



## DÉVELOPPEMENT D'OUTILS INTELLIGENTS SANS APPRENTISSAGE PRÉALABLE POUR LA CARACTÉRISATION DE SURFACES MICROSCOPIQUES

## DEVELOPMENT OF INTELLIGENT TOOLS WITHOUT PRIOR TRAINING FOR MICROSCOPIC SURFACE CHARACTERIZATION

*Etablissement* **Université Polytechnique Hauts de France**

*École doctorale* **Ecole Doctorale Polytechnique Hauts-de-France**

*Spécialité* **Informatique et applications**

*Domaine Scientifique* **Sciences pour l'ingénieur**

*Unité de recherche* **Laboratoire Interdisciplinaire des Transitions de Lille**

*Encadrement de la thèse* **Charles YAACOUB**

*Co-Encadrant* **Vivien SCOTTEZ**

**Financement** *origine* **50% Digital Surf + 50% ICL (FGES) Employeur ICL (FGES)**

Partenariat d'entreprises ou d'associations

*Début de la thèse le* **1er septembre 2026**

*Date limite de candidature (à 23h59)* **27 mai 2026**

### Grands défis sociétaux

---

Numérique, industrie, espace

### Objectifs de développement durable

---

Industrie, innovation et infrastructure

### Mots clés - Keywords

---

Transformée par diffusion d'ondelettes, Champs stochastiques non gaussiens, Analyse d'images microscopiques, Outils sans apprentissage, Caractérisation de surfaces

Wavelet Scattering Transform, Non-Gaussian stochastic fields, Microscopic image analysis, Training-free methods, Surface characterization

### Description de la problématique de recherche - Project description

---

L'estimation de paramètres à partir de champs stochastiques non gaussiens constitue un défi courant en science des données. Que ce soit dans l'étude de processus physiques intrinsèquement non-linéaires mais aussi dans l'analyse de données complexes ou d'images dans l'industrie, être capable d'extraire un ensemble compact de coefficients pouvant servir de statistiques récapitulatives est une étape incontournable.

Ces dernières années, cette tâche a de plus en plus été accomplie par l'utilisation de réseaux de neurones convolutifs (CNN), qui ont su démontrer leur grande efficacité dans leur capacité à décrire la non-linéarité mais dont les coefficients demeurent difficiles à interpréter. De plus, les CNN nécessitent une étape d'entraînement sur un jeu de données dédié qu'il faut donc constituer, stocker et protéger. Cette étape d'entraînement est souvent longue, gourmande en énergie et donc coûteuse.

Ce projet de thèse vise à aborder le problème de l'estimation de paramètres de champs stochastiques non gaussiens à l'aide de la transformée par diffusion d'ondelettes (wavelete scattering transform, WST), un outil statistique qui partage des idées avec les CNN, tout en ne nécessitant ni entraînement ni réglage fin de paramètres car elle est similaire aux estimateurs statistiques traditionnels. Elle génère un ensemble de statistiques récapitulatives qui caractérise efficacement la non-gaussianité. La WST construit des représentations de signaux invariants, stables et informatives pour la classification. Elle est calculée en dispersant l'information du signal le long de plusieurs voies, avec une cascade d'opérateurs de module d'ondelettes implémentés dans un réseau de convolution profond. Le développement et les applications récentes de cet outil sur des données et images astrophysiques sont particulièrement prometteurs et ont ainsi été reconnus à l'échelle nationale, notamment par les médailles de bronze et d'or du CNRS en 2025.

La WST est un outil généraliste au même titre qu'un réseau de neurone ou qu'une transformée de Fourier. Son intérêt réside dans sa capacité à caractériser la non-gaussianité sans entraînement préalable. Elle peut permettre de faire : de l'exploration de donnée, de la génération d'images, de la classification ou encore de la séparation de composantes.

Le doctorant aura pour mission de se familiariser avec cette approche afin de développer et d'implémenter cet outil dans le but d'améliorer le traitement et l'analyse des images en champ proche obtenues par microscopies à effet tunnel et à force atomique ou plus généralement avec des mesures de topographie  $z = f(x, y)$ . En effet, la qualité d'une image est définie par la fidélité avec laquelle la topographie mesurée reflète la surface réelle. Mais de nombreux facteurs, tels que la contamination de la sonde, la dérive instrumentale, l'instabilité du système d'asservissement, ou les changements soudains d'interaction pointe-échantillon, peuvent introduire des artefacts. Une fois identifiés, la thèse visera à réduire, voire à supprimer leur impact sur l'image. La microscopie à effet tunnel (STM) et la spectroscopie tunnel produisent également des cartes spectroscopiques de la densité locale d'états électroniques en fonction de la position et de l'énergie formant un ensemble tridimensionnel de données. Le contraste dans ces images, souvent appelées cartes d'interférence de quasi-particules, provient des variations spatiales locales de cette densité, qui fluctue en raison de la diffusion et l'interférence des électrons par des défauts, impuretés, dislocations, marches atomiques à la surface ou sous la surface du matériau étudié. La WST sera appliquée à de tels jeux de données pour révéler des informations physiques essentielles sur les relations de dispersion et les processus de diffusion dans des matériaux semi-conducteurs et des cristaux atomiques à deux dimensions

Parameter estimation from non-Gaussian stochastic fields is a common challenge in data science. Whether in the study of intrinsically nonlinear physical processes or in the analysis of complex data and images in industry, the ability to extract a compact set of coefficients that can serve as summary statistics is an essential step.

