

2010 *Mathematics Subject Classification.* 44A12, 65K10, 92C55, 49M29.

Key words and phrases. Image reconstruction, total variation, Poisson noise, tomography, MAP-EM, MLEM, dual approach.

The first author is supported by .

* Corresponding author: Voichiţa Maxim.

of algorithms aim to determine the image of mean intensities most likely to have produced the observed data. In the family of bayesian methods, the Stochastic Origin Ensemble (SOE) algorithm was more recently proposed by A. Sitek ([28]). For each voxel from the volume, SOE produces the Monte-Carlo estimation of the posterior distribution while MLEM produces only its mean.

When compared to algebraic algorithms, MLEM shows faster convergence and better robustness to the outliers. According to [14], this is explained by the fact that MLEM can be seen as a preconditioned gradient method. On the other hand, it tends to produce areas of intensity accumulation in regions where the intensity should be uniform. As emphasised *e.g.*, in [30, 27], this result is not related to the algorithm, but rather to the maximum likelihood solution itself. In the reconstructed image, the signal is thus embedded in the noise and at least it is necessary to interrupt the iterations well before convergence and to smooth the result (see *e.g.*, [33]). Early stopping of the iterations and post-smoothing reduces variance of the noise but also blurs the edges and may mask small sources. Moreover, it is known that early stopping of the iterations conducts to biased images. Finally, stopping the iterations does not ensure that the reconstructed image will have the desired properties from the point of view of the user's a priori information, and post-smoothing is likely to reduce concordance between solution and data. Snyder and Miller ([29]) constrain the solution to belong to the space of smooth functions, or "sieves". Each element of this space is the convolution of a function with a Gaussian kernel. They show how the MLEM iterations should be adapted to directly produce such a solution. The method is however quite sensible to the choice of the kernel (see [31]).

A smoothing step can be inserted between each two successive iterations of the algorithm ([27, 10]). The choice of the smoothing kernel is crucial, as the variance in the MLEM images is spatially not uniform and depends on the intensity. Using an uniform smoothing kernel, or even the usual total variation denoising ([23]), will result in over-smoothing in the low-intensity and under-smoothing in high-intensity regions. Moreover, it is not always easy to analyse the convergence properties of such an algorithm. In [27], the authors have succeeded in linking their algorithm to a penalized log-likelihood maximization by giving up the usual spatial invariance of the criterion.

A variational point of view is to include into the reconstruction process some a priori information that will ensure best compromise between smoothness and data fidelity, in the sense of the Maximum A Posteriori (MAP). As shown in [11], a small modification of the MLEM iteration formula allows to calculate the MAP estimate for smooth priors. In the case of total variation prior, a smooth approximation can be used ([21, 22]). In practice, this method gives unstable results and is restricted to small values of the regularization parameter where it is constructed as an explicit numerical scheme.

Shift-invariant priors give unsatisfactory results in terms of bias and resolution because the likelihood is shift-variant. Shift-variant quadratic priors were proposed *e.g.*, in [9] and [19]. Priors based on the total variation norm have been proposed ([16, 21, 22, 24, 35, 1, 25, 18]). They require advanced algorithms due to the complexity and non-differentiability of the total variation norm. However, the use of the total variation prior is appealing when one attempts to obtain images with smooth regions separated by sharp edges. The resolution of linear inverse problems with total variation regularization was tackled in the context of a Gaussian noise and

ℓ_2-norm (see *e.g.*, [26, 4]) but is still an issue when the data is Poisson distributed and thus a Kullback-Leibler distance is better suited. In this work we address the last topic, and we propose a solution in which the classical MLEM algorithm is used in conjunction with a custom-designed total-variation smoothing step.

1.1. **Principle of MLEM algorithm.** The volume containing the source of gamma photons is divided in J spatial locations called pixels or voxels, indexed on $j = 1, \ldots, J$. The emission counts at the various locations are independent and the value x_j of the j^{th} pixel is a realization of a Poisson random variable with mean value λ_j. The values $\boldsymbol{x}$ cannot be observed directly with a detector placed outside the object. For a given pixel j and a given detector element i, t_{ij} is the probability that a photon emitted at pixel j will be detected at detector i. The system matrix of the detection system (also called transition matrix) is then $T = (t_{ij})$. The element s_j, where $j = 1, \ldots, J$, of the sensitivity vector $\boldsymbol{s}$ is defined as the probability of a photon emitted at pixel j to be detected somewhere in the camera. For each pixel j, we obviously have the relation:

$$s_j = \sum_{i=1}^{I} t_{ij}. \tag{1}$$

For $i = 1, \ldots, I$, the numbers y_i of events detected by the detector elements also follow Poisson distributions with mean values μ_i and the observations are independent. The vector $\boldsymbol{y}$ of detected events approximately matches $T\boldsymbol{x}$, more precisely the means verify the equation

$$\boldsymbol{\mu} = T\boldsymbol{\lambda}. \tag{2}$$

In [8] the non observable variable $\boldsymbol{x}$ is called complete data and the observed $\boldsymbol{y}$ are called incomplete data. The aim of the MLEM algorithm is to estimate the means $\boldsymbol{\lambda}$ of the complete data from the observed values $\boldsymbol{y}$ of the incomplete data.

Given the observations $\boldsymbol{y} = (y_i)_{i=1,\ldots,I}$, an estimator of the vector $\boldsymbol{\lambda} = (\lambda_j)_{j=1,\ldots,J}$ is the solution of the maximum of the log-likelihood function

$$\ell(\boldsymbol{\lambda}|\boldsymbol{y}) = -\sum_{i=1}^{I}\sum_{j=1}^{J} t_{ij}\lambda_j + \sum_{i=1}^{I} y_i \ln\left(\sum_{j=1}^{J} t_{ij}\lambda_j\right) - \sum_{i=1}^{I} \ln(y_i!). \tag{3}$$

The elements of the Hessian matrix $\mathscr{H}_\ell(\boldsymbol{\lambda}|\boldsymbol{y})$ of ℓ are

$$\frac{\partial^2 \ell}{\partial\lambda_j \partial\lambda_k}(\boldsymbol{\lambda}|\boldsymbol{y}) = -\sum_{i=1}^{I} \frac{t_{ij}t_{ik}y_i}{\mu_i^2} \tag{4}$$

and for all $\boldsymbol{u} \in \mathbb{R}^J$,

$$\boldsymbol{u}^T \mathscr{H}_\ell(\boldsymbol{\lambda}|\boldsymbol{y})\boldsymbol{u} = -\sum_{i=1}^{I} \left(\frac{\sqrt{y_i}}{\mu_i}\sum_{j=1}^{J} t_{ij}u_j\right)^2 \leq 0. \tag{5}$$

The concavity of the log-likelihood implies that its maximum is global. When $\operatorname{rank}(T) = J$, the linear application of matrix T is injective and thus the Hessian is negative definite. The likelihood has then a strict maximum and an unique maximum solution. When $\operatorname{rank}(T) < J$, the Hessian is negative semi-definite and there may be multiple solutions for the maximum.

In practice, an iterative method is applied to approximate the minimum, and often the Expectation Maximization (EM) algorithm from [8] is used. One thus

obtain the maximum likelihood expectation maximization algorithm that was proposed for nuclear imaging applications in [15, 32]. Starting from an arbitrary vector $\boldsymbol{\lambda}^{(0)} \in (\mathbb{R}_+^*)^J$, the $(n+1)^{th}$ estimation of the maximum likelihood is given by:

$$\lambda_j^{(n+1)} = \lambda_j^{(n)} \frac{1}{s_j} \sum_{i=1}^{I} t_{ij} \frac{y_i}{\sum\limits_{k=1}^{J} t_{ik} \lambda_k^{(n)}}, \tag{6}$$

and the EM algorithm always decreases the value of the likelihood function. The convergence of the algorithm to the maximum of the likelihood was proved by Lange and Carson in [15] for strictly concave likelihood function and was proved in the general case by A.N. Iusem in [13].

The MLEM algorithm is both simple and efficient and thus is widely used. Its convergence speed is related to the system matrix. When the system matrix only accounts for the geometry of detection the convergence may be relatively fast and the result is noisy. For this reason the iterations must be stopped before too much high frequencies are introduced. On the contrary, when the system matrix also models some convolution introduced by *e.g.*, the detectors or the attenuation in the patient, the convergence is slow and the higher frequencies are hardly recoverable.

Introduction of a priori information helps in both cases to obtain a smooth and precise image in a reasonable time. This a priori information can be added easily in the Bayesian formalism and leads to maximum a posteriori (MAP) algorithms. When the prior is a smooth function, the proof of convergence of the algorithm naturally fits the EM developments from [8]. This is not any more the case when the prior is not smooth.

1.2. **A priori information and discrete total variation.** Total variation denoising is known to promote smoothness in images while still conserving sharp edges ([23, 6, 26, 4]). It tends to produce almost homogeneous regions separated by sharp frontiers, which makes that the images have a cartoon aspect. This type of regularization has a strong interest in particular for low dose acquisitions where the goal is to identify the shape of the objects in the volume ([22, 26, 24, 35, 1, 25]).

To facilitate the presentation and without loss of generality, all the developments hereafter are presented for the two-dimensional case. The three-dimensional case can then be easily derived and the results remain true with minor changes that we will specify when necessary.

Let Ω be an open subset of $\mathbb{R}^2$. The standard total variation is defined for functions $u \in L^1(\Omega)$ by

$$TV(u) = \sup \left\{ \int_\Omega u(x) \mathrm{div}\,(\varphi(x)) dx : \ \varphi \in C_c^1(\Omega; \mathbb{R}^2), \ |\varphi(x)| \leq 1 \ \forall x \in \Omega \right\}. \tag{7}$$

In particular, it is well known ([6]) that the functional TV is finite if and only if the distributional derivative Du of u is a finite measure on Ω. Moreover, if u has a gradient $\nabla u \in L^1(\Omega)$, then $TV(u) = \int_\Omega |\nabla u(x)| dx$. An other interesting particular case is $u = \chi_Q$ the characteristic function of a smooth set Q, when $TV(u)$ can be identified to the perimeter of Q : $TV(u) = \int_{\partial Q} 1 d\sigma$.

From the discrete point of view, we assume that the image $\boldsymbol{u}$ is a two-dimensional matrix of size $J = N \times N$ and we then denote by X the Euclidean space $\mathbb{R}^{N \times N} = \mathbb{R}^J$.

The discrete total variation of $\boldsymbol{u}$ is then defined by

$$TV(\boldsymbol{u}) = \sum_{1\leq i,j\leq N} |(\nabla \boldsymbol{u})_{i,j}|, \tag{8}$$

where $|y| := \sqrt{y_1^2 + y_2^2}$ for all $y = (y_1, y_2) \in \mathbb{R}^2$. Here, the discrete gradient $\nabla \boldsymbol{u}$ is a vector in $Y = X \times X$ given by $(\nabla \boldsymbol{u})_{i,j} = ((\nabla \boldsymbol{u})^1_{i,j}, (\nabla \boldsymbol{u})^2_{i,j})$ with

$$(\nabla \boldsymbol{u})^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j}, & \text{if } i < N \\ 0, & \text{if } i = N \end{cases}, \quad (\nabla \boldsymbol{u})^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j}, & \text{if } j < N \\ 0, & \text{if } j = N \end{cases}.$$

Notice that an equivalent definition of $TV(\boldsymbol{u})$ is

$$TV(\boldsymbol{u}) = \sup\{\langle \boldsymbol{\varphi}, \nabla \boldsymbol{u}\rangle_Y : \ \boldsymbol{\varphi} \in Y \text{ such that } |\varphi_{i,j}| \leq 1, \ i,j = 1,\ldots,N\} \tag{9}$$

where $\langle \boldsymbol{p}, \boldsymbol{q}\rangle_Y = \sum_{i,j}(p^1_{i,j}q^1_{i,j} + p^2_{i,j}q^2_{i,j})$. Finally, let us remark that $\langle \boldsymbol{\varphi}, \nabla \boldsymbol{u}\rangle_Y = -\langle \operatorname{div} \boldsymbol{\varphi}, \boldsymbol{u}\rangle_Y$ as soon as the discrete divergence $\operatorname{div} \boldsymbol{\varphi}$ is defined by

$$(\operatorname{div} \boldsymbol{\varphi})_{i,j} = \begin{cases} \varphi^1_{i,j} - \varphi^1_{i-1,j}, & \text{if } 1 < i < N \\ \varphi^1_{i,j}, & \text{if } i = 1 \\ -\varphi^1_{i-1,j}, & \text{if } i = N \end{cases} + \begin{cases} \varphi^2_{i,j} - \varphi^2_{i,j-1}, & \text{if } 1 < j < N \\ \varphi^2_{i,j}, & \text{if } j = 1 \\ -\varphi^2_{i,j-1}, & \text{if } j = N \end{cases}.$$

Let us also introduce the proximal operator $\operatorname{prox}_{\tau TV}[\boldsymbol{u}]$ defined by

$$\operatorname{prox}_{\tau TV}[\boldsymbol{u}] = (I + \tau \partial TV)^{-1}(\boldsymbol{u}) = \underset{\mathbf{v}\in X}{\arg\min} \left\{ \frac{1}{2\tau}\|\boldsymbol{u} - \mathbf{v}\|_X^2 + TV(\mathbf{v}) \right\},$$

where ∂TV is the sub-gradient of the TV function. Its value can be approximated for instance using the efficient dual approach proposed and analyzed by Chambolle in [6].

1.3. **State-of-the art of numerical methods for Total Variation regularization for Poisson distributed data.** In a recent work, Anthoine et al. [1] proposed some proximal methods for tomographic inverse problem with Poisson intensity and total variation regularization. Their idea consists to use some classical proximal methods to minimize an energy F in the form of:

$$F(\boldsymbol{\lambda}) = L(\boldsymbol{\lambda}) + G(\boldsymbol{\lambda}), \tag{10}$$

where L is the negative log-likelihood function (from which we withdraw the constant term) given in (3), that we re-express as:

$$L(\boldsymbol{\lambda}) = -\ell(\boldsymbol{\lambda}|\boldsymbol{y}) - \langle \log(\boldsymbol{y}!), \mathbf{1}\rangle = \langle T\boldsymbol{\lambda} - \boldsymbol{y}\log(T\boldsymbol{\lambda}), \mathbf{1}\rangle \tag{11}$$

and G is a total variational regularization term

$$G(\boldsymbol{\lambda}) = \alpha TV(\boldsymbol{\lambda}). \tag{12}$$

The operations on vectors from (11) are performed component-wise and the vector $\mathbf{1}$ is the vector of ones having the dimension of the data vector $\mathbf{y}$. A natural approach consists to apply the Chambolle-Pock [7] optimization algorithm that does not require the differentiability of the two operands. Denoting H the application given for all $\boldsymbol{\mu} \in (\mathbb{R}_+)^I$ by

$$H(\boldsymbol{\mu}) = \langle \boldsymbol{\mu} - \boldsymbol{y}\log(\boldsymbol{\mu}), \mathbf{1}\rangle, \tag{13}$$

that also verifies for all $\boldsymbol{\lambda}$ the equation $L(\boldsymbol{\lambda}) = H(T\boldsymbol{\lambda})$, the algorithm reads:

$$\begin{cases} \boldsymbol{\mu}^{(n+1)} &= (I + \sigma \partial H^*)^{-1}(\boldsymbol{\mu}^{(n)} + \sigma T\overline{\boldsymbol{\lambda}}^{(n)}) \\ \overline{\boldsymbol{\lambda}}^{(n+1)} &= (I + \tau \partial G)^{-1}(\overline{\boldsymbol{\lambda}}^{(n)} - \tau T^*\boldsymbol{\mu}^{(n+1)}) \\ \boldsymbol{\lambda}^{(n+1)} &= 2\overline{\boldsymbol{\lambda}}^{(n+1)} - \overline{\boldsymbol{\lambda}}^{(n)} \end{cases}, \tag{14}$$

and the sequence $(\boldsymbol{\lambda}^{(n)})$ converges to a minimum of F when the descent steps τ and σ verify $\tau\sigma\|T\|^2 < 1$. As F is well defined for $\boldsymbol{\lambda} > 0$, addition of a positivity constraint is also examined in [1]. Even if algorithms as (14) give interesting reconstructions, our approach is different in the sense that we want to exploit the idea and the good behaviour of the MLEM algorithm on highly noisy data.

With the same purpose, Sawatzky et al. [24, 25] proposed to successively apply:

- The (EM) step from (6), which can be expressed in matrix form as:

$$\boldsymbol{\lambda}^{(n+1/2)} = \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}} T^* \left[\frac{\boldsymbol{y}}{T\boldsymbol{\lambda}^{(n)}}\right], \tag{15}$$

- A weighted TV step

$$\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n+1/2)} - \boldsymbol{\omega}^{(n)} \partial TV(\boldsymbol{\lambda}^{(n+1)}), \tag{16}$$

  with

$$\boldsymbol{\omega}^{(n)} = \alpha \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}}. \tag{17}$$

Indeed, the gradient of $\ell(\boldsymbol{\lambda}|\boldsymbol{y})$ satisfies

$$\nabla_{\boldsymbol{\lambda}} \ell(\boldsymbol{\lambda}|\boldsymbol{y}) = T^*\mathbf{1} - T^*\left[\frac{\boldsymbol{y}}{T\boldsymbol{\lambda}}\right] \tag{18}$$

and the Euler equation associated to the minimization of $L + G$ is expressed as

$$0 \in T^*\mathbf{1} - T^*\left[\frac{\boldsymbol{y}}{T\lambda}\right] + \alpha \partial TV(\boldsymbol{\lambda}). \tag{19}$$

The inclusion of the positivity constraint as Karush-Kuhn-Tucker conditions then gives:

$$\boldsymbol{\lambda} \in \frac{\boldsymbol{\lambda}}{T^*\mathbf{1}} T^*\left[\frac{\boldsymbol{y}}{T\boldsymbol{\lambda}}\right] - \alpha \frac{\boldsymbol{\lambda}}{T^*\mathbf{1}} \partial TV(\boldsymbol{\lambda}). \tag{20}$$

According to this equality, Sawatzky et al. consider the iterative scheme

$$\boldsymbol{\lambda}^{(n+1)} = \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}} T^*\left[\frac{\boldsymbol{y}}{T\boldsymbol{\lambda}^{(n)}}\right] - \alpha \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}} \partial TV(\boldsymbol{\lambda}^{n+1}), \tag{21}$$

which appears to be equivalent to the previous weighted-TV one. Additionally, the authors explain how the weighted-TV step can be solved by minimizing the functional

$$\begin{aligned} \boldsymbol{\lambda}^{(n+1)} &= \operatorname{prox}_{\boldsymbol{\omega}^{(n)} TV}(\boldsymbol{\lambda}^{(n+1/2)}) \\ &= \arg\min_{\boldsymbol{\lambda}} \left\{ \frac{1}{2} \| \sqrt{\boldsymbol{\omega}^{(n)}} (\boldsymbol{\lambda} - \boldsymbol{\lambda}^{(n+1/2)}) \|^2 + TV(\boldsymbol{\lambda}) \right\}, \end{aligned}$$

using a slightly modified version of the Chambolle's algorithm ([6]) where $\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n+1/2)} - \boldsymbol{\omega}^{(n)} \operatorname{div} \boldsymbol{\varphi}$ and $\boldsymbol{\varphi}$ is calculated iteratively as

$$\boldsymbol{\varphi}^{(k+1)} = \frac{\boldsymbol{\varphi}^{(k)} + \tau \nabla \left(\boldsymbol{\omega}^{(n)} \operatorname{div} \boldsymbol{\varphi}^{(k)} - \boldsymbol{\lambda}^{(n+1/2)} \right)}{1 + \tau |\nabla \left(\boldsymbol{\omega}^{(n)} \operatorname{div} \boldsymbol{\varphi}^{(k)} - \boldsymbol{\lambda}^{(n+1/2)} \right)|}, \tag{22}$$

with the descent step τ verifying $0 < \tau < \dfrac{1}{4\|\boldsymbol{\omega}^{(n)}\|_\infty}$. To simplify the presentation we extended the notation $|\cdot|$ to vectors $\boldsymbol{\varphi} \in Y$, where $|\boldsymbol{\varphi}|$ is the vector of X with components $|\varphi_j|$, $j = 1, \ldots, J$.

1.4. **Outline of the paper.** Our motivation in this work is to introduce an algorithm which is similar to the one from Sawatzky et al., but using a variational point of view that allows to establish the convergence of the sequence of iterates to a minimizer of F. More precisely, we show in section 2 that an iterative sequence $(\boldsymbol{\lambda}^{(n)})_{n\geq 0}$ defined by

$$\boldsymbol{\lambda}^{(n+1)} \in \arg\min_{\boldsymbol{u}} \{\langle \boldsymbol{u}, \boldsymbol{s}\rangle - \langle \log(\boldsymbol{u}), \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\rangle + G(\boldsymbol{u})\}, \tag{23}$$

and $\boldsymbol{\lambda}^{(0)} > 0$ converges to a minimizer of F. Here, as in Sawatzky et al., $\boldsymbol{\lambda}^{(n+1/2)}$ is the result of the EM step (15) and we note $\boldsymbol{s} = T^*\mathbf{1}$ which is assumed to satisfy

$$\min_j s_j = s_{\min} > 0. \tag{24}$$

In section 3, we focus on the numerical computation of a solution of (23). In particular, we introduce a descent algorithm based on a dual approach and prove its stability under classical assumptions on the step size.

Finally, in the last section, we give numerical comparisons of our algorithm with the ones proposed by Anthoine et al. [1] and Sawatzky et al. [24]. We will see in particular that the MAP-EM approach gives better results in the sense that the energy F decreases faster.

2. **General variational framework for the Expectation Maximization approach with smooth or non smooth prior.** In this section we adapt the general framework of the Expectation Maximization algorithm introduced in [8] to non-smooth posterior distributions. The case of exponential families to which the Poisson distributions belong is then derived as a particular case.

