

RAPPORT D'ACTIVITÉ

**Centre de calculs
AZZURRA**



**UNIVERSITÉ
CÔTE D'AZUR**

2025

Maeva ANTOINE

DATA CENTER SUD

 **UNIVERSITÉ CÔTE D'AZUR**

 **SYNERGIE
SECS**

Table des matières

AZZURRA EN CHIFFRES.....	4
FAITS MARQUANTS.....	5
MISE À JOUR LOGICIELLE	6
DÉMÉNAGEMENT.....	7
MISE À JOUR MATÉRIELLE.....	10
UTILISATEURS.....	12
STATISTIQUES D'UTILISATION	13
HEURES CONSOMMÉES.....	13
TAUX D'OCCUPATION.....	15
STATISTIQUES PAR LABORATOIRE.....	16
FORMATIONS	17
FORMATIONS ET COURS AYANT EU ACCES À AZZURRA.....	18
BILAN SCIENTIFIQUE.....	19
RÉCAPITULATIF DES PROJETS DE RECHERCHE.....	21
RECAPITULATIF DES PUBLICATIONS.....	23
DESCRIPTIF DES PROJETS.....	30
Projets impliquant des ingénieurs MSI.....	31
Chimie	38
Physique non linéaire, fluides complexes	86
Optique	111
Hydraulique urbaine	115
Dynamique des fluides	122
Géophysique interne	124
Mathématiques, informatique.....	130
Magnétohydrodynamique.....	167
Astrophysique	177
Biologie, santé	224
Droit, économie, géographie	261

OPAL	270
Projets accrédités OPAL par l'université ne calculant pas sur Azzurra	271

AZZURRA EN CHIFFRES

82 nœuds dédiés au calcul CPU, représentant **2 872** cœurs.

6 nœuds dédiés au calcul GPU, soit **8** cartes GPU Tesla V100, **4** cartes A100 et **12** cartes H100.

560 To pour le stockage des données.

1 644 796,43 € HT de matériel, prestations d'installation / support et licence logicielle depuis 2019.

Depuis son lancement en mai 2020, **40 397 087** heures de calcul CPU et **202 682** heures de calcul GPU ont été utilisées.

En 2025 :

7 764 842 heures de calcul CPU ont été utilisées.

54 802 heures de calcul GPU ont été utilisées.

84 projets de recherche ont bénéficié d'une allocation de ressources.

211 personnels ont bénéficié d'un compte utilisateur sur Azzurra.

63 étudiants ont bénéficié d'un compte dans le cadre d'un cours.

217 étudiants et personnels UniCA ont été accrédités OPAL par l'université.

943 595,43 € HT dépensés en prestations de service et achats de matériel.

69% de ces dépenses ont été financées par des laboratoires de recherche de l'université.

FAITS MARQUANTS

Mars 2025 :

- Mise à jour logicielle du cluster (OS + Slurm + drivers + firmware).

Juillet 2025 :

- Déménagement du cluster depuis le Centre Inria d'Université Côte d'Azur vers le nouveau data center sur le campus Valrose.
- Ajout de nouveau matériel :
 - 1288 cœurs CPU
 - 12 cartes GPU H100
 - 240 To de stockage supplémentaires.

MISE À JOUR LOGICIELLE

La mise à jour logicielle a été réalisée en mars.

- L'OS est passé de CentOS 7.9 à Rocky 9.5 sur toutes les machines.
- Slurm est passé de la version 18.08.8 à 23.11.10.
- Les drivers et firmware ont été mis à jour.
- Une limitation (0.5 To) de l'espace de stockage par utilisateur sur le NFS (/home) est mise en place. Les utilisateurs impactés ont été contactés par e-mail.
- Des limites de charge sur login-hpc ont été instaurées, pour éviter que des utilisateurs exécutent du code directement sur cette machine et la ralentissent.

Suite à la mise à jour logicielle :

Certains codes compilés par les utilisateurs ont dû être recompilés à l'aide d'un nouveau module.

63 modules logiciels existants ont été mis à jour ou réinstallés afin d'être compatibles avec la mise à jour de l'OS.

Un module aocc (compilateurs AMD) a été ajouté pour optimiser l'utilisation des nouvelles machines pour le calcul CPU, équipées de processeurs AMD.

Un module avec les compilateurs Intel 2025 a également été ajouté.

DÉMÉNAGEMENT

Le cluster a été déménagé du Centre Inria d'Université Côte d'Azur en juillet 2025 pour être réinstallé dans le tout nouveau data center sur le campus Valrose :







4 armoires sont dédiées au cluster Azzurra, dans lesquelles nous avons pu ajouter également du nouveau matériel, décrit en pages suivantes.

MISE À JOUR MATÉRIELLE

Nouveau matériel

35 machines PowerEdge R6625 pour le calcul CPU, chacune équipée de :

- 2 processeurs AMD EPYC 9124, 3 GHz à 16 cœurs (= 32 cœurs par machine)
- 384 Go RAM

3 machines PowerEdge R760XA pour le calcul GPU, chacune équipée de :

- 4 cartes NVIDIA H100 NVL, PCIe, 350 W-400 W, 94 Go
- 2 processeurs Intel Xeon Gold 5420+ 2 GHz à 28 cœurs (= 56 cœurs par machine)
- 512 Go RAM

La mémoire de la machine SMP existante a été augmentée de **512 Go de RAM**, offrant désormais 1,5 To de RAM.

Le stockage BeeGFS a été doublé avec l'achat de **deux serveurs PowerEdge R760**, afin d'offrir 480 To en tout.

Les machines de calcul existantes en 2024 restent incluses au cluster.

Matériel remplacé

La machine frontale (pour la connexion des utilisateurs), dont la garantie expirait en janvier 2025, a été remplacée par **un serveur PowerEdge R660**.

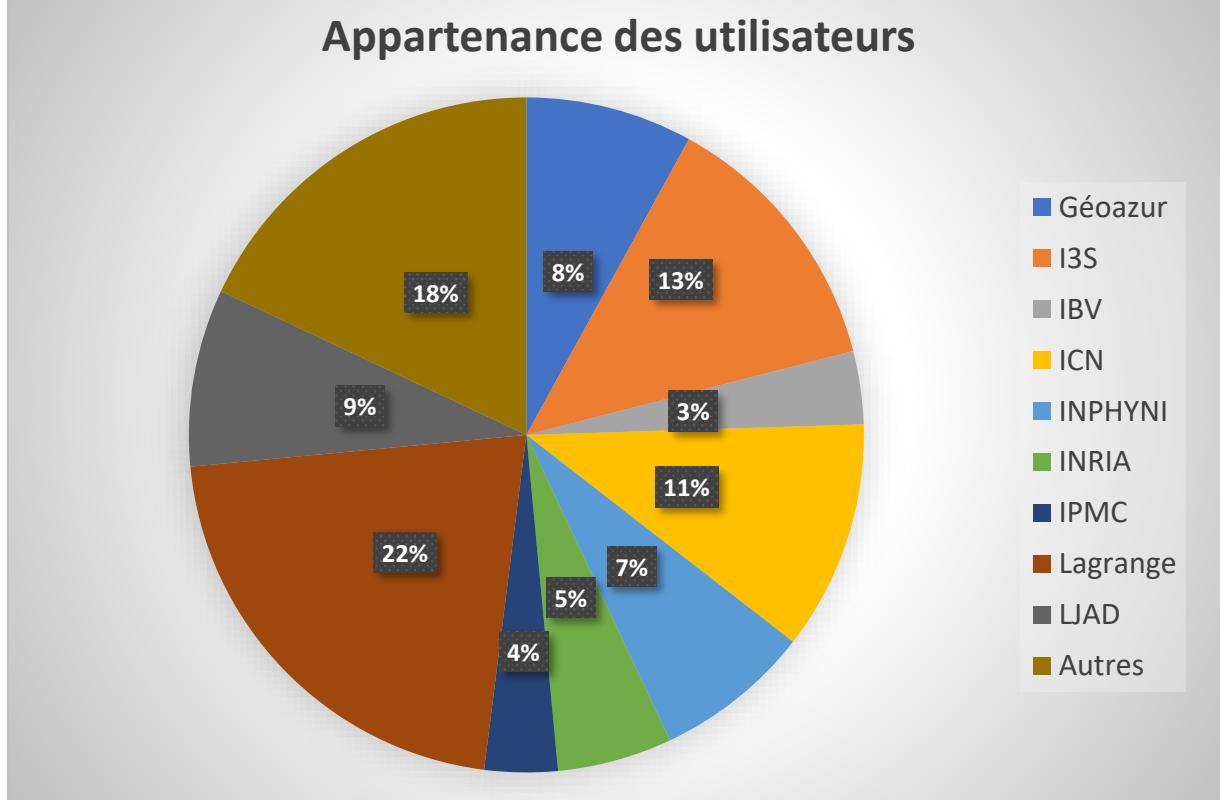
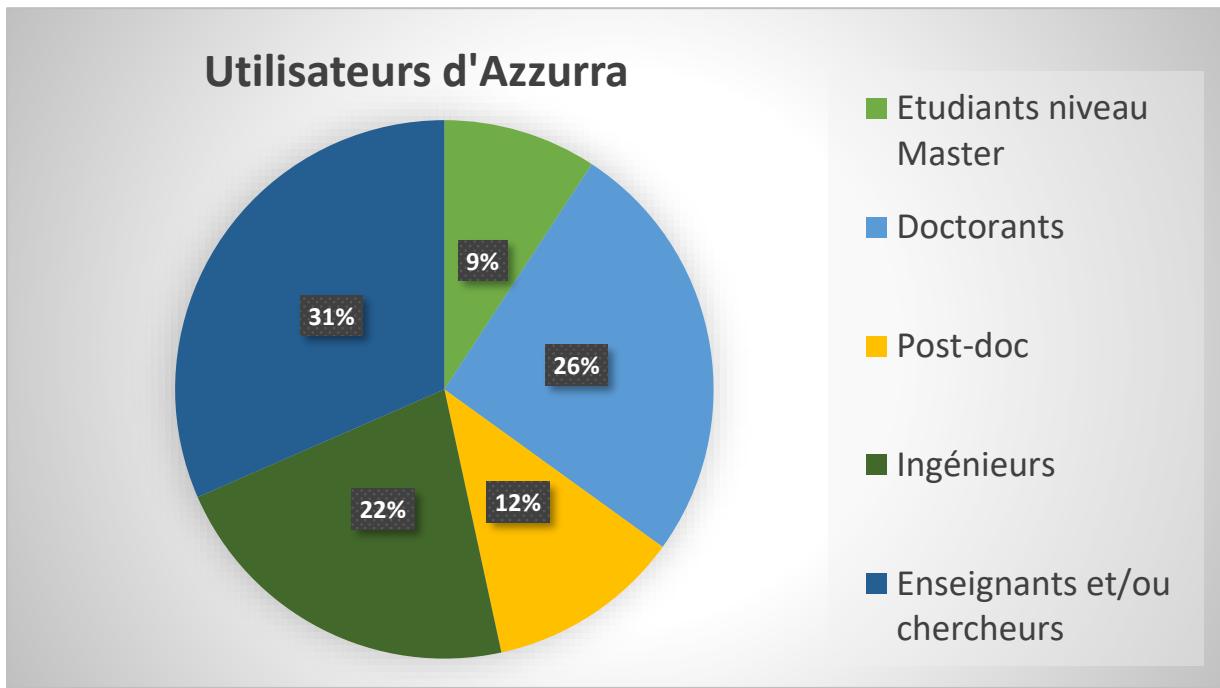
Le switch InfiniBand (IB), dont la garantie était aussi expirée depuis janvier 2025, a été remplacé par **deux switchs IB Mellanox Quantum QM8790 40 ports**.

