

# RAPPORT D'ACTIVITÉ

---

**Centre de calculs  
AZZURRA**

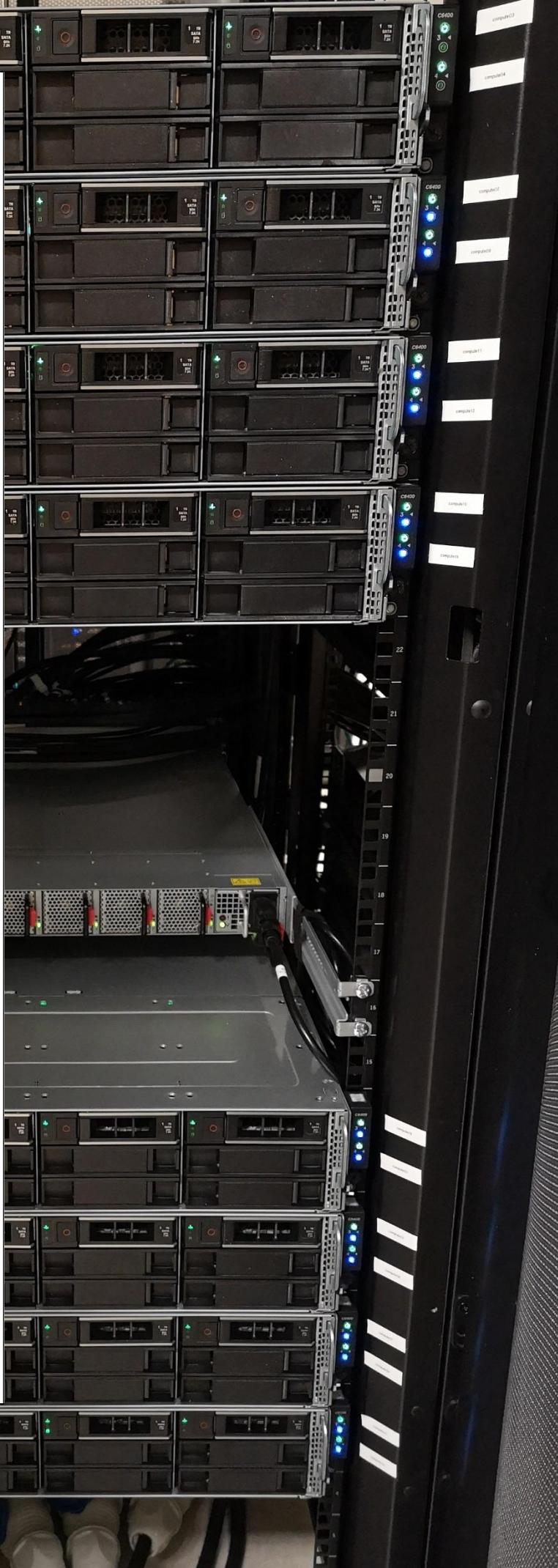


**UNIVERSITÉ  
CÔTE D'AZUR**

---

2023

**Maeva ANTOINE**



---

## Table des matières

AZZURRA EN CHIFFRES.....	3
FAITS MARQUANTS.....	4
UTILISATEURS.....	5
STATISTIQUES D'UTILISATION.....	6
FORMATIONS.....	8
FORMATIONS ET COURS AYANT EU ACCES À AZZURRA.....	9
COMMUNICATION AUTOUR D'AZZURRA.....	10
BILAN SCIENTIFIQUE.....	11
RÉCAPITULATIF DES PROJETS DE RECHERCHE.....	13
RECAPITULATIF DES PUBLICATIONS.....	15
DESCRIPTIF DES PROJETS.....	21
Projets impliquant des ingénieurs MSI.....	22
Chimie .....	49
Physique non linéaire, fluides complexes .....	73
Optique .....	95
Dynamique des fluides .....	100
Géophysique interne .....	103
Mathématiques, informatique.....	120
Magnétohydrodynamique.....	142
Astrophysique .....	151
Biologie, santé.....	174
Droit, économie .....	204
OPAL .....	210
Projets accrédités OPAL par l'université ne calculant pas sur Azzurra .....	211

---

# AZZURRA EN CHIFFRES

**49** nœuds dédiés au calcul CPU, soit **1 864** cœurs.

**3** nœuds dédiés au calcul GPU, soit **8** cartes GPU Tesla V100 et **4** cartes A100.

**320** To pour le stockage des données.

**701 201,00 €** HT de matériel, prestations d'installation / support et licence logicielle depuis 2019.

En 2023 :

**8 732 763** heures de calcul CPU ont été utilisées. (**+22%** par rapport à 2022)

**63** projets de recherche ont bénéficié d'une allocation de ressources. (**+50 %**)

**143** personnels ont bénéficié d'un compte utilisateur sur Azzurra. (**+34 %**)

**100** étudiants ont bénéficié d'un compte dans le cadre d'un cours. (**+809%**)

**156** étudiants et personnels UniCA ont été accrédités OPAL par l'université. (**+14 %**)

Depuis son lancement en mai 2020, **23 964 887** heures de calcul CPU ont été utilisées.

---

# FAITS MARQUANTS

## Février 2023 :

- Désactivation de l'authentification par mot de passe. Pour des raisons de sécurité, seule une clé SSH permet désormais la connexion au cluster.
- Utilisation d'Azzurra par les étudiants du M2 HYDROPROTECH pour leur participation au [challenge HydroEurope](#).

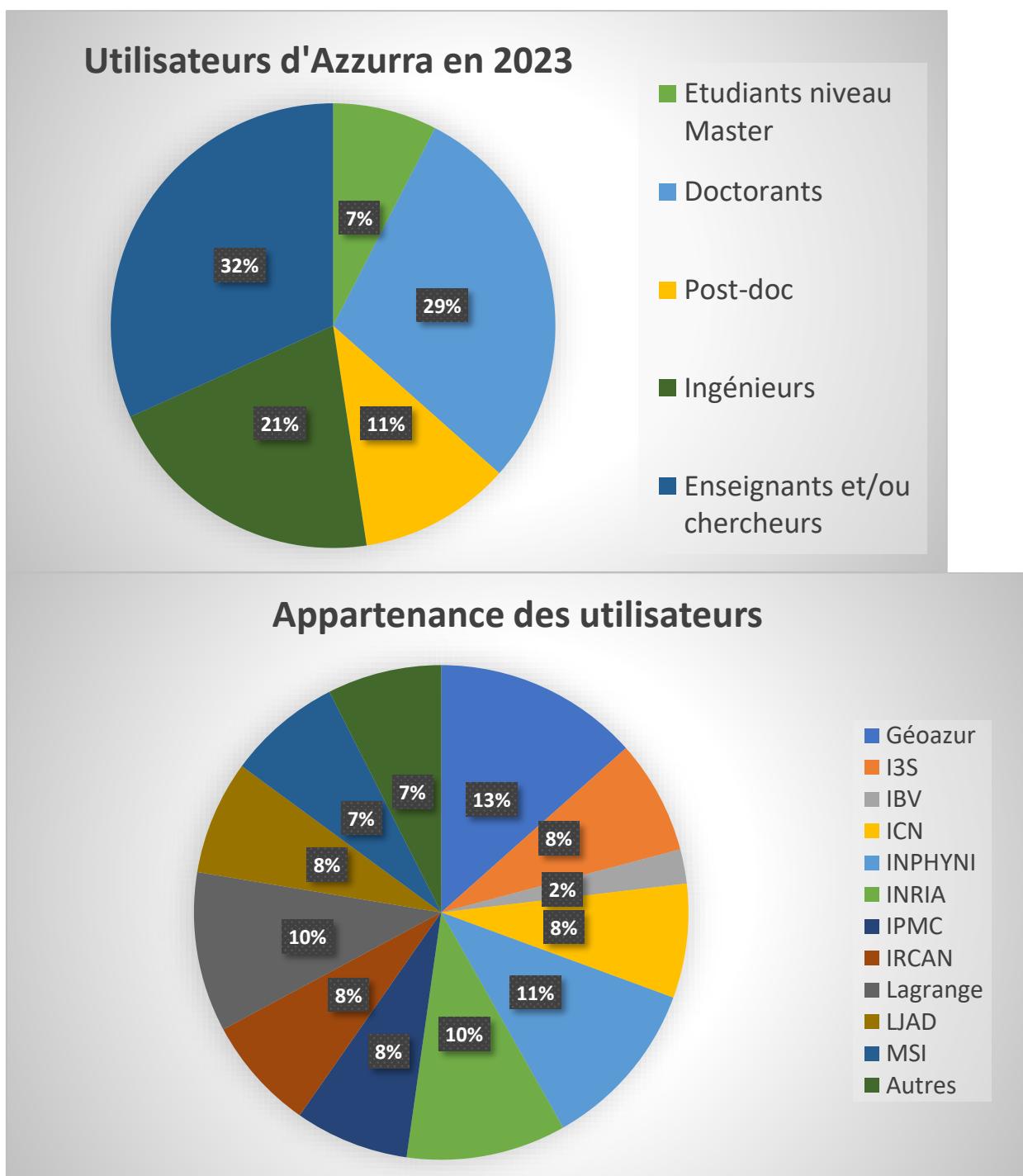
## Décembre 2023 :

Ouverture de comptes temporaires dans le cadre de cours pour 92 étudiants de Master 2 INUM, HydroProtech et GE5 Polytech'Nice-Sophia.

# UTILISATEURS

Durant l'année 2023, **143** personnes ont bénéficié d'un compte utilisateur sur Azzurra.  
**(+34 % par rapport à 2022)**

Le profil des utilisateurs est le suivant :



# STATISTIQUES D'UTILISATION

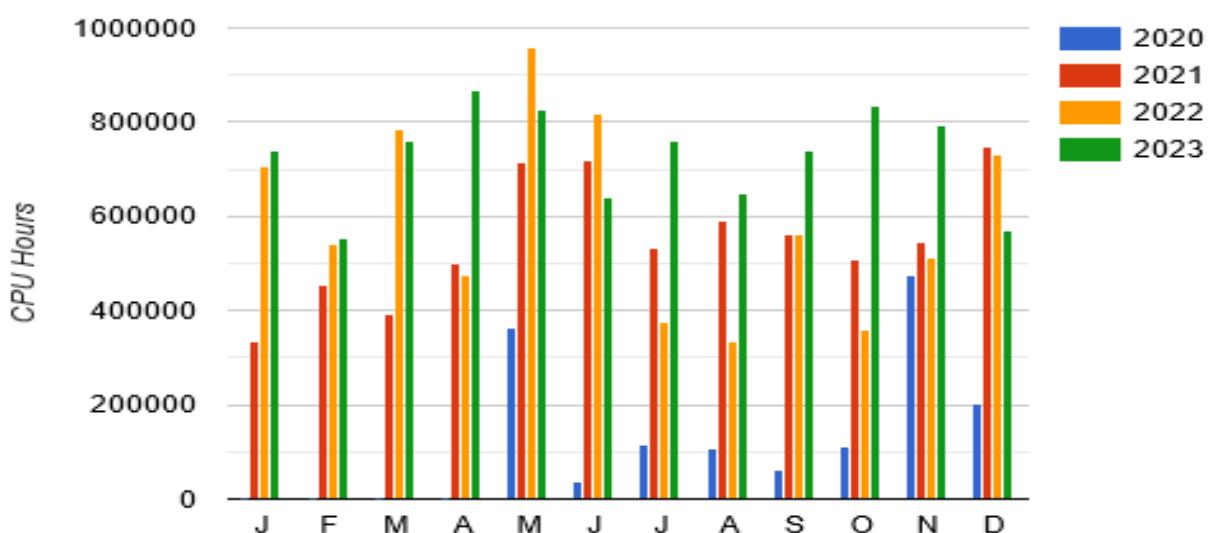
## Heures consommées

Depuis son lancement officiel en mai 2020, **23 964 887** heures de calcul CPU ont été consommées.

En 2023, ce sont **8 732 763** heures qui ont été utilisées (+ 22% par rapport à 2022).

**2023**

<i>Janvier</i>	741174
<i>Février</i>	551931
<i>Mars</i>	758381
<i>Avril</i>	868969
<i>Mai</i>	826597
<i>Juin</i>	640773
<i>Juillet</i>	759334
<i>Août</i>	648753
<i>Septembre</i>	739968
<i>Octobre</i>	832819
<i>Novembre</i>	794380
<i>Décembre</i>	569684



# OCCUPATION DU CLUSTER

## CPU

Environ **1 511 808** heures disponibles par mois soit 48768 heures de calcul disponibles quotidiennement.

Taux d'occupation du cluster (CPU) sur l'année : **49,07%**

**2023**

<i>janvier</i>	49.03%
<i>février</i>	40.42%
<i>mars</i>	50.23%
<i>avril</i>	59.39%
<i>mai</i>	54.68%
<i>juin</i>	43.80%
<i>juillet</i>	50.23%
<i>août</i>	42.91%
<i>septembre</i>	50.58%
<i>octobre</i>	55.09%
<i>novembre</i>	54.30%
<i>décembre</i>	37.68%

## GPU

8640 heures GPU disponibles par mois.

Taux d'occupation des ressources GPU sur l'année : **34,38%**

**2023**

<i>janvier</i>	32.97%
<i>février</i>	31.20%
<i>mars</i>	42.69%
<i>avril</i>	18.88%
<i>mai</i>	15.94%
<i>juin</i>	55.94%
<i>juillet</i>	39.07%
<i>août</i>	10.78%
<i>septembre</i>	24.30%
<i>octobre</i>	48.47%
<i>novembre</i>	43.33%
<i>décembre</i>	48.85%

# FORMATIONS



Depuis 2021, Azzurra est utilisée comme plateforme servant de support à différentes formations. Des comptes temporaires sur Azzurra sont fournis aux inscrits le temps du cours.

---

## FORMATIONS ET COURS AYANT EU ACCES À AZZURRA

- **Telemac et parallélisme**

**2022-2023 :**

- Cours à Polytech et challenge HydroEurope 2023
- M2 HydroProtech
- Nombre de participants : 8
- Informations : sur le [site du Master](#) et le [site du challenge HydroEurope](#)

**2023-2024 :**

- GE5 Polytech'Nice-Sophia
- M2 HydroProtech
- Nombre de participants : 82

- **Calcul haute performance**

- MASTER2-INUM
- Nombre de participants : 10

---

# COMMUNICATION AUTOUR D'AZZURRA

L'actualité relative à Azzurra est relayée via différents canaux de communication au sein d'UniCA. Quelques exemples cette année :

## **Compte X de la MSI :**

[https://x.com/MSI\\_UCA/status/1684180447633326081?s=20](https://x.com/MSI_UCA/status/1684180447633326081?s=20)

[https://x.com/MSI\\_UCA/status/1618629185316225026?s=20](https://x.com/MSI_UCA/status/1618629185316225026?s=20)

## **Mention d'articles dans les Newsletters INDEX de :**

Mars 2023

Novembre 2023

## **Articles régulièrement postés sur le site UniCA MSI :**

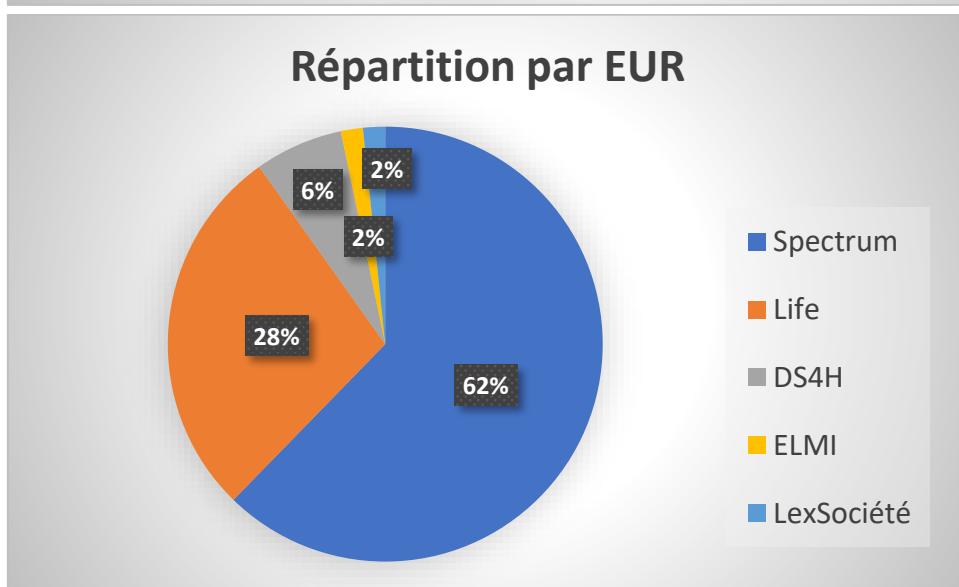
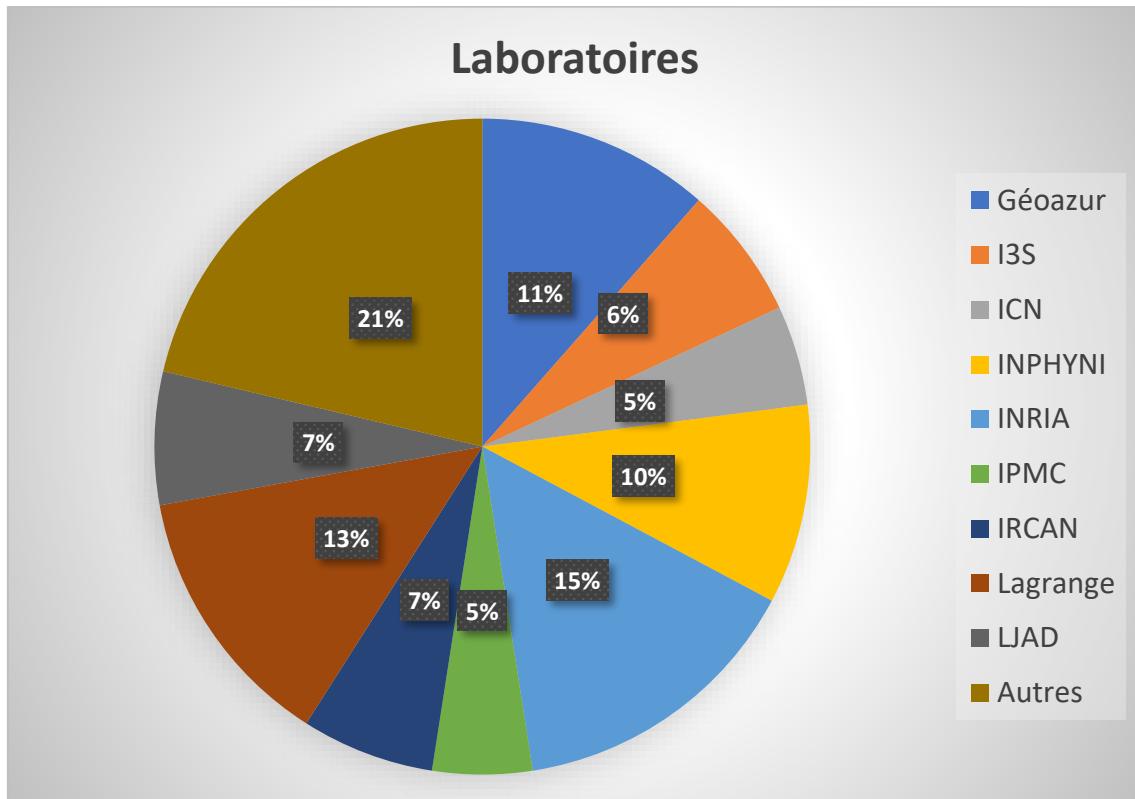
<https://msi.univ-cotedazur.fr/>

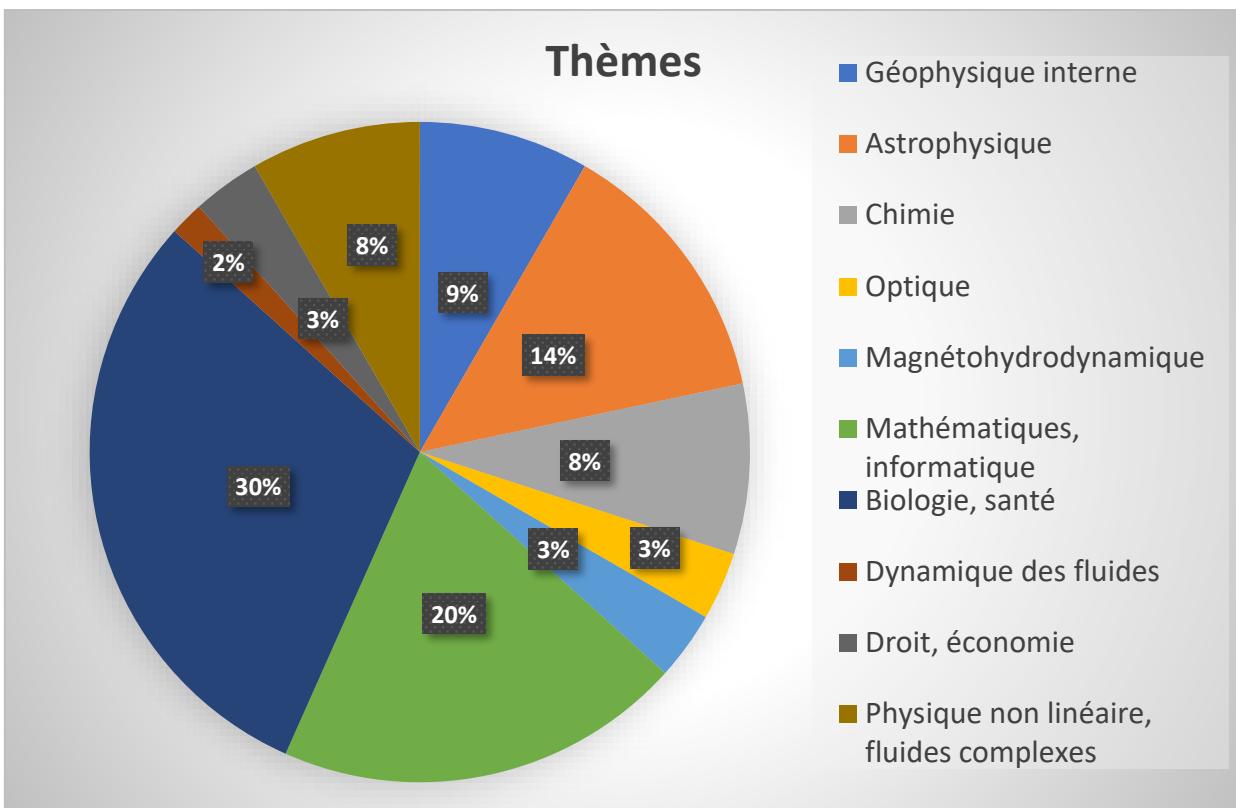
Mars : <https://newsroom.univ-cotedazur.fr/newsletter-index/mesocentre-de-calcul-azzurra-le-rapport-dactivite-2022-est-consultable-en-ligne>

Octobre : <https://msi.univ-cotedazur.fr/a-propos/vie-de-la-msi/actualites/reconnaissance-des-ecritures-imprimees-et-manuscrites-des-documents-chinois-anciens-le-projet-de-recherche-numerica-sinologica>

# BILAN SCIENTIFIQUE

En 2023, **63** projets ont bénéficié d'une allocation de ressources sur Azzurra (**+50%** par rapport à 2022). Ces projets se répartissent ainsi :





#### **Dans le détail :**

Géophysique interne (Géoazur; EUR Spectrum)	5
Astrophysique (Géoazur, Lagrange; EUR Spectrum)	8
Chimie (ICN, LJAD; EUR Spectrum)	5
Optique (Inphyni; EUR Spectrum)	2
Magnétohydrodynamique (Lagrange, LJAD; EUR Spectrum)	2
Mathématiques, informatique (I3S, Inria, LJAD; EUR DS4H, Spectrum)	12
Biologie, santé (IPMC, ISA; EUR Life)	18
Dynamique des fluides (CEMEF, LJAD; EUR Spectrum)	1
Droit, économie (ERMES; EUR LexSociété – INSTAR ; EUR ELMI)	2
Physique non linéaire, fluides complexes (Inphyni; EUR Spectrum)	6

# RÉCAPITULATIF DES PROJETS DE RECHERCHE

Porteur	Laboratoire	Discipline	EUR	CLASSIFICATION
Chiavassa	Lagrange	Astrophysique	Spectrum	Astrophysique
Fienga	Geoazur	Astrophysique	Spectrum	Astrophysique
Fienga	Geoazur	Astrophysique	Spectrum	Astrophysique
Mary	Lagrange	Astrophysique	Spectrum	Astrophysique
Méheut	Lagrange	Astrophysique	Spectrum	Astrophysique
Méheut	Lagrange	Astrophysique	Spectrum	Astrophysique
Parmentier	Lagrange	Astrophysique	Spectrum	Astrophysique
Rousseau	Lagrange	Astrophysique	Spectrum	Astrophysique
Becavin	IPMC	Pharmacologie	Life	Biologie, santé
Besse	IBV	Biologie	Life	Biologie, santé
Blin	LP2M	Médecine	Life	Biologie, santé
Bottini	INRAE	Biologie	Life	Biologie, santé
Bottini	INRAE	Médecine	Life	Biologie, santé
Bottini	MD lab / MSI	Médecine	Life	Biologie, santé
Clampitt	IRCAN / MSI	Biologie	Life	Biologie, santé
Cristofari	IRCAN	Médecine	Life	Biologie, santé
Croce	IRCAN	Médecine	Life	Biologie, santé
Danchin	Agrobiotech	Génomique	Life	Biologie, santé
Gautier	IPMC	Pharmacologie	Life	Biologie, santé
Glaichenhaus	IPMC	Pharmacologie	Life	Biologie, santé
Gauci	UR2CA-iULS	Médecine	Life	Biologie, santé
Hubstenberger	IBV	Biologie	Life	Biologie, santé
Hug	LAMHESS	Médecine	DS4H	Biologie, santé
Kozlowski	INRAE / MSI	Biologie	Life	Biologie, santé
Liti	IRCAN	Médecine	Life	Biologie, santé
Roux	IRCAN	Médecine	Life	Biologie, santé
Cassam-Chenaï	IJAD	Chimie	Spectrum	Chimie
Fiorucci	ICN	Chimie	Spectrum	Chimie
Fontaine-Vive	ICN	Chimie	Spectrum	Chimie
Nothias	ICN	Chimie	Spectrum	Chimie
Henry	Inria	Chimie/Mathématiques	Spectrum	Chimie
Kalugin	-	Chimie	-	Chimie
Kyrychenko	-	Chimie	-	Chimie
Constant	Ermes	Droit	LexSociété	Droit, économie
Czellar	INSTAR	Economie	ELMI	Droit, économie
Vincenzi	IJAD	Mécanique des fluides	Spectrum	Dynamique des fluides
Ampuero	Geoazur	Géophysique	Spectrum	Géophysique interne
Duclaux	Geoazur	Géophysique	Spectrum	Géophysique interne
Operto	Geoazur	Géophysique	Spectrum	Géophysique interne
Sigloch	Geoazur	Géophysique	Spectrum	Géophysique interne
Sladen	Geoazur	Géophysique	Spectrum	Géophysique interne
Marcotte	IJAD	Mathématiques	Spectrum	Magnétohydrodynamique
Ponty	Lagrange	Mécanique des fluides	Spectrum	Magnétohydrodynamique
Bossy	Inria	Mathématiques	Spectrum	Mathématiques, informatique

Dolean	IJAD	Mathématiques	Spectrum	Mathématiques, informatique
Comet	I3S	Biologie/Informatique	Spectrum	Mathématiques, informatique
Gandon	Inria	Informatique	Spectrum	Mathématiques, informatique
Lanteri	Inria	Mathématiques	Spectrum	Mathématiques, informatique
Malis	Inria	Informatique	Spectrum	Mathématiques, informatique
Martinet	I3S	Informatique	Spectrum	Mathématiques, informatique
Pasquier	I3S	Informatique	DS4H	Mathématiques, informatique
Simoes	I3S / MSI	Informatique	Spectrum	Mathématiques, informatique
Villata	Inria	Informatique	DS4H	Mathématiques, informatique
Villata	Inria	Informatique	DS4H	Mathématiques, informatique
Villata	Inria	Informatique	DS4H	Mathématiques, informatique
Aubry	InPhyNi	Physique	Spectrum	Optique
Doya	InPhyNi	Physique	Spectrum	Optique
Simonnet	InPhyNi	Physique	Spectrum	Physique non linéaire, fluides complexes
Bec	Inria	Physique/Mécanique des fluides	Spectrum	Physique non linéaire, fluides complexes
Krstulovic	Lagrange	Physique/Mécanique des fluides	Spectrum	Physique non linéaire, fluides complexes
Nazarenko	InPhyNi	Physique/Mécanique des fluides	Spectrum	Physique non linéaire, fluides complexes
Nazarenko	InPhyNi	Physique/Mécanique des fluides	Spectrum	Physique non linéaire, fluides complexes
Peters	InPhyNi	Physique/Mécanique des fluides	Spectrum	Physique non linéaire, fluides complexes

---

# RECAPITULATIF DES PUBLICATIONS

## Journaux :

- Bandgap fluctuations and robustness in two-dimensional hyperuniform dielectric materials. Luis Froufe, Geoffroy Aubry, Frank Scheffold, and Sofia Magkiriadou. <https://doi.org/10.1364/OE.484232>
- Bayesian test of the mass of the graviton with planetary ephemerides. Vincenzo Mariani, Agnès Fienga, Olivier Minazzoli, Mickaël Gastineau, Jacques Laskar. Accepted for publication in Physical Review D. 21 pages. <https://doi.org/10.48550/arXiv.2306.07069>
- Foiselle, M., S. Barbosa, O. Godin, C. L. Wu, W. Boukouaci, M. Andre, B. Aouizerate, F. Berna, C. Barau, D. Capdevielle, P. Vidailhet, I. Chereau, L. Davidovic, J. M. Dorey, C. Dubertret, J. Dubreucq, C. Faget, G. Fond, S. Leigner, P. M. Llorca, J. Mallet, D. Misdrahi, E. Martinuzzi, C. Passerieux, R. Rey, B. Pignon, M. Urbach, F. Schurhoff, N. Glaichenhaus, M. Leboyer, R. Tamouza and F.-S. Groups (2022). "Immuno-metabolic profile of patients with psychotic disorders and metabolic syndrome. Results from the FACE-SZ cohort." Brain Behav Immun Health 22: 100436.
- Barbosa, S., E. Martinuzzi, C. Gilet, L. Fillatre, O. Khalfallah, I. Sommer, S. Leucht, P. Dazzan, P. McGuire, C. Arango, C. M. Diaz-Caneja, W. Fleischhacker, D. Rujescu, B. Glenthøj, I. Winter, R. S. Kahn, R. Yolken, S. Lewis, R. Drake, L. Davidovic, M. Leboyer, N. Glaichenhaus and O. S. Group (2021). "Baseline Levels of C-Reactive Protein and Proinflammatory Cytokines Are Not Associated With Early Response to Amisulpride in Patients With First Episode Psychosis: The OPTiMiSE Cohort Study." Schizophrenia Bulletin Open, 2021, 10.1093/schizbulopen/s gab017. hal-03286370
- Mahmoud M R Elsawy, Anthony Gourdin, Mickael Binois, Régis Duvigneau, Didier Felbacq, et al. Multiobjective statistical learning optimization of RGB metalens. ACS photonics, 2021, 8 (8), pp.2498-2508. (10.1021/acsphotonics.1c00753). (hal-03212349v2)
- Mahmoud Elsawy, Mickael Binois, Régis Duvigneau, Stéphane Lanteri, Patrice Genevet. Optimization of metasurfaces under geometrical uncertainty using statistical learning. Optics Express, 2021, 29 (19), pp.29887. (10.1364/OE.430409). (hal-03356986)
- Pasquier, C., & Robichon, A. (2022). Evolutionary divergence of phosphorylation to regulate interactive protein networks in lower and higher species. International Journal of Molecular Sciences, 23(22), 14429. hal-03881865
- Pasquier, C., Guerlais, V., Pallez, D., Rapetti-Mauss, R., & Soriani, O. (2023). A network embedding approach to identify active modules in biological interaction networks. Life Science Alliance, 6(9). hal-04135439
- "First-Principles Analysis of CuMg<sub>2</sub>InS<sub>4</sub>: Insights into Optical, Piezoelectric, and Thermoelectric Properties", Sadouki, O., Khelfaoui, F., Fontaine-Vive F. et al., J. Electron. Mater., 2023, 52, 6778–6790
- "New phenylspirodrimanes from the Sponge-Associated Fungus *Stachybotrys chartarum* MUT 3308", M. Dayras, F. Fontaine-Vive, M. Mehiri et al., Marine Drugs, 2023, 21(3), 135

- Testing the mass of the graviton with Bayesian planetary numerical ephemerides B-INPOP. Vincenzo Mariani, A. Fienga, O. Minazzoli. Accepted by Astronomy & Astrophysics  
<https://doi.org/10.48550/arXiv.2310.00719>
- Fanny Simões, Charles Bouveyron, Frédéric Precioso, DeepWILD: Wildlife Identification, Localisation and estimation on camera trap videos using Deep learning, Ecological Informatics, Volume 75, 2023, 102095, ISSN 1574-9541,  
<https://doi.org/10.1016/j.ecoinf.2023.102095>
- Laure Martinelli, Camille Bihanic, Aurélie Bony, Florence Gros, Corentin Conart, Sébastien Fiorucci, Hervé Casabianca, Frédéric Schiets, Giorgiana Chietera, Benoît Boachon, Bernard Blerot, Sylvie Baudino, Frédéric Jullien, Denis Saint-Marcoux, Citronellol biosynthesis in pelargonium is a multistep pathway involving progesterone 5  $\beta$  -reductase and/or iridoid synthase-like enzymes, Plant Physiology, 2023;, kiad550,  
<https://doi.org/10.1093/plphys/kiad550>
- Ligand Binding Properties of Odorant-Binding Protein OBP5 from *Mus musculus*. L. Moitrier, C. Belloir, M. Lalil, Y. Hou, J. Topin, L. Briand. Biology, 2022, 12(1), 2.
- Extracellular loop 2 of G protein–coupled olfactory receptors is critical for odorant recognition. Y. Yu, Z. Ma, J. Pacalon, L. Xu, W. Li, C. Belloir, J. Topin, L. Briand, J. Golebiowski, X. Cong. Journal of Biological Chemistry, 2022, 298(9).
- Exploring Dihydroflavonol-4-reductase reactivity and selectivity by QM/MM-MD simulations. J. Diharce, E. Bignon, S. Fiorucci, S. Antonczak, ChemBioChem, 2022, 23(3), e202100553.
- Zhu Y, Semisalov B, Krstulovic G, et al. Direct and Inverse Cascades in Turbulent Bose-Einstein Condensates[J]. Physical Review Letters, 2023, 130(13): 133001. On the cover .  
<https://link.aps.org/doi/10.1103/PhysRevLett.130.133001>
- Picardo JR, Plan ELCVM, Vincenzi D. Polymers in turbulence: stretching statistics and the role of extreme strain rate fluctuations. Journal of Fluid Mechanics. 2023;969:A24. doi:10.1017/jfm.2023.524
- Wu, Y., Aghamiry, H. S., Operto, S., & Ma, J. (2023). Helmholtz-equation solution in nonsmooth media by a physics-informed neural network incorporating quadratic terms and a perfectly matching layer condition. *Geophysics*, 88(4), T185-T202.

## Conférences :

- Fallacious Argument Classification in Political Debates. Pierpaolo Goffredo, Shohreh Haddadan, Vorakit Vorakitphan, Elena Cabrio, Serena Villata. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence Main Track. Pages 4143-4149.  
<https://doi.org/10.24963/ijcai.2022/575>
- Testing the mass of the graviton with Bayesian planetary numerical ephemerides B-INPOP. Vincenzo Mariani, A. Fienga, O. Minazzoli. Rencontres de Moriond 2023 (GRAM).
- Jourdon, A., May, D. A., and Gabriel, A. A.: Generalization of the Nitsche method to apply oblique boundary conditions in regional geodynamic models, EGU General Assembly

---

2023, Vienna, Austria, 24–28 Apr 2023, EGU23-14864, <https://doi.org/10.5194/egusphere-egu23-14864>, 2023.

- Michel Orsi, Elisabeth Lemaire, Laurent Lobry, and François Peters. Simulation of pressure-driven flow of non-brownian suspensions : Local constitutive laws. In EuroMech Symposium- Suspension flows and rheology : inertia, shape and roughness matter, Nice, France, June 2023
- Michel Orsi, Laurent Lobry, and François Peters NON-BROWNIAN SUSPENSIONS IN PRESSURE-DRIVENFLOW: NUMERICAL SIMULATIONS WITHFRAME-INVARIANT SUB-GRID CORRECTIONS. 8TH PACIFIC RIM CONFERENCE ON RHEOLOGY, May 15-19, 2023
- M. Binois, R. Duvigneau, Mahmoud M. R. Elsayy, P. Genevet, S. Kadir, et al. Advanced Numerical Modeling Methods for the Characterization and Optimization of Metasurfaces. AT-AP-RASC 2022 - 3rd URSI Atlantic and Asia Pacific Radio Science Meeting, May 2022, Gran Canaria, Spain. pp.1-4, (10.23919/AT-AP-RASC54737.2022.9814395). (hal-03864889)
- Mahmoud M R Elsayy, Mickaël Binois, Régis Duvigneau, Samira Kadir, Anthony Gourdin, et al. Multiobjective optimization for highly efficient and robust metasurface designs. OPTIQUE Nice 2022, Jul 2022, Nice, France. (hal-03864836)
- Mahmoud M. R. Elsayy, Christina Kyrou, Elena Mikheeva, Rémi Colom, Jean-Yves Duboz, et al. Universal active metasurface modulation with ultimate performance in reflection. Nanophotonics and Micro/Nano Optics International Conference 2022, Oct 2022, Paris, France. (hal-03864856)
- Bastien Buchwalter, Veronika Czellar and Engin Iyidogan. Multifractal Cryptocurrencies. Présenté aux conférences et séminaires suivants :
  - Fifth International Workshop in Financial Econometrics, Vila Angatu Eco Resort, Santo André (Brazil). Octobre 2023
  - Conference on Forecasting Financial Markets: ADVANCES FOR EXCHANGE RATES, INTEREST RATES AND ASSET MANAGEMENT, Rennes (France). Juin 2023
  - French Inter-Business School Conference and Forecasting Financial Markets Conference, Toulouse (France). Mars 2023
  - Finance Seminar, University of Luxembourg (Luxembourg). Janvier 2023
  - Finance Seminar, Rennes School of Business, Rennes (France). Janvier 2023
  - SKEMA Finance Seminar, Paris (France). Novembre 2022
  - ESSEC Asset Pricing Breakfast, Paris (France). Novembre 2022
- Lombardi, R., Jozwiak, M., Dellamonica, J., & Pasquier, C. (2023, June). A machine learning approach to predict weaning outcome among ventilated patients in Intensive Care unit. In Reanimation 2023, the French Intensive Care Society International Congress (Vol. 13). hal-04160272
- Clustering of heavy particles in Keplerian turbulence. Presented by Fabiola Gerosa at Geophysical and Astrophysical Fluid Dynamics 2023.
- Clusters of heavy particles in two-dimensional Keplerian turbulence. Presented by Fabiola Gerosa at Particle Growth in Turbulence 2023 and EUROMECH COLLOQUIUM 628 Complex Particles in Turbulent Flow 2023.

- Minimal seeds for subcritical dynamos. Complex Days, Nice, janvier 2023 : contribution orale de F. Marcotte
- Matching receptor to odorant with protein language and graph neural networks. M. Hladiš, M. Lalis, S. Fiorucci, J. Topin. International Conference on Learning Representation (ICLR), 2023.
- Une IA peut-elle apprendre à sentir ? S. Fiorucci. IADATES - "Intelligence Artificielle, Arômes et Parfum", Grasse, 13 février 2023.
- Ligand-based virtual screening of membrane proteins. S. Fiorucci Winterschool Biomembranes Pharmacology and Therapeutics, Nice, 15 décembre 2022.
- Numerical models decipher odorant binding and receptor activation at the molecular level. M. Hladiš, J. Pacalon, M. Lalis, S. Fiorucci, J. Topin. Complex Day, Nice (France), February 2023.
- Wu, Yanqi & Aghamiry, Hossein & Operto, Stéphane & Ma, Jianwei. (2023). Frequency-Domain Wave Simulation Using Physics-Informed Neural Networks (PINNs) with Free Surface Boundary Condition. [84th EAGE Annual Conference & Exhibition](#), Jun 2023, Volume 2023, p.1 – 5. <https://doi.org/10.3997/2214-4609.202310403>
- E. Sopaci and J. P. Ampuero (2024), "Triggering of very shallow earthquakes by surface mass removal processes - case study of the 2019 Mw4.9 Le Teil, France earthquake", European Geophysical Union General Assembly 2024
- J. Premus and J. P. Ampuero (2024), "Dynamic earthquake source inversion with GAN priors, with application to the 2023 Mw 7.8 Kahramanmaraş, Turkey earthquake", European Geophysical Union General Assembly 2024

## Workshops et autres présentations orales :

- Pierpaolo Goffredo, Elena Cabrio, Serena Villata, Shohreh Haddadan and Jhonatan Torres Sanchez. DISPUTTool 2.0: A Modular Architecture for Multi-Layer Argumentative Analysis of Political Debates. AAAI 2023 Demo paper
- Olfactory stimulation for the treatment of apathy in neurodegenerative disorders. J. Topin, R. David. Annual Meeting of the NeuroMod Institute, Antibes (France), June 2022.
- Université de Newcastle, octobre 2023 : séminaire de F. Marcotte
- From genes to perception: decoding chemical senses by numerical and molecular modeling approaches. S. Fiorucci, Chimie de la matière complexe (UMR7140 CNRS) Strasbourg, 9 mars 2023.
- Café art/science regards scientifiques : la transmission de savoir-faire olfactifs. Le Hublot Nice, Décembre 2022.
- Forum scientifique - Café rencontre : L'IA au service de la création parfum, mythe ou réalité ? Musée International de la Parfumerie Grasse, Octobre 2022.

---

## **Thèses :**

- Dynamique de particules allongées dans des écoulements tourbillonnaires et turbulence élastique dans les solutions de polymères. Sumithra Reddy YERASI, Laboratoire Jean Alexandre Dieudonné, Université Côte d'Azur. Soutenue le 2 mai 2023.
- Thèse de Megan Clampitt : Indexation de l'état de santé des coraux par une approche basée sur l'intelligence artificielle. Soutenue le 04-09-2023 à l'Université Côte d'Azur , dans le cadre de l'École doctorale Sciences de la vie et de la santé.
- Thèse de Mattis Reisner : Études expérimentales de la multifractalité et des transitions de phase topologiques dans des réseaux de résonateurs micro-ondes. Soutenue le 01-03-2023 à l'Université Côte d'Azur, dans le cadre de l'École doctorale Sciences fondamentales et appliquées.
- Compétition entre la solvatation et l'agrégation ionique dans des systèmes ioniques-solvant: influence sur les propriétés de transport. Dmytro DUDARIEV. 2023. Université de Lille (dans le cadre de l'action de solidarité UniCA - Université de Kharkiv).

## **Posters :**

- Using bulk RNA-seq and iClip-seq analysis to investigate mRNA localization in synapses and uncover the role of the IMP protein. Hiba Laghrissi, Lauren Blot, Martina Halleger, Jernej Ule, Arnaud Hubstenberger & Florence Besse. POSTER JOBIM 2023 - Nice (FRANCE)
- New insights on planetesimal formation in turbulent disks. Fabiola Antonietta Gerosa, Héloïse Meheut, Jérémie Bec. The inner disk of young stars (SPIDI 2023)
- "Theoretical study of the vibrational spectra of Naphtalene and its cation with the VMFCI method", (NanoSpace Joint Scientific Meeting, Ljubljana, 9-11/02/2023, Slovénie).
- Comparisons between tidal deformation modeling from the semi-analytical ALMA3 and the finite element CitcomSVE. Agnes Fienga, Observatoire de la Côte d'Azur, Géoazur, Valbonne, France, Arthur Briaud, CNRS, Observatoire de la Cote d'azur, Geoazur, Valbonne, France, Shijie Zhong, University of Colorado at Boulder, Department of Physics, Boulder, CO, United States and Anthony Mémin, Université Côte d'Azur, CNRS, OCA, IRD, Géoazur, Nice, France. AGU 2023.
- Tidal deformations and internal structure: sensitivity to rheologies. Agnes Fienga, Observatoire de la Côte d'Azur, Géoazur, Valbonne, France, Alex Guinard, Géoazur - Université Nice Sophia Antipolis, Valbonne, France, Anthony Mémin, Université Côte d'Azur, CNRS, OCA, IRD, Géoazur, Nice, France, Arthur Briaud, CNRS, Observatoire de la Cote d'azur, Geoazur, Valbonne, France and Clément Ganino, Université Côte d'Azur, OCA, CNRS, Géoazur, Valbonne, France. AGU 2023.
- Recognizing emotions triggered by fragrances using virtual reality and physiological measurement. X. Meleqi, U Meierhenrich, A. Pegard, J. Topin. ECRO 2022, Berlin (Allemagne), 31 août-3 septembre 2022.

- 
- Deciphering the ligand binding properties of the mouse odorant-binding protein OBP5 from *Mus musculus*. L. Moitrier, C. Belloir, M. Lalis, Y. Hou, J. Topin, L. Briand. ECRO 2022, Berlin (Allemagne), 31 août-3 septembre 2022.

## Stages et projets étudiants niveau master :

- Carolina Paz Gutiérrez Morales. Inversión bayesiana en altas frecuencias de la distribución de deslizamiento del terremoto de Tocopilla 2007, Mw 7.7. Rapport de master « Magíster en Ciencias, Mención Geofísica », Universidad de Chile (2022).  
<https://repositorio.uchile.cl/handle/2250/187604>
- Romain Lombardi. Prédiction du succès du test de sevrage ventilatoire en service de réanimation par Machine Learning. Rapport de stage Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023
- Rapport de stage de Sonya Joyce Ines NDOMKEU TCHATTO « Construction d'une base de données et segmentation automatique d'avant-bras par Deep Learning: premières étapes de la planification préopératoire. » Polytech Grenoble, 2023.
- Mame Seynabou Fall. Développement d'un outil pour l'analyse des données multi-omiques afin d'améliorer la compréhension de la pathophysiologie du syndrome de l'X fragile. Rapport de stage Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023
- Ercan Seçkin. Origine, histoire évolutive et prédiction des gènes orphelins dans les génomes de nématodes de ravageurs de culture. Rapport de stage Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023
- Maxime Multari. Finding signatures associated to the mechanism of pathogen resistance bypass by network inference using multi-omics data integration. Rapport de stage Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023
- Youssef Boulaimen. VIOLA: a novel strategy for Variant prioritization using Latent space in the context of mitochondrial diseases. Rapport de stage Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023
- Rapport de stage de Anjana Bhat. MSc2 in Data Science and AI. Université Côte d'Azur 2023
- Rapport de stage de Madina Bekbergenova. Université Côte d'Azur 2023

---

## **DESCRIPTIF DES PROJETS**

---

## **Projets impliquant des ingénieurs MSI**

---

La Maison de la Modélisation, de la Simulation et des Interactions (MSI) doit sa création au projet UCA<sup>JEDI</sup> pour lequel en 2016 la ComUE Université Côte d'Azur a remporté le prix Initiatives d'Excellence (IDEX) du gouvernement français.

Son but : favoriser et structurer les échanges transdisciplinaires dans la recherche académique et les partenariats de recherche public privé dans la région SUD sur tous les domaines qui sollicitent des expertises fortes en matière de modélisation, de simulation et de traitement de données massives. La MSI intervient en appui du réseau d'acteurs qui constituent et animent Université Côte d'Azur.

La MSI est pilotée par un comité composé de huit chercheurs issus de différents laboratoires d'Université Côte d'Azur et présidé par un directeur nommé par le Président du projet IDEX UCA<sup>JEDI</sup> ; ce comité s'appuie sur une équipe de chercheurs et d'ingénieurs experts et sur une chargée de mission.

Les ingénieurs de la MSI s'investissent sur des projets scientifiques développés au sein de centres d'expertise de la MSI ou au niveau d'Université Côte d'Azur. De plus, dans le cadre d'appels à projets, la MSI soutient des projets de recherche interdisciplinaires et/ou tournés vers des partenariats industriels. Elle propose également une offre de formation pour chercheurs et ingénieurs via des séminaires réguliers et des écoles thématiques.

La MSI ayant financé Azzurra, ses ingénieurs ont accès au cluster pour leurs différents projets, décrits ci-dessous.

---

## ***IntegrOMICs analysis, a new angle of study for understanding the pathophysiology of the Fragile X Syndrome.***

Stage étudiant (Master bio-info UCA) dans le cadre du projet n°1 encadré par Silvia Bottini (MSI / INRAE).

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- Mame Seynabou Fall. Développement d'un outil pour l'analyse des données multi-omiques afin d'améliorer la compréhension de la pathophysiologie du syndrome de l'X fragile. Rapport de stage Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023

### **Description :**

The Fragile X-Syndrome (FXS) represents the most common inherited form of intellectual disability and the first monogenic cause of Autism Spectrum Disorders. This neurodevelopmental disorder is characterized by a broad range of neurologic/psychiatric phenotypes including mental impairment, autism, attention deficit, hyperactivity, social anxiety, epilepsy... In the majority of cases, this disease results from the absence of expression of the protein FMRP (Fragile-X Messenger RibonucleoProtein, former Fragile-X Mental Retardation Protein) encoded by the FMR1 gene. Alternatively, some patients present point mutations in the coding sequence of FMRP, which do not alter the expression of the protein but would impact its functions according to processes that often remain to be elucidated [1, 2].

To date, no effective therapy is available to treat FXS patients. The search for new FXS-specific drugs is focused on targeting bioactive molecules that act on cellular mechanisms altered in the absence of FMRP and that can be administered to patients - alone or in combination - during different life stages [3, 4]. Hence, the development of these compensatory therapies first requires a comprehensive understanding of the cellular functions of FMRP affected in FXS. Most of the FMRP protein is cytoplasmic and therefore the majority of the studies has focused on its role in this compartment. Canonically, FMRP is defined as a RNA-binding protein that positively or negatively modulates the translation of a wide range of target mRNAs, including mRNAs encoding proteins essential for the development and the function of synapses. In this context, FMRP is member of ribonucleoprotein complexes called "transport granules" that

---

promote the transport of its mRNA targets along the axons and the dendrites. Thus, the cognitive deficiencies observed in patients with FXS are thought to result, at least in part, from the deregulated expression of proteins encoded by the FMRP-bound mRNAs, thereby altering the cellular pathways in which they participate.

However, the protein also has a nuclear localization signal, a nuclear export signal and a nucleolar localization signal allowing it to navigate between the cytoplasm and the nucleus. Strikingly, a pathogenic mutation within the nuclear export signal of FMRP, R442Q, has been recently found in some FXS patients [5, 6]. Expression of the missense variant in cell line shows normal expression but significant retention of FMRP in the cell nucleus [5]. These findings suggest that the disruption of the nucleocytoplasmic shuttling of FMRP have a deleterious effect on neuronal functions, possibly through the interference of some FMRP-mediated cellular processes in the nucleus. Consistently, a growing number of studies now associate FMRP with nuclear functions, such as DNA repair [7-9], nuclear pore assembly [10], ribosomal RNA methylation [11] or certain steps in mRNA maturation and export [12-22]. Notably, splicing recently emerged as a key nuclear process regulated by FMRP [16, 17, 23-25]. Then, while the implication of FMRP in some nuclear steps of mRNA processing is now accepted, very few information is available concerning: (i) whose mRNAs are targeted by FMRP in the nucleus, (ii) how FMRP impacts the nuclear mRNA biogenesis processes and (iii) if this regulation varies according to the targets. To broaden our insight into the contribution of nuclear FMRP in mammalian neuronal physiology, we very recently identified a nuclear interactome of the protein by combining subcellular fractionation of rat brains with pull down affinity purification and mass spectrometry analysis [33]. In this publication, we listed 55 candidate nuclear partners. This unique dataset includes factors involved in different steps of mRNA biogenesis, as transcription, editing, splicing or nuclear export, revealing a potential central regulatory function of FMRP in the biogenesis of its target mRNAs.

It thus appears that the contribution of FMRP in neuronal physiology is extremely complex. Not only the protein is involved in various cellular processes (nuclear, cytoplasmic, synaptic) but it also affects a multitude of mRNA targets and thus potentially as many biological pathways. The development of efficient compensatory therapies to the FXS requires beforehand to compile the most comprehensive list of the cellular processes regulated by FMRP and dysregulated in the disease. In this context, large-scale OMIC screens appear particularly adapted to face this challenge. Consistently, literature provides several transcriptomics, proteomics and metabolomics studies evaluating the multiple perturbations of the cellular homeostasis in the context of the

---

FXS. However, it should be noted these individual analysis of OMICs approaches only offers a one-dimensional view of a complex system. Interestingly, the computational biology and bioinformatics community has recently developed several approaches and tools integrating results from multiple OMICs studies to associate phenotypes and extract recurrent features, in a so called integromics analysis. Such an integrative approach has never been conducted in the context of the FXS.

Thus, the present project aims to conduct an unprecedented multi-OMICs analysis by combining computational modelling and experimental validation and embedding OMICs data generated by the FXS community and our original research on the nuclear contribution of FMRP to neuronal physiology.

Multi-omics methods are emerging as valuable tools for understanding the functioning of complex diseases. Data integration is defined as a process by which data from different sources are combined statistically to make large-scale conclusions about a disease and to obtain a comprehensive view of biological processes. Omics integrative approaches increase the reliability of a biological discovery if it can be validated by concordant omics signatures (genomics, transcriptomics, and proteomics) [30]. Currently, we are working on developing a novel multi-omics integration tool based on an unsupervised learning approach [34]. Specifically, unsupervised approaches learn representations by identifying patterns in the data and extracting meaningful knowledge, while overcoming data complexities. Among such approaches, we are focusing on unsupervised learning methods via neural networks such as variational autoencoders (VAEs) [31] to capture a better representation of the data structure (i.e., intrinsic relationships between the data variables), and therefore will allow a more efficient multi-omics data integration. The project is divided in two tasks:

1. Extensive bibliographic research and database mining to collect single and multi-omics datasets

The FMR1 gene, encoding the FMRP protein, is highly conserved across species. As a result, there are many animal models for FXS: mouse, rat, zebrafish, frog, drosophila, and mollusk [32], in addition to immortalized Fmr1-KO cell lines. Besides, numerous studies were conducted on human material, such as post-mortem brain, fibroblasts, blood cells or iPSCs derived from patients. We have so far listed forty-four OMICs analyses directly connected to the FXS, through the interrogation of sites specialized in the collection of OMICs studies such as OmicsDI (<https://www.omicsdi.org/>), bibliographic exploration and

our unpublished studies. Ten of them identify mRNA or protein interactomes of FMRP and the others represent comparisons between wild-type and FXS models, at the level of their proteomes, translatomes, transcriptomes, epigenomes, metabolomes or lipidomes. The next step will be to classify these studies in terms of homogeneity of models and experimental conditions, and then in terms of accessibility and format of the data. The open science policy currently conducted worldwide guarantees the accessibility of most of these data via numerous repository sites (ProteoXchange <http://www.proteomexchange.org/>, Gene Expression Omnibus <https://www.ncbi.nlm.nih.gov/geo/> ...). A first test dataset will be selected based on its consistency at the biological level and its adequacy to meet the requirements of multi-OMICs analyses (same FXS model, adapted data format...). If necessary, analyses in a format that is not adequate for the purpose will be reformatted.

## 2. Development of a computational framework to integrate multi-omics data

A first exploratory phase in which existent multi-omics integration algorithm will be tested and benchmarked on our datasets to understand their limitation. In the second phase, we will take advantage of the previous phase to adapt our VAE integromics approach to be efficient for our kind of data. Here the challenge will be to identify the optimal VAE configuration through parameter tuning and the proper definition of the cost function, which requires the use of a risk estimation function. The multi-omics integration through VAE methods will allow to validate or to propose novel signatures to understand the etiology of the FXS. On the other side, the exploration of the latent space toward clustering algorithms will allow to identify the candidates with modulation at single-omic level or any combination of two or more multi-omics leading the establishment of a hierarchy of FMRP partners.