Let $q(\boldsymbol{x}|\boldsymbol{\phi})$ be the family of sampling densities depending on parameter $\boldsymbol{\phi} \in X$ characterizing the complete data, $p(\boldsymbol{y}|\boldsymbol{\phi})$ the sampling density of the incomplete data and $k(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\phi})$ the conditional density of $\boldsymbol{x}$ given $\boldsymbol{y}$ and $\boldsymbol{\phi}$. As in [8] we denote

$$P(\boldsymbol{\phi}) = \log(p(\boldsymbol{y}|\boldsymbol{\phi}), \tag{25}$$

$$Q(\boldsymbol{\phi}'|\boldsymbol{\phi}) = \mathbb{E}(\log q(\boldsymbol{x}|\boldsymbol{\phi}')|\boldsymbol{y}, \boldsymbol{\phi}), \tag{26}$$

$$K(\boldsymbol{\phi}'|\boldsymbol{\phi}) = \mathbb{E}(\log k(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\phi}')|\boldsymbol{y}, \boldsymbol{\phi}), \tag{27}$$

which implies that

$$Q(\boldsymbol{\phi}'|\boldsymbol{\phi}) = P(\boldsymbol{\phi}') + K(\boldsymbol{\phi}', \boldsymbol{\phi}). \tag{28}$$

Assume that the log-likelihood P of the complete data, described in our tomographic application by equation (6), is differentiable, concave and that its negative $-P$ is coercive,

$$\lim_{\|\boldsymbol{\phi}\|\to+\infty} P(\boldsymbol{\phi}) = \infty. \tag{29}$$

From the Jensen's inequality it follows that

$$\forall(\boldsymbol{\phi}, \boldsymbol{\phi}') \in X \times X \quad K(\boldsymbol{\phi}'|\boldsymbol{\phi}) \leq K(\boldsymbol{\phi}|\boldsymbol{\phi}), \tag{30}$$

and we also suppose that for all $\boldsymbol{\phi} \in X$, the application $\boldsymbol{\phi}' \mapsto Q(\boldsymbol{\phi}'|\boldsymbol{\phi})$ is concave and bounded above.

Instead of searching for an estimation of $\boldsymbol{\phi}$ as the maximum of the likelihood P, our strategy here consists to calculate the maximum a posteriori associated to some prior density whose negative log-likelihood is G. Equivalently, we search for the minimum of

$$F = -P + G, \tag{31}$$

where G is assumed to be a continuous convex function possibly non smooth.

The idea of the maximum a posteriori Expectation Maximization (MAP-EM) algorithm is to iteratively reach a maximum of F by solving a sequence of minimization problems

$$\boldsymbol{\phi}^{(n+1)} \in \underset{\boldsymbol{\phi}\in X}{\arg\min}\,\{U(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)})\}, \tag{32}$$

relaxed versions of (31), where the criterion is

$$U(\boldsymbol{\phi}'|\boldsymbol{\phi}) = F(\boldsymbol{\phi}') - K(\boldsymbol{\phi}'|\boldsymbol{\phi}) = G(\boldsymbol{\phi}') - Q(\boldsymbol{\phi}'|\boldsymbol{\phi}). \tag{33}$$

Hereafter we show that the sequence $(\boldsymbol{\phi}^{(n)})$ converges to a minimum of F. For this we revisit the proof given to a similar result in [8], using the non-smooth convex analysis formalism.

2.1. **Convergence of the iterative sequence produced by a GEM algorithm.** The following lemma characterizes the solution of the minimum of function F from (31).

Lemma 2.1. *A vector $\boldsymbol{\phi}^*$ is a minimum of F if and only if*

$$\boldsymbol{\phi}^* \in \arg\min\{U(\boldsymbol{\phi}|\boldsymbol{\phi}^*) \,:\, \boldsymbol{\phi} \in X\}. \tag{34}$$

Proof. Let $\boldsymbol{\phi}^*$ be a minimum of F. From (30) we know that for all $\boldsymbol{\phi} \in X$, $K(\boldsymbol{\phi}|\boldsymbol{\phi}^*) \leq K(\boldsymbol{\phi}^*|\boldsymbol{\phi}^*)$ thus $U(\boldsymbol{\phi}|\boldsymbol{\phi}^*) \geq U(\boldsymbol{\phi}^*|\boldsymbol{\phi}^*)$ and equation (34) is verified. Conversely, if $\boldsymbol{\phi}^*$ verifies (34) then $0 \in \partial U(\cdot\,|\boldsymbol{\phi}^*)(\boldsymbol{\phi}^*) = \partial F(\boldsymbol{\phi}^*) - \nabla K(\cdot|\boldsymbol{\phi}^*)(\boldsymbol{\phi}^*)$, where $\nabla K(\cdot|\boldsymbol{\phi}^*)(\boldsymbol{\phi}^*)$ is the gradient of $K(\cdot|\boldsymbol{\phi}^*)$ at $\boldsymbol{\phi}^*$. From Lemma 2 of [8], $\nabla K(\cdot|\boldsymbol{\phi})(\boldsymbol{\phi}) = 0$ for all $\boldsymbol{\phi}$ and in particular for $\boldsymbol{\phi} = \boldsymbol{\phi}^*$, thus $0 \in \partial F(\boldsymbol{\phi}^*)$ and $\boldsymbol{\phi}^*$ is a minimum of F. □

Hereafter we re-use but also re-define the term Generalized Expectation Maximization (GEM) algorithm from Dempster et al. [8].

Definition 2.2. For an application U bounded above and such that for all $\boldsymbol{\phi} \in X$, the mapping $\boldsymbol{\phi}' \in X \mapsto U(\boldsymbol{\phi}'|\boldsymbol{\phi})$ is concave, we call Generalized Expectation Maximization (GEM) algorithm a continuous mapping $M : X \to X$ such that if

$$\boldsymbol{\phi} \notin \arg\min\{U(\boldsymbol{\phi}'|\boldsymbol{\phi}) \,:\, \boldsymbol{\phi}' \in X\}, \tag{35}$$

then $M(\boldsymbol{\phi})$ should verify

$$U(M(\boldsymbol{\phi})|\boldsymbol{\phi}) < U(\boldsymbol{\phi}|\boldsymbol{\phi}), \tag{36}$$

otherwise $M(\boldsymbol{\phi}) = \boldsymbol{\phi}$.

The next theorem shows that each iteration of a GEM algorithm decreases the value of the cost function F except when the minimum is already attained, and establishes the link between the minima of F and the fixed points of the GEM algorithm.

Theorem 2.3. *For any GEM algorithm the following properties hold.*

(i) For all $\boldsymbol{\phi} \notin \arg\min\{U(\boldsymbol{\phi}'|\boldsymbol{\phi}) \,:\, \boldsymbol{\phi}' \in X\}$ we have

$$F(M(\boldsymbol{\phi})) < F(\boldsymbol{\phi}).$$

(ii) A vector $\boldsymbol{\phi}^*$ is a fixed point of M if and only if

$$\boldsymbol{\phi}^* \in \arg\min\{U(\boldsymbol{\phi}|\boldsymbol{\phi}^*) \,:\, \boldsymbol{\phi} \in X\}. \tag{37}$$

(iii) The set of fixed points of M coincides with the set of points where F attains its minimum: for all $\boldsymbol{\phi}^* \in X$,

$$\left[M(\boldsymbol{\phi}^*) = \boldsymbol{\phi}^* \;\Leftrightarrow\; \boldsymbol{\phi}^* \in \arg\min\{F(\boldsymbol{\phi}) \,:\, \boldsymbol{\phi} \in X\}\right]. \tag{38}$$

Proof. (i) From (33) and (30) it follows that

$$\begin{aligned} F(M(\boldsymbol{\phi})) &= U(M(\boldsymbol{\phi})|\boldsymbol{\phi}) + K(M(\boldsymbol{\phi})|\boldsymbol{\phi}) \\ &\leq U(M(\boldsymbol{\phi})|\boldsymbol{\phi}) + K(\boldsymbol{\phi}|\boldsymbol{\phi}). \end{aligned}$$

Then from (36) and again (33) we obtain the result.

(ii) If $\boldsymbol{\phi}^*$ is a fixed point of M then $U(M(\boldsymbol{\phi}^*)|\boldsymbol{\phi}^*) = U(\boldsymbol{\phi}^*|\boldsymbol{\phi}^*)$. Equation (37) follows from definition 2.2. The reciprocal is obvious by the definition of GEM.

(iii) This property immediately follows from (ii) and Lemma 2.1.

□

Let $(\boldsymbol{\phi}^{(n)})_{n\in\mathbb{N}}$ be a sequence produced by a GEM algorithm, *i.e.*, for all $n \in \mathbb{N}^*$, $\boldsymbol{\phi}^{(n+1)} = M(\boldsymbol{\phi}^{(n)})$ and $\boldsymbol{\phi}^{(0)}$ is some given initial value in X. We show now that the sequence $(\boldsymbol{\phi}^{(n)})_{n\in\mathbb{N}}$ converges to a minimizer of F at least when F is strictly convex. Note that the strict convexity of F was not required in the seminal paper [8], but the proof given therein is flawed. To the best of our knowledge there is no general proof of convergence for the MAP-EM algorithm, and the proof given for the particular case of the Poisson-MLEM algorithm by A.N. Iusem in [13] cannot be adapted easily to MAP. However, the theoretical strict convexity could be obtained by addition of a second regularization term with a very small coefficient that makes its numerical influence on the result negligible.

Theorem 2.4. *For any sequence $(\boldsymbol{\phi}^{(n)})_{n\in\mathbb{N}}$ produced by a GEM algorithm the following properties hold.*

(i) The sequence $(F(\boldsymbol{\phi}^{(n)}))$ is non-increasing and converges to the minimum of F.

(ii) If $(\boldsymbol{\phi}^{(n_k)})$ is a convergent sub-sequence of $(\boldsymbol{\phi}^{(n)})$ with limit $\boldsymbol{\phi}^*$, then $\boldsymbol{\phi}^* \in \arg\min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}$.

(iii) If F is strictly convex,

$$\lim_{n\to+\infty} \boldsymbol{\phi}^{(n)} = \arg\min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}.$$

Proof. (i) The fact that the sequence $(F(\boldsymbol{\phi}^{(n)}))$ is non-increasing is a direct consequence of theorem 2.3 (i). It is clear that a GEM sequence is bounded. Indeed, if this would not be the case, a sub-sequence $(\boldsymbol{\phi}^{(n_k)})$ such that $\lim_{k\to+\infty} \|\boldsymbol{\phi}^{(n_k)}\| = +\infty$ may be extracted. Since F is coercive, the sequence $(F(\boldsymbol{\phi}^{(n_k)}))$ would not be bounded either, which comes in contradiction with the fact that $(F(\boldsymbol{\phi}^{(n)}))$ is non-increasing and bounded below by the minimum of F. Let $(\boldsymbol{\phi}^{(n_k)})$ be a convergent sub-sequence of $(\boldsymbol{\phi}^{(n)})$ with limit $\boldsymbol{\phi}^*$. Then the sub-sequence $(\boldsymbol{\phi}^{(n_k+1)})$ is also convergent and tends to $M(\boldsymbol{\phi}^*)$. The sequence $(F(\boldsymbol{\phi}^{(n)}))$ being non-increasing and bounded below, it converges and

$$\lim_{n\to+\infty} F(\boldsymbol{\phi}^{(n)}) = \lim_{k\to+\infty} F(\boldsymbol{\phi}^{(n_k)}) = \lim_{k\to+\infty} F(\boldsymbol{\phi}^{(n_k+1)}),$$

thus $F(M(\boldsymbol{\phi}^*)) = F(\boldsymbol{\phi}^*)$. From theorem 2.3 (i) we then deduce that

$$\boldsymbol{\phi}^* \in \arg\min\{U(\boldsymbol{\phi}|\boldsymbol{\phi}^*) : \boldsymbol{\phi} \in X\},$$

and from the same theorem it results that $\boldsymbol{\phi}^*$ is a fixed point of M and $\boldsymbol{\phi}^* \in \arg\min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}$. Thus

$$\lim_{n\to+\infty} F(\boldsymbol{\phi}^{(n)}) = F(\boldsymbol{\phi}^*) = \min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}.$$

(ii) As an immediate consequence of the proof of (i) we have:

$$F(\boldsymbol{\phi}^*) = \lim_{k\to+\infty} F(\boldsymbol{\phi}^{(n_k)}) = \lim_{n\to+\infty} F(\boldsymbol{\phi}^{(n)}) = \min\{F(\boldsymbol{\phi}) \,:\, \boldsymbol{\phi}\in X\}.$$

(iii) If F is strictly convex there is an unique

$$\boldsymbol{\phi}^* \in \arg\min\{F(\boldsymbol{\phi}) \,:\, \boldsymbol{\phi}\in X\}.$$

From (ii), any convergent sub-sequence of $(\boldsymbol{\phi}^{(n)})$ has to converge to $\boldsymbol{\phi}^*$, thus the sequence converges to the same limit. □

A MAP estimator of the parameter $\boldsymbol{\phi}$ of the complete data density function can thus be obtained by choosing some initial value $\boldsymbol{\phi}^{(0)} \in X$ and then constructing a sequence of minimizers following the recursive relation (32). This condition, corresponding to the (M) step of the EM algorithm, is stronger than the GEM algorithm since it requires

$$U(\boldsymbol{\phi}^{(n+1)}|\boldsymbol{\phi}^{(n)}) \leq U(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)}) \tag{39}$$

for all $\boldsymbol{\phi}\in X$ and not only for $\boldsymbol{\phi} = \boldsymbol{\phi}^{(n)}$. However, when the solution of (32) is too difficult to obtain exactly, an approximate solution that verifies the GEM condition is sufficient to ensure the convergence of $(F(\boldsymbol{\phi}^{(n)}))$ to the minimum value of F.

We showed that a minimizing sequence for F can thus be obtained from (32) where U is defined by the law of the complete data and the regularization term. In the next subsection we apply this result to the particular case of tomography with low number of photons and data following Poisson distributions, a model largely employed in nuclear medicine imaging.

2.2. **Application to tomography with counted photons.** In emission tomography, the low emission rates makes that the photons are detected individually and their number follows a Poisson law. The same holds at emission level and the likelihood of the complete data belongs to an exponential family with logarithm that may be written as (for details see *e.g.*, [15, 32]):

$$\log q(\boldsymbol{x}|\boldsymbol{\lambda}) = -\langle\boldsymbol{\lambda}, \boldsymbol{s}\rangle + \langle\log(\boldsymbol{s}\boldsymbol{\lambda}), t(\boldsymbol{x})\rangle + \log b(\boldsymbol{x}), \tag{40}$$

where $\boldsymbol{s} = T^*\mathbf{1}$ and the product of two vectors $(\boldsymbol{s}\boldsymbol{\lambda})$ is a component-wise product. The (E) step of the EM algorithm consists to calculate its expectation conditionally on the observations $\boldsymbol{y}$ and on the previous estimate $\boldsymbol{\lambda}^{(n)}$ which is:

$$Q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(n)}) = -\langle\boldsymbol{\lambda}, \boldsymbol{s}\rangle + \langle\log(\boldsymbol{s}) + \log(\boldsymbol{\lambda}), \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\rangle + c, \tag{41}$$

where c is a constant and $\boldsymbol{\lambda}^{(n+1/2)}$ corresponds to the MLEM estimation and is calculated from $\boldsymbol{\lambda}^{(n)}$ following (15). To compute the MAP-EM estimate $\boldsymbol{\lambda}^{(n+1)}$ in case the minimum was not already reached in $\boldsymbol{\lambda}^{(n)}$, one must either choose

$$\boldsymbol{\lambda}^{(n+1)} \in \underset{\boldsymbol{u}\in(\mathbb{R}_+)^J}{\arg\min}\{\langle\boldsymbol{u}, \boldsymbol{s}\rangle - \langle\log(\boldsymbol{u}), \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\rangle + G(\boldsymbol{u})\}, \tag{42}$$

or at least find a $\boldsymbol{\lambda}^{(n+1)}$ such that

$$U(\boldsymbol{\lambda}^{(n+1)}|\boldsymbol{\lambda}^{(n)}) < U(\boldsymbol{\lambda}^{(n)}|\boldsymbol{\lambda}^{(n)}).$$

Equation (42) was previously derived in [35] for the differentiable total variation prior. The same equation results in the smooth case directly from [8] and was also derived in [11]. Here we obtain (42) as a particular case of the MAP-EM estimation with GEM algorithms for non-smooth priors.

Solutions for the numerical calculation of the MAP estimator were already proposed in the literature. In [11], an explicit scheme with complexity similar to the

one of the MLEM algorithm was proposed. The (smoothed) total variation regularization comes as a particular case and the same explicit scheme was studied in [21] and [22]. It was however observed that the explicit scheme is unstable and requires very low regularization parameters. The semi-implicit scheme from [35] also seems unstable. A stable explicit-implicit scheme was proposed in [24, 25].

In the following section, we introduce and give the stability analysis for an efficient dual algorithm for the resolution of (42) with non-smooth total variation regularization.

3. A dual algorithm to compute the MAP-EM estimator.

The objective of this section is to approximate numerically the minimizer

$$\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{u}^* = \arg\min_{\boldsymbol{u}} \left\{ H(\boldsymbol{u}) + G(\boldsymbol{u}) \right\},$$

where

- the function H is defined for all $\boldsymbol{u} \in (\mathbb{R}_+^*)^J$ by

$$H(\boldsymbol{u}) = \langle \boldsymbol{u} - \boldsymbol{\lambda}^{(n+1/2)} \log \boldsymbol{u}, \boldsymbol{s} \rangle, \tag{43}$$

  where $\boldsymbol{s}$ and $\boldsymbol{\lambda}^{(n+1/2)}$ are considered constant during the minimization procedure,
- the function G is the discrete total variation functional $G(\boldsymbol{u}) = \alpha TV(\boldsymbol{u})$, with TV defined in (9).

Some connection can be made between the functions defined by equations (43) and (13). The role of both is to avoid the linear transform induced by the matrix T and thus to simplify the minimization. However the two functions are fundamentally different, since (43) is defined for variables having the dimension of the complete data and makes use of the partial estimation $\boldsymbol{\lambda}^{(n+1/2)}$, whether (13) is defined for variables having the dimension of the incomplete data.

A first natural idea to minimize $H + G$ consists to use a splitting approach with implicit minimization of both H and G. More precisely, we may consider a Douglas-Rachford splitting consisting to iterate for $l > 0$:

$$\left\{ \begin{array}{lcl} \boldsymbol{u}^{(l)} & = & \text{prox}_{\sigma G}[\mathbf{v}^{(l)}] \\ \mathbf{v}^{(l+1)} & = & \mathbf{v}^{(l)} + \lambda_l(\text{prox}_{\sigma H}[2\boldsymbol{u}^{(l)} - \mathbf{v}^{(l)}] - \boldsymbol{u}^{(l)}) \end{array} \right. . \tag{44}$$

Here, $\lambda_l \in [\varepsilon, 2-\varepsilon]$ with $\varepsilon \in (0,1)$ can be viewed as an inertial parameter, $\sigma > 0$ is the descent step and the initial value $\mathbf{v}^{(0)}$ can be set to $\boldsymbol{\lambda}^{(n+1/2)}$. Then the sequence $(\boldsymbol{u}^{(l)})_{l \geq 0}$ is expected to converge to a minimum of $H + G$.

The proximal operator $\text{prox}_{\sigma H}[\boldsymbol{u}]$ defined by

$$\text{prox}_{\sigma H}[\boldsymbol{u}] = (I + \sigma \partial H)^{-1}(\boldsymbol{u}) = \arg\min_{\boldsymbol{v}} \left\{ \frac{1}{2\sigma} \|\boldsymbol{u} - \boldsymbol{v}\|^2 + H(\boldsymbol{v}) \right\},$$

can be calculated explicitly as

$$\text{prox}_{\sigma H}[\boldsymbol{u}] = \frac{1}{2} \left((\boldsymbol{u} - \sigma \boldsymbol{s}) + \sqrt{(\boldsymbol{u} - \sigma \boldsymbol{s})^2 + 4\sigma \mathbf{s} \boldsymbol{\lambda}^{(n+1/2)}} \right) \tag{45}$$

and the proximal operator $\text{prox}_{\sigma G}[\boldsymbol{u}]$ can be computed numerically using the Chambolle's algorithm ([6]),

$$\text{prox}_{\sigma G}[\boldsymbol{u}] = \boldsymbol{u} - \sigma \, \text{div}\, \boldsymbol{\varphi},$$

with $\boldsymbol{\varphi}$ calculated iteratively as:

$$\boldsymbol{\varphi}^{(k+1)} = \frac{\boldsymbol{\varphi}^{(k)} + \tau\nabla\left(\operatorname{div}\boldsymbol{\varphi}^{(k)} - \boldsymbol{u}/\sigma\right)}{1 + \tau|\nabla\left(\operatorname{div}\boldsymbol{\varphi}^{(k)} - \boldsymbol{u}/\sigma\right)|}, \tag{46}$$

for some descent step chosen as $0 < \tau < 1/4$. Even if this algorithm gives good solutions, its algorithmic cost is important as it requires three iteration loops: one in n for the calculation of the intermediate MLEM solution, one in l for the Douglas-Rachford splitting and one in k for the application of the Chambolle's algorithm.

In this paper, we propose to improve the efficiency of the method by using a complete dual approach to minimize $H + G$. In particular, as will be described later, the dual approach reduces the complexity as only the loops in n and k are necessary.