Le besoin d'un second switch s'explique par l'augmentation du nombre de machines de calcul. Pour cette même raison a été acheté un troisième switch ethernet, modèle **PowerSwitch E3248P-ON, 48 x 1G RJ-45**.

UTILISATEURS

Durant l'année 2025, **211** personnels ont bénéficié d'un compte utilisateur sur Azzurra.
(+17% par rapport à 2024)

Le profil des utilisateurs est le suivant :



STATISTIQUES D'UTILISATION

HEURES CONSOMMÉES

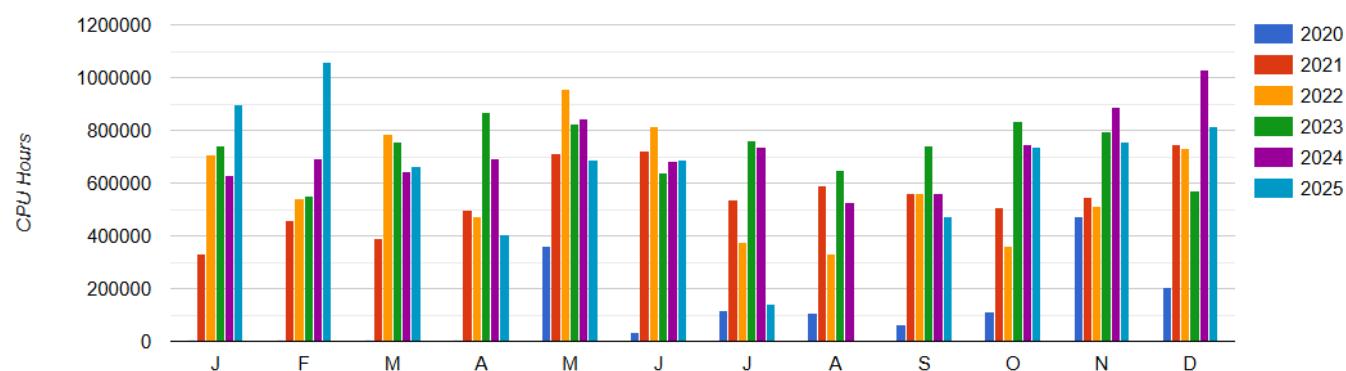
Heures CPU

Depuis son lancement officiel en mai 2020, **40 397 087** heures de calcul CPU ont été consommées.

En 2025, ce sont **7 764 842** heures qui ont été utilisées.

2025

<i>Janvier</i>	895156
<i>Février</i>	1057479
<i>Mars</i>	661097
<i>Avril</i>	405313
<i>Mai</i>	687058
<i>Juin</i>	685956
<i>Juillet</i>	142818
<i>Août</i>	831
<i>Septembre</i>	474350
<i>Octobre</i>	738150
<i>Novembre</i>	755332
<i>Décembre</i>	813093



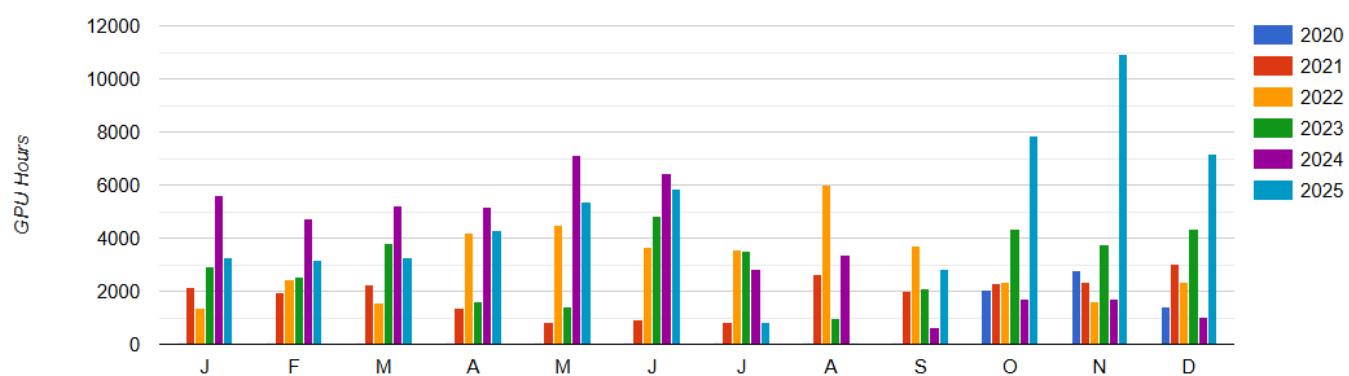
Heures GPU

Depuis son lancement officiel en mai 2020, **202 682** heures de calcul GPU ont été consommées.

En 2025, ce sont **54 802** heures qui ont été utilisées.

2025

<i>Janvier</i>	3265
<i>Février</i>	3176
<i>Mars</i>	3246
<i>Avril</i>	4314
<i>Mai</i>	5387
<i>Juin</i>	5840
<i>Juillet</i>	830
<i>Août</i>	-
<i>Septembre</i>	2816
<i>Octobre</i>	7865
<i>Novembre</i>	10903
<i>Décembre</i>	7160



TAUX D'OCCUPATION

CPU

Environ **1 511 808** heures disponibles par mois jusqu'en août, ensuite : **2 379 852** heures.

Taux d'occupation du cluster (CPU) sur l'année : **33.02%**

2025

<i>janvier</i>	59.21%
<i>février</i>	77.44%
<i>mars</i>	59.92%
<i>avril</i>	51.94%
<i>mai</i>	45.45%
<i>juin</i>	46.89%
<i>juillet</i>	6.78%
<i>août</i>	0.03%
<i>septembre</i>	19.93%
<i>octobre</i>	30.25%
<i>novembre</i>	31.98%
<i>décembre</i>	33.32%

GPU

8640 heures GPU disponibles par mois jusqu'en août, ensuite : **17 280** heures.

Taux d'occupation des ressources GPU sur l'année : **35.27%**

2025

<i>janvier</i>	36.57%
<i>février</i>	39.39%
<i>mars</i>	49.44%
<i>avril</i>	41.82%
<i>mai</i>	60.34%
<i>juin</i>	67.59%
<i>juillet</i>	6.33%
<i>août</i>	-
<i>septembre</i>	16.29%
<i>octobre</i>	44.05%
<i>novembre</i>	63.10%
<i>décembre</i>	40.1%

STATISTIQUES PAR LABORATOIRE

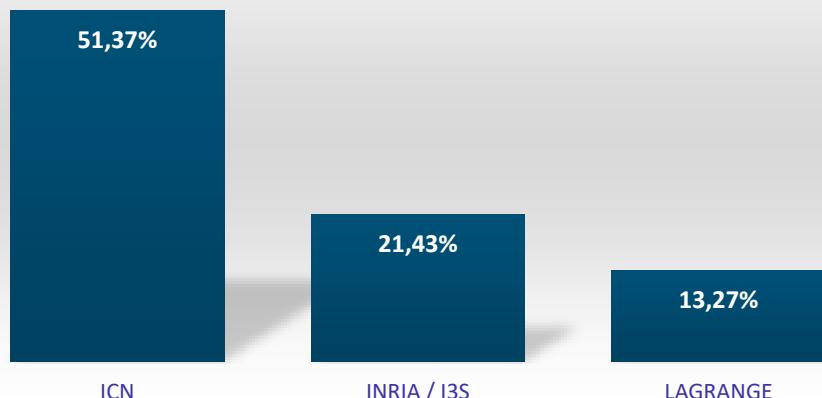
Laboratoires qui ont le plus utilisé le calcul CPU

TOP 3 : utilisation des CPUs



Laboratoires qui ont le plus utilisé le calcul GPU

TOP 3 : utilisation des GPUs



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

Année universitaire 2024-2025 :

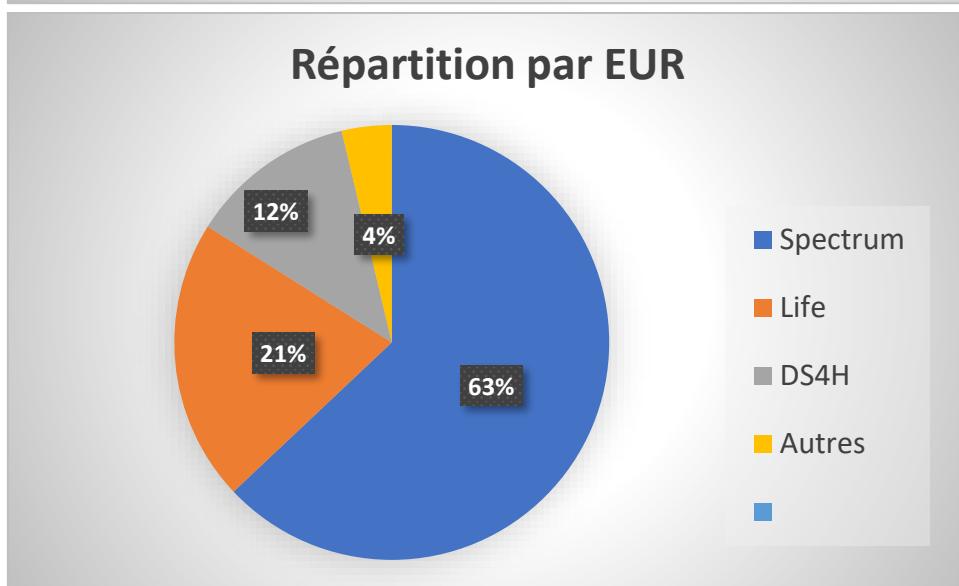
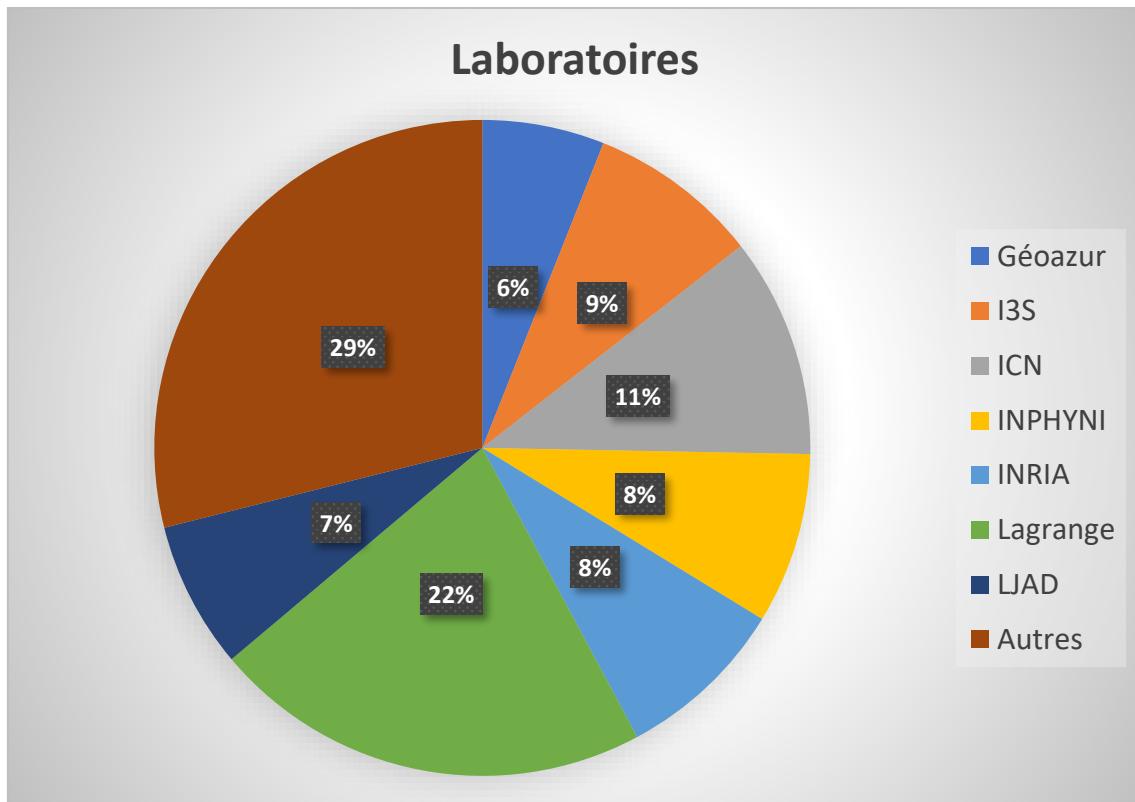
- **Telemac et parallélisme**
 - GE5 Polytech'Nice-Sophia
 - M2 HydroProtech
 - Nombre de participants : 33
- **Calcul haute performance**
 - MASTER2-INUM
 - Nombre de participants : 12