References: [1] Maurin, T., et al., *Front Mol Biosci*, 2018. 5: p. 113.DOI: 10.3389/fmolb.2018.00113.[2] Richter, J.D., et al., *Nat Rev Neurosci*, 2021. 22(4): p. 209-222.DOI: 10.1038/s41583-021-00432-0. [3] Castagnola, S., et al., *Front Synaptic Neurosci*, 2017. 9: p. 15.DOI: 10.3389/fnsyn.2017.00015. [4] Delhaye, S., et al., *Mol Psychiatry*, 2021. 26(9): p. 4570-4582.DOI: 10.1038/s41380-020-00997-9. [5] Zeidler, S., et al., *Gene*, 2021. 768: p. 145298.DOI: 10.1016/j.gene.2020.145298. [6] Mangano, G.D., et al., *Eur J Med Genet*, 2022. 65(3): p. 104441.DOI: 10.1016/j.ejmg.2022.104441. [7] Alpatov, R., et al., *Cell*, 2014. 157(4): p. 869-81.DOI: S0092-8674(14)00415-2. [8] Zhang, W., et al., *Hum Mol Genet*, 2014. 23(19): p. 5188-96.DOI: hmg/ddu241 [9] Chakraborty, A., et al., *Cell Rep*, 2020. 32(12): p. 108179.DOI: 10.1016/j.celrep.2020.108179. [10] Agote-Aran, A., et al., *EMBO J*, 2020. 39(20): p. e104467.DOI:

10.15252/embj.2020104467. [11] D'Souza, M.N., et al., iScience, 2018. 9: p. 399-411.DOI: 10.1016/j.isci.2018.11.007. [12] Bhogal, B., et al., Nat Neurosci, 2011. 14(12): p. 1517-24.DOI: nn.2950. [13] Shamay-Ramot, A., et al., PLoS Genet, 2015. 11(12): p. e1005702.DOI: 10.1371/journal.pgen.1005702.[14] Filippini, A., et al., RNA Biol, 2017. 14(11): p. 1580-1591.DOI: 10.1080/15476286.2017.1338232.[15] Tran, S.S., et al., Nat Neurosci, 2019. 22(1): p. 25-36. DOI: 10.1038/s41593-018-0287-x.[16] Didiot, M.C., et al., Nucleic Acids Res, 2008. 36(15): p. 4902-12. DOI: gkn472. [17] Zhou, L.T., et al., Neuroscience, 2017. 349: p. 64-75. DOI: 10.1016/j.neuroscience.2017.02.044. [18] Kim, M., et al., Mol Cell Biol, 2009. 29(1): p. 214-28. DOI: MCB.01377-08. [19] Edens, B.M., et al., Cell Rep, 2019. 28(4): p. 845-854 e5.DOI: S2211-1247(19)30852-6 [20] Hsu, P.J., et al., J Biol Chem, 2019. 294(52): p. 19889-19895.DOI: AC119.010078.[21] Westmark, C.J., et al., Sci Rep, 2020. 10(1): p. 10781.DOI: 10.1038/s41598-020-66394-y.[22] Kim, G.W., et al., J Virol, 2021. 95(13): p. e0009721.DOI: 10.1128/JVI.00097-21.[23] Brooks, A.N., et al., Genome Res, 2015. 25(11): p. 1771-80.DOI: 10.1101/gr.192518.115.[24] Shah, S., et al., Cell Rep, 2020. 30(13): p. 4459-4472 e6.DOI: 10.1016/j.celrep.2020.02.076.[25] Hale, C.R., et al., Elife, 2021. 10.DOI: 10.7554/eLife.71892.[26] Mientjes, E.J., et al., Neurobiol Dis, 2006. 21(3): p. 549-55.DOI: S0969-9961(05)00243-3. [27] Dahlhaus, R., Front Mol Neurosci, 2018. 11: p. 41.DOI: 10.3389/fnmol.2018.00041. [28] Maurin, T., et al., Nucleic Acids Res, 2018. 46(12): p. 6344-6355.DOI: 4970504. [29] Maeder, C.I., et al., Cell, 2018. 174(6): p. 1436-1449 e20.DOI: 10.1016/j.cell.2018.07.046. [30] Labory, J., et al., Front Mol Biosci, 2020. 7: p. 590842. DOI: 10.3389/fmolb.2020.590842.[31] Pratella, D., et al., Int J Mol Sci, 2021. 22(19). DOI: 10.3390/ijms221910891.[32] Drozd, M., et al., Front Mol Neurosci, 2018. 11: p. 124. DOI: 10.3389/fnmol.2018.00124. [33] Kieffer, K., et al., 2022 Frontiers in Molecular Biosciences. <https://www.frontiersin.org/articles/10.3389/fmolb.2022.954087/full>. [34] Labory, Le Bideau et al. Bioinformatics 2022,DOI: 10.1093/bioinformatics/btac603

---

## ***Origin, evolutionary history and prediction of orphan genes in the genomes of crop pest nematodes***

Stage étudiant (Master bio-info UCA) dans le cadre du projet n°2 encadré par Silvia Bottini (MSI / INRAE).

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- Ercan Seçkin. Origine, histoire évolutive et prédition des gènes orphelins dans les génomes de nématodes de ravageurs de culture. Rapport de stage Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023

### **Description :**

Ensuring food security for a population that will exceed 9 billion people by 2050 while preserving the environment is a major challenge. In this context, a key issue is to reduce agricultural losses caused by plant pests and pathogens. In particular, parasitic nematodes are responsible for considerable agricultural losses worldwide. The progressive banning of certain nematicides that are harmful to the environment and human health, and the ability to overcome resistance in the field, make it essential to develop new strategies for controlling these pests. This requires further fundamental knowledge of the evolution of the genomes of these organisms in relation to adaptation to plant parasitism. Preliminary work in the team suggests a very large number of potential orphan genes in plant-parasitic nematodes. These orphan genes may be involved in the ability to manipulate or parasitise plants. For example, the majority of parasitism effector proteins in nematodes are encoded by orphan genes. The origin, evolutionary history and impact of orphan genes in the genomes and biology of these species remain poorly understood. Different mechanisms may contribute to the presence of orphan genes in the genomes of plant parasitic nematodes. Horizontal gene transfer may explain the absence of homologs in related species, while gene duplication followed by divergence may contribute to the impossibility of finding homologs in all of life. A third mechanism that is increasingly recognised as important but has not yet been studied in plant parasites is de novo gene emergence. This is the process by which a new gene is formed from non-genic DNA and not from modifications of an existing gene. Long considered an unlikely event, this mechanism of generating novelty and protein diversity from non-genes has been demonstrated in many organisms, from yeast (84 de

---

novo genes specific to *Saccharomyces cerevisiae* in the latest study) to humans (60 de novo genes emerged since separation from chimpanzees).

## Objectives

The first objective of the project will be to identify and determine the origin and evolutionary history of orphan genes in the nematodes of the genus *Meloidogyne*, the most damaging to agricultural production worldwide. This group of pests is a very good model to study this question because good quality genomes for about ten *Meloidogyne* species are available. In addition, the transcriptomic data also available will make it possible to verify the functional character of the genes identified. The main steps will be to (i) identify orphan genes in plant-parasitic nematodes using comparative genomics approaches, (ii) validate the expression of these genes using transcriptomic data, (iii) reconstruct the chronology of emergence of these genes in the phylogeny of these organisms using parsimony or likelihood approaches and (iv) determine a possible de novo origin for these orphan genes.

The second objective of the project will be to use AI methods to develop predictors of orphan genes based on the features identified in the previous steps. The main steps will be to (i) test different algorithms for feature selection, (ii) train machine learning models based on supervised methods to develop an orphan gene prediction tool, (iii) evaluate the performance of this tool on a test dataset and estimate its generalization to other genomes. This project will determine the contribution of orphan genes to the composition of the genomes of various crop nematode pests and the evolutionary mechanisms that led to their presence.

## Références :

- 1 Grynberg P, Coiti Togawa R, Dias de Freitas L, Antonino JD, Rancurel C et al. (2020) Comparative genomics reveals novel target genes towards specific control of plant-parasitic nematodes. *Genes* 11, 1347.
- 2 McGowan J, Byrne KP, Fitzpatrick DA (2018) Comparative analysis of oomycete genome evolution using the Oomycete Gene Order Browser (OGOB). *Genome Biol Evol* 11, 189-206.
- 3 Mitchum MG, Hussey RS, Baum TJ, Wang X, Elling AA et al. (2013) Nematode effector proteins: An emerging paradigm of parasitism. *New Phytol* 199, 879-94.

- 
- 4 Danchin EG, Rosso M-N, Vieira P, de Almeida-Engler J, Coutinho P et al. (2010) Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *PNAS* 107, 17651–17656
- 5 McGowan J, Fitzpatrick DA (2020) Recent advances in oomycete genomics. *Adv Genet* 105, 175-228
- 6 Abad P, Gouzy J, Aury J-M, Castagnone-Sereno P, Danchin EGJ et al. (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat Biotechnol* 26: 909-915
- 7 Blanc-Mathieu R, Perfus-Babecoch L, Aury J-M, Da Rocha M, Gouzy J et al. (2017) Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genet* 13, e1006777
- 8 Van Oss SB, Carvunis, A-R (2019) De novo gene birth. *PLoS Genet* 15, e1008160.
- 9 Wu B, Knudson A (2018) Tracing the de novo origin of protein-coding genes in yeast. *mBio* 9, e01024-18
- 10 Wu D-D, Irwin DM, Zhang Y-P (2011) De novo origin of human protein-coding genes. *PLoS Genet*. 7, e1002379.

---

## **Finding signatures associated to the mechanism of pathogen resistance bypass by network inference using multi-omics data integration**

Stage étudiant (Master bio-info UCA) dans le cadre du projet n°2 encadré par Silvia Bottini (MSI / INRAE).

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- Rapport de stage de Maxime Multari. Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023

### **Description :**

Plants possess resistance genes to a variety of pathogens, allowing them to avoid deleterious effects of infection. Some pathogens were able to overcome these barriers during the course of evolution. One known example is the Mi-1.2 gene from the tomato *Solanum lycopersicum* conferring resistance to the root-knot nematode *Meloidogyne incognita*. Previous studies in our lab have tested the durability of Mi-1.2 resistance against *M. incognita*. Tomato plants carrying the resistance gene Mi-1.2 were inoculated with a load of avirulent nematodes. Collecting the initial few offspring and reinoculating on tomato plants bearing Mi-1.2, allowed the establishment of a resistance-breaking subpopulation after 5 to 10 generations. This same experience was carried out with two initially avirulent populations, from Mexico and Russia, both yielding virulent subpopulations in the lab. We sequenced both populations from Mexico and Russia, and both avirulent and virulent phenotypes, to study whether genomic variations accompany resistance breaking down. Using long-read sequencing, we were able to identify not only single-nucleotide polymorphisms, but also structural variants (> 50 bp) exclusively present in the two virulent subpopulations (Mexico and Russia). We also studied the genome expression profiles of these populations, and whether the different genomic variations could influence the expression of candidate virulence / avirulence genes.

To study whether there is a causal relationship between the presence of a variant in a particular genomic region X and the differential expression of a gene Y independently by their proximity we will implement a causal network inference method. Biological relationship can be represented as networks where nodes represent molecules and connections between nodes represent molecule-molecule interactions, whereas omics data levels can be seen as feature vectors of molecules. Biological network inference and

---

causal learning are methods that can be used to investigate the direct and indirect multi-layer associations and possible causal relations between omics data features in the system. Their particularity is that they allow to put directionality between nodes in the network (i.e.  $X \rightarrow Y$ ). Here we will use causal network inference algorithms based on constraints in order to reconstruct a direct network of interactions between variants and gene. The study of the inferred network structure will allow to characterize the molecular signatures of *M. Incognita* -circumvention of resistance and to gain insight toward the understanding of the elude the mechanisms of resistance.

---

## ***Toward a multi-omics approach for disease-causing molecules identification using graph convolutional network.***

Stage étudiant (Master bio-info UCA) dans le cadre du projet n°2 encadré par Silvia Bottini (MSI / INRAE).

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- Rapport de stage de Anjana Bhat. MSc2 in Data Science and AI. Université Côte d'Azur 2023
- Rapport de stage de Madina Bekbergenova. Université Côte d'Azur 2023

### **Description :**

Context:

Modern data-driven biology relies on the interpretation of large-scale molecular measurements (known as omics data) to understand and predict biological phenotypes, such as the characteristics of an organism or the state of an individual cell. Next-generation sequencing (NGS) biotechnologies can now comprehensively measure different types of molecules (e.g., DNA, RNA, and proteins) at a high-speed and low-cost, leading to an unprecedented generation of omics data from different biological samples or across thousands of individual cells. The ‘omics’ notion indicates that nearly all instances of the targeted molecules are measured in the assay providing holistic views of the biological system [1]. However, individual analyses of single-omics layers only offer a one-dimensional view of a complex system. Interestingly, the computational biology and bioinformatics community, has recently developed several approaches and tools integrating results from multiple omics studies to associate phenotypes and extract recurrent features, in a so-called integromics analysis. Integrative multi-omics approaches allow to reduce the complexity of the phenotypes by highlighting the most salient features. Compared to single omics interrogations, multi-omics can provide researchers with a greater understanding of the flow of information, from the original cause of disease to the functional consequences or relevant interactions. Multi-omics approaches integrate data from different omics levels to understand their combined influence on the biological process. Depending on the samples available and the aim of the study, several approaches have been proposed to identify these relationships. Inter- and intra-omics relationships may be explored either by low-throughput pairings or by high-throughput assessment of larger networks [2,3].

---

## Proposal:

Biological networks can be treated as graphs, where nodes represent genes and connections between nodes represent gene–gene interactions, whereas omics data levels can be seen as feature vectors of genes. Graph deep learning has recently emerged to incorporate graph structures into a deep learning framework. In particular, graph convolutional networks (GCNs) are able to classify unlabelled nodes in a network on the basis of both their associated feature vectors, as well as the network's topology, making it possible to integrate graph-based data with feature vectors in a natural way. Advances in feature interpretation strategies for deep neural networks make it also possible to investigate the decision of such methods, leveraging deep understanding of the underlying data [4].

## Activities to do:

The main task of the internship would be to develop a GCN model to perform multi-omics integration, specifically:

- Train the model on several biological pathways
- Identify the best weights to represent nodes interrelationships.
- Extend the model to handle multi-cohorts and multi-organisms data.
- Test on experimental data

**Keywords:** bioinformatics, multi-OMICS, graph convolutional network, deep learning, disease-causing molecules, complex diseases

## Some references:

A. Conesa and S. Beck, "Making multi-omics data accessible to researchers," *Sci Data*, vol. 6, no. 1, pp. 1–4, 2019.

Inouye M, Silander K, Hamalainen E, Salomaa V, Harald K, Jousilahti P, et al. An immune response network associated with blood lipid levels. *PLoS genetics* 2010;6:e1001113.

Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Molecular systems biology* 2010;6:441.

Schulte-Sasse et al. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. *Nature Machine Intelligence*, 2022

---

## **VIOLA: a novel strategy for Variant prioritization using LAtent space in the context of mitochondrial diseases**

Stage étudiant (Master bio-info UCA) dans le cadre du projet n°3 encadré par Silvia Bottini (MSI / MD Lab).

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- Rapport de stage de Youssef Boulaimen. Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023

### **Description :**

#### **Context**

Mitochondrial diseases (MD) are rare disorders caused by deficiency of the mitochondrial respiratory chain, which provides energy in each cell through oxidative phosphorylation (Elliot et al. 2008). These diseases are extremely heterogeneous both clinically and genetically, with a broad range of age onset and very different symptoms, that makes their diagnosis very challenging (Munnich et al. 2001). Among them, predominate neurological disorders associated with muscle symptoms and metabolic alterations such as Leigh's syndrome, Alpers disease or Leber's Optic atrophy (Di Mauro 2004). The estimated incidence is 1/5000 births, about 200 new cases per year in France.

MD are caused by the changes or alterations (variants) on genes involved in mitochondrial functions (Saneto et al. 2017). Although mitochondria possess their own genome (Robin et al. 1988), they need nuclear genes to encode proteins for their biogenesis. Hence, MD may be caused by pathogenic variants affecting either mtDNA or nuclear genes.

Nowadays, there are more than 300 nuclear genes known to be involved in MD and the list of candidate genes is continuously growing (Craven et al. 2017). High throughput sequencing (NGS) has highly improved the diagnosis of MD by allowing the sequencing of the entire mtDNA and of panels of more than 300 nuclear genes in patients (Vasli et al. 2012, Plutino et al. 2018). However, in one patient out of two the responsible gene is unknown. The advent of whole-exome sequencing (WES) and whole-genome sequencing (WGS) has greatly accelerated the identification of variants on previously unknown disease genes (Wortmann et al. 2015). Although these technologies are mainstays in

---

Mendelian disease diagnosis, their success rate for detecting causal variants is far from complete, ranging from 25 to 50% (Taylor et al. 2015). Several variants detected by WES remain as variants of unknown significance (VUS) or they are missed due to the inability to prioritize them. Furthermore, variants present in non-coding regions that could affect gene expression, cannot be identified by WES. While this limitation is overcome by WGS which enumerates variants in both coding and non-coding regions, WGS presents important (and currently unresolved) challenge of data interpretation for the more than 3 million variants per sample (Lohmann & Klein 2014). Recently the employ of RNA sequencing (RNA-seq) has been proposed (Byron et al. 2016). This technology provides a direct probing of RNA abundance and sequence of both coding and non-coding genome, including allele-specific expression and splice isoforms (Wang et al. 2009). Despite the very promising premises of RNA-seq to detect new variants, the pioneering works employing this technique improved the diagnostic power of only 10%, leading to the increase of the number of VUS and not providing new methods to prioritize variants (Kremer et al. 2017). In this context, the development of novel integrative computational approaches are essential to resolve diagnostic deadlock and improve our knowledge of mitochondrial disorders (Labory et al. 2020).

Nowadays, multi-omics studies have already been performed but the algorithms for data analysis have been tested mainly on other kind of diseases such as cancer. The differences are crucial: while in cancer research we dispose of several samples, very large cohorts are not available for MD. The other important point is the heterogeneity of MD: while in cancer research we look for common signatures among patients compared to healthy individuals, in MD each patient can be considered as a unique representative of his class because the responsible gene and its carried variant are usually specific of each patient. Finally and most importantly, multi-omics algorithms and tools already existing are developed for patient classification, instead we are interested to identify the candidate genes for the pathology. Auto-encoders (AE) and variational auto-encoders (VAE) are very promising technologies to integrate and analyze data from different sources (e.g., multi-omics, patient registries, ...) that can be used to overcome further challenges, such as low diagnostic rates, reduced number of patients and geographical dispersion.

#### Objectives:

1. To set up a standardized protocol to analyse sequencing data (WES and RNA-seq) of patients suspected of MD.

- 
2. Development of a novel computational strategy to stratify variants identified by sequencing data analysis based on VAE models
  3. Refine the most optimal parameters for the VAE configuration
  4. Apply on different dataset to choose the most performant and general architecture
  5. Bioinformatically validate the found candidate disease-genes (functional analysis, pathways analysis, etc..)
  6. Functional study of prioritized variants to identify candidate responsible gene for MD.
  7. Validation of novel algorithm on publicly available datasets and on in-house patient cohort.

### Expected results

The diagnosis of mitochondrial diseases is based on the identification of the disease responsible gene(s). This will allow to be able to offer genetic counseling, prenatal diagnosis, to consider therapeutic approaches or to improve the care of patients. With this project, we expect to enrich the currently deployed pipelines for mitochondrial disease diagnosis by expanding the range of mutations that can be identified with RNA-seq, in particular through a thorough exploration of non-coding transcripts and splice events that can affect gene structure and functionality. We dispose of an initial cohort of 20 patients with clinical evidences of MD in diagnostic stalemate. RNA-seq has been performed after WES did not allow to identify the responsible gene. Control samples from GTEx (Genotype Tissue Expression) database will be selected to identify tissue and disease specific transcripts (The GTEx consortium 2015). This cohort will be used to set up the bioinformatic and modeling framework, joint to a correspondent control cohort of healthy patients with similar physical parameters selected from GTEx database. We will use a machine learning approach to implement a novel method to stratify variants in order to identify potential candidate genes to test experimentally. The challenge will be to identify the optimal machine learning strategy to treat a large number of variables and few observations (patients), because usually these kind of algorithm perform better with the opposite data configuration. Since gene expression is tissue dependent, sequencing data from healthy individuals from multiple tissues (fibroblast, muscle, blood) will be included in the analysis in order to disentangle the tissue heterogeneity. The functional analysis of identified variants will allow to prioritize candidate genes for MD. To functional characterize variants, we need to integrate data from different sources: from sequencing data to clinical information and biological pathways. We will use co-expression network in order to identify functional co-expressed gene modules associated with mitochondrial diseases. This approach will allow the integration of

---

functional relationships from variants and gene expression to gene pathways, allowing to relate genetic variants measured in WES to metabolic pathways quantified by RNA-seq data, extracted for example from GO, RegNetwork and KEGG ontologies. We will take advantage of data from public databases to benchmark and validate our approaches, in particular using a training subset from the Geuvadis consortium (Lappalainen et al, Nature 2013), providing RNA-seq data for 460 samples for which WES are available from the 1000 Genome project. Finally, novel identified variants on the patient cohort will be further investigated by performing experiments to determine the pathogenicity of the variant and the function of the carrying gene in MD.

We foresee that this project will contribute to set up new diagnostic tools in the context of personalized medicine to reduce the number of patients with a diagnostic impasse. This study will settle the milestones to transfer the conjoint use of WES and RNA-seq from research fields to diagnostic environment. Finally, the development of a novel strategy to prioritize variants will address one of the biggest challenges in the domain of Mendelian diseases and will allow the development of personalized medicine to ameliorate patients healthcare.

---

## **Développement d'une approche de médecine personnalisée pour améliorer le diagnostic des maladies mitochondrielles**

Sujet de thèse de Justine Labory, dans le cadre du projet n°3 encadré par Silvia Bottini (MSI / MD Lab).

### **Contexte**

Les maladies mitochondrielles (MM) sont des pathologies rares provoquées par une déficience de la chaîne respiratoire (CR) mitochondriale, qui fournit l'énergie aux cellules par phosphorylation oxydative (1). Ce sont les plus fréquentes des maladies métaboliques avec une incidence estimée à 1/5000 naissances. Les MM sont extrêmement hétérogènes, tant sur le plan clinique que génétique, avec un large éventail d'âges et des symptômes très différents, ce qui rend leur diagnostic très difficile (2).

Les MM sont dues à des variants dans des gènes impliqués dans la biogénèse mitochondriale (3). Bien que les mitochondries possèdent leur propre génome (4), plus de 1000 gènes nucléaires sont nécessaires au fonctionnement de la CR (5). Par conséquent, les MM peuvent être causées par des variants pathogènes affectant soit l'ADNmt, soit les gènes nucléaires. Leur diagnostic nécessite l'identification du gène responsable.

Aujourd'hui, plus de 300 gènes nucléaires sont impliqués dans les MM et la liste ne cesse de s'allonger (6). Malgré le développement du séquençage à haut débit (NGS), plus d'un patient sur deux est en impasse diagnostique, le gène responsable restant inconnu. L'arrivée du séquençage de l'exome (WES) et du séquençage du génome (WGS) a considérablement accéléré l'identification de variants sur des gènes de maladies jusqu'alors inconnus (7). Néanmoins, leur taux de réussite pour la détection des variants responsables se situent seulement entre 25 et 50 %. Plusieurs variants détectés par le WES restent des variants de signification inconnue (VUS) ou sont omis en raison de l'impossibilité de les classer par ordre de priorité. En outre, les variants présents dans des régions non codantes ne peuvent pas être identifiés par WES. Le WGS détecte les variants dans les régions codantes et non codantes, mais il présente un défi important (et actuellement non résolu) d'interprétation des données pour les plus de 3 millions de variants par échantillon (8). Des études récentes ont montré l'intérêt du séquençage de l'ARN (RNAseq), pour sélectionner des gènes candidats présentant une expression altérée ou monoallélique et pour identifier des variants introniques profonds affectant l'épissage, lorsqu'il est utilisé en combinaison avec le WES (9). Pour les MM, l'utilisation du RNAseq chez 48 patients non diagnostiqués après un WES a permis d'augmenter le taux de réussite de 10% (10). Dans ce contexte, le développement de nouvelles approches bio-informatiques est essentiel pour résoudre l'impasse diagnostique et améliorer notre connaissance des maladies mitochondrielles (11).

### **Objectifs :**

1. Mise en place d'un protocole standardisé pour analyser les données de séquençage (WES et RNAseq) des patients avec suspicion de MM
2. Développement d'une nouvelle stratégie bio-informatique pour stratifier les variants identifiés par l'analyse des données de séquençage
3. Étude fonctionnelle des variants prioritaires pour identifier le gène responsable candidat des MM
4. Validation de l'outil nouvellement développé sur des données publiques et sur une cohorte de patients réalisée en interne.

## Résultats attendus

L'identification des gènes responsables de MM permettra de proposer un conseil génétique, un diagnostic prénatal et d'envisager de nouvelles approches thérapeutiques. Avec ce projet, nous espérons développer de nouveaux outils en élargissant la gamme des variants qui peuvent être identifiés par RNAseq, en particulier par une analyse approfondie des transcrits non codants et des événements d'épissage qui peuvent affecter la structure et la fonctionnalité des gènes. Grâce au Centre de Référence sur les Maladies Mitochondriales, nous disposons d'une cohorte de 20 patients en impasse diagnostique (80 patients supplémentaires sont en cours d'inclusion). Le RNAseq a été réalisé après l'échec du WES pour identifier le gène responsable. Des échantillons provenant d'individus sains de la base de données GTEx (Genotype Tissue Expression) seront sélectionnés pour identifier les transcrits spécifiques des tissus sains et ceux spécifiques de la pathologie. Nous utiliserons une approche de machine learning pour développer une nouvelle méthode de stratification des variants afin d'identifier des gènes candidats potentiels à tester ensuite expérimentalement. Le défi consistera à identifier la meilleure stratégie pour traiter un grand nombre de variables et peu d'observations (patients), car généralement ce type d'algorithme fonctionne mieux avec la configuration de données inverse (peu de variables et beaucoup d'observations). Comme l'expression des gènes dépend des tissus, les données de séquençage d'individus sains provenant de multiples tissus (fibroblastes, muscles, sang) seront incluses dans l'analyse afin de distinguer l'hétérogénéité des tissus. L'analyse fonctionnelle des variants identifiés permettra de hiérarchiser les gènes candidats. Pour caractériser les variants de manière fonctionnelle, nous devons intégrer des données provenant de différentes sources : des données de séquençage aux informations cliniques en passant par les voies métaboliques. Nous utiliserons les réseaux de co-expression afin d'identifier les modules de gènes fonctionnels co-exprimés associés aux MM. Cette approche permettra de relier les variants génétiques trouvés en WES aux voies métaboliques identifiées par les données RNAseq, extraites par exemple des bases de données de la Gene Ontology, RegNetwork et KEGG. Nous exploiterons les données des bases de données publiques pour étalonner et valider nos approches, en particulier pour entraîner nos algorithmes. Nous utiliserons notamment des données du consortium Geuvadis (12), fournissant des données RNAseq pour 460 échantillons pour lesquels le WES est disponible dans le cadre du projet « 1000 Genome ». Enfin, la validation fonctionnelle des variants identifiés se fera sur des modèles cellulaires et murins en lien avec l'équipe « Génétique des MM » à l'IRCAN.

Nous espérons que ce projet contribuera à la mise en place de nouveaux outils de diagnostic dans le cadre de la médecine personnalisée afin de réduire le nombre de patients en impasse diagnostique. Cette étude sera une étape importante dans le transfert de l'utilisation conjointe du WES et du RNaseq dans le domaine de la recherche au diagnostic. Enfin, le développement d'une nouvelle stratégie pour prioriser les variants permettra de relever l'un des plus grands défis dans le domaine des maladies mendéliennes pour améliorer la prise en charge des patients.

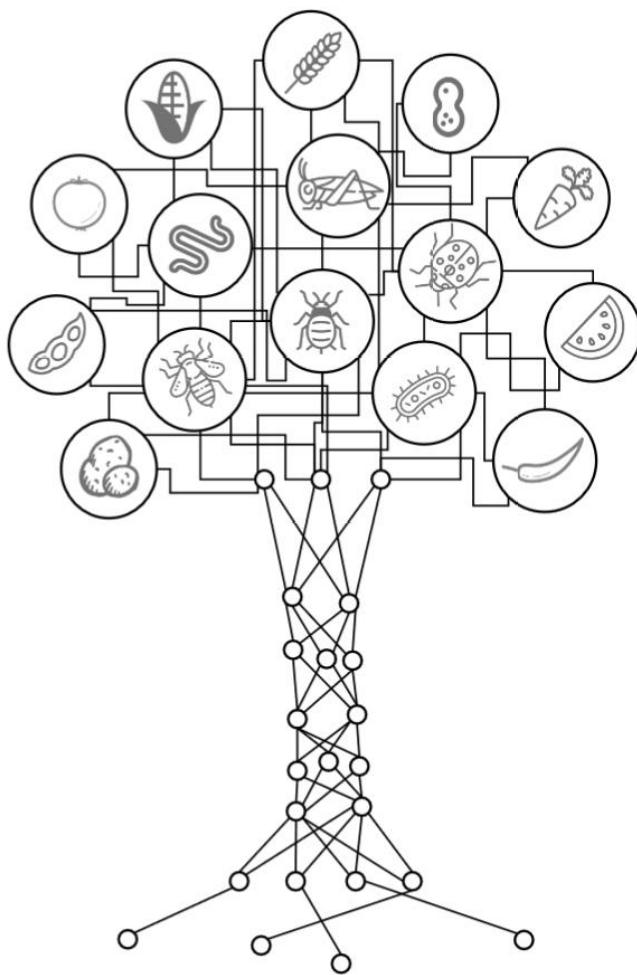
## Valeurs ajoutées

Ce projet transdisciplinaire sera développé avec le "Centre de Référence sur les Maladies Mitochondriales" dont le Professeur Paquis, le superviseur de ce projet de doctorat, est le coordinateur, en collaboration avec le « Medical Data Laboratory (MDLab) ». Le MDLab, dont le Dr. Bottini, co-superviseur de ce projet, est le directeur, est un centre d'excellence en « data sciences » appartenant à la Maison de la Modélisation, de la Simulation et des Interactions (Université Côte d'Azur, UCA). Le MDLab a été créée pour renforcer le dialogue entre les chercheurs en « data sciences » et ceux du domaine médical. Enfin l'équipe « Génétique des Maladies Mitochondriales » à l'IRCAN assurera la validation fonctionnelle *in vivo* grâce à ses compétences sur l'étude des fonctions mitochondrielles et aux modèles cellulaires et murins disponibles.

## Références

1. Elliott H, Samuels D, Eden J, Relton C, Chinnery P. Pathogenic Mitochondrial DNA Mutations Are Common in the General Population. 2008;
2. Munnich A, Rustin P. Clinical spectrum and diagnosis of mitochondrial disorders. 2001;
3. Saneto R. Genetics of Mitochondrial Disease. In 2017.
4. Robin DE, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. 1988;
5. Rath S, Sharma R, Gupta R, Ast T, Chan C, Durham TJ, et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Research. 8 janv 2021;49(D1):D1541-7.
6. Craven L, Alston C, Taylor R, Turnbull D. Recent Advances in Mitochondrial Disease. 2017;
7. Wortmann S, Koolen D, Smeitink J, van den Heuvel L, Rodenburg R. Whole exome sequencing of suspected mitochondrial patients in clinical practice. 2015;
8. Lohmann K, Klein C. Next Generation Sequencing and the Future of Genetic Diagnosis. Neurotherapeutics. 1 oct 2014;11(4):699-707.
9. Rahman J, Rahman S. Mitochondrial Medicine in the Omics Era. 2018;
10. Kremer L, Bader D, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. 2017;
11. Labory J, Fierville M, Ait-El-Mkadem S, Bannwarth S, Paquis-Flucklinger V, Bottini S. Multi-omics approaches to improve mitochondrial disease diagnosis: challenges, advances and perspectives. Frontiers. 2020;
12. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 1 sept 2013;501(7468):506-11.

## IA et multi-omiques en santé des plantes



### Description :

Produire des ressources agricoles suffisantes pour alimenter une population mondiale estimée à plus de 9 milliards d'individus d'ici 2050, tout en réduisant l'impact sur notre planète est l'un des principaux défis auxquels l'humanité est confrontée.

Les plantes sont engagées dans divers types d'interactions avec les organismes présents dans leur environnement. Ces interactions peuvent être soit bénéfiques pour leur développement (par exemple, la symbiose avec les champignons et les bactéries), soit préjudiciables pour leur santé et leur survie (par exemple, les ravageurs, les agents pathogènes et les parasites).

pathogènes et les parasites). Les ravageurs agricoles sont eux-mêmes impliqués dans des interactions avec des prédateurs, des pathogènes et des parasites qui peuvent être utilisés comme agents de bio-contrôle (par exemple, les guêpes parasitoïdes).

L'approfondissement de nos connaissances sur le fonctionnement et le succès de ces différentes interactions biotiques est ainsi d'une importance capitale pour améliorer la santé des plantes et éviter une crise agricole à venir, tout en limitant notre impact environnemental.

Dans ce contexte Université Côte D'Azur (UCA) via la MSI et l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) ont développé un projet de recherche commun visant à étudier des interactions entre organismes en lien avec la santé des plantes grâce à l'intégration de données multi-omiques et à l'utilisation d'algorithme d'Intelligence Artificielle (IA).

Les principales thématiques abordées concernent l'étude :

- Des interactions entre les plantes et divers bio-agresseurs (nématodes, oomycètes, insectes).
- Des relations symbiotiques bénéfiques à la plante telles que celles entretenues avec différentes bactéries et champignons.

- 
- Des interactions entre insectes parasitoïdes et bio-agresseurs des plantes dans le cadre de leur bio-contrôle.

Pour y parvenir, divers pipelines d'analyse de données biologiques devront être mis en place, ce qui nécessitera l'accès à des ressources informatiques conséquentes telles que celles proposées par le cluster Azzurra. La réalisation de certaines des études précédemment citées nécessitera en effet d'avoir accès :

- i) à une capacité de stockage importante compte tenu de la quantité de données à traiter
- ii) à une grande puissance de calcul via l'utilisation de GPU afin d'entrainer et utiliser des modèles d'intelligence artificielle.

## Projet DeepWILD



© Parc national du Mercantour

Projet MSI et 3IA Côte d'Azur en partenariat avec le Parc national du Mercantour.

### Résultats obtenus grâce aux moyens informatiques attribués :

Fanny Simões, Charles Bouveyron, Frédéric Precioso, DeepWILD: Wildlife Identification, Localisation and estimation on camera trap videos using Deep learning, Ecological Informatics, Volume 75, 2023, 102095, ISSN 1574-9541, <https://doi.org/10.1016/j.ecoinf.2023.102095>

### Description :

L'objectif du projet est de localiser, classifier et compter les espèces du Parc national du Mercantour à l'aide d'un algorithme de deep learning. Les images utilisées sont issues de vidéos extraites de pièges photographiques installés dans la Vallée de la Roya. L'ensemble des vidéos et des images utilisées pour ce projet sont stockées sur Azzurra. Azzurra a permis le traitement de ce gros volume de données et l'application d'un modèle de détection d'objet (Faster R-CNN avec Inception-ResNet-v2) capable de reconnaître les différentes espèces présentent dans le Parc national du Mercantour.

---

L'équipe travaillant sur ce projet est composée de Charles Bouveyron, Frédéric Precioso et Fanny Simões pour UCA, et Nathalie Siefert et Stéphane Combeaud pour le Parc national du Mercantour.

---

## Projet Tara



© Megan Clampitt

### Résultats obtenus grâce aux moyens informatiques attribués :

Thèse de Megan Clampitt : Indexation de l'état de santé des coraux par une approche basée sur l'intelligence artificielle. Soutenue le 04-09-2023 à l'Université Côte d'Azur , dans le cadre de l'École doctorale Sciences de la vie et de la santé.

### Description :

Projet de l'IRCAN dont le but est d'évaluer si une approche IA permet d'estimer la santé des colonies corallines à l'aide d'un dispositif automatisé et facilement applicable.

On utilise en premier lieu des images sous-marines collectées au cours de l'expédition Tara Pacific. Cette expédition a produit ~14000 photos (correspondant à 2216 colonies individuelles) qui étaient annotées manuellement. Ces annotations comprennent plusieurs facteurs : l'identification taxonomique jusqu'au rang du genre, la prédation, le contact avec d'éventuelles algues, la présence d'organismes perforants, le contact avec les sédiments, ainsi que d'autres facteurs directement liés à la santé des coraux (i.e. coloration, blanchissement, maladies, pigmentation). Grâce aux études de l'expédition Tara, plusieurs

---

autres bases de données sont pareillement disponibles, elles contiennent notamment des facteurs biologiques (biomarqueurs et télomères) et environnementaux.

Nous avons débuté une approche IA pour extraire les catégories d'annotations qui avaient été faites manuellement. Plus spécifiquement, nous utilisons la méthode dite du « Convolutional neural network (CNN) ». La première étape de cette approche est d'identifier le genre auquel appartient les colonies photographiées. Avec notre réseau qui comprenait les photographies et les annotations manuelles (qui ne visaient pas à identifier le genre), nous avons réussi à identifier les genres avec une exactitude de 66.28%.

L'étape en cours est d'augmenter l'exactitude de cette prédiction avec des approches IA comme le « transfer learning » ou les modèles génératifs. De plus, nous prévoyons de compléter ce réseau avec davantage de variables qui étaient annotées manuellement (blanchissement, présence de maladies, prédation, etc.) et nous espérons aussi pouvoir utiliser l'algorithme pour prédire les facteurs biologiques (télomères et biomarqueurs) qui sont considérablement plus difficile à obtenir. Mais il nous faudra plus d'échantillons pour construire l'algorithme.

---

**Chimie**

**EUR Spectrum**

---

## **Deep Reinforcement Learning for Mass Spectrometry Data Acquisition in Metabolomics**

**Responsable scientifique :** Louis-Félix Nothias

**Laboratoire :** ICN

### **Description :**

Mass spectrometry-based untargeted metabolomics is crucial for identifying small molecules in complex biological samples. However, current data acquisition methods have limitations in capturing all relevant molecules. To address this issue, we propose using artificial intelligence (AI) to optimize mass spectrometry data acquisition in real-time, maximizing the number and quality of identified metabolites. First, we will utilize large amounts of publicly available mass spectrometry data to develop a deep neural network that can predict the quality of generated fragmentation spectra based on instrument configurations. Second, we will use offline reinforcement learning to explore novel instrument configurations to enhance the data acquisition process. A critical focus will be placed on defining a suitable reward function that guides the AI agent's exploration, considering factors such as spectrum quality, novelty of acquired spectra, and resource utilization. Third, we will use a virtual mass spectrometry environment to simulate the fragmentation process and allow the AI agent to control data acquisition. This will enable thorough assessment and comparison against baseline approaches and alternative strategies.

Once fully trained and validated, the AI agent will be deployed onto a mass spectrometer to autonomously control the data acquisition process in real time, evaluating its performance in detecting putative metabolites compared to traditional approaches. By utilizing AI to optimize molecular discovery from untargeted metabolomics experiments, we will enhance the identification of metabolites that were previously overlooked, unlocking valuable biological insights. These advances will have transformative implications for precision medicine, drug discovery, environmental adaptation research, and other areas of the life sciences, ultimately impacting human health and climate change remediation.

---

## ***Simulations de complexes biomoléculaires impliqués dans les sens chimiques (olfaction, gustation).***

**Responsable scientifique du projet :** Sébastien Fiorucci

**Laboratoire :** ICN

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Publications dans des revues à comités de lecture :

ChemoDB: A Database of Olfactory Receptor-Odorant Pairs for Understanding the Molecular Mechanisms of Olfaction. M. Lalíš, M. Hladiš, J. Pacalon, S. Fiorucci, L. Briand, J. Topin. 2023, in prep.

TAARs response to amine are largely affected by sequence variants. J. Pacalon, C. Belloir, S. Fiorucci, L. Briand, J. Topin. 2023, submitted

Laure Martinelli, Camille Bihanic, Aurélie Bony, Florence Gros, Corentin Conart, Sébastien Fiorucci, Hervé Casabianca, Frédéric Schiets, Giorgiana Chietera, Benoît Boachon, Bernard Blerot, Sylvie Baudino, Frédéric Jullien, Denis Saint-Marcoux, Citronellol biosynthesis in pelargonium is a multistep pathway involving progesterone 5 $\beta$ -reductase and/or iridoid synthase-like enzymes, Plant Physiology, 2023; <https://doi.org/10.1093/plphys/kiad550>

Matching receptor to odorant with protein language and graph neural networks. M. Hladiš, M. Lalíš, S. Fiorucci, J. Topin. International Conference on Learning Representation (ICLR), 2023, accepted.

Ligand Binding Properties of Odorant-Binding Protein OBP5 from *Mus musculus*. L. Moitrier, C. Belloir, M. Lalíš, Y. Hou, J. Topin, L. Briand. Biology, 2022, 12(1), 2.

Extracellular loop 2 of G protein-coupled olfactory receptors is critical for odorant recognition. Y. Yu, Z. Ma, J. Pacalon, L. Xu, W. Li, C. Belloir, J. Topin, L. Briand, J. Golebiowski, X. Cong. Journal of Biological Chemistry, 2022, 298(9).

Exploring Dihydroflavonol-4-reductase reactivity and selectivity by QM/MM-MD simulations. J. Diharce, E. Bignon, S. Fiorucci, S. Antonczak, ChemBioChem, 2022, 23(3), e202100553.

Conférences sur invitations :

Une IA peut-elle apprendre à sentir ? S. Fiorucci. IADATES - "Intelligence Artificielle, Arômes et Parfum", Grasse, 13 février 2023.

---

Ligand-based virtual screening of membrane proteins. S. Fiorucci Winterschool Biomembranes Pharmacology and Therapeutics, Nice, 15 décembre 2022.

Communications orales :

Numerical models decipher odorant binding and receptor activation at the molecular level. M. Hladiš, J. Pacalon, M. Lalis, S. Fiorucci, J. Topin. Complex Day, Nice (France), February 2023.

Olfactory stimulation for the treatment of apathy in neurodegenerative disorders. J. Topin, R. David. Annual Meeting of the NeuroMod Institute, Antibes (France), June 2022.

Séminaires :

From genes to perception: decoding chemical senses by numerical and molecular modeling approaches. S. Fiorucci, Chimie de la matière complexe (UMR7140 CNRS) Strasbourg, 9 mars 2023.

Posters :

Recognizing emotions triggered by fragrances using virtual reality and physiological measurement. X. Meleqi, U Meierhenrich, A. Pegard, J. Topin. ECRO 2022, Berlin (Allemagne), 31 août-3 septembre 2022.

Deciphering the ligand binding properties of the mouse odorant-binding protein OBP5 from *Mus musculus*. L. Moitrier, C. Belloir, M. Lalis, Y. Hou, J. Topin, L. Briand. ECRO 2022, Berlin (Allemagne), 31 août-3 septembre 2022.

Communications grand public :

Café art/science regards scientifiques : la transmission de savoir-faire olfactifs. Le Hublot Nice, Décembre 2022.

Forum scientifique - Café rencontre : L'IA au service de la création parfum, mythe ou réalité ? Musée International de la Parfumerie Grasse, Octobre 2022.

## Description du projet :

Nous cherchons à déchiffrer les bases moléculaires de la perception chimiosensorielle (i.e. les sens chimiques : olfaction et gustation) qui est par nature un processus neuronal complexe. Pour tenter de déchiffrer le code combinatoire de la perception chimiosensorielle nous mettons au point des modèles numériques basés sur des approches de modélisation moléculaire, de machine learning, ou encore de

bioinformatique structurale. Pour décrire les mécanismes moléculaires sous-jacents, nous tirons profit de l'expertise de nos collaborateurs, notamment en biologie moléculaire et neurobiologie (Figure 1).

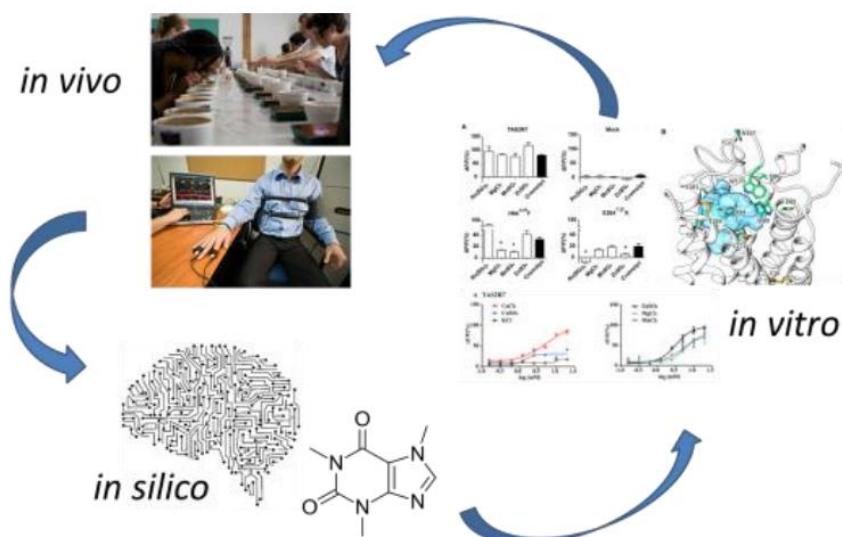


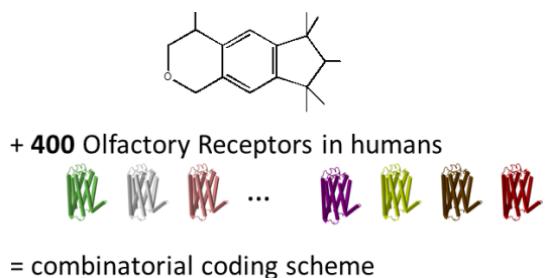
Figure 1 Synergie de l'approche pluridisciplinaire

## 1. Décrypter le codage de l'information chimique par notre système sensoriel.

Notre cerveau est fait de milliards de neurones et est considéré comme l'organe le plus complexe de notre corps. Parmi nos cinq sens, l'odorat et le goût sont les deux sens chimiques permettant aux êtres vivants d'identifier et de réagir aux substances présentes dans leur environnement. L'extraordinaire capacité de discrimination de notre cerveau provient de l'organisation de notre système de détection. Nous possédons environ 400 récepteurs olfactifs et 30 récepteurs gustatifs. Sachant qu'une molécule (odorante ou sapide) peut activer plusieurs récepteurs sensoriels, qu'un récepteur peut répondre à plusieurs molécules différentes et que la perception d'une odeur ou d'une saveur peut provenir d'une seule molécule ou d'un mélange de molécules, la combinatoire est virtuellement infinie<sup>1</sup> (Figure 2). L'homme peut par exemple distinguer plus de mille milliards de molécules odorantes. L'espace chimique des molécules pouvant activer notre odorat ou notre goût est immense et l'un des objectifs de nos travaux est d'identifier les

<sup>1</sup> Topin et al. Submitted. <https://doi.org/10.1101/2020.10.23.348706>

Figure 2 Codage combinatoire de la perception des odeurs



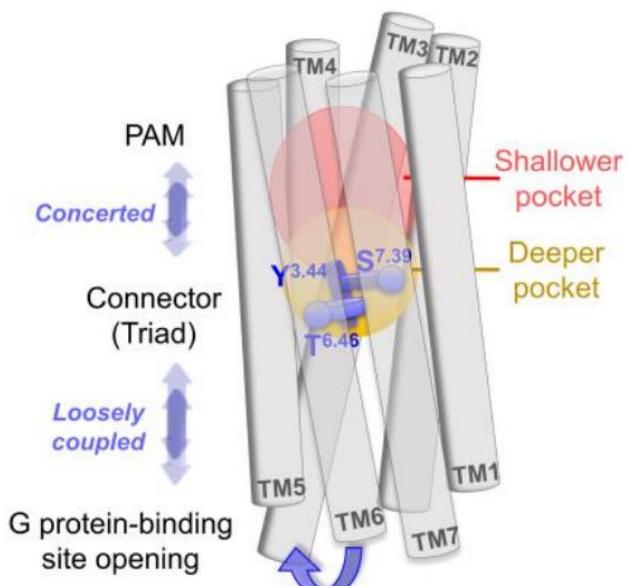
liens chémo-génomiques de ces espaces de grande dimensionnalité et de mettre au point des modèles moléculaires et numériques capables de prédire de façon rationnelle des composés olfactifs et sapides nouveaux.

Faits marquants : Les travaux de l'équipe ont notamment permis de mieux comprendre les relations entre la structure d'un composé et son activité biologique<sup>2</sup> et permis d'élargir le spectre de reconnaissance de récepteurs olfactifs humains<sup>3</sup>, de poissons<sup>4</sup> ou encore d'insectes<sup>5</sup> mais également de récepteurs gustatifs humains<sup>6,7</sup>.

## 2. Développer de nouveaux composés pour moduler l'activation des récepteurs chimiosensoriels.

La détection des odeurs et des saveurs sucré, amère et umami nécessitent l'activation des récepteurs de la famille des Récepteurs Couplés aux Protéines G (RCPG) dont le mécanisme d'activation est similaire à celui de la Rhodopsine, récepteur prototypique des RCPG de classe A. Ils agissent comme des « pinces » moléculaires qui peuvent « s'ouvrir » lorsqu'un agoniste (odorant ou molécule sapide) s'associe avec eux, ou « se refermer » s'il s'agit d'un agoniste inverse (Figure 3). La nature versatile de ces récepteurs (subtil équilibre entre différents états conformationnels dont la forme active même en l'absence de ligand, i.e. état basal)

Figure 3 Mécanisme de modulation allostérique d'un RCPG (ici de classe C)



<sup>2</sup> Bushdid et al. Science, 2014, 343, 1370-1372

<sup>3</sup> Licon et al. Plos Comput. Biol. 2019, 15(4), e1006945

<sup>4</sup> Cong et al. J. Biol. Chem., 2019, 294, 6762-6771

<sup>5</sup> Caballero-Vidal et al. Sci. Rep. 2020, 10, 1655

<sup>6</sup> Chéron et al. Food. Chem. 2017, 221, 1421-1425

<sup>7</sup> Bouyssot et al. Food. Chem. 2020, 324, 126864

rend leur pharmacologie complexe. En l'absence de structure cristallographique de récepteur chimiosensoriel, le développement rationnel de nouveaux composés olfactifs et gustatifs reste un défi. L'un des objectifs de nos travaux est de tirer parti des outils de simulation moléculaire (validés par des études de mutagénèse dirigée) pour étudier la dynamique des interactions ligands-récepteurs et développer de nouveaux modulateurs du goût et de l'odorat. Faits marquants : Les travaux de l'équipe ont notamment montré qu'il était possible de prédire le mécanisme d'activation et le spectre de reconnaissance de récepteurs chimiosensoriels à l'aide de modèles moléculaires et numériques.

On peut mettre en avant les résultats sur les récepteurs olfactifs de mammifères<sup>8 9 10 11</sup> mais également sur la famille des récepteurs gustatifs<sup>12 13</sup> et plus largement sur les RCPG<sup>14 15 16</sup>, famille de récepteurs impliqués dans la signalisation cellulaire et naturellement une des premières cibles pharmacologiques pour de nombreuses pathologies.

### 3. Collaborations et importance pour le tissu socio-économique

Le projet s'appuie sur une approche pluridisciplinaire chimie-informatique-neurobiologie. Il permettra d'envisager la mise au point d'un « nez ou d'une langue computationnelle » physiologiquement inspiré(e) capable de prédire l'odeur, la saveur et/ou la réponse émotionnelle d'un composé sur la base de sa structure moléculaire. Pour l'industrie des arômes et parfums, la recherche de nouvelles molécules est un enjeu majeur. Par analogie avec d'autres molécules connues, il serait possible de prédire les structures de nouvelles molécules ayant des effets de modulation de l'odorat ou du goût (inhibition notamment). De plus, les récepteurs chimiosensoriels sont également exprimés dans d'autres organes que ceux de nos sens du goût et de l'odorat (expression ectopique) et leur rôle sur notre santé est encore mal connu. On peut donc envisager des applications de notre projet dans de grands secteurs économiques aussi divers que la santé, l'industrie des arômes et parfums ou l'industrie agroalimentaire, particulièrement bien représentés dans la région PACA. Notre travail de recherche mobilise de nombreux partenaires au niveau local

<sup>8</sup> Yu et al. PNAS, 2015, 112(48), 14966-1497

<sup>9</sup> de March et al. JACS, 2015, 137(26), 8611-8616

<sup>10</sup> de March et al. Angew. Chem. Int. Ed. 2018, 57, 4554-4558

<sup>11</sup> Bushdid et al. CMLS, 2019, 76(5), 995-1004

<sup>12</sup> Chéron et al. Chem. Senses, 2019, 44, 303-310

<sup>13</sup> Wang et al. Chem. Senses, 2019, 44, 339-347

<sup>14</sup> Cong et al. J. Chem. Theory Comput., 2018, 14(8), 4467-4473

<sup>15</sup> Cong et al. J. Chem. Info. Model. 2019 59(6), 2871-2878

<sup>16</sup> Cong et al. PCCP, 2018, 20, 24915-24920

---

avec des experts du domaine de la santé (CHU Nice, IPMC) et de l'intelligence artificielle (thèse 3IA qui démarrera en octobre 2020 en collaboration avec l'I3S), au niveau national avec des experts des sens chimiques chez l'homme (CRNL-Lyon, CSGA-Dijon) ou chez l'insecte (INRAEVersailles) ainsi qu'au niveau international avec des experts en neurosciences à Duke (USA), au Monell (USA) ou encore au DGIST (Corée). A noter également que le projet entre dans les thématiques prioritaires de l'Université Côte d'Azur liées à la science des odorants à travers l'Institut d'Innovation et de Partenariat en Arômes, Parfums & Cosmétiques et aux neurosciences computationnelles et aux sciences de la cognition notamment soutenues à travers l'institut Neuromod ou encore la structure CoCoLab.

#### 4. Méthodologies employées

D'un point de vue computationnel, nos travaux s'appuient sur des outils de modélisation moléculaire et d'apprentissage statistique (machine / deep learning). La modélisation d'un récepteur chimiosensoriel dans un environnement physiologiquement inspiré (prise en compte explicite de la membrane, du solvant ou encore des ions) nécessite des simulations moléculaires d'environ 60 000 à 100 000 atomes et pouvant aller jusqu'à 400 000 atomes pour des systèmes multimériques (Figure 4). Nous étudions les changements de conformations qui ont lieu lors des processus de liaison d'un odorant ou d'une molécule sapide avec un récepteur. Cela nécessite des simulations de dynamique moléculaire de plusieurs dizaines de microsecondes. Pour cela nous utilisons les logiciels Amber et Gromacs, par ailleurs déjà installées sur l'ancien supercalculateur de l'Université (CICADA), que ce soit pour des calculs sur CPU ou GPU. Les analyses des trajectoires de simulations se font souvent sur les machines de l'équipe et il peut arriver que ponctuellement nous ayons besoin de réaliser une analyse (calculs de mode normaux par ex.) qui nécessite une ressources mémoire plus importante. Dans ce cas, Azzurra nous permettra de réaliser ces calculs. Par ailleurs, les structures de ces récepteurs et les interactions ligand-récepteurs sont déduites de contraintes expérimentales et théoriques par des approches de reconstruction par homologie et de docking.

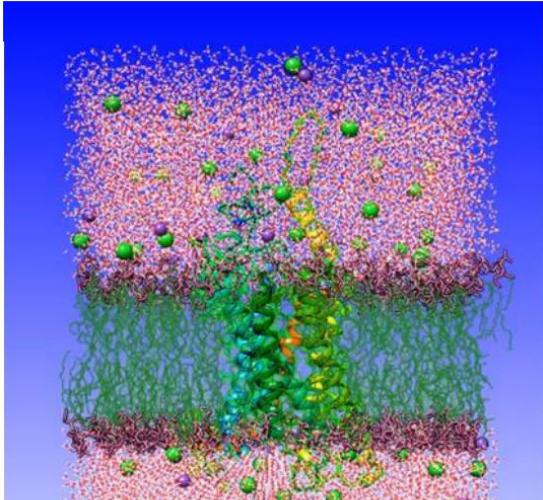
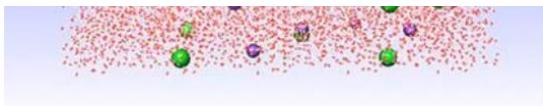


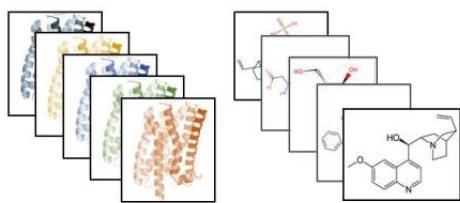
Figure 4 Boîte de simulation d'un RCPG dans environnement physiologique.



Pour cela nous utilisons les logiciels Modeller et Autodock, également installés sur CICADA et parfaitement adaptés pour du calcul parallèle et distribué sur CPU. Ponctuellement nous pouvons utiliser des logiciels de chimie quantique (Gaussian) pour le calcul de propriétés électroniques des systèmes étudiés. L'identification et la prédiction des liens chémogénomiques et des relations structure-activité nécessite d'intégrer de grandes bases de données sur la structure des molécules odorantes ou saines et sur la structure des récepteurs chimiosensoriels. Les descripteurs moléculaires permettront d'encoder les structures chimiques et serviront de millions d'entrées aux modèles de machine learning (SVM, RF, ...) ou de deep learning qui seront ensuite validés par des expériences *in vitro* chez nos collaborateurs (Figure 5). Ces nouveaux résultats alimenteront *in fine* les bases de données initiales et permettront selon un principe de cercle vertueux d'améliorer les modèles *in silico*. L'ensemble de ces méthodes sont notamment implémentées dans différentes librairies Python (Scikit-learn, Pytorch) optimisées pour le calcul CPU et GPU ou inclus dans les extensions du logiciel R.

Pour cela nous utilisons les logiciels Modeller et Autodock, également installés sur CICADA et parfaitement adaptés pour du calcul parallèle et distribué sur CPU. Ponctuellement nous pouvons utiliser des logiciels de chimie quantique (Gaussian) pour le calcul de propriétés électroniques des systèmes étudiés. L'identification et la prédiction des liens chémogénomiques et des relations structure-activité nécessite d'intégrer de grandes bases de données sur la structure des molécules odorantes ou saines et sur la structure des récepteurs chimiosensoriels. Les descripteurs moléculaires permettront d'encoder les structures chimiques et serviront de millions d'entrées aux modèles de machine learning (SVM, RF, ...) ou de deep learning qui seront ensuite validés par des expériences *in vitro* chez nos collaborateurs (Figure 5). Ces nouveaux résultats alimenteront *in fine* les bases de données initiales et permettront selon un principe de cercle vertueux d'améliorer les modèles *in silico*. L'ensemble de ces méthodes sont notamment implémentées dans différentes librairies Python (Scikit-learn, Pytorch) optimisées pour le calcul CPU et GPU ou inclus dans les extensions du logiciel R.