As the functions H and G are proper convex and lower semi-continuous, the Fenchel-Rockafellar duality theorem states that if H^* and G^* are the convex conjugates of H and G,

$$H^*(\boldsymbol{p}) = \sup_{\boldsymbol{u}}\left\{\langle\boldsymbol{u},\boldsymbol{p}\rangle - H(\boldsymbol{u})\right\} \tag{47}$$

and

$$G^*(\boldsymbol{p}) = \sup_{\boldsymbol{u}}\left\{\langle\boldsymbol{u},\boldsymbol{p}\rangle - G(\boldsymbol{u})\right\}, \tag{48}$$

then

$$\inf_{\mathbf{u}}\left\{H(\boldsymbol{u}) + G(\boldsymbol{u})\right\} = -\inf_{\mathbf{p}}\left\{H^*(-\mathbf{p}) + G^*(\mathbf{p})\right\}. \tag{49}$$

If $\boldsymbol{u}^*$ is a minimizer of $H + G$ then there exists a solution $\boldsymbol{p}^*$ of the dual problem such that $-\boldsymbol{p}^* \in \partial H(\boldsymbol{u}^*)$ and $\boldsymbol{u}^* \in \partial G^*(\boldsymbol{p}^*)$. The minimizer $\boldsymbol{u}^*$ of the primal problem should verify $\boldsymbol{u}^* > 0$. Since H is differentiable for $\boldsymbol{u}^* > 0$, $-\boldsymbol{p}^* = \nabla H(\boldsymbol{u}^*)$ and we get

$$\boldsymbol{u}^* = \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \boldsymbol{p}^*}. \tag{50}$$

A minimizer $\boldsymbol{u}^*$ of $H + G$ can then be deduced from a minimizer $\boldsymbol{p}^*$ of $H^* + G^*$.

3.1. **Formulation of the dual problem.** The Fenchel-Legendre transform of H can be calculated explicitly as follows. From the Euler equation associated to the maximization in $\boldsymbol{u}$ of $\langle\boldsymbol{u},\mathbf{p}\rangle - H(\boldsymbol{u})$, we obtain

$$H^*(\mathbf{p}) = \langle\boldsymbol{u}^*[\mathbf{p}],\mathbf{p}\rangle - H(\boldsymbol{u}^*[\mathbf{p}]) \tag{51}$$

with

$$\boldsymbol{u}^*[\mathbf{p}] = \max\left\{\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} - \mathbf{p}}, 0\right\}. \tag{52}$$

Thus $H^*(\mathbf{p})$ is infinite for $\mathbf{p} \geq \boldsymbol{s}$. For $\mathbf{p} < \boldsymbol{s}$ its value is:

$$\begin{aligned} H^*(\mathbf{p}) &= -\left\langle\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s}-\mathbf{p}}, \boldsymbol{s}-\mathbf{p}\right\rangle + \left\langle\log\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s}-\mathbf{p}}, \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\right\rangle \\ &= \left\langle\log(\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}) - \mathbf{1}, \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\right\rangle - \left\langle\log(\boldsymbol{s}-\mathbf{p}), \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\right\rangle \\ &= C - \left\langle\log(\boldsymbol{s}-\mathbf{p}), \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\right\rangle, \end{aligned} \tag{53}$$

with C independent from $\mathbf{p}$. On the other side, it was shown in [6] that the Fenchel-Legendre transform of the total variation functional G is

$$G^*(\boldsymbol{p}) = \chi_{\alpha K}(\boldsymbol{p}) = \begin{cases} 0 & \text{if } \boldsymbol{p} \in \alpha K \\ +\infty & \text{otherwise,} \end{cases} \tag{54}$$

where K is the set

$$K = \{\operatorname{div} \boldsymbol{\varphi} : \ \boldsymbol{\varphi} \in Y, \ |\boldsymbol{\varphi}| \leq 1\}.$$

Now from (53) and (54) it follows that the solution p^* of the dual problem can be expressed as $p^* = \alpha \operatorname{div} \boldsymbol{\varphi}^*$, where $\boldsymbol{\varphi}^*$ is the solution of

$$\inf\{h(\boldsymbol{\varphi}) : \ \boldsymbol{\varphi} \in Y, \ |\boldsymbol{\varphi}| \leq 1\}. \tag{55}$$

The function h is defined for $\boldsymbol{\varphi}$ such that $\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi} > 0$ by

$$h(\boldsymbol{\varphi}) = H^*(-\alpha \operatorname{div} \boldsymbol{\varphi}) - C = -\langle \log(\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}), \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)} \rangle \tag{56}$$

and is infinite elsewhere. Finally, by the duality equation (50), the minimizer $\boldsymbol{u}^*$ of the primal problem can be obtained from the minimizer $\boldsymbol{\varphi}^*$ of the dual problem as:

$$\boldsymbol{u}^* = \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}^*}. \tag{57}$$

Remark 1. From a discrete point of view, note that if for all $j = 1 \ldots, J$, $|\varphi_j| \leq 1$, we have $\|\operatorname{div} \boldsymbol{\varphi}\|_\infty \leq 4$ and then for $\alpha < s_{\min}/4$ we obtain

$$\boldsymbol{s} - \alpha \operatorname{div} \boldsymbol{\varphi} \geq s_{\min} - 4\alpha > 0,$$

where $s_{\min}$ is defined in (24). The functional h is thus finite on $S = \{\boldsymbol{\varphi} \in Y : \ |\boldsymbol{\varphi}| \leq 1\}$. In the rest of the paper we assume that $\alpha < s_{\min}/4$. This condition is also assumed for the proof of convergence given in theorem 3.2.

Remark 2. In a generalization from two-dimensional images to three-dimensional volumes we will have $\|\operatorname{div} \boldsymbol{\varphi}\|_\infty \leq 6$ and the condition on the regularization parameter becomes $\alpha < s_{\min}/6$.

3.2. **An iterative scheme to solve the dual problem.** In this subsection we derive an algorithm to compute a solution of (55), that is

$$\boldsymbol{\varphi}^* = \arg\min_{\boldsymbol{\varphi} \in S} h(\boldsymbol{\varphi}), \tag{58}$$

with $S = \{\boldsymbol{\varphi} \in Y : \ |\boldsymbol{\varphi}| \leq 1\}$. The gradient of h is

$$\nabla h(\boldsymbol{\varphi}) = \alpha \nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}} \right). \tag{59}$$

Now, given $\boldsymbol{\mu} \in X$ the Lagrange multiplier associated to the constraint $|\boldsymbol{\varphi}| \leq 1$, the Karush-Kuhn-Tuker condition reads:

$$\alpha \nabla \left(\frac{s\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}} \right) + \boldsymbol{\mu}\boldsymbol{\varphi} = 0, \tag{60}$$

with either $|\varphi_j| = 1$ and $\mu_j > 0$, or $|\varphi_j| < 1$ and $\mu_j = 0$. Moreover, as in the latter case we also have

$$\left| \left(\nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}} \right) \right)_j \right| = 0,$$

we see that

$$\boldsymbol{\mu} = \alpha \left| \nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}} \right) \right|. \tag{61}$$

Following the idea of Chambolle's algorithm, we consider a semi-implicit gradient descent scheme and we show that it converges to the solution $\boldsymbol{\varphi}^*$ of (58). With

some minimization step $\tau > 0$ and the initial value $\boldsymbol{\varphi}^{(0)} = 0$ we then consider the scheme

$$\boldsymbol{\varphi}^{(k+1)} = \boldsymbol{\varphi}^{(k)} - \tau \nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}^{(k)}} \right) - \tau \left| \nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}^{k}} \right) \right| \boldsymbol{\varphi}^{(k+1)}, \tag{62}$$

which can be written equivalently as

$$\boldsymbol{\varphi}^{(k+1)} = \frac{\boldsymbol{\varphi}^{(k)} - \tau \mathbf{z}^{(k)}}{1 + \tau |\mathbf{z}^{(k)}|}, \quad \text{with } \mathbf{z}^{(k)} = \nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}^{(k)}} \right). \tag{63}$$

Hereafter we will show the convergence of the sequence of iterates to the solution of (58), which requires the following auxiliary result.

Lemma 3.1. *If $\alpha < s_{\min}/4$, the function h defined in (56) is convex and continuously differentiable on its domain. Its gradient is Lipschitz on S with constant*

$$L_h = 8\alpha^2 \frac{\|\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\|_\infty}{(s_{\min} - 4\alpha)^2}. \tag{64}$$

Proof. The first derivative of h satisfies for $\boldsymbol{\varphi} \in Y$ such that $\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi} > 0$ and $\boldsymbol{\psi} \in X$ the equation:

$$h'(\boldsymbol{\varphi})(\boldsymbol{\psi}) = -\alpha \left\langle \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}}, \mathrm{div}\, (\boldsymbol{\psi}) \right\rangle,$$

thus

$$\begin{aligned}
\|\nabla h(\boldsymbol{\varphi}) - \nabla h(\boldsymbol{\varphi}')\|_2 &= \sup_{\|\boldsymbol{\psi}\|_2 = 1} |h'(\boldsymbol{\varphi})(\boldsymbol{\psi}) - h'(\boldsymbol{\varphi}')(\boldsymbol{\psi})| \\
&= \alpha^2 \sup_{\|\boldsymbol{\psi}\|_2 = 1} \left| \left\langle \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)} \mathrm{div}\, (\boldsymbol{\varphi} - \boldsymbol{\varphi}')}{(\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi})(\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}')}, \mathrm{div}\, (\boldsymbol{\psi}) \right\rangle \right| \\
&\leq \alpha^2 \|\mathrm{div}\, (\boldsymbol{\varphi} - \boldsymbol{\varphi}')\|_2 \sup_{\|\boldsymbol{\psi}\|_2 = 1} \left\| \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)} \mathrm{div}\, \boldsymbol{\psi}}{(\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi})(\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}')} \right\|_2.
\end{aligned}$$

As shown in [6], $\|\mathrm{div}\, (\boldsymbol{\varphi} - \boldsymbol{\varphi}')\|_2^2 \leq 8 \|\boldsymbol{\varphi} - \boldsymbol{\varphi}'\|_2^2$ and $\|\mathrm{div}\, \boldsymbol{\psi}\|_2^2 \leq 8$, thus

$$\|\nabla h(\boldsymbol{\varphi}) - \nabla h(\boldsymbol{\varphi}')\|_2 \leq 8\alpha^2 \left\| \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{(\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi})(\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}')} \right\|_\infty \|\boldsymbol{\varphi} - \boldsymbol{\varphi}'\|_2.$$

Since we have $\|\mathrm{div}\, \boldsymbol{\varphi}\|_\infty \leq 4$ on S, it follows that

$$\|\nabla h(\boldsymbol{\varphi}) - \nabla h(\boldsymbol{\varphi}')\|_2 \leq 8\alpha^2 \frac{\|\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\|_\infty}{(s_{\min} - 4\alpha)^2} \|\boldsymbol{\varphi} - \boldsymbol{\varphi}'\|_2,$$

thus the gradient of h is Lipschitz with constant given in (64).

The second derivative of h is

$$h''(\boldsymbol{\varphi})(\boldsymbol{\psi})(\widetilde{\boldsymbol{\psi}}) = \left\langle \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{(\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi})^2}, \alpha^2 \mathrm{div}\, (\boldsymbol{\psi}) \mathrm{div}\, (\widetilde{\boldsymbol{\psi}}) \right\rangle,$$

and thus h is convex as

$$h''(\boldsymbol{\varphi})(\boldsymbol{\psi})(\boldsymbol{\psi}) = \left\langle \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{(\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi})^2}, \alpha^2 \mathrm{div}\, (\boldsymbol{\psi})^2 \right\rangle \geq 0.$$

□

Remark 3. When the coefficient of regularization is sufficiently small, i.e $\alpha << 1$, the constant L_h can also be approximated by

$$L_h \simeq 8\alpha^2 \|\boldsymbol{\lambda}^{(n+1/2)}/\boldsymbol{s}\|_\infty. \tag{65}$$

We can now state the main result of this section.

Theorem 3.2. *Let $\alpha < s_{\min}/4$ and $\tau < \alpha/L_h$. Then the sequence $(h(\boldsymbol{\varphi}^{(k)}))$ with $\boldsymbol{\varphi}^{(k)}$ defined in (63) is decreasing and converges to the minimum of h on S.*

Proof. It is easy to show by induction that $|\boldsymbol{\varphi}^{(k)}| \leq 1$ for all $k \in \mathbb{N}$. As the functional h is convex and with gradient L_h-Lipschitz on K, it follows from a classical result (see *e.g.*, [20]) that for all $k \in \mathbb{N}$,

$$h(\boldsymbol{\varphi}^{(k+1)}) \leq h(\boldsymbol{\varphi}^{(k)}) + \langle \nabla h(\boldsymbol{\varphi}^{(k)}), \boldsymbol{\varphi}^{(k+1)} - \boldsymbol{\varphi}^{(k)} \rangle + \frac{L_h}{2} \|\boldsymbol{\varphi}^{(k+1)} - \boldsymbol{\varphi}^{(k)}\|_2^2.$$

Let us note $\boldsymbol{\eta} = (\boldsymbol{\varphi}^{(k+1)} - \boldsymbol{\varphi}^{(k)})/\tau$. As $\nabla h(\boldsymbol{\varphi}^{(k)}) = \alpha \boldsymbol{z}^{(k)}$, we obtain

$$h(\boldsymbol{\varphi}^{(k+1)}) - h(\boldsymbol{\varphi}^{(k)}) \leq \tau \left(\alpha \langle \boldsymbol{z}^{(k)}, \boldsymbol{\eta} \rangle + \tau \frac{L_h}{2} \|\boldsymbol{\eta}\|_2^2 \right),$$

From (62) it follows that

$$\boldsymbol{\eta} = -\boldsymbol{z}^{(k)} - |\boldsymbol{z}^{(k)}| \boldsymbol{\varphi}^{(k+1)}, \text{ with } \boldsymbol{z}^{(k)} = \nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}^{(k)}} \right),$$

which is well defined as $\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}^{(k)} > 0$. From the last two equations we then obtain

$$\begin{aligned} h(\boldsymbol{\varphi}^{(k+1)}) - h(\boldsymbol{\varphi}^{(k)}) \quad \leq \quad & -\alpha (\|\boldsymbol{z}^{(k)}\|_2^2 + \langle \boldsymbol{z}^{(k)}, |\boldsymbol{z}^{(k)}| \boldsymbol{\varphi}^{(k+1)} \rangle) \\ & + \tau \frac{L_h}{2} \left(\|\boldsymbol{z}^{(k)}\|_2^2 + 2 \langle \boldsymbol{z}^{(k)}, |\boldsymbol{z}^{(k)}| \boldsymbol{\varphi}^{(k+1)} \rangle + \|\boldsymbol{z}^{(k)} \boldsymbol{\varphi}^{(k+1)}\|_2^2 \right). \end{aligned}$$

Now, as $\|\boldsymbol{\varphi}^{(k+1)}\|_\infty \leq 1$, it holds $\|\boldsymbol{z}^{(k)} \boldsymbol{\varphi}^{(k+1)}\|_2^2 \leq \|\boldsymbol{z}^{(k)}\|_2^2$ thus

$$h(\boldsymbol{\varphi}^{(k+1)}) - h(\boldsymbol{\varphi}^{(k)}) \leq (-\alpha + \tau L_h) \left(\|\boldsymbol{z}^{(k)}\|_2^2 + \langle \boldsymbol{z}^{(k)}, |\boldsymbol{z}^{(k)}| \boldsymbol{\varphi}^{(k+1)} \rangle \right).$$

Finally, as $\langle \boldsymbol{z}^{(k)}, |\boldsymbol{z}^{(k)}| \boldsymbol{\varphi}^{(k+1)} \rangle \geq -\|\boldsymbol{z}^{(k)}\|_2^2$ and $\alpha - \tau L_h > 0$, we have shown that $h(\boldsymbol{\varphi}^{(k+1)}) \leq h(\boldsymbol{\varphi}^{(k)})$ for all $k \in \mathbb{N}$.

Let $\ell = \lim_{k \to \infty} h(\boldsymbol{\varphi}^{(k)})$. The bounded sequence $(\boldsymbol{\varphi}^{(k)})$ has a converging subsequence $(\boldsymbol{\varphi}^{(k_m)})$ with limit some $\boldsymbol{\varphi}^* \in Y$. The subsequence $(\boldsymbol{\varphi}^{(k_m+1)})$ also converges to some $\widetilde{\boldsymbol{\varphi}} \in Y$. Passing to the limit in (63) we get

$$\widetilde{\boldsymbol{\varphi}} = \frac{\boldsymbol{\varphi}^* - \tau \mathbf{z}^*}{1 + \tau |\mathbf{z}^*|}, \quad \text{with } \mathbf{z}^* = \nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}^*} \right). \tag{66}$$

With the same arguments as above it can be shown that

$$h(\tilde{\varphi}) - h(\boldsymbol{\varphi}^*) \leq (-\alpha + \tau L_h) \left(\|\boldsymbol{z}^*\|_2^2 + \langle \boldsymbol{z}^*, |\boldsymbol{z}^*| \tilde{\varphi} \rangle \right) \leq 0.$$

Moreover, as $\ell = \lim_{m \to \infty} h(\boldsymbol{\varphi}^{(k_m)}) = \lim_{m \to \infty} h(\boldsymbol{\varphi}^{(k_m+1)})$, we deduce that $h(\widetilde{\boldsymbol{\varphi}}) = h(\boldsymbol{\varphi}^*)$. Thus $\|\boldsymbol{z}^*\|_2^2 + \langle \boldsymbol{z}^*, |\boldsymbol{z}^*| \tilde{\varphi} \rangle = 0$, which implies that $\boldsymbol{z}^* + |\boldsymbol{z}^*| \widetilde{\boldsymbol{\varphi}} = 0$. From (66) it follows that $\boldsymbol{\varphi}^* = \widetilde{\boldsymbol{\varphi}}$ and it satisfies the equation

$$\nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}^*} \right) + \left| \nabla \left(\frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \operatorname{div} \boldsymbol{\varphi}^*} \right) \right| \boldsymbol{\varphi}^* = 0,$$

which is the Euler equation for the dual problem. Finally,

$$\lim_{k \to \infty} h(\boldsymbol{\varphi}^{(k)}) = \ell = \min_{\boldsymbol{\varphi} \in S} h(\boldsymbol{\varphi}).$$

□

Remark 4. Notice that the case $\alpha > s_{\min}/4$ can be treated numerically using the following scheme

$$\boldsymbol{\varphi}^{(k+1)} = \frac{\boldsymbol{\varphi}^{(k)} + \tau \boldsymbol{z}^{(k)}}{1 + \tau |\boldsymbol{z}^{(k)}|}, \tag{67}$$

with

$$\boldsymbol{z}^{(k)} = \nabla \left(\max \left\{ \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}^{(k)}}, 0 \right\} \right), \tag{68}$$

which seems to stabilize the previous scheme for large value of α. Indeed, this modified dual approach can be viewed as a projection of the primal variable

$$\boldsymbol{u}^{(k)} = \frac{\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}}{\boldsymbol{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}^{(k)}},$$

on $(\mathbb{R}_+)^J$ (see definition of $\boldsymbol{u}^*[p]$ in (52)).

Remark 5. For the three-dimensional case, the hypotheses of theorem 3.2 becomes $\alpha < s_{\min}/6$ and $\tau \leq \alpha / L_h$ where

$$L_h = 12\alpha^2 \frac{\|\boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)}\|_\infty}{(s_{\min} - 6\alpha)^2}. \tag{69}$$

4. Numerical experiments.

4.1. The algorithms.

In this section, we give some numerical experiments of our global scheme which is expressed as:

- (EM) step from MAP-EM

$$\boldsymbol{\lambda}^{(n+1/2)} = \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}} T^* \left[\frac{\boldsymbol{y}}{T\boldsymbol{\lambda}^{(n)}} \right]$$

- TV minimization

$$\boldsymbol{\lambda}^{(n+1)} \in \underset{\boldsymbol{u} \in (\mathbb{R}_+)^J}{\arg\min} \{ \langle \boldsymbol{u}, \boldsymbol{s} \rangle - \langle \log(\boldsymbol{u}), \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)} \rangle + G(\boldsymbol{u}) \}.$$

We also test a FISTA version. Indeed, in a certain sense, the up-mentioned scheme can be viewed as an ISTA approach ([3]) which alternates minimization of a smooth and non smooth energy, and could further be accelerated by adding a FISTA step. The algorithm becomes :

- ISTA Scheme

$$\widetilde{\boldsymbol{\lambda}}^{(n+1)} \in \underset{\boldsymbol{u} \in (\mathbb{R}_+)^J}{\arg\min} \{ \langle \boldsymbol{u}, \boldsymbol{s} \rangle - \langle \log(\boldsymbol{u}), \boldsymbol{s}\boldsymbol{\lambda}^{(n+1/2)} \rangle + G(\boldsymbol{u}) \},$$

  with $\boldsymbol{\lambda}^{(n+1/2)} = \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}} T^* \left[\frac{\boldsymbol{y}}{T\boldsymbol{\lambda}^{(n)}} \right]$.
- Inertial parameter step

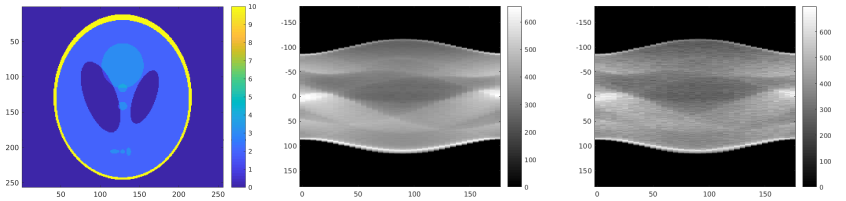
$$t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2},$$

- FISTA update

$$\boldsymbol{\lambda}^{(n+1)} = \widetilde{\boldsymbol{\lambda}}^{(n+1)} + \frac{(t_n - 1)}{t_{n+1}} (\widetilde{\boldsymbol{\lambda}}^{(n+1)} - \widetilde{\boldsymbol{\lambda}}^{(n)})$$

where $t_1 = 1$ and $\widetilde{\boldsymbol{\lambda}}^{(0)} = \mathbf{1}$.