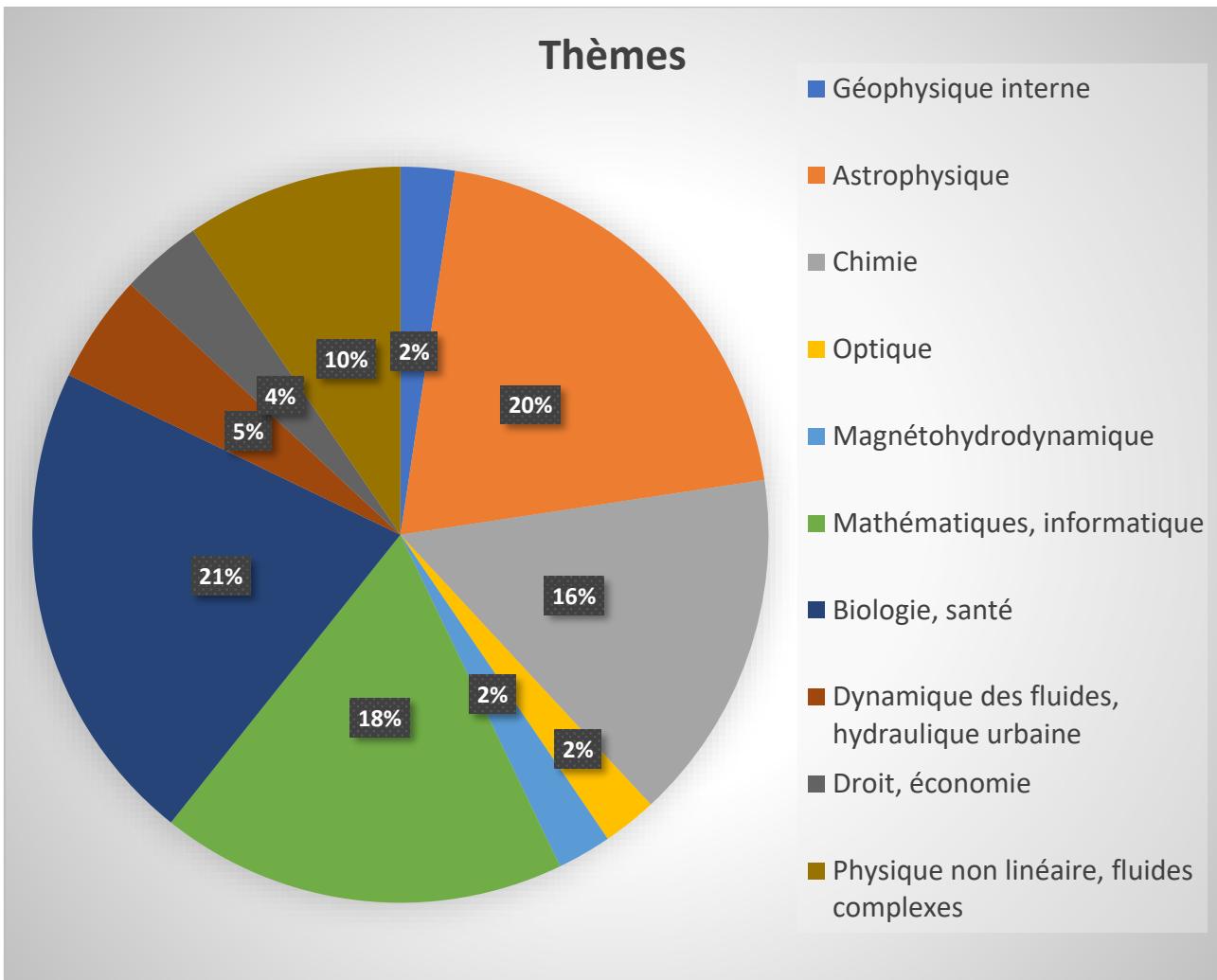
Année universitaire 2025-2026 :

- **Calcul haute performance**
 - MASTER2-INUM
 - Nombre de participants : 18

BILAN SCIENTIFIQUE

En 2025, **84** projets ont bénéficié d'une allocation de ressources sur Azzurra (**+14%** par rapport à 2024). Ces projets se répartissent ainsi :





Dans le détail :

Géophysique interne (Géoazur; EUR Spectrum)	2
Astrophysique (Géoazur, Lagrange; EUR Spectrum)	17
Chimie (ICN, LJAD; EUR Spectrum)	13
Optique (Inphyni; EUR Spectrum)	2
Magnétohydrodynamique (Lagrange, LJAD; EUR Spectrum)	2
Mathématiques, informatique (I3S, Inria, LJAD; EUR DS4H, Spectrum)	15
Biologie, santé (IPMC, ISA; EUR Life)	18
Dynamique des fluides, hydraulique urbaine (LJAD, Géoazur ; EUR Spectrum)	4
Droit, économie, géographie (ERMES; EUR LexSociété – INSTAR ; EUR ELMI – EUR ODYSSEE)	3
Physique non linéaire, fluides complexes (Inphyni; EUR Spectrum)	8

RÉCAPITULATIF DES PROJETS DE RECHERCHE

Porteur	Laboratoire	Discipline	EUR	CLASSIFICATION
Carry	Lagrange	Astrophysique	Spectrum	Astrophysique
Chiavassa	Lagrange	Astrophysique	Spectrum	Astrophysique
Fienga	Geoazur	Astrophysique	Spectrum	Astrophysique
Fienga	Geoazur	Astrophysique	Spectrum	Astrophysique
Lega	Lagrange	Astrophysique	Spectrum	Astrophysique
Mary	Lagrange	Astrophysique	Spectrum	Astrophysique
Méheut	Lagrange	Astrophysique	Spectrum	Astrophysique
Méheut	Lagrange	Astrophysique	Spectrum	Astrophysique
Michel	Lagrange	Astrophysique	Spectrum	Astrophysique
Michel	Lagrange	Astrophysique	Spectrum	Astrophysique
N'Diaye	Lagrange	Astrophysique	Spectrum	Astrophysique
N'Diaye	Lagrange	Astrophysique	Spectrum	Astrophysique
Niccolini	Lagrange	Astrophysique	Spectrum	Astrophysique
Parmentier	Lagrange	Astrophysique	Spectrum	Astrophysique
Slezak	Lagrange	Astrophysique	Spectrum	Astrophysique
Tanga	Lagrange	Astrophysique	Spectrum	Astrophysique
Tanga	Lagrange	Astrophysique	Spectrum	Astrophysique
Becavin	IPMC	Pharmacologie	Life	Biologie, santé
Besse	IBV	Biologie	Life	Biologie, santé
Blin	LP2M	Médecine	Life	Biologie, santé
Bottini	MD lab / MSI	Médecine	Life	Biologie, santé
Braendle	IBV	Biologie	Life	Biologie, santé
Cerezo	C3M	Médecine	Life	Biologie, santé
Cristofari	IRCAN	Médecine	Life	Biologie, santé
Danchin	Agrobiotech	Génomique	Life	Biologie, santé
Delaunay	IBV	Biologie	Life	Biologie, santé
Dominguez	C3M	Médecine	Life	Biologie, santé
Gauci	UR2CA-iULS	Médecine	Life	Biologie, santé
Glaichenhaus	IPMC	Pharmacologie	Life	Biologie, santé
Hubstenberger	IBV	Biologie	Life	Biologie, santé
Hudry	IBV	Biologie	Life	Biologie, santé
Hug	LAMHESS	Médecine	DS4H	Biologie, santé
Liti	IRCAN	Médecine	Life	Biologie, santé
Milanesio	IRCAN / MSI	Biologie	Life	Biologie, santé
Mograbi	IHU-Respirera	Médecine	Life	Biologie, santé
Antonczak	ICN	Chimie	Spectrum	Chimie
Azoulay	ICN	Chimie	Spectrum	Chimie
Cassam-Chenaï	LJAD	Chimie	Spectrum	Chimie
Dumont	ICN	Chimie	Spectrum	Chimie
Fiorucci	ICN	Chimie	Spectrum	Chimie
Fiorucci	ICN	Chimie	Spectrum	Chimie
Fontaine-Vive	ICN	Chimie	Spectrum	Chimie
Mija	ICN	Chimie	Spectrum	Chimie