1. Database (DB) of chemosensory receptors and odorants/tastants



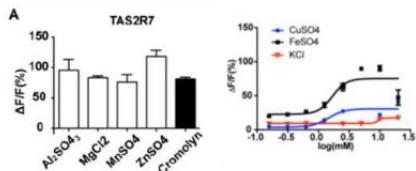
2. Encoding molecular structures into numerical descriptors

U_42	U_18	U_37	U_35	U_38	U_58	U_36	U_46
0.17	0.29	0.63	0.30	0.70	0.26	0.39	0.08
0.80	0.08	0.18	0.23	0.25	0.68	0.03	0.71
0.29	0.53	0.51	0.49	0.61	0.64	0.00	0.13
0.13	0.15	0.50	0.13	0.61	0.28	0.33	0.08
0.15	0.34	0.75	0.15	0.99	0.00	0.33	0.08
0.07	0.17	0.17	0.48	0.11	0.86	0.03	0.21
0.26	0.22	0.50	0.34	0.49	0.43	0.36	0.17
0.57	0.12	0.30	0.23	0.36	0.55	0.14	0.29
0.20	0.15	0.16	0.44	0.12	0.85	0.00	0.33

Computational modeling  
of olfaction & taste

5. Feeding DB

4. Prediction & *in vitro* validation



3. Machine learning  
(SVM, RF, ...)

3'. Deep Learning  
(ANN, ...)

Figure 5 Différentes étapes de la construction du modèle numérique prédictif basé sur des approches d'apprentissage automatique.

---

## **Modélisation de molécules d'intérêt biologique, catalytique et éco-compatible**

**Responsable scientifique :** Fabien Fontaine-Vive

**Laboratoire :** ICN

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- "First-Principles Analysis of CuMg<sub>2</sub>InS<sub>4</sub>: Insights into Optical, Piezoelectric, and Thermoelectric Properties", Sadouki, O., Khelfaoui, F., Fontaine-Vive F. et al., J. Electron. Mater., 2023, 52, 6778–6790
- "New phenylspirodrimanes from the Sponge-Associated Fungus Stachybotrys chartarum MUT 3308", M. Dayras, F. Fontaine-Vive, M. Mehiri et al., Marine Drugs, 2023, 21(3), 135

### **Description :**

Le Service de Modélisation Moléculaire est un service de la Plateforme Technologique de Chimie dont les missions premières sont d'appuyer les travaux de recherche de l'ICN en créant des modèles numériques corroborant les expériences et permettant ainsi de les interpréter et les prédire. J'ai la responsabilité (maintenance, mise en œuvre, conception de méthodes et de modèles numériques, disponibilité et formation) pour les équipes de recherche de l'ICN ainsi que pour le tissu industriel local et national. Les problématiques que l'on est à même d'étudier avec les matériels et logiciels du service peuvent se classer comme suit :

Déterminer les conformations d'énergie minimale, dynamique moléculaire (AMBER)

Calculs de propriétés moléculaires et spectroscopiques (UV, ECD, IR, RMN, neutrons)

Etude de la réactivité chimique par les méthodes de la chimie quantique.

Applications aux sciences des matériaux.

Depuis 10 ans, le service de modélisation moléculaire est spécialisé dans l'étude des propriétés catalytiques de complexes métalliques et de nanoparticules et la compréhension des mécanismes d'inhibition de molécules antivirales.

Les 3 types de calculs de modélisation moléculaire peuvent être présentés ci-après :

Calculs ab-initio Optimisation de structures, calculs d'énergie électroniques, spectres électroniques de dichroïsme circulaire et UV, spectres vibrationnels IR	Calculs semi- empiriques Détermination des conformations d'énergie minimale, dynamique moléculaire	Calculs empiriques Détermination des conformations d'énergie minimale, dynamique moléculaire
---	--	--

### Présentation scientifique du projet :

Pour le groupe « Produits Naturels Marins » de l'équipe « Molécules Bioactives », les calculs de TD-DFT avec GAUSSIAN permettent d'élucider la configuration absolue de molécules bioactives chirales grâce aux calculs de spectres électroniques de dichroïsme circulaire électronique. Au préalable, une recherche conformationnelle au moyen de méthodes de dynamique moléculaire empirique et semi-empirique avec AMBER est nécessaire afin d'identifier les conformations les plus probables des molécules.

La figure 1 illustre la recherche de la configuration absolue de molécules extraites d'éponges marines : les anchinopeptolides qui diffèrent par les groupements R et R'.

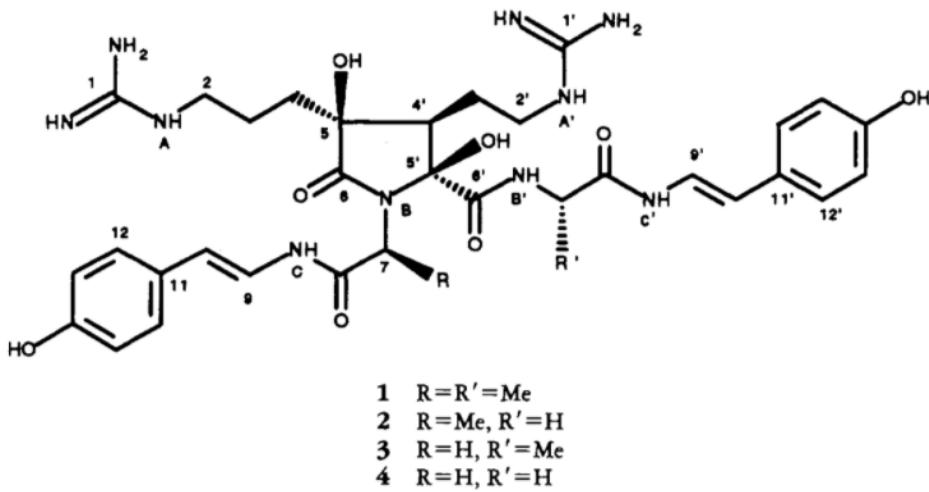


Fig. 1. Stéréochimie relative des molécules anchinopeptolides extraites d'éponges marines

Dix conformations pour chaque molécule conduisent à un calcul d'optimisation DFT puis à un calcul d'états excités TD-DFT, et la moyenne des spectres pondérée par le facteur de Boltzmann a été affichée dans la figure 2 pour une molécule. La publication sur ces travaux est en cours de rédaction et la configuration absolue des trois autres molécules restent à identifier.

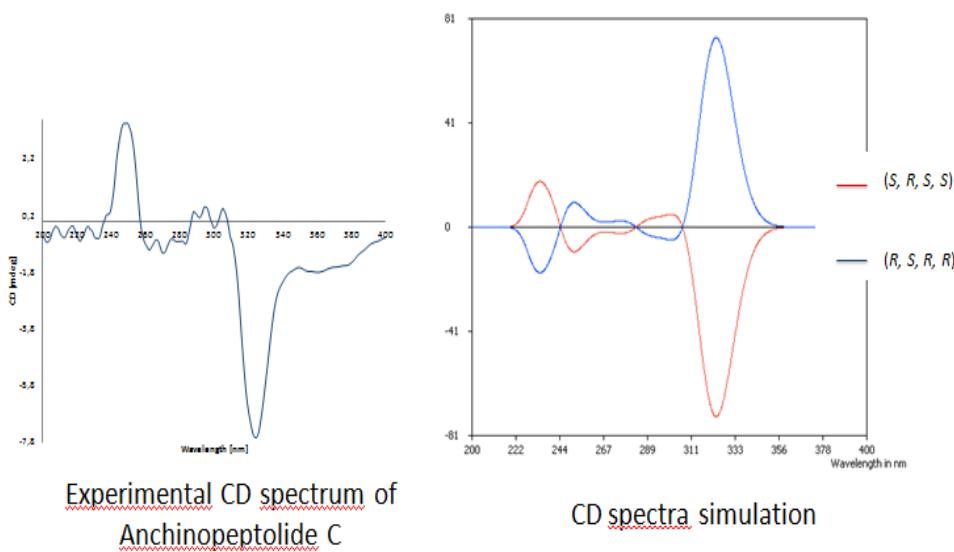


Fig. 2. Spectres ECD expérimental et calculé par TD-DFT de la molécule chirale anchinopeptolide C.

Des calculs similaires, conduits au sein du cluster CICADA ont dès lors et déjà permis de publier plusieurs articles de rang A référencés ci-dessous<sup>17,18,19</sup>.

Pour le groupe « Catalyse, Méthodologies de Synthèse et Chimie Fine » de l'équipe « Arômes Parfums Synthèse et Modélisation », les calculs DFT et TD-DFT permettront d'élucider la configuration absolue de molécules chirales et d'effectuer une étude mécanistique afin d'élucider le mécanisme réactionnel. La figure 3 illustre le projet en cours concernant la cyclisation catalysée par un complexe d'or. Le but du calcul sera de proposer une étude mécanistique permettant d'expliquer la catalyse asymétrique et déterminer si la réaction est cinétiquement ou thermodynamiquement limitante.

<sup>17</sup> New bioactive chlorinated cyclopentene derivatives from the marine-derived Fungus Phoma sp., M. Elsebai, H.A. Ghabbour, N. Legrave, F. Fontaine-Vive, M. Mehiri, *Med Chem Res* 27, 1885–1892 (2018).

<sup>18</sup> Pan-genotypic Hepatitis C Virus Inhibition by Natural Products Derived from the Wild Egyptian Artichoke, M. Elsebai, G. Koutsoudakise, V. Saludese, G. Pérez-Vilaró, A. Turpeinen, S. Mattila, A. M. Pirtilä, F. Fontaine-Vive, M. Mehiri, A. Meyerhans and J. Dieze, *J. of Virology*, 2016, vol. 90 no. 4 1918-1930.

<sup>19</sup> New isocyanide antibiotics with broad spectrum against Gram-negative pathogens produced by cultures of an unidentified fungus, N El Aouad, V Gonzalez, M De la Cruz, L Lorenzo, JR Tormo, J Martin, I Perez-Victoria, C Diaz, F Vicente, F Fontaine-Vive, OP Thomas, O Genilloud, F Reyes, *Planta Med* 201480-PIL72.

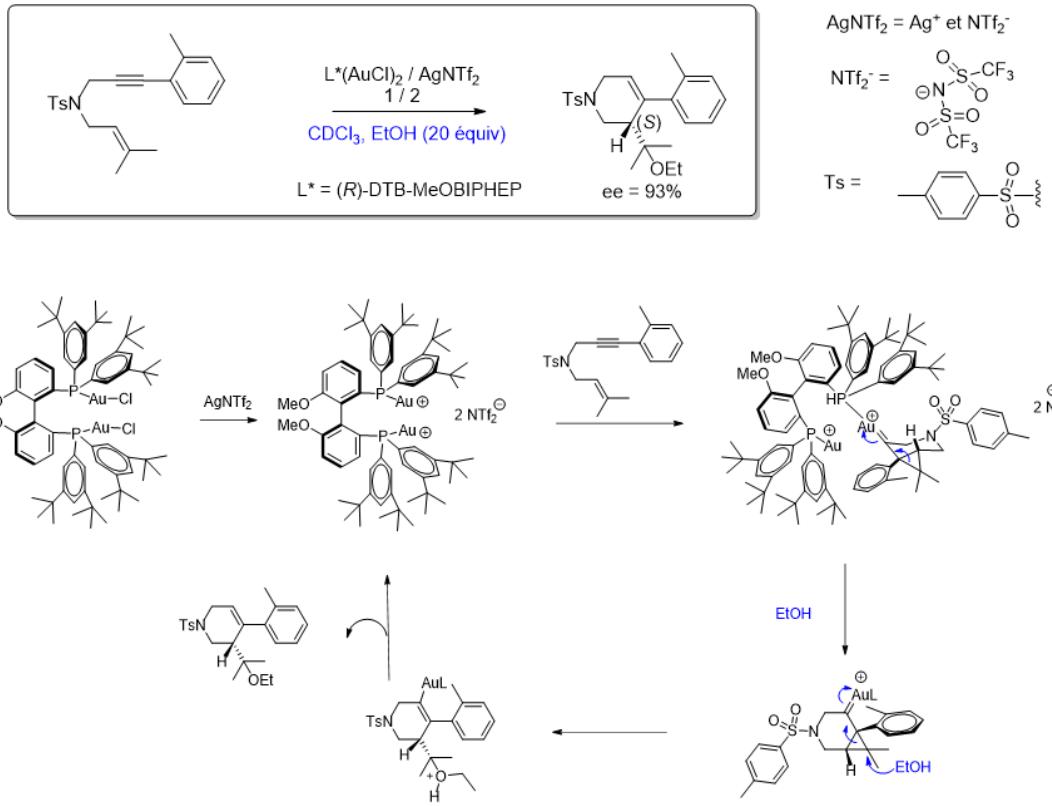


Fig. 3. Mécanisme de cyclisation asymétrique catalysée par des complexes organiques d'or

Des calculs similaires, conduits au sein du cluster CICADA ont dès lors et déjà permis de publier plusieurs articles de rang A référencés ci-dessous<sup>20,21,22,23,24,25</sup>.

Pour l'équipe « Matériaux et Polymères Eco-Compatibles », les calculs DFT dans l'état solide avec le logiciel VASP nous permettront d'interpréter les spectres vibrationnels et de

<sup>20</sup> Gold-catalyzed cycloisomerization of 1,6-cyclohexenylalkyne: an efficient entry to bicyclo[3.2.1]oct-2-ene and bicyclo[3.3.1]nonadiene, Davenel V., Nisole C., Fontaine-Vive F., Fourquez J.-M., Chollet A.-M., Michelet V., The Journal of Organic Chemistry, 2020.

<sup>21</sup> Silver-catalyzed intramolecular [4+2] cycloaddition reaction of amide-1,6-enynes, X. Chen, F. Fontaine-Vive, S. Martini, V. Michelet, *Catalysis Communications*, 2020, 106117.

<sup>22</sup> In(OTf)3-Catalysed Easy Access to Dihydropyranocoumarin and Dihydropyranochromone Derivatives, Boufroua N., Dunach E., Fontaine-Vive F., Achouche-Bouzroura S., Poulain-Martini S., New Journal of Chemistry, 2020.

<sup>23</sup> Synergies in the catalytic activity of bimetallic nanoparticles and new synthetic methods for the preparation of fine chemicals, I. Francesco, F. Fontaine-Vive, S. Antoniotti, Chem. Cat. Chem., 2014, 6, 2784–2791.

<sup>24</sup> Novel radical tandem 1,6-enynes thioacetylation / cyclisation : Au-Pd nanoparticles catalysis versus thermal activation as a function of the substrate specificity, I. Francesco, J. Giauffret, F. Fontaine-Vive, J. Edwards, G. Hutchings, S. Antoniotti, *Tetrahedron*, 2014, Volume 70, Issue 51, 9635-9643.

<sup>25</sup> Access to polycyclic derivatives by triflate-catalysed intramolecular hydroarylation, B. Cacciuttolo, S. Poulain-Martini, F. Fontaine-Vive, M. A. H. Abdo, H. El Kashef, E. Dunach, *EuroJOC* 2014, Issue 33, 7458-7468.

prédir la structure de polymères biosourcés destinés à remplacer les dérivés de produits pétroliers. Les changements conformationnels induits par des étirements pourront ainsi

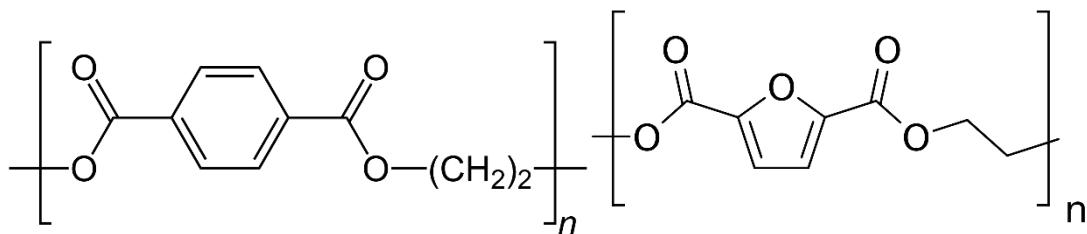


Fig. 4. Unités de répétition de PET (gauche) et PEF (droite)

être élucidés. Par exemple, le PET est un polymère thermoplastique pétrosourcé dont la structure chimique est donnée à la figure 4. Nous travaillons actuellement sur un projet qui vise à l'industrialisation de son « analogue » 100% biosourcé, le PEF. Les structures chimiques de ces 2 polymères sont proches mais diffèrent par la présence d'un noyau furanique dans le PEF à la place du noyau aromatique du PET. Il a été montré expérimentalement que certaines propriétés du PEF sont très différentes de celles du PET, ce qui s'explique par les différences de structures. Il a notamment été proposé que ces différences pourraient provenir du fait que le cycle benzénique peut tourner alors que le cycle furanique ne le peut pas. La présence de l'oxygène sur le cycle furanique est également un élément important (interactions dipolaires) pour expliquer les différences de propriétés. Nous souhaitons donc réaliser un travail plus fondamental, en modélisant les structures de ces composés et en faisant les prédictions des spectres IR, afin d'essayer d'expliquer de façon plus académique les différences de comportement observées expérimentalement.

---

## ***Calcul ab initio de spectres de molécules réservoirs pour le carbone interstellaire et circumstellaire***

**Responsable scientifique :** Patrick Cassam-Chenaï

**Laboratoire :** LJAD

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Poster : "Theoretical study of the vibrational spectra of Naphtalene and its cation with the VMFCI method", (NanoSpace Joint Scientific Meeting, Ljubljana, 9-11/02/2023, Slovénie).

### **Description :**

Le projet se concentrera sur l'étude de 2 familles de molécules considérées comme piégeant une grande partie du carbone interstellaire et circumstellaire, et qui pour cela jouent un rôle clé dans l'astrochimie organique, et possiblement prébiotique. Il s'agit des hydrocarbures aromatiques polycycliques (HAPs) et des fullerènes.

Les HAPs forment une famille de composés organiques constitués de cycles aromatiques d'atomes de carbones dont les valences périphériques sont saturées par des atomes d'hydrogènes. Les HAPs (ou des espèces dérivées proches) sont considérés comme les porteurs des "Bandes Infra-rouges Aromatiques" (BIAs) à ~3.3, 6.7, 7.7, 8.6, 11.3 μm. Ces bandes caractéristiques sont parmi les plus fortes observées en infra-rouge (IR) dans le milieu interstellaire et présentent une certaine variabilité en fonction des sources observées. Ainsi, la spectroscopie IR des HAPs est d'importance capitale pour l'astrochimie et l'astrophysique: une meilleure compréhension de celle-ci est essentielle pour déduire des observations IR, des informations sur la composition des HAPs et les paramètres physiques de leur environnement [1].

Le fullerène le plus emblématique est le buckminsterfullerène, C<sub>60</sub>, parfois également appelé footballène. Ses propriétés physiques uniques sont l'objet de recherches très actives. Elle a été détectée dans une grande variété d'environnements circumstellaires et interstellaires, y compris dans des nébuleuses protoplanétaires [2], grâce à son spectre dans l'infra-rouge (IR) résultant de transitions entre ses niveaux d'énergie vibrationnels. De plus, une expérience récente de spectroscopie haute résolution sur molécules isolées vient de livrer des données d'une qualité sans précédent [3], dont l'interprétation reste

---

très incomplète. C'est pourquoi, une étude théorique du spectre de rotation-vibration du C<sub>60</sub> est du plus vif intérêt.

Nous avons mis au point au laboratoire un code, CONVIV, qui résout l'équation de Schrödinger avec une précision qui correspond à l'état de l'art, pour les degrés de liberté de rotation-vibration moléculaires [4]. Il peut être utilisé pour calculer ab initio le spectre IR de molécules de tailles arbitraires, y compris le C<sub>60</sub> qui comporte 174 degrés de liberté internes. Il a permis de tester les champs de force actuellement disponibles pour ce système dans la littérature [5]. Les résultats ab initio obtenus avec CONVIV seront confrontés aux données expérimentales et observationnelles, afin d'extraire le plus d'informations possible de ces dernières, en particuliers les données de l'interféromètre Matisse et surtout du James Webb space telescope qui devrait être mis en orbite cette année.

#### Références :

- [1] G. Mulas, C. Falvo, P. Cassam-Chenai, C. Joblin, J. Chem. Phys. 149, 144102 (2018). [2] Y. Zhang, S. Kwok, Astrophys. J. 730, 126 (2011).
- [3] P. B. Changala, M. L. Weichman, K. F. Lee, M. E. Fermann, J. Ye, Science 363, 49 (2019).
- [4] P. Cassam-Chenai, J. Liévin, Journal of Computational Chemistry 27, 627-640 (2006).
- [5] Daniel A. Jelski, Laszlo Nemes, and S Broughton Journal of Cluster Science 16, 1 (2005).

## **Uncertainty in nano-particule agglomeration**

**Responsable scientifique :** Christophe Henry

**Laboratoire :** Inria (équipe Calisto)

### **Description :**

#### **Aim and Objectives :**

The objective of this project is to refine existing models for particle agglomeration using Molecular Dynamics simulations. In particular, we aim at analyzing these results to extract statistical information on the morphology of the aggregates formed and the sensitivity of these results to some of the input parameters. The analysis of these results will be used to refine existing macroscopic models based on Population Balance Equation (PBE).

#### **Methodology :**

The methodology used here consists in simulating the agglomeration of nanoscopic carbon quantum dots undergoing purely diffusive motion (Brownian motion).

Carbon quantum dots (CQD) are particles formed by carbon sheets (see Paloncyova et al, 2018). As displayed in Figure 1, we consider in this study CQDs composed of 3 benzene rings on the edge size and of a total of 5 layers. This means that each CQD has a diameter of the central layer equal to 1.6 nm and a height of 1.74 nm. Each CQD is composed of 288 atoms.

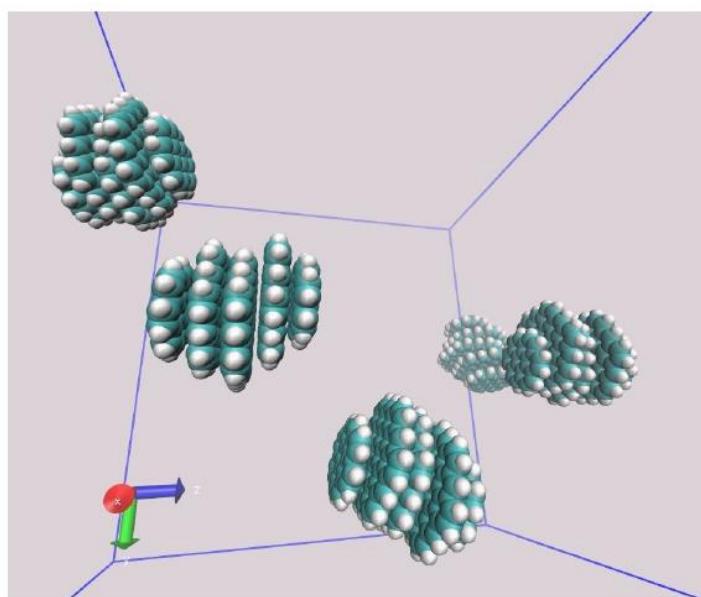


Figure 1 : snapshot of 5 carbon quantum dots (CQD) in a simulation box.

---

Each simulation is performed by randomly introducing 5 CQDs within a cube of size  $L = 10$  nm filled with water molecules (not shown in Figure 1 for the sake of clarity). Periodic boundary conditions are applied in all directions. The temperature is fixed at 300 K. Then, Molecular Dynamics (MD) simulations are performed using the GROMACS simulation package with a simulation time step of 2 fs over an equilibration period of 20 ns. This total simulation time has been chosen long enough to ensure that a single aggregate is formed by the end of the simulation (i.e. all 5 CQDs form one large cluster).

With this project, we aim at running several simulations starting from various random initial configurations of the CQDs and to extract statistical information on the agglomeration dynamics. Among other outputs that will be measured, we plan to measure the aggregation time (i.e. the time needed to have a single cluster formed of all 5 CQDs), the spatial and temporal correlations between consecutive aggregation events, the aggregate morphology (overall shape, porosity, orientation, fractal dimension).

### **Outcome and impact :**

This work is part of a project funded from the European Union's Horizon 2020 research and innovation programme through grant agreement no. 760907, Virtual Materials Marketplace. Within the framework of this EU project, we aim at running here a new scientific workflows that coupled both microscopic approaches (MD here) and macroscopic approaches (PBE here).

The results and analysis of these simulations will be used to write a scientific paper.

### **Bibliography :**

- [1] Palonciová, M., Langer, M. and Otyepka, M., 2018. Structural dynamics of carbon in water and n, n-dimethylformamide probed by all-atom molecular dynamics simulations. *Journal of chemical theory and computation*, 14(4), pp.2076-2083. <https://pubs.acs.org/doi/10.1021/acs.jctc.7b01149#citeThis>
- [2] Jauja-Ccana, V. R., Cordova-Huaman, A. V., Feliciano, G. T., & Gómez, A. L. R. T. (2021). Experimental and molecular dynamics study of graphene oxide quantum dots interaction with solvents and its aggregation mechanism. *Journal of Molecular Liquids*, 335, 116136. <https://doi.org/10.1016/j.molliq.2021.116136>
- [3] Izvekov, S., & Violi, A. (2006). A coarse-grained molecular dynamics study of carbon nanoparticle aggregation. *Journal of chemical theory & computation*, 2(3), 504-512. <https://doi.org/10.1021/ct060030d>

## **pH-Dependent Adsorption of poly(Acrylic Acid) onto Amorphous TiO<sub>2</sub> Nanoparticles**

Projet ayant accès à Azzurra dans le cadre de la solidarité UniCA – Université de Kharkiv (voir R.A. 2022)

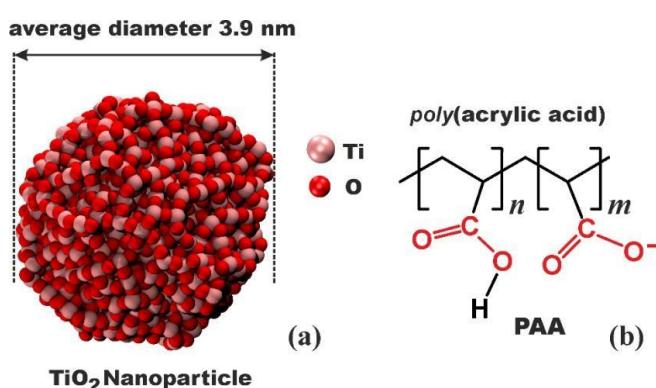
**Responsable scientifique :** Oleg N. Kalugin, Alexander Kyrychenko

**Laboratoire :** V.N. Karazin Kharkiv National University, Department of Inorganic Chemistry

**Description :**

### **1. Introduction**

Titanium (IV) oxide nanoparticles (TiO<sub>2</sub>-NPs) have promising perspectives in a broad range of applications, such as nuclear science, the development of advanced reactors, the pigment industry, and nanomedicine. TiO<sub>2</sub>-NPs have technologically essential properties that can be used as substrates to build up Grätzel-type solar cells, which prove to be an economical material for photovoltaic, photocatalytic, and sensing applications. There is a crucial need for the stabilization of titanium dioxide, which is of significant importance in manufacturing high-quality products. Moreover, developing a computational model of TiO<sub>2</sub>-NPs with a stabilization agent and studying chemical behavior is very important.

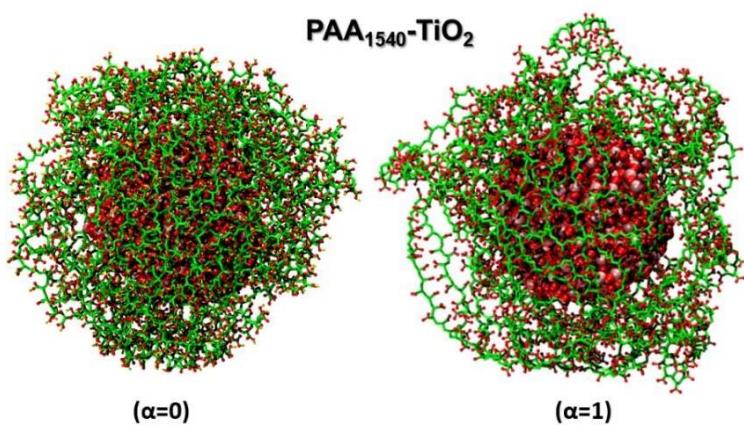


**Figure 1.** (a) Amorphous TiO<sub>2</sub> nanoparticle with the average diameter of 3.9 nm. (b) Molecular structure of poly(acrylic acid) (PAA) monomers in a protonated neutral (n) and dissociated charged (m) forms, respectively.

### **2. Project description**

The project is devoted to studying the pH-dependent adsorption of polyacrylic acid onto amorphous TiO<sub>2</sub> nanoparticles (Figure 1) by classical MD simulations. The structure and adsorption dynamics of single-chain PAA1540 in an aqueous solution at neutral and protonated charged states were considered (Figure 1). PAA is a weak polyelectrolyte with pKa 4.6 [1]. So, to mimic pH changes, the protonation degrees  $\alpha$  of PAA are set to 0 and 1, which corresponds to pH<2 and pH>10, respectively. The G53a6 force field and SPC model were used to treat the PAA polymer and explicit water molecules, respectively.

To model TiO<sub>2</sub>-NPs, the repulsion and dispersion terms of nonbonded interactions were computed using the Lennard-Jones (LJ) 12-6 pairwise potential energy function adopted from [2]. The performance of these 12-6 LJ pair potentials has recently been validated by modeling the adsorption behavior of small ligands and peptides onto TiO<sub>2</sub> nanoparticles. All MD simulations and analyses were carried out using the GROMACS set of programs, version 2021.4.



**Figure 2.** MD simulations of PAA1540-TiO<sub>2</sub> nanoparticles at the two degrees of PAA ionization  $\alpha = 0$  and 1, respectively.

Our MD simulations revealed that polymer adsorption onto amorphous TiO<sub>2</sub> nanoparticles was critically dependent on the ionization degree of PAA (Figure 2). The tightly bound structure was formed at  $\alpha = 0$ , whereas the loose polymer coating was observed upon PAA dissociation at  $\alpha = 1$ . Our study demonstrates that the pH-tuned PAA coating of TiO<sub>2</sub> nanoparticles opens up the opportunity for their use as “smart” TiO<sub>2</sub>-polymer nanocomposites [3].

---

**References:**

1. Kyrychenko A., Blazhynska M. M., Slavgorodska M. V., Kalugin O. N. Stimuli-responsive adsorption of poly(acrylic acid) onto silver nanoparticles: Role of polymer chain length and degree of ionization. *J. Mol. Liq.* 2019. 276, 243-254.
2. Luan B., Huynh T., Zhou R. Simplified tio2 force fields for studies of its interaction with biomolecules. *J. Chem. Phys.* 2015. 142, 234102.
3. Prud M. V., Blazhynska M. M., Kyrychenko A., Kalugin O. N. pH- Dependent Adsorption of Poly(Acrylic Acid) onto Titanium Dioxide Nanoparticles. Manuscript in preparation for RSC Advances.

---

## ***Competition of solvation and ionic aggregation in ionic-molecular systems of different nature: influence on transport properties***

Projet ayant accès à Azzurra dans le cadre de la solidarité UniCA – Université de Kharkiv (voir R.A. 2022)

**Responsable scientifique :** Oleg N. Kalugin

**Laboratoire :** V.N. Karazin Kharkiv National University, Department of Inorganic Chemistry

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Successfully defended the PhD thesis on the project topic by Dmytro Dudariev : « Compétition entre la solvatation et l'agrégation ionique dans des systèmes ioniques-solvant: influence sur les propriétés de transport »

The work can be accessed via thesis.fr or directly by following address:

<https://pepite-depot.univ-lille.fr/LIBRE/EDSMRE/2023/2023ULILR009.pdf>

Currently, two research papers are in the process of being written with the intention of subsequent publication:

- Microstructure and Transport Properties of Lithium Hexafluorophosphate in Binary Mixture of Dimethyl Carbonate With Ethylene Carbonate From Molecular Dynamics Simulation (D.S. Dudariev, Y.A. Holubenko, O.N. Kalugin)
- Computational Insights into Structural and Transport Properties: 1-1'-Spiropyrrolidinium Tetrafluoroborate in Acetonitrile (D.S. Dudariev, I.S. Vovchynskyi, K.O. Lohachova O.N. Kalugin)

### **Description :**

The objective of this project is to analyze the microscopic structure of the series ion-molecular systems that widely used for practical electrochemistry and to characterize the effect of the ion aggregation on the transport properties of these systems. By using molecular dynamics simulation, the following systems were investigated: (i) the solutions of LiPF<sub>6</sub> in dimethyl carbonate / ethylene carbonate mixture (1:1), (ii) the solutions of SBPBF<sub>4</sub> in acetonitrile, and (iii) the mixtures of room-temperature ionic liquids (ILs) C<sub>4</sub>mimX (X= BF<sub>4</sub><sup>-</sup>, PF<sub>6</sub><sup>-</sup>, TFO<sup>-</sup>, TFSI<sup>-</sup>) with dipolar aprotic solvents such as acetonitrile,  $\gamma$ -butyrolactone and propylene carbonate.

Two approaches for microscopic structure investigations were explored:

- (i) in terms of radial distribution functions (RDF), running coordination numbers (RCN)

---

and (ii) Voronoi polyhedra (VP) analysis. The former ones were used for the detailed description of ion subsystems in terms ionaggregates. Special attention was paid to the algorithm of calculation of transport properties: viscosity and diffusion.

For all the systems the aggregate analysis showed the formation of the ionic continuous network with the increase of electrolyte concentration. This affects significantly diffusivity and viscosity in these solutions.

Voronoi polyhedra analysis of ILs-solvent mixtures showed that below the IL mole fraction of about 0.2, the ions are well solvated by the solvent molecules, but above this mole fraction they start to form contact pairs, while the solvent molecules, expelled from the vicinity of the ions, self- associates.

---

## **Physique non linéaire, fluides complexes**

**EUR Spectrum**

---

## **Convex integration for Euler 2-D**

**Responsable scientifique :** Eric Simonnet

**Laboratoire :** INPHYNI

### **Description du projet :**

La finalité du projet est de calculer une solution faible dissipative de classe C0 des équations d'Euler 2-D.

Des résultats récents (Isset, 2018, Buckmaster et al. Comm.Math, 2019, Giri & Radu, 2023) ont permis de démontrer la conjecture d'Onsager pour les équations d'Euler en 2-D et 3-D en utilisant des techniques d'intégration convexe introduite par Nash, Kuiper (1953) et Gromov (80s) et adaptées aux équations d'Euler par De Lellis et Szekelyhidi à la fin des années 2000s. Il s'agit d'une approche itérative d'homogénéisations successives permettant de contrôler la norme C0 du tenseur de Reynolds via une décomposition de Nash.

Cette approche permet de calculer des classes de solutions faibles admissible, i.e. convergeant vers un profil d'énergie spécifiée avec une régularité entre 0 et 1/3 (Onsager) pour les équations d'Euler 3D (via Mikado). Une technique similaire permet d'avoir le même résultat pour les équations d'Euler 2D mais seulement avec une régularité C0 (Choffrut, Szekelyhidi).

Du fait de la forte composante algorithmique des techniques d'intégration convexe, il est possible de les implémenter numériquement:

dans le contexte géométrique pour l'embedding C1 du tore par Borelli et al. PNAS, 2012,  
dans le contexte des équations d'Euler 3D par T.Matsumoto (2020s).

Ce projet consiste donc à aborder le cas des équations d'Euler 2D pour la classe C0. L'avantage est l'aspect 2D permettant l'utilisation de résolution beaucoup plus fine qu'en 3D. Ce calcul n'a jamais été effectué jusqu'à présent.

Le projet a plus long terme envisage 1) de modifier l'algorithme dans un contexte stochastique, et/ou de Mikado intermittents et/ou 2) d'introduire des réseaux de neurones pour les itérations de Nash et/ou 3) de transposer la méthode aux équations SQG (surface quasigeostrophic eqs).

---

## Rhéologie des suspensions concentrées

**Responsable scientifique :** François Peters

**Laboratoire :** INPHYNI

### Résultats obtenus grâce aux moyens informatiques attribués :

Article en préparation :

Orsi Michel, Laurent Lobry, Elisabeth Lemaire, and François Peters. Mass and momentum detailed balance during particle migration in the pressure driven flow of a non-Brownian suspension.

Communications orales :

Michel Orsi, Elisabeth Lemaire, Laurent Lobry, and François Peters. Simulation of pressure-driven flow of non-brownian suspensions : Local constitutive laws. In EuroMech Symposium- Suspension flows and rheology : inertia, shape and roughness matter, Nice, France, June 2023.

Michel Orsi, Laurent Lobry, and François Peters NON-BROWNIAN SUSPENSIONS IN PRESSURE-DRIVENFLOW: NUMERICAL SIMULATIONS WITHFRAME-INVARIANT SUB-GRID CORRECTIONS. 8TH PACIFIC RIM CONFERENCE ON RHEOLOGY, May 15-19, 2023

### Description du projet :

L'équipe Rhéologie des Suspensions Concentrées de l'Institut de Physique de Nice s'intéresse aux propriétés d'écoulement des suspensions de particules solides en écoulement lent. Du point de vue macroscopiques, ces propriétés sont prises en compte par des lois de comportement de fluides continus décrivant des relations mathématiques entre grandeurs macroscopiques (contraintes, taux de déformation, flux de particules). Les fonctions matérielles (viscosité généralisées, coefficient de diffusion de particules etc.) mises en jeu reflètent, de façon encore partiellement incomprise, les mécanismes physiques actifs à l'échelle des particules (interactions hydrodynamiques, interactions directes entre particules, en particulier contact). Ces mécanismes microscopiques mettent en jeu des paramètres microscopiques assez divers (propriétés rhéologiques du fluide suspendant, taille, forme des particules, ainsi que leur élasticité et leurs propriétés de surface). Depuis une dizaine d'année, l'équipe RSC tente d'approfondir la compréhension

---

des liens entre ces mécanismes à l'échelles microscopiques et le comportement mésoscopique à une échelle supérieure. Pour ce faire, nous nous appuyons sur des approches combinées d'expériences de rhéologie macroscopique classique, de mesures locales (à l'échelle de la particule), de simulations résolues à l'échelle des particules, de modélisation et de calcul d'écoulement en milieux continus. Plusieurs de ces approches nécessitent la mise en œuvre de calculs assez lourds, c'est pourquoi le projet présenté ici se décline en plusieurs volets qui seront détaillés ci-dessous. On notera qu'une partie de ces travaux fait l'objet d'un projet ANR (AMARhéo 2019- 2022) qui a notamment financé la machine SMP du cluster Azzurra.

### Simulations discrètes de suspensions

Les simulations discrètes ont pour but de décrire la dynamique des particules et du fluide en tenant compte des interactions pertinentes au sein de la suspension, et d'en déduire, au moyen de moyenne, les lois de comportements macroscopiques permettant en retour de décrire le matériau comme un milieu continu. La difficulté de cette démarche réside dans le haut niveau de couplage existant entre la dynamique du fluide (partie « fluide » du code) et celle des particules (partie DEM: Discrete Element Method). Nous utilisons pour ce faire une méthode de domaines fictifs, qui permet de calculer un écoulement dans tous le volume représentatif de la suspension. La présence des particules, notamment leur caractère indéformable, est prise en compte via une densité de force calculée à chaque pas de temps, et qui traduit les relations de conservations mécaniques fondamentales. L'écoulement ainsi déterminé permet de calculer les interactions hydrodynamiques entre particules qui en résultent. Les écoulements à une échelle plus petite que la maille ne sont évidemment pas accessibles et donnent lieu à des corrections de force et de moment hydrodynamiques « sousmaille » sur les particules. Du point de vue des particules, outre ces interactions hydrodynamiques, les interactions directes sont de première importance. Dans notre cas, il s'agit d'interactions de contact, mettant en jeu à la fois des forces élastiques, de frottement et d'adhésion. Cette approche, qui nécessite de décrire des échelles inférieures à la taille des particules dans des systèmes comptant un nombre suffisant de particules pour obtenir des statistiques suffisantes, demande une

---

puissance de calcul importante. Elle a contribué ces dernières années à clarifier l'influence du frottement entre particules sur la rhéologie des suspensions.<sup>26 27 28 29</sup>

Deux codes différents fondés sur deux méthodes légèrement différentes, seront mis en oeuvre. Le premier utilise la méthode FCM (Force Coupling Method). L'écoulement est calculé en utilisant des méthodes spectrales (typiquement F.F.T.) implémenté dans le langage Fortran. Ce code, faiblement parallélisé (procédure OpenMP), permet de gérer des systèmes de petite taille (typique 1000 particules) pour des durées de calcul d'une quinzaine de jours sur 4 coeurs. Ce code a déjà été validé et utilisé sur le cluster Cicada, et a donné lieu à plusieurs publications, . Un deuxième code (Domaine fictif sous OpenFOAM) a donné lieu sous une forme différente à plusieurs publication, . Son adaptation à l'environnement OpenFOAM est en cours de mise au point. Il s'agit ici de paralléliser massivement les calculs en vue de simuler des systèmes de très grande taille. Ceci permettra de décrire la dynamique des particules jusqu'à des échelles intermédiaires (mésoscopiques) entre la taille des particules et les distances typiques caractéristiques des écoulements pertinents pour les applications. C'est notamment crucial pour les écoulements où la fraction volumique de solide et le taux de cisaillement présentent des gradients spatiaux et imposent donc au système une telle échelle intermédiaire. Ces gradients sont à l'origine de flux de particules qui, s'ils ont été étudié depuis longtemps du point de vue expérimental, sont actuellement mal compris du point de vue microscopique, et imparfaitement décrits par les modèles de milieux continus actuels. La majorité des simulations s'est jusqu'ici limitée à des systèmes homogènes. La simulation discrète de tels systèmes hétérogènes est donc d'un très grand intérêt dans ce domaine. Le calcul des écoulements est réalisé par la méthode des volumes finis dans l'environnement libre OpenFOAM qui offre des outils standards de résolution des équations de Navier-Stokes / Stoke de façon parallèle (procédure MPI). L'adaptation de la méthode des domaines fictifs à cet environnement, avec en particulier la parallélisation

---

<sup>26</sup> Stany Gallier, Elisabeth Lemaire, Laurent Lobry, and François Peters. A fictitious domain approach for the simulation of dense suspensions. *Journal of Computational Physics*, 256:367–387, 2014.

<sup>27</sup> Stany Gallier, Elisabeth Lemaire, François Peters, and Laurent Lobry. Rheology of sheared suspensions of rough frictional particles. *Journal of Fluid Mechanics*, 757:514–549, 2014.

<sup>28</sup> François Peters, Giovanni Ghigliotti, Stany Gallier, Frédéric Blanc, Elisabeth Lemaire, and Laurent Lobry. Rheology of non-brownian suspensions of rough frictional particles under shear reversal: A numerical study. *Journal of rheology*, 60(4):715–732, 2016.

<sup>29</sup> Laurent Lobry, Elisabeth Lemaire, Frédéric Blanc, Stany Gallier, and François Peters. Shear thinning in non-brownian suspensions explained by variable friction between particles. *Journal of Fluid Mechanics*, 860:682–710, 2019.

---

de la partie DEM de la méthode, constitue une partie du travail de thèse de Michel Orsi, dont le contrat est également financé par l'ANR AMARhéo. Il est prévu que des jobs puissent être lancés sur typiquement 100-200 coeurs.

### Simulations en milieux continus

L'objet de cette partie est de développer et utiliser des modèles en milieux continus capables de décrire l'écoulement des suspensions, y compris la diffusion des particules, dans des géométries d'écoulement complexes rencontrées soit dans des expériences de recherche menées dans l'équipe, soit dans le traitement de problématiques industrielles. Le verrou principal de cette activité réside dans la nécessité de décrire les équations de transport de la quantité de mouvement et de la masse de la suspension vue comme un milieu continu, ainsi que le transport des particules (diffusion) par rapport à ce milieu continu. L'approche suivie, classique, consiste en la définition, en plus de la phase moyenne (suspension), d'une phase solide, certes transportée par la suspension, mais également capable de mouvement relativement à elle. La problématique principale dans ce domaine consiste à définir les lois de comportement décrivant le transport de cette phase solide à partir de contraintes développées dans la phase particulaire. Les modèles actuels peinent en effet à reproduire les résultats expérimentaux de l'équipe RSC et plus largement de la communauté scientifique qui s'intéresse à ces problèmes. Les équations aux dérivées partielles issues des modèles adoptés sont résolus par la méthode des volumes finis dans l'environnement OpenFOAM, qui comme cela a déjà été rappelé, permet une parallélisation par la procédure MPI. Ce travail, déjà en cours depuis quelques années<sup>30</sup>, a été mis en oeuvre au sein du cluster Cicada, et constitue le travail de thèse d'Ayoub Badia, dirigé conjointement par Yves D'Angelo au laboratoire J.A.D. et Laurent Lobry au laboratoire InPhyNi.

### Traitement d'images expérimentales par des méthodes d'intelligence artificielle

L'équipe RSC a développé des méthodes de rhéologie locale capable de mesurer, en plus des grandeurs rhéologiques intégrées classiques (couple, taux de cisaillement), des grandeurs locales au sein de la suspension (fraction volumique locale, taux de

---

<sup>30</sup> Talib Dbouk, Elisabeth Lemaire, Laurent Lobry, and F Moukalled. Shear-induced particle migration: Predictions from experimental evaluation of the particle stress tensor. *Journal of Non-Newtonian Fluid Mechanics*, 198:78–95, 2013.

---

déformation local). Ces mesures locales constituent une très grande valeur ajoutée pour l'étude quantitative des lois responsables du transport de la suspension et de la phase solide au sein d'un écoulement. Elles permettent en effet de mesurer directement les gradients de vitesse, taux de cisaillement, fraction volumique qui déterminent les différents flux (matière, quantité de mouvement etc). Elles ont notamment permis de mettre en évidence le rôle primordial des contacts entre particules dans les propriétés rhéologiques des suspensions non-browniennes.<sup>31</sup> <sup>32</sup>

Ceci se fait au prix d'expériences complexes et d'un traitement de données assez lourd. En effet les suspensions utilisées doivent être rendues transparentes en adaptant l'indice des particules à celui du fluide, qui est lui-même rendu visible par l'adjonction d'un colorant fluorescent activé par une nappe laser, dont l'image est enregistré par une caméra. L'épaisseur de la nappe laser étant plus fine que le taille des particules, celles-ci apparaissent comme des disques noirs sur les images enregistrées par la caméra. L'essentiel du traitement de ces images consiste à les segmenter pour séparer l'intérieur des particules du liquide avant de déterminer la position des centres des particules. Ce traitement est classiquement réalisé par des méthodes morphologiques, mais cela nécessite le réglage de nombreux paramètres qui sont fortement dépendant des conditions d'éclairages liées à de subtiles variations de l'indice de réfraction de la suspension. Pour accélérer le traitement et ne plus dépendre du réglage de ces paramètres, nous nous sommes tournés vers des techniques d'intelligence artificielle. Ce travail, effectué en collaboration avec Bruno Figliuzzi du Centre de Morphologie Mathématique de Mines ParisTech, est basé sur un algorithme de réseau neuronal convolutif<sup>33</sup> (CNN). « L'apprentissage » du réseau se fait en utilisant des images de nos expériences de rhéologies locales, segmentées par une méthode classique. Une image alimente l'entrée du réseau et la sortie calculée par le réseau est comparée à la « vérité terrain » (i.e l'image segmentée). Une fonction de « coût » calcule alors l'écart entre la sortie et cette vérité terrain. Pour minimiser ce « coût », les nombreux paramètres du réseau sont subtilement modifiés par un processus de « descente de gradient ». Ce

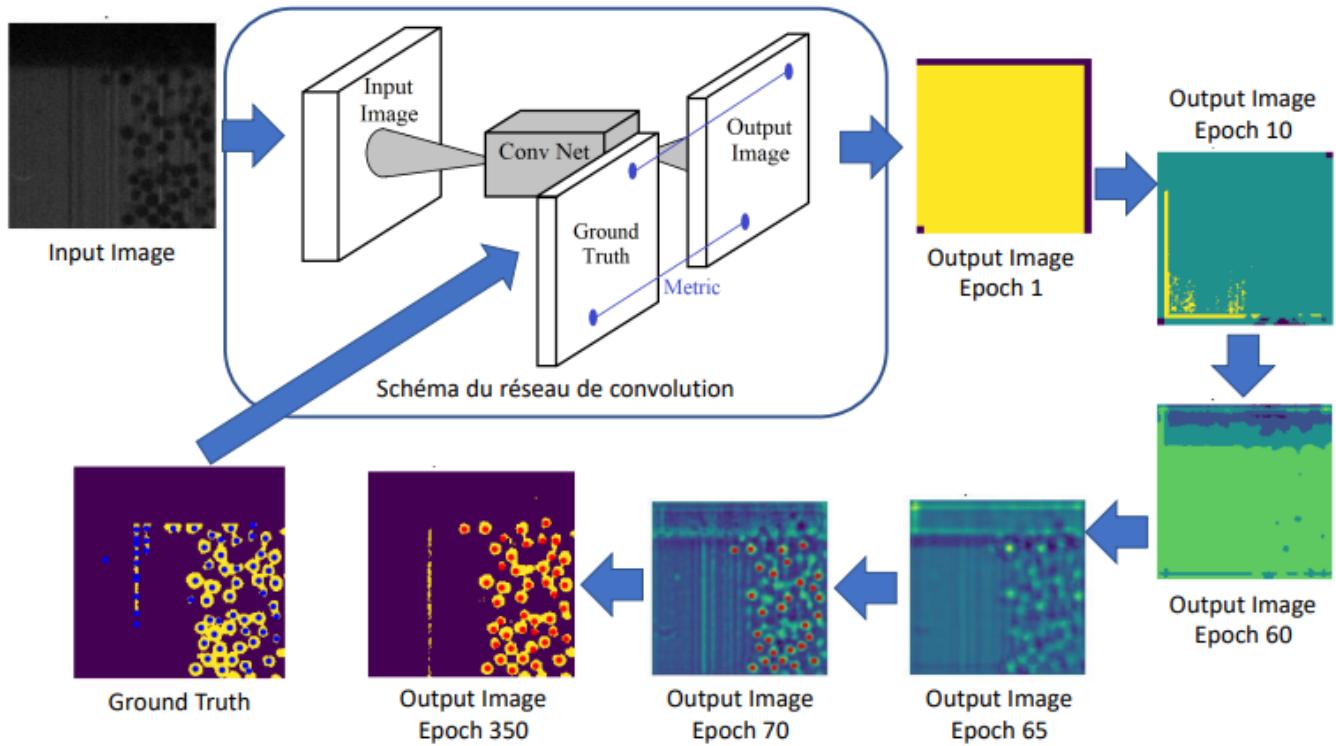
---

<sup>31</sup> Frédéric Blanc, François Peters, and Elisabeth Lemaire. Experimental signature of the pair trajectories of rough spheres in the shear-induced microstructure in noncolloidal suspensions. *Physical review letters*, 107(20):208302, 2011.

<sup>32</sup> Frédéric Blanc, Elisabeth Lemaire, Alain Meunier, and François Peters. Microstructure in sheared non-brownian concentrated suspensions. *Journal of rheology*, 57(1):273–292, 2013.

<sup>33</sup> Chen, Qifeng, Jia Xu, and Vladlen Koltun. "Fast image processing with fully-convolutional networks." *Proceedings of the IEEE International Conference on Computer Vision*. 2017.

processus est itéré avec une série d'images différentes jusqu'à minimiser le « coût ». Le schéma ci-dessous illustre cet apprentissage. Les images de sorties sont calculées à partir d'une même image d'entrée sur la base des paramètres d'apprentissages obtenus au bout d'un nombre d'itération (epoch). Les points rouges sur l'image de l'epoch 350 correspondent aux particules détectées par le réseau sur l'image segmentée. Ce résultat contient moins d'erreurs de détection que sur la vérité terrain (points bleus).



Fort de ces premiers résultats prometteurs, nous orientons nos travaux vers l'apprentissage sur une banque d'images de synthèse dans le but d'adapter la détection à tout type de forme de particules (cubes, fibres, etc ...) ainsi qu'à une extension à des images en trois dimensions. La programmation et l'apprentissage supervisé du réseau sont réalisés en python 3.7 à l'aide des bibliothèques Pytorch-Cuda, Numpy, scikit-learn et scikit-image. Les calculs sont effectués sur GPU (Pytorch gère CUDA de manière transparente). La base de données d'images nécessite une volumétrie d'environ une dizaine de Go (voir plus dans le cas de la 3D) stockée transitoirement sur Azzurra.

---

## Wave turbulence simulations and theory

**Responsable scientifique :** Sergey Nazarenko

**Laboratoire :** INPHYNI

### Résultats obtenus grâce aux moyens informatiques attribués :

Zhu Y, Semisalov B, Krstulovic G, et al. Direct and Inverse Cascades in Turbulent Bose-Einstein Condensates[J]. Physical Review Letters, 2023, 130(13): 133001. On the cover .

<https://link.aps.org/doi/10.1103/PhysRevLett.130.133001>

### Description :

The proposed research project is part of the Simons Foundation international collaboration project "Wave Turbulence". Our research aims at developing solid theoretical foundations for wave turbulence systems and modelling experiments in wave turbulence performed in laboratory by partners of this collaboration. In this work, large-scale numerical simulations of different wave systems will be performed. We focus on those systems (e.g. Bose-Einstein Condensate (BEC) and nonlinear optics) which can be described by the Gross-Pitaevskii equation(GPE). Then we propose to confront the numerical and theoretical results coming from Weak Wave Turbulence theory (WWT) for BEC, and also numerical and experimental results for nonlinear optics. The Gross-Pitaevskii equation

$$i\hbar \frac{\partial \psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi(x,t) + g|\psi(x,t)|^2 \psi(x,t) + F + D$$

is also known as the nonlinear Schrödinger equation, where  $\psi^{(x,t)}$  is the order parameter indicating the condensate wave function in BEC (  $\psi^{(x,t)}$  may represent different physical quantities in other physical contexts), F and D represent possible external forcings and dissipation mechanisms,  $\hbar= h / 2\pi$  and  $g=(4 \pi \hbar^2 a_s) / m$  is the coupling constant fixed by the s-wave scattering length as that models a local interaction between bosons. Here we set  $g=1$  corresponding to the defocusing GP.

---

## ***Etude de la turbulence d'ondes dans les fluides stratifiés et en rotation***

**Responsable scientifique :** Giorgio Krstulovic

**Laboratoire :** Lagrange / OCA

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Le cluster Azzurra nous a permis de réaliser de nombreuses simulations numériques directes d'écoulements stratifiés dans un domaine périodique à l'aide du code pseudo-spectral Fluidsim. Nous avons réalisé deux types de simulations : celles utilisant les équations dynamiques complètes, et celles où des structures non propagatives (les modes vorticaux) sont retirées artificiellement de l'écoulement à l'aide d'une projection dans l'espace spectral. Notre calculateur local, Azzurra et le calculateur national Jean-Zay nous ont permis d'effectuer une étude paramétrique en faisant varier les nombres de Froude et de Reynolds de flottabilité sur plusieurs décades en réalisant plus 80 simulations.

Nous préparons actuellement 3 articles que nous comptons soumettre à des revues à comité de lecture :

- Outre l'étude des ondes, les simulations réalisées constituent une base de données utile pour étudier d'autres problématiques des écoulements stratifiés (compréhension des spectres d'énergies, détermination du coefficient de mélange, ...). Celle-ci contient les spectres spatiaux et spatio-temporelles des quantités d'intérêt, ainsi que les mesures des quantités intégrales de l'écoulement (taux de dissipation d'énergie, rms de la vitesse, ...). Nous préparons donc un article "Regimes in stratified turbulence forced by horizontally divergent modes analyzed from a comprehensive open dataset" présentant notre jeu de données, que nous prévoyons de rendre public.
- Nous avons caractérisé la présence d'ondes dans les écoulements avec et sans modes vorticaux. Nous avons observer que retirer les modes vorticaux permettait de se rapprocher d'un système d'ondes, mais n'était pas suffisant pour que ces ondes interagissent de façon faiblement non-linéaire. Nous préparons un article "Internal gravity waves in stratified flows with and without vortical modes" concernant ces travaux.
- En changeant le rapport d'aspect et la fréquence du forçage, nous avons remarqué qu'il était possible d'avoir une meilleure excitation des ondes, indépendamment de la présence de modes vorticaux. Nous prévoyons donc la préparation d'un troisième article "Toward a weak turbulence regime in stratified flows simulations". Nous comptons utiliser les heures de calcul restantes à notre projet pour réaliser des simulations analogues dans le cas des fluides en rotation.

## Résumé

Comme de nombreux écoulements géophysiques, les écoulements atmosphériques et océaniques sont dans un état *turbulent*, et sont fortement impactés par la *stratification* et la *rotation*. La compréhension de ces écoulements est encore aujourd’hui un sujet de recherche fondamentale en turbulence et physique statistique. Il s’agit également d’un enjeu majeur dans l’amélioration des prédictions faites par les modèles météorologiques et climatiques.