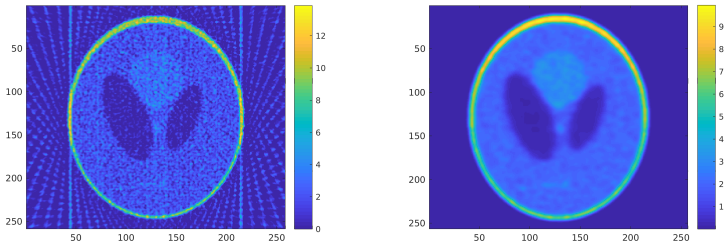
4.2. **Results.** Ideal Radon projections of the Shepp-Logan phantom from Matlab (figure 1.a) sampled in 256^2 pixels, were calculated for angles running from 0° to 175° in steps of 5°. Random values for empirical projections were drawn from a multi-variate Poisson law having mean the theoretical projections. The total number of counts is then about 3×10^6. Ideal projections and their noisy counterparts are shown for comparison in figure 1 (b) and (c).



(a) Phantom (b) Ideal projections (c) Empirical projections

FIGURE 1. (a) Reference figure, the Shepp-Logan phantom and its sinograms: (b) without noise, (c) with Poisson random noise.

In this test, where the number of projections and the number of counts are low, the quality of the image reconstructed without any smoothing is low too. For comparison, in figure 2 we show the analytic reconstruction including the Hamming filter from Matlab and the MLEM reconstruction after 50 iterations and Gaussian post-smoothing.



(a) Analytic (b) MLEM

FIGURE 2. (a) The analytic reconstruction from noisy projections including Hamming filtering and (b) the MLEM solution after 50 iterations and Gaussian smoothing with $\sigma = 2$.

We add a total variation penalty term to the negative log-likelihood and solve numerically for the minimum of the energy with different algorithms mentioned in the paper. The value of the regularization coefficient α is set to 0.025, where this value gave in a subjective analysis the best compromise between smoothness of the homogeneous regions and image contrast. We do not deal in this work with the choice of the regularization parameter and we redirect the reader to specific literature that tackles this task (see *e.g.*, [2, 17, 34]). The noise realizations are different between figures, but identical inside all of the comparison figures. As figures of merit we plot the total energy calculated from equation (10) and the

mean squared error,

$$MSE(\widehat{\boldsymbol{\lambda}}) = \frac{1}{J}\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\|_2. \tag{70}$$

We first compare in figures 3 and 4 the proposed dual method with the Douglas-Rachford splitting described in section §3 after 200 MLEM iterations. The two methods give very similar results but the computing time is significantly larger for Douglas-Rachford due to an additional internal loop (see section §3 for details). In our numerical comparisons the computing time was reduced by a factor of ten with the proposed dual algorithm.

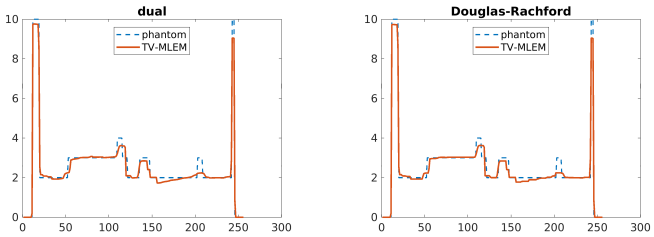


FIGURE 3. Comparison between Douglas-Rachford splitting and dual algorithm. We show the central vertical profiles from the reconstructed images (left and center) and the total energy curves for the two methods (right). The time necessary to run the 200 iterations is 10 times less for the dual algorithm compared to Douglas-Rachford.

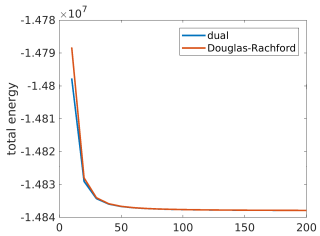


FIGURE 4. Decrease of the cost function with the EM iterations for the Douglas-Rachford splitting and for the dual algorithm. The results are very similar and the mean squared error is not shown since the curves look identical.

The method proposed by Anthoine et al. does not use the capability of the EM algorithm to rapidly increase the likelihood and converges slowly compared to the proposed dual and the Sawatzky et al. methods, as it can be seen in figure 5. After 1000 iterations the reconstructed images and extracted central vertical profiles are relatively close to each other.

In figure 7 we evaluate the influence of the FISTA acceleration technique on the convergence speed of the dual algorithm. After 200 iterations (this number was set arbitrarily for homogeneity with the previous comparisons) the results with and without FISTA are the same. However, the FISTA technique allows to reach

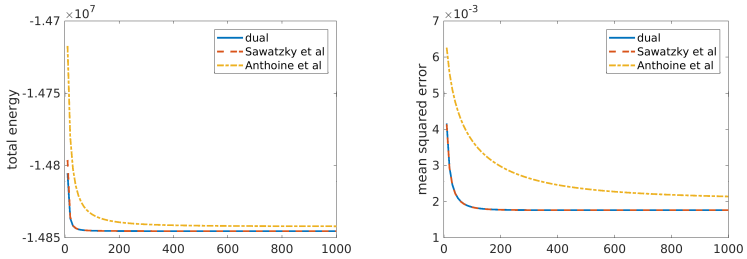


FIGURE 5. Figures of merit for the comparison between the proposed dual method (blue line), Sawatzky et al. method (red dashed-line) and Anthoine et al. method (yellow dotted-line).

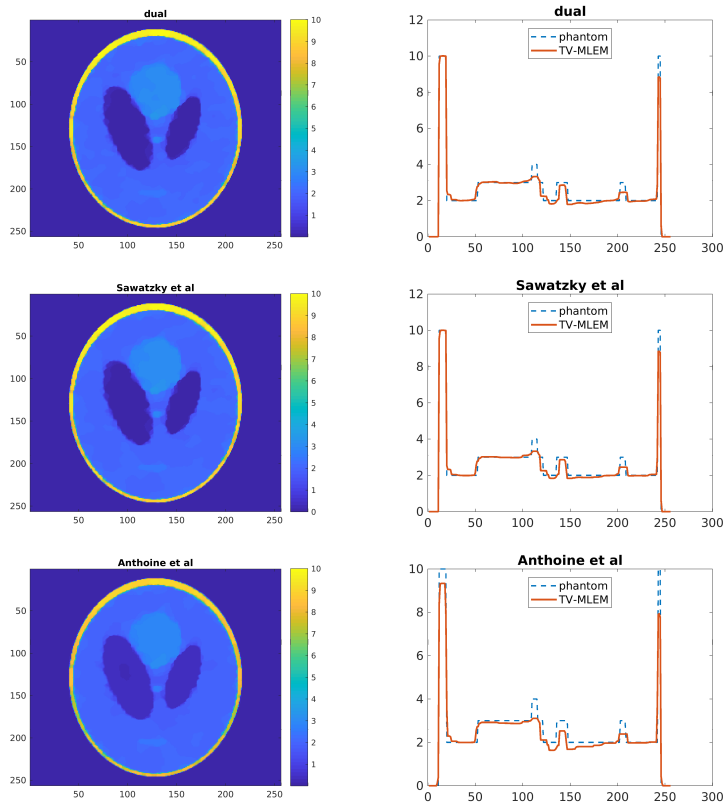


FIGURE 6. Comparison between the proposed dual method (first row), Sawatzky et al. (second row) and Anthoine et al. (third row) methods after 1000 MLEM iterations. We recall that the noise realization is different from figure 3 and thus the two figures should not be compared. The second column shows the central vertical profiles plotted along the images in the first column.

the numerical convergence in about 30 iterations whereas it requires more than 50 iterations for the dual method without FISTA technique.

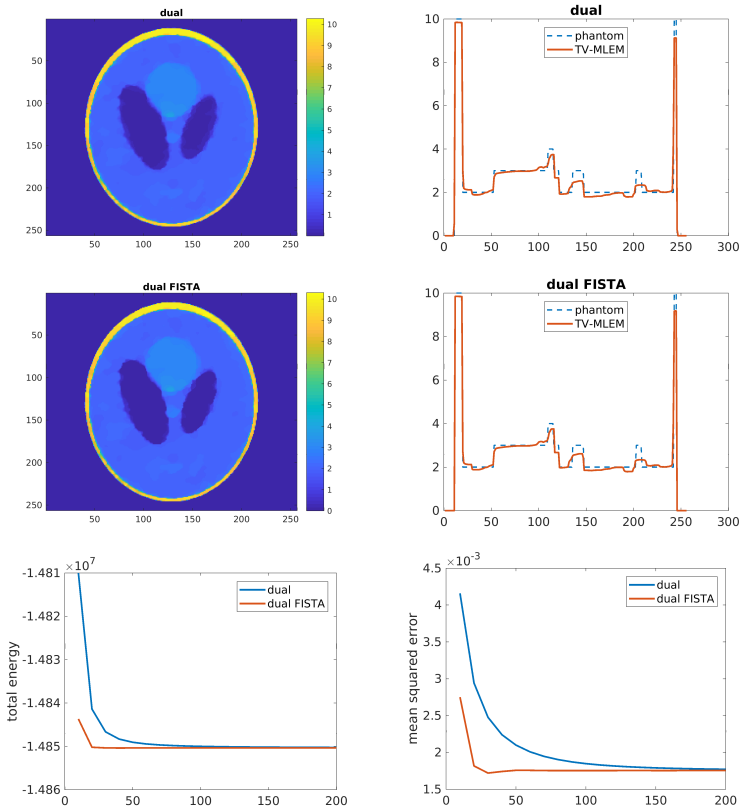


FIGURE 7. FISTA acceleration strongly improves the convergence of the dual algorithm.

5. **Conclusions.** In this paper we adapt the maximum-a-posteriori expectation maximization framework to non-smooth convex priors, aiming to maximize an energy composed of a likelihood function and a prior distribution. Its specificity is to split the search for the optimal value in an expectation step that allows to move from an optimization problem in the domain of incomplete data (in our case the projections) to a simpler one in the domain of complete data (in our case the image), and a maximization step for the new criterion. We then deduce consistency and convergence results for the MAP-EM algorithms. Total variation regularization of tomographic images calculated from Poisson distributed data is then derived from as a particular case. For the maximization step we lean on the Fenchel-Rockafellar duality principle and we propose a simple and efficient algorithm developed following the ideas that A. Chambolle first introduced for fidelity terms expressed with the ℓ_2-induced distance. We then succeed to prove its convergence to the solution at least for regularization parameters that do not exceed a given upper bound. The resulting MAP-EM algorithm is both consistent and relatively fast and successfully

competes experimentally with other algorithms proposed in the literature for the resolution of the same problem. Our results also tend to show that FISTA acceleration allows to further improve the convergence speed although we are aware that our setup is different from the one where FISTA was originally proposed.

REFERENCES

[1] S. Anthoine, J.-F. Aujol, Y. Boursier and C. Melot, Some proximal methods for Poisson intensity CBCT and PET, *Inverse problems and Imaging*, **6** (2012), 565–598.

[2] J. Bardsley and J. Goldes, Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation, *Inverse Problems*, **25** (2009), 095005, URL http://stacks.iop.org/0266-5611/25/i=9/a=095005.

[3] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM journal on imaging sciences*, **2** (2009), 183–202.

[4] A. Beck and M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, *Trans. Img. Proc.*, **18** (2009), 2419–2434, URL http://dx.doi.org/10.1109/TIP.2009.2028250.

[5] J. Browne and A. De Pierro, A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography, *IEEE transactions on medical imaging*, **15** (1996), 687–699.

[6] A. Chambolle, An algorithm for total variation minimization and applications, *Journal of Mathematical imaging and vision*, **20** (2004), 89–97.

[7] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *Journal of Mathematical Imaging and Vision*, **40** (2011), 120–145, URL https://doi.org/10.1007/s10851-010-0251-1.

[8] A. Dempster, N. Laird and D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, **39** (1977), 1–38.

[9] J. A. Fessler and W. L. Rogers, Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs, *IEEE Transactions on Image processing*, **5** (1996), 1346–1358.

[10] E. Garduno and G. T. Herman, Superiorization of the ML-EM algorithm, *IEEE Transactions on Nuclear Science*, **61** (2014), 162–172.

[11] P. J. Green, Bayesian reconstructions from emission tomography data using a modified EM algorithm, *IEEE Transactions on Medical Imaging*, **9** (1990), 84–93.

[12] H. Hudson and R. Larkin, Accelerated image reconstruction using ordered subsets of projection data, *IEEE transactions on medical imaging*, **13** (1994), 601–609.

[13] A. N. Iusem, A short convergence proof of the EM algorithm for a specific Poisson model, *Brazilian Journal of Probability and Statistics*, **6** (1992), 57–67, URL http://www.jstor.org/stable/43601445.

[14] L. Kaufman, Implementing and accelerating the EM algorithm for positron emission tomography, *IEEE transactions on medical imaging*, **6** (1987), 37–51.

[15] K. Lange and R. Carson, EM reconstruction algorithms for emission and transmission tomography, *J Comput Assist Tomogr*, **8** (1984), 306–316.

[16] E. Levitan and G. T. Herman, A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography, *IEEE Transactions on Medical Imaging*, **6** (1987), 185–192.

[17] F. Lucka, K. Proksch, C. Brune, N. Bissantz, M. Burger, H. Dette and F. Wübbeling, Risk estimators for choosing regularization parameters in ill-posed problems - properties and limitations, *Inverse Problems & Imaging*, **12** (2018), 1121–1155.

[18] A. Mikhno, E. D. Angelini, B. Bai and A. F. Laine, Locally weighted total variation denoising for ringing artifact suppression in pet reconstruction using psf modeling, in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, IEEE, 2013, 1252–1255.

[19] J. Nuyts and J. A. Fessler, A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution, *IEEE transactions on medical imaging*, **22** (2003), 1042–1052.

[20] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, vol. 30, Siam, 1970.

[21] V. Y. Panin, G. L. Zeng and G. T. Gullberg, Total variation regulated EM algorithm, *IEEE Transactions on Nuclear Science*, **46** (1999), 2202–2210.
[22] M. Persson, D. Bone and H. Elmqvist, Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography, *Physics in Medicine & Biology*, **46** (2001), 853, URL http://stacks.iop.org/0031-9155/46/i=3/a=318.
[23] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D: nonlinear phenomena*, **60** (1992), 259–268.
[24] A. Sawatzky, C. Brune, F. Wubbeling, T. Kosters, K. Schafers and M. Burger, Accurate EM-TV algorithm in PET with low SNR, in *Nuclear Science Symposium Conference Record, 2008. NSS'08. IEEE*, IEEE, 2008, 5133–5137.
[25] A. Sawatzky, C. Brune, T. Koesters, F. Wuebbeling and M. Burger, Em-tv methods for inverse problems with poisson noise, in *Level set and PDE based reconstruction methods in imaging*, Springer, 2013, 71–142.
[26] E. Y. Sidky, C.-M. Kao and X. Pan, Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT, *Journal of X-ray Science and Technology*, **14** (2006), 119–139.
[27] B. Silverman, M. Jones, J. Wilson and D. Nychka, A smoothed EM approach to indirect estimation problems, with particular, reference to stereology and emission tomography, *Journal of the Royal Statistical Society. Series B (Methodological)*, **52** (1990), 271–324.
[28] A. Sitek, Representation of photon limited data in emission tomography using origin ensembles, *Physics in Medicine & Biology*, **53** (2008), 3201.
[29] D. L. Snyder and M. I. Miller, The use of sieves to stabilize images produced with the EM algorithm for emission tomography, *IEEE Transactions on Nuclear Science*, **32** (1985), 3864–3872.
[30] D. L. Snyder, M. I. Miller, L. J. Thomas and D. G. Politte, Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography, *IEEE Transactions on Medical Imaging*, **6** (1987), 228–238.
[31] S. Stute and C. Comtat, Practical considerations for image-based PSF and blobs reconstruction in PET, *Physics in medicine and biology*, **58** (2013), 3849.
[32] Y. Vardi, L. Shepp and L. Kaufman, A statistical model for positron emission tomography, *Journal of the American Statistical Association*, **80** (1985), 8–20.
[33] E. Veklerov, J. Llacer and E. Hoffman, MLE reconstruction of a brain phantom using a Monte Carlo transition matrix and a statistical stopping rule, *IEEE Transactions on Nuclear Science*, **35** (1988), 603–607.
[34] Y.-W. Wen and R. H. Chan, Using generalized cross validation to select regularization parameter for total variation regularization problems, *Inverse Problems & Imaging*, **12** (2018), 1103–1120.
[35] M. Yan, J. Chen, L. A. Vese, J. Villasenor, A. Bui and J. Cong, *EM+TV Based Reconstruction for Cone-Beam CT with Reduced Radiation*, 1–10, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, URL https://doi.org/10.1007/978-3-642-24028-7_1.

E-mail address: maxim@creatis.insa-lyon.fr
E-mail address: feng@creatis.insa-lyon.fr
E-mail address: banjak@creatis.insa-lyon.fr
E-mail address: elie.bretin@insa-lyon.fr

Ultramicroscopy 189 (2018) 109–123

Contents lists available at ScienceDirect

Ultramicroscopy

journal homepage: www.elsevier.com/locate/ultramic

Evaluation of noise and blur effects with SIRT-FISTA-TV reconstruction algorithm: Application to fast environmental transmission electron tomography

Hussein Banjak[a], Thomas Grenier[a], Thierry Epicier[b,*], Siddardha Koneti[b], Lucian Roiban[b], Anne-Sophie Gay[c], Isabelle Magnin[a], Françoise Peyrin[a], Voichita Maxim[a]

[a] Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CREATIS, CNRS UMR 5220 INSERM U1206, 69621 Villeurbanne Cedex, France
[b] MATEIS, UMR 5510, CNRS, Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, 69621 Villeurbanne Cedex, France
[c] IFP Energies nouvelles, B.P. 3, rond-point de l'échangeur de Solaize, F-69360 Solaize, France

ARTICLE INFO

Article history:
Received 19 December 2017
Revised 21 March 2018
Accepted 28 March 2018

Keywords:
Electron tomography
Iterative reconstruction
Environmental TEM
Nanocatalysts
Fast acquisition

ABSTRACT

Fast tomography in Environmental Transmission Electron Microscopy (ETEM) is of a great interest for in situ experiments where it allows to observe 3D real-time evolution of nanomaterials under operating conditions. In this context, we are working on speeding up the acquisition step to a few seconds mainly with applications on nanocatalysts. In order to accomplish such rapid acquisitions of the required tilt series of projections, a modern 4K high-speed camera is used, that can capture up to 100 images per second in a 2K binning mode. However, due to the fast rotation of the sample during the tilt procedure, noise and blur effects may occur in many projections which in turn would lead to poor quality reconstructions. Blurred projections make classical reconstruction algorithms inappropriate and require the use of prior information. In this work, a regularized algebraic reconstruction algorithm named SIRT-FISTA-TV is proposed. The performance of this algorithm using blurred data is studied by means of a numerical blur introduced into simulated images series to mimic possible mechanical instabilities/drifts during fast acquisitions. We also present reconstruction results from noisy data to show the robustness of the algorithm to noise. Finally, we show reconstructions with experimental datasets and we demonstrate the interest of fast tomography with an ultra-fast acquisition performed under environmental conditions, i.e. gas and temperature, in the ETEM. Compared to classically used SIRT and SART approaches, our proposed SIRT-FISTA-TV reconstruction algorithm provides higher quality tomograms allowing easier segmentation of the reconstructed volume for a better final processing and analysis.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Electron Tomography (ET) is a powerful high-resolution imaging tool for visualization of three-dimensional (3D) structure of objects down to the atomic level [1,2]. This technique has been first applied using conventional bright field imaging in biological sciences [3] and then in the field of material sciences since the late 1980s. ET reconstructs the 3D volume of the sample from a tilt series of 2D projections acquired in the Transmission Electron Microscope (TEM) at different viewing angles around a tilt axis. For an ideal reconstruction, a full $+/-90^{\circ}$ tilt angle range is required. Usually, the sample cannot be tilted at high angles due to shadowing effects caused by the sample holder or the sample shape itself and space limitation in the chamber. The tilt angle range is typically limited to $+/-70^{\circ}$ and the remaining missing angular range is commonly called missing wedge. The projection series is first aligned and then combined to derive the 3D structure of the object either using an analytical reconstruction algorithm like the Weighted Back-Projection (WBP) [4] algorithm or an iterative one like the Algebraic Reconstruction Technique (ART) [5], the Simultaneous Algebraic Reconstruction Technique (SART) [6] and the Simultaneous Iterative Reconstruction Technique (SIRT) [7]. Due to the missing wedge problem, image reconstruction in ET is considered as an ill-posed problem and suffers from severe artifacts when using classical reconstruction algorithms. Recently, reconstruction algorithms based on compressive sensing have been used in ET to reduce missing wedge artifacts in the final reconstruction [8]. These compressed sensing reconstruction methods require image sparsity as prior knowledge and that is usually provided

* Corresponding author.
E-mail address: thierry.epicier@insa-lyon.fr (T. Epicier).

https://doi.org/10.1016/j.ultramic.2018.03.022
0304-3991/© 2018 Elsevier B.V. All rights reserved.

by the image gradient. The most popular regularization approach within ET has been Total Variation (TV) regularization [9]. In addition to these sparsity-based reconstruction methods, another advanced regularization technique named Discrete Algebraic Reconstruction Technique (DART) has been introduced [10]. This method combines reconstruction with segmentation by exploiting a prior knowledge on the gray levels of the object and can produce accurate reconstruction results under extremely limited data conditions like limited-angle data and a few number of projections.