Nothias	ICN	Chimie	Spectrum	Chimie
Ronco	ICN	Chimie	Spectrum	Chimie
Kalugin	-	Chimie	-	Chimie
Kalugin	-	Chimie	-	Chimie
Kyrychenko	-	Chimie	-	Chimie
Constant	Ermes	Droit	LexSociété	Droit, économie
Czellar	INSTAR	Economie	ELMI	Droit, économie
Fox	ESPACE	Géographie	Odyssee	Géographie
Vincenzi	LJAD	Mécanique des fluides	Spectrum	Dynamique des fluides
Brenner	LJAD	Mécanique des fluides	Spectrum	Hydraulique urbaine
Delestre	LJAD	Mécanique des fluides	Spectrum	Hydraulique urbaine
Petit	Geoazur	Mécanique des fluides	Spectrum	Hydraulique urbaine
Ampuero	Geoazur	Géophysique	Spectrum	Géophysique interne
Duclaux	Geoazur	Géophysique	Spectrum	Géophysique interne
Marcotte	LJAD	Mathématiques	Spectrum	Magnétohydrodynamique
Ponty	Lagrange	Mécanique des fluides	Spectrum	Magnétohydrodynamique
Antonini	I3S	Informatique	DS4H	Mathématiques, informatique
Bossy	Inria	Mathématiques	Spectrum	Mathématiques, informatique
Comet	I3S	Biologie/Informatique	Spectrum	Mathématiques, informatique
Faron	I3S	Informatique	Spectrum	Mathématiques, informatique
Gandon	Inria	Informatique	Spectrum	Mathématiques, informatique
Goudon	LJAD	Mathématiques	Spectrum	Mathématiques, informatique
Lanteri	Inria	Mathématiques	Spectrum	Mathématiques, informatique
Nguyen	I3S	Informatique	DS4H	Mathématiques, informatique
Pasquier	I3S	Informatique	DS4H	Mathématiques, informatique
Rey	I3S	Informatique	DS4H	Mathématiques, informatique
Sassatelli	I3S	Informatique	DS4H	Mathématiques, informatique
Villata	Inria	Informatique	DS4H	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
Blanc	InPhyNi	Physique/Mécanique des fluides	Spectrum	Physique non linéaire, fluides complexes
Homann	Lagrange	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

- Planetesimal gravitational collapse in a gaseous environment: Thermal and dynamic evolution Segretain, P.; Méheut, H.; Moreira, M.; Lesur, G.; Robert, C. & Mauxion, J., A&A, EDP Sciences, 2024, 692, A118
- Combining machine learning and electrophysiology for insect odorant receptor studies. A. Comte, S. Fiorucci, E. Jacquin-Joly, in Olfactory receptors: methods and protocols, Methods Mol. Biol. (2025), accepted.
- Accelerating Ligand Discovery for Insect Odorant Receptors. A. Comte, M. Lalis, L. Brajon, R. Moracci, N. Montagné, J. Topin, E. Jacquin-Joly, S. Fiorucci. Int. J. Biol. Sci. 2025, 21(5), 2101-2117.
- Ligand discrimination in hOR1Al based on the capacitive response. A. Lagunas, C. Belloir, M. Lalis, L. Briand, J. Topin, P. Gorostiza, J. Samitier. Biosens Bioelectron. 2025, 271, 117000.
- Study and Optimization of the Selectivity of an Odorant Binding Protein-based Bioelectronic Nose. M. El Kazzy, M. Lalis, C. Hurot, J.S. Weerakkody, R. Mathey, C. Saint-Pierre, A. Buhot, T. Livache, J. Topin, L. Moitrier, C. Belloir, L. Briand, Y. Hou. Biosens. Bioelectron 2025, 268, 116879.
- A Comprehensive Analysis of Olfactory Receptor Deorphanization. M. Lalis, M. Hladiš, S. Abi Khalil, C. Deroo, C. Marin, M. Bensafi, N. Baldovini, L. Briand, S. Fiorucci, J. Topin. Chem. Senses 2024, 49, bjae037.
- Michel Orsi, Laurent Lobry, Elisabeth Lemaire, and François Peters. Mass and momentum balance during particle migration in the pressure-driven flow of frictional non-brownian suspensions. Journal of Fluid Mechanics, 998 :A16, 2024.
- Michel Orsi, Laurent Lobry, Elisabeth Lemaire, François Peters; The effect of coupled friction and adhesion on the rheology of non-Brownian dense suspensions. *J. Rheol.* 1 July 2025; 69 (4): 515–539
- Jourdon, A., May, D. A., Hayek, J. N., & Gabriel, A.-A. (2025). 3D reconstruction of complex fault systems from volumetric geodynamic shear zones using medial axis transform. *Geochemistry, Geophysics, Geosystems*, 26, e2025GC012169. <https://doi.org/10.1029/2025GC012169>
- Simonnin, A; Parmentier, V. et al. Time-resolved absorption of six chemical species with MAROON-X points to a strong drag in the ultra-hot Jupiter TOI-1518 b *Astronomy & Astrophysics*, Volume 698, id.A314, 24 pp. <https://hal.science/hal-05134227v1>
- Dernoncourt, F., Avrillon, S., Logtens, T., Cattagni, T., Farina, D. and Hug, F. (2025), Flexible control of motor units: is the multidimensionality of motor unit manifolds a sufficient condition?. *J Physiol*, 603: 2349-2368. <https://doi.org/10.1113/JP287857>
- "Development of a reliable, sensitive, and convenient assay for the discovery of new eIF5A hypusination inhibitors", O. Benaceur, P. Ferreira Montenegro, M. Kahi, F. Fontaine-Vive, N. M. Mazure, M. Mehiri, F. Bost, P. Peraldi, PLOS One, 2025, 20(2), e0308049

- "CuMg₂GaS₄: A novel photocatalyst with promising properties", Hocine H., Khelfaoui F., Al-Douri Y., Boudia K., Fontaine-Vive F. et al., Multidiscipline Modeling in Materials and Structures, 2025, 21 (3): 501–519
- "Silver- and Gold-Catalyzed Divergent Cascade Cycloisomerization / [3 + 2] versus [2+2+1] Cycloaddition towards a Stereoselective Access to Heterohexacyclic Derivatives", E. Gentilini, A. Bouschon, V. Davenel, F. Fontaine-Vive, J.-M. Fourquez, V. Michelet, Chemical Science, 2025
- "Ruthenium-Catalyzed Enyne Metathesis: an Entry to Functionalized Azaborine Heterocycles", Mao K., Fontaine-Vive F., Melot R, Michelet V., Organic Letters, 2025, 27, 12, 2952–2957
- A robust, discrete-gradient procedure for optimisation with time-dependent PDE and norm constraints. Mannix, P.M., Skene, C.S., Auroux, D. & Marcotte, F. SMAI Journal of Computational Mathematics 10, pp 1-28 (2024).
- Singlan, Nina, Fadi Abou Choucha, and Claude Pasquier. "A new Similarity Based Adapted Louvain Algorithm (SIMBA) for active module identification in p-value attributed biological networks." Scientific Reports 15.1 (2025): 11360. HAL : hal-05020436 v1 DOI : 10.1038/s41598-025-95749-6
- Yannick Ponty, Helene Politano, Annick Pouquet. Intermittency assessed through a model of kurtosis-skewness relation in MHD in fast dynamo regimes, Journal of Plasma Physics. 91(2):E69 2025 <https://doi.org/10.1017/S0022377825000169>
- Pouquet, A., Marino, R., Politano, H., Ponty, Y., and Rosenberg, D.: Intermittency in fluid and MHD turbulence analyzed through the prism of moment scaling predictions of multifractal models, Nonlin. Processes Geophys., 32, 243–259, 2025 <https://doi.org/10.5194/npg-32-243-2025>
- Hug F, Dernoncourt F, Avrillon S, Thorstensen J, Besomi M, van den Hoorn W, Tucker K. Non-homogeneous distribution of inhibitory inputs among motor units in response to nociceptive stimulation at moderate contraction intensity. J Physiol. 2025 Jun;603(11):3445-3461. doi: 10.1113/JP288504. Epub 2025 May 31. PMID: 40449022; PMCID: PMC12148210.
- Lega, E. , Morbidelli, A., Masset ,F. and Béthune W. Exploring the formation of multiple dust-trapping rings in the inner Solar System. A&A, 702, A45 (2025) <https://doi.org/10.1051/0004-6361/202555284>
- A mixture-like model for tumor-immune system interactions. C. Tayou Fotso, S. Girel, F. Anjuere, V. M. Braud, T. Goudon, F. Hubert. J. Theor. Biol., 111738, 2024.
- Balossi, R., P. Tanga, M. Delbo, A. Cellino, and F. Spoto 2025. An ancient L-type family associated with (460) Scania in the middle main belt as revealed by Gaia DR3 spectra. Astronomy and Astrophysics 702, A239. <https://doi.org/10.1051/0004-6361/202555527>
- Minker, K., and 22 colleagues 2025. A dynamical dichotomy in large binary asteroids. Astronomy and Astrophysics 701, A42. <https://doi.org/10.1051/0004-6361/202554973>
- Aurélien Crida, Clément Baruteau, Jean-François Gonzalez, Frédéric Masset, Paul Segretain, et al.. The reflex instability: exponential growth of a large-scale $m = 1$ mode in astrophysical discs. Open Journal of Astrophysics, 2025, 8, [\(10.33232/OJ.A.143230\)](https://doi.org/10.33232/OJ.A.143230). hal-05222976

- Labarre, V., Krstulovic, G., & Nazarenko, S. (2025). Wave-kinetic dynamics of forced-dissipated turbulent internal gravity waves. *Physical Review Letters*, 135(1), 014101.
- Kyrychenko A., Prud M. V., Lohachova K. O., Prezhdo O. V., Kalugin O. N. Native folding stability of the main protease of coronavirus SARS-CoV-2 on a gold surface. *J. Phys. Chem. Lett.* 2025, 43 (16), 11222-11229. doi: 10.1021/acs.jpclett.5c02743
- Dikarieva K., Koverga V., Kalugin O. Local structure and Li-ion transport mechanism in LiFSI/DME/BTFE electrolyte revealed by molecular dynamics simulation. *Kharkov University Bulletin; Chemical Series*, No 45(68) 2025.

Conférences

- Planetary gravitational collapse in a gaseous environment, Auteurs: Segretain, P.; Méheut, H.; Moreira, M. Conférence de la Société Française d'Astronomie et d'Astrophysique (Marseille).
- A Deep Metric Learning Approach for Olfactory Mixture Discriminability. M. Lalis, S. Fiorucci, J. Topin, M. Hladiš. RECOMB/ISCB Conference on Regulatory & Systems Genomics with DREAM Challenges (RSGDREAM 2024). Madison (Wisconsin, USA), 1-3 October 2024.
- Molecular basis of mammalian sweet taste detection. S. Fiorucci, L. Briand. 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024
- Olfaction under the lens of a computational microscope. J. Topin, J. Pacalon, M. Hladiš, C. Belloir, L. Briand, S. Fiorucci. 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024.
- Encoding Odor: Bridging Smell and Combinatorial code using M2OR database. M. Lalis, M. Hladiš, S. A. Khalil, L. Briand, S. Fiorucci, J. Topin, 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024.
- Celian Ringwald, Fabien Gandon, Catherine Faron, Franck Michel, Hanna Abi Akl. Kastor: Fine-tuned Small Language Models for Shape-based Active Relation Extraction. Extended Semantic Web Conference 2025, Jun 2025, Portoroz, France. (hal-05078493)
- Fienga et al. Updates on the INPOP ephemerides. Symposium 401 de l'IAU (2025)
- Murray, A PCA based effective model for asteroid accelerations. Symposium 401 de l'IAU (2025)
- V. Mariani et al. Boosting decision trees for Main Belt Asteroid selection in planetary ephemerides. Symposium 401 de l'IAU (2025)
- Michel Orsi, Elisabeth Lemaire, Laurent Lobry, and François Peters. Detailed mass and momentum balances during particle migration in the pressure-driven channel flow of frictional non-brownian suspensions. In 12th International Conference on Multiphase flow, Toulouse, France, May 2025.
- Michel Orsi, Elisabeth Lemaire, Laurent Lobry, and François Peters. Rheology of non-brownian suspensions of adhesive and frictional particles via particle-resolved numerical simulations. In AERC 2025 - The Annual European Rheology Conference., Lyon, France, 2025.