Une particularité des écoulements stratifiés et/ou en rotation est leur capacité à supporter la propagation d’*ondes internes* : les *ondes de gravité* dans les fluides stratifiés et les *ondes d’inertie* dans les fluides en rotation. L’énergie de l’écoulement est alors répartie entre trois composantes : l’énergie cinétique contenue par les ondes, l’énergie cinétique contenue par les tourbillons, et l’énergie potentielle. En plus de la dynamique des tourbillons, qui constituent les écoulements incompressibles, il est donc nécessaire de décrire le comportement des ondes, et de leurs interactions avec les tourbillons et les structures aux grandes échelles.

La *Théorie de la Turbulence d’Ondes* est utilisée pour décrire les systèmes d’ondes dispersives, interagissant de façon faiblement non-linéaire, avec une dissipation négligeable. Elle est utilisée pour étudier de nombreux systèmes d’ondes, comprenant les écoulements stratifiés et/ou en rotation dans la limite des faibles nombres de *Froude* (forte stratification) et de *Rossby* (forte rotation) et d’un grand nombre de *Reynolds* (régime turbulent). Cette théorie permet, entre autres, d’écrire une *équation cinétique* qui décrit l’évolution de l’amplitude des modes, et fournit des prédictions concernant les quantités statistiques d’intérêt.

Les campagnes de mesures atmosphériques et océaniques permettent d’obtenir les spectres d’énergie cinétique et de température, mais ne permettent de quantifier ni les proportions des différents types d’énergie, ni les échanges entre celles-ci. Bien que des avancées expérimentales importantes ont permis d’améliorer notre compréhension des écoulements stratifiés et/ou en rotation, les expériences ne permettent pas encore d’observer les régimes d’intérêt géophysique, notamment à cause des effets de tailles finies et de la présence de parois. Ces limitations motivent l’utilisation de simulations numériques pour tenter d’obtenir plus d’informations sur ces écoulements.

Le but de ce projet est de confronter les prédictions fournies par la Théorie de la Turbulence d’Ondes à des simulations numériques directes idéalisées, et de quantifier l’importance des ondes dans les différents régimes turbulents, c.-à-d. lorsque les nombres de Froude, de Rossby et de Reynolds varient. Nous utiliserons les équations de Navier-Stokes dans l’*approximation de Boussinesq* en imposant des *conditions aux limites périodiques*. Elles seront résolues à partir d’un *code pseudo-spectral* parallélisé avec *MPI*.

## 2 Présentation générale

La stratification et la rotation rendent les écoulements turbulents anisotropes. La force de gravité tend à ramener les particules fluides à leur niveau de flottabilité moyen. Dans le cas d’une stratification stable, la gravité tend donc à inhiber les mouvements verticaux. L’écoulement est alors marqué par la présence de structures allongées horizontalement en forme de “pancakes”, et une partie importante de l’énergie est transférée vers les modes de cisaillement verticaux. Dans le cas purement tournant, quand le nombre de Rossby est faible, l’écoulement est caractérisé par un équilibre selon l’horizontale appelé équilibre géostrophique (le gradient de pression horizontal équilibre la force de Coriolis). Si cet équilibre est réalisé exactement, on montre que l’écoulement est invariant selon la verticale (théorème de Taylor-Proudman), ce qui explique la présence de structures en forme de colonnes, dites de Taylor. Une partie importante de l’énergie de l’écoulement est alors contenu dans ces colonnes.

Par définition des écoulements turbulents, la dissipation est négligeable sur une grande gamme d'échelle, appelée gamme inertie. L'équilibre entre stratification d'une part, et l'inertie d'autre part, est atteints à une échelle intermédiaire, appelée échelle d'Ozmidov. En dessous de cette échelle, l'écoulement présente les caractéristiques d'une turbulence homogène et isotrope. En revanche, aux grandes échelles, la turbulence est fortement anisotrope.

Les écoulements stratifiés et/ou en rotation ont la capacité à supporter la propagation d'*ondes internes* : les ondes de gravité dans les fluides stratifiés et les ondes d'inertie dans les fluides en rotation. L'énergie de l'écoulement est alors répartie entre l'énergie cinétique contenue par les ondes, l'énergie cinétique contenue par les tourbillons, et l'énergie potentielle. On peut donc tenter de décomposer l'écoulement en ondes et en tourbillons et d'étudier les interactions entre ces structures [Lelong and Riley, 1991].

Les écoulements océaniques sont caractérisés pas une forte stratification (faible Froude), une forte rotation (faible Rossby), et une turbulence développée (grand Reynolds). Il est maintenant reconnu que les ondes internes jouent un rôle important dans ce régime, et donc dans les processus océaniques aux grandes échelles [MacKinnon et al., 2017], notamment dans le transport de masse, d'impulsion, et d'énergie. La compréhension de la dynamique de ces ondes est donc importante pour l'amélioration des modèles atmosphériques et océaniques [Gregg et al., 2018]. Bien qu'une équation d'état réaliste soit nécessaire pour modéliser l'atmosphère, les équations de Navier-Stokes dans l'approximation de Boussinesq sont souvent utilisée pour étudier les fluides stratifiés [Brethouwer et al., 2007]. Elles sont même utilisées dans certains modèles océaniques.

D'un point de vue théorique, la théorie de la turbulence d'ondes [Zakharov et al., 2012, Nazarenko, 2011] fournit un cadre d'étude des systèmes d'ondes dispersives, interagissant de façon faiblement non linéaire, faiblement impactées par la dissipation. Elle s'applique aux écoulements stratifiés [Caillol and Zeitlin, 2000] et/ou en rotation [Galtier, 2003, Medvedev and Zeitlin, 2007] dans la limite des faibles nombres de *Froude* et de *Rossby* et d'un grand nombre de *Reynolds*. Cette théorie permet, entre autres, d'écrire une équation cinétique qui décrit l'évolution de l'amplitude des modes (qui est l'analogue de l'équation de Boltzmann pour un système d'ondes, dans le domaine spectral). Elle fournit des prédictions concernant les spectres d'énergie, la direction de la cascade d'énergie, les fonctions de structures et des autres objets statistiques d'intérêt.

Bien que des avancées expérimentales importantes ont permis de vérifier certaines prédictions de la turbulence d'onde dans les écoulements stratifiés [Davis et al., 2020] et/ou en rotation [Monsalve et al., 2020], les expériences ne permettent pas d'observer les régimes d'intérêt géophysique, notamment à cause des effets de tailles finis et de la présence de parois. Les informations expérimentales sur les ondes internes viennent principalement des capteurs de vitesse et de températures immergées, utilisées lors des campagnes de mesures maritimes [Garrett and Munk, 1979, Polzin and Lvov, 2011]. Certaines mesures de spectres océaniques sont cohérentes avec la théorie de la turbulence d'ondes [Lvov et al., 2004]. Il existe également de nombreuses simulations de turbulence fortement stratifiée [Waite and Bartello, 2006, Pouquet et al., 2017, Maffoli, 2017]. La principale difficulté à observer le régime de turbulence d'ondes vient du fait qu'il faut un faible nombre de Froude, mais aussi un grand nombre de Reynolds de flottabilité [Brethouwer et al., 2007]. Si l'on souhaite étudier les propriétés statistiques des ondes, il faut également que les simulations soient suffisamment longues devant le temps typique d'interaction des ondes.

Le but de ce projet est d'effectuer des simulations numériques des équations de Navier-Stokes dans l'approximation de Boussinesq pour les fluides stratifiés et en rotation afin de vérifier les prédictions de la théorie de la turbulence d'ondes. Le projet se décompose en 3 parties :

## 1. Fluides stratifiés :

- Nous souhaitons réaliser des écoulements statistiquement stationnaires en utilisant la projection poloidale. Le but est de se rapprocher au mieux d'un système d'ondes de gravité en éliminant les modes vorticaux et les modes de cisaillement de la dynamique.
- Nous voulons également réaliser des simulations des équations sans projection poloïdale, à faible Froude. Ces simulations sont importantes pour observer le comportement des ondes dans un écoulement plus réaliste. En particulier, ces simulations nous aideront à comprendre l'effet des modes vorticaux et le rôle des ondes de gravité dans le mélange d'énergie.

## 2. Fluides en rotation :

- Nous souhaitons réaliser des écoulements statistiquement stationnaires en éliminant les colonnes de Taylor de la dynamique. Le but est de se rapprocher au mieux d'un

système d'ondes d'inertie en éliminant les colonnes de Taylor de la dynamique.

- Nous voulons également réaliser des simulations des équations complètes, à faible Rossby. Ces simulations sont nécessaires pour observer le comportement des ondes d'inertie dans un écoulement plus réaliste. En particulier, ces simulations nous aideront à comprendre l'effet des colonnes de Taylor et le rôle des ondes d'inertie dans le mélange d'énergie.

### 3. Fluides stratifiés et en rotation :

Dans un dernier temps, nous souhaiterions faire une simulation avec une forte rotation et une forte stratification pour comprendre le comportement des ondes gravito-inertielles. Ce cas est le plus pertinent pour les applications géophysiques.

Ce travail théorique et numérique vise à complémenter les expériences menées actuellement aux laboratoires FAST (Orsay) par l'équipe de Pierre Philippe Cortet, et au LEGI (Grenoble) par l'équipe de Nicolas Mordant. Il s'inscrit dans le cadre de notre participation au projet de la Simons Foundation sur la turbulence d'ondes.

Nous utiliserons le code pseudo-spectral fluidsim (<https://fluidsim.readthedocs.io/en/latest/index.html>) développé par Pierre Augier (LEGI) avec qui nous collaborons directement. Ce code a déjà utilisé sur différents clusters pour effectuer des calculs intensifs.

## 3 Méthode

### 3.1 Modèle théorique

Le modèle le plus simple pour décrire les fluides stratifiés en rotation consiste à considérer un fluide Newtonien dans l'approximation de Boussinesq, dans lequel le gradient de densité moyen est constant. Plus précisément, on considère que le champ de densité  $\rho$  peut se décomposer sous la forme suivante :

$$\rho = \rho_0 + \frac{d\bar{\rho}}{dz}(z - z_0) + \rho' \quad (1)$$

où  $\rho_0 + \frac{d\bar{\rho}}{dz}(z - z_0)$  représente la densité moyenne (profil linéaire), et  $\rho'$  est la perturbation de densité. Dans ce cas, les équations du mouvement du fluide s'écrivent

$$\nabla \cdot \mathbf{v} = 0, \quad (2)$$

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\nabla p - b \mathbf{e}_z - 2\Omega \times \mathbf{v} + \nu \Delta \mathbf{v}, \quad (3)$$

$$\frac{\partial b}{\partial t} + \mathbf{v} \cdot \nabla b = N^2 v_z + \kappa \Delta b, \quad (4)$$

où  $\mathbf{v} = (v_x, v_y, v_z)$  est le champ de vitesse,  $p$  est la pression totale (incluant la pression hydrostatique et la force centrifuge),

$$N \equiv \sqrt{-\frac{g}{\rho_0} \frac{d\bar{\rho}}{dz}}$$

est la fréquence de Brunt-Väisälä,  $g$  est l'accélération de pesanteur,

$$b \equiv -\frac{g\rho'}{\rho_0}$$

est la flottabilité,  $\Omega = \Omega \mathbf{e}_z$  est la vitesse angulaire du référentiel,  $\nu$  est la viscosité cinématique, et  $\kappa$  est la diffusivité. Ici, on considère le cas où l'axe de rotation est parallèle à l'axe de stratification. Nous utiliserons des conditions aux limites périodiques en espace.

On considère un écoulement de vitesse horizontale caractéristique  $U_h$  (par exemple la fluctuation typique de la vitesse horizontale) et de taille  $L_h$ . Lorsque la fréquence de Brunt-Väisälä devient grande devant la fréquence de retournement à grande échelle, i.e.  $N \gg U_h/L_h$ , la stratification influence fortement l'écoulement. L'effet de la stratification est donc contrôlé par le *nombre de Froude*

$$Fr = \frac{U_h}{NL_h}. \quad (5)$$

De façon similaire, l'effet de la rotation est quantifié par le *nombre de Rossby*

$$Ro = \frac{U_h}{2\Omega L_h}. \quad (6)$$

L'effet de la viscosité et de la diffusion sont quantifiés par les nombres de Reynolds et de Peclet :

$$Re = \frac{U_h L_h}{\nu} \quad \text{et} \quad Pe = \frac{U_h L_h}{\kappa}. \quad (7)$$

Nous considérerons un fluide pour lequel  $\nu = \kappa$ . Par conséquent, les paramètres de contrôle de l'écoulement sont  $Fr$ ,  $Ro$ , et  $Re$  uniquement. Il a été montré [Brethouwer et al., 2007] que le régime de turbulence fortement stratifiée est atteint quand le nombre de Reynolds de flottabilité

$$R_b = Re Fr^2 \quad (8)$$

est grand, et que le nombre de Froude est faible (typiquement  $< 0.02$ ). Ce régime est donc difficile à atteindre dans les expériences et dans les simulations numériques [Pouquet et al., 2017].

La relation de dispersion des ondes gravito-inertielles, dans la limite inviscide, est

$$\omega_{\mathbf{k}} = \pm \sqrt{N^2 \sin^2 \theta_{\mathbf{k}} + 4\Omega^2 \cos^2 \theta_{\mathbf{k}}}, \quad (9)$$

où  $\omega_{\mathbf{k}}$  représente la pulsation de l'onde ayant pour vecteur d'onde  $\mathbf{k} = (k_x, k_y, k_z)$  et  $\theta_{\mathbf{k}}$  est l'angle entre le vecteur d'onde et l'axe  $\mathbf{e}_z$ . La vitesse de phase de ces ondes est

$$\mathbf{c}_{p\mathbf{k}} = \frac{\omega_{\mathbf{k}}}{k} \mathbf{e}_{\mathbf{k}} = \pm \frac{\sqrt{N^2 \sin^2 \theta_{\mathbf{k}} + 4\Omega^2 \cos^2 \theta_{\mathbf{k}}}}{k} \mathbf{e}_{\mathbf{k}}, \quad (10)$$

tandis que leur vitesse de groupe est

$$\begin{aligned} \mathbf{c}_{g\mathbf{k}} = \nabla_{\mathbf{k}} \omega_{\mathbf{k}} &= \frac{\partial \omega_{\mathbf{k}}}{\partial k} \mathbf{e}_{\mathbf{k}} + \frac{1}{k} \frac{\partial \omega_{\mathbf{k}}}{\partial \theta_{\mathbf{k}}} \mathbf{e}_{\theta_{\mathbf{k}}} + \frac{1}{k \sin \theta_{\mathbf{k}}} \frac{\partial \omega_{\mathbf{k}}}{\partial \varphi_{\mathbf{k}}} \mathbf{e}_{\varphi_{\mathbf{k}}} \\ &= \pm \frac{(N^2 - 4\Omega^2) \sin \theta_{\mathbf{k}} \cos \theta_{\mathbf{k}}}{k \sqrt{N^2 \sin^2 \theta_{\mathbf{k}} + 4\Omega^2 \cos^2 \theta_{\mathbf{k}}}} \mathbf{e}_{\theta_{\mathbf{k}}}. \end{aligned} \quad (11)$$

où l'on a introduit le repère sphérique  $(\mathbf{e}_{\mathbf{k}}, \mathbf{e}_{\theta_{\mathbf{k}}}, \mathbf{e}_{\varphi_{\mathbf{k}}})$  associé au vecteur d'onde  $\mathbf{k}$ .

La transformée de Fourier du champ de vitesse peut alors se décomposer comme

$$\hat{\mathbf{v}}_{\mathbf{k}} = \hat{v}_{\mathbf{k}\theta} \mathbf{e}_{\theta_{\mathbf{k}}} + \hat{v}_{\mathbf{k}\varphi} \mathbf{e}_{\varphi_{\mathbf{k}}} \quad (12)$$

où  $\hat{v}_{\mathbf{k}\theta}$  est la composante poloïdale et  $\hat{v}_{\mathbf{k}\varphi}$  est la composante toroïdale. La composante selon  $\mathbf{e}_{\mathbf{k}}$  est nulle par incompressibilité. Pour un fluide purement stratifié, on montre que les ondes ne sont pas portées par la composante toroïdale. La projection du champ de vitesse sur sa composante poloïdale permet d'inhiber la formation de mode vorticaux, et ainsi favoriser l'observation d'un régime de turbulence d'ondes. Pour la même raison, il est possible de ne forcer que la partie poloïdale du champ de vitesse pour favoriser l'excitation des ondes.

Les modes de cisaillement correspondent aux modes de la vitesse horizontale pour lesquels  $k_x = k_y = 0$  (invariance de la vitesse horizontale selon l'horizontale), tandis que les colonnes de Taylor correspondent aux modes de la vitesse horizontale pour lesquels  $k_z = 0$  (invariance selon la verticale). Comme dans des expériences récentes [Monsalve et al., 2020], nous souhaitons dans un premier temps retirer ces modes de la dynamique pour simplifier le problème.

Deux possibilités sont envisagées pour favoriser l'observation de la gamme inertie dans ces régimes. La première est d'utiliser un rapport d'aspect plus faible pour limiter la taille de la grille de résolution. La seconde est d'utiliser une hyperviscosité et une hyperdiffusivité

$$-\nu_{\alpha}(-\Delta)^{\alpha/2} \mathbf{u} \quad \text{et} \quad -\kappa_4(-\Delta)^{\alpha/2} b. \quad (13)$$

dans les équations (2-4) avec, typiquement,  $\alpha = 4$ . Cela a pour effet de repousser la dissipation à des plus petites échelles, pour une dissipation totale égale. L'hyperviscosité sera ajustée de façon à stabiliser les simulations quelle que soit la résolution choisie.

Nous effectuerons d'abord des simulations à des résolutions modérées (typiquement  $160 \times 640 \times 640$  ou  $240 \times 960 \times 960$ ) depuis l'état de repos pour atteindre un régime statistiquement stationnaire. Nous raffinerons ensuite progressivement la résolution de ces simulations. Cela permettra de limiter le temps de calcul durant le régime transitoire.

### 3.2 Modèle numérique

Les équations (2-4) seront résolues avec le code pseudo-spectral **Fluidsim** [Mohanam et al., 2019b] du projet Fliddyn [Augier et al., 2019]. Il est principalement écrit en Python. La pression est traitée de façon implicite par projection du champ de vitesse sur l'espace des champs à divergence nulle. Nous utilisons la méthode de Runge-Kutta d'ordre 4 pour le schéma temporel, mais le code permet aussi d'utiliser la méthode de Runge-Kutta d'ordre 2. Le code a été testé sur la machine Jean-Zay lors d'un accès préparatoire (projet AP012A13209).

Les tâches coûteuses en calculs sont effectuées par des fonctions optimisées et compilées, ce qui permet au code d'avoir des performances comparables à celles d'un code Fortran ou C++ [Augier et al., 2021]. En particulier, il utilise le module Fluidfft [Mohanam et al., 2019a] qui gère l'utilisation des librairies standards de transformées de Fourier rapides (fftw1d, fftw3d, pfft, et p3dfft) déjà installées dans le cluster, ou dans le répertoire de travail de l'utilisateur. Fluidfft est écrit en C++, Cython et Python. Le code utilise Pythran, un compilateur Python qui produit des exécutables via C++11. Les fichiers de données sont sauvés au format hdf5. La librairie hdf5/1.12.0-mpi est utilisée pour gérer l'enregistrement parallélisé des données de sortie.

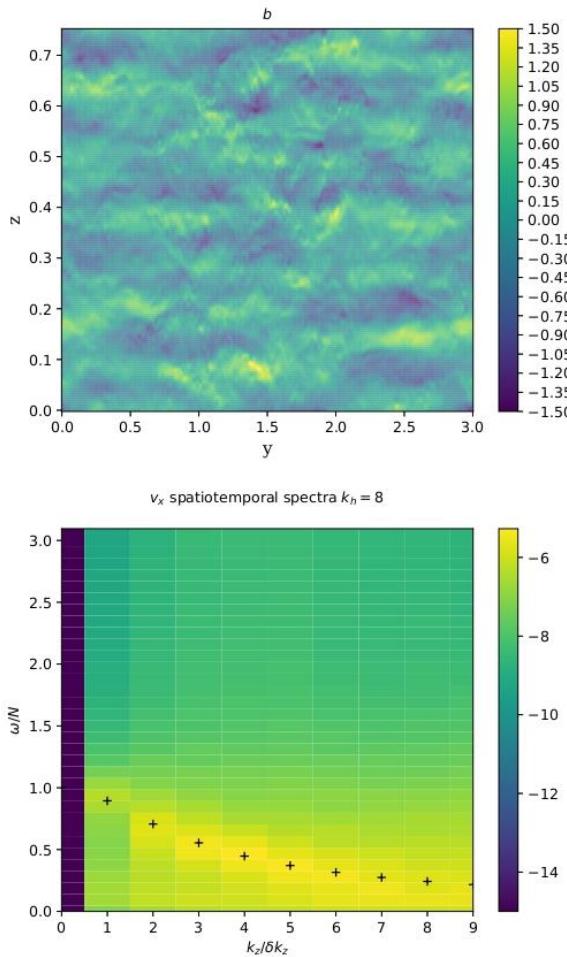


FIGURE 1 – Coupe verticale du champ de flottabilité (haut) et spectre spatiotemporel  $k_h - \omega$  (bas) pour une simulation avec  $N = 40$ ,  $R_b = 40$  et la projection poloidale réalisée sur Jean-Zay. Les croix représentent la relation de dispersion théorique pour les ondes de gravité.

Beaucoup d'outils sont déjà implémentés dans le code : les projections sur la composante polaire, les forçages des différents champs (différentes composantes de la vitesse et flottabilité), la suppression des modes de cisaillement, l'utilisation d'une hyperviscosité, et le traitement de données. En particulier, le code permet d'enregister les spectres spatiaux, temporels et spatio-

temporels (figure 1). Fluidsim a déjà utilisé sur le cluster Occigen (CINES) pour effectuer des calculs intensifs avec une bonne parallélisation jusqu'à  $\sim 10^4$  processeurs pour une grille de  $1152^3$ . Il a été utilisé pour effectuer des simulations de turbulence stratifiée à 2D sur ce cluster [Calpe Linares, 2020]. Il est aussi utilisé sur le mésocentre CRIMSON de l'Observatoire de la Côte d'Azur et sur le cluster du LEGI.

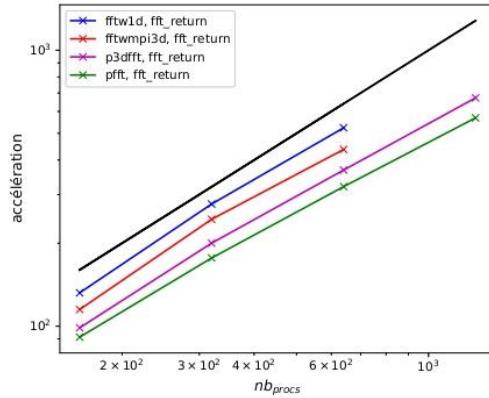


FIGURE 2 – Facteur d'accélération en fonction du nombre de processeurs pour réaliser les transformées de Fourier rapides par Fluidfft . Le test a été réalisé sur la machine Jean-Zay pour une grille de  $1280 \times 1280 \times 640$  points et 4 librairies : fftw1d, fftw3d, pf3dfft, et p3dfft.

La figure 2 montre que les opérations de transformées de Fourier rapides sont bien parallélisées, avec une loi d'échelle quasi-optimale jusqu'à  $640 - 1280$  processeurs pour une grille de taille  $1280 \times 1280 \times 640$ . La librairie fftw1d est la plus efficace, mais est limité en termes de nombre de processeurs utilisables, car elle utilise une décomposition en tranche. Les librairies avec une décomposition en crayon (pf3dfft et p3dfft) sont plus lentes à nombre de processeurs égal, mais permettent d'utiliser plus de processeurs. La librairie p3dfft est plus rapide que pf3dfft, et est environ 1.5 fois plus lente que fftw1d. La figure 3 montre que le solveur utilisé pour la résolution des équations (ns3d.strat) présente des propriétés de parallélisation très similaires à celles de Fluidfft.

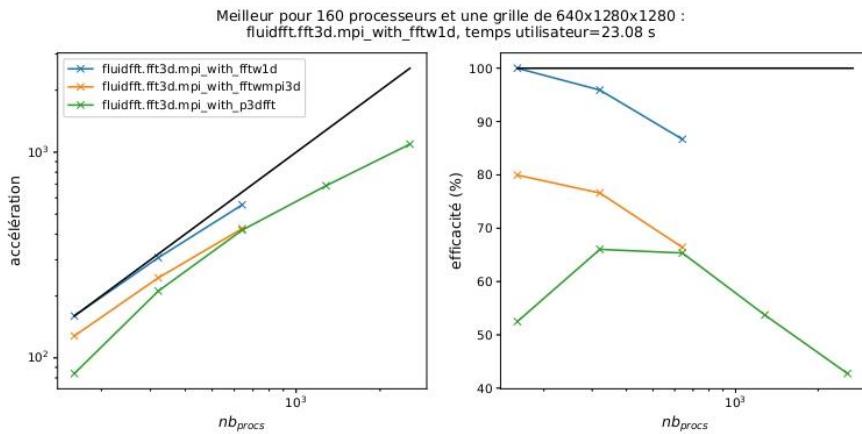


FIGURE 3 – Propriétés de parallélisation du solveur ns3d.strat de Fluidsim. Le test a été réalisé sur la machine Jean-Zay pour une grille de  $1280 \times 1280 \times 640$  points et 100 pas de temps. À gauche : accélération relative du temps d'exécution  $t$  par rapport au cas de référence (1 noeud = 40 processeurs),  $t/t_0$ . A droite : efficacité de la parallélisation  $(tN_p)/(tN_p)_0$ , où  $N_p$  est le nombre de processeurs utilisés.

### 3.3 Justification de l'usage des ressources sur les machine demandées

Ce projet fait l'objet d'une demande d'attribution d'heures sur la machine Jean-Zay (projet DARI AP012A13209). Afin de préparer au mieux l'utilisation du cluster national, nous souhaiterions disposer de ressources supplémentaires pour effectuer des simulations en faisant varier les paramètres de contrôle (les nombres sans dimensions  $Fr$ ,  $Ro$ , et  $Re$ ). Ces simulations auront pour but de calibrer les détails du forçage afin de favoriser l'observation du régime de turbulence d'ondes, de faire une première étude paramétrique du problème, et de générer des champs dans les états statistiquement stationnaires. Nous pourrons ensuite augmenter la résolution de ces champs pour relancer les simulations sur Jean-Zay. Nous avons déjà réalisé plusieurs simulations à une résolution de  $640 \times 640 \times 160$  sur notre mésocentre.

Pour adresser les différents volets de ce projet, nous comptons réaliser les simulations suivantes :

#### 1. Fluides stratifiés :

- Nous souhaitons réaliser des écoulements statistiquement stationnaires en utilisant la projection poloïdale. Nous voulons obtenir une vingtaine de simulations dans le cas purement stratifié avec une résolution de  $640 \times 640 \times 160$ . Le coup de chaque simulation est estimé à  $10 k$  heures CPU.
- Nous voulons également réaliser des simulations des équations sans projection poloïdale, à faible Froude. Ces simulations sont importantes pour observer le comportement des ondes dans un écoulement plus réaliste. Le coup de chaque simulation est estimé à  $20 k$  heures CPU.

**Nous estimons le coût de cette partie à  $600 k$  heures CPU.**

#### 2. Fluides en rotation :

- Nous souhaitons réaliser des écoulements statistiquement stationnaires en éliminant les colonnes de Taylor de la dynamique. Nous voulons réaliser une vingtaine de simulations dans le cas purement rotatif avec une résolution de  $640 \times 640 \times 160$ . Le coup de chaque simulation est estimé à  $8 k$  heures CPU.
- Nous voulons également réaliser des simulations des équations complètes, à faible Rossby. Le coup de chaque simulation est estimé à  $16 k$  heures CPU.

**Nous estimons le coût de cette partie à  $480 k$  heures CPU.**

#### 3. Dans un dernier temps, nous souhaiterions faire varier la rotation et la stratification, toujours à résolution modérée.

**Nous estimons le coût de cette partie à  $600 k$  heures CPU.**

Nous souhaitons réaliser ces simulations conjointement sur la machine Azzurra et sur notre mésocentre (CRIMSON de l'Observatoire de la Côte d'Azur). Par conséquent, **nous demandons  $500 k$  heures CPU sur la machine Azzura.**

## 4 Plan de gestion de données

Pendant la phase de production, un volume de données sera généré, estimé à 5 To. Ces données correspondent aux champs de vitesse et de flottabilité, ainsi que les spectres spatiotemporels qui seront analysés en post-traitement. Une partie du post-traitement sera faite directement sur les machines de calcul, y compris la visualisation (si possible), et d'autres analyses numériques seront effectués localement dans notre mésocentre. Une fois que les données auront été traitées, ce volume pourra être réduit pour ne garder que les données les plus essentielles.

## Références

- [Augier et al., 2021] Augier, P., Bolz-Tereick, C. F., Guelton, S., and Mohanan, A. V. (2021). Reducing the ecological impact of computing through education and Python compilers. *Nature Astronomy*, 5(4) :334–335.
- [Augier et al., 2019] Augier, P., Mohanan, A. V., and Bonamy, C. (2019). FluidDyn : A python open-source framework for research and teaching in fluid dynamics by simulations, experiments and data processing. *Journal of Open Research Software*, 7.

- [Brethouwer et al., 2007] Brethouwer, G., Billant, P., Lindborg, E., and Chomaz, J.-M. (2007). Scaling analysis and simulation of strongly stratified turbulent flows. *Journal of Fluid Mechanics*, 585 :343–368.
- [Caillol and Zeitlin, 2000] Caillol, P. and Zeitlin, V. (2000). Kinetic equations and stationary energy spectra of weakly nonlinear internal gravity waves. *Dynamics of Atmospheres and Oceans*, 32(2) :81–112.
- [Calpe Linares, 2020] Calpe Linares, M. (2020). *Numerical study of 2D stratified turbulence forced by internal gravity waves*. Theses, Université Grenoble Alpes [2020-....].
- [Davis et al., 2020] Davis, G., Jamin, T., Deleuze, J., Joubaud, S., and Dauxois, T. (2020). Succession of resonances to achieve internal wave turbulence. *Phys. Rev. Lett.*, 124 :204502.
- [Galtier, 2003] Galtier, S. (2003). Weak inertial-wave turbulence theory. *Phys. Rev. E*, 68 :015301.
- [Garrett and Munk, 1979] Garrett, C. and Munk, W. (1979). Internal waves in the ocean. *Annual Review of Fluid Mechanics*, 11(1) :339–369.
- [Gregg et al., 2018] Gregg, M., D'Asaro, E., Riley, J., and Kunze, E. (2018). Mixing efficiency in the ocean. *Annual Review of Marine Science*, 10(1) :443–473. PMID : 28934598.
- [Lelong and Riley, 1991] Lelong, M. P. and Riley, J. J. (1991). Internal wave—vortical mode interactions in strongly stratified flows. *Journal of Fluid Mechanics*, 232 :1–19.
- [Lvov et al., 2004] Lvov, Y. V., Polzin, K. L., and Tabak, E. G. (2004). Energy spectra of the ocean's internal wave field : Theory and observations. *Phys. Rev. Lett.*, 92 :128501.
- [MacKinnon et al., 2017] MacKinnon, J. A., Zhao, Z., Whalen, C. B., Waterhouse, A. F., Trossman, D. S., Sun, O. M., Laurent, L. C. S., Simmons, H. L., Polzin, K., Pinkel, R., Pickering, A., Norton, N. J., Nash, J. D., Musgrave, R., Merchant, L. M., Melet, A. V., Mater, B., Legg, S., Large, W. G., Kunze, E., Klymak, J. M., Jochum, M., Jayne, S. R., Hallberg, R. W., Griffies, S. M., Diggs, S., Danabasoglu, G., Chassignet, E. P., Buijsman, M. C., Bryan, F. O., Briegleb, B. P., Barna, A., Arbic, B. K., Ansong, J. K., and Alford, M. H. (2017). Climate process team on internal wave–driven ocean mixing. *Bulletin of the American Meteorological Society*, 98(11) :2429–2454.
- [Maffioli, 2017] Maffioli, A. (2017). Vertical spectra of stratified turbulence at large horizontal scales. *Phys. Rev. Fluids*, 2 :104802.
- [Medvedev and Zeitlin, 2007] Medvedev, S. and Zeitlin, V. (2007). Turbulence of near-inertial waves in the continuously stratified fluid. *Physics Letters A*, 371(3) :221–227.
- [Mohanam et al., 2019a] Mohanam, A. V., Bonamy, C., and Augier, P. (2019a). FluidFFT : Common API (c++ and python) for fast fourier transform HPC libraries. *Journal of Open Research Software*, 7.
- [Mohanam et al., 2019b] Mohanam, A. V., Bonamy, C., Linares, M. C., and Augier, P. (2019b). FluidSim : Modular, Object-Oriented Python Package for High-Performance CFD Simulations. *Journal of Open Research Software*, 7.
- [Monsalve et al., 2020] Monsalve, E., Brunet, M., Gallet, B., and Cortet, P.-P. (2020). Quantitative experimental observation of weak inertial-wave turbulence. *Phys. Rev. Lett.*, 125 :254502.
- [Nazarenko, 2011] Nazarenko, S. (2011). *Wave Turbulence*. Lecture Notes in Physics. Springer Berlin Heidelberg.
- [Polzin and Lvov, 2011] Polzin, K. L. and Lvov, Y. V. (2011). Toward regional characterizations of the oceanic internal wavefield. *Reviews of Geophysics*, 49(4).
- [Pouquet et al., 2017] Pouquet, A., Marino, R., Mininni, P. D., and Rosenberg, D. (2017). Dual constant-flux energy cascades to both large scales and small scales. *Physics of Fluids*, 29(11) :111108.
- [Waite and Bartello, 2006] Waite, M. L. and Bartello, P. (2006). Stratified turbulence generated by internal gravity waves. *Journal of Fluid Mechanics*, 546 :313–339.
- [Zakharov et al., 2012] Zakharov, V., L'vov, V., and Falkovich, G. (2012). *Kolmogorov Spectra of Turbulence I : Wave Turbulence*. Springer Series in Nonlinear Dynamics. Springer Berlin Heidelberg.

## Modelling one-dimensionnal optical turbulence

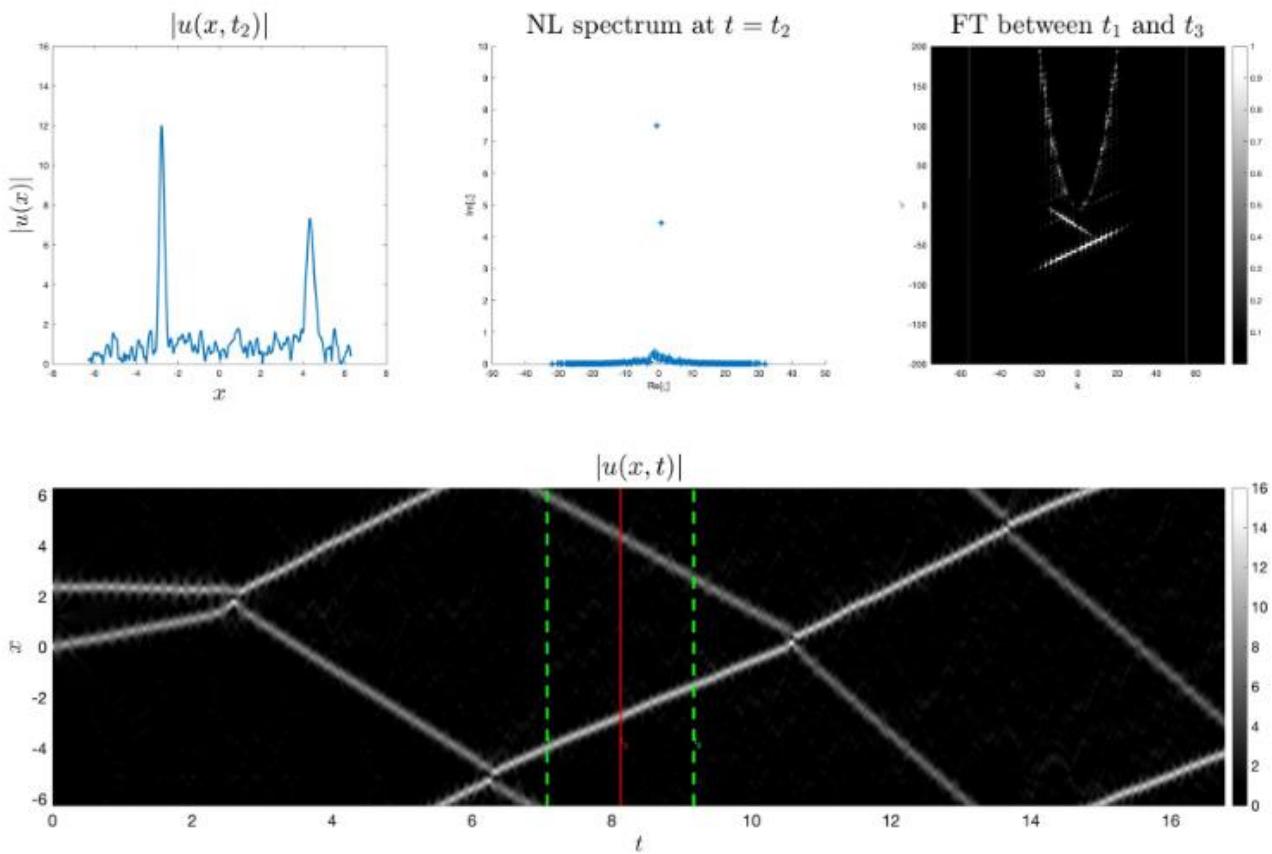
**Responsable scientifique :** Sergey Nazarenko

**Laboratoire :** INPHYNI

### Description :

I study a given system which I let evolve in time and which I characterize once I have all its evolution. This system is governed by the Non-Linear Schrödinger equation modified and I use tools such as Direct Scattering Transform and omega k plots to identify the solitons in the system. My goal is to understand the soliton turbulence, which is the final state where only one big soliton remains, after it has gained all the wave action of the system. This evolution is possible thanks to the modification to NLSE which makes the collisions not elastic anymore.

I join a figure which illustrates the simulations I am running and how I study them:



My system depends on a parameter which controls the importance of the modification in the equation. The results obtained are very sensitive to a modification of this parameter. That is why I would like to start a lot of simulations in parallel with slightly different parameters.



---

## ***Spontaneous stochasticity in surface quasi-geostrophic turbulence***

**Responsable scientifique :** Jérémie Bec

**Laboratoire :** Inria

### **Description :**

Developed turbulence usually features three distinctive anomalies: the persistence of a finite energy dissipation in the limit of vanishing viscosity (the dissipative anomaly), intermittency (anomalous scaling), and the explosive separation of fluid particles. These three anomalies point towards three spontaneously broken symmetries in the limit of vanishing viscosity: time reversal, scale invariance and uniqueness of the Lagrangian flow, a fascinating phenomenon that is known as spontaneous stochasticity.

The introduction of a small noise, sent to zero together with a small-scale regularisation leads to the explosive separation between fluid particles, no matter how close they initially are. This phenomenon, known as Richardson's super diffusion, has been substantially observed for Navier-Stokes turbulence. Moreover, for simplified models (e.g. random velocities as in the Kraichnan model), this procedure has allowed to build well-defined probability measures over Lagrangian trajectories, even if the notion of trajectory is broken in a deterministic sense. On the other hand, the Eulerian flow is defined as the deterministic mapping between initial and final fields over prescribed time intervals. However, it is conjectured since the seminal work of Lorenz in the late sixties, that multiscale fluid flows could have a drastically unpredictable behaviour. Spontaneous stochasticity provides a modern perception of this idea : the dynamical evolution of singular velocity fields could be a mathematically ill-posed problem, therefore leading to finite-time separation of initially undistinguishable fields. This suggests that turbulent velocity fields should perhaps only be considered in terms of probability measures.

To date, the complex interplay between time irreversibility, intermittency and spontaneous stochasticity remains a challenge for the three-dimensional Navier-Stokes equations. The purpose of this project is to discuss the case of Surface Quasi Geostrophic (SQG) flows, which describe the two-dimensional active transport of a temperature scalar field in a strongly stratified and rotating environment. Besides its relevance to geophysics, SQG bears formal resemblance with various flows of interest, including the 3D Navier-Stokes in the direct cascade regime. Preliminary numerics (see Valade, Thalabard & Bec, Ann. H. Poincaré 2023, arXiv.2210.12366) reveal that SQG flow share with Navier-Stokes the same three broken symmetries. Moreover, the mechanisms bridging

---

Eulerian and Lagrangian descriptions also share features with Kraichnan flows : there is evidence that fluid particles separate in an explosive manner. This spontaneously stochastic behaviour, evidenced both for forward and backward separations, connects to the anomalous dissipation of the advecting fields. However, our preliminary study displays non-universal statistics in the double limit of vanishing viscosity and perturbation, telling us that SQG stochasticity might be present in a tempered version only.

The objective of this project is to shed light on this question by performing unprecedented large-size simulations of the randomly forced SQG equations in a turbulent statistical steady-state. This will first allow us to provide a state-of-the-art description of turbulent SQG statistics, varying the viscosity (or equivalently the Reynolds number). As a second step, we will study the impact of various types of perturbations on these solutions in order to assess whether they display a strong or tempered spontaneously stochastic behaviour.

---

**Optique**

**EUR Spectrum**

---

## ***Comportement spatial de la propagation d'une pompe hautement multimode dans des amplificateurs et lasers à fibre***

**Responsable scientifique :** Valérie Doya

**Laboratoire :** INPHYNI

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Depuis octobre 2022, Lucas Zanaglia est en thèse sur le sujet 'Condensation et thermalisation de la lumière dans des fibres optiques multimodes chaotiques'.

Un résumé a été proposée à la conférence Journées Nationales de l'Optique Guidée (juillet 2023).

### **Description :**

Les lasers à haute puissance sont aujourd'hui utilisés dans une grande variété d'applications industrielles et environnementales. Afin d'améliorer leur robustesse et leur efficacité, ces systèmes sont maintenant majoritairement fibrés, c'est-à-dire que l'ensemble des éléments qui composent le laser est contenu dans une fibre optique. Le milieu amplificateur est obtenu par un dopage aux ions terres rares de la fibre et les miroirs (appelé réseaux de Bragg) qui constituent la cavité sont en général photo-inscrits dans le cœur de la fibre. Afin d'optimiser le schéma de pompage du laser, des réflecteurs sont en général employés pour recycler la pompe dans la cavité laser. L'un des enjeux actuels est de pouvoir les fabriquer directement dans la gaine de la fibre optique. L'aspect fortement multimode de la pompe se propageant dans la gaine rend l'interaction avec un réseau de Bragg non triviale. Les relations couramment employées dans le cadre d'un faisceau monomode ne sont plus valables et une connaissance fine du contenu modale de la pompe devient alors indispensable.

L'objectif de ce projet est donc de calculer numériquement l'évolution du signal de pompe au sein d'une fibre optique amplificatrice afin de pouvoir optimiser le design d'un réflecteur de pompe fibré. La géométrie du système et le schéma numérique employé nécessite actuellement un temps de calcul de plusieurs semaines empêchant toute étude paramétrique. Parallèlement à un travail d'optimisation du code, nous sollicitons les ressources dédiées du centre de calcul pour diminuer notablement le temps de calcul.

Ce projet s'inscrit dans le cadre d'une collaboration actuelle avec le Centre Optique Photonique et Laser (Université Laval, Québec) et TeraXion (entreprise québécoise spécialisée dans la fabrication de composants et lasers fibrés). Nous avons aussi amorcé une collaboration avec OptiWave, une entreprise canadienne qui édite un logiciel de

---

calcul optique dédié à l'interaction laser/réseaux de Bragg.

---

## ***Light scattering in disordered systems with correlations***

**Responsable scientifique :** Geoffroy Aubry

**Laboratoire :** INPHYNI

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Un article accepté :

Bandgap fluctuations and robustness in two-dimensional hyperuniform dielectric materials. Luis Froufe, Geoffroy Aubry, Frank Scheffold, and Sofia Magkiriadou. <https://doi.org/10.1364/OE.484232>

Thèse de Mattis Reisner : Études expérimentales de la multifractalité et des transitions de phase topologiques dans des réseaux de résonateurs micro-ondes. Soutenue le 01-03-2023 à l'Université Côte d'Azur, dans le cadre de l'École doctorale Sciences fondamentales et appliquées.

### **Description :**

In this project, I plan to realize different kinds of numerical simulations in order to better understand the influence of correlated disordered dielectrics on the propagation of electromagnetic waves (light) in these systems.

Disordered dielectrics like clouds, foams, powders, etc. appear white because light propagate into it not in a straight line, like in vacuum, but get scattered, ie., changes its propagation direction much more than once in the medium. This explains the reason why the sun position can't be seen on a cloudy day, although the sun light still arrive to our eyes. In a cloud, there is, a priori, no correlation between the scatterers (i.e. the water droplets) positions: this medium is referred to as an "uncorrelated disordered system". In a "correlated disordered system", the positions of the scatterers are not completely independent from each other. The limit of this class of system are photonic crystals, where all the positions of the scatterers are linked to each other, and no disorder is present anymore.

Using correlated disorder, it is possible to tune light transport in order to observe a wide variety of regimes: from transparency, to light hindered to enter the medium (like in a photonic bandgap) including light diffusion or even light stuck in the medium, the so called Anderson localized regime.

In my previous work done during my postdocs in Konstanz (Germany) and Fribourg (Switzerland), I already did some experimental and numerical work on these kinds of systems (see selected references at the end of this paragraph). As a new permanent researcher in the Institut de Physique de Nice, I plan to continue these studies that are of

---

great interest for applications such as supercontinuum fiber lasers and for data processing using optical modules based on 2D silicon photonic crystal technology. Technological applications for 3D PBG bandgap materials offer great potential in lasing, wave-guiding, light-harvesting and even optical computers.

Selected references:

- Aubry, G. J.; Schertel, L.; Chen, M.; Weyer, H.; Aegeerter, C. M.; Polarz, S.; Cölfen, H. & Maret, G. Resonant transport and near-field effects in photonic glasses  
Phys. Rev. A 96, 043871 (2017) <http://dx.doi.org/10.1103/PhysRevA.96.043871>
- Aubry, G. J.; Froufe-Pérez, L. S.; Kuhl, U.; Legrand, O.; Scheffold, F. & Mortessagne, F. Experimental Tuning of Transport Regimes in Hyperuniform Disordered Photonic Materials Phys. Rev. Lett. 125, 127402 (2020) <http://dx.doi.org/10.1103/PhysRevLett.125.127402>
- Aeby, S.; Aubry, G. J.; Muller, N. & Scheffold, F. Scattering from controlled defects in woodpile photonic crystals Advanced Optical Materials, 9, 2001699 (2021)  
<http://dx.doi.org/10.1002/adom.202001699>
- Yazhgur, P.; Aubry, G. J.; Froufe, L. S. & Scheffold, F. Light scattering from colloidal aggregates on a hierarchy of length scales Opt. Express, 29, 14367 (2021)  
<http://dx.doi.org/10.1364/OE.418735>

---

# Dynamique des fluides

**EUR Spectrum**

---

## Dynamique de polymères en turbulence et turbulence élastique

**Responsable scientifique :** Dario Vincenzi

**Laboratoire :** LJAD

### Résultats obtenus grâce aux moyens informatiques attribués :

Nous avons étudié l'effet d'un gradient de vitesse non-gaussien sur la déformation des polymères. Nous avons également examiné la relaxation de la distribution de l'extension des polymères vers sa forme d'équilibre dans un écoulement turbulent.

Ces résultats ont fait l'objet d'un papier :

Picardo JR, Plan ELCVM, Vincenzi D. Polymers in turbulence: stretching statistics and the role of extreme strain rate fluctuations. *Journal of Fluid Mechanics*. 2023;969:A24. doi:10.1017/jfm.2023.524  
et disponible sur HAL : <https://cnrs.hal.science/hal-03919174v1/document>

Thèse :

Dynamique de particules allongées dans des écoulements tourbillonnaires et turbulence élastique dans les solutions de polymères. Sumithra Reddy YERASI, Laboratoire Jean Alexandre Dieudonné, Université Côte d'Azur. Soutenue le 2 mai 2023.

Des travaux sur la simulation numérique de la turbulence élastique sont en cours. Un papier sur ce sujet sera soumis dans les mois à venir.

### Description :

Le projet est composé de deux volets. Le premier porte sur la dynamique lagrangienne de polymères élastiques dans un écoulement turbulent isotrope et sur la transition coil-stretch. Nous étudierons l'effet des interactions hydrodynamiques et de volume exclu. Le calcul numérique consiste en la simulation d'équations différentielles stochastiques et en l'analyse statistique des solutions.

Le deuxième volet porte sur la turbulence élastique, une régime chaotique que l'on observe dans les solutions de polymères visqueuses. Nous effectuerons des simulations numériques pseudo-spectrales du modèle FENE-P, qui décrit l'évolution couplée du tenseur de conformation des polymères et du champ de vitesse (donnée par les équations de Navier-Stokes). Les simulations seront d'abord effectuées en deux dimensions dans un carré avec conditions de bord périodiques. Le projet à plus long terme est de réaliser des simulations en trois dimensions.



---

## Géophysique interne

**EUR Spectrum**

---

## **Dynamics of large earthquakes**

**Responsable scientifique :** Jean-Paul Ampuero

**Laboratoire :** Géoazur

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- M. T. Herrera, J. P. Ampuero, J. G. F. Crempien (2024a), "A proxy implementation of thermal pressurization for earthquake cycle modeling on rate-and-state faults", manuscript in review in Geophysical Journal International, preprint: <https://doi.org/10.31223/X5S377>.
- M. T. Herrera, J. P. Ampuero, J. G. F. Crempien (2024b), "Conditions leading to low apparent fault strength under dynamic weakening", to be submitted to Journal of Geophysical Research.
- E. Sopaci and J. P. Ampuero (2024), "Triggering of very shallow earthquakes by surface mass removal processes - case study of the 2019 Mw4.9 Le Teil, France earthquake", European Geophysical Union General Assembly 2024
- J. Premus and J. P. Ampuero (2024), "Dynamic earthquake source inversion with CAN priors, with application to the 2023 Mw 7.8 Kahramanmaraş, Turkey earthquake", European Geophysical Union General Assembly 2024

### **Description :**

Earthquakes are a fundamental component of plate tectonics, one of the main processes that shape the surface of our planet, and one of the most damaging natural hazards. Elucidating the physics of earthquakes remains a grand challenge. Thanks to recent advances in data gathering, computational seismology, laboratory rock mechanics and theoretical fault mechanics, the earthquake research community has made major advances towards answering first-order questions such as: How do earthquakes start? What determines the wide spectrum of fault slip behavior? What controls the speed and complexity of rupture propagation? How do seismic and aseismic slip interact? However, many key aspects of these questions still remain unsolved.

My team's research at Geoazur aims at answering these open questions by developing a fundamental understanding of the mechanical behavior of tectonic faults, across temporal and spatial scales spanning rupture dynamics, earthquake cycles, and the evolution of fault zone structure. Our vision is to achieve this goal by combining state-of-the-art theoretical modeling, high-performance computing and novel geophysical observation techniques. In this computational project, we aim at advancing our

---

understanding of earthquake cycles and rupture dynamics on very large and segmented faults.

Our main recent achievement in earthquake theory was to establish an “equation of motion” for large earthquakes, whose rupture spans the whole seismogenic depth and then propagates horizontally. The distinctive feature of these ruptures is their large aspect ratios. We found through computational experiments that such elongated ruptures have radically different behavior than predicted by previous theories, which were limited to ruptures in 2D and to almost-circular ruptures in 3D. We thus developed a new theory of the evolution of elongated ruptures (Weng and Ampuero, *J. Geophys. Res.*, 2019), leading to an equation of motion and a “rupture potential”, a quantity that encapsulates the most important factors controlling rupture arrest. If we are able to estimate this rupture potential along a fault, we can map the fault areas that are most prone to generate a large earthquake. Motivated by results of our high-performance computer simulation experiments, we further developed the implications of the new theory to understand what controls the rupture speed of earthquakes (Weng and Ampuero, *Nature Geoscience*, 2020), a parameter that strongly affects their damaging potential. This work reconciled earthquake theory with the puzzling observation of earthquakes propagating at speeds that are forbidden in the previous 2D theory of dynamic fracture.

In this project, we propose to extend the theory of elongated ruptures in complementary directions that will allow us to extract more implications for real earthquakes:

- We will study the effects of heterogeneities of stress and strength on the dynamics of elongated ruptures. Our current theory accounts already for lateral heterogeneities, but we still need to consistently include depth-dependent heterogeneities. With this capability, we will study how systematic and stochastic heterogeneities affect the statistical properties of earthquakes, such as scaling laws and the variability of source time functions.
- We will extend the theory to account for variable rupture width in the theory of elongated ruptures. In long asperities surrounded by creep, dynamic rupture can penetrate inside the creeping region and thus the rupture width is not fixed. The idea we will pursue is to use 1D fracture mechanics problems on depth cross-sections to solve for the rupture width, then use that laterally variable width in the equation of motion of elongated ruptures. We will validate this approach by comparison to results of 3D dynamic rupture simulations.

- We will conduct 2.5D earthquake cycle simulations on long faults with heterogeneous frictional properties. We will use these simulations as a testbed to validate the developments above. We will further use the modified theory to gain insights on the predictability of earthquake rupture. In particular, we will search for practical proxies of the spatio-temporal evolution of the rupture potential which could be used to forecast the location, size and timing of the next large earthquake in our (idealized) earthquake cycle models. Those proxies are quantities such as the geodetic coupling rates.
- Large faults are divided into segments that can break separately or together. Because the number of broken segments controls the final earthquake size, it is essential to understand the mechanical interactions between fault segments. Can we relate the probability of multi-segment ruptures to observable fault properties? Can we distinguish between temporary and permanent barriers? Can we anticipate the size of the next large earthquake on a given fault? We will conduct a parametric study based on 2.5D earthquake cycle modeling to quantify the efficiency of fault barriers to stop ruptures, as a function of their length and frictional strength.

The 3D dynamic rupture simulations will use the SPECFEM3D software, inwhich our team leads the development of earthquake dynamics simulation capabilities. The 2.5D earthquake cycle simulations will use the SEM2DPACK software, fully developed by our team.

#### References:

- H. Weng and J. P. Ampuero (2020), Continuum of earthquake rupture speeds enabled by oblique slip, *Nature Geo.*, 13, 817–821, doi:10.1038/s41561-020-00654-4
- H. Weng and J. P. Ampuero (2019), The dynamics of elongated earthquake ruptures, *J. Geophys. Res.*, 124 (8), 8584-8610, doi: 10.1029/2019JB017684

---

## ***Continental Transpression: Deciphering Deformation Digitally (CT3D)***

**Responsable scientifique :** Guillaume Duclaux

**Laboratoire :** Géoazur

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Les calculs réalisés sur Azzurra ont permis le développement de nouvelles conditions aux limites dans le code open-source pTatin3D :

Jourdon, A., May, D. A., and Gabriel, A. A.: Generalization of the Nitsche method to apply oblique boundary conditions in regional geodynamic models, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-14864, <https://doi.org/10.5194/egusphere-egu23-14864>, 2023.

Ces travaux ont eu un impact fort pour l'évolution du code et ont permis à Anthony Jourdon de déposer un projet d'ERC Starting Grant retenu pour la phase d'auditions (audition programmée le 22 juin 2023).

Des travaux de modélisation 3D de la dynamique des systèmes obliques liés à la thèse de Josselin Gremmel sont en cours.

### **Description :**

Ce projet propose d'étudier le partitionnement de la déformation dans les systèmes continentaux en transpression à l'aide de modèles numériques thermomécaniques en trois dimensions (3D). Cette approche computationnelle nous permettra 1) de mieux comprendre la manière dont les fabriques des roches métamorphiques apparaissent et sont reliées les unes aux autres dans l'espace et dans le temps, et 2) de déterminer le niveau de couplage entre les différents niveaux structuraux de la croûte continentale dans les systèmes transpressifs. Dans les domaines orogéniques transpressifs anciens que nous connaissons, les roches ont subi une histoire de déformation polyphasée, résultant de la compétition des forces tectoniques et de volume évoluant au cours du temps, qui conduit au développement de fabriques finies complexes et variées à toutes les échelles. Dans les niveaux supérieurs de la croûte continentale on observe des systèmes de failles qui forment des structures en fleur, de zones de relais transpressifs et transtensifs, ainsi que des bassins étroits dont la géométrie semble fortement influencée par les fabriques pré-existantes. Dans la croûte intermédiaire à profonde la déformation est marquée par le développement de grandes zones de cisaillement ductiles verticales limitant des dômes marqués par un aplatissement sub-horizontal et un étirement sub-

---

parallèle aux zones de cisaillement. Des plis d'axes obliques à la direction de raccourcissement principal se forment par fluage longitudinal synchrone de la transpression. Les questions que nous souhaitons adresser peuvent-être résumés comme suit : Comment ces fabriques dans les roches métamorphiques sont-elles reliées les unes aux autres dans l'espace et dans le temps ? Quel niveau de couplage existe entre les différents niveaux structuraux dans les systèmes orogéniques transpressifs ?

L'évolution de la nature et de l'orientation des structures au sein de la croûte sera modélisée à partir d'un code thermomécanique 3D (*pTatin 3D* : *May et al., 2014*<sup>34</sup> ; *Duclaux et al., 2019*<sup>35</sup>) et analysée en post-processing à partir de scripts Python que nous avons développés. Ces travaux répondent également à un enjeu sociétal majeur et visent à améliorer notre compréhension du couplage mécanique entre les niveaux profonds des grandes zones de cisaillement ductiles et les systèmes de failles décrochantes et décro-chauchantes sismogènes observées en surface dans les zones de convergence obliques.