Considering the current development of the so-called Environmental Transmission Electron Microscopy (ETEM), applying ET under environmental (varying gas and temperature) conditions would be a breakthrough to follow in 3D dynamic events during in situ/operando observations of nanocatalysts and to study their morphological evolution almost in real time during chemical reactions. In contrast to conventional TEM that requires high vacuum conditions, environmental TEM microscopes are designed to maintain some partial pressure in the specimen area [11]. ETEM is very well adapted to the study of nanocatalysts: a typical example which will be further illustrated in the present work concerns the depollution of exhaust gases from diesel automotive engines [12]. A catalyst (here, zirconia, i.e. zirconium dioxide ZrO_2) is used to reduce carbon soot emissions which are dangerous for both human health and environment; it helps burning carbon particles by favouring their oxidation at low temperature. Whereas in situ ETEM observations could be performed in almost real time in 2D conditions, following the soot combustion in 3D requires to speed up the tomographic approach and to acquire tilt series in a few seconds. Such speeds cannot be reached with the conventional acquisition procedure where images are acquired during pauses of the rotation performed step-by-step with an incremental angular step of typically 1 or 2°. Usual TEM cameras with low readout speeds are also severely limiting the speed of acquisition. Consequently, the goniometer has to be continuously rotated at its fastest speed to highly accelerate the acquisition of tilted series for a very short total time of acquisition of the order of five seconds or even less [13,14]. Simultaneously, fast recording is made possible with the availability of modern cameras like the OneView camera from GATAN (www.gatan.com) that can record at least 25 frames per second (fps) with a 4K × 4K image resolution and up to 300 fps with a 512 × 512 image resolution. Intuitively, with this high rotation speed, a rotational blur is most probably induced. Additionally, the object moves during rotation in the microscope because of mechanical instabilities of the goniometer system which causes a translational blur in the 2D projections during acquisition. Thus, these are the main concerns of fast ETEM that we address here. More precisely, we evaluate the influence of blur effects on the final quality of reconstruction obtained with our proposed GPU-based reconstruction toolbox.

Our developed reconstruction toolbox consists of algebraic iterative algorithms (SART and SIRT) with total variation (TV) [15] penalization incorporated as a regularization step to suppress most noise and handle blur effects. In addition, it integrates a Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [16]. Firstly, numerical simulations are performed using a 3D model to evaluate the robustness of the algorithm to noisy data. The blurring effect is also studied by means of these numerical simulations to evaluate the influence of the goniometer rotation speed during image acquisition on 3D reconstruction quality. Finally, we report the first sets of experimental data collected in a few seconds on nanocatalytic systems studied in 3D in the environmental TEM.

This paper is organized as follows. In the second section, we give an overview of the structure of our proposed reconstruction algorithm named SIRT-FISTA-TV. Numerical comparisons with synthetic noisy and blurred data are presented in Section 3 to test the performance of the algorithm. Reconstruction results from experimental tomographic tilt series are presented in Section 4, and finally conclusions and perspectives are given in Section 5.

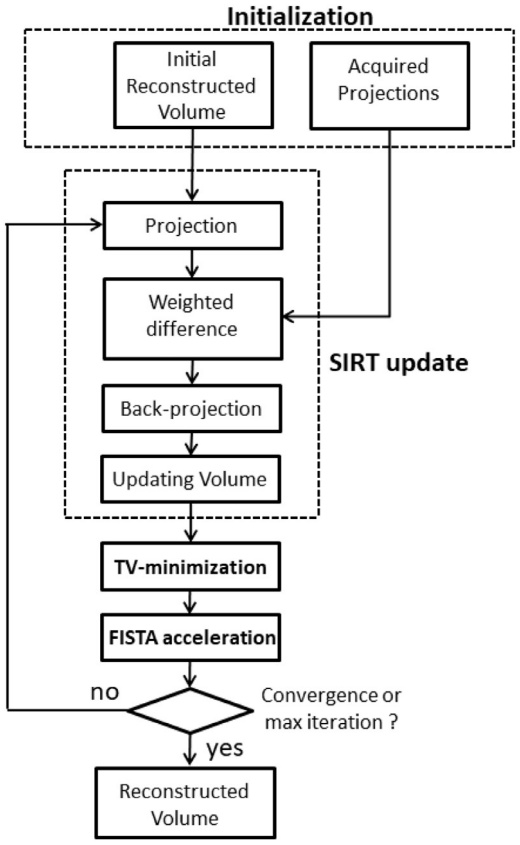


Fig. 1. Schematic view of SIRT-FISTA-TV reconstruction process.

2. Description of reconstruction algorithm: SIRT-FISTA-TV

We present in this section a regularized algebraic reconstruction technique named SIRT-FISTA-TV. Algebraic methods represent the reconstruction problem as a linear system of equations $\mathbf{p} = A.\mathbf{f}$ where $\mathbf{p}$ refers to projection data, A is the projection system model and $\mathbf{f}$ defines the 3D object that we aim to reconstruct. These methods tend to solve the system by minimizing the ℓ_2 error $\|A.\mathbf{f} - \mathbf{p}\|$ where direct inversion of the system is extremely time consuming and memory intensive since the matrix A is very large. As illustrated in Fig. 1, the global iterative process of SIRT-FISTA-TV consists of three main steps: i) classical SIRT reconstruction, ii) TV minimization and iii) volume update following FISTA technique. In the following subsections we will explain each step in details and we will present the pseudo-code of the algorithm in the last subsection.

2.1. SIRT algorithm

In electron tomography, SIRT [7] is a commonly used reconstruction algorithm that tends to be more stable than ART in case of noisy projections. Its structure is described by the following equation

$$f_j^{(k)} = f_j^{(k-1)} - \lambda \cdot \frac{\sum_{i=1}^{M} \frac{\sum_{n=1}^{N} a_{in} f_n^{(k-1)} - p_i}{\sum_{n=1}^{N} a_{in}} \cdot a_{ij}}{\sum_{i=1}^{M} a_{ij}}, \quad (1)$$

where a_{ij} represents an element of the matrix A with i the index of a pixel in the projection $\mathbf{p}$ and j the index of a voxel in the image

f. Here, N is the total number of voxels in **f**, M is the total number of pixels in **p**, k is the iteration number and λ is a relaxation parameter.

As illustrated in Fig. 1, each iteration of SIRT consists of three phases: (i) projection of the current volume to produce synthesized projections, (ii) computing the difference between synthesized projections and experimentally acquired projections and (iii) back-projection of the weighted difference to update the volume. Note that Eq. (1) may be also written as

$$\mathbf{f}^{(k)} = \mathbf{f}^{(k-1)} - \lambda \cdot \frac{1}{A^*\mathbf{1}} A^* \left[\frac{A\mathbf{f}^{(k-1)} - \mathbf{p}}{A\mathbf{1}} \right], \tag{2}$$

where A^* is the adjoint (or conjugate transpose) of A and $\mathbf{1}$ is a column vector of ones. SIRT is thus a weighted version of the gradient descent algorithm for the minimization of $\|A.\mathbf{f} - \mathbf{p}\|_{l_2}^2$ which is expressed as

$$\mathbf{f}^{(k)} = \mathbf{f}^{(k-1)} - \lambda A^*(A\mathbf{f}^{(k-1)} - \mathbf{p}). \tag{3}$$

2.2. TV minimization

TV minimization is a regularization method that can remove noise while preserving the edges of the image. This technique, originally proposed for image denoising [15], has been widely used in computed tomography and significantly reduces artifacts for image reconstruction with few-views or missing wedges [17]. The TV optimization problem consists to solve

$$\hat{\mathbf{f}} = \underset{\mathbf{f}\in\mathbb{R}^N}{\operatorname{argmin}} \left\{ \|A.\mathbf{f} - \mathbf{p}\|_{l_2}^2 + \mu \|\mathbf{f}\|_{TV} \right\}, \tag{4}$$

where the positive parameter μ controls the degree of smoothness and $\|\mathbf{f}\|_{TV}$ is the l_1-norm of the image gradient magnitudes $|\nabla f|_{i,\,j,\,k}$. This penalty function is expressed by

$$\|\mathbf{f}\|_{TV} = \sum_{i,j,k} \sqrt{(f_{i,j,k} - f_{i-1,j,k})^2 + (f_{i,j,k} - f_{i,j-1,k})^2 + (f_{i,j,k} - f_{i,j,k-1})^2} \tag{5}$$

where $f_{i,\,j,\,k}$ is the value of the voxel with index $(i,\,j,\,k)$.

In this paper we use a proximal forward-backward iterative scheme with improved convergence rate called FISTA [16], for general optimization problems

$$\min_{\mathbf{f}\in\mathbb{R}^N} \{\phi(\mathbf{f}) + \psi(\mathbf{f})\}, \tag{6}$$

where ψ is a continuous convex function possibly nonsmooth as the TV norm (5) and ϕ is a convex function continuously differentiable with Lipschitz continuous gradient $L(\phi)$. In our case, $\phi(\mathbf{f}) = \|A.\mathbf{f} - \mathbf{p}\|_{l_2}^2$ and $\psi(\mathbf{f}) = \mu\|\mathbf{f}\|_{TV}$. FISTA stands for Fast ISTA and the optimization step inherited from ISTA consists to split problem (6) into a gradient descent step on the smooth function ϕ,

$$\mathbf{g}^{(k)} = \mathbf{f}^{(k-1)} - \frac{1}{L}\nabla\phi(\mathbf{f}^{(k-1)}), \tag{7}$$

for some step size $L > 0$, followed by a search in the vicinity of this new point for a better estimation of the minimum of (6) that also reduces the value of ψ. This new estimation is calculated as

$$\mathbf{f}^{(k)} = \underset{\mathbf{f}\in\mathbb{R}^N}{\operatorname{argmin}} \left\{ \psi(\mathbf{f}) + \frac{L}{2}\|\mathbf{f} - \mathbf{g}^{(k)}\|_{l_2}^2 \right\}. \tag{8}$$

In convex analysis notation, $\mathbf{f}^{(k)} = \operatorname{prox}_{\psi/L}(\mathbf{g}^{(k)})$. Instead of the gradient descent step from (7), we will use the SIRT update step (1) with $\lambda = 1/L$ and L bounded below by $L(\phi)$, the spectral norm of A^*A.

With the usual trick consisting to add a small positive constant ε to the gradient from (5), the TV-norm becomes differentiable and a gradient descent method is used to solve (8). Thus, there are two levels for the iteration performed in this algorithm: (k) defining the overall iteration number and (n) defining the sub-iteration number in the gradient descent step. The current image $\mathbf{f}^{(k,\,n)}$ is updated at each iteration (n) of the gradient step method as follows:

$$\mathbf{f}^{(k,n)} = \mathbf{f}^{(k,n-1)} - \alpha \cdot d_p^{(k)} \cdot \frac{\mathbf{v}^{(k,n-1)}}{\|\mathbf{v}^{(k,n-1)}\|_{l_2}}, \tag{9}$$

with

$$d_p^{(k)} = \|\mathbf{f}^{(k)} - \mathbf{f}^{(k-1)}\|_{l_2} \tag{10}$$

and

$$\mathbf{v}^{(k,n-1)} = \left. \frac{d\|\mathbf{f}\|_{TV}}{d\mathbf{f}} \right|_{\mathbf{f}=\mathbf{f}^{(k,n-1)}}. \tag{11}$$

Here, α is a positive parameter controlling the extent of smoothing. The value of $d_p^{(k)}$ is computed initially and depends on the values of $\mathbf{f}^{(k-1)}$ and $\mathbf{f}^{(k)}$ corresponding to the image values before and after the SIRT update step, respectively, during overall iteration (k).

2.3. FISTA acceleration

FISTA [16] is an acceleration method used to speed up the algorithm convergence and consists to replace $\mathbf{f}^{(k-1)}$ in (7) with a very specific linear combination of the previous two estimates $\bar{\mathbf{f}}^{(k-1)}$ and $\bar{\mathbf{f}}^{(k-2)}$, where $\bar{\mathbf{f}}^{(k)}$ denotes from now on the solution of (8). The new value of $\mathbf{f}^{(k-1)}$ becomes

$$\mathbf{f}^{(k)} = \bar{\mathbf{f}}^{(k)} + \left(\frac{t^{(k-1)} - 1}{t^{(k)}} \right) \left(\bar{\mathbf{f}}^{(k)} - \bar{\mathbf{f}}^{(k-1)} \right), \tag{12}$$

where $\bar{\mathbf{f}}^{(k)}$ is the image obtained after TV-minimization during iteration (k). The parameter t is updated at each iteration as follows

$$t^{(k)} = \frac{1 + \sqrt{1 + 4\left(t^{(k-1)}\right)^2}}{2}, \tag{13}$$

where $t^{(0)} = 1$ is the initial given value. With this technique, the total number of iterations used to produce an image of similar quality is potentially reduced compared to ISTA algorithm and thus to SIRT-TV. As proved in [16], the complexity rate of FISTA is of $O(1/k^2)$ (with k the iteration number) to compare to a worst-case complexity of $O(1/k)$ for ISTA.

Different methods have been developed to solve optimization problems involving TV regularization, in particular in tomography. The method described in this paper is more closely related to the one from [17], where the ART algorithm was used without FISTA acceleration. In a previous work [9], a method based on the alternating minimization of the augmented Lagrangian function was proposed with application in ET. However our method requires significantly lower memory and a reduced number of parameters that normally do not vary significantly from one experiment to the other.

2.4. Pseudo-code of SIRT-FISTA-TV

The pseudo-code for our proposed SIRT-FISTA-TV algorithm is listed in Table 1.

Thus, this algorithm is easy to use where there are only four user-defined parameters: i) SIRT relaxation parameter (λ), ii) overall number of iterations (Niter), iii) TV regularization parameter (α) and iv) number of iterations for TV (NTViter). A standard choice of these parameters is: $\lambda = 0.2$, $\alpha = 0.2$, Niter $= 10$ and NTViter $= 10$. However, TV regularization parameters may change depending on noise level in projection data.

Table 1
Pseudo-code of SIRT-FISTA-TV reconstruction algorithm.

Initialization:
$f^{(0)} \leftarrow 0$, Niter $\leftarrow 50$, NTViter $\leftarrow 10$,
$\lambda \leftarrow 0.2$, $\alpha \leftarrow 0.2$, $t^{(0)} \leftarrow 1$
SIRT-FISTA-TV iterations: 3-19
for k=1,2,...,Niter **do**
  SIRT update: 5-8
  $f^{(k)} = f^{(k-1)} - \lambda \cdot \frac{1}{A^*\mathbf{1}} A^* \left[\frac{Af^{(k-1)} - p}{A\mathbf{1}} \right]$
  if $f_j^{(k)} < 0$ **then**
    $f_j^{(k)} \leftarrow 0$ (positivity constraint)
  end if
  TV minimization: 10-14
  Initialization:
  $f^{(k,0)} \leftarrow f^{(k)}$, $d_p^{(k)} = \|f^{(k)} - f^{(k-1)}\|_{l_2}$
  for n=1,2,...,NTViter **do**
    $v^{(k,n-1)} = \frac{d\|f\|_{TV}}{df}\Big|_{f=f^{(k,n-1)}}$
    $f^{(k,n)} = f^{(k,n-1)} - \alpha \cdot d_p^{(k)} \cdot \frac{v^{(k,n-1)}}{\|v^{(k,n-1)}\|_{l_2}}$
  end for
  FISTA acceleration: 16-18
  Initialization:
  $\bar{f}^{(k)} \leftarrow f^{(k,NTViter)}$
  $t^{(k)} = \frac{1 + \sqrt{1 + 4\left(t^{(k-1)}\right)^2}}{2}$
  $f^{(k)} = \bar{f}^{(k)} + \left(\frac{t^{(k-1)} - 1}{t^{(k)}} \right) \left(\bar{f}^{(k)} - f^{(k-1)} \right)$
end for

3. Performance evaluation of SIRT-FISTA-TV using noisy and blurred simulated data

In this section, we first show a numerical validation of SIRT-FISTA-TV then we evaluate its performance with noisy and blurred data. The numerical comparisons are based on data simulated from a 3D model of a grey cylinder with white spherical and dark cuboid inclusions of different sizes. Fig. 2 shows 3D view and cross-sections (in XY- and XZ- planes) of this model. It intends to represent a composite nanomaterial made of an homogeneous matrix containing pores and particles; no intensity variation was applied to any of these three phases: matrix, pores, nanoparticles during the tilting procedure.

The computation was carried out on a computer under Windows 7 with 128GB of RAM and an Intel Xeon 3.5 GHz processor. Our proposed SIRT-FISTA-TV algorithm was coded in MATLAB with Graphics Processing Unit (GPU) implementations using C-CUDA on Nvidia Quadro M5000 card to highly accelerate the reconstruction algorithm. The projection and back-projection operations of ASTRA toolbox [18] are used. This projection-operation is based on the Joseph interpolation kernel [19] and the backprojection-operation uses a pixel driven method [20] with linear interpolation. In our implementation, we follow the method of Bleichrodt et al. [21] which uses another MATLAB toolbox named Spot [22] and combines it with the ASTRA toolbox in order to represent the problem in MATLAB in matrix notation. For numerical validation of the proposed algorithm, noise-free projections were obtained over a full tilt range of ±90° in order to avoid any missing wedge effect and to ensure a perfect reconstruction. The angular increment is 0.36°, resulting in 500 projections. Each 2D projection is of size $M = 256 \times 256$ pixels and the reconstructed volume is of size $N = 256 \times 256 \times 512$ voxels. In what follows, we apply 200 iterations in order to ensure that convergence is reached. The reconstructed values are enforced to be positive in each iteration. The relaxation parameter of SIRT update step is always set to $\lambda = 0.2$ in our experiments. For TV regularization, there are two parameters to define: i) the number of TV outer iterations which is always set to 10 and ii) the TV relaxation parameter set to $\alpha = 0.2$. In this simulation, the total reconstruction time for SIRT is 5 minutes and for SIRT-FISTA-TV is 83 minutes. The reconstruction results of SIRT-FISTA-TV are given in Fig. 3 that shows cross-sections in XY- and XZ- planes and the histograms of the white inclusions size distribution for original (first row) and reconstructed (second row) volumes. We focused here on white spherical inclusions because their analysis is straightforward through a single parameter, i.e. their diameter, allowing to compare easily the quality of our reconstructions under various conditions. The two cross-sections in the reconstructed volume look identical with the original ones and the two histograms are in excellent agreement.

The value of peak signal-to-noise ratio (PSNR) obtained is 30.6decibels (dB). The PSNR value is given by

$$\mathrm{PSNR}(\hat{f}) = 10 \log \left(\frac{DR^2(f)}{MSE(\hat{f})} \right), \tag{14}$$

where DR is the image dynamic range and MSE is the mean Euclidean distance between reconstructed volume $\hat{f}$ and original one f and is defined by:

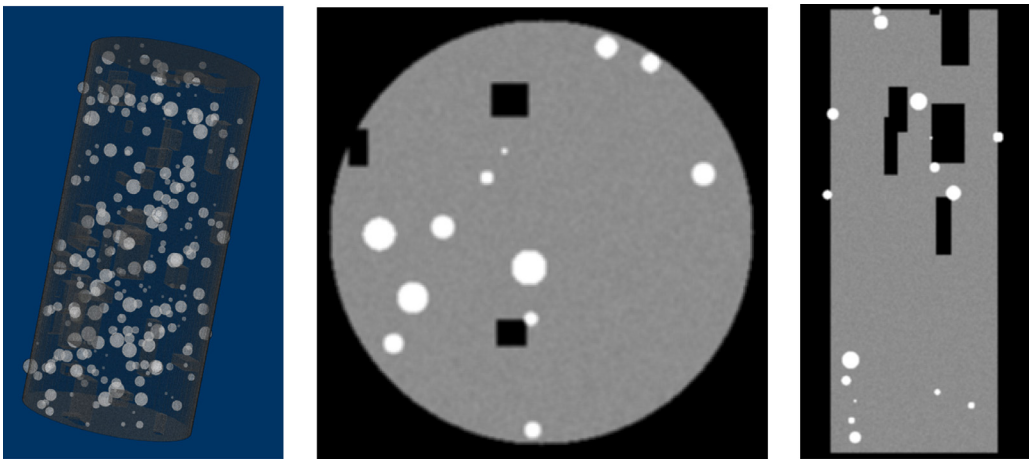
$$\mathrm{MSE}(\hat{f}) = \frac{1}{N} \sum_{j=1}^{N} \left(\hat{f}_j - f_j \right)^2, \tag{15}$$

with N is the total number of voxels in f and $\hat{f}$.

The volume is also reconstructed with SIRT-TV algorithm in order to show the advantages of FISTA acceleration technique. Fig. 4 displays PSNR values of volumes reconstructed by both SIRT-FISTA-TV (blue curve) and SIRT-TV (red curve) as function of different number of iterations. It is clearly demonstrated that SIRT-FISTA-TV performs better than SIRT-TV algorithm and improves the PSNR value by around 4 dB. SIRT-FISTA-TV is then tested with noisy and blurred data and the reconstruction results are given in the following two subsections.

3.1. Reconstruction with noisy data

We add to projections a Gaussian noise of zero mean and standard deviation $\sigma = 2\%$ of the maximum intensity in noiseless data. In this simulation, only 70 projections were generated with a tilting angle range from $-70°$ to $+70°$ in order to mimic experimental acquisitions. The reconstruction results of SIRT-FISTA-TV are compared with that of SIRT combined with different filtering techniques: median, Gaussian and anisotropic diffusion. These filtering methods are used to denoise projections prior to reconstruction. Anisotropic diffusion is a filter that tends to remove noise while maintaining image edges. This filter was proposed in 1990 [23] and controls the rate of diffusion as a function of image gradient. For example, there is no diffusion or reduced diffusion in regions where the gradient is high, whereas there is more diffusion



(a) 3D view (b) XY cross-section (c) XZ cross-section

Fig. 2. (a) 3D ghost model used to carry on numerical simulations and evaluate the performance of the developed algorithm.