- M. Elsawy, A. Gobé, G. Leroy, S. Lanteri and C. Scheid. Advanced computational design of complex nanostructured photonic devices using high order discontinuous Galerkin methods and statistical learning global optimization. Proceedings of the Emerging Technologies in Computational Sciences for Industry, Sustainability and Innovation - Math to Product 2023 Conference. Lecture Notes in Computational Science and Engineering, Springer, Volume 146, to appear (2025)
- M. Elsawy, A. Gobé, G. Leroy and S. Lanteri. Inverse Design of Photonic Devices with Statistical Learning-Based Global Optimization Algorithms. 46th PIERS (Photonics and Electromagnetics Research Symposium), Abu Dhabi, UAE, 4-8 May 2025
- "A Remarkable Catalyst-Free Photochemical Alkene Hydrophosphination with Bis(trimethylsilyl)phosphonite", E. Bréger, J. Dussart-Gautheret, R. Taakili, F. Fontaine-Vive, C. C. Cummins, R. Benhida, JACS Au, 2025, 5, 6, 2861–2870
- L. Pradi, T. Jiang, M. Feraud, M. Bekbergenova, Y. Taghzouti and L. F. Nothias, An AI Pipeline for Scientific Literacy and Discovery: a Demonstration of Perspicacité-AI integration with Knowledge Graphs, ISWC2025
- Balossi, R. and P. Tanga 2025. Post-impact self-gardening: a poorly known mechanism with consequences on the properties of asteroid surfaces. EPSC-DPS Joint Meeting 2025 (EPSC-DPS2025 EPSC-DPS2025-486. <https://doi.org/10.5194/epsc-dps2025-486>
- Liu, Z., D. Hestroffer, J. Desmars, R. Lallemand, P. David, and L. Liberato 2025. Binary system dynamics and physical property analysis using Gaia astrometry. EPSC-DPS Joint Meeting 2025 (EPSC-DPS2025 EPSC-DPS2025-1274. <https://doi.org/10.5194/epsc-dps2025-1274>
- Liberato, L., D. Mary, P. Tanga, J. Desmars, R. Lallemand, Z. Liu, and D. Hestroffer 2025. Updated method for binary asteroid search in Gaia FPR. EPSC-DPS Joint Meeting 2025 (EPSC-DPS2025 EPSC-DPS2025-814. <https://doi.org/10.5194/epsc-dps2025-814>
- Tores, J., Ancarani, E., Sun, R., Sassatelli, L., Wu, H.-Y., & Precioso, F. (2025). Improving concept-based models for challenging video interpretation. In Proc. ACM Int. Conf. Multimedia, Dublin, Ireland.
- Elisa Ancarani, Julie Tores, Rémy Sun, Lucile Sassatelli, Hui-Yin Wu, and Frederic Precioso. 2025. Leveraging Concept Annotations for Trustworthy Multimodal Video Interpretation through Modality Specialization. In Proceedings of the 1st International Workshop on Cognition-oriented Multimodal Affective and Empathetic Computing (CogMAEC '25). Association for Computing Machinery, New York, NY, USA, 11–19. <https://doi.org/10.1145/3746277.3760415>
- Guinard, A., Fienga, A., Mémin, A., Ganino, C. 2024. Coupled tidal tomography and thermal constraints for probing Mars viscosity profile. AGU Fall Meeting Abstracts.
- ENHANCED EFFICACY OF A SPECIFIC HDAC3 INHIBITOR IN COMBINATION WITH 5-AZACITIDINE AGAINST DIFFUSE LARGE B-CELL LYMPHOMA. Emilio M. Serrano-Lopez, Calixte Calaud, Leonie A. Cluse, Izabela Todorovski, Thomas Mardivirin, Matt Teater, Abdelillah Maarad, Noura Tawfic, Sophia Rutaquio, Peter J. Fraser, Killian Guiraud, David Yoannidis, Tim Semple, Emeline Kerreneur, Arnaud Jacquel, Ari M Melnick, Ricky W. Johnstone, Pilar M. Dominguez. 2025 FASEB Hematological Malignancies conference (Southbridge MA, USA)

- H. Meheut, F. Gerosa, J. Bec. A unique solution to overcome the barriers to planetesimal formation at low dust-to-gas ratio. *Semaine de l'astrophysique française*, 2025, Toulouse, France. ([hal-05343429](#))
- Maëline Baillet, Matteo Geraci, Tiberio Potenza, Théo Salinas, Mohammed Assaba, Florian Cordier, Pierfranco Costabile, Olivier Delestre, Margherita Lombardo and Roger Moussa. Novel methodology for strategic mesh refinement in two-dimensional (2D) rainfall-runoff simulations. SIMHYDRO 2025

Workshops et autres présentations orales

- Planetesimal gravitational collapse in a gaseous environment, Auteurs: Segretain, P.; Méheut, H.; Moreira, M. Présenté à :
 - Idefix User days (Grenoble)
 - Biphasic meeting (Nice)
 - Les disques et la formation planétaire à l'ère du JWST et de l'ELT (Roscoff)
- Présentation de Rila Hilma donnée lors de International Conference on Sustainable Coral Reefs, Manado, North. Sulawesi, Indonesia, 13-15 December 2024.
- How chirality affects the sense of smell ? A study by a combination of AI and sensory analysis. J. Topin, M. Hladiš, M. Lalis, S. A. Khalil, S. Fiorucci, Chirality, Kyoto (Japan) 2024.
- Navigating the unknown: unearthing novel ligand of insect odorant receptor through a predictive docking workflow. A. Comte, M. Lalis, E. Jacquin-Joly, S. Fiorucci. Club de Neurobiologie des Invertébrés, Cif-sur-Yvette, 30 et 31 mai 2024
- In silico approaches expand the chemical space of insect odorant receptors A. Comte, M. Lalis, G. Caballero-Vidal, N. Montagné, C. Meslin, S. Fiorucci, E. Jacquin-Joly. XXVII International Congress of Entomology (ICE2024); Dubai (Quatar), 25-30 August 2024.
- Mahe Cloé, Eric Slezak - Séminaire Institut Interdisciplinaire Intelligence Artificielle (3IA) le 16 mai 2025 - "First robust census of binary supermassive black holes : Deep Learning Techniques applied to the Euclid space mission."
- Colin Brisson, Ayoub Kahfy, Marc Bui, Frédéric Constant, « Named Entity Recognition in Context: Edit_Dunhuang team Technical Report for Evahan2025 NER Competition », Proceedings of the Second Workshop on Ancient Language Processing, pp 176–181.
- Blind Identification of Subcritical Dynamo Equilibria. F. Marcotte, P. Mannix, Y. Ponty, C. Skene, S. Tobias . ICTW Durham 2025.
- L. Pradi, T. Jiang, M. Feraud, M. Bekbergenova, Y. Taghzouti and L. F. Nothias, Perspicacité-AI. Metabolomics Symposium, Jul 2025, Nice, France. ([hal-05233151](#))
- HAMONOU Elliott, PAPAICONOMOU Nicolas, KUNZ Werner, ANTONCZAK, TOURAUD Didier. Coffee, tea and chocolate: which methylxanthine do you prefer? The International Symposium on Gr*een Chemistry ISCG2025, 12-16 mai 2025, La Rochelle, France.

Thèses

- Matej Hladiš, soutenue le 19/06/2024 et intitulée Graph neural networks and protein language models to reveal the combinatorial code of olfaction.
- Maxence Lalis, soutenue le 11/12/2024 et intitulée Numerical and analytical modeling of olfaction.
- Modeling of the interactions between tumor growth and immune response, C. Tayou Fotso, PhD thesis, Univ. Côte d'Azur, july 2025, <https://hal.science/tel-05322358v1>
- Antoine Collin. Annotation cellulaire automatique pour la construction d'un atlas cellulaire. Médecine humaine et pathologie. Université Côte d'Azur, 2024. Français. ([NNT : 2024COAZ6039](#)). ([tel-04991529](#))
- Adrien Simonnin. Thèse de Doctorat: *Exploration de la diversité physique des exoplanètes et des naines brunes à travers la spectroscopie infrarouge à très haute résolution au VLT et à l'ELT*. Ecole Doctorale Sciences Fondamentales et appliquées (EDSFA). Laboratoire Lagrange.
- Thèse de Justine Labory. Développement d'une approche de médecine personnalisée pour améliorer le diagnostic des maladies mitochondrielles. Université Côte d'Azur, 2025. <https://theses.fr/2025COAZ6018>
- Thèse de doctorat de Nicolas Valade. Turbulence quasi-géostrophique de surface. 2025, Université Côte d'Azur. <https://theses.fr/s345052>

Posters

- In silico approaches expand the chemical space of insect odorant receptors, A. Comte, M. Lalis, G. Caballero-Vidal, N. Montagné, C. Meslin, S. Fiorucci, E. Jacquin-Joly, ICE 2024 Kyoto, 25-30 aout 2024
- Combining AI and mass spectrometry to analyze odorant molecules. J.B. Coffin, R. Krafft, S. Antoniotti, S. Fiorucci. 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024
- Navigating the unknown: unearthing novel ligand of insect odorant receptor through a predictive docking workflow. A. Comte, M. Lalis, E. Jacquin-Joly, S. Fiorucci. 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024
- Mahe Cloé - Colloque des doctorants de deuxième année de l'Ecole Doctorale SFA, le 20 mai 2025 - Poster "First robust census of binary supermassive black holes : Deep Learning Techniques applied to the Euclid space mission."
- Poster dans la Conference "Why Galaxies Care About AGB Stars" November 17–21, 2025 in Santiago, Chile. Evolution vers des modèles binaires d'une étoile AGB. Tomassini et al.
- L. Pradi, T. Jiang, M. Feraud, M. Bekbergenova, Y. Taghzouti and L. F. Nothias, Perspicacité-AI: An AI-Pipeline for Accelerating Scientific Education and Research. Metabolomics 2025, Jun 2025, Prague, Czech Republic. ([hal-05232923](#))
- "Automatic Segmentation of Forearm Bones Using Deep Learning Approaches". T. Aguilar Vidal, R. Cremese, R. Winter, I. Costantini, T. Poujade, J.-B. Masson, M.-O. Gauci. The 25th

- Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery (CAOS) in Davos, Switzerland, from June 16th to 19th 2025.
- Segmentation automatique des os de l'avant-bras à l'aide d'approches d'apprentissage profond. Théo Aguilar Vidal ; Jean-Baptiste Masson ; Robin Cremese ; Remy Winter ; Isa Costantini ; Thibault Poujade ; Marc-Olivier Gauci. Troisième édition du Colloque Français d'Intelligence Artificielle en Imagerie Biomédicale (IABM 2025) du 17 au 18 Mars 2025 à Nice.
 - Inès Krawczyk et al., Poster meeting ICON "A systems biology approach to analyze the synchronization of cellular circadian clocks " 2025
 - Inès Krawczyk et al, Poster meeting DYNABIO "A systems biology approach to analyze intercellular coupling between peripheral circadian clocks" 2026
 - KUUSK Katarina-Liyza, BASSARD Jean-Etienne, ANTONCZAK Serge. CHIL, the hidden architect of the isoflavonoid metabolon? XXXII International Conference on Polyphenols, 7-10 July, Turku, Finland.
 - DAVID Lucas, ANTONCZAK Serge, HASELMAIR-GOSCH Christian, HALBWIRTH Heidi. Study of Enzyme substrate specificity by molecular dynamics approaches applied to DFR, a key enzyme in polyphenols biosynthesis. XXXII International Conference on Polyphenols, 7-10 July, Turku, Finland.