In recent years, this task has increasingly been accomplished using convolutional neural networks (CNNs), which have demonstrated strong effectiveness in capturing nonlinearity, but whose coefficients remain difficult to interpret. Moreover, CNNs require a training phase on a dedicated dataset that must be created, stored, and secured. This training step is often time-consuming, energy-intensive, and therefore costly.

This PhD project aims to address the problem of parameter estimation for non-Gaussian stochastic fields using the wavelet scattering transform (WST), a statistical tool that shares conceptual similarities with CNNs while requiring neither training nor fine-tuning, as it is closer to traditional statistical estimators. It generates a set of summary statistics that efficiently characterize non-Gaussianity. The WST builds invariant, stable, and informative signal representations for classification. It is computed by propagating signal information along multiple paths through a cascade of wavelet modulus operators implemented within a deep convolutional network. Recent developments and applications of this tool to astrophysical data and images are particularly promising and have been recognized at the national level, notably by CNRS bronze and gold medals in 2025.

The WST is a general-purpose tool, similar in scope to neural networks or the Fourier transform. Its strength lies in its ability to characterize non-Gaussianity without prior training. It can be used for data exploration, image generation, classification, and component separation.

The PhD candidate will be tasked with becoming familiar with this approach in order to develop and implement this tool to improve the processing and analysis of near-field images obtained by scanning tunneling microscopy (STM) and atomic force microscopy (AFM), or more generally from topographic measurements  $z = f(x, y)$ . Indeed, image quality is defined by how faithfully the measured topography reflects the actual surface. However, many factors such as probe contamination, instrumental drift, instability in the feedback system, or sudden changes in tip-sample interaction can introduce artifacts. Once identified, the thesis will aim at reducing or even eliminating their impact on the images.

Scanning tunneling microscopy (STM) and tunneling spectroscopy also produce spectroscopic maps of the local density of electronic states as a function of position and energy, forming a three-dimensional dataset. The contrast in these images, often referred to as quasiparticle interference maps, arises from local spatial variations in this density, which fluctuates due to the scattering and interference of electrons by defects, impurities, dislocations, and atomic steps at or below the surface of the studied material. The WST will be applied to such datasets to reveal essential physical information about dispersion relations and scattering processes in semiconductor materials and two-dimensional atomic crystals.

## Thématique / Contexte

---

Le contexte scientifique de cette thèse s'inscrit à l'interface entre la science des données, le traitement du signal et la physique des matériaux, où l'analyse de données complexes issues de surfaces microscopiques constitue un enjeu majeur. L'estimation de paramètres à partir de champs stochastiques non gaussiens, omniprésents dans les phénomènes physiques non linéaires et les images expérimentales, requiert des outils capables d'extraire des descripteurs compacts et pertinents. Si les réseaux de neurones convolutifs se sont imposés ces dernières années, leurs limites en termes d'interprétabilité et de coût d'entraînement motivent l'exploration d'approches alternatives. Dans ce contexte, la transformée par diffusion d'ondelettes (WST) apparaît comme une méthode prometteuse, combinant la richesse descriptive des approches profondes avec la rigueur des outils statistiques, tout en s'affranchissant de la phase d'apprentissage. Son application à l'analyse d'images issues de microscopies avancées ouvre ainsi de nouvelles perspectives pour la caractérisation fine des surfaces et l'extraction d'informations physiques essentielles.

## Références bibliographiques

---

1. Mallat (2012), DOI : 10.48550/arXiv.1101.2286
2. Bruna & Mallat (2013), DOI : 10.48550/arXiv.1203.1513
3. Sifre & Mallat (2013), Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, [https://openaccess.thecvf.com/content\\_cvpr\\_2013/papers/Sifre\\_Rotation\\_Scaling\\_and\\_2013\\_CVPR\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2013/papers/Sifre_Rotation_Scaling_and_2013_CVPR_paper.pdf)
4. Allys et al. (2019) DOI : <https://doi.org/10.1051/0004-6361/201834975>

## Contexte du poste : Modalités d'encadrement, de suivi de la formation et d'avancement des recherches du doctorant - Details on the thesis supervision

---

Co-directeur: Bruno Grandidier

Taux d'encadrement: Viven Scottez (50%), Bruno Grandidier (40%), Charles Yaacoub (10%)

## Conditions scientifiques matérielles et financières du projet de recherche

---

Le transfert de compétences issues de la recherche vers l'industrie et l'ajout de fonctionnalités au logiciel Mountains de Digital Surf fera intégralement partie des missions du doctorant.

## Objectifs de valorisation des travaux de recherche du doctorant : diffusion, publication et confidentialité, droit à la propriété intellectuelle,...

---

Publications scientifiques et présentations en conférences internationales

Protection des résultats (confidentialité, brevets, logiciels)

Valorisation industrielle et transfert technologique

## Profil et compétences recherchées - Profile and skills required

---

Master (ou équivalent) en data/IA avec une forte appétence pour la physique

ou

Master (ou équivalent) en physique des matériaux avec une forte appétence pour l'informatique

Master's degree (or equivalent) in data science/AI with a strong interest in physics

or

Master's degree (or equivalent) in materials physics with a strong interest in computer science

Dernière mise à jour le 6 mai 2026