---

<sup>34</sup> *Duclaux, G., Huismans, R. S., & May, D. A. (2020). Rotation, narrowing, and preferential reactivation of brittle structures during oblique rifting. Earth and Planetary Science Letters, 531, 115952.*

<sup>35</sup> *May, D. A., Brown, J., & Le Pourhiet, L. (2014, November). pTatin3D: High-performance methods for long-term lithospheric dynamics. In SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (pp. 274-284). IEEE.*

---

## **WIND (Waveform Inversion of Node Data)**

**Responsable scientifique** : Stéphane Operto

**Laboratoire** : Geoazur

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- Wu, Yanqi & Aghamiry, Hossein & Operto, Stéphane & Ma, Jianwei. (2023). Frequency-Domain Wave Simulation Using Physics-Informed Neural Networks (PINNs) with Free Surface Boundary Condition. *84th EAGE Annual Conference & Exhibition*, Jun 2023, Volume 2023, p.1 – 5. <https://doi.org/10.3997/2214-4609.202310403>
- Wu, Y., Aghamiry, H. S., Operto, S., & Ma, J. (2023). Helmholtz-equation solution in nonsmooth media by a physics-informed neural network incorporating quadratic terms and a perfectly matching layer condition. *Geophysics*, 88(4), T185-T202.

### **Description :**

#### Présentation générale du projet

Ce projet porte sur le développement d'une méthode d'imagerie sismique par inversion des formes d'ondes complètes (FWI: Full Waveform Inversion). La FWI consiste à reconstruire les propriétés mécaniques du sous-sol gouvernant la propagation des ondes en résolvant un problème d'optimisation sous contrainte par des approches locales (méthodes de gradient). Le problème direct résout l'équation d'onde, une équation aux dérivés partielles (EDP) linéaire dont la solution et le second membre sont respectivement le champ d'onde et la source. Les observables sont les enregistrements des champs d'onde aux capteurs. La FWI minimise une distance entre ces observables et leur analogues simulés numériquement pour mettre à jour les paramètres contenus dans les coefficients de l'EDP.

Les applications sismiques concernent (1) l'imagerie des bassins sédimentaires pour l'exploration pétrolière (projet WIND financé par un consortium industriel) et (2) l'imagerie de la croûte terrestre des zones de subduction (chantiers Geoazur). Ces deux applications sont effectuées à partir d'acquisitions de fond de mer mises en oeuvre avec des sources artificielles et des câbles sous-marins (OBC) ou des stations sous-marines multi-composantes (OBSs). En sismologie (sources passives), l'imagerie de la croûte et du manteau supérieur des Alpes à partir de téléseismes (projet LisAlps financé par l'ANR).

---

Nous implémentons la FWI dans les domaines temps-espace (codes GeoInv3D\_fwk et LITHOS) et fréquence-espace (code FFWI). Le problème direct y est respectivement décrit par un problème d'évolution avec conditions initiales et un problème de conditions aux limites nécessitant la résolution de systèmes d'équations linéaires creux à seconds membres multiples. Dans ce dernier cas, nous utilisons le solveur multifrontal MUMPS et un solveur hybride direct/itératif préconditionné par la méthode de décomposition en domaine ORAS (Optimized Restricted Additive Schwarz). Suivant les codes et les applications, nous utilisons des schémas de discrétisation fondés sur des différences finies, des éléments finis sur maillage tétraédrique et des éléments spectraux.

Les codes GeoInv3D\_fwk, LITHOS et FFWI sont parallélisés avec MPI et OpenMP et ont été exécutés en production sur le mésocentre CRIMSON, sur Occigen et sur Jean Zay. Ils sont basiquement écrits en Fortran 90 avec des ingérences de C++. Nos travaux récents portent sur une reformulation de la FWI visant à étendre son régime linéaire vers une méthode de Lagrangien augmenté et l'interfaçage de régularisations hybrides pour réduire son espace nul. Ces travaux ont été implémentés sous Matlab et doivent être transférés dans les trois codes pour tester les concepts sur des cas d'étude de grande dimension.

### Demande sur Azzura

La présente demande portant spécifiquement sur l'accès à Azzura concerne deux aspects:

- Le premier concerne le portage progressif du code GeoInv3D\_fwk sur GPU.
- Le deuxième concerne une application d'intelligence artificielle fondée sur des "physics-informed neural networks". Dans le cadre de la thèse de Yanqi WU à Geoazur, nous souhaitons commencer à évaluer si la résolution de l'équation d'onde (le problème direct) et la mise à jour des paramètres constitutifs du milieu (problème inverse) peuvent être effectués avec des approches de deep learning où la fonction cout est respectivement formée par la norme L2 de l'erreur sur l'équation d'onde (problème direct) et une fonction de pénalité ou de Lagrangien augmenté combinant la norme L2 de l'écart entre les données observées et simulées et la norme L2 de l'erreur sur l'équation d'onde. Des tests préliminaires ont été effectués sur les CPUs d'un serveur local de Geoazur. Néanmoins, nous avons maintenant besoin d'accéder à des GPUs pour traiter des applications de dimension plus représentatives.

References sur PINN (Physics-Informed Neural Networks with applications to wave equation solution and seismic waveform inversion)

- 
- M. Raissi, P. Perdikaris and G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*, 378, 686--707, 2019.
- C. Song and T. Alkhalifah and U. B. Waheed, Solving the frequency-domain acoustic VTI wave equation using physics-informed neural networks, *Geophysical Journal International*, 225(2), 846--859, 2021.
- C. Song and T. Alkhalifah, Wavefield reconstruction inversion via machine learned functions, *SEG Technical Program Expanded Abstracts 2020*, 1710-1714, 2020.

---

## ***Study of the rheological controls on vertical sinking of subducting slabs through numerical modeling***

**Responsable scientifique :** Karin Sigloch

**Laboratoire :** Geoazur

### **Description :**

This is a numerical investigation of the modes in which the solid earth recycles material and heat between its surface and its deep interior, on geological time scales. Specifically, we study whether thermo-mechanical modeling of the subduction of tectonic plates into the earth's mantle can match recent geometric constraints on the deformed plate geometries in the Earth's mantle, as observed by geophysical imaging. We use the fluid-dynamics finite-element code Fluidity [Davies et al., 2011], which solves for the conservation of mass and momentum in 2-D geometries, with non-linear rheologies suitable for the study of rock deformation. The code uses an adaptive remeshing technique to tackle the strong viscosity contrasts that characterize geodynamical flows (up to 4 orders of magnitudes). Tectonic plates form in ocean basins from upwelling mantle rock, and spend several tens of millions of years cooling at the surface. Then these ocean plates dive back into the earth's mantle at subduction zones and slowly sink towards the core, cooling the ambient mantle in the process. This material cycle, which makes the earth the only habitable planet in the solar system, is thought to be driven by the "slab pull" force, i.e., the excess density of plate margins that have entered a subduction zone relatively recently, and which keep pulling down after them the tectonic surface plate they are still attached to [Forsyth and Uyeda, 1975].

Although geodynamic conceptual and numerical models often show these slabs as remaining plate-like, thin sheets to great mantle depths, geophysical imaging (seismic 3-D tomography) is increasingly clear that from relatively shallow depths of only a few hundred kilometers, the slabs appear thickened to several times lithospheric thickness. This is a fundamental geometry difference that implies completely different rheological, advection and cooling characteristics for the earth's mantle. Presumably the thickening occurs through slab folding, as seen in Figs. 2&3, although these spatial details are currently not resolvable by geophysical imaging. In order to constrain the plausible range of rheologies and deformation styles, we add a number of observations from plate kinematic and from

the land geology of accretionary mountain belts, which form the surface record of subduction.

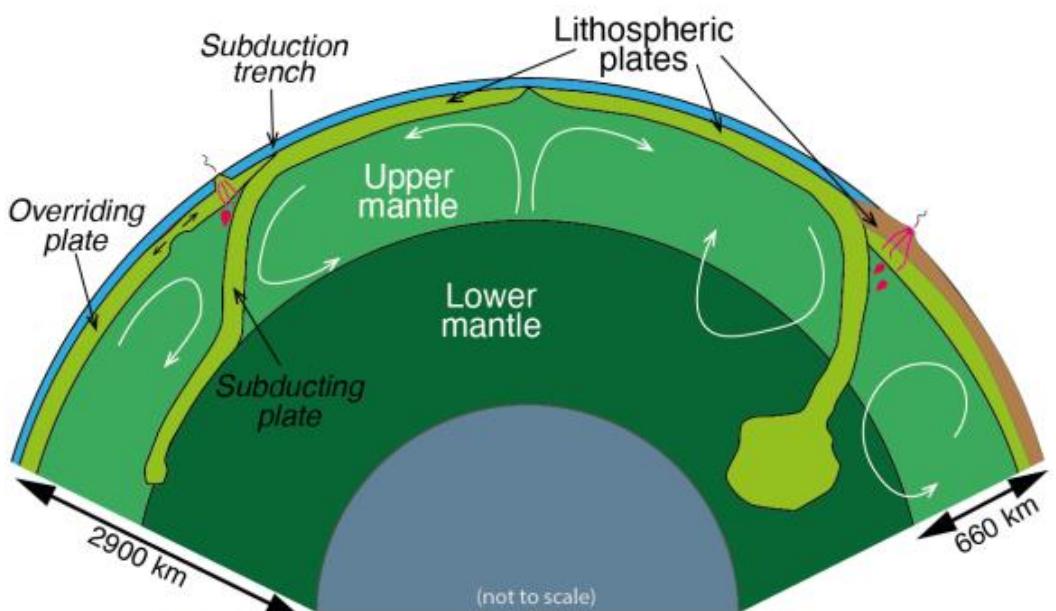


Figure 1 : Schematic view of two subduction zones.

Over the past decade, such “tomo-tectonic” analysis has revealed that many past subduction zones were quasi-stationary in an absolute reference frame [Sigloch and Mihalynuk, 2013; Mohammadzaheri et al., 2021]. The tomographic images also showed that stationary trenches are associated with wide “slab walls” in the lower mantle (400–700-km wide), a special-case geometry that rather strongly points to slab thickening (folding?) at the mantle transition zone (Fig. 2). In contrast to these observations, state-of-the-art numerical models of subduction zone dynamics

suggest that non-stationary trenches (slab-rollback and subduction-trench retreat) are the most common dynamical response of slab sinking in the mantle (fig. 3, [Garel et al., 2014]). Hence, a strong conceptual discrepancy has recently opened between observational constraints and current numerical geodynamic models, which has not been addressed quantitatively. Here we want to close this gap and numerically explore the mantle and slab properties that would result in thickened slabs under stationary trenches, while also reproducing realistically fast subduction rates.

This work is conducted as part of the PI’s ERC Starting Grant “DEEPTIME” (639003). In May 2021, I moved from the University of Oxford to a Directeur de Recherche position at the CNRS and Geoazur. My ERC grant has moved with me to the CNRS, as has the postdoctoral associate who has been working on the sub-project presented here, Dr. Nestor Cerpa. With this request for computing time on Azzurra, we are hoping to replace the institutional

computing resources that were available to us in Oxford, in order to successfully conclude and publish our investigation.

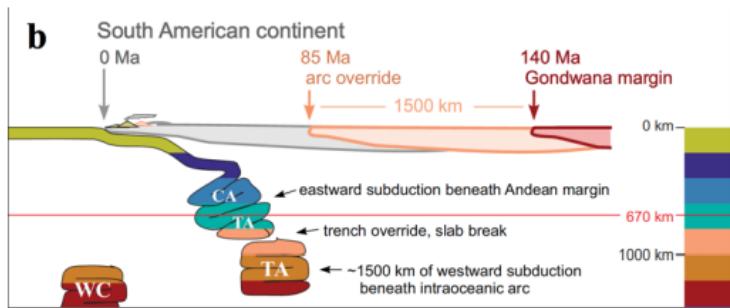


Figure 2 : Tomo-tectonic interpretation of the evolution of the South American subduction zone [from Mohammadzaheri et al. 2021].

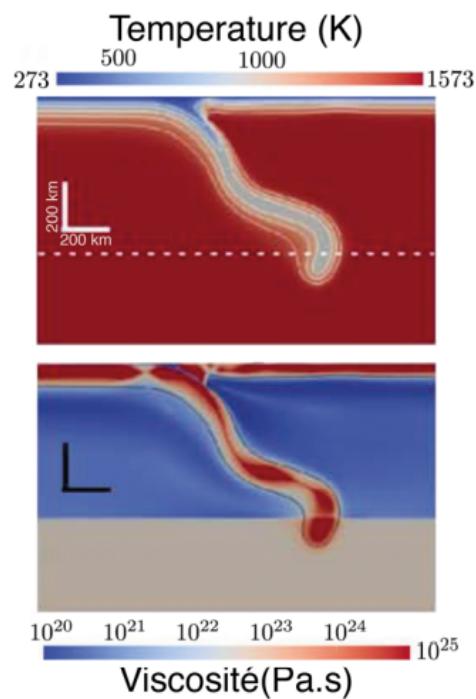


Figure 3 : Two-dimensional numerical model of a subduction zone [from Garel et al., 2014].

## References

- Davies, D. R., Wilson, C. R., & Kramer, S. C. (2011). Fluidity: A fully unstructured anisotropic adaptive mesh computational modeling framework for geodynamics. *Geochemistry, Geophysics, Geosystems*, 12(6).
- Forsyth, D, and Uyeda, S. (1975). On the relative importance of the driving forces of plate motion. *Geophysical Journal International* 43, no. 1: 163-200.
- Garel, F., Goes, S., Davies, D. R., Davies, J. H., Kramer, S. C., & Wilson, C. R. (2014). Interaction of subducted slabs with the mantle transition - zone: A regime diagram from 2 - D thermo - mechanical models with a mobile trench and an overriding plate. *Geochemistry, Geophysics, Geosystems*, 15(5), 1739-1765.
- Mohammadzaheri, A., Sigloch, K., Hosseini, K., & Mihalynuk, M. G. (2021). Subducted Lithosphere under South America from Multi - frequency P - wave Tomography. *Journal of Geophysical Research: Solid Earth*, 126, e2020JB020704.
- Sigloch, K., & Mihalynuk, M. G. (2013). Intra-oceanic subduction shaped the assembly of Cordilleran North America. *Nature*, 496(7443), 50-56

---

## **Imaging high-frequency radiation of large earthquakes**

**Responsable scientifique :** Anthony Sladen

**Laboratoire :** Géoazur

Projet en partenariat avec l'[Universidad de Chile \(FCSM\)](#).

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Carolina Paz Gutiérrez Morales. Inversión bayesiana en altas frecuencias de la distribución de deslizamiento del terremoto de Tocopilla 2007, Mw 7.7. Rapport de master « Magíster en Ciencias, Mención Geofísica », Universidad de Chile (2022).

<https://repositorio.uchile.cl/handle/2250/187604>

### **Description :**

*A project led by Carolina Gutierrez, research engineer at University of Chile (Santiago), and supervised by [Sergio Ruiz](#), Professor at University of Chile (Santiago), and [Anthony Sladen](#) CNRS researcher at University Côte d'Azur.*

Characterizing the high-frequency distribution of earthquakes is critical for earthquake hazard assessment and rapid response

The high-frequency radiation generated during subduction earthquakes is a process that is still poorly understood in seismology. Because this type of emission causes the most damage in the near-field it is essential to understand its behavior to assess the seismic hazard in a specific region correctly. Multiple methodologies have been developed to model the strong-ground motion of large magnitude earthquakes, including the use of stochastic Green functions. This approach successfully simulates high frequencies ( $> 1 \text{ Hz}$ ) in the borehole and on the ground surface incorporating soil amplification transfer functions to account for the amplifying effect of the soil. Such methodology opens the way for improved and rapid assessment of earthquake damage, and a better understanding of the underlying physical processes.

### Our modeling approach

Most studies of fault inversion use low-frequency data, leaving out the short-period emission of the rupture process. In this project, we use high-frequency acceleration records in the near-field, filtered between 0.05 and 20 Hz, to invert the rupture model for two large events, the 2007 Tocopilla earthquake (Mw 7.7) and the 2017 Valparaíso earthquake (Mw 6.9). Due to the high-frequency content of the data, it is difficult to simulate this type of motion with a deterministic approach, hence we use a stochastic simulation technique proposed by Otarola and Ruiz (2016),

---

based on Boore's method from 1983. This method generates synthetic accelerograms for P, SV, and SH waves from a finite fault model where each sub-fault is a point source. For each point source, a window of Gaussian white noise is modulated following Brune's w source spectrum. Then in the time domain, they are rotated and added to obtain the time-series acceleration in the EW, NS, and UD components.

### Solving the inverse problem

The inverse problem of determining the rupture model given the acceleration data is nonlinear and ill-conditioned, hence we use a Bayesian methodology to be able to explore the full parameter space and obtain the ensemble of probable models. To sample the space efficiently we use a TMCMC sampler.

### Results obtained for the 2007 Tocopilla earthquake (Mw 7.7)

For the Tocopilla earthquake, we model four acceleration records from stations in boreholes, using a parametrization of two circular asperities for the source model. With the methodology described above, we infer the most probable rupture model given the observed data. The methodology is first validated with synthetic tests and then applied to the observed data of the Tocopilla earthquake. The results indicate that our methodology allows us to properly estimate a simple rupture consisting of two circular asperities with a similar radius of 40-50 km. The center of the first asperity was imposed at the hypocenter location, but the location of the second one is properly located beneath the Mejillones peninsula (Figure 1). The predicted accelerations generated from this model fit the data well, especially in the period range between 0.2 and 1 s (Figure 2).

### Results obtained for the 2017 Valparaíso earthquake (Mw 6.9)

For the Valparaíso earthquake, we model records from four stations in boreholes and three stations on the ground surface, where we use soil amplification transfer functions based on the local velocity profile. We use a parametrization of a single circular asperity for the source model and infer the parameters that define the most probable rupture model given the data. The current results show a rupture model consisting of a circular asperity with a 14.5 km, fixed at the hypocenter and with a maximum slip of 2.1 m (Figure 3). The predicted accelerations generated from this model show a better fit in the EW and UD components, reflected in the goodness of fit in Figure 4. For this event, we have more stations, which are located at different distances from the source. Achieving a good fit for the near-field and distant stations is challenging, considering the same high-frequency decay behavior. This can be partially solved by including 2 different gamma parameters to model the decay of high frequencies in Brune's spectral model. We are currently working on incorporating new approaches to improve the results presented.

What has been achieved thanks to Azzurra and the way forward

All the results shown here, as well as the synthetic tests, were obtained using Azzurra's resources. These have been essential in developing and completing my Master's thesis "High-frequency bayesian inversion of the slip distribution of the 2007 Tocopilla earthquake, Mw 7.7" at the University of Chile, and have allowed us to expand the analysis and apply this methodology to the Valparaíso earthquake.

An important aspect of the methodology of this project is the high computational cost of the forward model. For the two modeled earthquakes, simulating the acceleration data takes between 1-3 seconds. During the inverse problem, the sampler runs  $\sim 10^4$  parallel chains to obtain the ensemble of models from the posterior distribution. At the end of the inversion, the forward model has been executed approximately  $3 \times 10^6$  times, resulting in a total execution time of  $\sim 2\text{-}3$  days. Nonetheless, this amount of chains is not enough to correctly represent the posterior distribution of rupture models. We need to solve the inverse problem with a larger amount of chains, closer to  $10^6$  to achieve a more robust solution.

Our objective is to publish the results of this project in an article and present it at Seismology conferences. We have submitted an abstract to present this project at the Latin American and Caribbean Seismological Commission IV Assembly, which will take place in Quito, Ecuador, from October 3rd to October 5th, 2022.

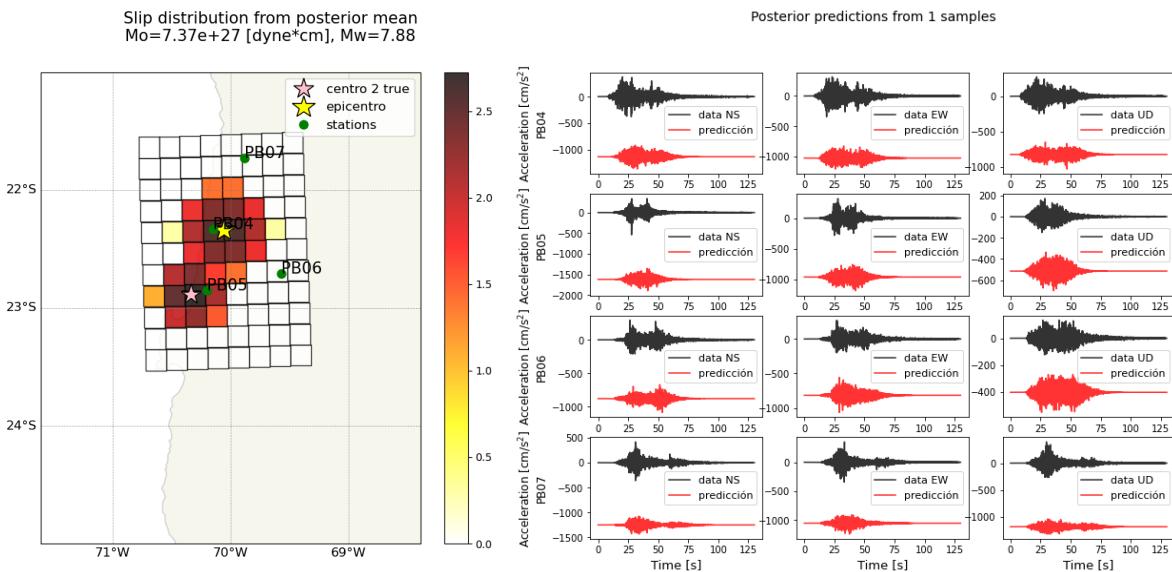


Figure 1: a) Estimated rupture model for the Tocopilla earthquake. The color scale shows the slip distribution for the two circular asperities. The north asperity is fixed in the hypocenter represented by the yellow star. The pink star is the estimated location of the center of the second asperity, right above the Mejillones peninsula. The estimated radii are 49 km and 38 km, and the maximum slips are 3 m and 2.6 m for the north and south asperity respectively. We estimate a rupture velocity of 3.4 km/s, a stress drop of 30 bars, and a gamma (high-frequency spectral decay in Brune's model) of 1.55. b) Comparison between the Tocopilla acceleration records in black and the acceleration predictions from the estimated rupture model in red, for each station and component.

### Goodness of Fit

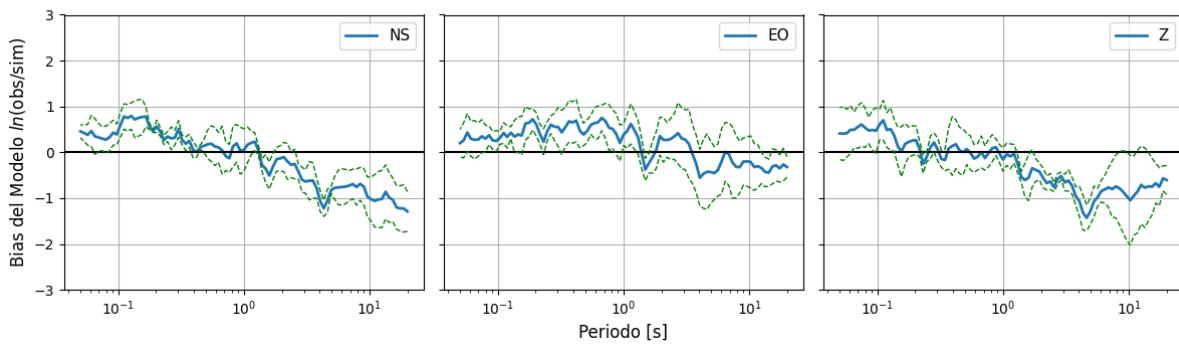


Figure 2: Goodness of fit (GOF) between the observed and predicted acceleration response spectra of 5% damping for the Tocopilla earthquake. The blue line shows the bias, and the region between the dashed lines shows the standard error.

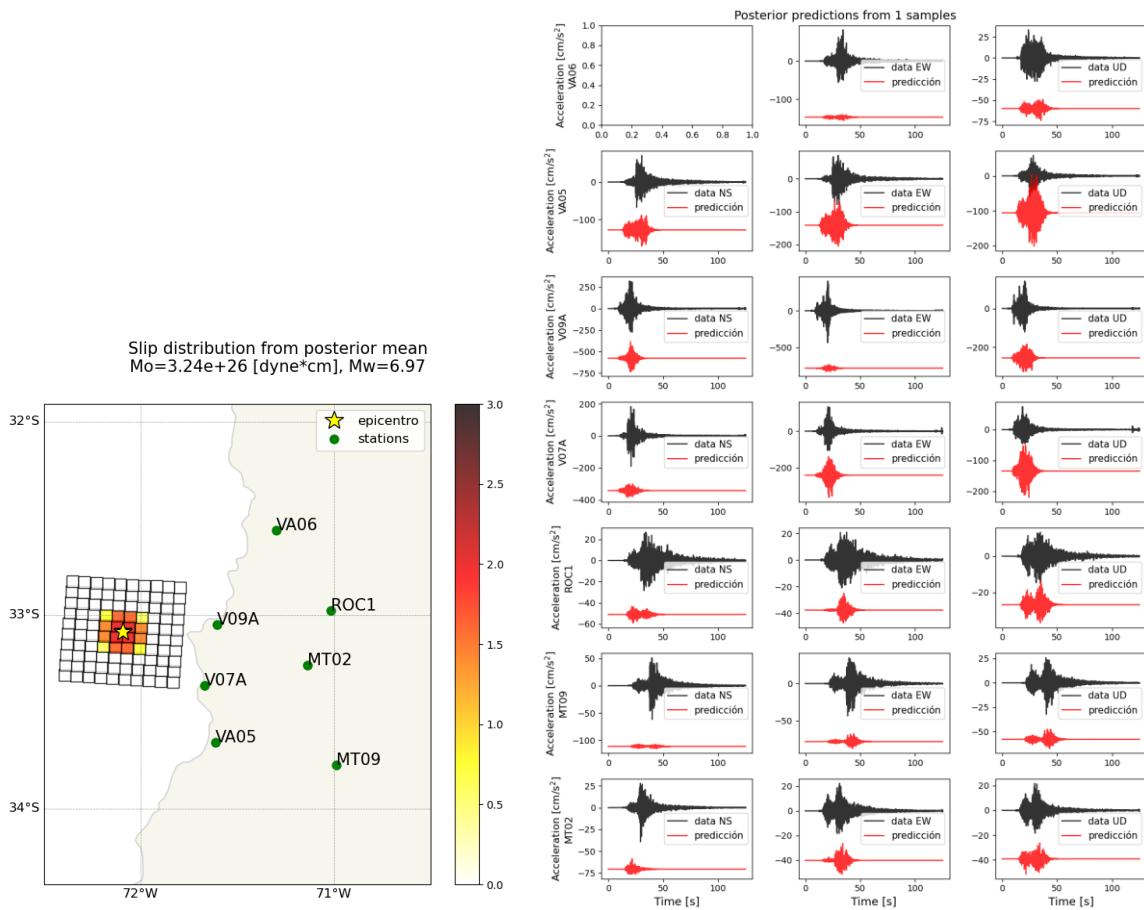


Figure 3: a) Estimated rupture model for the Valparaíso earthquake. The color scale shows the slip distribution for the circular asperity, which is fixed at the hypocenter of the event, represented by a yellow star. The estimated radius is 14.5 km, and the maximum slip is 2.1 m. We estimate a rupture velocity of 1.63 km/s, a stress drop of 31 bars, and a gamma (high-frequency spectral decay in Brune's model) of 1.85 for near stations (V09A, V07A, and VA05) and of 2.0 for far-away stations (VA06, ROC1, MT02, and MT09). b) Comparison between the Valparaíso acceleration records in black and the acceleration predictions from the estimated rupture model in red, for each station and component. The NS component is not well recorded for the VA06 station and it is not considered in the analysis.

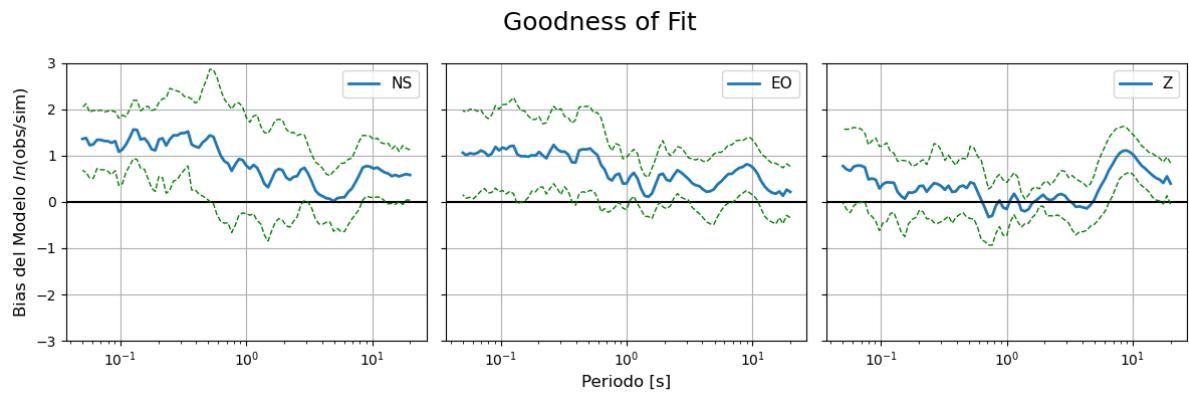


Figure 4: Goodness of fit (GOF) between the observed and predicted acceleration response spectra of 5% damping for the Valparaíso earthquake. The blue line shows the bias, and the region between the dashed lines shows the standard error.

---

## **Mathématiques, informatique**

**EUR DS4H, Spectrum**

---

## **BioEvolNet**

**Responsable scientifique :** Jean-Paul Comet

**Laboratoire :** I3S

### **Description :**

Dans le cadre de la thèse de Romain Michelucci (voir résumé ci-dessous), nous sommes amenés à développer de nouvelles approches de type MCTS (reinforcement learning) pour la recherche de paramètres de modèles de systèmes biologiques.

Nous avons initié une collaboration avec le LAMSADE (Paris Dauphine, Tristan Cazenave) et les calculs à faire devraient être publiés dans le cadre de cette collaboration.

**Objectives:** In molecular biology, a large range of interactions between genes, proteins, and molecular entities has been widely studied, and large maps have been built. Nevertheless, the knowledge of inter-actions is not sufficient to explain the dynamics of the system and to hope being able to control it : the dynamics is also governed by parameters explaining the relative strengths between regulators. The goal of this PhD is to develop methodologies able to automate the identification of parameter values associated to a gene regulatory network, from the knowledge of gene regulatory graphs and from some behavioural properties.

The research will focus on formal methods and artificial intelligence (bio-inspired methods and reinforcement learning) in order to automate the identification of parameters of formal models that can best explain the observations. Current collaborations with biologists will give rise to fruitful applications by, for example, designing a temporised model of the coupling between circadian clock, cellular cycle and metabolism regulation.

**Scientific Context :** The group « Bioinfo formelle » of I3S laboratory has proposed different formal frameworks to study the behaviour of gene regulatory networks. A regulatory network, modelled as a graph, defines the static interactions of a biological system : in the gene regulatory networks modelling setting, each interaction abstracts the individual influence of a gene x on the expression of another gene y. The dynamic of the network is governed by numerous unknown parameters that we want to identify. René Thomas has proposed a discrete modelling framework that allows approximations of

---

sigmoid functions (that usually represent the evolution of target gene expression level according to the level of the regulatory entity) by step functions. Using this modelling framework, the exploration of the qualitative dynamics of the system can be described as path exploration in a finite state space and formal methods from computer science are useful [1].

In order to take into account temporal information, which plays crucial role in a wide range of biological systems, we also have developed, for several years, a hybrid modelling framework [2,3], which measures the time to go from a state to another one. This framework consists in a particular class of hybrid automata, but the crucial problem remains in the determination of accurate values for all numerous parameters. In order to address this question, we developed a "weakest precondition calculus" inspired by Hoare's Logic (initially dedicated to imperative programs) which leads to constraints on parameters which has to be fulfilled in order to make the model's dynamics to be consistent with the observations [4, 5, 6]. However, the exploitation of the constraints generated by the Hoare Logics is not so easy : classical constraints solvers are not able to extract solutions.

References :

- [1] G. Bernot, J.-P. Comet, A. Richard, J. Guespin, Application of Formal Methods to Biological Regulatory Networks : Extending Thomas' Asynchronous Logical Approach with Temporal Logic. *J.T.B.* 229(3) :339-347, 2004.
- [2] J. Behaegel, J.-P. Comet, G. Bernot, E. Cornillon, F. Delaunay, A hybrid model of cell cycle in mammals, 6th International Conference on Computational Systems-Biology and Bioinformatics (CSBio'2015), Bangkok, Thailand. November 22-25, 2015.
- [3] J. Behaegel, J.-P. Comet, G. Bernot, E. Cornillon and F. Delaunay, A hybrid model of cell cycle in mammals. *J. of Bioinformatics and Computational Biology.* 14(1) :1640001 [17 pp.], 2016.
- [4] G. Bernot ; J.-P. Comet ; Z. Khalis ; A. Richard ; O.F. Roux, A Genetically Modified Hoare Logic. *Theoretical Computer Science.* 765 :145-157, 2019. DOI 10.1016/j.tcs.2018.02.003.
- [5] J. Behaegel, J.-P. Comet and M. Folschette, Constraint Identification Using Modified Hoare Logic on Hybrid Models of Gene Networks, International Symposium on Temporal Representation and Reasoning (TIME), Mons, Belgium. 16-18 October 2017. pp. 5 :1-5 :21
- [6] J. Behaegel, J.-P. Comet and M. Pelleau. Identification of Dynamic Parameters for Gene Networks, the 30th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Volos, Greece. November 5-7, 2018.

---

## **Real-time neuromorphic semantic scene segmentation**

**Responsable scientifique :** Jean Martinet

**Laboratoire :** I3S

### **Description :**

Image segmentation is a computer vision task in the field of AI in which specific regions of an image must be labelled according to their semantic contents. The objective of this PhD project is to design a real-time semantic scene segmentation model using a neuromorphic approach. More specifically, the project will combine a new kind of vision sensor, event cameras, with a brain-inspired machine learning paradigm, spiking neural networks. These two complementary neuromorphic models are both based on binary discrete events, sparse and asynchronous, allowing for a low latency, thus making them suitable for real-time applications. Moreover, recently developed neuromorphic hardware allow for low-power implementations, which extends the use of these models to embedded devices such as self-driving vehicles or robots.

Studies in biological vision have often been a great source of inspiration for design of computer vision algorithms. Thus, several successful methods were designed with varying functions, such as: object detection, image classification and image segmentation. Image segmentation is a computer vision task in which specific regions of an image are labelled according to its semantic contents. More specifically, the goal of semantic image segmentation is to label each pixel of an image with a corresponding class of what is being represented.

Within this context, several architectures have been proposed, such as: DeconvNet [1], Fully Convolutional Networks [2], DeepLab [3] and the autoencoder SegNet [4, 5]. The latter consists of an encoder and a decoder based on Fully CNNs (by removing all the fully connected layers) which stores maxpooling indices instead of feature maps during encoding process in order to be used during upsampling at the decoder. This mechanism made method outperform all others in terms of accuracy and memory space occupation.

However, since this method uses Artificial Neural Networks model that leads to huge computational costs with high memory consumption, it makes its application limited to systems with large power budget without generalization to embedded ones equipped with power-restricted components like: IoT devices, Automotive, AR/VR and

Industrial automation.

To overcome this challenge, many recent work have turned to Spiking Neural Networks, a brain-inspired paradigm that represents a special class of artificial neural networks, where neurons communicate by sequences of spikes [6]. Contrary to deep convolutional networks, spiking neurons do not fire at each propagation cycle, but rather fire only when their activation level (or membrane potential, an intrinsic quality of the neuron related to its membrane electrical charge) reaches a specific threshold value. When a neuron fires, it generates a non-binary signal that travels to other neurons, which in turn increases their potentials. The activation level either increases with incoming spikes, or decays over time. Neuromorphic hardware implementing SNN can be built with CMOS technology, and typically uses low power (under the threshold voltage), enabling to reduce energy dissipation by several orders of magnitude, compared to standard digital architectures [7].

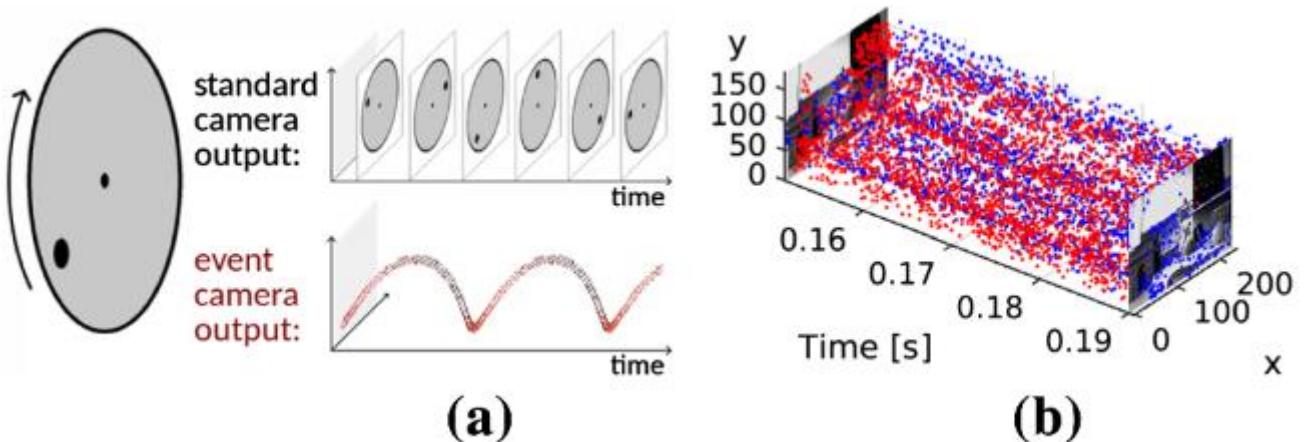


Figure 1: (From Gehrig et al. [8]) a. Comparison of the output of a standard frame based camera and an event camera when facing a black dot on a rotating disk. The standard camera outputs frames at a fixed rate, thus sending redundant information when there is no motion in the scene. Event cameras respond to pixel-level brightness changes with microsecond latency. b. A combined frame and event-based sensor.

Events are colored according to polarity: blue (brightness increase) and red (brightness decrease).

Also, because of the asynchronous operation of SNNs, they are suitable for processing temporal data from scenes captured by event cameras (or silicon retinas) inspired by the biological retina. Contrary to standard frame-based cameras, event cameras only report the per-pixel intensity changes, represented by 4-tuples: the x, y position of the pixel, the sign of the intensity change p (polarity), and the timestamp t. While standard frame-based sensors capture a fixed number of images per second, making them motion blind between frames (as illustrated in Figure 1), which is a strong limitation in dynamic scenes with fast moving objects, event sensors show a number of advantages

---

over standard sensors namely:

- High temporal resolution ( $\sim 1 \mu\text{ s}$ )
- High dynamic range (140 dB versus 60 dB)
- Ultra-low power (mW versus W)
- Low-latency
- No motion blur

State-of-the-art approaches in machine learning provide excellent results for vision tasks with standard cameras, however, asynchronous event sequences require special handling, and spiking networks can take advantage of this asynchrony. The above-mentioned properties of each spiking neuron networks and event based cameras make them ideal for a better accurate semantic segmentation which then can be used for:

- Video segmentation [9]
- Object tracking [10]
- Annotations generations for unlabelled images by propagating ground truth labels from labeled to unlabeled frame using events between two consecutive frames. [11]
- Real-time semantic scenes segmentation with fast attention [12]
- Federated learning that enables embedding devices to collaboratively learn a shared prediction model while keeping all the training data on device. [13]

In summary, SNN and event-based cameras combination is likely to help machines such as Unmanned Autonomous Vehicles (UAV), autonomous vehicles, or robots gain high-level understanding from visual scenes, therefore participating in developing the next generation AI paradigm that will eventually compete with deep learning.

[1] Hyeyoung Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE international conference on computer vision, pages 1520–1528, 2015.

[2] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440, 2015.

[3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062, 2014.

[4] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12):2481–2495, 2017.

[5] Inigo Alonso and Ana C Murillo. Ev-segnet: Semantic segmentation for event-based cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 0–0, 2019.

[6] Filip Ponulak and Andrzej Kasinski. Introduction to spiking neural networks: Information processing, learning and applications. Acta neurobiologiae experimentalis, 71(4):409–433, 2011.

- 
- [7] Paul A Merolla, John V Arthur, Rodrigo Alvarez-Icaza, Andrew S Cassidy, Jun Sawada, Filipp Akopyan, Bryan L Jackson, Nabil Imam, Chen Guo, Yutaka Nakamura, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197):668–673, 2014.
- [8] Daniel Gehrig, Henri Rebucq, Guillermo Gallego, and Davide Scaramuzza. Eklt: Asynchronous photometric feature tracking using events and frames. *International Journal of Computer Vision*, 128(3):601–618, 2020.
- [9] Wenguan Wang, Tianfei Zhou, Fatih Porikli, David Crandall, and Luc Van Gool. A survey on deep learning technique for video segmentation. *arXiv preprint arXiv:2107.01153*, 2021.
- [10] Qiang Wang, Li Zhang, Luca Bertinetto, Weiming Hu, and Philip HS Torr. Fast online object tracking and segmentation: A unifying approach. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 1328–1338, 2019.
- [11] Yi Zhu, Karan Sapra, Fitsum A Reda, Kevin J Shih, Shawn Newsam, Andrew Tao, and Bryan Catanzaro. Improving semantic segmentation via video propagation and label relaxation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8856–8865, 2019.
- [12] Ping Hu, Federico Perazzi, Fabian Caba Heilbron, Oliver Wang, Zhe Lin, Kate Saenko, and Stan Sclaroff. Real-time semantic segmentation with fast attention. *IEEE Robotics and Automation Letters*, 6(1):263–270, 2020.
- [13] Yeshwanth Venkatesha, Youngeun Kim, Leandros Tassiulas, and Priyadarshini Panda. Federated learning with spiking neural networks. *IEEE Transactions on Signal Processing*, 69:6183–6194, 2021.
- [14] Daniel Gehrig, Henri Rebucq, Guillermo Gallego, and Davide Scaramuzza. Asynchronous, photometric feature tracking using events and frames. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 750–765, 2018.
- [15] Youngeun Kim, Joshua Chough, and Priyadarshini Panda. Beyond classification: Directly training spiking neural networks for semantic segmentation. *arXiv preprint arXiv:2110.07742*, 2021.
- [16] Kinjal Patel, Eric Hunsberger, Sean Batir, and Chris Eliasmith. A spiking neural network for image segmentation. *arXiv preprint arXiv:2106.08921*, 2021.
- [17] Peter Blouw, Xuan Choo, Eric Hunsberger, and Chris Eliasmith. Benchmarking keyword spotting efficiency on neuromorphic hardware. In *Proceedings of the 7<sup>th</sup> annual neuro-inspired computational elements workshop*, pages 1–8, 2019.
- [18] Wolfgang Maass. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9):1659–1671, 1997.
- [19] Francois Danneville, Christophe Loyez, K Carpentier, I Sourikopoulos, E Mercier, and A Cappy. A sub-35 pw axon-hillock artificial neuron circuit. *Solid-State Electronics*, 153:88–92, 2019.
- [20] Emre O Neftci, Hesham Mostafa, and Friedemann Zenke. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36(6):51–63, 201

---

## ***Learning RDF pattern extractors from natural language and knowledge graphs - application to Wikipedia and the LOD***

**Responsable scientifique :** Fabien Gandon

**Laboratoire :** Université Côte d'Azur, Inria, CNRS, I3S

### **Description :**

Whether automatically extracted from articles' structured elements or produced by bots or crowdsourcing, the open and linked data published by DBpedia and Wikidata now offer rich and complementary views of the textual descriptions found in Wikipedia. However, the text of Wikipedia articles still contains a lot of information that it would be interesting to extract for improving the coverage and quality of those bases. Until recently, relation extraction questions were solved by multiple-step processes. The latest improvement in deep learning and the development of large language models have shown their abilities in many downstream and complex tasks, and directly impacted the information extraction field. However, using and restricting these approaches to a given knowledge domain is still an open question.

More information: [http://calculs.univ-cotedazur.fr/wp-content/uploads/2023/07/GeMSS\\_poster-a40fb56a5e235c77960d8dde0f1815a0.pdf](http://calculs.univ-cotedazur.fr/wp-content/uploads/2023/07/GeMSS_poster-a40fb56a5e235c77960d8dde0f1815a0.pdf)

---

## **Contre-argumentation en langage naturel pour lutter contre la désinformation en ligne**

**Responsable scientifique :** Serena Villata

**Laboratoire :** INRIA, I3S, CNRS, UCA, 3IA

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Fallacious Argument Classification in Political Debates. Pierpaolo Goffredo, Shohreh Haddadan, Vorakit Vorakitphan, Elena Cabrio, Serena Villata. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence Main Track. Pages 4143-4149.

<https://doi.org/10.24963/ijcai.2022/575>

Pierpaolo Goffredo, Elena Cabrio, Serena Villata, Shohreh Haddadan and Jhonatan Torres Sanchez. DISPUTool 2.0: A Modular Architecture for Multi-Layer Argumentative Analysis of Political Debates. AAAI 2023 Demo paper

### **Description :**

Online social media platforms try to limit the virality of disinformation spread for example through content moderation. While these measures show some kind of effectiveness in limiting the diffusion of misleading or fake information, one big issue still remains unsolved: identifying disinformation and reporting its status to the users is not enough to counter it.

The goal of the PhD program is to address that urgent need by designing intelligent (semi-)automated ways to generate counter-arguments to fight the spread of disinformation online. The idea is to avoid the undesired effects that come with content moderation, such as overblocking, when dealing with disinformation online, and to directly intervene in the discussion (e.g., Twitter threads) with textual arguments that are meant to counter the fake content as soon as possible, and prevent it from further spreading. The idea is to obtain high quality counter-arguments while reducing efforts and supporting human fact-checkers in their everyday activities.

---

## **Descente d'échelle à très haute résolution pour la prévision de vent local - Dimensionnement de simulation de bassin de navigation sportive**

**Responsable scientifique :** Mireille Bossy

**Laboratoire :** EPC CALISTO (Inria)

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Les principes et avancées scientifiques du projet Aventage sont décrits sur la page web dédiée :

<https://windpos.inria.fr/projects/avantage-downscaling-methodologies/>

En particulier, les calculs Azzurra sont illustrés dans cette page web et dans le rapport de stage de Master 2 de Thomas Ponthieu "Haute résolution pour la simulation du vent par descente d'échelle. Évaluation du modèle SDM-WindPoS, pour une utilisation dans le domaine des sports nautiques" (2021).

Après une période de pause, le projet va être relancé cette année.

### **Description :**

#### **Contexte scientifique**

La prévision météorologique est un domaine d'expertise qui s'appuie sur un modèle numérique qui résout les équations de la physique et de la dynamique de l'atmosphère, alimenté par des données d'observation. Le modèle numérique produit une donnée de prévision pour des périodes de plusieurs jours à large échelle (25 à 50 km) sur l'ensemble du globe. Ensuite, une chaîne de traitement numérique dite de descente d'échelle va raffiner la résolution spatiale et temporelle de la prévision sur une région géographique en particulier.

Dans ce projet nous nous intéressons plus particulièrement à la prévision de condition de vent dans les basses couches de l'atmosphère, là où les effets de la turbulence atmosphérique sont les plus prépondérants et où la description des conditions de terrains (topographie et type de surface, urbaine, mer, végétation) peut apporter beaucoup plus de finesse et mieux résoudre la variabilité des conditions de vent prédites à l'échelle locale.

SDM-WindPoS est un logiciel développé dans l'EPC CALISTO qui met en œuvre une méthode innovante de simulation du vent qui s'appuie sur la formulation lagrangienne des équations de la turbulence pour proposer un solveur numérique qui supporte des résolutions spatiales très fines. Ce solveur numérique peut être utilisé dans le domaine ou

---

sous-domaine pertinent pour l'analyse du vent, la dynamique au bord de ce domaine pouvant être extrapolée à partir d'une simulation météorologique de plus grande échelle.

SDM-WindPoS est notamment conçu pour être utilisé en association avec les sorties d'un solveur météorologique fournissant une prévision de vent dans les premières centaines de mètre de l'atmosphère, et seulement dans la zone où l'information locale de vent est décisive pour l'application visée. Des informations topographiques sont également introduites en entré de la simulation pour enrichir la donnée de plus grande échelle, et ainsi saisir le plus finement possible les effets du relief et du terrain.

### Description et objectifs

Le projet vise à caractériser les performances du modèle SDM-WindPoS sur de la prévision de vent à de très hautes résolutions spatiales, qui sont complètement innovantes dans une chaîne de prévision. Le domaine cible de cette descente d'échelle est le bassin de navigation sportive de la ville de Marseille, sur une plage horaire de quelques heures correspondant à la durée d'un événement sportif sur l'eau.

Différentes situations de vent, par secteur et force sont fournis par le modèle WRF en chaîne opérationnel de prévision. Le modèle WRF est descendu jusqu'à 1km de résolution sur la région de Marseille. Un run de WindPoS prend ensuite le relai de cette descente à quelques dizaines de mètres de résolution.

L'accès à Azzurra pour ce projet nous aidera à évaluer l'apport de SDM-windPoS dans la chaîne de prévision, via

- une validation qualitative des simulations (impact des reliefs naturels et urbains, sillages, recirculations, turbulence).
- une évaluation contre des mesures de vent (sous réserve de leur disponibilité dans la durée du projet).

Azzurra nous aidera également à dimensionner et à d'évaluer plus exhaustivement l'apport et l'impact des simulations emboitées intermédiaires.

---

## **Complex Network Mining**

**Responsable scientifique :** Claude Pasquier

**Laboratoire :** I3S

### **Résultats obtenus grâce aux moyens informatiques attribués :**

La grille de calcul « Azzurra » a été utilisée cette année dans le cadre d'un stage de master 2 et de deux thèses. Le mésocentre est également utilisé dans le cadre de deux projets :

- ARN CROQUIS (Collecting, Representing, cOmpleting, merging and Querying heterogeneous and Uncertain wastewater and stormwater network data) dont l'objectif est de développer un modèle informatique avancé pour la gestion et la fusion de données hétérogènes avec une application aux réseaux d'eaux pluviales et d'eaux usées,
- projet NewgenTOXiv dont l'objectif est de développer des tests de toxicologie *in vitro* afin de remplacer les tests effectués sur les animaux.

Ces travaux ont été décrits dans plusieurs publications :

Pasquier, C., & Robichon, A. (2022). Evolutionary divergence of phosphorylation to regulate interactive protein networks in lower and higher species. International Journal of Molecular Sciences, 23(22), 14429. hal-03881865

Pasquier, C., Guerlais, V., Pallez, D., Rapetti-Mauss, R., & Soriani, O. (2023). A network embedding approach to identify active modules in biological interaction networks. Life Science Alliance, 6(9). hal-04135439

Lombardi, R., Jozwiak, M., Dellamonica, J., & Pasquier, C. (2023, June). A machine learning approach to predict weaning outcome among ventilated patients in Intensive Care unit. In Reanimation 2023, the French Intensive Care Society International Congress (Vol. 13). hal-04160272

Romain Lombardi. Prédiction du succès du test de sevrage ventilatoire en service de réanimation par Machine Learning. Rapport de stage Master 2 Sciences du Vivant Parcours Biologie, informatique et mathématiques (BIM). Université Côte d'Azur 2023

### **Description :**

La modélisation des données sous forme de réseau s'est répandue dans divers domaines en tant qu'approche unifiée pour représenter des données relationnelles complexes.

---

Formellement, ces réseaux complexes sont des multigraphes dans lesquels les nœuds et les arêtes sont (multi-)étiquetés. Le cœur de notre activité de recherche se concentre sur l'analyse de ces réseaux complexes à des fins d'extraction d'informations. Nous développons des méthodes de fouille de données visant à identifier les motifs cachés, à mettre en évidence des relations entre des attributs ou à détecter des modules (ou communautés de nœuds) d'intérêt. Nous mettons l'accent sur trois aspects : (1) l'intégration de données ou de connaissances externes, telles que les métadonnées, les connaissances métier ou les ontologies, (2) la prise en compte de l'évolution temporelle des réseaux et (3) la transparence et l'explicabilité des résultats.

En parallèle, étant donné que notre domaine d'application privilégié est la biologie moléculaire, nous développons des méthodologies spécifiques pour l'analyse de données issues d'analyse transcriptionnelle unicellulaire (sc-RNAseq) ou de séquençages de type « long-read ».

Nos travaux sont réalisés dans le cadre de plusieurs programmes de recherche :

- Le projet RIDAP, financé par l'IDEX de l'Université Côte d'Azur, vise à comprendre, grâce à une approche associant la biologie et l'Intelligence Artificielle, le rôle de SigmaR1 (Sig1R), une protéine chaperon des canaux ioniques activée par le stress, dans le compartiment stromal de l'adénocarcinome canalair pancréatique (AdKP) pour identifier une nouvelle cible thérapeutique.
- Le projet NewgenTOXiv, financé par le Programme d'investissements d'avenir (PIA) a pour but de développer des tests de toxicologie *in vitro* afin de remplacer les tests effectués sur les animaux.
- Le projet CROQUIS, soutenu par l'Agence Nationale de la Recherche (ANR), vise à concevoir des solutions innovantes pour la collecte, la représentation, la fusion, la complémentation et l'interrogation de données hétérogènes et incertaines relatives aux réseaux d'eaux usées et d'eaux pluviales.

Nous prévoyons d'utiliser les installations informatiques d'Azzurra dans le cadre de différents sujets de thèse / post-doc, notamment : Développement d'une approche basée sur l'ARNseq d'une cellule unique pour tester des substances cancérigènes non génotoxiques (avec l'IPMC).

---

Le travail quotidien de routine consiste à :

- Analyse des données de séquences d'ARN de cellules uniques, caractérisation du profil différentiel des types de cellules répondant à des cancérogènes non génotoxiques.
- Assemblage de pipelines basés sur nextflow pour étudier l'épissage alternatif aberrant sur des données RNAseq en vrac et unicellulaires.
- Développement d'un logiciel de visualisation des isoformes d'ARN épissé alternativement.
- Faire des tests statistique des données RNAseq
- Faire du mapping

La quantité de données et la nature des ensembles de données exigent l'utilisation de GPU et d'installations de calcul à haute performance pour permettre le multitâche (test et contre-test de plusieurs flux de travail en développement en même temps) et des tâches spécifiques telles que l'intégration d'échantillons d'ARNseq unicellulaires, l'analyse de trajectoires unicellulaires et d'autres.

---

## ***Simulations numérique de nanophotonique computationnelle avec le solveur Galerkin discontinue massivement parallèle DIOGENeS***

**Responsable scientifique :** Stéphane Lanteri

**Laboratoire :** INRIA

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- Mahmoud M R Elsawy, Anthony Gourdin, Mickael Binois, Régis Duvigneau, Didier Felbacq, et al. Multiobjective statistical learning optimization of RGB metalens. *ACS photonics*, 2021, 8 (8), pp.2498-2508. ([10.1021/acspophotonics.1c00753](https://doi.org/10.1021/acspophotonics.1c00753)). ([hal-03212349v2](https://hal.archives-ouvertes.fr/hal-03212349v2))
- Mahmoud Elsawy, Mickael Binois, Régis Duvigneau, Stéphane Lanteri, Patrice Genevet. Optimization of metasurfaces under geometrical uncertainty using statistical learning. *Optics Express*, 2021, 29 (19), pp.29887. ([10.1364/OE.430409](https://doi.org/10.1364/OE.430409)). ([hal-03356986](https://hal.archives-ouvertes.fr/hal-03356986))
- M. Binois, R. Duvigneau, Mahmoud M. R. Elsawy, P. Genevet, S. Kadir, et al. Advanced Numerical Modeling Methods for the Characterization and Optimization of Metasurfaces. AT-AP-RASC 2022 - 3rd URSI Atlantic and Asia Pacific Radio Science Meeting, May 2022, Gran Canaria, Spain. pp.1-4, ([10.23919/AT-AP-RASC54737.2022.9814395](https://doi.org/10.23919/AT-AP-RASC54737.2022.9814395)). ([hal-03864889](https://hal.archives-ouvertes.fr/hal-03864889))
- Mahmoud M R Elsawy, Mickaël Binois, Régis Duvigneau, Samira Khadir, Anthony Gourdin, et al. Multiobjective optimization for highly efficient and robust metasurface designs. OPTIQUE Nice 2022, Jul 2022, Nice, France. ([hal-03864836](https://hal.archives-ouvertes.fr/hal-03864836))
- Mahmoud M. R. Elsawy, Christina Kyrou, Elena Mikheeva, Rémi Colom, Jean-Yves Duboz, et al. Universal active metasurface modulation with ultimate performance in reflection. Nanophotonics and Micro/Nano Optics International Conference 2022, Oct 2022, Paris, France. ([hal-03864856](https://hal.archives-ouvertes.fr/hal-03864856))

### **Description :**

Atlantis (<http://www-sop.Inria.fr/atlantis/>) est une équipe-projet commune entre l'Inria, le CNRS et l'Université Côte d'Azur par son association avec le Laboratoire de Mathématiques J.A. Dieudonné (UMR 7351).