Stages et projets étudiants niveau master

- Analyse des données exowinds de KELT-20b: Rapport de stage de Valentin De Lia, master MAUCA (2025).
- Etude de l'atmosphère de WASP-121b et comparaison aux données JWST: rapport de Stage Pablo Drake, master MASS (2025).
- Rapport de stage de Lucie Bolelli « Analyse expérimentale des oscillations biologiques de 12 heures ». Master 1 Sciences de la Vie ; Parcours Bioinformatique et Biologie Computationalle (2025)
- Rapport de stage de Théo Llopis. Deep Learning pour la Détection d'Exoterres. Ecole : ENSTA Bretagne (2025)
- Rapport de stage de Loup Doinel « Apprentissage automatique appliqué au raisonnement argumentatif. », GRENOBLE INP Esisar (2025).
- Rapport de stage de Yani Kedjar « Dynamique de reconnaissance de récepteurs olfactifs : une cartographie à l'échelle moléculaire ». Master Sciences du Vivant : parcours recherche en biologie et santé, U. Lorraine (2025)
- Rapport de stage de Jorge Carreno Donoso "Dynamics and statistics of quantized vortices in a disk". Master OAM: Fundamental physics and applications (2025).

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^{JEDI} 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^{JEDI} ; 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.

Développement d'une approche de médecine personnalisée pour améliorer le diagnostic des maladies mitochondriales

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

Résultats obtenus grâce aux moyens informatiques attribués :

Thèse de Justine Labory. Développement d'une approche de médecine personnalisée pour améliorer le diagnostic des maladies mitochondriales. Université Côte d'Azur, 2025.
<https://theses.fr/2025COAZ6018>

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 Mitochondrielles, 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, UniCA). 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.

Monitoring coral reef health based on non-invasive photo colony analysis by AI machines

Description :

Coral health, like human health, is complex and multifactorial, making it difficult, if not impossible, to describe based on a single variable. Disease and/or bleaching are commonly used as the sole determinant of coral colony health. In accordance with the World Health Organization (WHO) definition of health ("physical, mental and social well-being, not merely the absence of disease"), it is necessary to consider other factors than the presence/absence of pathology to define coral health.

In other words, a reference diagnosis on coral health is still lacking.

A combination of visual attributes related to the health of a coral colony, which can easily be extracted from colony photographs, could be used to define the health status of a given colony: bleaching, contact with algae (macroalgae, turf and CCA), boring organisms, predation, contact with sediment, tissue loss, pigmentation, and disease. Thus, IRCAN analyzed these visual attributes in more than 1000 colonies sampled across the Pacific during the Tara-Pacific expedition for which IRCAN has a vast collection of molecular data. In two different genera (Pocillopora and Millepora), these attributes can explain specific transcriptional signatures linked to coral health and disease (immunity, extracellular matrix, signal transduction...), reinforcing the idea that they can serve as surrogate markers of coral colonies health.

How these coral attributes can be used to define the health status of coral reefs was studied in collaboration with Universitas Papua (UNIPA) / Faculty of Fisheries and Maritime Studies – in partnership with Misool Foundation at sites of contrasting reefs around Manokwari and the south of Misool (intact, damaged, polluted, restored). A statistical analysis showed that the distribution of the 9 attributes among many coral colonies per site is a characteristic of the original reef, allowing image analysis to be used as a rapid way to track health status and the evolution of coral reefs. Thus, in parallel, from the photos taken in Manokwari and Misool, the Maison de la Modélisation, de la Simulation et des Interactions (MSI, University Côte d'Azur) built AI-based machines to automatically recognize the 9 health-related attributes that IRCAN annotated manually underwater.

The effectiveness of these tools in detecting and extracting attributes from coral photos in seconds is comparable to the levels of two different annotators/divers who would manually annotate the same colony after diving. A publication on these results is currently being prepared for submission.

In addition, a database is constructed to analyze photos in real time and with a rendering of the results per listed site. Such an approach should be useful for future coral ecophysiology studies and reef monitoring.

Chimie

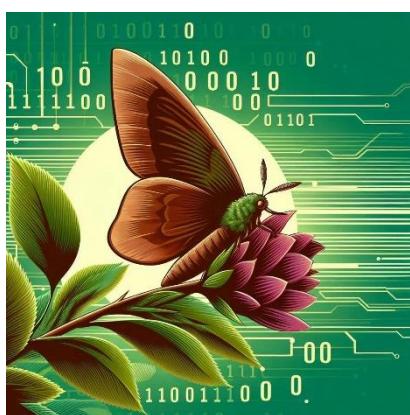
EUR Spectrum

Ecologie chimique inverse : criblage virtuel assisté par l'IA des récepteurs olfactifs d'insectes

Responsable scientifique du projet : Sébastien Fiorucci

Laboratoire : ICN

Description :



Le projet vise à accélérer la découverte de sémiochimiques pour la lutte biologique contre les insectes en utilisant les approches de criblage virtuel guidée par l'IA des récepteurs olfactifs des insectes. L'objectif est de définir l'espace olfactif des espèces étudiées, de mieux comprendre le fonctionnement de leurs récepteurs olfactifs et d'identifier de nouvelles molécules actives sur leur comportement. Les retombées attendues sont importantes dans des domaines aussi variés que la protection des cultures, la santé humaine et animale, la pollinisation ou la bioconservation, via le contrôle ciblé des comportements olfactifs des insectes.

Dynamics of protein-ligand and peptides-ligands elucidated at the molecular and atomic levels

Responsable scientifique : Elise Dumont

Laboratoire : ICN

Résultats obtenus grâce aux moyens informatiques attribués :

- Un article soumis à Inorg. Chem. "Site Specific Study of the Supramolecular Interaction Between Amyloid Beta Peptides and the Polyoxometalate [α -SiW₁₁O₃₉]₈- Using NMR and Computational Methods" , A. Roux, L. Decremoux, C. Aldoma, I. C. Stoltze, S. Bentaleb, X. Chen, G. Bertho, C. Caradeuc, E. Dumont, C. Hureau, S. Blanchard, N. Giraud
- Deux articles en cours de rédaction (avec Xiaojing WU)
- Rapport de stage de Yani Kedjar « Dynamique de reconnaissance de récepteurs olfactifs : une cartographie à l'échelle moléculaire ». Master Sciences du Vivant : parcours recherche en biologie et santé, U. Lorraine (2025)

Description du projet en pages suivantes.

Project summary

This project aims at capturing the dynamics of association of amyloid peptides with lanthanide complexes, in the framework of an ANR programme blanc collaborative project (DREAMY, which began in January 2023). We will seek to investigate the interaction of amyloid peptides with a lanthanide complex, so-called crystallophore, in direct cross-talk with paramagnetic NMR experiments that have already taken place in the lab of Prof. Nicolas Giraud (U. Paris Cité) as well as self-aggregation bioinorganic experiments performed by Dr. C. Hureau (LCC Toulouse).

Context

This project was initiated from several related studies which have established the action of lanthanide complexes (tris-dipicolinate and a new, patterned, hepta-coordinated complex shown in Figure 1 and named Xo4) to assist co-crystallization of proteins (lysozyme, ubiquitin, but also many more complex proteins sometimes refractory to crystallization). Up to now, ca. 40 such proteins have been co-crystallized with Xo4 lanthanide complexes, yet the molecular mechanism into play from the first one-to-one interactions between a given protein and Xo4 remains unknown.

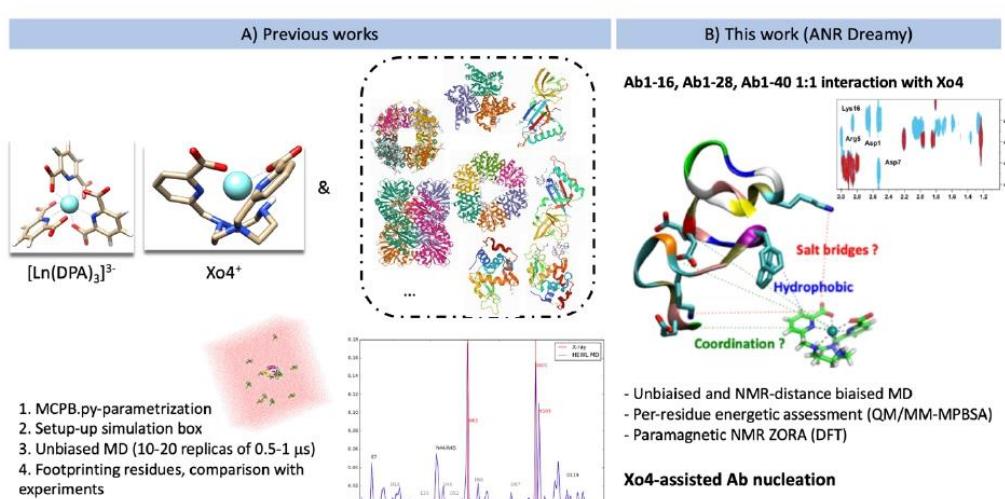


Figure 1. Graphical summary of the Xo4 research line. The previous study on Xo4-protein interactions have grounded a methodology and MCPB.py-based force field parameters for the lanthanide complexes.

Xo4 was designed to interact not only electrostatically as it is observed for other “molecular glues” such as tris-dipicolinate charged -3 and negatively-charged calixarenes [1,2] but to develop coordination with available surface carboxylates (Asp or Glu), but also electrostatic interactions between Xo4’s carboxylate and positively-charged residues (Lys and Arg) and π-stacking interactions with aromatic residues. Xo4 can potentially develop interactions with 7 residues out of 20, and in previous work we have assessed the free energy governing such supramolecular interactions. [3,4]

Beyond the context of protein crystallization “catalysis”, Xo4 can also interact with peptides which can be assessed accurately combining MD simulations and paramagnetic NMR, including the accurate estimation of the binding free energies [3]. We have proposed in the framework of the Dreamy ANR proposal to take advantage of Xo4 to tune the self-assembly of amyloid peptides, due to their biological and medicinal importance. In particular, we think Xo4 can be useful auxiliaries to catch some high-energy intermediates involved in the nucleation, before the formation of fibrils which is somehow more understood. Bioinorganic experiments have corroborated the action of Xo4. Paramagnetic NMR can obtain high-accuracy spectroscopic signatures (TOCSY maps) during the nucleation phase which will be compared to molecular dynamics interactions to propose on a firm basis 3D structure of nuclei.

A first step of this project will be to characterize the 1:1 interaction of Xo4 with three amyloid peptides (Ab-1-16, Ab1-28 and Ab1-40). Preliminary results have been obtained for Ab-1-16 that comfort us in the methodology to describe these supramolecular interactions. Indeed, we note that the lanthanide complex spans several interactions site and can “hope” between stable energy minima, with no need to introduced an explicit bias for the 1:1 interaction. Our parametrization also captures faithfully the carboxylate...Ln binding: an example of agreement between X-ray our dynamic, in solution MD-based footprinting for lysozyme is given in Figure 1. We have obtained such agreement for various proteins, yet peptides are typically more dynamics and hence more difficult to sample. Xo4 can hope between different binding sites, and trajectories of 0.5-2 μ s are insightful with 5-20 replica to assess time residency.