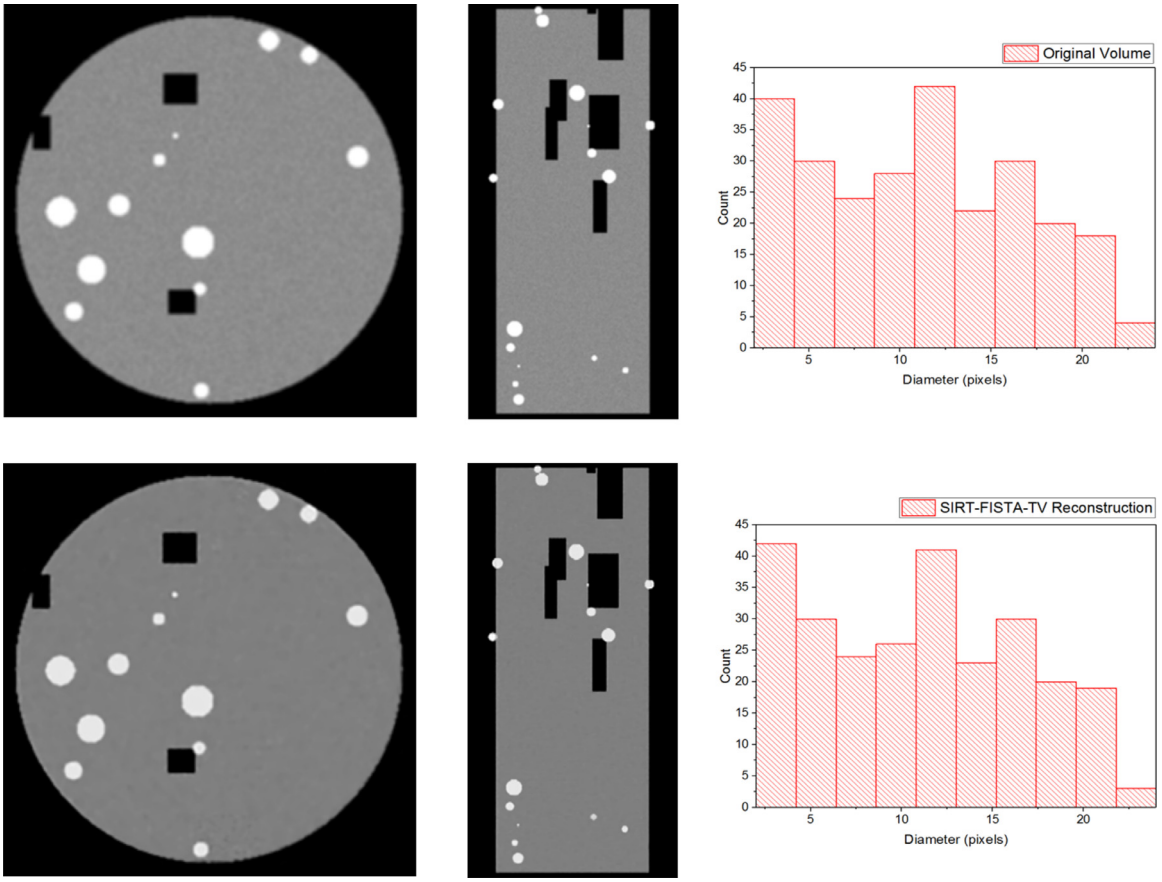


Fig. 3. SIRT-FISTA-TV reconstruction from complete and noiseless data. The first and second rows refer to original and reconstructed volume, respectively. Central cross-sections in XY- and XZ- planes are displayed in the first and second columns, respectively, and histograms for distribution of white inclusion diameters are shown in the third column.

in certain directions where the gradient value is small. Gaussian filter is often effective in reducing noise but it is a low pass filter and is expected to blur edges. Median filter is a non-linear filter that preserves edges better than Gaussian filter. However, its performance is not that much better than Gaussian filter for high levels of noise. The choice of parameters for these filtering methods was exhaustively optimized according to mean-square-error (MSE) values. It is observed in the XY-plane cross-section of each reconstructed volume shown in Fig. 5, that the noise is significantly suppressed with SIRT-FISTA-TV compared to SIRT and spatial resolution is preserved. For the other SIRT reconstructions with data filtered prior to reconstruction, noise artifacts are reduced but the edges are smoothed and spatial resolution is reduced.

To further evaluate the robustness of SIRT-FISTA-TV to noisy data, Gaussian noise of various standard deviations was added to projections. The graph displayed in Fig. 6 plots PSNR against standard deviation of added noise. As expected, this graph shows that PSNR decreases as the noise level increases. We can see that the red curve which refers to reconstruction from SIRT-FISTA-TV gives the highest PSNR value compared to the other curves that corre-

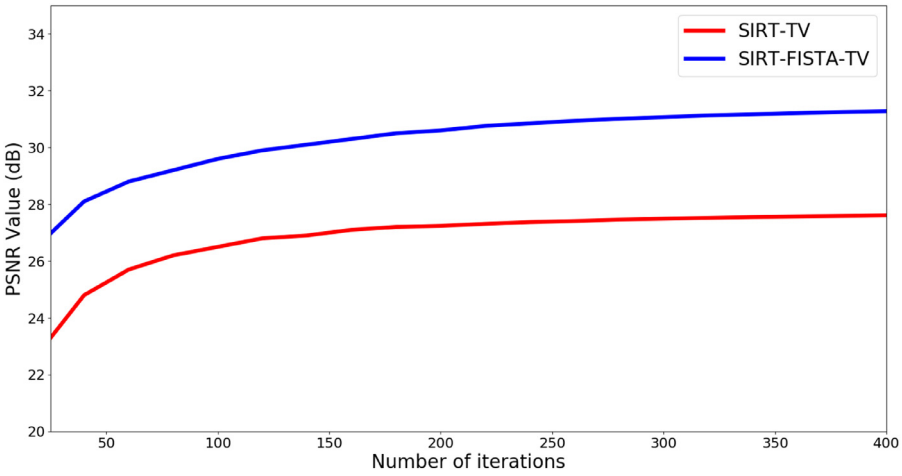
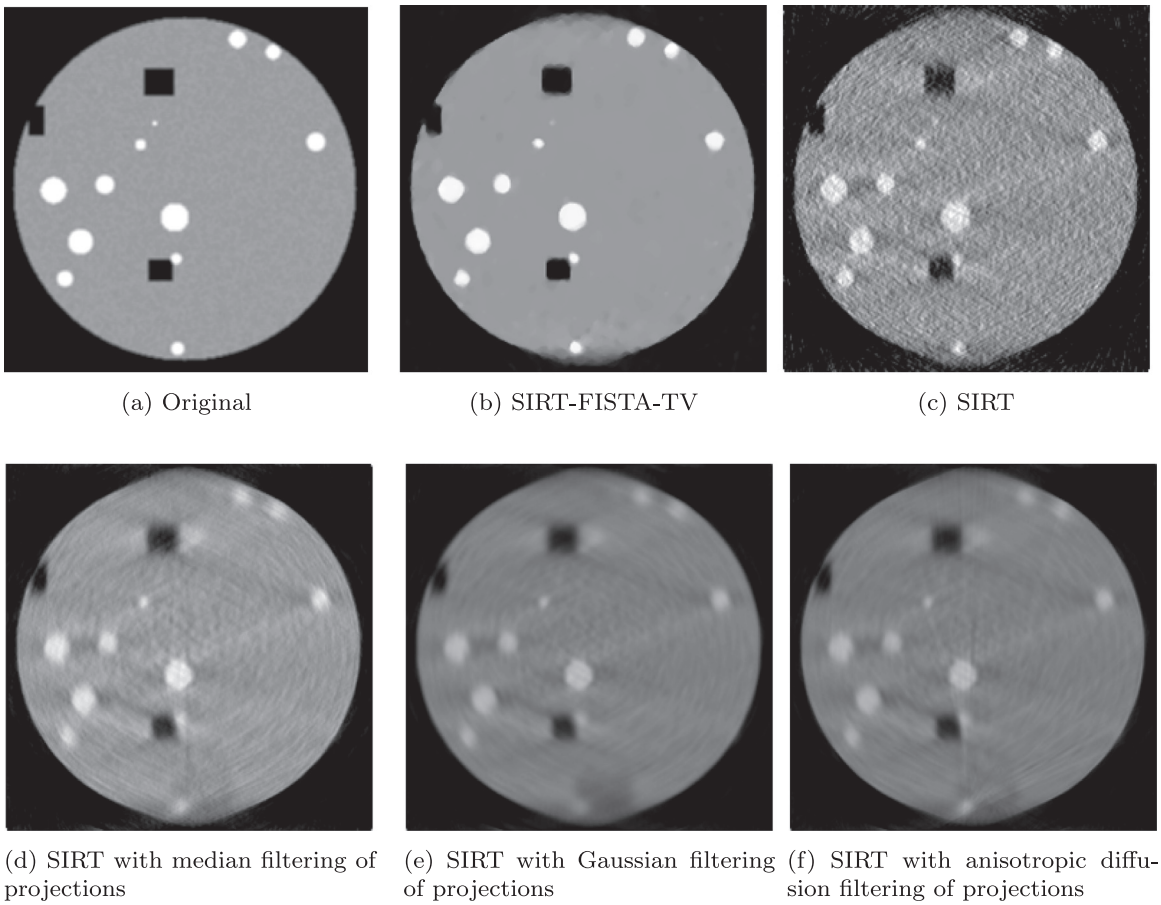


Fig. 4. PSNR values for volumes reconstructed by SIRT-FISTA-TV and SIRT-TV as function of number of iterations.



(a) Original (b) SIRT-FISTA-TV (c) SIRT

(d) SIRT with median filtering of projections (e) SIRT with Gaussian filtering of projections (f) SIRT with anisotropic diffusion filtering of projections

Fig. 5. Numerical comparison of reconstruction from noisy data. We display a cross-section in XY plane of volumes reconstructed by SIRT, SIRT-FISTA-TV and SIRT with projections denoised using different filtering techniques.

spond to SIRT followed by different filtering techniques applied after reconstruction process. This emphasizes again the robustness of SIRT-FISTA-TV to noise.

3.2. Reconstruction with blurred data

SIRT-FISTA-TV is tested with blurred projections. In this study, both translational and rotational blurs are applied on the projections. In the first simulation, the blur is translational and of variable amplitude between 25 and 50 pixels where the detector is 256^2 pixels. This translational blur mimics the lateral motion of a real sample during a tilt series when experiencing displacements due to mechanical instabilities of the goniometer. A motion filter that can be expressed as a linear filter with impulse response describing the displacement is used to simulate blurred images. These images are chosen randomly and different magnitudes of blurring are set in order to mimic mechanical instabilities of the goniometer during fast rotation. We show in Fig. 7 an original projection and a blurred one obtained using the motion filter.

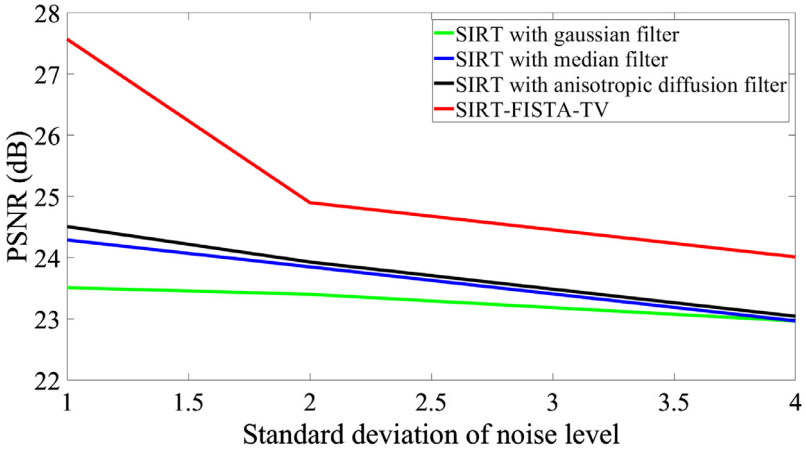
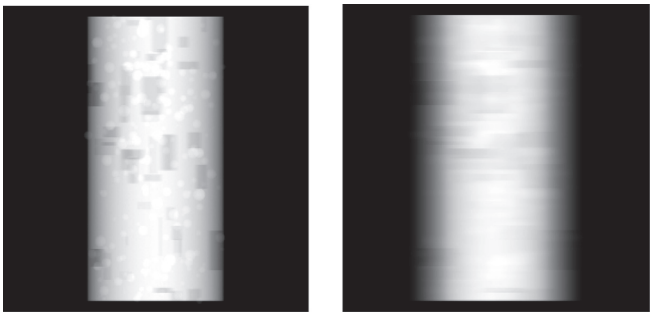


Fig. 6. Peak signal-to-noise ratio as function of noise level for reconstructions with SIRT-FISTA-TV and SIRT combined with different filtering techniques.



(a) Original (b) Blurred

Fig. 7. Example of a simulated blurred projection with the amplitude of blur is 50 pixels and the size of the projection is 256×256 pixels.

In these simulations, 500 projections are acquired with a tilting angular range between -70° and $+70^\circ$ and some percentage of projections (10%, 30%, or 50%) are blurred randomly with the translational blur described previously. The results are shown in Fig. 8 that displays cross-sections in XY- and XZ- plane of each reconstructed volume. The cross-sections of reconstructed volume obtained with non-blurred data are displayed in the first column and taken as a reference in this numerical comparison. When 10% of projections are blurred, the cross-sections illustrated in the second column seem identical to the references shown in first column. When 30% of projections are blurred, there is a slight blurring effect in the reconstructed cross-sections, but there is no severe influence on the reconstruction quality. Finally, when 50% of projections are blurred, there are artifacts due to blurring effects and it can be observed in the XY-plane cross-section of fourth column that a tiny white inclusion (defined by the red rectangular region) has disappeared.

The quantitative evaluation is carried on the distribution of sizes of reconstructed white inclusions. We illustrate in Fig. 9 histograms representing the distribution of the diameter of white inclusions for the reconstructed volumes shown previously in Fig. 8. The size of inclusions is determined using the "3D object counter" plugin [24] of ImageJ. This operation is manual where the user has to select and adjust a threshold value that allows then an automatic segmentation and delivers a results table containing all measured inclusions with their volume and respective position. The histogram corresponding to the reconstruction with 10% of blurred projections is very similar to the histogram of reconstruction from non-blurred data except for very tiny white inclusions of diameter less than 5 pixels. A very few number of these white inclusions are not reconstructed due to blurring effects. The histogram obtained from the reconstruction with 30% of blurred projections is not significantly different from that obtained from reconstruction with non-blurred data. On the other hand, it is observed from the histogram obtained with 50% of blurred projections that many inclusions especially small ones are no more reconstructed.

SIRT-FISTA-TV is further tested with rotational blur to evaluate the blur effects within the projections inevitably induced by the high rotation speed of the goniometer. In our experimental applications, for a complete tilt of the sample over 140° in 5 seconds, the angular rotation speed is 28°/sec. This means that the object will be in rotational motion while recording its projections, which produces the rotational blur. With these conditions, if 100 frames are acquired per second, a rotational blur effect of 0.28° is obtained. We consider in this simulation a rotational blur of 0.3°, 2° and 5°. This rotational blur is simulated by averaging together several projection images which were calculated every 0.1° during the numeric generation of the tilt series. For each simulation, the total number of projections acquired is computed as follows: number of projections = (angular range)/(rotational blur). As a result, the total number of projections for a rotational blur of 0.3°, 2° and 5° is 466, 70 and 28, respectively. The numerical comparisons are illustrated in Fig. 10 that shows a XY-plane cross-section of each volume reconstructed from non-blurred and blurred projections. In addition, histograms representing the distribution of white inclusions in each reconstructed volume are also shown. In case of a rotational blur of 0.3°, the corresponding results displayed in the second column of Fig. 10 are in good agreement with those obtained from non-blurred data and displayed in the first column. For 2° of rotational blur, very slight artifacts are observed in the third column and a small difference error is obtained when calculating the histogram for distribution of white inclusion diameters. Finally, severe artifacts are obtained when the rotational blur is 5° and a large difference error is obtained in the histogram (see fourth column of Fig. 10).

To show the advantages of SIRT-FISTA-TV and its robustness to blurred data compared to conventional reconstruction algorithms, we show in Fig. 11 reconstructions of SIRT and SIRT-FISTA-TV together in the case of a rotational blur of 2°. In particular, we illustrate SIRT-FISTA-TV reconstruction with non-blurred data in Fig. 11a, SIRT-FISTA-TV reconstruction with blurred data in Fig. 11b and SIRT reconstruction with blurred data in Fig. 11c. It is evident from this numerical comparison that SIRT-FISTA-TV can handle blur effects, performs much better than SIRT and facilitates the segmentation process after reconstruction.

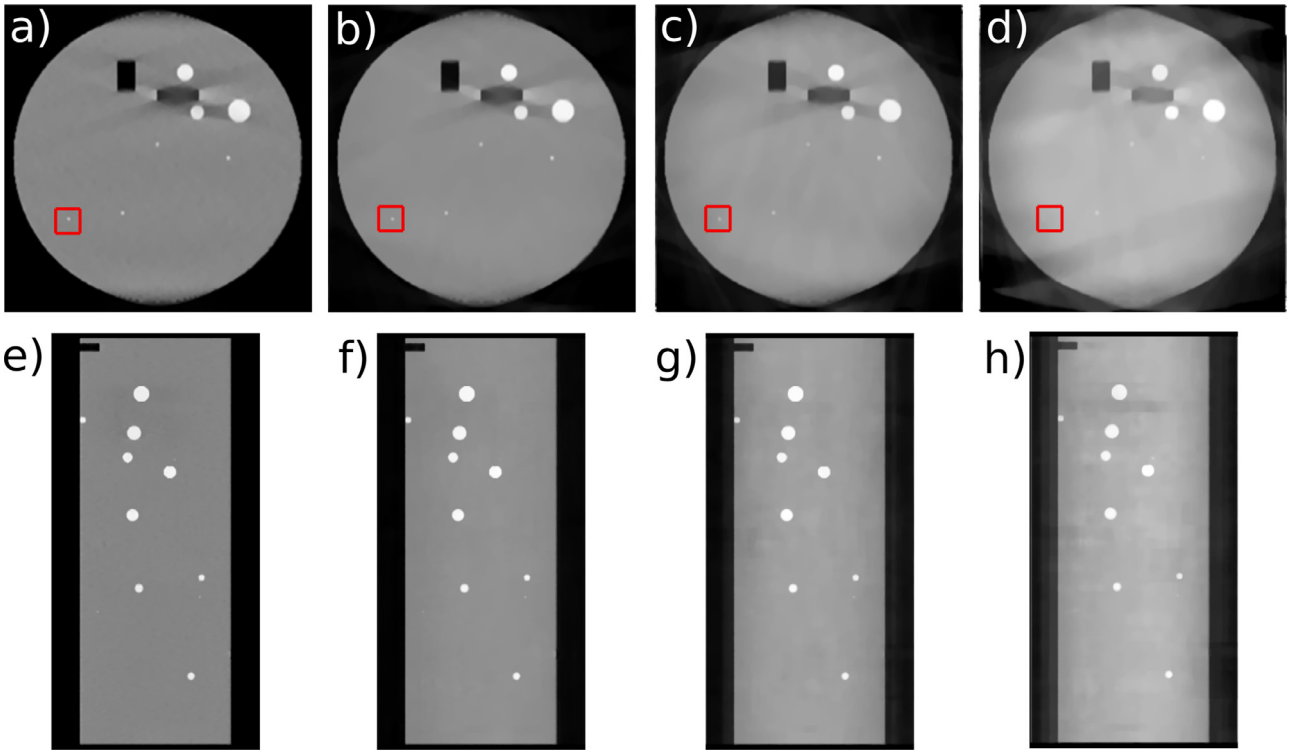


Fig. 8. SIRT-FISTA-TV reconstructions from non-blurred and blurred projections using a translational blur. (a): XY-plane cross-section in the volume reconstructed with non-blurred projections. (b-d): XY-plane cross-sections in reconstructions with randomly blurred projections; (b) with 10%, (c) 30% and (d) 50% of blurred projections. (e): XZ-plane cross-section in the volume reconstructed with non-blurred projections. (f-h): XZ-plane cross-sections in reconstructions with randomly blurred projections; (f) with 10%, (g) 30% and (h) 50% of blurred projections.

We also evaluate SIRT-FISTA-TV with both translational and rotational blur simultaneously. The rotational blur is fixed to 0.3° and the percentage of projections on which we apply a translational blur is 10%, 30% and 50%. Similarly, we show in Fig. 12 a XY-plane cross-section of each reconstructed volume and histograms representing the distribution of white inclusion diameters. It can be observed that for 10% and 30% of blurred projections, the results obtained are still acceptable. However, for 50% of blurred projections, there are some blurring artifacts in the reconstruction and the histogram values do not agree well with that obtained in the case of non blurred projections.

A segmentation is then applied to evaluate how large is the shape error for rectangular holes in reconstructed volumes with and without blurred projections. K-means clustering method [25] is used for segmentation and the results obtained are illustrated in Fig. 13 which demonstrates that these rectangular holes (in white) are segmented relatively properly even if the projections are blurred. For a quantitative evaluation of these segmented volumes, we compute the Dice's similarity coefficient [26] which is one of the most frequently used evaluation measures in medical image segmentation. The Dice coefficient is defined as the size of the intersection of two sets A and B divided by the average of their sizes. It is measured as follows:

$$\mathrm{Dice}(A,B) = \frac{2|A \cap B|}{(|A| + |B|)},$$

where the dice score is between 0 and 1 and it is equal to 1 when A and B are identical. Here, A corresponds to the segmented original volume and B to the segmented reconstructed one. The Dice values obtained are: 0.9922 for reconstruction with non-blurred projections, 0.9895, 0.9814 and 0.9603 for reconstruction with 10%, 30% and 50% of blurred projections, respectively. These high values of dice score show again that the reconstruction error of rectangular holes error is still small even if projections are blurred.