Ses activités de recherche se concentrent sur la conception, l'analyse, le développement et l'application de techniques de calcul avancées pour des systèmes d'équations aux dérivées partielles (EDP) modélisant des problèmes d'interaction lumière-matière à l'échelle du nanomètre. Le domaine correspondant est celui de la nanophotonique, aussi connue sous le nom de nano-optique, la nanophotonique vise à exploiter les interactions lumière-matière à des échelles nanométriques afin d'atteindre un haut niveau de contrôle

---

de la lumière. La nanophotonique englobe une grande variété de sujets, notamment les métamatériaux, la plasmonique, l'imagerie haute résolution, la nanophotonique quantique et les matériaux photoniques fonctionnels.

C'est dans ce contexte que nous travaillons sur différentes problématiques et des applications en collaboration avec des chercheurs physiciens, notamment de laboratoires SNRS du périmètre UCA (notamment le CRHEA et INPHYNI) et des partenaires industriels (notamment, STMicroelectronics).

Les recherches de l'équipe projet ATLANTIS sont appuyées et soutenues par la suite logicielle DIOGENeS (<http://diogenes.Inria.fr>). Cette suite logicielle comprend plusieurs solveurs numérique type élément fini basés sur la méthode Galerkin discontinue. Cette méthode, particulièrement adaptée au calcul parallèle, permet de simuler des configurations réalistes de structures nanophotoniques. Ces simulations peuvent permettre de valider des nouveaux modèles physiques pour la représentation des caractéristiques des matériaux à des échelles nanométriques, peuvent également permettre de valider de nouvelles méthodes numériques (schéma en temps particulier, etc..) ou bien encore de faire de l'optimisation paramétrique.

---

## **Numerical solution for a hand-held high-resolution medical scanner (MEDSCAN)**

**Responsable scientifique :** Victorita Dolean

**Laboratoire :** LJAD

### **Description :**

PhD project MEDSCAN (BoostUrCareer program - cofunded by EU, IDEX, Région PACA)

Microwave tomography is a novel, early development stage way of imaging with a number of potentially attractive medical applications. With microwave tomography in the frequency range of several hundred MHz to several GHz, tissues are well differentiated and, consequentially can be imaged, based on differences in dielectric properties (i.e. contrast). It has been proven in a number of studies that dielectric properties of biological tissues are a strong indicator of their functional and pathological conditions. A very small data acquisition time (within 10-20 msec range) is another attractive feature of microwave tomography. The possibility of functional and fast imaging combined with decent spatial resolution (5-7 mm), cost efficiency and use of safe, very low dose non-ionizing radiation make microwave tomography very competitive with traditional imaging modalities, widely used in clinical practice, such as CT, MRI, PET, SPECT and ultrasound. Microwave imaging (MI) has attracted significant research interest in recent times. By exposing tissues to low-level microwave incident field and capturing the scattered field by an array of antennas, the estimation of the dielectric properties of the tissues can be approximated and a diagnosis inferred. Physicians primarily rely on medical imaging systems such as computed tomography (CT) and magnetic resonance imaging (MRI) for confirmation of the diagnosis. There are, however, not fast, cost effective nor portable, and nor carried by first response paramedical teams. Furthermore, the increase use of CT scans has brought to light health concerns due to the ionizing radiation the technology relies upon. However, imaging the human tissues is an ambitious task; it is composed of multi-layered dielectrics of high loss.

From computational point of view, we expect a development of a parallel computing simulation tool which will be based on FreeFEM++ (and thus open source). The linear systems resulting from the discretization of the three-dimensional time harmonic Maxwell equations using continuous or discontinuous finite element methods designed on

---

tetrahedral meshes are characterized by large sparse, complex coefficients and irregularly structured matrices. Classical preconditioned iterative methods (such as the GMRES Krylov method preconditioned by an incomplete LU factorization) generally behave poorly on these linear systems. A standard alternative solution strategy calls for parallel sparse direct solvers, but they generally fail to solve very large three-dimensional problems, due to the potentially huge memory requirements for these cases. Iterative methods can be used to overcome this memory problem. In the past years, a class of parallel solution methods namely two-level overlapping Schwarz methods already developed for some scalar partial differential equations was recently extended to Helmholtz and Maxwell equations in some particular cases. These equations are notoriously difficult, and more investigation is needed especially in heterogeneous cases.

There are two major challenges at the numerical level.

1. Developing a numerical tool for solving the forward problem to realize a realistic model of the data acquired by the microwave imaging system is novel and challenging. The modeling must have to take account accurately of the high heterogeneity and complexity of human tissues (skin, fat, muscle, veins and arteries...) for normal cases and for different possible pathology cases. The tissues are dissipative dielectric media in the microwave domain, i.e. they exhibit a complex permittivity. They are moreover dispersive (the complex permittivity varies versus frequency). The wave/matter interaction must take also in account accurately of the field transmitted and received from imaging system (Evanescent Imaging Matrix).
2. Developing a new numerical method for solving the inverse problem able to reconstruct tomographic microwave images from experimental data (for different realistic phantoms models) acquired with the system.

Both points rely on the solution of large problems of millions of unknowns for which scalable methods from the state of the art need to be tested and others to be developed.

---

## **Metis**

**Responsable scientifique :** Serena Villata

**Laboratoire :** 3IA Côte d'Azur/ 3IA techpool

### **Description :**

Le Secrétariat général de la défense et de la sécurité nationale (SGDSN) a proposé au Lab IA de l'aider à identifier et classifier les informations importantes sur Internet grâce à l'Intelligence Artificielle. L'emploi de l'intelligence artificielle répond aux priorités suivantes :

1. Optimiser la veille des menaces et des risques : aider les agents dans leur mission de veille, et notamment de détecter, d'identifier et de classifier automatiquement les informations à transmettre immédiatement.
2. Passage à l'échelle sur la quantité de flux traitée : possibilité d'évaluer une plus grande quantité d'informations, qui aujourd'hui est bloqué par le nombre d'agents du bureau, et qui sera débloqué par l'outil, par la capacité de stockage de données et la puissance de calcul ;
3. Automatisation d'une tâche de faible valeur ajoutée : cela permettra aux agents de consacrer plus de temps à des tâches à plus forte valeur ajoutée, en particulier d'accompagner la montée en puissance du Bureau de veille et d'alerte (BVA) sur la synthèse et l'analyse.

L'objectif est de développer une preuve de concept permettant d'extraire et analyser les données textuelles issues d'une agence de presse, pour les classifier comme étant des informations présentant un intérêt fort sur les questions de défense et de sécurité en général, ou sur des thèmes particuliers (par exemple, veille sur la pandémie, veille sur la résilience).

## The environment representation for autonomous driving

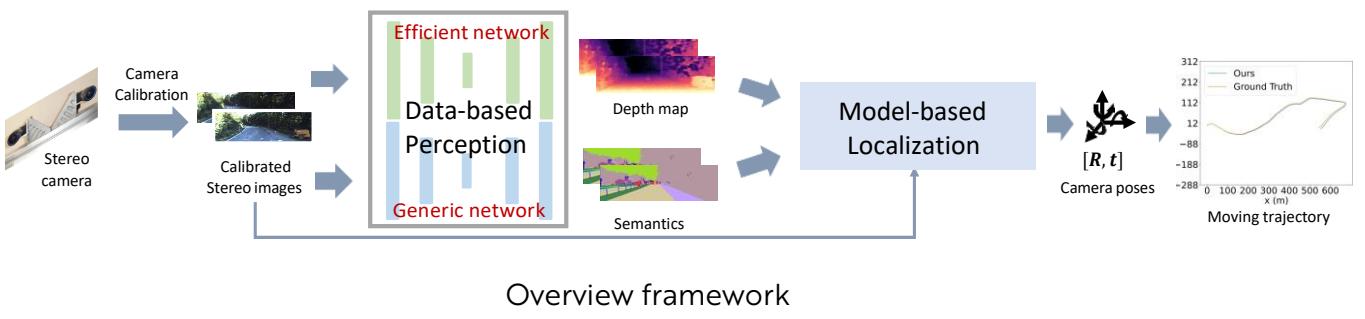
**Responsable scientifique :** Ezio Malis

**Laboratoire :** Inria

### Description :

#### Context

Automated vehicles have been mature with the new technologies in sensors and algorithms. For an autonomous driving vehicle, the basic requirement is to know where it is, then make planning and controlling in the environment. When we want an autonomous vehicle to localize itself in real-time, it is required to enrich the extracted information, including geometry information and abstract semantic information. The ability to detect and recognize all encountered situations and then decide the appropriate behavior is essential. In the end, autonomous driving is running in real-time. The system should have the ability to perceive the environment and to give an accurate self-localization like a human in real-time.



Overview framework

#### Objectives

The main objective of this project is to set up a global perception and localization framework. The overview framework is shown in Figure 1. The objectives include:

1. Build environment representation which includes two kinds of information:
  - Geometry (Depth information)
  - Semantics
2. Autonomous driving application.

- 
- Localization (where the autonomous vehicle is)
  - Environment perception (mobile objects and free space detection)

According to the above objectives, our methodology is shown as follows:

1. Combine the data-based and model-based approaches.
2. Generic and efficient algorithms for real-time applications.

---

## ***Augmenting participation, co-creation, trust and transparency in Deliberative Democracy at all scales***

**Responsable scientifique :** Serena Villata

**Laboratoire :** INRIA, I3S, CNRS, UCA, 3IA

### **Description :**

ORBIS addresses the disconnects between ambitious ideas and collective actions at a large socio-technical scale. It responds to the profound lack of dialogue between citizenship and policy making institutions by providing a theoretically sound and highly pragmatic socio-technical solution to enable the transition to a more inclusive, transparent and trustful Deliberative Democracy in Europe. The project shapes and supports new democratic models that are developed through deliberative democracy processes; it follows a socio-constructive approach in which deliberative democracy is not a theory which prescribes new democratic practices and models, but rather the process through which we can collectively imagine and realize them. ORBIS provides new ways to understand and facilitate the emergence of new participatory democracy models, together with the mechanisms to scale them up and consolidate them at institutional level. It delivers: (i) a sound methodology for deliberative participation and co-creation at scale; (ii) novel AI-enhanced tools for deliberative participation across diverse settings; (iii) a novel socio-technical approach that augments the articulation between deliberative processes and representative institutions in liberal democracies; (iv) new evidence-based democratic models that emerge from the application of citizen deliberation processes; (v) demonstrated measurable impact of such innovations in real-world settings. The project builds on cutting-edge AI tools and technologies to develop a sustainable digital solution, and bridges theories and technological solutions from the fields of political and social science, social innovation, Artificial Intelligence, argumentation and digital democracy. The achievement of the project's goal is validated through six use cases addressing contemporary issues at different scales and settings, experimenting with different civic participation and deliberation models, and involving diverse types of stakeholders.

---

## Magnétohydrodynamique

**EUR Spectrum**

---

## ***Effet Dynamo et Instabilités Magnétohydrodynamiques***

**Responsable scientifique :** Yannick Ponty

**Laboratoire :** Lagrange / OCA, CNRS

### **Description :**

Ce projet comporte trois volets, qui utilisent différentes techniques numériques et deux codes pseudo-spectraux. Ces projets seront conduits pendant au moins deux ans 2020 et 2021. Nous avons déjà obtenu du temps de calcul à GENCI (2020), notamment sur les machines Occigen (500 000 h/mono procs) et Curie (400 000 h/mono procs). La production de long et coûteux runs seront essentiellement effectués sur les clusters de GENCI. Le mésocentre Azzurra sera utilisé pour des tests et des simulations de plus faible consommation.

- I ) Effet dynamo dans le Von Karman ( Yannick Ponty Lagrange)
- II) Effet dynamo de fluctuation (Dario Vincenzi LJAD & Yannick Ponty)
- III) Optimisation par méthodes adjointes pour la dynamo (Florence Marcotte INRIA/LJAD & Yannick Ponty).

### **Contexte et historique :**

L'origine des champs magnétiques est abordée par l'effet dynamo, c'est à dire l'amplification spontanée d'un germe magnétique dans des fluides électriquement conducteurs en mouvement. Depuis de nombreuses années, plusieurs projets de dynamo expérimentales voient le jour tant en France qu'à l'étranger. Deux projets de dynamo expérimentales (Riga (novembre 1999 et juillet 2000) et Karlsruhe (décembre 1999) ont obtenu un effet dynamo dans des écoulements hélicitaires très contraints dans des tubes, présentant un champ de vitesse avec très peu de fluctuations [1]. L'équipe française de Von Kàrmàn Sodium dont l'expérience est basée à Cadarache a essayé d'obtenir l'effet dynamo avec un écoulement très fluctuant et turbulent. Ils ont mené plusieurs campagnes d'expériences depuis les années 2000, où seuls les phénomènes d'induction avaient pu être étudiés. Mais en septembre 2006, le projet a obtenu ce qui semble être la première dynamo turbulente de l'histoire [2]. Une moisson de résultats expérimentaux a été obtenue malgré la difficulté des mesures dans les métaux liquides. Les simulations numériques qui ont accès à toutes les quantités physiques en même temps vont donc jouer un rôle important dans la compréhension de l'effet dynamo et de sa saturation [3].

---

L'avenir expérimental de la dynamo se trouve actuellement en Allemagne dans le centre Helmholtz zentrum à Dresden [4], où une expérience est en train de se construire dans un nouveau bâtiment. Un cylindre de 2m de diamètre rempli de sodium liquide sera forcé par un système de précession sur une énorme plateforme tournante. Ils espèrent obtenir un effet dynamo, dans un écoulement turbulent. En collaborant directement avec des équipes expérimentales et dans un souci permanent de comparer les simulations et les résultats expérimentaux, ce projet numérique commencé depuis 2002, a permis explorer et de comprendre les propriétés d'un tel système, et notamment l'influence de la turbulence sur le seuil critique de l'instabilité dynamo, la relation entre la topologie émergente à grande échelle et les fluctuations à petite échelle. Cette étude paramétrique, en nombre de Reynolds cinétique et magnétique nécessite des simulations temporelles longues et des résolutions spatiales élevées pour approcher les régimes turbulents, que seuls des codes parallèles peuvent atteindre.\*

[1] Special issue : Magnetohydrodynamics (2002).

[2] R. Monchaux, M. Berhanu, M. Bourgoin, M. Moulin, Ph. Odier, J.-F. Pinton, R. Volk, S. Fauve, N. Mordant, F. Pétrélis, A. Chiffaudel, F. Daviaud, B. Dubrulle, C. Gasquet, L. Marié, and F. Ravelet."Generation of a Magnetic Field by Dynamo Action in a Turbulent Flow of Liquid Sodium"Phys. Rev. Lett. 98, 044502 (2007).

<http://dx.doi.org/10.1103/PhysRevLett.98.044502>

[3] M. Berhanu, R. Monchaux, S. Fauve , N. Mordant, F. Pétrélis, A. Chiffaudel, F. Daviaud F. Ravelet, M. Bourgoin , Ph. Odier, J.-F. Pinton, R. Volk, B. Dubrulle, L. Marie "Magnetic field reversals in an experimental turbulent dynamo" Europhys. Lett. 77, 59001 (2007). doi:10.1209/0295-5075/77/59001

[4] André Giesecke, Tobias Vogt, Thomas Gundrum, and Frank Stefani. "Nonlinear Large Scale Flow in a Precessing Cylinder and Its Ability To Drive Dynamo Action" Phys. Rev. Lett. 120, 024502 – Published 12 January 2018.

I) Effet dynamo dans le Von Karman.

{ Holger Homann & Yannick Ponty (Lagrange, OCA)}

Forçage par des disques en mouvement avec un perméabilité magnétique variable: Effet dynamo fluide-structure.

Dans cette approche nous simulons complètement le forçage des disques qui tournent en contra-rotation pour produire un écoulement turbulent proche des écoulements observés en laboratoire [2-3]. Nous avons obtenu des résultats similaires aux expériences dans des simulations purement hydrodynamiques. Un article sur des résultats hydrodynamiques de ce système est publié dans New Journal of Physics (La méthode de pénalisation est expliquée dans notre l'article publié). Nous avons introduit le champs magnétique pour étudier l'effet dynamo dans un tel système et le comparer aux dynamos trouvées dans l'expérience de VKS [2, 3]. Il s'agit surtout de modéliser l'effet de la perméabilité magnétique de la structure disques-pales sur le mode dynamo. Nous avons tous les ingrédients hydrodynamiques (champ moyen, tourbillons le long des pales). Il nous reste à explorer et à comprendre l'interaction fluide-structure.

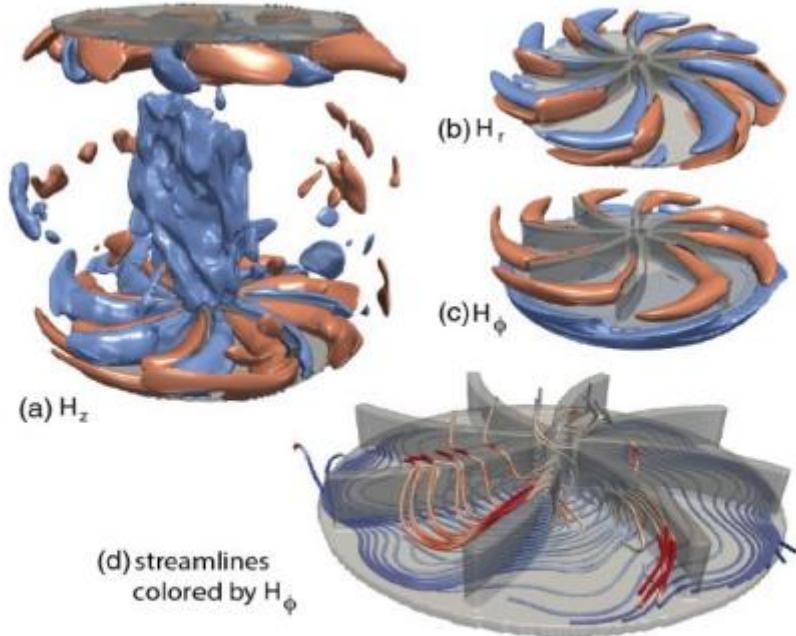


FIG. 4. (a)  $H_z$ , (b)  $H_r$ , (c)  $H_\phi$  isoamplitude contours (red, positive value; blue, opposite negative value), and (d) magnetic streamlines colored by the  $H_\phi$  value.  $Re = 1500$ ,  $Rm = 500$ , and  $\mu_r = 14$ .

Fig] : Visualisation des composantes du champ magnétique  $H_z$ ,  $H_r$ ,  $H_\phi$  et des lignes de champ autour de la structure solide du disques-pales, du mode dynamo ( $m=0$ ) .

En 2016-2019:

Nous avons pu faire une série de simulations où un effet dynamo est produit par notre von Karman numérique. Lors de cette série, nous avons varié la perméabilité magnétique de la structure solide de l'expérience (les disque-pales). Nous avons pu montrer que la géométrie des pâles et le saut de perméabilité joue un rôle moteur pour obtenir le mode

dynamo ( $m=0$ ) observé dans les expériences (voir Fig 1). Le saut de perméabilité magnétique pousse les lignes de champ vers le disque, qui devient un attracteur et un amplificateur de champ. Cet effet est fortement présent dans le régime de saturation. Notre article est publié dans Phys. Rev. Lett<sup>36</sup>

Pour 2021 et ...

- Nous devons continuer à explorer la dynamique non linéaire présente dans le régime de saturation (inversion erratique du champ magnétique) avec des simulations plus longues. Nous avons observé des inversions du champ magnétique, comme dans l'expérience VKS. Nous devons produire des simulations longues pour étudier la dynamique de ces inversions, leurs statistiques et les différents mécanismes associées (voir Fig 2).
- Nous pourrons aussi explorer une autre géométrie expérimentale qui n'a pas produit de dynamo (sphère) avec deux disques-pâles identiques à VKS.

### Magnetic inversions with differential rotation $\theta = 0.23$

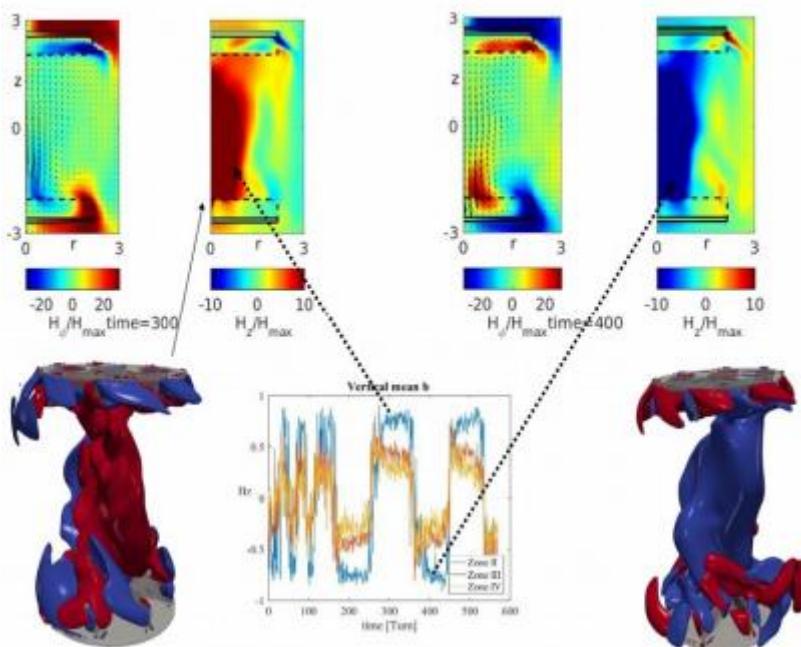


Fig 2: Inversion du champ magnétique dans une simulation où les deux disques ne tournent pas à la même vitesse :  $\theta = (\Omega_2 - \Omega_1)/(\Omega_2 + \Omega_1)$ .

## II) Effet dynamo de fluctuation

<sup>36</sup> Dynamo Enhancement and Mode Selection Triggered by High Magnetic Permeability S. Kreuzahler, Y. Ponty, N. Plihon, H. Homann, and R. Grauer Phys. Rev. Lett. 119, 234501 – Published 6 December 2017 [PDF]  
Une illustration de cette publication a fait l'objet de la COVER PAGE du journal Phys. Rev. Lett. (PRL).

---

{ Dario Vincenzi (LJAD) & Yannick Ponty (Lagrange, OCA)}

En collaboration avec Dario Vincenzi (Lab Dieudonné Nice), nous voulons explorer numériquement la dynamo de type Kasensev. Une contribution fondamentale à la théorie de la dynamo turbulente (ou de fluctuation) a été donnée par Kazantsev<sup>37</sup> (1968), qui a étudié analytiquement la croissance de l'énergie magnétique dans un écoulement aléatoire incompressible, gaussien et à temps de corrélation très court. Le modèle de Kazantsev continue à revêtir une grande importance dans la théorie de la dynamo<sup>38 39</sup>, car il est l'un des rares modèles de dynamo turbulente qui peut être traité analytiquement. Il a ainsi fait l'objet d'un grand nombre d'études et plusieurs généralisations du modèle ont été proposées afin de prendre en compte des propriétés d'un écoulement turbulent qui ne figuraient pas dans le modèle original. Même s'il est très utile pour une compréhension qualitative de l'effet dynamo turbulent, le modèle de Kazantsev est néanmoins assez limité à cause des hypothèses d'isotropie et de décorrélation temporelle. La compréhension de la dynamo turbulente dans des conditions réalistes se fonde donc nécessairement sur la simulation numérique et la comparaison avec les expériences de laboratoire. Ici nous proposons d'étudier numériquement différents écoulements de type Kraichnan soit avec un forçage delta corrélé en temps donc avec du bruit additif<sup>40</sup> soit une système PDE stochastique où l'écoulement (le bruit) est multiplicatif, où le schéma numérique prendra en compte de la stochasticité du système. De plus, l'effet de la compressibilité sera aussi étudié et comparé aux résultats analytiques<sup>41</sup>.

Ce projet utilisera le code Cubby. En utilisant la bibliothèque BOOST::RANDOM ou std::random, il y a une implémentation d'un contrôle de champ de vecteur aléatoire qui sont reproductibles avec un système de "seed" contrôlés et reproductibles. Cette propriété sera importante pour cette étude sur de la dynamo produite par des écoulements aléatoires et où le contrôle des séquences aléatoires seront importantes.

### III) Optimisation par méthodes adjointes pour la dynamo

{Florence Marcotte (INRIA/LJAD) & Yannick Ponty(Lagrange, OCA)}.

---

<sup>37</sup> Kazantsev, AP (1968) Sov. Phys. JETP 26, 1031

<sup>38</sup> Rincon, F (2019) J. Plasma Phys. 85, 205850401

<sup>39</sup> D. Vincenzi, The Kraichnan–Kazantsev Dynamo J. Stat. Phys. 106, 1073–1091 (2002)

<sup>40</sup> A A Schekochihin et al 2007 New J. Phys. 9 300

<sup>41</sup> M. Martins Afonso, D. Mitra & D. Vincenzi "Kazantsev dynamo in turbulent compressible flows" Proc. R. Soc. A 475, 20180591 (2019).

---

Un autre volet du projet concerne le développement et la mise en œuvre d'un code de contrôle optimal d'écoulements MHD incompressibles. Le Direct Adjoint Looping (contrôle optimal) permet aujourd'hui d'optimiser dans des espaces de grande dimension des systèmes fluides instationnaires et gouvernés par une dynamique entièrement non-linéaire. Fondé sur des méthodes adjointes, le DAL détermine itérativement la variable d'optimisation (par exemple un forçage) minimisant ou maximisant une quantité d'intérêt (fonctionnelle de coût) tout en prescrivant un certain nombre de contraintes (équations d'évolution, conditions de bord, etc) à l'aide de multiplicateurs de Lagrange<sup>42</sup>. En pratique, sa mise en œuvre consiste, en partant d'une première estimation arbitraire de la variable d'optimisation, à intégrer les équations d'évolution gouvernant le problème physique considéré (« problème direct ») à partir d'un instant initial jusqu'au temps cible T (où la quantité d'intérêt est évaluée), puis à remonter par une intégration en temps rétrograde du problème adjoint (gouvernant les multiplicateurs de Lagrange) à une condition permettant de mesurer la distance à l'optimalité. Une meilleure estimation de la variable d'optimisation est alors construite, et la procédure est répétée jusqu'à convergence de l'algorithme. Chaque itération (chaque passage dans la boucle DAL) représente en terme de coût de calcul l'équivalent de 2 simulations numériques directes (3 lorsque la mémoire requise nécessite l'utilisation de checkpointing). La méthode DAL a déjà été utilisée dans de nombreux contextes en hydrodynamique, notamment pour optimiser le mélange dans un fluide stratifié<sup>43</sup>. L'application de la méthode DAL pour la première fois à des écoulements instationnaires gouvernés par le système MHD non-linéaire complet (équations de Navier-Stokes, équation d'induction et conditions solénoïdales) nécessite le développement et la validation des outils numériques appropriés. Elle permettra d'étudier les conditions optimales pour l'obtention d'une dynamo fluide (en terme d'abaissement du seuil de l'instabilité dynamo par un forçage approprié). Ce type d'étude est notamment motivé par l'intensification, au niveau international, de l'effort de recherche pour le développement de dynamos expérimentales, notamment à Dresden (projet DRESDYN<sup>44</sup>) ou à Madison (projet MPDX<sup>45</sup>). Ce projet utilisera le code Cubby qui permet facilement d'implémenter de nouvelles physiques ou des systèmes d'équations.

---

<sup>42</sup> D. Hill, Adjoint systems and their role in the receptivity problem for boundary layers, *J. Fluid Mech* (1995) 292.

<sup>43</sup> F. Marcotte & C.P. Caulfield, Optimal mixing in 2D stratified plane Poiseuille flow at finite Peclet and Richardson numbers. *J. Fluid Mech.* (2018) 853.

<sup>44</sup> F. Stefani et al., Dresdyn - a new facility for mhd experiments with liquid sodium, *Magnetohydrodynamics* (2014) 48

<sup>45</sup> C. Cooper et al., The Madison plasma dynamo experiment: a facility for studying laboratory plasma astrophysics, *Physics of Plasmas* (2014) 21

---

Sa construction modulaire est adapté à la procédure DAL : DNS-Adjoint ou les équation adjointes changent à chaque changement de contraintes. L'algorithme d'optimisation nécessite de nombreuses boucles DNS-Adjoint pour converger, soit un cycle de 30/40 simulations pour une seule contrainte. Méthodes numériques: Les calculs seront effectués sur le codes «Latu» et «Cubby» qui sont deux codes pseudospectraux qui calculent des EPD dans un espace périodique. Le code «Latu» est un code pseudo-spectral parallélisé avec la bibliothèque MPI construit par Holger Homann (Laboratoire Lagrange, OCA). Il peut donc tourner sur toutes les plateformes. Il est écrit en C++ et utilise les transformées de Fourier (FFTW3) et P3DFFT. Ce code est efficace, il a été utilisé pour un projet PRACE en 2011 avec un scaling pour des runs en  $4096^3$  jusqu'à 64000 cores . Il permet d'étudier la dynamique des fluides, de MHD, de particules inertielles, modèles bifluide ou de traceur. Il y a aussi une implémentation des techniques de pénalisation pour des parois immersives et des objets en mouvement comme la rotation les disques à pales de VKS. Il est maintenu par un suiveur de version GIT. Récemment, il y a eu un effort récent pour le mettre le système de compilation sur CMAKE, et implémenter des tests automatiques. Le code « Cubby » est un code pseudo-spectral parallélisé avec MPI et pthread. Il y a donc une double parallélisation . Il a été originellement construit par Yannick Ponty (Laboratoire Lagrange OCA) et ensuite il a été re-designé par un IR Alain Miniussi (OCA) en C++ (design pattern), de façon extrêmement modulaire. Il a un très bon scaling. Il est maintenu par un suiveur de version Subversion, et va passer prochainement sous GIT. Sa construction modulaire permet l'implémentation d'un nouveau système de PDE ou de schéma numérique très facilement. Il utilise CMAKE comme système de compilation, avec plus de 150 tests automatiques unitaires ou complets.

---

## **Contrôle optimal de dynamos sous-critiques**

**Responsable scientifique :** Florence Marcotte

**Laboratoire :** LJAD / Inria

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Présentations en conférence/séminaires invités :

- Minimal seeds for subcritical dynamos. Complex Days, Nice, janvier 2023 : contribution orale de F. Marcotte
- Université de Newcastle, octobre 2023 : séminaire de F. Marcotte

Les résultats obtenus dans le cadre du précédent projet DYNSEED ont été déterminants pour obtenir le financement StG-ERC associé au projet CIRCE, lancement en janvier 2024.

### **Description du projet :**

Le projet vise à développer une méthode d'analyse de stabilité non-linéaire d'écoulements magnétohydrodynamiques (MHD) fondée sur le contrôle optimal, en considérant dans un premier temps le cas particulier des instabilités dynamos sous-critiques. Le code de contrôle optimal développé dans le cadre de ce projet s'appuie sur un code déjà existant, validé, massivement parallèle et librement distribué (DEDALUS) pour intégrer les équations de la MHD incompressible instationnaire en 3D (équations de Navier-Stokes + équation d'induction + contraintes solénoidales pour le champ de vitesse et le champ magnétique) en temps avant, ainsi que leurs équations adjointes en temps rétrograde. Ce code permettra d'étudier les conditions optimales pour l'obtention d'une instabilité dynamo fluide dans un domaine à géométrie cylindrique, en particulier dans un écoulement linéairement stable. Ce type d'étude est notamment motivé par l'intensification, au niveau international, de l'effort de recherche pour le développement de dynamos expérimentales, notamment à Dresde (projet DRESDYN) ou à Madison (projet MPDX). Il doit permettre à terme de caractériser, dans un contexte plus général, la stabilité non-linéaire d'autres types d'écoulements MHD (applications en fusion à confinement magnétique notamment).

---

## Astrophysique

**EUR Spectrum**

---

## **EXOWINDS: High-resolution spectra of hot exoplanets**

**Responsable scientifique :** Vivien Parmentier

**Laboratoire :** Lagrange

### **Description :**

Among the thousands of known exoplanets, a subset of hot, close-in planets are the best ones for atmospheric characterisation. These hot Jupiters span a very large range of temperature, metallicity, gravity or rotation period. Using transit and eclipse spectroscopy we can access the chemical composition of their atmosphere and deduce elemental abundance ratios (e.g. C/O ratio, Fe/O) that are key to understanding their formation mechanisms (Guillot et al. 2022).

Hot Jupiters are often tidally locked, leading to intrinsically 3D atmospheres. Strong day-to-night winds transport heat and chemical species from the hot dayside to the cold nightside. The spatial variation of temperature and chemical composition can strongly bias the measurements of the abundances of their atmospheric constituents and even lead to spurious molecular detections (Taylor et al. 2020). That is because the observations often probe the spatially averaged spectra of these planets. Composition, temperature and their spatial inhomogeneities are entangled in the measured signal and one cannot be measured correctly without properly understanding the other. This is particularly true for the incoming JWST observations, for which the quality of the observations means that 3D effects will become dominant (Pluriel et al. 2020).

The atmospheric physics of hot Jupiters and the degree of spatial inhomogeneities of these planets are still not fully understood. Their horizontal temperature contrast depends on the competition between advection of heat from day to night by the winds, that tends to homogenize the thermal and chemical structure, and the radiative cooling of the atmosphere that tends to increase the day/night contrast. The main mechanisms setting the wind speed in hot Jupiter atmospheres are currently debated, with different dissipation mechanisms (e.g. hydrodynamic instabilities or Ohmic drag) leading to different predictions as to how the wind speed should vary across the population of planets (Koll and Komacek 2018). Understanding the atmospheric circulation of exoplanets is therefore fundamental to correctly measuring their atmospheric properties, including their chemical composition.

---

Ground-based high-spectral resolution observations able to resolve individual lines ( $R > 40,000$ ) are key to better understand hot Jupiter atmospheres because they give us a direct access to dynamical quantities and allow for local measurements. The Doppler shifts created by the planet's rotation and the atmospheric winds allow lines formed in different part of the planet atmosphere to be spectroscopically separated, turning spectral separation into spatially resolved spectra. The technique has recently been used to infer chemical and thermal gradient in the planet WASP-33b (Cont et al. 2021; van Sluijs et al. 2022) and measure the east/west wind speed difference on the planet WASP-76b (Gandhi et al. 2022). However, to obtain a holistic understanding of hot Jupiter atmospheres, we need to carry out these studies at a larger, population scale.

The goal of this proposal is to determine the 3D thermal, chemical and dynamic structure of the population of hot Jupiter atmospheres using high-spectral resolution observations. For this we propose to combine a forward modelling approach, using 3D non-grey global circulation models (WP1) and a reverse approach by developing a multi-dimensional atmospheric retrieval framework (WP2). This proposal is supported by several observational programs and its legacy will inform the future E-ELT observations (WP3).

---

## ***Red supergiant and AGB atmosphere and winds using 3D radiative hydrodynamics simulations***

**Responsable scientifique :** Andrea Chiavassa

**Laboratoire :** Lagrange / OCA

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Radiative hydrodynamics (RHD) simulations of rotation stars have been computed on AZZURRA and LICALLO (OCA). The results are going to be submitted to A&A (Astronomy and Astrophysics, referred journal) by the end of the month of October showing that the prototypical red supergiant star Betelgeuse is likely not rotating but the turbulent convection mimics the observed rotational profile. The paper will be J.Z. Ma, A. Chiavassa et al.. J-Z. Ma is a PhD student working in Germany.

Description du projet en page suivante.

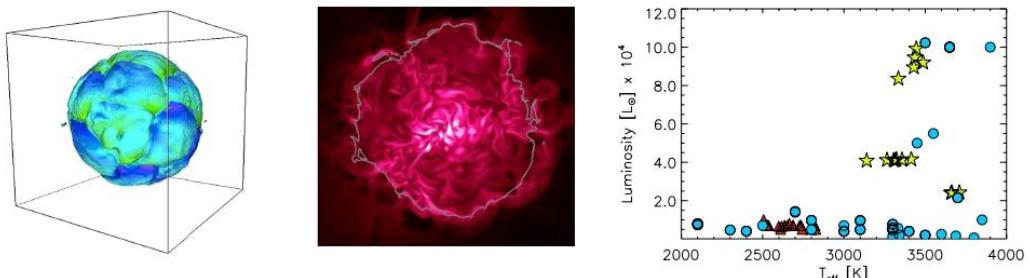
## Red supergiant and AGB atmosphere and winds

In the last years it has become possible to produce multi-dimensional radiative hydrodynamics (RHD) simulations of the gas flow in the surface layers including the effect of radiation. The energy exchange between the gas and the radiation field is critical as it determines the temperature stratification and is ultimately what drives the convective motion. 3D simulations are ab initio, time-dependent, multi-dimensional, and not-local. These simulations are of two kinds: (i) *box-in-a-star* models (for main sequence stars) cover only a small section of the surface layers of the deep convection zone, and the numerical box includes some fixed number of convective cells, which are large enough to not constrain the cells by the horizontal (cyclic) boundaries; (ii) *star-in-a-box* models (Fig., left panel) cover the whole convective envelope of the star and are used to model RSG and AGB stars. The transition where the box-in-a-star models become inadequate to represent the star occurs around  $\log(g) \sim 1$ , when the convective structures become too large (e.g. Chiavassa et al. 2010) and the influence of sphericity becomes important. The numerical code we use and for which we have a solid expertise is CO5BOLD-code (Freytag et al. 2012). This code is state-of-the-art for the *star-in-a-box* simulations. CO5BOLD is computer-time demanding. However, with the increased computational power, it has become possible to compute numerical grids of RSG and AGB stars (Fig., right panel). The computational time varies with stellar parameters, the numerical resolution, the physical ingredients included, the numerical approaches and the directive of parallelization taken into account. Until recently, CO5BOLD' calculations were performed in parallel using Open Multi-Processing (OpenMP) that limits the shared memory and the number of computing threads to a single server. Our most recent version of the code allows us to tackle an hybrid parallelised approach (i.e. Message Passing Interface - MPI - and OpenMP) for the calculations that increases the speed by a factor of 8 at least. In the end, the computation of a simulation depends on the numerical resolution required, the treatment of the radiative transfer, and the temporal follow up. Nowadays, it takes about 15 days to obtain a relaxed simulation on 320 cores (8 nodes with 2 CPU, 20 cores/CPU).

The presence of a magnetic field in stars is intimately linked to the convection across the stellar photosphere. Convection generates a field and at the same time increases the atmospheric velocities and the chromospheric temperatures. As a consequence, the overall structure of the stellar atmosphere is affected. For example, in the case of RSGs, local dynamos are expected to appear in relation with the large convective cells, as shown using CO5BOLD with their first toy models (Freytag et al. 2002). The numerical scheme used for the solution of equations of Magneto-HydroDynamics (MHD) in CO5BOLD is different from the one employed for the case of pure RHD case (see Freytag et al. 2012 for more details). Recently, CO5BOLD has been used to calculate the first Magneto-HydroDynamics simulations of evolved stars (Fig. 2, central panel), demonstrating the feasibility of this kind of approach.

We will work on the computation/development/analysis of a new grid of (at least 20) RSG and AGB simulations including the magnetic field. The already existent 3D grid will be used as starting simulations for the new grid. Since the WP aim at covering the parameters space of the observations (Fig., right panel), it will be very important to make stringent choices of the initial parameters and numerical resolution. This aspect will be tackled in detail working on: (i) higher numerical resolution to resolve small-scale features at the photospheric surface, (ii) large temporal coverage, (iii) characterisation of the effects of magnetic field in the outer layers, (iv) simulations including rotation (already undertaken in Freytag et al. 2017 for AGBs and by Chiavassa for RSGs, work in progress) will also be performed to study the enhancement of the dipolar component of the magnetic field.

**These simulations are of paramount importance because they will provide our knowledge basis on mechanisms launching evolved stars winds, assuming that combined effects of magneto-acoustic shocks, convection, radiation pressure on gas can create the required conditions.** The grid across the different stellar parameters will be crucial to puzzling out how the mass-loss rate and other wind properties depend on the stellar evolution.



**Fig.** *Left panel:* Example of *star-in-a-box* simulation. *Central panel:* 3D simulation of an RSG with magnetic field (Freytag, private communication). The image displays the magnetic field strength in the interior, increasing towards the core. The approximate position of the surface (indicated with the gray line) shows the emergence of many small magnetic loops, which develop in the outer layers. *Right panel:* actual grid of RSG (yellow stars) and AGB (red triangles) simulations compared to the actual sample (blue circles) of our observations

---

## **Couplage turbulent gaz-grains solides dans les disques protoplanétaires**

**Responsable scientifique :** Héloïse Méheut

**Laboratoire :** Lagrange / OCA

### **Résultats obtenus grâce aux moyens informatiques attribués :**

I was able to test the modifications I've done to the code Snoopy. I've performed with it 3D simulations that I presented in various conferences:

- Clustering of heavy particles in Keplerian turbulence. Presented by Fabiola Gerosa at Geophysical and Astrophysical Fluid Dynamics 2023.
- Clusters of heavy particles in two-dimensional Keplerian turbulence. Presented by Fabiola Gerosa at Particle Growth in Turbulence 2023 and EUROMECH COLLOQUIUM 628 Complex Particles in Turbulent Flow 2023.
- New insights on planetesimal formation in turbulent disks. Fabiola Antonietta Gerosa, Héloïse Meheut, Jérémie Bec. The inner disk of young stars (SPIDI 2023)

I'm also performing 2D simulations with another code (ShearPart). I'm preparing a paper with these results and I plan to write a second one too.

### **Description :**

Etat de l'art

Les disques protoplanétaires sont des pouponnières de planètes, en permettant leur formation à partir de la poussière interstellaire. Il s'agit de systèmes particulièrement complexes composés de gaz stratifié, cisaillé et en rotation autour d'une étoile jeune et de solides de tailles allant de la centaine de microns jusqu'à celles des planètes. Les observations récentes (ALMA, SPHERE) de disques protoplanétaires ont montré la présence de structures complexes dans la distribution de poussière (spirales, croissants, asymétries). Leur interprétation nécessite de mieux comprendre la dynamique des solides dans ces disques gazeux. Cette problématique rejoint des questions encore ouvertes pour la compréhension de la formation des planètes. D'une part, le transport dans les disques protoplanétaires ne peut pas être expliqué exclusivement par la turbulence magnéto-rotationnelle, mais nécessite d'autres effets. D'autre part, il est nécessaire de mieux comprendre les mécanismes conduisant à une croissance rapide des planétésimaux. Le couplage dynamique entre la poussière et la turbulence du gaz constitue une réponse

---

possible à ces questions. Par exemple, l'instabilité d'entraînement (streaming instability) permet d'expliquer en partie une agrégation rapide. Elle résulte de la rétroaction de la poussière sur le gaz et se matérialise par une excitation locale de la dynamique et l'apparition de filaments avec de très fortes concentrations de solides. Bien que de nombreuses études aient porté sur le développement macroscopique de cette instabilité, les propriétés de la turbulence générée lors du couplage entre le gaz et la poussière, ainsi que son impact sur la concentration de poussière et le transport dans les disques, restent à ce jour mal compris. Aussi, l'effondrement gravitationnel de ces concentrations de solides est encore mal connu, du fait des limites des méthodes numériques à ce jour utilisées. Ce projet s'appuiera sur les résultats obtenus lors du stage d'Anand Kapoor financé par l'EUR Incise qui a permis de démontrer la faisabilité de ce projet.

## OBJECTIFS

Le but de ce projet est d'apporter de nouvelles perspectives à ce problème en proposant une approche novatrice, transdisciplinaire de l'étude du couplage entre la dynamique de la poussière et la turbulence du gaz porteur. Nous développerons des outils de pointe pour la simulation directe du fluide et d'une phase solide qui sera modélisée soit par un grand nombre de particules ponctuelles, soit par une approche hydrodynamique. La collaboration interdisciplinaire entre l'OCA et l'INRIA est ici cruciale pour étudier la turbulence avec les outils théoriques et numériques utilisés en mécanique des fluides et ses applications industrielles, et ainsi débloquer un verrou scientifique dans la compréhension de la formation des planètes.

Les travaux proposés s'articulent en trois objectifs principaux :

1. Étude de la dynamique et de la concentration des particules solides dans un fluide turbulent cisailé et en rotation, en fonction de la fraction massique de particules solides et de leur taille. Les particules sont déjà prises en compte dans notre code, et cet objectif ne nécessitera pas de nouveaux développements.
2. Quantification des modulations de la turbulence incompressible par ces petites particules. Cet objectif nécessite d'implémenter dans le code existant l'effet retour des particules sur le fluide. Ces développements seront achevés dans les mois qui viennent.
3. Extension au cas d'écoulements faiblement compressibles. L'idée est de reproduire les conditions propres aux disques protoplanétaires. Il sera alors possible d'étudier les différents mécanismes de concentration des solides ainsi que leur effet sur le transport dans les disques.

---

Le but de cette demande est de s'attaquer au premier objectif. Les deux autres objectifs, à plus long terme, seront l'objet de demandes de temps complémentaires sur les moyens de calcul régionaux ou nationaux.

## OUTILS

Ce projet utilisera notre code spectral développé en C++ pour l'étude de la turbulence d'un fluide cisaillé. Il s'agit d'un code à mémoire partagée (utilisant FFTW3 et POSIX threads), ce qui est adapté pour cette étude bi-dimensionnelle avec une charge mémoire légère. Nous prévoyons de faire une étude de paramètres (vitesse de rotation et temps d'arrêt des particules) en deux dimensions, sur des grilles de taille  $2048^2$  et avec 106 particules. Nous envisageons d'effectuer 36 simulations correspondant à 6 valeurs de vitesse de rotation et 6 valeurs de temps d'arrêt. L'écoulement doit être intégré sur plusieurs dizaines de temps de retournement pour garantir des statistiques suffisamment convergées. Le temps de calcul d'une simulation est estimé à environ 5000 hCPU.

---

## ***Construction of the Bayesian next generation of planetary ephemerides, B-INPOP***

**Responsable scientifique :** Agnès Fienga

**Laboratoire :** Geoazur

### **Résultats obtenus grâce aux moyens informatiques attribués :**

- Testing the mass of the graviton with Bayesian planetary numerical ephemerides B-INPOP. Vincenzo Mariani, A. Fienga, O. Minazzoli. Rencontres de Moriond 2023 (GRAM).
- Testing the mass of the graviton with Bayesian planetary numerical ephemerides B-INPOP. Vincenzo Mariani, A. Fienga, O. Minazzoli. Accepted by Astronomy & Astrophysics <https://doi.org/10.48550/arXiv.2310.00719>
- Bayesian test of the mass of the graviton with planetary ephemerides. Vincenzo Mariani, Agnès Fienga, Olivier Minazzoli, Mickaël Gastineau, Jacques Laskar. Accepted for publication in Physical Review D. 21 pages. <https://doi.org/10.48550/arXiv.2306.07069>

Description du projet pages suivantes.

---

# CONSTRUCTION OF THE BAYESIAN NEXT GENERATION OF PLANETARY EPHEMERIDES, B-INPOP

**Supervior: A. Fienga, Géoazur**

agnes.fienga@oca.eu

**Co-fundings: ERC AstroGeo (J. Laskar, IMCCE, Observatoire de Paris)**

jacques.laskar@imcce.fr

## Abstract

This project aims at initiating paths for the future development of planetary ephemerides. With the income of new very accurate and very numerous data from space missions such as ExoMars, Bepi-Colombo and JUICE, the maintenance and the development towards new scientific applications of INPOP planetary ephemerides will be very challenging. The INPOP planetary ephemerides are worldwide famous and used planetary ephemerides allowing, beside the regular delivery of positions and velocities of about 14,000 objects of the solar system (major and minor planets as well as the Earth and Moon orbital and rotational states) to common users by also space agencies (such as CNES and ESA), to test general relativity and alternative theories of gravitation and to constraint the mass distribution in the solar system by determining masses for main belt and Kuiper belt objects and consequently the possible existence of an unobserved planet P9. INPOP is also used as an important benchmark for long-term ephemerides dating geological samples.

INPOP is currently built using about 200,000 observations from sources very different in technics but also in accuracies and on the basis of a classic least-square inversion. However because of the important number of unknowns in the dynamical modeling (i.e. about 800 parameters for the planetary orbits, 102 for the Moon libration and 42,000 for the minor planets) the inversion of such a system is non-linear and, since (Fienga *et al.*, 2015), new approaches have been tested for improving the inversion ( Fienga *et al.*, 2020a). In this project, we propose to investigate new directions starting from algorithms developed in the frame of artificial intelligence and deep learning. For example we plan to investigate the Metropolis Algorithm and the Highly Parallelized Neighborhood Algorithm (Cambridge, 1999a) for improving the determination of about 350 main belt asteroid masses that perturb heavily the inner planet orbits. These perturbations are the most limitative factor for the accuracy of the modern planetary ephemerides (Kuchynka *et al.*, 2010) but are also an important source of information for measuring the masses of these poorly known objects (only about 50 asteroids have their masses known in precision). By improving the characterization of these masses we will improve the accuracy of the Mercury orbit (important for the Bepi-Colombo mission (Iess *et al.*, 2021)), the orbit of Mars (fundamental for ExoMars missions) but also for the Earth (which is a critical point for paleoclimatology and the ERC AstroGéo). We will also apply these algorithms for constraining the distribution of mass in the Kuiper belt. We will do this application together with a densification and re-analysis of old observations for Uranus and Neptune which are very sensitive to the perturbations induced by these big but far objects. This aspect is also of crucial interest for the ERC AstroGéo as an important keypoint for the improvement of long-term ephemerides.

Finally, as in ( Fienga *et al.*, 2020b; Bernus *et al.*, 2019; 2020), we will apply these AI approaches to test alternative theories of gravitation.

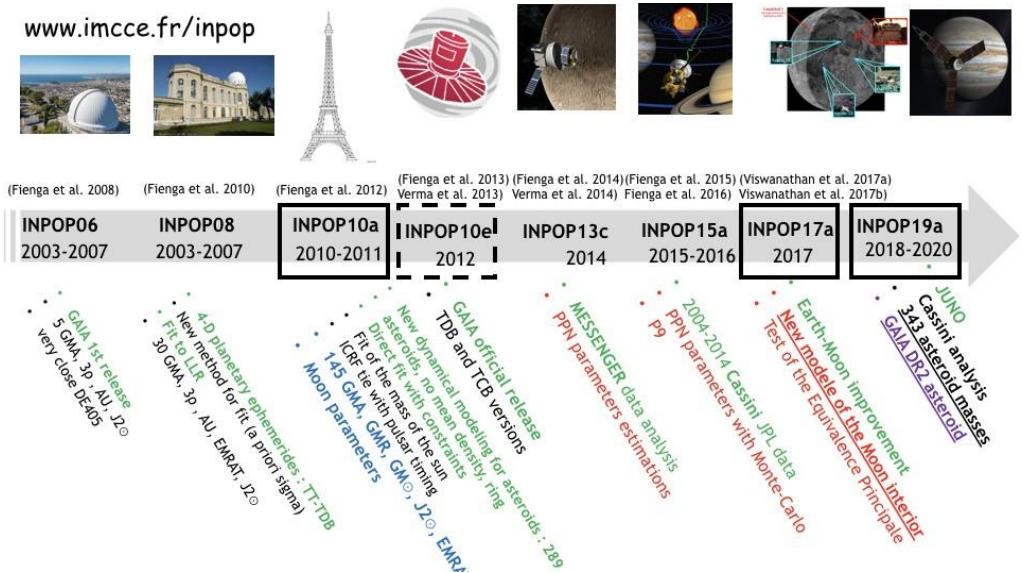


Figure 1: INPOP planetary, lunar and asteroid ephemerides Evolution

## 1 INTRODUCTION

The INPOP (Intégrateur Numérique Planétaire de l’Observatoire de Paris) planetary ephemerides have started to be built in 2003 by J. Laskar and A. Fienga and their team. With the years, INPOP has become a worldwide standard acknowledged by the community (Archinal *et al.*, 2018; Hilton *et al.*, 2016). On Figure 1, are summarized the different evolutions of the INPOP planetary ephemerides: from the first 4-D position, velocity and time-scale ephemerides in 2013 (INPOP10a and its version delivered for the Gaia DPAC, INPOP10e) and the improved Earth-Moon modeling in 2017 (INPOP17a) and its applications in selenophysics (Viswanathan *et al.*, 2019) to the first partially derivative-frees inversion algorithms applied to fundamental physics testings and asteroid mass determinations (Fienga *et al.*, 2015; Bernus *et al.*, 2019; 2020; Fienga *et al.*, 2020a) and the implementation of the almost 2 millions of Gaia solar system object observations in INPOP19a (Deram *et al.*, 2021).

The planetary ephemerides are used for:

- a) Providing very accurate positions, velocities for 14,000 objects of the solar system (major and minor planets as well as the Earth and Moon orbital and rotational states and relativistic time-scales) to common users by also space agencies, such as CNES and ESA. The INPOP planetary ephemerides are used by ESA for the navigation and the data analysis of the missions Gaia, Bepi-Colombo and JUICE. The INPOP planetary ephemerides are accessible to the whole public on the IMCCE website [www.imcce.fr/inpop](http://www.imcce.fr/inpop).
- b) Enhancing the scientific return of space missions such as Mars orbiters (ExoMars, MEX, MRO...), Venus Express (VEX) but also Cassini and Juno missions. For example, with INPOP and the Cassini data, the mass of the Kuiper belt and constraint the existence of unknown and still undiscovered planet P9 have been estimated(Di Ruscio *et al.*, 2020; Fienga *et al.*, 2020b). VEX data were also used to improve mass estimations of main belt asteroids together with Mars orbiters navigation (Verma *et al.*, 2013) as well as solar physics parameters.
- c) Characterizing solar physics parameters such as the sun oblateness ((Fienga *et al.*, 2015)), the electronic densities of the fast and slow winds ((Verma *et al.*, 2013))
- d) Testing laws of gravitation in the solar system. Since 2015, INPOP has been intensively used for providing constraints on possible violation of General Relativity (Verma *et al.*, 2014; Fienga *et al.*, 2015) or for testing alternative theories of gravitation in the solar system (Bernus *et al.*, 2019; 2020).
- e) Estimating the mass distribution in the solar system with the introduction of the latest observations obtained with the Cassini and Juno missions ((Di Ruscio *et al.*, 2020)). With ( Fienga *et al.*, 2020a), 100 main belt asteroid masses have been estimated with an accuracy better than 30%. This result doubles the number of asteroid masses obtained with such accuracy.
- f) Calibrating the datation of geological samples for paleoclimatologic studies. This latest aspect is at the center of the ERC AstroGeo. Indeed, for dating geologic samples, a long-term ephemeris LaX is used for estimating the Earth insolation (Laskar *et al.*, 2011) and LaX is calibrated in using INPOP. Furthermore as it has been demonstrated in

(Laskar, 1999), uncertainties on present ephemeris can have an important impact on its long-term extension, LaX. So INPOP improvements will impact LaX and geological datation.

g) Limiting the source of uncertainties for the detection of Cosmological Background (CMB) gravitational waves with pulsar timing arrays. The CMB gravitation waves are different from the one detected by VirgoLirgo as they are issued from the first ages of the Universe (expansion). As it has been demonstrated (Caballero *et al.*, 2018; Vallisneri *et al.*, 2020), the positions and velocities uncertainties of giant planet orbits are a limiting factor in the bayesian detection of CMB signatures in all pulsar timing arrays. Collaborations between the INPOP team and the European Pulsar Timing array (EPTA) and the US one (Nanograv) have started since 2019 on the question of the assessment of the planetary ephemerides accuracies.

## 2 STATE-OF-THE-ART

The regularly improved accuracies of the spacecraft navigation or radio experiment (below 1 m for Mars orbiters, about 1 cm for Bepi-Colombo, 20 m for Cassini and Juno missions) as well as the increasing number of data (i.e. 50,000 Mars orbiters observations and about 2 millions for the Gaia DR2) make the inversion of this complex dynamical modeling counting for up to 42,000 unknowns very challenging. When for the case of the construction of the 14,000 main belt asteroids, efficient linear algebra algorithm has been implemented successfully (Deram *et al.*, 2021), some cases remain complex to handle in the frame of the classic least-square minimization.

A good example is the problem of the main belt asteroid perturbations on the inner planets. As one can see on Figure 2, the main belt gathers hundreds of thousand of objects of different sizes and of unknown masses. The projection of these object perturbations on the inner planet orbits along the line-of-observations (i.e. the earth-mars distance as on Fig. 2) is non linear and induces correlated displacements that can reach several hundreds of meters, far above the present accuracies of the Mars orbiter observations (80 cm over 10 years). The individual contributions are then very difficult to disentangle and the determination of the masses becomes impossible without introducing specific algorithms. In 2020 (Fienga *et al.*, 2020a), we introduced a Monte Carlo Least-Square (MCLS) algorithm as a first attempt to combine Monte Carlo exploration of possible density constraints extracted from spectroscopic surveys and Least-Square minimization. We obtained 100 asteroid masses over 350 with an accuracy better than 30% but 250 asteroid masses are still not estimated accurately. Furthermore one can question the selection of 350 asteroids proposed by (Williams, 1984) as the most efficient representation of the asteroid perturbations.

The same type of approach combining Monte Carlo (MC) or Genetic Algorithm (GA) and least-square minimizations has been used for the estimation of highly correlated parameters imposed by alternative theories of gravitation (Fienga *et al.*, 2015; Bernus *et al.*, 2020) or for the localisation of the unknown and still undiscovered planet P9 (Fienga *et al.*, 2020b).