State-of-the-art and objectives

Amyloid self-assembly is a hot topic in computational Biochemistry, and has been investigated widely both experimentally and computationally (also relying on coarse-grained methods). Insights have been gained into the formation of fibrillation towards organized structures, but the upstream formation of nuclei remains elusive (see Figure 2). NMR combined to MD is particularly appealing to catch such fast, rare events especially as lanthanides are introduced as a supramolecular, paramagnetic probe.

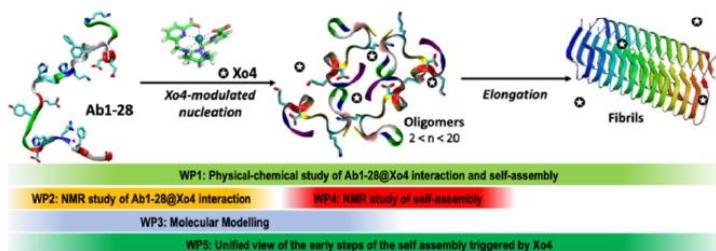


Figure 2 : Scheme of the amyloid peptide Ab1-28 (PDB 1HZ3) and the Xo4 along nucleation and elongation steps leading to fibrils formation.

We will first investigate more in-depth the 1:1 interaction between Xo4 (which co-exists as two enantiomers Δ and Λ), which we already obtained for Ab1-16 according to our preliminary simulations. Structures are shown in Figure 3 leftside. They are in neat agreement

with the 2D NMR TOCSY spectra, which are difficult to interpretate and call for MD simulations footprints: in turn, our preliminary MD simulations have already guided his assignment. For Ab1-28, our first MD exploration is also in good agreement. For Ab1-40, the NMR spectra is too elusive and our MD simulations will be important to provide guidance for NMR assignment. We aim at rationalizing this supramolecular interaction before to tackle the Xo4-assisted nucleation, which is more challenging and computationally-demanding. Our simulations are performed with Amber 2022 program, to benefit from its efficient GPU performance but also to be able to introduce as restraints the distances inferred from NMR in the second step of the project.

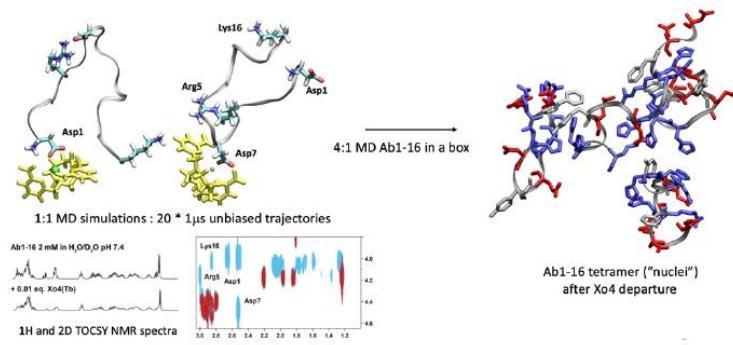


Figure 3 : Xo4-Ab1-16 interaction (one-to-one)

The main objective of the project will be to generate structures of nuclei, i.e. between 6-20 peptides to compare with NMR data. Such simulations require a high sampling to be able to catch high-lying energy intermediates. Preliminary calculations on the shortest peptide, Ab1-16, have provided a few 3D structures of tetramers as we have performed as test calculations for 10 replica of 2 μ s each. Such simulations typically require large water boxes (described with the OPC force field to be compatible with the IDP specific force field used for amyloid peptides). [5] We note that such brute-force, all-atom MD simulations to investigate self-assemble processes have been attempted, up to 50 peptides [6]. The direct input from NMR experiments will be crucial to guide our simulations, and for the larger oligomers whereas unbiased MD can provide structures and characterize key residues up to n=4 as suggested by our preliminary MD, even if it is not yet clear for which size of the amyloid oligomer the input from NMR will become necessary. Solving purely amyloid tetramers by all-atom, explicit-solvent MD simulations [6b] even for the full-length Ab1-42 amyloid peptide back in 2016. This tends to proves the feasibility of such dynamics, together with our preliminary simulations on Ab1-16.

Classical MD simulations will be refined at the QM/MM-MD level to restore a “true” lanthanide center and refine the description of the lanthanide f electrons prior to NMR DFT calculations (relying on DFT+ZORA Hamiltonian, and performed in collaboration with Dr. B. Le Guennic as the ADF code is the sole QM code that can provide computational paramagnetic NMR chemical shifts). These calculations will be performed on the National cluster.

A recruited post-doctoral fellow (Dr. Xiaojing WU) will begin in March 2024 and will be in charge of most simulations and their analysis.

Outcomes

We already have secured a few preliminary results that have guided the NMR interpretation and that are currently under redaction. We expect to publish independently the 1:1 interaction, the NMR-orientated, methodological approach to generate on-the-fly paramagnetic spectra and the self-assembly study. The post-doctoral fellow to be recruited will also present her results in conferences.

References:

- [1] P. B. Crowley, Acc. Chem. Res., 2022, 55 :2019 [doi](#)
- [2] S. Engilberge et al., ACS Nano, 2019, 13:10343 [doi](#)
- [3] E. Dumont et al., PCCP 2013, 15:18235 [doi](#) ; S. Denis-Quanquin et al., PCCP 2021, 23 :11224
- [4] S. Engilberge et al., Chem. Eur. J., 2018, 24 :9737 [doi](#)
- [5] P. S. Shabane et al., J. Chem. Theor. Comput., 2019, 15 :2620 [doi](#)
- [6] a) D. Matthes et al., Sci. Rep., 2016, 6:33156 [doi](#) ; b) A. M. Brown, Biophys. J., 2016, 111:937 [doi](#) ; c) S. Jalali et al., J. Mol. Liq., 2022, 347 :118283 [doi](#)
- [7] A. Bartocci et al., J. Chem. Inf. Model., 2022, 62 :6739 [doi](#)

Simulations Numériques de Systèmes Biomoléculaires Impliqués dans la Biosynthèse de Métabolites Spécialisés

Responsable scientifique : Serge Antonczak

Laboratoire : ICN

Résultats obtenus grâce aux moyens informatiques attribués :

Conférences et posters (en lien avec les ressources d'Azzurra-NEF). Azzurra a été remercié à l'occasion des présentations suivantes :

*** Oral Communications :

HAMONOU Elliott, PAPAICONOMOU Nicolas, KUNZ Werner, ANTONCZAK, TOURAUD Didier. Coffee, tea and chocolate: which methylxanthine do you prefer? The International Symposium on Gr*een Chemistry ISCG2025, 12-16 mai 2025, La Rochelle, France.

*** Poster Communications :

KUUSK Katarina-Liyza, BASSARD Jean-Etienne, ANTONCZAK Serge. CHIL, the hidden architect of the isoflavanoid metabolism? XXXII International Conference on Polyphenols, 7-10 July, Turku, Finland.

DAVID Lucas, ANTONCZAK Serge, HASELMAIR-GOSCH Christian, HALBWIRTH Heidi. Study of Enzyme substrate specificity by molecular dynamics approaches applied to DFR, a key enzyme in polyphenols biosynthesis. XXXII International Conference on Polyphenols, 7-10 July, Turku, Finland.

Description :

Ce projet vise à mieux comprendre l'organisation des enzymes de la famille de la biosynthèse des polyphénols dans le règne végétal. Diverses méthodes de modélisation moléculaires seront mises en œuvre pour obtenir une vision rationnelle de leur organisation et de leur efficacité réactionnelle.

Contexte :

Les composés de la famille des polyphénols, et plus particulièrement de la famille des flavonols, des flavonoïdes ou des anthocyanes, sont produits au sein des plantes à la suite

de cascades enzymatiques. Ces composés sont importants pour les plantes car, selon leur nature, ils protègent le végétal des agressions extérieures (résistance aux composés radicalaires ou oxydants, protection contre les agressions animales, résistance aux UV...) et permettent leur bon développement (attraction des polliniseurs...). De par leurs propriétés, ces composés intéressent tout particulièrement la filière cosmétique mais peuvent également présenter de l'intérêt dans le cadre de la filière santé (prévention contre les maladies cardiovasculaires...). Leur présence et même leur concentration dans les cellules végétales sont dépendantes de l'organisme et de ses cycles de maturation.

En effet, partant d'un précurseur commun, une suite de transformations enzymatiques au sein de la cellule permet la production de ces composés (voir Fig1).

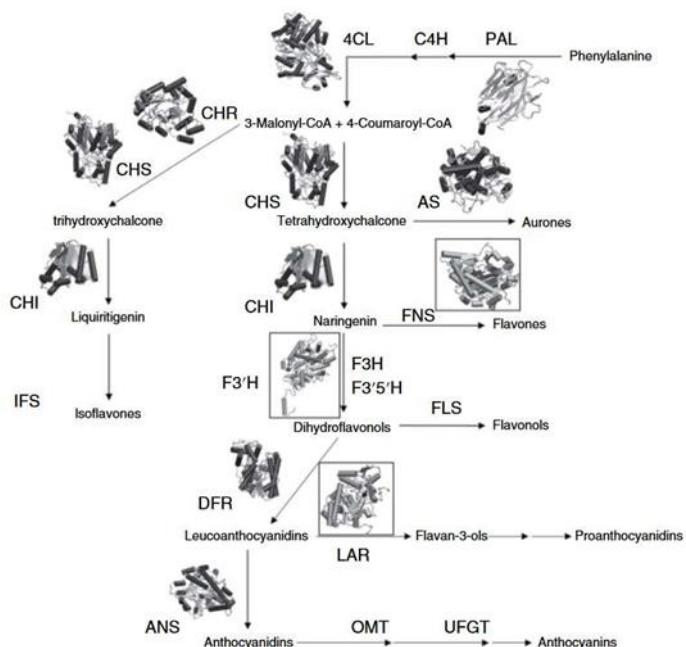


Fig1 : Schéma partiel des chemins de biosynthèse des polyphenols. voir ref 1.

Les enzymes impliquées à chaque étape de cette biosynthèse présentent de subtiles différences structurales qui entraînent des spécificités de substrat menant à la production de certains métabolites spécifiques au dépend d'autres de la même famille². Ainsi, chaque organisme végétal, génétiquement unique, possède ses propres spécificités qu'il est important de décrypter afin de comprendre le schéma général de l'intérêt de ces métabolites spécialisés pour les plantes. De récentes investigations ont par ailleurs mis en avant la formation de complexes multienzymatiques, transitoires, qui permettraient

d'établir une efficacité générale plus importante de la production de ces polyphénols, mais aussi seraient susceptibles de protéger en les séquestrant les molécules intermédiaires de réactions les plus « fragiles ». Ces complexes multienzymatiques sont appelés métabolon (Fig2).