Table 2
Normalized Euclidean distance for the histograms of white inclusion diameters within volumes reconstructed with different types of blur effects.

Blur effect	NED
Rotational blur of 0.3°	0.7
Rotational blur of 2°	2.1
Rotational blur of 5°	14.5
Translational blur over 10% of projections	2
Translational blur over 30% of projections	2.6
Translational blur over 50% of projections	5.8
Rotational blur of 0.3° + translational blur over 10% of projections	2.3
Rotational blur of 0.3° + translational blur over 30% of projections	3.6
Rotational blur of 0.3° + translational blur over 50% of projections	8.4

For quantitative evaluation, we show in Table 2 the normalized Euclidean distance (NED) between histograms for white inclusions diameters of volumes reconstructed without blur and with different types of blur effects. The value of NED between two histograms is given by:

$$\mathrm{NED}(H_1, H_2) = \sqrt{\sum_{i=1}^{K} \frac{(H_1(i) - H_2(i))^2}{K}}, \quad (16)$$

where $H_1(i)$ and $H_2(i)$ are count values for the two histograms at bin index i and K is the total number of bins for each histogram. The results obtained show that NED becomes significant in case of a rotational blur of 5° or when a translational blur is applied over 50% of projections.

Finally, we compare reconstruction quality when removing blurred projections before reconstruction. This was done because in practical situations, the alignment of blurred projections is not precise and may cause additional artifacts in reconstruction. In this simulation, a translational blur is applied over 50% of projections. Then, SIRT-FISTA-TV reconstruction is done using a) overall pro-

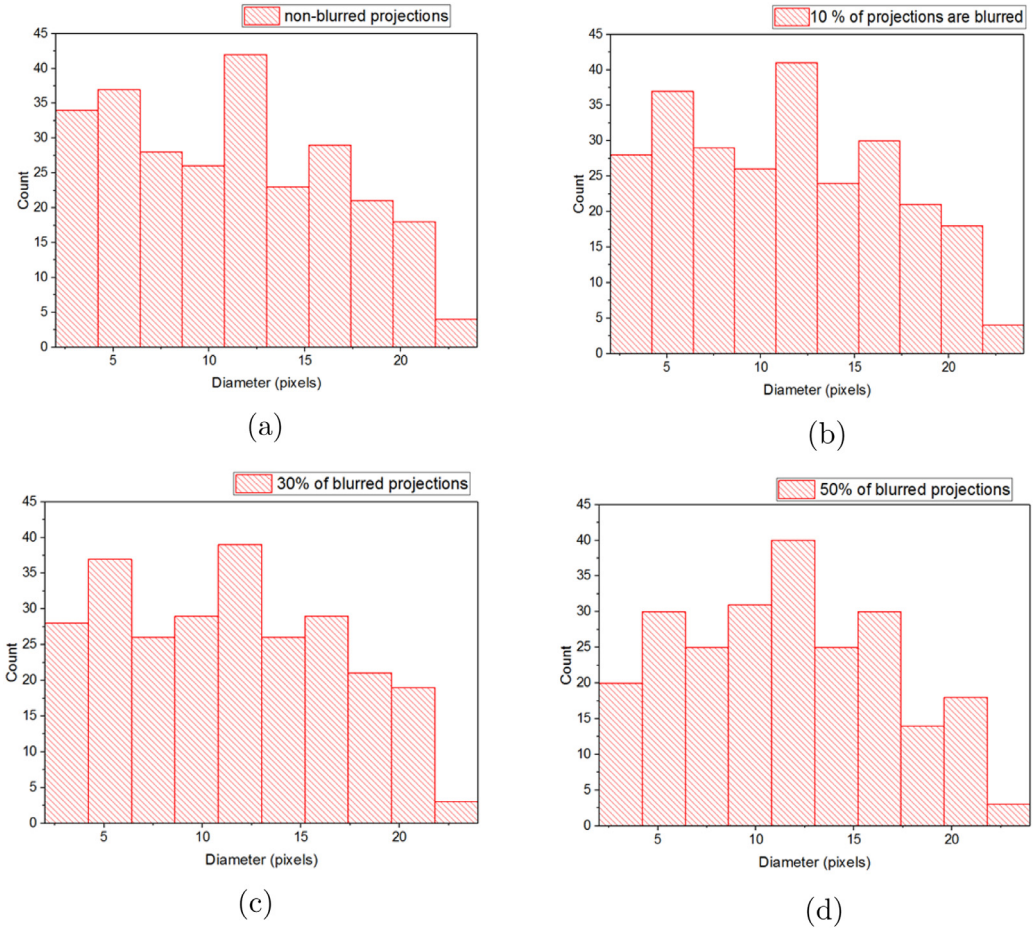


Fig. 9. Histograms for the distribution of white inclusion diameters within volumes reconstructed by SIRT-FISTA-TV with (a) non-blurred, (b) 10%, (c) 30% and (d) 50% of blurred projections.

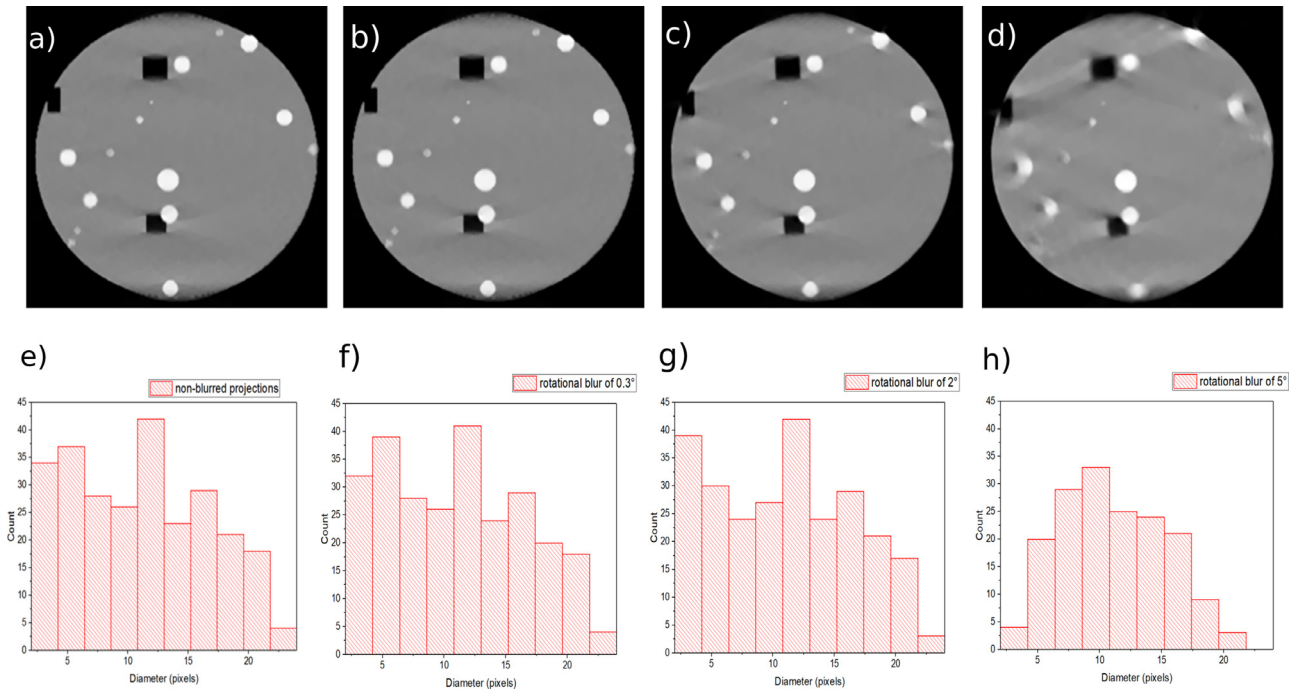


Fig. 10. SIRT-FISTA-TV reconstructions from non-blurred and blurred projections using a rotational blur of different degrees: 0.3°, 2° and 5°. (a): XY-plane cross-section in volume reconstructed with non-blurred projections. (b-d): XY-plane cross-sections in reconstructions with different degrees of rotational blur; (b) with 0.3°, (c) 2° and (d) 5°. (e): Histogram for distribution of white inclusion diameters within the volume reconstructed with non-blurred projections. (f-h): Histograms for distribution of white inclusion diameters within volumes reconstructed with different degrees of rotational blur; (f) with 0.3°, (g) 2° and (h) 5°.

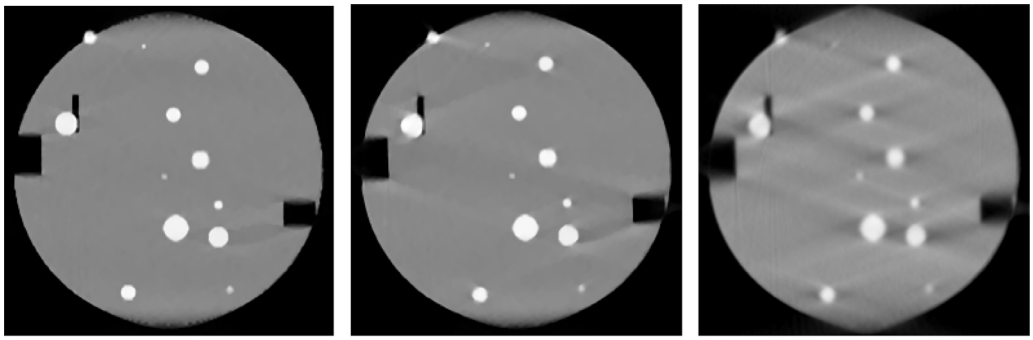


Fig. 11. XY-plane cross-sections in volumes reconstructed by (a) SIRT-FISTA-TV with non-blurred data, (b) SIRT-FISTA-TV with blurred data and (c) SIRT with blurred data. The type of blur applied here is a rotational blur of 2°.

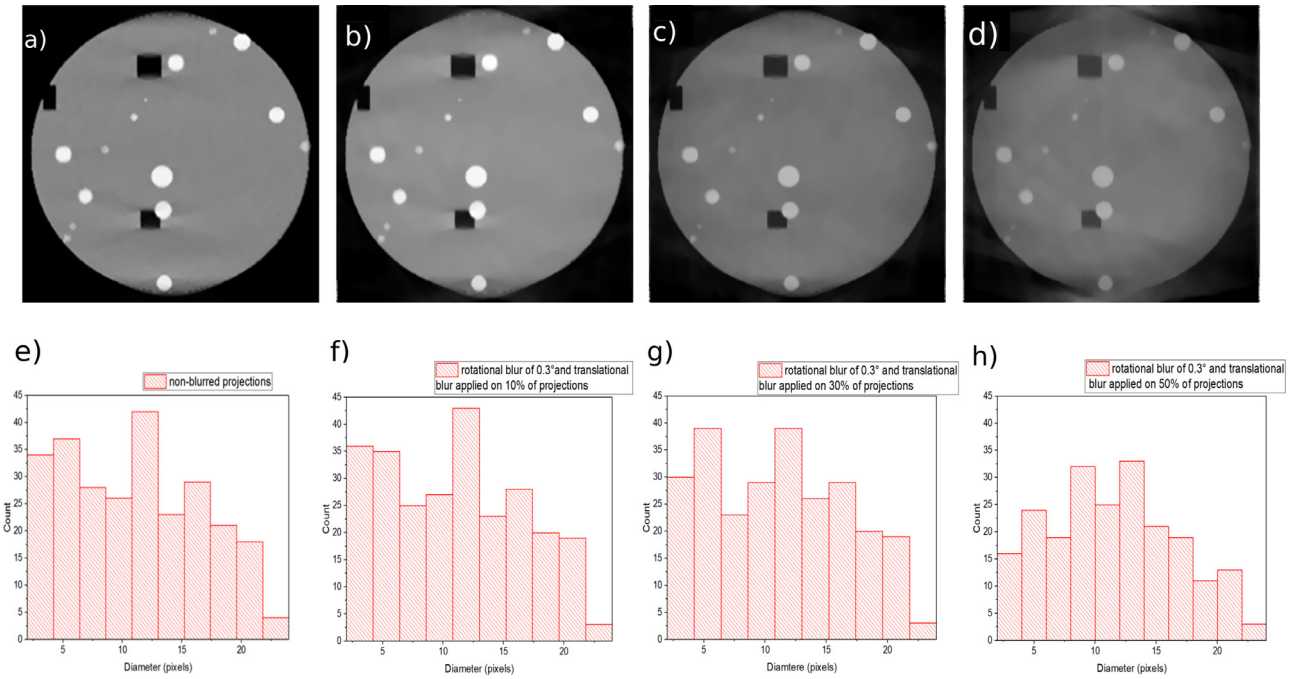


Fig. 12. SIRT-FISTA-TV reconstructions from non-blurred and blurred projections. (a): XY-plane cross-section in volume reconstructed with non-blurred projections. (b-d): XY-plane cross-sections in volumes reconstructed with a rotational blur of 0.3° and a translational blur which was applied on different percentage of projections; (b) with 10%, (c) 30% and (d) 50% of projections. (e): Histogram for distribution of white inclusion diameters within the volume reconstructed with non-blurred projections. (f-h): Histograms for distribution of white inclusion diameters within volumes reconstructed with a rotational blur of 0.3° and a translational blur which was applied on different percentage of projections; (f) with 10%, (g) 30% and (h) 50% of projections.

jections including blurred ones and b) only non blurred projections. XY-plane cross-section of each reconstructed volume is illustrated in Fig. 14. When using these blurred projections, we obtain blur effects in the reconstruction (see Fig. 14a) and very small white inclusions are no more reconstructed. On the other hand, it can be observed in Fig. 14b that these blur effects are eliminated when removing blurred projections before reconstruction and that all white inclusions are reconstructed. These indicative simulations show the advantages of removing the most blurred projections before reconstruction for both facilitating the alignment process (for experimental dataset) and improving reconstruction quality.

4. Experimental results

We present in this section some reconstructions from experimental data obtained with a differentially pumped FEI Titan ETEM operating at 300 keV and aberration-corrected. The microscope is equipped with a high sensitive Gatan OneView 16 Megapixels CMOS camera of 15 μm pixel size enabling ultra-high-definition image resolution. This advanced camera is of high readout speed and can record 25 fps at full 4K x 4K resolution, 100 fps at 2K x 2K resolution and even 300 fps if the image resolution is reduced to 512 x 512 pixels. Two tilt series are acquired in the TEM bright field mode (TEM-BF) on two different catalytic systems both studied under environmental gas and high temperature conditions. The first series is a fast step-by-step acquisition while the second one is an ultra-fast acquisition with a continuous rotation. These two methods are described in details in [27].

The sample used during the first experiment is constituted by palladium nanoparticles supported on alumina-based grains (Pd/Al_2O_3) that have been studied in ETEM [28] for its practical interest in various industrial reactions in petrochemistry. The tilt series was acquired by fast step-by-step TEM under gas and high temperature (P = 2 mbar O_2 and T = 450°C) that is under real environmental conditions. The total acquisition time of this sequence was around 144 seconds and the overall number of projections is

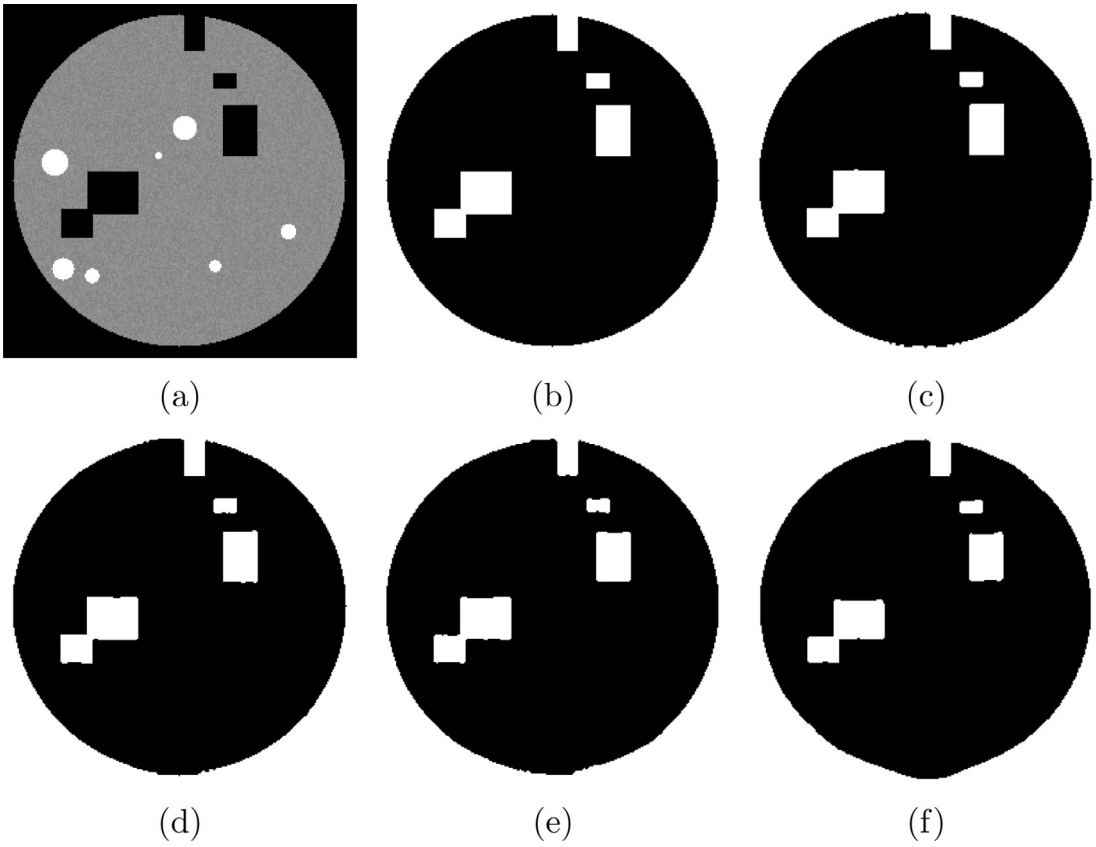


Fig. 13. Segmentation of rectangular holes in SIRT-FISTA-TV reconstructed volumes obtained from blurred and non-blurred projections. (a-c): XY-plane cross-sections of (a) original volume (b) segmented original volume and (c) segmented volume that is reconstructed from non-blurred projections. (d-f): XY-plane cross-sections in segmented volumes that are reconstructed with a rotational blur of 0.3° and a translational blur which is applied on different percentage of projections; (d) with 10%, (e) 30% and (f) 50%.

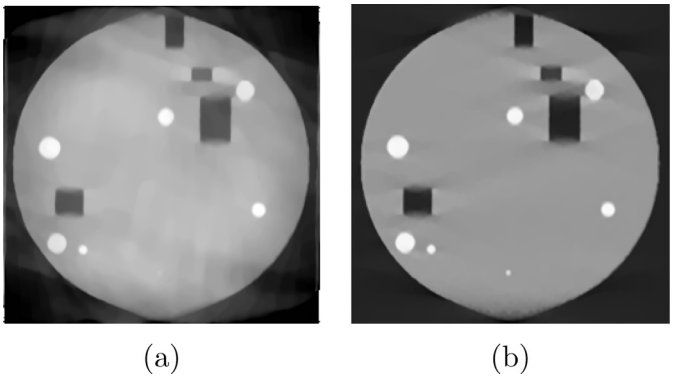


Fig. 14. XY-plane cross-section in volumes reconstructed by SIRT-FISTA-TV using (a) all projections including blurred ones (a translational blur is applied to 50% of projections) and (b) only non-blurred projections.

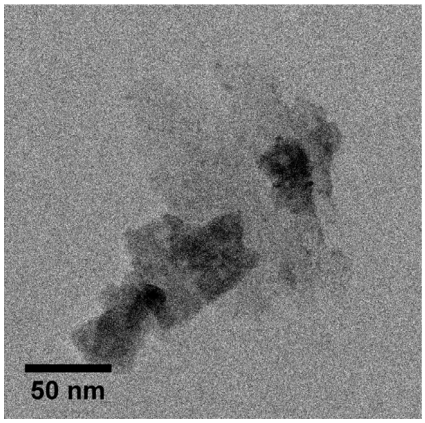


Fig. 15. 2D projection of Pd/Al_2O_3 sample.

72 with tilting angular range from −66° to +74°. The projections are in 4K mode and binned by a factor of 2 before performing reconstruction. A 2D projection of the sample is illustrated in Fig. 15. The projections are aligned using IMOD [29]. SIRT and SIRT-FISTA-TV algorithms are both used to produce reconstructed volumes of 1024^3 voxels with the size of each voxel is $0.3\,nm^3$. Fig. 16 displays a horizontal cross-section of each reconstructed volume. The results obtained show that SIRT-FISTA-TV significantly suppresses noise compared to SIRT while still preserving sharp edges.