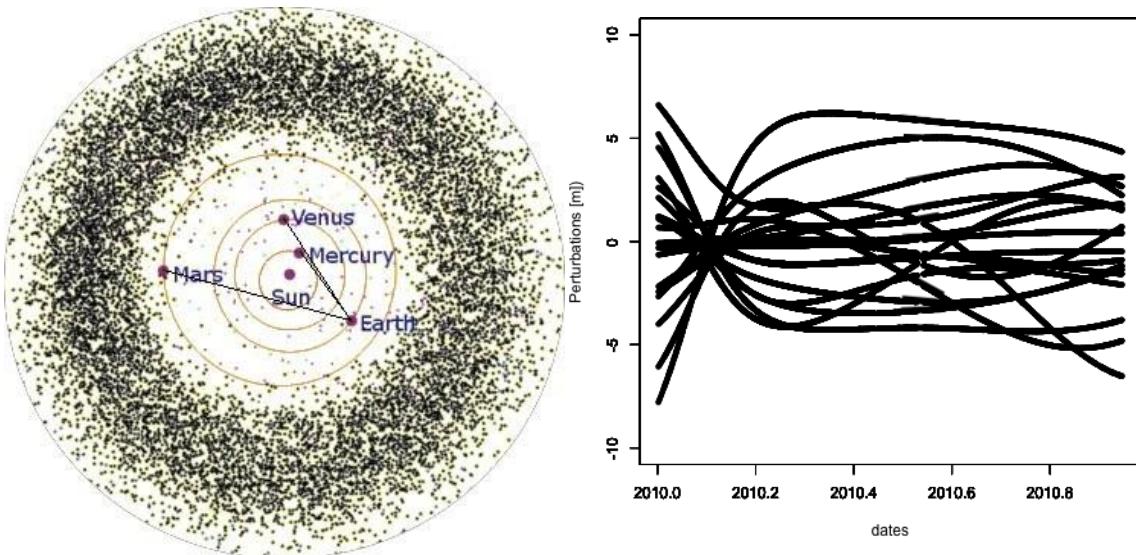
The good side of these mixed methods is the extension of the parameter domain exploration via the MC or the GA with a limited cost of time as these random walks are operated for a reduced number of parameters (the parameters of interest such as the angular positions and distances of P9) but not for all the INPOP parameters. The inconvenient of such methods is the use of least-square minimization for a non-linear model which can lead to local minima even if associated with a wider exploration of the parameter domain. Furthermore, up to now, the analysis of the obtained results were operated in using frequentist paradigm ( $\chi^2$  likelihood).

The goal of the proposed PhD is to develop a complete bayesian approach, never applied to the construction of ephemerides, in order to test new ways for disentangling signatures and exploring wider range of parameters in two major fields of application: the distribution of mass in the solar system and tests of alternative theories of gravitation.

## 3 DESCRIPTION OF THE PROJECT

The goal of this project is to create a new planetary ephemerides called B-INPOP, which will be the bayesian version of INPOP. The basic dynamical modeling will remain the same but the least-squares adjustment will be associated or totally replaced by a derivative-free approach. Several strategies will be investigated.

1) At the beginning, the student will have to investigate the case of the Post Parameterized Newtonian (PPN) parameters  $\beta$  and  $\gamma$  in the frame of the Bepi-Colombo simulations. This work is currently under development. Results have already been obtained with a frequentist method ( $\chi^2$  likelihood). PDFs of the problem have been obtained and a preliminary computation of the bayesian integrals has been done. From there, the student will have to implement a Metropolis algorithm (Tarantola, 1987) and see how efficient it could be in comparison with the results obtained with the likelihood. In a second step highly parallelized Neighborhood Algorithm (Sambridge, 1999a;b) will also be tested. Mixed strategies (MC for the two PPN parameters and least-square minimization for the rest of the parameters) or fully random samplings (priors for all the INPOP parameters, from planetary orbit initial conditions, oblateness and mass of the Sun, Earth-Moon mass ratio to asteroid masses but also PPN) will be tested and compared to the likelihood



**Figure 2: Main belt perturbations on inner planer orbits. On the right-hand side, is plotted an example of perturbations induced by 20 asteroids on the Earth-Mars distances over 1 year.**

results. In terms of time of computation, the random sampling is faster than the mixed strategy (1 h for the MCLS versus 10 mn for the full random sampling when running on 40 processors Intel Xeon 2.3GHz) but requires a more important number of runs. A first publication should come out from this first step.

2) In a second step, a specific focus will be address the determination of the asteroid masses. This problematic has already been experimented with a MCLS approach ( Fienga *et al.*, 2020a) but in this project the student will be asked to generalize the results that he or she will obtain with the PPN parameter determination to a more complex problem as the number of parameters to consider will be of about 350. Using the method developed by (Kuchynka and Folkner, 2013), the student will investigate if additional asteroids are not required in the model and will complete the INPOP dynamical modeling accordingly. Tests involving additional ring implementation will also be considered as a way to disentangle more easily the point-mass object signatures from signatures induced by objects too small to be detected. The same approach will be implemented and tested for the Kuiper belt objects. This task will be associated with an update of the Uranus and Neptune old optical observations based on a new analysis of their satellites observations, taking into account the Gaia catalog to provide improved positions for the planets. At least a second publication should come out from the step 2.

3) Additional applications to other types of theories of gravitation such as the dilaton theory would also be investigated.

4) During the PhD a Monte Carlo Markov Chain will be implemented in order to assess the planetary ephemerides accuracies. The assessment will be firstly compared to the INPOP covariance matrix estimated with the partial derivatives and distributed to the Pulsar timing array communities (EPTA and Nanograv). Publications in collaboration are planned.

We also expect the student to propose other strategies that can emerge from an intensive work of bibliography and discussion with UCA colleagues (i.e. Géoazur, Lagrange, Ecole des mines) using derivative-free inversion methods. In terms of computational power, we are currently using the PSL mesocentre but we will start using the UCA OPAL cluster this fall.

#### 4 EXPECTED OUTPUTS

With the implementation of such bayesian approach and the creation of B-INPOP, we expect three major breakthroughs in planetology, fundamental physics and paleoclimatology:

1) With the determination of 350 asteroid masses, at least as this list could be extended, we will obtain the first lot of consistent mass determinations of this size. In collaboration with M. Delbo (Lagrange, OCA) we will combine the obtained masses with volumes extracted from the MP3C database for the

---

350 main belt bodies among which we have

planetesimals (expected 170) and collisionally generated asteroids. From the comparison of body bulk density to the composition and the mineralogical analogues we will infer the porosity, which informs about the internal structure. We will use composition and internal structure of planetesimals to constrain the properties of the different stages of their formations and we will infer properties of the protoplanetary disk collisional environments.

**2)** With the determination of the B-INPOP mass distribution in the Kuiper Belt, besides the new input for the modele of solar system formation, we will improve significantly the accuracy on the Uranus and Neptune orbits which are key elements for the paleoclimatological datation. The mass distribution in the Kuiper belt plays an important role for the localisation of the whole solar system as it can induce a significant displacement on the solar system barycenter (SSB) relative to the sun. A displacement of about 100 km of the SSB (this is what was obtained with INPOP20a and the latest update of the Kuiper belt mass) leads to a modification of the Earth isolation of several degrees. The topic is directly associated with the ERC AstroGo, co-funder of this project, its aim being the improvement of the Astronomical Solutions over Geological Time.

**3)** With the B-INPOP implementation and its bayesian approaches for testing GR and alternative theories, we will increase the scientific return of the Bepi-Colombo mission by considering complexed theories such as the dilaton and SME (Standard Model Extension) theories. Based on recent simulations not yet published, we also expect to gain a factor 35 for the determination of the PPN parameter  $\gamma$  and about a factor 4 on PPN  $\beta$  thanks to the combination of Bepi-Colombo radio-science measurements and to derivative-frees technics.

**4)** A better assessment of the accuracy of the INPOP planetary ephemerides, specifically for the outer planet orbits, will also help for a better description of the pulsar timing sensitivity to the planetary orbits and for a better description of the planetary priors into the CMB gravitational wave detections.

In conclusion, this project aims at developing B-INPOP, the new generation of INPOP planetary ephemerides, a worldwide used and ERC granted ephemeris, that has shown its important impact in a wide field of topics from planetology, fundamental physics, paleoclimatology. With the new approaches that are proposed to be investigated in this project, B-INPOP will increase the scientific return of ESA missions such as Bepi-Colombo. It will participate at a new understanding of the stages of the solar system formation as well as a better calibration of geological datation, a mandatory step for the long-term studies in Earth climate evolution. Finally it will contribute to the detection of the relic CMB gravitational waves. The risks for this project are very low because of the well-established INPOP team and its strong collaboration in the field and a clear program of publication.

## 5 TECHNICAL ASSESSMENTS

The technical assessments for this project are based on the configuration used on the MesoPSL cluster since 2016. In the least-square configuration, each job is composed by the iteration of the following loop: the numerical integration (parallelized) of the planet and asteroid equations of motions over 1 century (10 mn), the computation of the residuals by comparison with observations (3 mn), the construction of the linearized Newton-Gauss equation (1 mn) and its inversion (3 mn). The stop of the iteration is given by the convergency of the  $\chi^2$  to a minimum. It usually takes 4 hours and each solution occupies, after removing auxiliary files, 150 Mb. The analysis of the results are very fast, based on R routines. In the bayesian configuration, the inversion of the jacobian matrix and the iterations are not necessary, leading to a job duration of about 10 mn and a size of the solution of less than 15 Mb. After analysis, only few Kb remains per solutions. In term of memory, 100 runs in the least square configurations use less than 32 Gb, the matrix inversion consuming the maximum of the memory. With the bayesian approach, the use is even less and should not be a limiting factor. A first step of development and construction of the bayesian approach will be done on our local cluster in Goazur (machine gpm with 72 nodes and 32 Gb of memory). An installation on Azzura will then be operated. By considering 4 cores of 32 nodes each, we will launch 160 runs per 10mn for the bayesian configuration. In asking for 400,000 hours this year over 4 cores, we guarantee 2,4 millions of runs for the bayesian approach which should be enough for the installation and the validation of the procedure as well as the first operational runs of the project. In the next three years, we should continue to work on the same basis.

## REFERENCES

- A. Fienga, J. Laskar, P. Exertier, H. Manche, and M. Gastineau, *Celestial Mechanics and Dynamical Astronomy* **123**, 325 (2015).
- A. Fienga, C. Avdellidou, and J. Hanus\*, *MNRAS* **492**, 589 (2020a).
- M. Sambridge, *Geophysical Journal International* **138**, 479 (1999a).
- P. Kuchynka, J. Laskar, A. Fienga, and H. Manche, *A&A* **514**, A96 (2010), arXiv:1004.3119 [astro-ph.EP].
- B. A. Archinal, C. H. Acton, M. F. A'Hearn, and al., *Celestial Mechanics and Dynamical Astronomy* **130**, 22 (2018).
- J. L. Hilton, C. Acton, J.-E. Arlot, and al., *Transactions of the International Astronomical Union, Series A* **29A**, 22 (2016).
- V. Viswanathan, N. Rambaux, A. **Fienga**, J. Laskar, and M. Gastineau, *Geophys. Res. Lett.* **46**, 7295 (2019), arXiv:1903.07205 [physics.geo-ph].
- P. Deram, A. Fienga, A. K. Verma, M. Gastineau, and J. Laskar, *Celestial Mechanics and Dynamical Astronomy* (2021).
- A. Di Ruscio, A. Fienga, D. Durante, L. Iess, J. Laskar, and M. Gastineau, *A&A* **640**, A7 (2020).
- A. K. Verma, A. Fienga, J. Laskar, K. Issautier, H. Manche, and M. Gastineau, *A&A* **550**, A124 (2013), arXiv:1206.5667 [physics.space-ph].
- A. K. Verma, A. Fienga, J. Laskar, H. Manche, and M. Gastineau, *A&A* **561**, A115 (2014), arXiv:1306.5569 [astro-ph.EP].
- L. Iess, S. W. Asmar, P. Cappuccio, G. Cascioli, and al., *Space Sci. Rev.* **217**, 21 (2021).
- A. Fienga, A. Di Ruscio, L. Bernus, P. Deram, D. Durante, J. Laskar, and L. Iess, *A&A* **640**, A6 (2020b).
- L. Bernus, O. Minazzoli, A. Fienga, M. Gastineau, J. Laskar, and P. Deram, *Phys. Rev. Lett.* **123**, 161103 (2019), arXiv:1901.04307 [gr-qc].
- L. Bernus, O. Minazzoli, A. Fienga, M. Gastineau, J. Laskar, P. Deram, and A. Di Ruscio, *Phys. Rev. D* **102**, 021501 (2020), arXiv:2006.12304 [gr-qc].
- J. Laskar, A. Fienga, M. Gastineau, and H. Manche, *A&A* **532**, A89 (2011), arXiv:1103.1084 [astro-ph.EP].
- J. Laskar, *Philosophical Transactions of the Royal Society of London Series A* **357**, 1735 (1999).
- R. N. Caballero, Y. J. Guo, K. J. Lee, and al., *MNRAS* **481**, 5501 (2018), arXiv:1809.10744 [astro-ph.EP].
- M. Vallisneri, S. R. Taylor, J. Simon, and al., *ApJ* **893**, 112 (2020), arXiv:2001.00595 [astro-ph.HE].
- J. G. Williams, *Icarus* **57**, 1 (1984).
- A. Tarantola, *Inverse problem theory: methods for data fitting and model parameter estimation*, edited by Elsevier (1987).
- M. Sambridge, *Geophysical Journal International* **138**, 727 (1999b).
- P. Kuchynka and W. M. Folkner, *Icarus* **222**, 243 (2013).

---

## **Groupe prospective IA du laboratoire Lagrange**

**Responsable scientifique :** Sylvain Rousseau

**Laboratoire :** Lagrange

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Formation du personnel ingénieur en vue de traiter des données de mesures astronomiques.

### **Description :**

Nous formons un petit groupe d'ingénieurs informaticien au sein du laboratoire Lagrange, encouragés par la direction à acquérir un savoir-faire dans différentes méthodes de Machine Learning (dont les réseaux de neurones), afin d'être appliquées à des problématiques actuelles émergeant dans notre domaine (traitement du signal en astronomie/astrophysique).

En plus d'une plateforme où l'on pourra mettre en pratique notre expérience des codes acquis lors de nos formation en IA, l'utilisation d'Azzurra nous permettrait d'aborder des problèmes d'astrophysique exigeants en terme de puissance de calcul. Par exemple, l'extraction et la reconnaissance de signal dans les données des "data challenges" des futurs satellites LISA (détection d'onde gravitationnelle) nécessiteront le développement de méthodes "deep learning" sur de gros volumes de données.

---

## **Détection d'exoplanètes par imagerie directe exploitant le machine learning**

**Responsable scientifique :** David Mary

**Laboratoire :** Lagrange

### **Description :**

L'imagerie d'exoplanètes nous offre des vues uniques sur d'autres mondes. Par l'observation de leurs environnements, nous pouvons étudier comment de jeunes planètes naissent et se forment au sein de leur nuage primordial, puis comment elles évoluent en un système planétaire tel que le nôtre. Grâce aux images, nous percevons directement les photons qu'elles nous renvoient, qui transportent la signature de la composition de leurs atmosphères.

Ces signaux sont extrêmement ténus : pour chaque photon émis par une planète, c'est un milliard de photons envoyé par leur étoile qui contaminent nos images. La découverte d'exoplanètes est ainsi suspendue non seulement à la qualité de nos instruments, mais également à celle des algorithmes de détection utilisées pour les identifier.

Cette thèse, à la frontière entre l'astrophysique et les approches statistiques de type data mining et Intelligence Artificielle, développera des méthodes de détection d'exoplanètes à la fois puissantes et robustes aux erreurs de détection. Par l'exploitation du grand nombre d'images accessibles dans les archives des instruments, nous développerons une méthodologie de détection s'appuyant sur une analyse statistique solide de la contamination stellaire, qui assurera l'identification des planètes les plus indécélables dans nos images.

Nous démontrerons les performances de cette approche avec les données d'instruments de premier plan : l'emblématique télescope spatial Hubble de la NASA, sur lequel notre équipe a déjà validé avec succès des algorithmes d'imagerie d'exoplanètes, et son successeur NASA-ESA le James Webb Space Telescope (JWST, lancé fin 2021). Cette thèse coïncidera avec l'avènement à court et moyen terme de plusieurs instruments au sol et spatiaux capables d'imager des planètes toujours plus petites et plus proches de leur étoile comme l' ELT (2027, ESO) et NRST (2027, NASA).

Cette thèse alliera l'expertise des deux plus grands laboratoires d'astrophysiques de la région SUD, le laboratoire Lagrange à Nice et le Laboratoire d'Astrophysique de Marseille, en partenariat avec une entreprise locale (basée à Sophia Antipolis) en traitement des

---

images incluant l’Intelligence Artificielle locale, ACRI-ST, spécialisée dans l’observation d’une planète particulière – la nôtre. Cette thèse bénéficiera ainsi d’une émulation régionale et d’une dynamique internationale qui la placent en excellente position pour avoir un fort impact sur l’astrophysique moderne de l’observation de planètes et d’exoplanètes.

---

## **Moon CitcomSVE**

**Responsable scientifique :** Agnès Fienga

**Laboratoire :** Geoazur

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Posters :

- Comparisons between tidal deformation modeling from the semi-analytical ALMA3 and the finite element CitcomSVE. Agnes Fienga, Observatoire de la Côte d'Azur, Géoazur, Valbonne, France, Arthur Briaud, CNRS, Observatoire de la Cote d'azur, Geoazur, Valbonne, France, Shijie Zhong, University of Colorado at Boulder, Department of Physics, Boulder, CO, United States and Anthony Mémin, Université Côte d'Azur, CNRS, OCA, IRD, Géoazur, Nice, France. AGU 2023.
- Tidal deformations and internal structure: sensitivity to rheologies. Agnes Fienga, Observatoire de la Côte d'Azur, Géoazur, Valbonne, France, Alex Guinard, Géoazur - Université Nice Sophia Antipolis, Valbonne, France, Anthony Mémin, Université Côte d'Azur, CNRS, OCA, IRD, Géoazur, Nice, France, Arthur Briaud, CNRS, Observatoire de la Cote d'azur, Geoazur, Valbonne, France and Clément Ganino, Université Côte d'Azur, OCA, CNRS, Géoazur, Valbonne, France. AGU 2023.

### **Description :**

This project takes place in the context of an international collaboration between Géoazur and the German Space Agency (DLR) founded in 2020 by the ANR/DFG contract ANR-19-CE31-0026. The aim of this project is to better understand the internal structure of the Moon as seen by the Lunar Laser Ranging in one hand and satellite altimetry in the other hand. These two techniques give inconsistent descriptions of the Moon tidal deformation.

The question at which this DLR/Geoazur collaboration has proposed to give an answer is about the cause of such inconsistency: Is it an observational issue from one of the two methods (LLR versus altimetry) ? Or is it a more fundamental problem linked to our perception of the lunar inner structure ? Could it be related to the Moon asymmetry between a thin very basaltic and hotter near side and a thick very cratered farside ? In particular, it has been proposed to test the existence of plumes in the lunar mantle that could explain the different regimes of deformation between the near side (observed with LLR and altimetry) and the far side (observed only with space altimetry).

---

To model such heterogeneities in the lunar mantle, close to the one known in the Earth mantle, we proposed to use a finite element method such as the one developed by Prof. Zhong. Prof. Zhong (who will visit our group in 2023 for 6 weeks) has coded an important software (CitcomSVE) able to model viscoelastic deformation at the regional and global scale. For now, this software has been used to study the impact of deglaciation of the Earth surface (Geruo et al. 2013, Kang et al. 2022), Earth lateral plumes motion (Li and Zhong 2019) and some attempt for the Moon tidal deformation (Zhong et al. 2012, Qin et al. 2016).

With this software adapted to the Moon we will be able to test the hypothesis of the existence of Moon mantle plumes for different regions and configurations, in particular those favoring nearside / farside asymmetry. The work to do with Prof. Zhong is then to adapt CITCOMSVE to the Moon case and the tidal deformations induced by the Earth and the Sun on the Moon. Our team has been in contact with Prof. Zhong since 2021 but due to the complexity of the planned modifications, a work with direct interactions will be of highly benefit for this project. Prof. Zhong' stay and the realization of this project will significantly benefit Geoazur by increasing its activities and collaborations with Lagrange in planetary geophysics. It will demonstrate how the skills and tools developed in Earth sciences are also crucial tools for studying the evolution of the solar system. Finally, CitcomSVE is a highly parallelized code involving massive computation. AI optimization methods will be considered at some key stage of the computation. This work with Prof. Zhong will increase the skills developed in Geoazur in intensive computation and Aioptimization.

In the future, besides Moon applications, Mars will also be of very intense interest after the Moon. An hypothesis of mantle plume below the Tharsis Mons has been conjectured recently. With this new version of CitcomSVE we will be able to model the consequences of such a plume on the global and regional deformations of the planet on different time-scales. Finally, our team being part of the mission Bepi-Colombo that will reach Mercury in 2026, we also planned to produce unique modelling of the Mercury deformations thanks to this new version of CitcomSVE.

---

## **Formation des planétésimaux par effondrement gravitationnel**

**Responsable scientifique :** Héloïse Méheut

**Laboratoire :** Lagrange / OCA

### **Description :**

#### **État de l'art**

La formation des planètes est actuellement un problème astrophysique clé, en relation avec les processus de transport et d'accrétion au sein des disques protoplanétaires. Le projet proposé se place dans le cadre d'une croissance incrémentale des solides à partir de la poussière présente dans le disque, et nous nous intéressons plus particulièrement au problème de la formations des planétésimaux à partir de grains de taille millimétrique, qui reste une des principales inconnues de ce scénario. Plusieurs mécanismes qui permettraient de former des amas de poussières liés gravitationnellement sont actuellement discutés, parmi lesquels l'instabilité de flux (plus connue sous le nom de "streaming instability", cf Youdin and Goodman (2005) ; Carrera and Simon (2022)), ou bien un piégeage de la poussière dans des structures en vortex (Gerosa et al., 2022). Cependant, l'étape suivante, qui consiste en l'effondrement gravitationnel de ces nuages denses de particules dans le gaz environnant n'est que peu étudiée, alors que celle-ci est cruciale pour comprendre la structure des planétésimaux formés. Plusieurs études, comme celles de Jansson et al. (2017) ou Nesvorný et al. (2019) se sont intéressées à l'effondrement du nuage de particules, en considérant plutôt les collisions, et en négligeant l'effet du gaz environnant.

#### **Objectifs**

L'étude de l'effondrement d'un nuage de particules couplé au gaz a été menée dans un cadre simplifié par Shariff and Cuzzi (2015), suivant une modélisation 1D sphérique, où le gaz est considéré isotherme, et la poussière est modélisée comme un fluide sans pression. Le projet vise à étendre cette étude, dans plusieurs directions :

- la poussière est modélisée de manière lagrangienne, afin de voir les différences entre les deux approches ;
- une évolution adiabatique du gaz, prise en compte de l'équation d'énergie, est étudiée, afin de voir quel est l'effet du chauffage par la friction entre gaz et poussière sur l'effondrement ;
- un passage en 3D, afin d'étudier l'effet d'une dispersion initiale des particules, ou bien l'effet d'un mouvement de rotation initial du gaz.

#### **Méthodes**

Pour mener à bien ce projet, le code hydrodynamique IDEFIX (Lesur and al., 2023) est utilisé.

Ce nouveau code est écrit en langage C++ et se base sur la librairie Kokkos (Trott and al., 2022)

---

pour la portabilité, ce qui permet de l'utiliser sur CPU ou GPU. La parallélisation est également possible via l'utilisation de MPI. Le code IDEFIX résout les équations hydrodynamiques selon une méthode à volumes finis de type Godunov sur une grille cartésienne, cylindrique ou sphérique, à 1, 2 ou 3D. Un module pour prendre en compte l'auto-gravité a été développé, en permettant de résoudre l'équation de Poisson sur n'importe quelle grille, via l'inversion du laplacien discret par des méthodes algébriques. Un module de particule, basé sur une approche lagrangienne est également disponible. Le code a été testé sur les grands centres de calculs nationaux (Irene Rome, utilisant des CPU AMD, Jean Zay, avec des GPU Nvidia V100, et AdAstra, avec des GPU AMD Mi250), avec une extension du problème d'Orszag-Tang en 3D, avec une résolution atteignant 2563 cellules par cœur. Les performances obtenus sont résumées ci-après :

- Les performances atteignent  $2 \times 10^7$  cellules par seconde sur un nœud AMD de 128 cœurs CPU sur Irene Rome, avec une accélération de l'ordre de 20 sur Jean Zay et de l'ordre de 50 sur AdAstra.
- L'efficacité de la parallélisation est supérieure à 80 % jusqu'à 1024 nœuds CPU (131072 cœurs), et jusqu'à 128 nœuds GPU (512 cœurs) sur Jean Zay. La performance est même de l'ordre 90 à 95 % avec 256 nœuds GPU (1024 cœurs) sur AdAstra.

## Références

- Daniel Carrera and Jacob Simon. Challenge of forming planetesimals from mm grains in dust rings. *Bulletin of the AAS*, 54(8), December 2022. URL <https://baas.aas.org/pub/2022n8i102p06/release/1>.
- Fabiola Antonietta Gerosa, Héloïse Meheut, and Jérémie Bec. Clusters of heavy particles in two-dimensional Keplerian turbulence, October 2022. URL <http://arxiv.org/abs/2210.13147>. arXiv :2210.13147 [astro-ph, physics :nlin, physics :physics].
- Karl Wahlberg Jansson, Anders Johansen, Mohtashim Bukhari Syed, and Jürgen Blum. THE ROLE OF PEBBLE FRAGMENTATION IN PLANETESIMAL FORMATION. II. NUMERICAL SIMULATIONS. *ApJ*, 835(1) :109, January 2017. ISSN 1538-4357. doi : 10.3847/1538-4357/835/1/109. URL <https://iopscience.iop.org/article/10.3847/1538-4357/835/1/109>.
- G. R. J. Lesur and al. Idefix : a versatile performance-portable godunov code for astrophysical flows. *Astronomy and Astrophysics*, 2023. submitted.
- David Nesvorný, Rixin Li, Andrew N. Youdin, Jacob B. Simon, and William M. Grundy. Trans-Neptunian Binaries as Evidence for Planetesimal Formation by the Streaming Instability, June 2019. URL <http://arxiv.org/abs/1906.11344>. arXiv :1906.11344 [astro-ph].
- Karim Shariff and Jeffrey N. Cuzzi. The Spherically Symmetric Gravitational Collapse of a Clump of Solids in a Gas. *ApJ*, 805(1) :42, May 2015. ISSN 1538-4357. doi : 10.1088/0004-637X/805/1/42. URL <http://arxiv.org/abs/1409.2541>. arXiv :1409.2541 [astro-ph].
- Christian R. Trott and al. Kokkos 3 : Programming Model Extensions for the Exascale Era. *IEEE Transactions on Parallel and Distributed Systems*, 33(4) :805–817, April 2022. ISSN 1558-2183. doi : 10.1109/TPDS.2021.3097283. Conference Name : IEEE Transactions on Parallel and

---

Distributed Systems.

Andrew N. Youdin and Jeremy Goodman. Streaming Instabilities in Protoplanetary Disks. *ApJ*, 620(1) :459–469, February 2005. ISSN 0004-637X, 1538-4357. doi : 10.1086/426895. URL <https://iopscience.iop.org/article/10.1086/426895>.

---

## **Biologie, santé**

**EUR Life**

---

## **PRA 2023 - A novel neural framework to assess movement control in health and disease**

**Responsable scientifique :** François Hug

**Laboratoire :** LAMHESS

### **Description :**

Successful movement is critical to facilitate basic and vital functions of living, for interaction with each other and the world around us. However, understanding how movement is controlled remains one of the greatest challenges we face to date. My project lies in the nexus between discovery and clinical science, using an innovative multidisciplinary approach to address critical gaps in our current understanding of the neural control of human movement. The scientific breakthrough of my project is to change the scale at which we observe and interpret the control of movement, i.e. from the muscles to the spinal (alpha) motor neurons. Knowledge gained from this project will have broad applications in neurosciences, rehabilitation, robotics.

Within the context of this project, we aim to understand how synaptic inputs are distributed across a population of motor neurons to produce movement.

One Master student from the Neuromod Master is testing deep learning models to address our scientific question.

Specifically, we develop a set of models in the form of simple variational autoencoders with less than 8500 parameters. We would like to train 3 models on 5 conditions for 6 subjects, 3-fold cross validated, 5 times for a relatively stable random state, for two different datasets, i.e. muscles. This makes for total of 2700 models. Locally we can manage to train 3 models in the span of 5 minutes at most on a single thread 8th generation i7 laptop processor.

---

## ***Caractérisation de la diversité des ostéoclastes multinucléés pour comprendre leur rôle dans les maladies inflammatoires chroniques***

**Responsable scientifique :** Claudine Blin

**Laboratoire :** LP2M (Laboratoire de Physiomédecine Moléculaire)

**Description :**

### **Contexte et impact scientifique**

Les maladies inflammatoires chroniques sont très souvent associées à une destruction osseuse qui peut entraîner une fragilité osseuse, un risque accru de fracture, une invalidité et un décès prématuré. Les ostéoclastes (OCL) sont les seules cellules spécialisées dans la résorption osseuse physiologique nécessaire à la croissance et au maintien du squelette. Elles sont aussi responsables de la destruction osseuse dans des contextes pathologiques, ce qui en fait une cible thérapeutique majeure<sup>1</sup>. Cependant, les traitements actuels ne sont pas satisfaisants, soit en raison de l'apparition de résistances (biothérapies), soit parce qu'en bloquant complètement la résorption osseuse, ils inhibent également la formation osseuse, entraînant une mauvaise qualité osseuse et des effets indésirables.

Les OCL appartiennent à la lignée myéloïde et, en tant que tels, ils sont dotés d'une fonction immunitaire innée<sup>1–3</sup>. Comme d'autres cellules myéloïdes (telles que les macrophages), les OCLs ont des origines multiples, selon le stade de développement où le contexte pathologique<sup>1,4</sup>. Cependant, les conséquences de ces différentes origines sur la diversité fonctionnelle des OCLs ne sont pas connues. De plus, les OCL sont les seules cellules monocytaires physiologiquement multinucléées qui se différencient par fusion de cellules myéloïdes et qui, pour maintenir un nombre de noyaux stable, sont aussi capables de fission<sup>1,4,5</sup>. Ces caractéristiques font des OCLs une cellule extrêmement dynamique. Malgré cela, leur diversité reste négligée.

Notre équipe a déjà montré que les OCL dérivés des monocytes et des cellules dendritiques sont distincts et ont des fonctions immunitaires différentes<sup>1–3</sup>. Grâce à des analyses transcriptomiques à large échelle, nous avons confirmé chez la souris l'existence de sous- populations distinctes d'OCLs, identifié plusieurs marqueurs et certaines caractéristiques spécifiques de ces sous- populations. Par conséquent, les OCL semblent beaucoup plus complexes que prévu et ces nouvelles propriétés ouvrent des

perspectives thérapeutiques très nouvelles et prometteuses. L'identification de biomarqueurs de l'activation précoce des OCL pathologiques permettra de mieux prendre en charge les patients avant que la perte osseuse ne soit détectée. La compréhension des caractéristiques spécifiques des diverses populations d'OCL pourrait conduire à des thérapies ciblées pour les OCL pathologiques sans affecter les OCL physiologiques, préservant ainsi la qualité de l'os.

## **Objectifs**

La diversité des OCLs et sa contribution à l'évolution des propriétés de ces cellules au cours de la vie est une question très nouvelle et aucune donnée n'est encore publiée dans ce domaine. Notre objectif est donc d'explorer en détail cette diversité au cours du développement et du vieillissement ainsi que dans les maladies liées à la destruction osseuse inflammatoire (ostéoporose, arthrite, cancer), chez la souris. Les cellules multinucléées sont peu explorées en analyse computationnelle et nous sommes intéressés par l'analyse de ces cellules afin de :

- 1) comparer les populations d'OCLs murins avec d'autres cellules myéloïdes afin de mieux les situer dans la lignée myéloïde et de prédire comment leur origine et leur fusion avec différents progéniteurs myéloïdes ou cellules matures participent à leur diversité.
- 2) comparer les populations d'OCLs murins entre eux pour définir leur spécificité aux différents âges de la vie et en condition pathologique, de comprendre comment ces populations sont ou non liées entre elles et d'identifier de nouveaux biomarqueurs ou des cibles thérapeutiques pour les ostéoclastes pathologiques.

## **Faisabilité**

L'équipe dirigée par Claudine Blin (directrice de recherche, LP2M, Nice) est spécialiste de l'ostéoimmunologie et de la biologie des ostéoclastes. Le LP2M dispose de tous les équipements et de l'expertise technique pour mener des expérimentations sur les OCLs (isolement, cytométrie multispectrale, culture, tests fonctionnels sur des OCLs matures). Tous les modèles murins sont disponibles au laboratoire ou grâce aux collaborations établies par l'équipe.

L'analyse bioinformatique des données est assurée par Valeriia Rezapova, doctorante dans l'équipe (bourse 3IA, EUR-Life) qui applique différents outils existants à l'analyse de la diversité des OCLs. Le volume de données traitées est croissant et les capacités de calcul des ordinateurs du laboratoire sont insuffisantes. Le recours au centre de calcul Azzurra est donc devenu indispensable pour continuer notre projet.

Ces analyses computationnelles seront basées sur les données transcriptomiques que nous avons déjà obtenues ou que nous allons réaliser dans ces différents contextes. Ce projet tirera également profit des données transcriptomiques disponibles dans les bases de données publiques sur différentes cellules myéloïdes. Outre les données transcriptomiques, le projet intégrera également des données provenant de la cytométrie multispectrale, de l'histologie multiplex et de l'analyse métabolomique.

## Perspectives

Ce projet est à l'intersection de l'immunologie, du remodelage des tissus et de la bioinformatique. Il permettra de comprendre comment l'origine et la multinucléation des ostéoclastes influencent leur phénotype et leur fonction. Il fournira également de nouveaux outils pour l'analyse bioinformatique des cellules multinucléées. Enfin, il fournira des informations très novatrices sur les caractéristiques spécifiques des ostéoclastes pathologiques, telles que des biomarqueurs ou de nouvelles cibles thérapeutiques, ouvrant ainsi de nouvelles perspectives pour le diagnostic, le suivi ou le traitement de la destruction osseuse inflammatoire.

## References

1. Madel, M.-B. et al. Immune Function and Diversity of Osteoclasts in Normal and Pathological Conditions. *Frontiers in Immunology* 10, (2019).
2. Ibáñez, L. et al. Inflammatory osteoclasts prime TNF $\alpha$ -producing CD4(+) T cells and express CX3 CR1. *J. Bone Miner. Res.* (2016) doi:10.1002/jbmr.2868.
3. Madel, M.-B. et al. Dissecting the phenotypic and functional heterogeneity of mouse inflammatory osteoclasts by the expression of Cx3crl. *eLife* 9, e54493 (2020).
4. Jacome-Galarza, C. E. et al. Developmental origin, functional maintenance and genetic rescue of osteoclasts. *Nature* 568, 541–545 (2019).
5. Wakkach, A. et al. Bone marrow microenvironment controls the in vivo differentiation of murine dendritic cells into osteoclasts. *Blood* 112, 5074–5083 (2008).
6. Madel, M.-B. et al. Osteoclasts contribute to early development of chronic inflammation by promoting dysregulated hematopoiesis and myeloid skewing. 2020.12.09.418137 (2020) doi:10.1101/2020.12.09.418137.

---

## ***Automatic Generation of a Three-Dimensional Model of the Forearm using a Statistical Shape Model for Preoperative Planning***

**Responsable scientifique :** Marc-Olivier Gauci

**Laboratoire :** UR2CA-iULS

**Résultats obtenus grâce aux moyens informatiques attribués :**

Rapport de stage de Sonya Joyce Ines NDOMKEU TCHATO « Construction d'une base de données et segmentation automatique d'avant-bras par Deep Learning: premières étapes de la planification préopératoire. » Polytech Grenoble, 2023.

### **Description :**

The PerFORMS project aims to:

- Automatic segmentation of the addressed bones (ulna and radius), starting from CT-scans, which is often the first step in 3D modeling.
- Prediction of the missing anatomy of the forearm due to a potential truncation by the scanner field of view (FOV) which is generally focused on the wrist and not the whole forearm.
- Prediction of the pronosupination movement of the forearm which is needed to standardize the position of the forearm in anatomical position and to understand the potential bony conflicts that could exist.

---

## **Impact des éléments génétiques mobiles dans le cancer et le vieillissement**

**Responsable scientifique :** Gael Cristofari

**Laboratoire :** IRCAN

**Description :**

Au-delà des petites variations nucléotidiques tels que les SNPs (single nucleotide polymorphisms) ou les indels (insertions-délétions de quelques bases), les variations structurales de l'ADN (larges insertions et délétions, amplifications, translocations, inversions) contribuent considérablement à la plasticité du génome humain. L'insertion d'éléments transposables en est un mécanisme prépondérant. La plupart de ces séquences sont des rétrotransposons, c'est-à-dire des éléments génétiques mobiles, qui prolifèrent par un mécanisme de copier-coller, appelé rétrotransposition, via un intermédiaire ARN et une étape de transcription inverse. Les progrès des technologies de séquençage à haut débit ont jeté un nouvel éclairage sur la variation du génome par rétrotransposition. Des centaines de copies L1 actives sont présentes dans notre génome et produisent de nouvelles insertions, contribuant pour ~20 % de toutes les variantes structurelles chez l'Homme. En outre, il est maintenant bien établi que la rétrotransposition ne se limite pas à la lignée germinale - entraînant des variations génétiques héréditaires et parfois de nouvelles maladies génétiques, mais peut également entraîner des réarrangements somatiques importants du génome au cours de l'embryogenèse, du développement neural, ou dans de nombreux cancers. En outre, ils peuvent contribuer directement au vieillissement normal ou au développement de maladies neurodégénératives.

L'objectif principal de ce projet est d'intégrer des données multi-omiques pour prédire à travers des approches de machine learning l'impact de ces éléments sur les gènes, y compris lorsqu'ils s'insèrent dans des régions non-codantes (introns, régions intergéniques). Plus précisément, nous analyserons l'effet:

1. sur l'épigénome (marques d'histones, méthylation de l'ADN grâce à des données de séquençage nanopore);

2. sur la transcription grâce à des données de transcription native et de eQTLs;

Nous comparerons de larges jeux de données générés dans le laboratoire ou obtenus dans des bases de données publiques (ENCODE, GTex, etc). Différentes approches de machine learning seront mises en oeuvre, y compris des "large language models".

---

Dans l'ensemble, ce projet établira un nouveau cadre pour prédire l'impact potentiel de milliers de rétrotransposons identifiés dans les projets de séquençage, et élargira notre compréhension d'un processus mutagène majeur chez l'Homme.

---

## ***Etude de populations de nématodes parasites de plantes Meloidogyne enterolobii***

**Responsable scientifique :** Etienne Danchin

**Laboratoire :** INRAE

**Description :**

Les nématodes à galles du genre *Meloidogyne* sont un groupe de nématodes parasites de plantes impactant lourdement l'approvisionnement alimentaire mondial. Ils sont responsables de plusieurs milliards d'euros de pertes économiques chaque année dans le domaine agricole. Ces nématodes sont les plus répandus et les plus agressifs. Ils peuvent parasiter virtuellement n'importe quelles plantes cultivées et ainsi détruire des hectares entiers de cultures, pourtant les méthodes de luttes actuelles sont très réduites. Les nematicides chimiques ont été interdits dans la plupart des pays en raison de leur toxicité, laissant comme méthodes principales le déploiement de plantes non-hôtes ou plantes portant des gènes de résistance contre les nématodes du genre *Meloidogyne*. Cependant, cette dernière méthodologie manque de fiabilité : peu de gènes de résistance sont connus, certains de ces gènes se font contourner au bout d'un certain temps ou bien sont même totalement inefficaces contre certains nématodes. C'est le cas en Europe avec *Meloidogyne enterolobii*, une espèce identifiée pour la première fois en 2008 en Suisse mais qui depuis cette année a été classée organisme de quarantaine.

A l'heure actuelle, le potentiel adaptatif de *Meloidogyne enterolobii* envers une large gamme d'hôte ainsi que ces caractéristiques moléculaires nous sont encore inconnus. Pour cela nous avons produit des données génomiques de haute-fidélité avec le séquenceur PacBio pour différentes souches de *M. enterolobii* d'origines géographiques et de gamme de plantes hôtes différentes. Ces données génomiques de haute-fidélité ont ensuite été combinée avec un transformeur, DeepConsensus afin d'améliorer leur qualité. Ce transformeur a augmenté en moyenne de 4 % le nombre de lectures ayant une qualité Q20, de 10 % les lectures ayant une qualité Q30 et enfin de 70 % les lectures ayant une qualité Q40, chez les 5 isolats testés au sein de la plateforme de bio-informatique et de génomique de l'Institut Sophia Agrobiotech (ISA). Les génomes de ces différentes souches ont ensuite été assemblés et nous avons sélectionné comme nouveau génome de référence l'assemblage le plus contigu. Il a une taille de 273Mbp avec 556 contigs et une valeur N50 de 2,11 Mb. Nous avons ensuite aligné les lectures des

---

autres populations contre le nouveau génome de référence afin d'identifier et de décrire les différentes variations présentes au sein de cette espèce. Nous avons actuellement trouvé 7416 variations parmi les 5 isolats testés. Les plus représentées étant les délétions pour 57 % et les insertions avec 35 %. Quand toutes nos populations auront été séquencées, nous étudierons ensuite si les patterns des variations corrèlent avec les différences dans la gamme d'hôtes compatibles et la répartition géographique ce qui nous permettra de mieux appréhender la menace que représente ce nouvel organisme de quarantaine.

A ce jour, d'autres populations sont en cours de séquençage et devraient pouvoir bénéficier du même traitement afin d'être rajoutées dans les analyses de variants. Néanmoins l'étape limitante de notre pipeline d'analyse se trouve pendant l'utilisation du transformeur DeepConsensus. Cette étape requiert une grande puissance de calcul pour traiter les données brutes (fichiers subreads) et générer les nouveaux fichiers fastq qui seront utilisés pour la suite. Pour le moment, nous avons utilisé la plateforme de bio-informatique et de génomique de l'ISA pour lancer nos analyses. Deepconsensus met en moyenne 255h avec 250 Cpus et 1 Gpu pour tourner sur une seule population et utilise environ 1,5 To d'espace de stockage. Nous prévoyons de lancer DeepConsensus sur 6 nouvelles populations, si possible avec plusieurs en parallèle. De ce fait, dans le cadre de ce projet, nous aimerais obtenir l'accès au cluster de calcul Azzurra pour pouvoir lancer de manière plus efficace DeepConsensus sur nos populations restantes.

---

## **Cibler l'hétérogénéité de réponse cellulaire aux anticancéreux pour augmenter leur efficacité thérapeutique (brevet Phénotype-Avancé Génotype)**

**Responsable scientifique :** Jérémie Roux

**Laboratoire :** IRCAN

### **Description :**

The project is about the development of an algorithmic framework of tumor cells dynamic analyses, integrating machine learning and signal processing.

The objective is to develop a new algorithmic architecture integrating deep learning, and signal processing to predict drug response phenotypes of each cell forming a tumor clone, from dynamic imaging data acquired in the laboratory. To study these partial responses, the evolution of tumor cells treated with cancer drugs is imaged at regular intervals by microscopy. The sequence of images produced makes it possible to observe changes in cell properties (shape, radiometry), their division, and their phenotypic response (cell death or survival). Cells are plated on 96 well plates than imaged with the DV microscope. The output for one experiment, are images from up to 48 hours of acquisition of up to 40 sequences. Each sequence is a stack of 3200 images (about 1 GB): it is composed of up to 4 channels and 800 frames per channel (the number of channels can be increased depending on the experiment).

The aim will be to create a framework for semi-automatic or automatic dynamic analysis of these image sequences in order to understand, and therefore predict, the phenotypic responses to cancer treatments of each cell constituting a tumor clone.

The framework is currently under development on python using recent packages of data representation, image processing, signal processing, and machine learning (Tensorflow, keras, sklearn).

These analysis processes will make it possible to better understand the molecular origin of the variability of cellular responses to cancer drugs, and to highlight new therapeutic targets and biomarkers of sensitivity to the drugs tested.

---

## ***Etude de la contribution d'une protéine de liaison aux ARNs conservée dans la localisation des ARNs synaptiques et recherche systématique de ses cibles***

**Responsable scientifique :** Florence Besse

**Laboratoire :** IBV

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Using bulk RNA-seq and iClip-seq analysis to investigate mRNA localization in synapses and uncover the role of the IMP protein. Hiba Laghrissi, Lauren Blot, Martina Halleger, Jernej Ule, Arnaud Hubstenberger & Florence Besse. POSTER JOBIM 2023 - Nice (FRANCE)

### **Description :**

Le projet consiste à analyser différents jeux de données de séquençage :

- RNA-Seq sur des échantillons de synaptosomes de drosophiles sauvages et mutants
- iCLIP sur des protéines de liaison aux ARNs sauvages et mutantes
- HyperTRIBE sur des protéines de liaison aux ARNs sauvages et mutantes

Les ressources Azzurra sont essentielles à l'alignement des séquences sur les génomes de références.

---

## **Automatic cell type annotation for cell atlas construction**

**Responsable scientifique :** Christophe Bécavin

**Laboratoire :** IPMC

### **Résultats obtenus grâce aux moyens informatiques attribués :**

A single-cell atlas of the COPD lung, unpublished

scPermut : building a latent space reinforcing celltype identity, unpublished

### **Description :**

Chronic respiratory diseases (COPD, fibrosis, asthma, cystic fibrosis, ciliopathies) affect several hundred million people worldwide and the COVID pandemic has recently drawn world attention on the impact of infectious respiratory diseases. The tissue destruction that is associated with these diseases is often followed by a strong tissue remodeling of the airway tract, one of body's first line of defense. To better understand the importance of the different resident cell types in these processes and determine useful new targets for developing the therapeutical arsenal, we are developing single-cell analyses.

Single-cell technologies are becoming essential in many research projects to analyze complex biological systems. The power of these approaches, which allow the determination of cell composition and delineation of complex transcriptional regulatory networks in cellular subsets, involves a large panel of numerical methods to integrate properly many different biological modalities (information about RNA expression, splicing, editing, proteins, epigenome, genome, spatiotemporal expression). Many questions are emerging when working such massive data to optimize properly the integration of several separate datasets and be able to make comparisons between different experimental situations. Deep learning methods show great promise to automate classification or dimensionality reduction. We plan to use them in order to build the cellular atlas of the respiratory tree, after identification of the most appropriate deep network structure able to identify the different cell types.

The research project aims to:

- (I) determine the most suitable model of deep neural network to automate cell classification;

- 
- (2) develop a visualization algorithm for dynamically annotate the cell types, linked to the deep learning model;
  - (3) use these approaches for creating a lung cellular atlas that integrates *in vivo* and *in vitro* datasets, and document the variations observed under specific phenotypic/genotypic situations.

---

## ***Exploring associations between serum cytokine levels and behavior abnormalities in children***

**Responsable scientifique :** Nicolas Glaichenhaus

**Laboratoire :** Institut de Pharmacologie Moléculaire et Cellulaire

### **Résultats obtenus grâce aux moyens informatiques attribués :**

Foiselle, M., S. Barbosa, O. Godin, C. L. Wu, W. Boukouaci, M. Andre, B. Aouizerate, F. Berna, C. Barau, D. Capdevielle, P. Vidailhet, I. Chereau, L. Davidovic, J. M. Dorey, C. Dubertret, J. Dubreucq, C. Faget, G. Fond, S. Leigner, P. M. Llorca, J. Mallet, D. Misrahi, E. Martinuzzi, C. Passerieux, R. Rey, B. Pignon, M. Urbach, F. Schurhoff, N. Glaichenhaus, M. Leboyer, R. Tamouza and F.-S. Groups (2022). "Immuno-metabolic profile of patients with psychotic disorders and metabolic syndrome. Results from the FACE-SZ cohort." *Brain Behav Immun Health* 22: 100436.

Barbosa, S., E. Martinuzzi, C. Gilet, L. Fillatre, O. Khalfallah, I. Sommer, S. Leucht, P. Dazzan, P. McGuire, C. Arango, C. M. Diaz-Caneja, W. Fleischhacker, D. Rujescu, B. Glenthøj, I. Winter, R. S. Kahn, R. Yolken, S. Lewis, R. Drake, L. Davidovic, M. Leboyer, N. Glaichenhaus and O. S. Group (2021). "Baseline Levels of C-Reactive Protein and Proinflammatory Cytokines Are Not Associated With Early Response to Amisulpride in Patients With First Episode Psychosis: The OPTiMiSE Cohort Study." *Schizophrenia Bulletin Open*, 2021, 10.1093/schizbulopen/sgab017. hal-03286370

### **Description :**

Les cytokines sont de petites protéines solubles produites par les cellules du système immunitaire. A ce titre, elles jouent un rôle déterminant dans la défense de l'organisme contre les agents pathogènes. Toutefois, ces molécules peuvent également perturber le fonctionnement du cerveau et provoquer des anomalies comportementales telles que l'hyperactivité ou des difficultés d'interactions sociales. Afin de mieux comprendre le rôle des cytokines sur le fonctionnement du cerveau, nous avons eu accès à des données cliniques et des prélèvements biologiques qui ont été collectés dans le cadre de la cohorte EDEN, une cohorte française généraliste dont l'objectif est l'étude des déterminants pré et post nataux précoce du développement psychomoteur et de la santé de l'enfant. Ainsi, dans le cadre de la cohorte EDEN, plus de 2000 mères ont été recrutées pendant leur grossesse et suivies, ainsi que leur enfant, pendant plus de 5 ans.

De très nombreuses données psychosociales ont été collectées pendant la grossesse et après la naissance de l'enfant parmi lesquelles, la consommation de tabac et d'alcool de la mère pendant la grossesse, les niveaux d'éducation du père et de la mère, le revenu du foyer, le nombre de frères et de sœurs, les antécédents médicaux des parents, les

---

événements de santé entre la naissance et l'âge de 5 ans, le poids et la taille de l'enfant. De plus, les enfants ont été évalués à l'âge de 5 ans par un(e) psychologue ce qui a permis de définir, pour chacun d'entre eux, 5 scores cliniques mesurant les performances de l'enfant dans 5 dimensions comportementales: les problèmes comportementaux, les problèmes d'interactions avec les autres, les problèmes émotionnels, l'hyperactivité et les problèmes d'inattention, et le comportement pro-social.

Parallèlement au recueil des données psychosociales, l'étude EDEN a permis de récolter des échantillons de sang qui ont été prélevés chez la mère pendant la grossesse, à la naissance dans le cordon ombilical, et chez l'enfant à l'âge de 5 ans. Ces échantillons ont ensuite été analysés pour la concentration de 33 cytokines différentes.

L'objectif de notre projet est d'effectuer des études d'association de prédiction, afin d'identifier les variables cliniques et biologiques associées aux différentes dimensions comportementales étudiées dans la cohorte EDEN.

#### Méthode(s) utilisée(s)

Nous utiliserons 3 méthodes/algorithmes de régression pour comparer les résultats :

1- La régression logistique standard dans la cadre des modèles linéaires généralisés qui utilise l'estimateur usuel des moindres carrés.

2- Elastic Net est une méthode de contraction des coefficients de régression qui permet d'ajuster des modèles de régression logistique pénalisée. Elastic Net est normalement utilisé dans des problèmes de grande dimension ( $n \ll p$ ), un cas où des méthodes plus classiques ne fonctionnent pas. Elastic Net utilise des algorithmes consommateurs de temps de calcul et de stockage. En pratique, Elastic Net est testé de manière itérative pour différentes valeurs des paramètres lamda et alpha de manière à identifier une solution optimale par validation croisée et une mesure de la performance par calcul de l'aire sous la courbe (AUC).

3 – Algorithmes de Gradient boosting (XGBoost, etc) pour créer des modèles prédictifs. L'objectif principal de Gradient Boost est de minimiser la fonction de perte en ajoutant des apprenants faibles à l'aide d'un algorithme d'optimisation de la descente de gradient.

---

L'optimisation des paramètres de l'algorithme peut prendre un certain temps même si celui-ci a été spécialement conçu pour améliorer la vitesse et les performances.

Exécuter ces algorithmes dans un PC prend du temps car la parallélisation est réduite à quelques cœurs et les performances générales de la machine seront réduites.

---

## **Détection de signatures dans les données brutes de séquençage**

**Responsable scientifique :** Olivier Croce

**Laboratoire :** IRCAN

**Description :**

Les séquences constituées de motifs répétés courts sont fréquemment trouvés dans les génomes des eucaryotes. Pourtant leurs emplacements dans les génomes n'est pas toujours bien compris. C'est le cas par exemple des motifs répétés en tandem comme SAT3 ou T2AG3 dont ce dernier se trouve majoritairement au niveau des télomères, mais qui a aussi été retrouvé dans d'autres secteurs du génome. Un autre problème apparaît également lorsqu'il s'agit d'estimer la proportion réelle de ces motifs dans les génomes. En effet, leur assemblage difficile fait qu'il existe des incertitudes quant à leur abondance réelle par rapport à celle estimée à partir des génomes de référence.

Nous souhaitons identifier et estimer l'abondance des motifs répétés directement à partir de jeux de reads issus de séquences WGS haut débit du type Illumina HiSeq X. L'objectif est ensuite d'établir des corrélations entre signatures SSR et annotations phénotypiques.

## **Etude des régions non structurées des protéines des sites de contacts membranaires**

**Responsable scientifique :** Romain Gautier

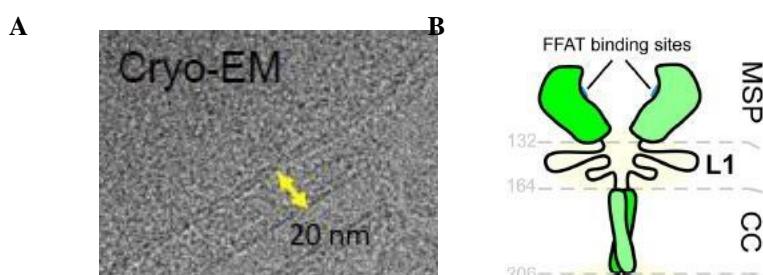
**Laboratoire :** IPMC

**Description :**

Introduction

De nombreuses protéines qui transportent les lipides de manière non-vésiculaire opèrent dans des régions subcellulaires étroites, appelées sites de contact membranaire (MCS). Dans les MCS, les membranes de deux organelles sont séparées d'environ 20 nm<sup>1</sup> (Figure 1A). Ces zones permettent la réalisation de diverses fonctions, comme l'échange lipidique par des protéines de transfert telle que OSBP, le positionnement des organelles dans la cellule, et bien d'autres. Les contraintes au niveau des MCS sont très variées. Par exemple, les MCS entre le réticulum endoplasmique (RE) et le trans-Golgi (MCS RE-Golgi) sont dynamiques et ont une durée de vie assez courte (quelques secondes), tandis que les MCS RE- mitochondries sont bien plus stables (plusieurs minutes à plusieurs heures).

Alors que les domaines protéiques bien repliés sont responsables de fonctions spécifiques telles que le transport de lipides ou les interactions avec les membranes, des résultats récents sur le complexe entre la protéine de transfert de lipides OSBP, et le récepteur général du RE VAP-A, suggèrent que des régions intrinsèquement désordonnées (IDR) influencent l'architecture et la dynamique des MCS (Figure 1B et 1C). Les IDRs du complexe OSBP/VAP garantiraient par exemple la flexibilité latérale et verticale des protéines des MCS<sup>2</sup>, ainsi que l'adaptation aux contraintes liées à l'organisation et la durée de vie des différents MCS<sup>3</sup>.