One may wonder about the influence of diffraction contrast during bright field (BF) TEM tomography experiments. It is beyond the scope of this paper to discuss in details about diffraction contrast unavoidable in the case of crystalline samples, which is however not a problem specific to 'fast' tomography. It is well-known that conventional imaging has drawbacks as compared to imaging modes less sensitive to diffraction effects, such as in Scanning TEM [30]. The context to which we refer here is environmental tomography, where the priority is to record the tilting series in the shortest possible time to provide a time-resolved 3D analysis [27]. Under those conditions, acquisitions of the order of a few seconds are aimed at [13,14]. STEM imaging, or precession intending to average diffraction contrasts in BF Tomography [31] are till today by far too slow and we then have no other choice than dealing with possible diffraction contrast variations in BF-TEM. The experimen-

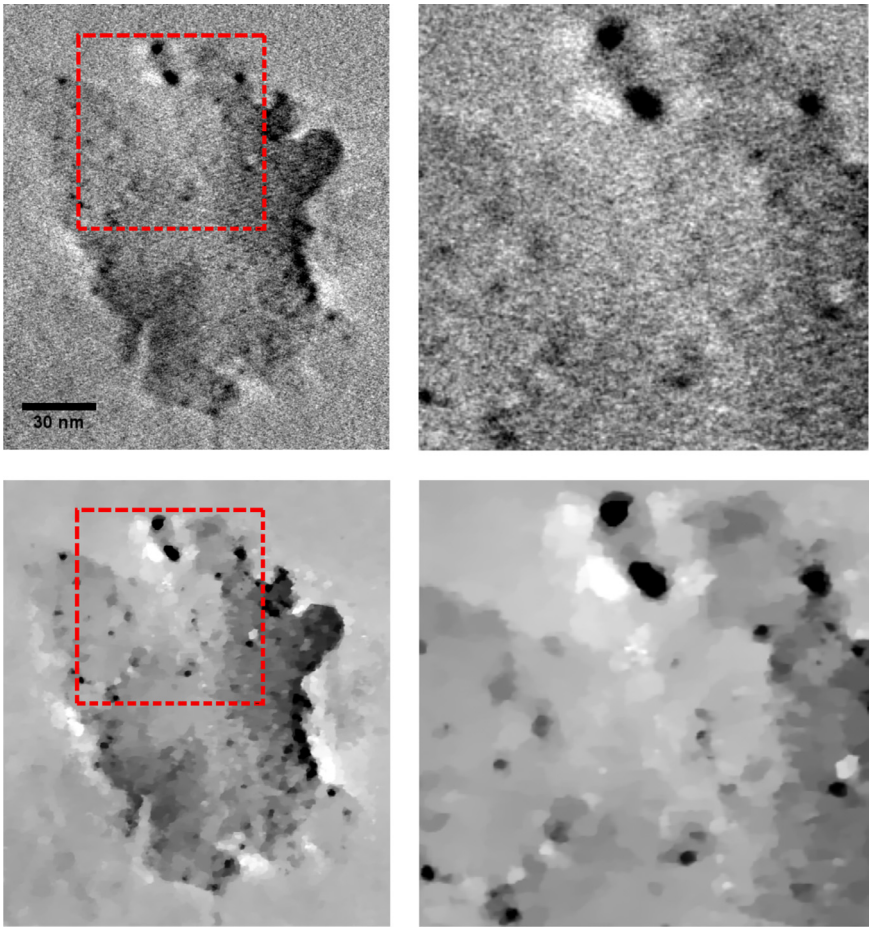


Fig. 16. Reconstruction of Pd/Al_2O_3 sample using SIRT (first row) and SIRT-FISTA-TV (second row). XY-plane cross-section in each reconstructed volume is shown in first column and a zoom-in on the red rectangular region is illustrated in second column. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tal examples reported in the present section show that reasonable volumes can indeed be reasonably reconstructed despite diffraction contrast.

The second experiment concerns the soot-zirconia catalytic system mentioned in the Introduction section. It was acquired under environmental conditions ($T = 300°C$ and $P = 5 \times 10^{-5}$ mbar O_2) that are just not severe enough to avoid the soot combustion. A rapid continuous rotation of the goniometer was performed to highly accelerate the data acquisition and images were recorded with the Gatan Oneview camera in the 2K x 2K bining mode at 100 fps. The tilt series contains 507 projections ($2K \times 2K$ resolution) with a tilting angular range from $-69°$ to $71°$ and was acquired in almost 5 seconds. A 2D projection is shown in Fig. 17a. The projections are binned by a factor of 2 and only 71 aligned projections are used for reconstruction. In particular, the sharpest projections are selected in order to facilitate the alignment procedure where the alignment of the severely blurred projections might be unreliable and is expected to produce misalignment artifacts in the reconstructed volumes. These aligned projections are selected manually based on visual inspection of the images when setting markers for alignment procedure.

SIRT and SIRT-FISTA-TV are used for reconstruction and a horizontal cross-section of each reconstructed volume (1024^3 voxels) is illustrated in Fig. 17. It is clear that SIRT-FISTA-TV gives better reconstruction quality than SIRT and we see again how the incorporated TV regularization step of SIRT-FISTA-TV can suppress noise artifacts without affecting spatial resolution.

After tomographic reconstruction, an additional segmentation step is required to produce a 3D model from the reconstructed volume. However, this segmentation step is sensitive to noise and becomes challenging to apply when data are very noisy. Segmentation of the reconstructed volume obtained by SIRT-FISTA-TV is much easier than that obtained by SIRT, where the thresholding step during segmentation is straightforward because the noise level is decreased significantly in the reconstructed volume and the edges are well defined. Fig. 17b shows the segmented model obtained from SIRT-FISTA-TV reconstruction where carbon soot is represented in gray and zirconia-based nanocatalyst in blue.

In order to evaluate reconstruction quality if the acquisition was even faster than the performed one, we blurred these experimental projections of the soot-zirconia agglomerate. As a remark, the current limitation to achieve faster acquisitions is not the readout of the camera, but rather the rotation speed of the goniometer. A translational blur of random amplitude between 10 and 20 pixels (field of view 380 pixels) is applied firstly to 10% and then to 50% of projections chosen randomly. Then, 3D reconstructions are done using SIRT and SIRT-FISTA-TV and the results obtained are shown in Fig. 18. Again, we observe that artifacts are obtained when reconstructing with 50% of blurred projections. However, SIRT-FISTA-TV is more robust to blur and gives better results than SIRT. For quantitative evaluation, MSE is measured for SIRT and SIRT-FISTA-TV reconstructions when 50% of projections are blurred. The volume reconstructed with non-blurred projections is considered as a reference in these measurements and the obtained values of MSE

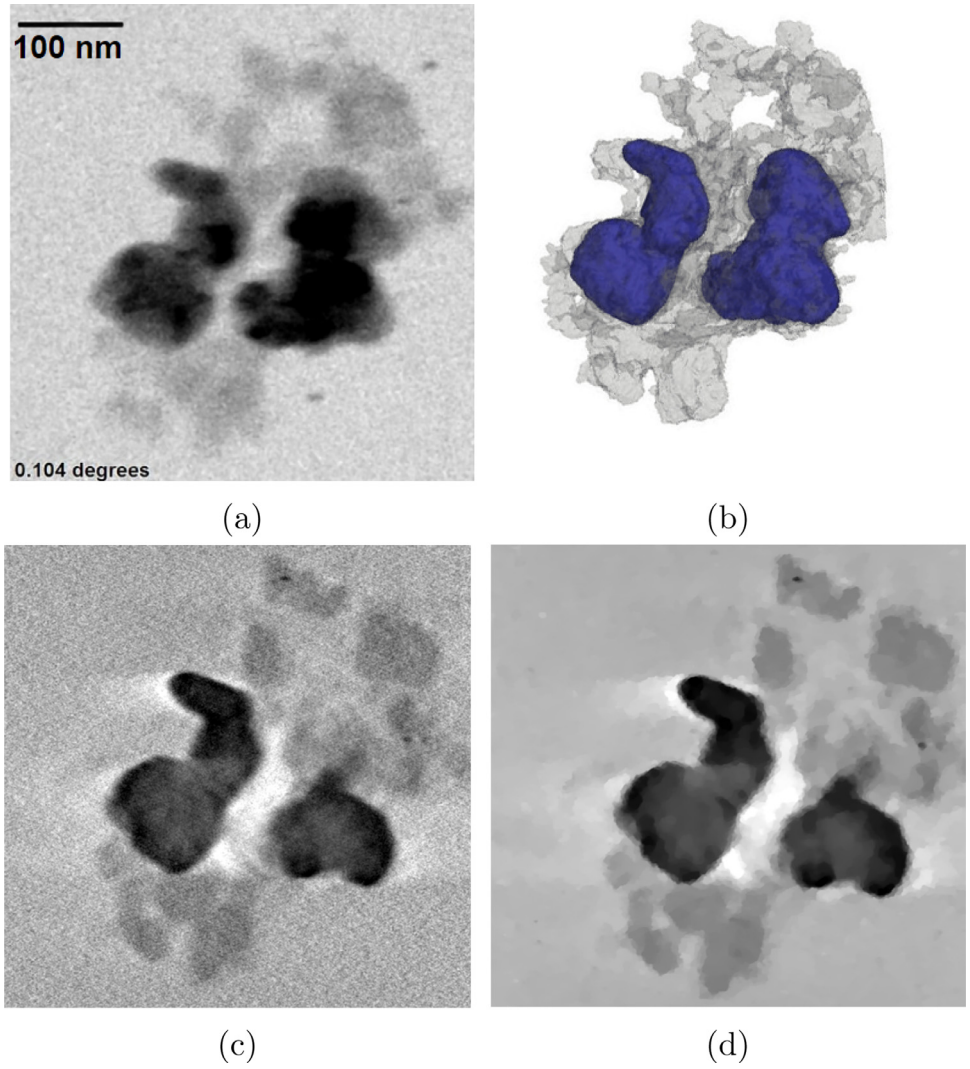


Fig. 17. (a): 2D projection of the soot-zirconia sample. (b) Segmented model of the volume reconstructed by SIRT-FISTA-TV. (c-d) Horizontal cross-sections within the reconstructed volume; (c) using SIRT and (d) SIRT-FISTA-TV.

are 15 for volume reconstructed by SIRT and 7.7 for that reconstructed by SIRT-FISTA-TV.

5. Conclusion

The main topic of this work deals with fast tomography inside an Environmental Transmission Electron Microscope in order to follow in situ and in 3D the evolution of nanomaterials such as nano-catalysts studied under dynamic gas and temperature conditions. In this context, the acquisition of the tilt series has to be accelerated to perform a real time analysis, or at least to avoid that the object of interest evolves during the data collection. High speed rotation of the sample and high speed image recording have then to be used, which may induce rotation and lateral blur in the projections. A new iterative reconstruction algorithm named SIRT-FISTA-TV was introduced in this paper with the aim to improve reconstructions from noisy and blurred projections induced by the high rotation speed of the goniometer during fast acquisitions. This algebraic algorithm with TV regularization was evaluated by use of both computer-simulated and experimental data sets.

It was demonstrated that SIRT-FISTA-TV produces reconstructed volumes with clear and sharp edges, and thus, the segmentation step was quite easy to perform after 3D reconstruction. Applying the algorithm on very noisy data, we have shown that it is very robust to noise and can provide excellent results compared to conventional algorithms of reconstruction. Numerical simulations were also carried to evaluate the impact of blurred projections on reconstruction quality by simulating both rotational and translational blur. The obtained results demonstrated that SIRT-FISTA-TV is robust to blurred data and gives better results than conventional reconstruction algorithms. It has however some limitations and cannot completely avoid artifacts in the 3D reconstruction when a large percentage (around 50%) of projections are highly blurred. With such a poor quality of data, it is advised to suppress blurred data before reconstruction.

Two experimental data sets were presented. The first one consisted in a fast step-by-step TEM and its objective was only to show the advantages of SIRT-FISTA-TV over classical reconstruction algorithms when applied on experimentally acquired data. The second was a fast continuous TEM acquisition that was performed under environmental conditions in 5 seconds over a rotation amplitude of 140°. The sample was a nanocatalyst studied in the context of Diesel motors soot oxidation catalysis. Dedicated results regarding the fast 3D analysis of its evolution under different temperature conditions will be submitted separately. The reconstruction results obtained with the SIRT-FISTA-TV algorithm show that quantitative 3D analysis becomes feasible even under operando ETEM conditions owing to the reasonable quality of reconstructed vol-

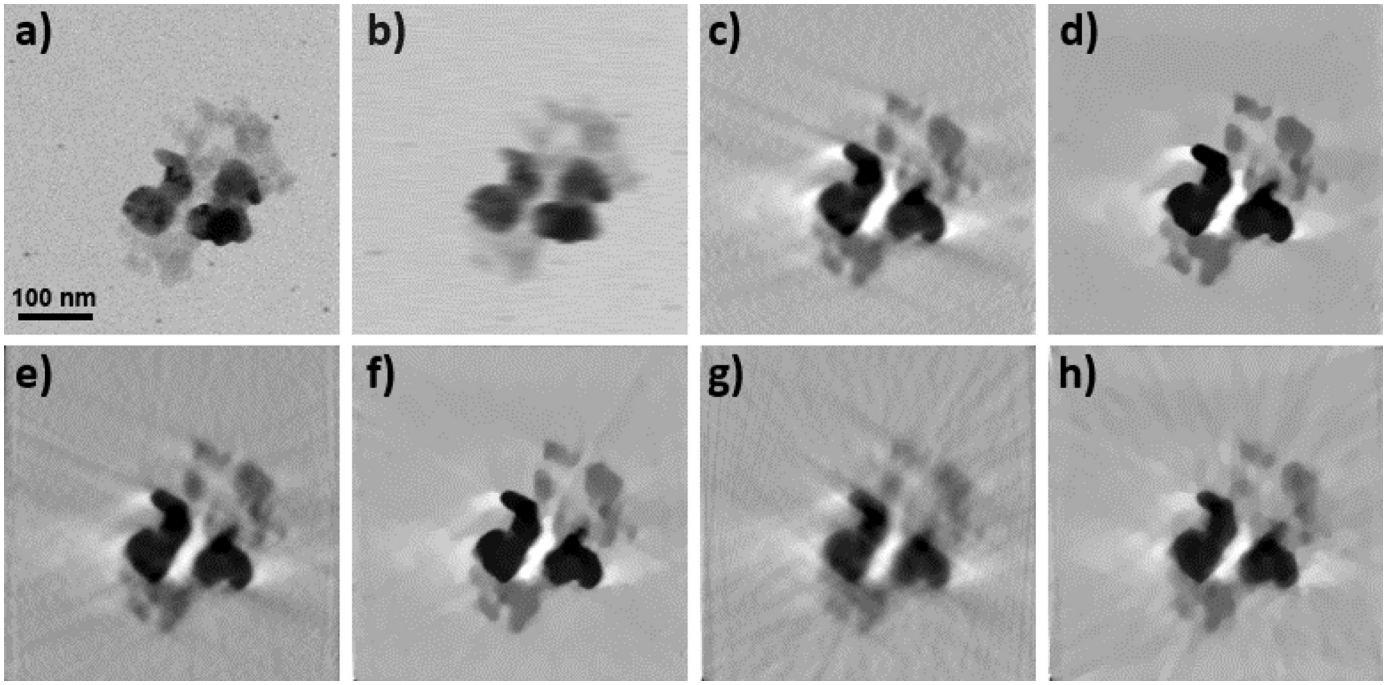


Fig. 18. (a): Experimental ETEM projection of the soot-zirconia agglomerate and (b): blurred projection with a translational blur of 20 pixels along Ox. (c-h): Horizontal sections within the reconstructed volume; (c) using SIRT and (d) SIRT-FISTA-TV without any blur; (e) using SIRT and (f) SIRT-FISTA-TV with translational blur applied randomly to 10% of projections; (g) and (h) same as (e-f) with 50% of the projections are randomly blurred.

umes even when obtained from tilt series recorded in a few seconds.

Works in progress concern the fast prediction and calibration of the goniometer trajectories and the improvement of alignment procedures in the presence of translational blurs.

Supplementary information

We provide three videos corresponding to simulated tilt series with a rotational blur fixed to 0.3° and the percentage of projections on which we apply a translational blur is 10%(Sino_0.3° rotational blur and 10% of projections with translational blur.avi), 30% (Sino_0.3° rotational blur and 30% of projections with translational blur.avi) and 50% (Sino_0.3° rotational blur and 50% of projections with translational blur.avi). In addition, the 3D model of the soot-zirconia agglomerate studied in Section 4 is provided as a video file (H-Banjak_et-al_SIRT-FISTA-TV_Soot-ZrO2_scalebar100nm.avi); the scalebar is 100 nm.

Acknowledgments

The authors acknowledge the financial support of the French National Research Agency through the ANR project 3DCLEAN (ANR-15-CE09-0009), the LABEX IMUST (ANR-10-LABX-0064) and LABEX PRIMES (ANR-11-IDEX-0007) of Université de Lyon within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the ANR and the BQR SPEE3D granted by INSA-Lyon. Thanks are due to P. Vernoux (IRCELYON, Univ. Lyon) for providing the soot-zirconia sample. We gratefully acknowledge the CLYM (www.clym.fr) for the access to the ETEM instrument, the NVIDIA Corporation for its support with the donation of the Quadro M5000 GPU used for this research and Gatan for its constant interest in the development of fast environmental tomography.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ultramic.2018.03.022.

References

[1] B. Goris, S. Bals, W. Van den Broek, E. Carbó-Argibay, S. Gómez-Graña, L.M. Liz–Marzán, G. Van Tendeloo, Atomic-scale determination of surface facets in gold nanorods, Nat. Mater. 11 (11) (2012) 930–935.
[2] M. Hayashida, M. Malac, Practical electron tomography guide: recent progress and future opportunities, Micron 91 (2016) 49–74.
[3] D. De Rosier, A. Klug, Reconstruction of three dimensional structures from electron micrographs, Nature 217 (5124) (1968) 130–134.
[4] M. Radermacher, Weighted back-projection methods, in: Electron Tomography, Springer, 2007, pp. 245–273.
[5] R. Gordon, R. Bender, G.T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography, J. Theor. Biol. 29 (3) (1970) 471IN1477–476IN2481.
[6] A.H. Andersen, A.C. Kak, Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm, Ultrason. Imaging 6 (1) (1984) 81–94.
[7] P. Gilbert, Iterative methods for the three-dimensional reconstruction of an object from projections, J. Theor. Biol. 36 (1) (1972) 105–117.
[8] R. Leary, Z. Saghi, P.A. Midgley, D.J. Holland, Compressed sensing electron tomography, Ultramicroscopy 131 (2013) 7091.
[9] B. Goris, W. Van den Broek, K.J. Batenburg, H.H. Mezerji, S. Bals, Electron tomography based on a total variation minimization reconstruction technique, Ultramicroscopy 113 (2012) 120–130.
[10] K.J. Batenburg, J. Sijbers, DART: a practical reconstruction algorithm for discrete tomography, IEEE Trans. Image Process. 20 (2011) 254253.
[11] T.W. Hansen, J.B. Wagner, Controlled Atmosphere Transmission Electron Microscopy, Springer, 2016.
[12] A. Serve, T. Epicier, M. Aouine, F.C.S. Aires, E. Obeid, M. Tsampas, K. Pajot, P. Vernoux, Investigations of soot combustion on yttria-stabilized zirconia by environmental transmission electron microscopy (ETEM), Appl. Catal., A 504 (2015) 74–80.
[13] L. Roiban, S. Koneti, K. Tran, Y.-M. Feng, T. Grenier, V. Maxim, T. Epicier, Rapid tomography in environmental TEM: how fast can we go to follow the 3d evolution of nanomaterials in situ? Microsc. Microanal. 22 (S5) (2016) 8–9.
[14] V. Migunov, H. Ryll, X. Zhuge, M. Simson, L. Strüder, K.J. Batenburg, L. Houben, R.E. Dunin-Borkowski, Rapid low dose electron tomography using a direct electron detection camera, Sci. Rep. 5 (2015).

[15] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1–4) (1992) 259–268.
[16] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[17] E.Y. Sidky, C.-M. Kao, X. Pan, Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT, J. X-ray Sci. Technol. 14 (2) (2006) 119–139.
[18] W. van Aarle, W.J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K.J. Batenburg, J. Sijbers, The ASTRA toolbox: a platform for advanced algorithm development in electron tomography, Ultramicroscopy 157 (2015) 35–47.
[19] P.M. Joseph, An improved algorithm for reprojecting rays through pixel images, IEEE Trans. Med. Imaging 1 (3) (1982) 192–196.
[20] R. Guedouar, B. Zarrad, A new reprojection method based on a comparison of popular reprojection models, Nucl. Instrum. Methods Phys. Res., Sect. A 619 (1) (2010) 270–275.
[21] F. Bleichrodt, T. van Leeuwen, W.J. Palenstijn, W. van Aarle, J. Sijbers, K.J. Batenburg, Easy implementation of advanced tomography algorithms using the ASTRA toolbox with spot operators, Numer. Algorithms 71 (3) (2016) 673–697.
[22] E. van den Berg, M. Friedlander, Spot-a linear-operator toolbox, URL http://www.cs.ubc.ca/labs/scl/spot (2014).
[23] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Anal. Mach. Intell. 12 (7) (1990) 629–639.
[24] S. Bolte, F. Cordelieres, A guided tour into subcellular colocalization analysis in light microscopy, J. Microsc. 224 (3) (2006) 213–232.
[25] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory 28 (2) (1982) 129–137.
[26] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.
[27] L. Roiban, S. Li, M. Aouine, A. Tuel, D. Farrusseng, T. Epicier, Fast 'operando' electron nanotomography, J. Microsc. 269 (2) (2018) 117–126.
[28] S. Koneti, L. Roiban, A.-S. Gay, P. Avenier, F. Dalmas, T. Epicier, Calcination of Pd nanoparticles on delta alumina: ex-situ analysis versus in-situ environmental TEM, Microsc. Microanal. 22 (Suppl5) (2016) 58–59.
[29] J.R. Kremer, D.N. Mastronarde, J.R. McIntosh, Computer visualization of three-dimensional image data using IMOD, J. Struct. Biol. 116 (1) (1996) 71–76.
[30] J.M. Thomas, P.A. Midgley, T.J. Yates, J.S. Barnard, R. Raja, I. Arslan, M. Weyland, The chemical application of high-resolution electron tomography: bright field or dark field? Angewandte Chemie 116 (48) (2004) 6913–6915.
[31] J. Rebled, L. Yedra, S. Estrade, J. Portillo, F. Peiro, A new approach for 3D reconstruction from bright field TEM imaging: beam precession assisted electron tomography, Ultramicroscopy 111 (9–10) (2011) 1504–1511.