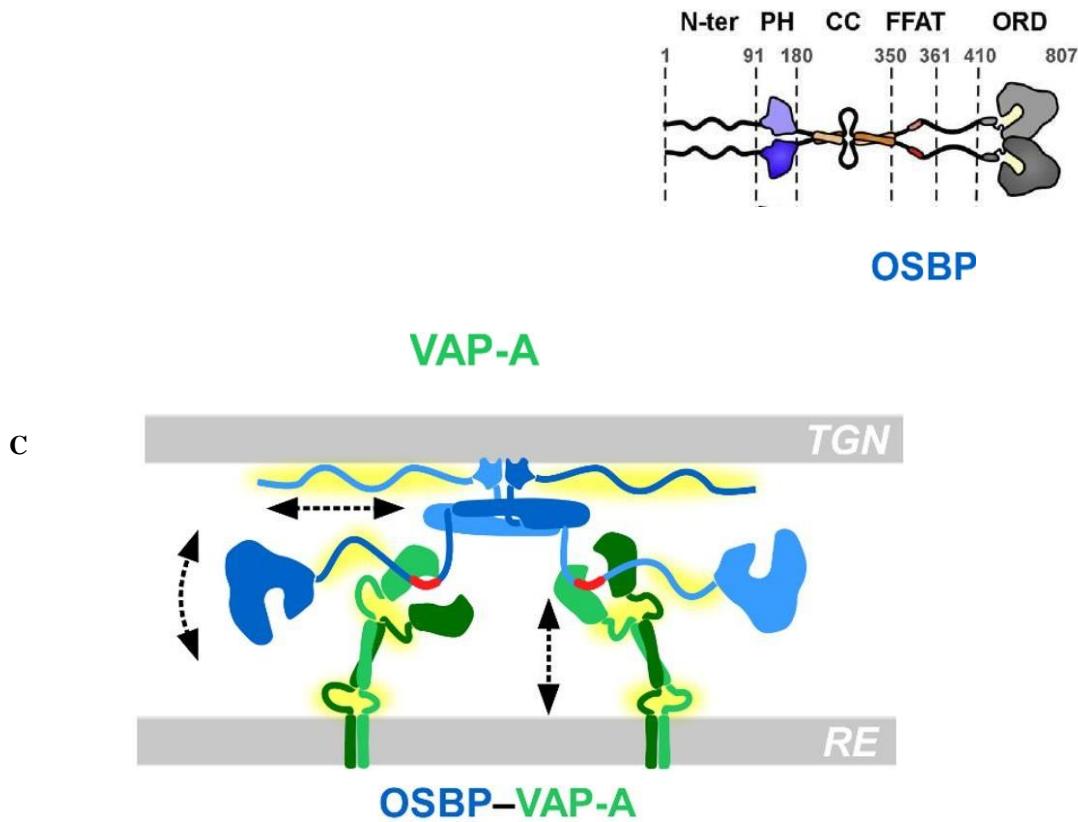


Figure 1: Le complexe OSBP/VAP-A dans les MCS. **A.** Image obtenue par Cryo-EM d'un site de contact membranaire (équipe Daniel Lévy). **B.** A gauche, architecture de VAP-A<sup>3</sup>. MSP = Major Sperm Protein Domain, le domaine de liaison à OSBP, CC = Coiled-Coil, TM = domaine transmembranaire, L1 = linker1 (IDR1), L2 = Linker2 (IDR2), ER = réticulum endoplasmique. Adroite, architecture de OSBP<sup>2</sup>. PH = Pleckstrin Homology domain, CC = Coiled-Coil, FFAT = Two phenylalanines in an acidic tract, le domaine de liaison à VAP-A, ORD = OSBP Related Domain, le domaine de liaison aux lipides, L1 = IDR terminale. **C.** Le complexe OSBP/VAP-A, un système flexible, les IDRs sont en jaune. RE = réticulum endoplasmique, TGN = Réseau trans-golgien.

D'autre part, nos études récentes suggèrent qu'il existe des hétérogénéités moléculaires et fonctionnelles entre VAP-A et son proche homologue VAP-B<sup>3</sup>, qui est également largement utilisé comme ancre par des protéines de transfert lipidique aux MCS. Nous avons remarqué que malgré un pourcentage d'identité élevée entre les parties structurées de ces deux protéines (~80%), leurs IDRs sont peu conservés (~40%). De plus, des mutations de VAP-B sont connues pour leur rôle dans le développement de la sclérose latérale amyotrophique (SLA de type 8), une maladie neurodégénérative ; et certaines de ces mutations sont localisées dans les IDRs<sup>4</sup>. De manière générale, comprendre les mécanismes mis en jeu dans les MCS est important car leurs dysfonctionnements sont impliqués dans de nombreuses maladies telles que le cancer, l'obésité, les troubles neurologiques et plusieurs infections virales (Hépatite C, Zika etc.)<sup>5</sup>

---

Le rôle des IDR est encore mal compris, et leur nature flexible rend difficile leur étude. Un sujet de thèse est mené depuis octobre 2022 dans notre laboratoire par la doctorante Zoé GRIMANELLI, portant, entre autres, sur l'analyse informatique des IDR du complexe OSBP/VAP. Le but est de décrire les caractéristiques, la fonction et l'organisation des IDR dans différents contextes de MCS, par une approche bioinformatique. Les simulations de dynamique moléculaire notamment permettent l'étude de la dynamique de molécules au niveau atomique. Elles représentent donc un outil intéressant dans l'étude des zones flexibles que sont les IDR, et permettront de développer les premiers modèles *in silico* de MCS.

## Projet

L'idée est d'utiliser les simulations de dynamique moléculaire sur des systèmes composés de VAP-A, VAP-B et OSBP avec des bicouches membranaires.

Nous souhaitons étudier l'influence des IDR de VAP-A sur son domaine de liaison à OSBP, appelé MSP (Major Sperm Protein domain) et plus largement sur son comportement général. Pour ce faire, des simulations de la protéine composée de ses IDR flexibles, ou mutés pour être plus courts ou plus rigides seront effectuées.

Nous souhaitons également déterminer l'influence de la composition lipidique sur les domaines transmembranaires de VAP-A et leur dimérisation. Pour cela, des membranes à diverses concentrations en cholestérol et composées de lipides à différents niveaux de saturation seront simulées avec VAP-A. Le motif de dimérisation présent sur le domaine transmembranaire (GXXXG) sera muté afin d'observer son impact.

Aussi, une étude comparative de VAP-A et VAP-B pourrait aider à comprendre le rôle des mutations observées sur les IDR de VAP-B dans le développement de la SLA. Nous simulerons des chimères VAP-A/VAP-B et modéliserons des mutations ponctuelles sur les IDR de VAP-A. Les résultats intéressants pourront guider les mutations à tester *in vitro*.

Ainsi, voici les systèmes que nous souhaitons simuler :

- VAP-A et VAP-B mutés au niveau du premier linker
- VAP-A et VAP-B mutés au niveau du deuxième linker
- VAP-A et VAP-B mutés au niveau des deux linkers
- VAP-A et VAP-B mutés au niveau du motif GXXXG

- 
- Différents types de chimères VAP-A/VAP-B
  - VAP-A et VAP-B dans des membranes à 0, 5 et 20% de cholestérol
  - VAP-A et VAP-B dans des membranes de compositions lipidiques différentes (à déterminer)

Soit une vingtaine de systèmes.

Nous avons pour l'instant réalisé 500 ns de simulations des dimères de VAP-A (Figure 2 gauche) et de VAP-B dans une bicouche membranaire, sur une carte GPU NVIDIA A40. Les systèmes tout atome, composés de 164332 atomes pour VAP-A, et 164186 pour VAP-B, ont été simulés en 25 jours soit une performance de 20ns/j. Le système gros-grain de VAP-A, composé de 78454 grains, a été simulé en 9h44 soit une performance de 1232ns/j (Figure 2 droite). Cependant, nous pensons que les systèmes étaient légèrement trop petits, et l'idéal est de faire des simulations de 1  $\mu$  s. Nous estimons donc 60 jours de calcul par système tout-atome sur nos cartes, soit pour nos systèmes :

20 systèmes x 3 répliques x 60 jours = 86.400 heures GPU,

+ essais gros-grains + marge d'erreurs ~ 100.000 heures GPU.

Nous utilisons l'outil de simulation de Dynamique Moléculaire Gromacs2021<sup>7,8</sup> compilé GPU (Cuda) avec les champs de force Chamm36 (pour les systèmes tout atome)<sup>9</sup> et Martini3<sup>10</sup> (pour les systèmes gros grain).

Nos ressources de calculs étant limitées, nous souhaitons déposer une demande d'accès régulier DARI A15 pour des heures disponibles au 1er novembre 2023. Nous avons obtenu 40 000 heures GPU en accès Dynamique sur la machine Jean-Zay (V100 et A100) en janvier 2023. Il a fallu faire des benchmarks sur différents systèmes et tester les paramètres Gromacs GPU sur les différents types de cartes. Nous avons consommé environ 40% de cette demande ce qui justifie notre demande sur Azzura en attendant l'accès régulier A15. Pour finaliser cette demande, nous espérons pouvoir effectuer des premières simulations grâce aux heures GPU de Azzura afin de tester les systèmes proposés. Nous demandons pour cela 50.000 heures GPU à Azzura. Ce complément d'heure nous est nécessaire pour la poursuite de ce projet de recherche.

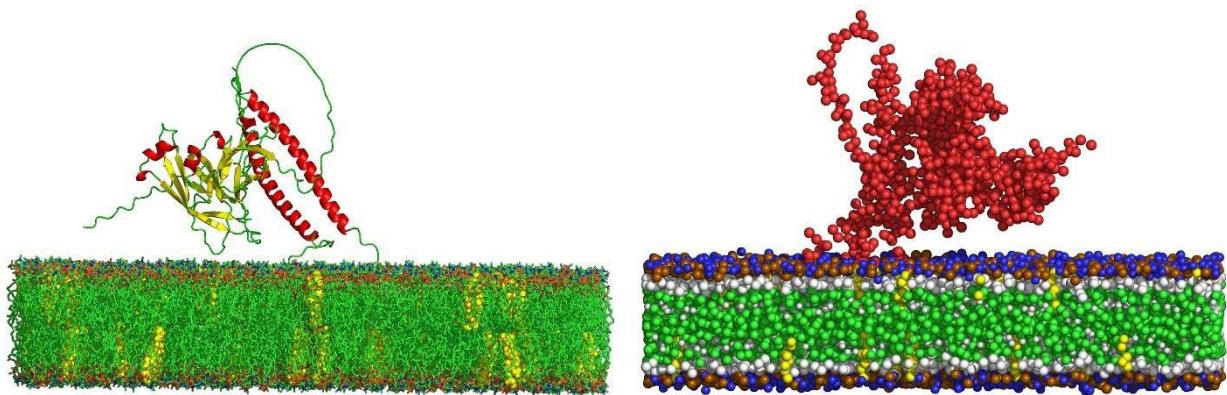


Figure 2 : système tout-atome (gauche) et gros-grain (droite) du dimère de VAP-A dans une bicouche membranaire, avant simulation

## Bibliographie

1. Prinz, W.A., Toulmay, A., and Balla, T. (2020). The functional universe of membrane contact sites. *Nat. Rev. Mol. Cell Biol.* 21, 7–24. 10.1038/s41580-019-0180-9.
2. Jamecna, D., Polidori, J., Mesmin, B., Dezi, M., Levy, D., Bigay, J., and Antonny, B. (2019). An Intrinsically Disordered Region in OSBP Acts as an Entropic Barrier to Control Protein Dynamics and Orientation at Membrane Contact Sites. *Dev. Cell* 49, 220-234.e8. 10.1016/j.devcel.2019.02.021.
3. Subra, M., Dezi, M., Bigay, J., Lacas-Gervais, S., Di Cicco, A., Dias Araújo, A.R., Abélanet, S., Fleuriot, L., Debayle, D., Gautier, R., et al. (2022). VAP-A intrinsically disordered regions enable versatile tethering at membrane contact sites. *bioRxiv*, 2022.05.13.491839. 10.1101/2022.05.13.491839.
4. Dudás, E.F., Huynen, M.A., Lesk, A.M., and Pastore, A. (2021). Invisible leashes: The tethering VAPs from infectious diseases to neurodegeneration. *J. Biol. Chem.* 296. 10.1016/j.jbc.2021.100421.
5. Strating, J.R.P.M., van der Linden, L., Albulescu, L., Bigay, J., Arita, M., Delang, L., Leyssen, P., van der Schaar, H.M., Lanke, K.H.W., Thibaut, H.J., et al. (2015). Itraconazole inhibits enterovirus replication by targeting the oxysterol-binding protein. *Cell Rep.* 10. 10.1016/j.celrep.2014.12.054.
6. Nchoutmboube, J., Ford-Siltz, L.A., and Belov, G.A. (2015). Enterovirus replication: GO with the (counter)flow. *Trends Microbiol.* 23. 10.1016/j.tim.2015.02.008.

- 
7. Bekker, H., Berendsen, H., Dijkstra, E.J., Achterop, S., Drunen, R., van der Spoel, D., Sijbers, A., Keegstra, H., Reitsma, B., and Renardus, M.K.R. (1993). Gromacs: A parallel computer for molecular dynamics simulations. *Phys. Comput.* 92, 252–256.
8. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
9. Huang, J., and Mackerell, A.D. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* 34, 10.1002/jcc.23354.
10. Souza, P.C.T., Alessandri, R., Barnoud, J., Thallmair, S., Faustino, I., Grünewald, F., Patmanidis, I., Abdizadeh, H., Bruininks, B.M.H., Wassenaar, T.A., et al. (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* 18, 10.1038/s41592-021-01098-3.

---

***Investigating the genome and transcriptome dynamics of early interspecific hybridization***

**Responsable scientifique :** Gianni Liti

**Laboratoire :** IRCAN

**Description du projet en pages suivantes.**

## Investigating the genome and transcriptome dynamics of early interspecific hybridization

### 1. Excellence

#### 1. *Quality and pertinence of the project's research and innovation objectives.*

**Background** The proposed project will investigate the genomic and gene expression dynamics of the early stages of interspecies hybridization using *Saccharomyces* yeasts as a model. Hybridization has shaped the evolutionary history of every kingdom of life: it is estimated that 2-5 percent of the human genome derives from archaic admixture events and that upwards of 10 percent of the genome of certain fish, plant, butterfly, and avian taxa originate from interspecies hybridization events<sup>1</sup>. Hybrids are more than the additive combination of their parental founders: by combining evolutionarily diverged genomes, hybridization leads to innovations through the emergence of new genetic interactions and phenotypes<sup>2</sup>. Hybrids can display superior phenotypes when compared to their founding parental lineages. This hybrid vigour has been widely exploited to improve agricultural crops<sup>3</sup>. However, many hybrids either do not form, or if they do, they are evolutionary dead-ends due to their intrinsic sterility, likely deriving from the high sequence divergence between parental subgenomes and incompatibility between subsets of genes<sup>4</sup>. The biological features and evolutionary mechanisms that contribute to the variation in the occurrence, persistence, and evolution of hybrids remain poorly understood in both plants and animals. **Understanding hybridization has far-reaching impacts** 1) to understand the evolutionary history of organisms 2) to characterize the genetic interactions and genome structural organization that drive gene expression and associated phenotypes 3) increase phenotypic diversity for biotechnologies.

**Problem statement:** Hybrid genomes carry signatures of severe past genomic instability such as chromosomal rearrangements, repeat amplification, widespread transposable element activation, chromatin reorganization, and gene expression misregulation. Merging highly diverged parental subgenomes that have independently evolved for millions of years is theorized to trigger “genomic shock”, a response to an unanticipated genomic challenge that leads to genome restructuring and destabilization of gene regulation<sup>5</sup>. Surprisingly, many studies of hybrid genomes have found that hybrid genomes are remarkably stable. The supervisor, Dr. Gianni Liti, found that mutation rates of yeast interspecific hybrids after 2500 mitotic generations were comparable to those of pure parental backgrounds<sup>6</sup>. This is unexpected, given the signatures of past genomic instability that typifies hybrids<sup>6</sup>. Thus, there is a paradox in the understanding of hybrid genome dynamics, that must be reconciled to achieve a mechanistic comprehension of hybridization events.

**Hypothesis:** To reconcile the genome stability of late hybrids with signatures of past genome instability, I hypothesize that hybrid genome evolution is not constant over time. I propose that hybrids experience an initial transient genome shock soon after zygote formation which manifests as the observed high genome instability and lethality. This “shock phase”, caused by negative epistatic interactions and regulatory incompatibilities between parental subgenomes, likely brings revolutionary changes that stabilizes many of the incompatibilities between the diverged subgenomes, and is followed by a stabilization phase. In this second phase, hybrid genomes are remarkably stable and recombination events are suppressed by the sequence divergence between the two subgenomes. Hybridization studies in eukaryotes have largely focused on extant hybrids with undefined evolutionary histories. I propose to investigate genomic instability in the earliest post-zygotic phases, which have been largely unexplored.

**The Objective** is to investigate early hybrid history through a comprehensive characterization of the implications of genome divergence on hybrid viability, genome stability, and gene expression. I propose to generate various interspecies hybrids across a range of sequence divergence and employ cutting-edge DNA and RNA sequencing to investigate the genomic and transcriptomic consequences of early hybridization over evolutionary time. The objective is subdivided as follows: **RO1: Studying the early phase of hybridization using a reporter-based assay to characterize heterogeneity of hybrid colonies (WP1), RO2: Genome-wide investigation of genome instability signatures in hybrids (WP2), RO3: Gene expression dynamics of early hybridization (WP3).**

**Original scientific questions:** The question of when genomic instability arises in the evolutionary history of interspecies hybrids remains unresolved as studies of interspecies hybrids after 50 to 2000 mitotic generations have found hybrid genomes to be stable<sup>6</sup>. This proposal has the ambition to resolve key aspects of the consequences of hybridization that have fascinated biologists for many decades. The central goal of advancing our understanding of how hybrid genomes emerge, evolve and overcome incompatibilities is at the forefront of the fields of genetics and evolution. I propose to investigate genomic instability in the

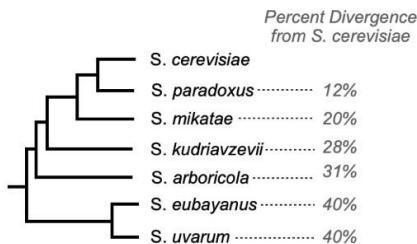


Figure 1: Schematic representation of the phylogeny of the genus *Saccharomyces*. The percentage divergence from *S. cerevisiae* is indicated on the right.

earliest post-zygotic phases that have, to date, been largely unexplored making insights gained from the exploration of this hybridization phase highly impactful.

**Innovative experiments:** This is the first study where such a comprehensive panel of hybrids will be used to mechanistically investigate hybrid genome instability. I will investigate thousands of single-cell zygotes in a time-resolved manner, evolve them under defined conditions, and monitor their genome evolution with base pair resolution to provide an unprecedented view of genome stability, cell viability, and gene expression from the onset of hybridization. This is a considerable departure from current knowledge based on hybrids commonly generated

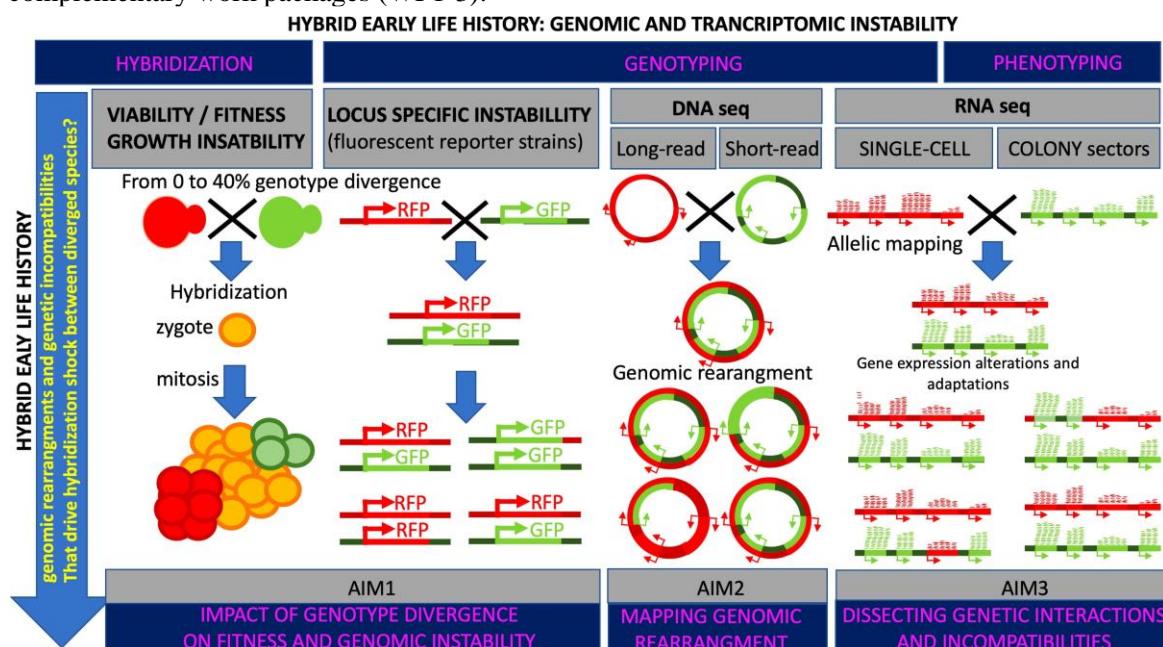
by mass-mating and marker selections.

**Cutting edge methods:** Methodologies employ state-of-the-art computational and molecular approaches that the Liti lab and others have pioneered. I will utilize an approach that I have recently adopted in the lab that allows me to track hybrid yeast cells deriving from single mating events over each mitotic generation considering both genomic and gene expression consequences of hybridization. The combination of the microscopy fluorescence assay and single-cell, short and ultra-deep long-read sequencing is at the cutting-edge of studies in interspecies hybridization and genome dynamics studies and will enable us to monitor genome instability during this transient developmental phase. Applying the “omics” approaches to the expanding zygote will reveal novel insights on the molecular dynamics and factors that contribute hybrid incompatibilities.

## 1.2 Soundness of the proposed methodology

**Model:** To investigate our hypothesis that genome stability is not constant over time, a time-resolved analysis of hybridization starting with zygote formation is essential. However, this approach is unfeasible in most species due to ethical reasons, the length of generation time, low recombination rate and the practical constraints of maintaining, breeding and genotyping large population sizes.

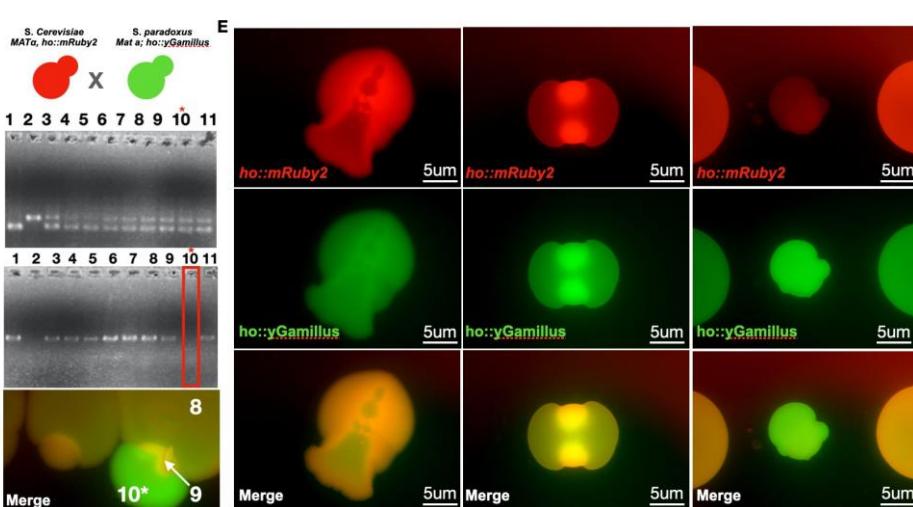
**Saccharomyces yeasts are an ideal model system to mechanistically study hybrids** as they can form despite extreme sequence divergence<sup>4</sup>. *Saccharomyces* hybrids between diverged parents naturally occur and have numerous industrial applications<sup>78</sup>. Hybrids can be generated and maintained in the lab, evolved under defined conditions with precise control over their sexual and asexual phases, and their genome evolution can be analyzed in real-time. Thus, yeasts offer a unique opportunity to combine genetics, genomics, and experimental evolution to investigate the very early hybrid phases. **This represents a major innovation** as time-resolved studies of early hybridization have never been conducted. The experimental plan consists of three distinct, but complementary work packages (WP1-3).



**WP1. Studying the early phase of hybridization using a reporter-based assay to characterize heterogeneity of hybrid colonies (Months 1-12).** Haploid cells from distinct *Saccharomyces* species will be combined to form interspecies hybrids<sup>9</sup>. I propose to pair haploid cells of opposite mating types to generate independent hybrids that cover the full range of sequence divergence, and track arrays of individual matings in a time-resolved manner starting from zygote formation. This is in contrast to the more commonly used mass mating approaches used to generate hybrids. To generate interspecies hybrid combinations with a broadrange of sequence variation, I will pair parental strains that diverge from 12% (*S.cerevisiae* x *S.paradoxus*)to 35% (*S.cerevisiae* x *S.eubayanus*) (figure1)<sup>9101112</sup>. Intraspecies heterozygous hybrids with <1% divergencewill be generated as controls for all the experiments described below. All experiments outlined in this proposal will be conducted in early phase hybrids from 0 (time of zygote formation) to 25 mitotic generations.

**WP1.1 assessing loss of heterozygosity (LOH) and aneuploidy rates at targeted loci (Months 1-8).** I will first measure genomic instability in clonally expanding hybrids by engineering haploid, parental strains, with fluorescent markers at the native HO locus. This locus will contain the heterozygous combination of a red and green fluorescent markers in resulting diploids. URA3 mutant and LYS2 mutant auxotrophic marker alleles will be swapped into the native URA3 and LYS2 loci to allow for the selection of on media lacking uracil and lysine. Similar assays have been extensively used in the lab<sup>131415</sup>. Fluorescent marker alleles will permit high-throughput microscopic screening for the emergence of exclusively green or red sectors that denote the occurrence of early-phase (~25 generations) genetic instability such as LOH, or aneuploidies in our various hybrid clones. PCR analysis will be used to determine gene loss (as demonstrated in fig2), qPCR will be used to determine iff luorescence instability phenotypes can be observedat the RNA level. Preliminary data indicates that there is indeed genome instability ine a r l y *Saccharomyces* interspecies hybrids relative to diploid parental strains (fig 2 ). As a hybrid

incompatibilities can result in ploidy changes I will also assay for ploidy. Contour-clamped homogenous electric field (CHEF) gel electrophoresis<sup>16</sup>, will be used to assess karyotes of hybrid clones to determine if hybrids have signatures of instability elsewhere in the genome. **Impact:** This reporter assay will provide rapid, systematic quantification of genomic instability in early interspecies hybrids which remains a mystery. These assays can also be used to whether instability signatures in hybrids stabilize with increasing mitotic generations as we hypothesize.



incompatibilities can result in ploidy changes I will also assay for ploidy. Contour-clamped homogenous electric field (CHEF) gel electrophoresis<sup>16</sup>, will be used to assess karyotes of hybrid clones to determine if hybrids have signatures of instability elsewhere in the genome. **Impact:** This reporter assay will provide rapid, systematic quantification of genomic instability in early interspecies hybrids which remains a mystery. These assays can also be used to whether instability signatures in hybrids stabilize with increasing mitotic generations as we hypothesize.

**Figure2:** Genome instability signatures of early interspecific hybrids using fluorescence reporter assay. A) Schematic of interspecies hybrid cross of *S. cerevisiae* and *S. paradoxus* B) Mating type PCR showing haploid parents lane1=Sc.mruby, lane2=Sp.yGam. and diploid hybrid lanes3-11 C) mRuby2 PCR. lane1=Sc.mRuby, lane2=Sp.yGam. Lanes3-9 and lane 11=interspecies, double fluorescent diploid hybrids. Lane10=colony sector that lost red fluorescence. No mRuby band in lane10 (diploid hybrid) is indicative of genome instability in the interspecies hybrid. D) Interspecies hybrid colony where samples 8-10 were taken. Numbers labelled indicate colony sector PCR samples derive. E) Examples of putative interspecies hybrid genome instability signatures from fluorescent marker assay.

**WP1.2 Evaluating impact of genomic divergence on hybrid viability (Months 2-10).** I will assess the viability of hundreds of clonally expanding diploid hybrids by using the characteristic morphology of zygotes to identify and re-array their resulting daughter cells<sup>1718</sup>. The percent of re-arrayed daughter cells thatsurvive from the interspecies cross will be compared to parental control crosses. By only selecting daughter cells from zygotes we can ensure that the viability rates are not impacted by lower mating efficiencies that can arise from highly diverged parental founder strains. **Impact:** Evaluating the impact of genomic divergence on zygote viability is a critical step towards understanding how genomic interactions constrain hybrid evolution.

**WP1.3-Assessing impact of environment on instability and viability of interspecies hybrids (Months 2-12).** Genomic instability in hybrids may be driven by environmental factors. To determine whether environmental stress can contribute to genomic instability in hybrids I will determine viability and instability

rates of the clonally expanding hybrids presented in WP1.1 after subjugating hybrids to growth in respiratory media, fermentation media, high and low temperatures. **Impact:** These experiments will elucidate how stress and selection shape the viability and instability of asexually reproducing hybrids over evolutionary time in early hybrid history. These assays will provide rapid, systematic quantification of genomic instability.

**WP2. Genome-wide investigation of genome instability signatures in hybrids (Months 7-22).** Here, I propose to characterize genome stability genome-wide in early-stage hybrids.

**Objective 2.1- Genome-wide investigation of genome instability signatures in hybrids (Months 7-22).** To determine when and where instability in the genome arises in the first 25 mitotic generations of the expanding zygote, I will map recombination events using long-read ultra-deep sequencing to identify balanced and unbalanced structural variation. MinION technology will be employed. Given the high level of divergence between the two parental subgenomes, I can unambiguously assign each read to a parental genome. Because chimeric reads are indicative of genomic rearrangement I can detect low-frequency genome instability in the expanding zygote population by searching for chimeric reads that bridge the two parental subgenomes. Short-read Illumina sequencing will be used for SNP discovery. I will conduct these experiments in expanding clones of the various hybrid combinations. The lab has extensive expertise in obtaining comprehensive genomic annotation of long and short-read DNA sequencing and using these datasets to track genome instability in diploid hybrids<sup>6919</sup>. **Impact:** By combining these two techniques I can globally assess genomic changes in hybrids, when they occur, and how they are dispersed across the genome. This is a major advance as it will represent the first genome-wide analysis of early phase hybrids.

**Objective 2.2-Single-cell sequencing to detect genome instability (Months 10-22).** 10X genomic droplet-based single-cell RNA sequencing technology has recently been optimized to allow analysis in yeast<sup>20</sup>. As there are only 2 genome copies per cell in diploid hybrids, allelic dropout and non-uniform sequencing depth remain major drawbacks for utilizing single-cell DNA sequencing technologies. Instead, to capture the heterogeneity of genome instability in the hybrid population I will conduct single-cell RNA-sequencing from the various parental founder cross combinations outlined above. While, single-cell RNA sequencing will not provide a fine-grained genotype view of the hybrids, LOH and aneuploidies will be detectable. **Impact:** The single-cell RNA-Seq will allow us to assess both genome stability and genomic heterogeneity across populations of cells with single-cell resolution. These experiments will allow variation in genome dynamics of hybrid populations over evolutionary time to be captured. They will establish whether recombination is stochastic and randomly dispersed throughout the genome or whether individuals or subpopulations of hybrids converge towards similar genomic changes. For the first time interspecific hybridization will be studied with single-cell resolution.

**WP3. Gene expression dynamics of early hybridization (Months 9-22):** The new genetic interactions and allelic combinations of highly heterozygous hybrids may disrupt the RNA expression programs of the parental founders and consequently trigger transcriptome shock which would manifest as differential gene expression from their parental founders. Strong divergence from parental gene expression could be counter-selected on, or new expression profiles, that did not exist in parental strains, could be positively selected upon, leading to the emergence of phenotypes not found in the parental strains. Thus, gene expression analysis can reveal the evolvability of hybrids. I will profile RNA expression of hybrid combinations described in WP1.1 at distinct time-points from the onset of hybridization to capture differential gene expression patterns between parental founders and hybrid clones (outlined WP1.1) over increasing mitotic generations. I will assess the molecular functions and biological processes of RNAs with similar transcriptomic regulation patterns to determine if specific complexes, pathways or functions are modulated in the early hybrids. Analyses will consider allelic specific expression as it may provide key mechanistic insights into genetic interactions between subgenomes. **Impact:** Altogether, time-resolved RNA-Sequencing will serve as a highly quantitative approach to map the evolution of gene expression in yeast early hybrid history, and divulge whether the two parental subgenomes converge toward similar transcript levels to enter a stabilization phase. This work will reveal the gene expression alterations that highly divergent hybrids must overcome to reach genomic stability. Beyond the contributions to the field of evolution, interspecies hybrids are powerful tools to dissect the genetic interactions that control gene expression networks, as hybrids provide a means to explore how variation in organization and sequence can impact the interactions of various complexes and supramolecular assemblies.

---

## **Tests et analyse bio informatique**

**Responsable scientifique :** Arnaud Hubstenberger

**Laboratoire :** IBV

**Description :**

La plateforme de bio informatique de l'iBV aide les équipes de l'iBV dans leurs analyses de données. La première année a été une période d'essai et la volonté est maintenant d'utiliser systématiquement Azzurra pour les nouveaux projets soumis à la plateforme . L'utilisation se fera en interactif, avec des notebook Jupyter.

L'objectif est d'utiliser par la suite la plateforme pour des analyses omiques et d'images.

Les projets envisagés sont:

- \* Analyse du développement de cellules cancéreuses à partir de données "single cell RNAseq" pour l'équipe de Thierry Virolle
- \* Analyse de l'expression de gènes au cours du développement embryonnaire de la souris à partir de données "single cell RNAseq" pour l'équipe d'Andreas Schedl
- \* D'autres projets pourront également être traités suivant les demandes en cours d'années.

---

## Droit, économie

**EUR ELMI, LexSociété**

---

## ***Estimation de modèles complexes en économie et finance à base de simulations***

**Responsable scientifique :** Veronika Czellar

**Laboratoire :** INSTAR, SKEMA

### **Résultats obtenus grâce aux moyens informatiques attribués :**

The Azzurra computing facilities allowed us to develop a powerful model that accurately captures cryptocurrency returns for the five major cryptocurrencies: Bitcoin, Dash, Ethereum, Litecoin and Ripple. Results are summarized in the working paper described below. To extend our results, we now need to estimate the returns of the next five major cryptocurrencies, i.e. Doge, Ethereum Classic, Monero, Nem and Stellar.

Title: Multifractal Cryptocurrencies

Authors: Bastien Buchwalter, Veronika Czellar and Engin Iyidogan

Abstract:

We propose a Markov-switching multifractal (MSM) model for the high-frequency return of cryptocurrencies. Maximum likelihood estimation in MSM would be prohibitively long to compute, especially in large samples such as high-frequency data. We therefore use Approximate Maximum Likelihood (AML; Czellar, Frazier, and Renault 2022) which is computationally feasible even in high-frequency data. Using AML we estimate all the parameters of the MSM model, including the number  $k$  of hidden states. The extended model has powerful out-of-sample performance over its GARCH-type alternatives. Finally, we create a trading strategy based on dynamic Sharpe ratio signals generated through the MSM-based model at high-frequency. The strategy generates substantial returns in high trading fee regimes of cryptocurrency exchanges even in the presence of execution delays.

The paper has been/will be presented at the following conferences/seminars:

October 23, 2023: Fifth International Workshop in Financial Econometrics, Vila Angatu Eco Resort, Santo André (Brazil)

June 15, 2023: Conference on Forecasting Financial Markets: ADVANCES FOR EXCHANGE RATES, INTEREST RATES AND ASSET MANAGEMENT, Rennes (France)

March 29, 2023: French Inter-Business School Conference and Forecasting Financial Markets Conference, Toulouse (France)

January 24, 2023: Finance Seminar, University of Luxembourg (Luxembourg)

---

January 17, 2023: Finance Seminar, Rennes School of Business, Rennes (France)

November 28, 2022: SKEMA Finance Seminar, Paris (France)

November 18, 2022: ESSEC Asset Pricing Breakfast, Paris (France)

### **Description :**

Le but du projet est de développer et d'utiliser des méthodes à base de simulations qui construisent des modèles capables de reproduire les données et de les prédire. Les applications en économie et finance sont particulièrement gourmandes en calculs car les données financières utilisées sont de grande taille, et ces méthodes à base de simulations nécessitent la simulation de pseudo-données de taille similaire, puis la minimisation de la distance entre les pseudo-données simulées et les vraies données.

---

## **OCR chinois**

**Responsable scientifique :** Frédéric Constant

**Laboratoire :** ERMES, UFR de Droit

### **Description :**

Les membres du groupe de recherche/du projet Numerica Sinologica se sont réunis à partir du constat partagé de la nécessité de développer des outils de reconnaissance de l'écriture manuscrite et xylographique spécifiques au chinois classique, et plus largement aux langues asiatiques utilisant des sinogrammes, comme préalable au développement des humanités numériques dans les études sur l'histoire et les langues de l'Asie orientale. Les chercheurs associés au projet sont membres des trois principales unités de recherche sur la Chine impériale en France, le CCJ (UMR 8173 Chine, Corée, Japon), le CRCAO (Centre de recherche sur les civilisations de l'Asie orientale, UMR 8155) et l'IAO (Institut d'Asie orientale, UMR 5062). Au-delà des laboratoires concernés, des centres de recherche et des bibliothèques en Europe nous ont également fait part de leur intérêt et de leur soutien à notre projet.

Les implications du développement de modèles de reconnaissance de caractères fiables pour notre discipline sont très larges et permettraient à la sinologie de véritablement prendre le tournant des humanités numériques et de rattraper son retard par rapport à d'autres champs des sciences humaines et sociales. Outre la constitution de bases de données, l'analyse quantitative, la fouille de textes ou encore le traitement automatique du langage, de tels outils permettraient de diffuser auprès du grand public des collections de documents dont certaines ne sont pour le moment accessibles que via la souscription d'abonnements auprès d'organismes privés. Au-delà de l'innovation technologique, il y a donc un véritable enjeu scientifique et de diffusion du savoir. Notre projet s'inscrit par ailleurs dans une démarche de science ouverte, et les codes informatiques seront rendus publics sans aucune restriction (par exemple sur GitHub).

La reconnaissance automatique de l'écriture manuscrite du chinois fait face à des contraintes particulières qui expliquent pourquoi, alors que des plateformes (Transkribus, eScriptorium, Calfa Vision) existent déjà pour les écritures de type alphabétique, rien de tel n'est pour le moment accessible à la communauté des sinologues. Cette situation est due tout d'abord aux spécificités du chinois, pour lequel il

est difficile de transposer les méthodes déjà utilisées pour d'autres langues sans les adapter. Le chinois est composé de milliers de signes ce qui nécessite une puissance de calcul beaucoup plus importante que pour une langue alphabétique ou de développer de nouvelles approches afin d'entraîner des modèles performants. Les outils existants ont en outre été conçus pour des écritures horizontales, ce qui oblige à développer des modèles de reconnaissance des zones de texte et lignes de base adaptés aux écritures verticales. Le contexte académique est également peu favorable, la reconnaissance des textes en chinois classiques n'a pour le moment pas été jugée prioritaire en France ou en Europe, et les chercheurs chinois n'ont de leur côté pas encore développé de solutions efficaces et accessibles librement à la communauté des chercheurs. C'est ainsi que les projets existants offrent soit des résultats limités, soit des solutions propriétaires et extrêmement coûteuses, dont il est par ailleurs difficile d'évaluer les performances.

Notre groupe s'est constitué en mai 2021, en collaboration avec le projet eScriptorium, dirigé par Daniel STOEKL (EPHE), qui a développé une infrastructure de reconnaissance optique des langues à partir de Kraken, un logiciel open-source d'OCR basé sur l'intelligence artificielle et développé pour les langues écritures non-latines. Nous avons, grâce à leurs outils, entraîné tout d'abord des modèles de segmentation et d'analyse de la mise en page des textes chinois afin d'identifier les zones de texte à reconnaître dans les documents analysés. Nous avons également constitué un jeu de données de correspondances entre des images et leur transcription suffisamment large pour couvrir l'ensemble des caractères chinois usuels dans les sources classiques. Cette étape préalable à l'entraînement d'un premier modèle a pu être financée par des contributions financières du CCJ, du CRCAO et de l'IAO. Au cours des prochaines semaines, les ingénieurs d'eScriptorium modifieront Kraken afin de l'adapter aux spécificités du chinois, et nous serons alors en mesure d'entraîner un premier modèle. Fort de ce modèle nous serons en mesure, en fonction de son efficacité, de démontrer la viabilité de notre démarche et notre capacité à développer un outil performant.

En l'état, les infrastructures existantes ne permettront cependant pas à des utilisateurs inexpérimentés (c'est-à-dire ne maîtrisant pas les langages informatiques) d'utiliser ce modèle pour obtenir la transcription d'un corpus d'images donné, encore moins de l'adapter pour un usage spécifique, en raison par exemple des caractéristiques propres à un type particulier de sources (variété des écritures manuscrites et xylographiques). Nous souhaitons donc aller plus loin et offrir à toute personne la possibilité d'utiliser sans contraintes techniques ces outils (comme le fait par exemple Transkribus et son outil de saisie assistée par ordinateur permettant l'entraînement de modèles par l'utilisateur).

---

Notre objectif est ainsi de nous organiser pour réunir les fonds nécessaires pour: 1. finaliser les phases d'entraînement nécessaires aux perfectionnement d'un premier modèle ; 2. proposer un outil permettant la reconnaissance de caractères chinois anciens et le développement de modèles adaptés aux besoins des chercheurs sans qu'il soit nécessaire d'avoir des compétences informatiques particulières. Cet outil pourrait prendre la forme soit d'un portail en ligne offrant à la fois une interface facile d'accès et une puissance de calcul suffisante, soit d'une suite logicielle pouvant être installée sur un poste de travail et réutilisée au sein des laboratoires à l'aide d'un support technique classique.

---

# OPAL

Azzurra fait partie de l'infrastructure OPAL, dont une présentation peut être consultée [sur cette page](#).

Au cours de l'année 2023, **156** utilisateurs ont pu bénéficier d'une accréditation OPAL **par l'université** (+13,87 % par rapport à 2022), parmi lesquels :

- **107** utilisateurs disposant d'un compte sur Azzurra pour un projet en cours de validité.
  - **4** d'entre eux ont aussi demandé un compte sur le cluster Licallo (OCA)
  - **13** d'entre eux ont aussi demandé un compte sur le cluster NEF (Inria)
- **49** utilisateurs sans compte sur Azzurra, dans le but de calculer sur un autre cluster OPAL (tous sur NEF). Ce sont **6** projets de recherche et **1** cours, décrits en page suivante.

**16** personnels accrédités par une autre entité (11 par OCA et 5 par Inria) disposent quant à eux d'un compte sur Azzurra.

L'accréditation OPAL donnée par UCA est valable un an, renouvelable après [envoi d'un rapport des publications de l'année écoulée](#).

---

## **Projets accrédités OPAL par l'université ne calculant pas sur Azzurra**

---

## ***ANR TRACTIVE: AI-driven video interpretation of character representation in films***

**Responsable scientifique :** Lucile Sassatelli

**Laboratoire :** Laboratoire I3S - Université Côte d'Azur, CNRS

**Plateforme OPAL :** NEF (Inria)

**Nombre d'accrédités :** 3

### **Description :**

Surveys on the quantitative representation of women in visual media are insufficient to grasp the issue of gender inequality. In film studies, techniques of mise-en-scène have been identified to represent a character as an object of desire rather than a subject of action. These cinematographic gestures are difficult to pinpoint but correspond to complex, subtle and wide-spread visual discourse patterns, that may convey biased gender representation. TRACTIVE is a four-year research project funded by the French National Research Agency (ANR) and involving six academic laboratories in computer science and social science and humanities. The scientific objective of TRACTIVE is to characterize and quantify gender representation and objectification in films and visual media by designing AI-driven multimodal (visual and textual) discourse analyses.

To address this objective, TRACTIVE brings together researchers from computer science, media studies, linguistics, and gender studies. Learning approaches are taken to address the task of multimodal objectification detection, based on graph-based transformers and pre-trained models for weakly-supervised learning.

The project involves the computer science partners I3S and Inria in Sophia Antipolis, and access to OPAL is meant to grant access to computing resources of NEF of staff (engineer and postdoc) hired at I3S, as their TRACTIVE collaborators at Inria.

---

## **Inference-aware federated learning**

**Responsable scientifique :** Giovanni Neglia

**Laboratoire :** Inria, NEO team

**Plateforme OPAL :** NEF (Inria)

**Nombre d'accrédités :** 1

**Description :**

An increasing number of applications rely on complex inference tasks that are based on machine learning (ML). Currently, there are two options to run such tasks: either they are served directly by the end device (e.g., smartphones, IoT equipment, smart vehicles), or offloaded to a remote cloud. Both options may be unsatisfactory for many applications: local models may have inadequate accuracy, while the cloud may fail to meet delay constraints. During Tareq Si Salem's PhD thesis, we proposed the novel idea of inference delivery networks (IDNs), networks of computing nodes that coordinate to satisfy ML inference requests achieving the best trade-off between latency and accuracy. IDNs bridge the dichotomy between device and cloud execution by integrating inference delivery at the various tiers of the infrastructure continuum (access, edge, regional data center, cloud).

We are currently working on the distributed training (federated learning) of such inference delivery networks taking into account the statistical heterogeneity due to the inference forwarding process: less (resp. more) powerful clients will tend to see simpler (resp. more difficult) samples at inference.

---

## **Design de metasurfaces pour système flashlidar**

**Responsable scientifique :** Stéphane Lanteri

**Laboratoire :** Inria

**Plateforme OPAL :** NEF (Inria)

**Nombre d'accrédité :** 1

### **Description :**

Les métasurfaces sont des composants optiques planaires constitués d'une assemblée de structures sub-longueur d'onde réparties sur une grille dont la période est également inférieure à la longueur d'onde d'illumination. Ces structures ont des tailles, des formes et des orientations différentes, afin que l'onde qui traverse la métasurface puisse accumuler une phase qui n'est pas la même en tout point de l'interface. La modulation du profil de phase permet ainsi de contrôler le comportement de la lumière en sortie. Selon l'application souhaitée, il suffit d'adapter ce profil afin de fabriquer une lentille pour focaliser la lumière, ou un élément optique diffractif pour projeter des images holographiques, ou pour ajuster la réflexion, la réfraction, la polarisation de l'onde, ... Ce projet a pour but de concevoir et de fabriquer des métasurfaces qui seront intégrées à des dispositifs d'imagerie LIDAR (light detection and ranging), le tout, en ayant les efficacités les plus élevées possibles. Il s'inscrit dans le cadre d'une collaboration en cours entre l'équipe de Patrice Genevet au CRHEA et l'équipe Atlantis du Centre de Recherche Inria d'Université Côte d'Azur, avec Stéphane Lanteri comme correspondant. L'accès à ces ressources de calcul nous permettra de réaliser des simulations électrodynamiques type FDTD (Finite Difference Time Domain), DGTD (Discontinuous Galerkin Time Domain) ou bien encore GDM (Green Dyadic Method) pour déterminer les propriétés optiques des structures qui constituent les métasurfaces. Ces simulations sont très coûteuses en temps de calcul et en mémoire RAM, cependant elles peuvent être fortement accélérées avec la parallélisation des codes. D'autre part, nous envisageons d'utiliser une approche de deep learning pour accélérer la prédiction des propriétés optiques de nos composants micrométriques. La possibilité d'accéder à un grand nombre de nœuds GPU, à disposition au centre de calcul Nef, serait un réel atout pour accroître la vitesse d'apprentissage des réseaux neuronaux.

---

## **Seismic imaging of the Alps (LISALPS project - ANR)**

**Responsable scientifique :** Stéphane Operto

**Laboratoire :** Geoazur

**Plateforme OPAL :** NEF (Inria)

**Nombre d'accrédités :** 2

### **Description :**

Scientific objective:

The goal of this project is to develop new 3D quantitative models of the crustal and upper mantle of the Alps from the teleseismic data of the AlpArray experiment (<http://www.alpparray.ethz.ch/en/home>) by Full Waveform Inversion (FWI). The Alp Array experiment is an European initiative that allowed for the deployment of a dense array of broadband seismometers covering the entire Alpine Chain. In this framework, our goal is to develop high-resolution models of the Alpine chain by FWI. More precisely, we first want to build a new model of the crust and upper mantle of the Alpine chain with a wavelength-scale resolution and reconstruct several classes of parameters such as P and S anisotropic wavespeeds, density and attenuation. In a second step we will focus on a more focused target centred on the Ligurian knot in the western Alps by pushing the inversion at higher frequencies.

### Methods:

FWI is a seismic imaging technique which is formulated mathematically as a partial-differential (PDE)-constrained optimization problem, which aims at reconstructing the constitutive properties of the Earth's interior from seismic wave measurements made at the Earth's surface. The constraint is the elastodynamic equations, which control the propagation of waves in the frame of linear elasticity. The parameters describing the Earth's properties are contained in the coefficients of the elastodynamic equations and are estimated by fitting the seismic measurements with the simulated counterparts. Due to the size of the data and parameter spaces, this optimization problem is solved with local approaches where the gradient of the objective function can be computed efficiently with the adjoint state method. In the framework of the teleseismic configuration, the target to be imaged is located below the station layout, while the

---

source (the teleseismic event) is located thousand km away from the target. To deal with this configuration, seismic wave simulation is subdivided in two steps. First we simulate the 3D wavefields from the event to the edges of the target in a simplified axisymmetric Earth with four 2D simulations (AXISEM software). Then, we use these wavefields at the edges of the 3D heterogeneous target to simulate waves inside it with the spectral element method (This two step approach is referred to as wavefield injection technique). Once the forward problem is solved we compute the data misfit at receiver and build the gradient of the objective function by correlation between the incident wavefields and the adjoint wavefields computed by time-reversed modeling of the data residuals. The Hessian can be taken into account with limited-memory quasi-Newton method (l-GFGS) or second-order adjoint state method and linear conjugate gradient solver.

Codes:

To tackle this application, we have 3D FWI codes that work on CPUs and GPUs. The codes are parallelized by domain decomposition of the unstructured hexahedral meshes discretizing the Earth's model and right-hand sides distribution (the teleseismic sources, keeping in mind that the image of the Earth is built by summation of the contribution of each source). Ideally a MPI process is assigned to each subdomain and each right-hand side. On GPU, wavefield simulation in a domain is speedup by multithreading. In a pure CPU configuration, we may use OpenMP when available cores are available for this third level of parallelism.

For running the code to image the entire Alpine chain, we have estimated that we need 2 500 000 scalar hours on CPUs. Depending on the CPU generation, we expect a speedup of at least one order of magnitude compared to CPU time (As an example, running SPECFEM3D code on GPU generates a speedup of 25). In terms of storage, we need 30Tb to store the AXISEM solutions at the edges, the datasets, and the FWI results.

This research will be performed in the framework of the LISALPS project, which is funded by ANR. It involves three partners Geoazur, LMA and ISTerre. The PI is Stephane Operto. A PhD thesis and a postdoc position at Geoazur were funded by ANR. The PhD thesis is performed by Najmeh Mohammadi at Geoazur. The postdoc position has been attributed to Stephen Beller, who is an expert on FWI, teleseismic imaging and high performance computing. The PhD thesis is mainly supervised by Stephen Beller and Vadim Monteiller from LMA.

---

## **CPER Ubiquarium**

**Responsable scientifique :** Michel Riveill

**Laboratoire :** I3S

**Plateforme OPAL :** NEF (Inria)

**Nombre d'accrédités :** 16

### **Description :**

Dans le cadre du Cadre du CPER 2014-2020, les acteurs du campus SophiaTech se sont regroupés afin de poursuivre l'équipement du campus. Une demande pour un budget global de 3.100 K€ a été ainsi validée autour de 4 projets :

- RANDOM : Plateforme de mesure de Rayonnements électromagnétique d'ANtennes et de Diffraction jusqu'aux Ondes Millimétriques, d'objets communicants, de capteurs et de systèmes radar- aménagement de la chambre anéchoïde dont les aspects bâtiments ont été réalisés dans le cadre du précédent CPER
- PLATON : PLATeforme ouverte hybride pour la cONception et l'expérimentation des réseaux du futur
- Véhicule autonome en déplacement urbain
- Ubiquarium : Espace d'expérimentation pour l'informatique ambiante et les objets connectés

Le projet qui nous concerne ici, est Ubiquarium et regroupe principalement 2 aspects :

- Un ensemble de matériels de nature diverse pour équiper la salle d'expérimentation domotique du même nom,
- L'achat de gros serveurs de calcul (5 serveurs Power EDGE T640) pour le traitement massif de données (big data) en vue notamment de leur exploitation dans le cadre des études sur l'intelligence artificielle.

C'est évidemment, ce dernier volet qui nous intéresse ici. En accord avec les différentes tutelles et pour des raisons spécifiques (climatisation, administration, maintenance...). Ce matériel a été installé dans les locaux Inria comme différents nœuds du cluster NEF.

L'objet cette demande est de permettre aux membres de l'équipe SPARKS de l'I3S d'avoir accès de manière privilégiée selon la gestion des nœuds au sein de NEF à ce matériel en vue d'y mener les travaux prévus. Nous rappelons que tous ceux qui ont un accès à NEF (via Inria ou via OPAL) peuvent bénéficier de la puissance de calcul offerte par ces différents nœuds.

- Les membres des EPC Inria-I3S ont déjà accès à NEF mais pas à ces nœuds de calcul de manière prioritaire. Cela se fera via la mise en place d'un groupe d'utilisateur SPARKS au sein de NEF.

Voici la liste des travaux prévus pour cette première tranche :

- Analyse de rapport médicaux (Correction orthographique, Annotation, Extraction de relations)
- Extraction non supervisée de relations entre entités nommées - Application à l'analyse de rapports médicaux
- Analyse d'image et interprétation des décisions
- Correction orthographique et annotations de rapport médicaux
- Prédiction du score d'imputabilité associé à une déclaration d'effet indésirable
- Detection and classification of the most important actions in videos, using Deep Learning and multimodal features
- Réseaux de neurones à spikes appliqués à des données évènementielles, selon des apprentissages non supervisés, par renforcement ou bien encore par imitation
- Réseaux de neurones à spikes appliqués à des données évènementielles, selon des apprentissages non supervisés, par renforcement ou bien encore par imitation
- Aide et conseil aux utilisateurs et métaanalyse pour la recommandation de workflows d'apprentissage automatique (suite du projet ROCKFlows)
- Building a recommender system to suggest learning resources to user of e-learning platforms, based on information extracted from Knowledge Graphs
- Étude des méthodes d'analyse des données sur les MOOCs en informatique pour l'amélioration de leur contenu et l'augmentation de leur taux de réussite et conception d'une chaîne de traitement des données en collaboration avec la plateforme FUN (France Université Numérique) pour transformer les données liées au processus d'apprentissage vers un format adapté pour l'analyse et l'ouverture de ses données

---

## **SpikingAI (NeuroMod project)**

**Responsable scientifique :** Gilles Scarella

**Laboratoire :** LJAD

**Plateforme OPAL :** NEF (Inria)

**Nombre d'accrédités :** 13

**Résultats obtenus grâce aux moyens informatiques attribués :**

- James, A., Reynaud-Bouret, P., Mezzadri, G., Sargolini, F., Bethus, I., & Muzy, A. (2023). Strategy inference during learning via cognitive activity-based credit assignment models. *Scientific Reports*, 13(1), 9408.
- A. Colas-Gattepaille, F. Dagang, A. Muzy, (submitted, 2022). Delay Neural Networks for temporal learning in deep neural networks
- C. Mascart, G. Scarella, P. Reynaud-Bouret, A. Muzy, Scalability of large neural network simulations via activity tracking with time asynchrony and procedural connectivity, *Neural computation.*, 2022, 34 (9), pp.1915-1943. <https://hal.science/hal-03258798v2>

Rapports de stage :

- Grünwald, J. (2023). "Hyperbolic graph based spiking neural networks with improved storage capacity". In : Université Côte d'Azur, Master's thesis. Encadré par Roland Diel et Alexandre Muzy.
- Grünwald, J. (2024). "Geometric Connectivity Increases Spiking Neural Network Performances". In : Université Côte d'Azur, Master's thesis. Encadré par Roland Diel et Alexandre Muzy.

### **Description du projet :**

Research projects at NeuroMod institute aim at gathering together computer scientists and mathematicians with biologists and cognitivists. On the long term, the SpikingAI project aims at developing new algorithms for understanding the learning dynamics of rodents and humans.

The data set used by these algorithms consists of the bio-cognitive data collected on rodents navigating in a maze and eeg data collected on humans during cognitive tasks. At behavioral level, the actions granularity (paths, turns, choices etc.) used by the rodents or participants to take their decisions is automatically inferred from animal position data [MZ17] or recorded answers. At neuronal level, the functional connectivity between the neurons is analyzed statistically [RMB21] from the neuron spike trains collected *in vivo*

---

inside the brain of the animals or from the eeg data recorded on humans. Finally, the relationship between behavioral and neuronal data is studied statistically and computationally. In particular, new machine learning algorithms are developed to integrate neuronal data.

We need High Performance Computing (HPC) for:

- Running large scale simulations of the electrical activity in neuronal networks [MMR20].
- Running the statistical analysis approach upon: the neuronal spike trains (generated by the large scale simulations, or collected on the animals) or the eeg data and behaviors.
- Running the machine learning algorithm at both behavioral and neuronal levels.
- Running new theoretical/artificial neuronal networks aiming, on the long term, to reproduce the learning activity *in vivo*.

## References

[MZ17] Alexandre Muzy and Bernard P. Zeigler. "Activity-based Credit Assignment (ACA) Heuristic for Simulation-based Stochastic Search in a Hierarchical Model-base of Systems". In: IEEE Systems Journal 11.4 (2017), pp. 1916–1927. doi: 10.1109/JYST.2014.2342534. url: <https://hal.archives-ouvertes.fr/hal-01315156>

[MMR20] Cyrille Mascart, Alexandre Muzy, and Patricia Reynaud-Bouret. "Discrete event simulation of point processes: A computational complexity analysis on sparse graphs". working paper or preprint. Jan.2020. url: <https://hal.archives-ouvertes.fr/hal-02455422>

[RMB21] Patricia Reynaud-Bouret, Alexandre Muzy, and Ingrid Bethus. "Towards a mathematical definition of functional connectivity". In: Comptes Rendus de l'Académie des Sciences - Series I – Mathematics (2021). url: <https://hal.archives-ouvertes.fr/hal-03093516>

---

## ***MSc INDEX Data Science and AI***

**Responsable pédagogique :** Giovanni Neglia

**Plateforme OPAL :** NEF (Inria)

**Nombre d'accrédités :** 13

### **Description :**

Dans le cadre du cours "Optimization for Data Science" du master UCA Data Science and AI, nous permettons aux étudiants d'utiliser le cluster NEF pour implémenter et tester des algorithmes d'entraînement (e.g. avec PyTorch).