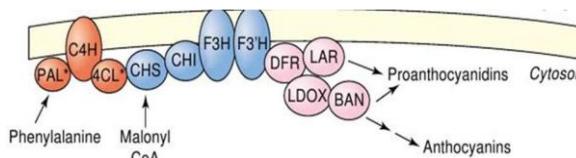


Fig2 : Organisation d'un métabolon présent dans la biosynthèse des anthocyanes. Adapté de la référence 3

Il apparaît désormais incontournable de devoir prendre en compte non plus les enzymes isolées pour une étude globale de la formations des polyphénols mais plutôt de tenir compte de l'ensemble de la machinerie enzymatique. Ceci est d'autant plus pertinent car lors de la formation de ces complexes multienzymatique, chaque enzyme est soumise à des interactions provenant de la proximité des enzymes voisines, modifiant de fait sa structure et potentiellement modifiant de façon plus ou moins importante son efficacité.

Par ailleurs, l'environnement moléculaire eu sein de la cellule végétale est associé à de fortes concentrations en molécules naturelles qui peuvent présenter des propriétés de solvatation ou de solubilisation très particulières. Ces dernières années ont vu l'émergence des Solvants Eutectiques Profonds (DES pour Deep Eutectic Solvent) constitués à partir de molécules naturelles, et qui viennent en remplacement des solutions élaborées à bases de solides ioniques (voir tableau 1).

Eutectic mixtures		Component 1	Component 2	Molar Ratio	Water content (wt%)	Density (g/mL)
Polar	GG	Glycerol	Glucose	2:1	20 %	1.303 ± 0.002
Non-polar	MC8	Menthol	Octanoic acid	5:1	—	0.885 ± 0.003
	MC12		Lauric acid	4:3	—	0.881 ± 0.003
	MO		1,2-Octanediol	1:1	—	0.907 ± 0.001
	MP		1,3-Propanediol	1:2	—	0.963 ± 0.005
	MT		Thymol	1:1	—	0.931 ± 0.002

Tableau 1 : composition de NaDES à partir de molécules hydrophiles et hydrophobes, tiré de la ref 4.

Ces nouveaux solvants peuvent être définis pour jouer le rôle d'extractant de molécules naturelles à haute valeur ajoutée (principes cosmétiques, compléments alimentaires...), les polyphénols qui sont le centre de nos investigations font partie de ces molécules qui intéressent ces domaines pharmaceutiques et cosmétiques. De plus, dans nos systèmes,

les métabolites secondaires produits par les plantes, au-delà d'être des molécules cibles à extraire, peuvent également être des constituants actifs de DES formés naturellement : les NaDES (Natural Deep Eutectic Solvent). De par les propriétés spécifiques qu'ils confèrent à l'environnement, il est supposé qu'il puissent jouer un rôle sur la structure et sur les efficacités des enzymes impliquées à chaque étape de la production de ces métabolites secondaires de la famille des polyphénols.

En résumé, il est nécessaire d'étudier chacune des étapes enzymatiques, les potentialités d'agrégation des enzymes mais aussi l'influence de l'environnement pour décrire précisément ces systèmes.

Dans ce cadre, nous profitons actuellement de trois financements déjà acquis pour lesquels trois étudiants en thèse développeront des activités de modélisation de ces systèmes : i) dans le cadre d'une Miti CNRS, nous bénéficions d'une thèse en cotutelle avec l'IBMP de Strasbourg où le travail de la doctorante est dédié à l'étude de l'importance de protéines chaperonnes impliquées dans la stabilisation du métabolon des isoflavonoides, ii) d'une thèse en cotutelle avec l'Université de Regensburg, Allemagne, cofinancée par l'EUR Spectrum sur l'importance des NaDES dans la solubilisation et l'extraction des polyphénols, iii) d'un financement de thèse ANR dans le cadre d'une ANR binationale France-Autriche au cours de laquelle le doctorant devra modéliser la spécificité d'une enzyme en fonction de l'organisme dont elle est issue mais aussi en fonction des effets de son environnement (enzymatique et de solvatation).

Deux autres projets sont déposés, dans le cadre des appels ANR, et se focaliseront plus particulièrement sur l'importance des NaDES sur la structure et l'efficacité des enzymes ou complexes enzymatiques impliqués dans ces voies de biosynthèse.

Justification de la demande :

i) Dans le cadre du financement Miti Prime80, Katarina Kuusk, doctorante Prime80, développe une recherche à la fois expérimentale (au sein du laboratoire IBMP du CNRS, à Strasbourg) et théorique (au sein de notre unité). Cette recherche vise à définir quel rôle joue une protéine chaperonne spécifique, la protéine CHIL, à la fois sur la stabilisation de l'édifice multienzymatique impliqué dans la biosynthèse des isoflavonoides, mais aussi sur l'efficacité des enzymes qui sont en interaction avec celle-ci. Dans le cadre des travaux de

modélisation qu'elle mène au sein de notre unité, elle est amenée à réaliser de nombreuses simulations de dynamique moléculaire, à la fois sur les enzymes isolées, mais aussi sur les complexes multienzymatiques, en interaction avec la membrane. Nous avons décidé dans un premier temps de nous limiter à un complexe comprenant 4 enzymes (CHS-CHI-IFS et CHIL) en interaction avec la membrane cellulaire mais nous n'excluons pas d'augmenter ce nombre (voir Fig 3).

Ainsi, les systèmes moléculaires comprennent de 60000 atomes (une seule enzyme considérée) à plus de 500000 atomes (4 enzymes et la membrane). De plus, les simulations de dynamique moléculaires devront non seulement être longues afin de pouvoir capturer les événements lent de réarrangement

structuraux mais aussi répétées afin d'augmenter l'ergodicité de nos résultats et de s'assurer que nous échantillonons les événements recherchés.

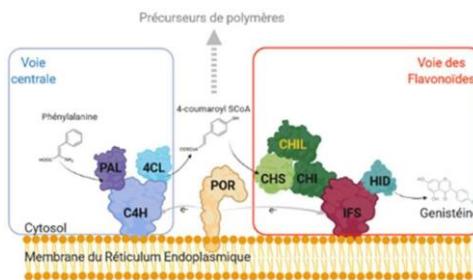


Fig 3. Voie de biosynthèse simplifiée des phénylpropanoïdes et localisation subcellulaire des enzymes d'intérêt et de la protéine non-catalytique CHIL

Méthodes employées : dynamiques moléculaires (MD) classiques avec les logiciels AMBER ou GROMACS sur GPU. Calculs quantique pour élaborations de paramètres avec le logiciel Gaussian sur CPU.

ii) Dans le cadre de notre projet financé par l'EUR Spectrum, Eliott Hamonou, doctorant ayant débuté sa thèse au 1er décembre 2023, sera amené à étudier des mélanges complexes de molécules naturelles, notamment de la famille des polyphénols. Les propriétés particulières de ces molécules, à la fois très hydrophobes du fait des leurs cycles conjugués mais aussi permettant de nombreuses liaisons hydrogène, du fait de leurs fonctions hydroxyles, en font des candidates de choix, à la fois pour tenter de les extraire de matrices naturelles mais aussi pour être elles-mêmes des molécules

participant à la création de ces solvant particuliers. Ces mélanges de molécules naturelles, constituant de nouveaux solvants « verts », appellés NaDES (Natural Deep Eutectic Solvent), font l'objet d'un intérêt grandissant comme alternative des solvants classiques, à base de molécules organiques possédant parfois des propriétés délétères pour l'environnement.

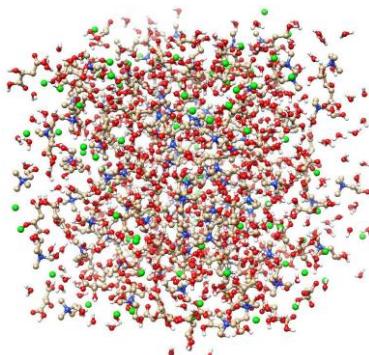


Fig 4. Boîte de solvant de type NaDES à base de deux molécules naturelles, le chlorure d'acétylcholine et l'acide malique

Le travail d'Elliott consistera alors à simuler les interactions intermoléculaires au sein de ces solvants afin d'en comprendre leurs propriétés (voir Fig 4.). Pour cela, il sera amené non seulement à tester un nombre de couples de molécules conséquent, en fonction de ses résultats expérimentaux, mais aussi de vérifier quels impacts peut avoir le ratio entre celles-ci. Ainsi, de nombreuses simulations de dynamiques moléculaires devront être conduites sur des systèmes comportant de 60000 à 100000 atomes. Parallèlement, il devra mener de nombreux calculs de Chimie quantique afin de décrire très finement les interactions entre ces molécules. Nous envisageons de réaliser des calculs hybrides QM/MM-MD afin d'évaluer l'impact de l'environnement sur les structures électroniques des molécules naturelles retenues.

Méthodes employées : dynamique moléculaire (MD) classiques avec les logiciels AMBER ou GROMACS sur GPU. Calculs quantique pour élaborations de paramètres avec le logiciel Gaussian sur CPU. Calculs hybrides de types QM/MM-MD (quantique/classique) avec le logiciel AMBER sur CPU.

iii) Dans le cadre de notre ANR franco-autrichienne, Lucas DAVID, doctorant embauché en octobre 2023 va être amené à déchiffrer les raisons de la spécificité de substrat de l'enzyme DFR. Cet enzyme, en fonction de l'organisme végétal d'où elle provient, ne catalyse pas la transformation des mêmes substrats avec la même efficacité. Pour cela, il devra multiplier les dynamiques moléculaire enzyme- substrat tenant compte d'une quinzaine de mutations possible au sein de l'enzyme et en tenant compte de trois substrats différents (voir Fig 5.). Il sera amené également à rechercher si des modifications allostériques de l'enzyme dues à des interactions avec d'autres protéines (complexe multienzymatique) entraînent un altération des taux de conversion. Par ailleurs, il analysera l'impact que pourraient avoir des concentrations élevées en molécules naturelles (NaDES) dans l'environnement de cette enzyme lors de la réaction enzymatique. Les systèmes moléculaires comprendront de 60000 à 200000 atomes.

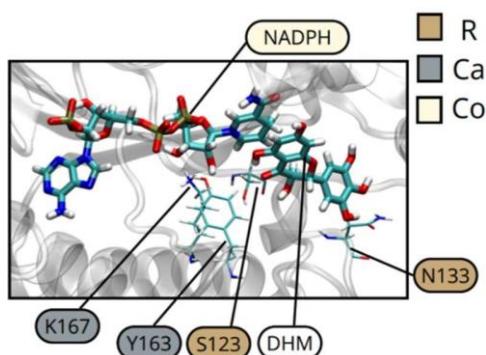


Fig 5 : organisation du site actif de l'enzyme DFR

Méthodes employées : dynamique moléculaire (MD) classiques avec les logiciels AMBER ou GROMACS sur GPU. Calculs hybrides de types QM/MM-MD (quantique/classique) avec le logiciel AMBER sur CPU. Calculs quantique pour élaborations de paramètres avec le logiciel Gaussian sur CPU.

iv) les projets que nous déposons actuellement (notamment ANR), en collaboration avec les Universités de Strasbourg et de Rennes visent à rechercher quelles influence pourraient avoir des solutions de type NaDES dans l'environnement immédiat de système moléculaires de type métabolon (voir Fig 6.).

Dans ce projet, nous profiterons des résultats acquis lors des recherches développées ci-dessus. Nous prévoyons l'embauche d'un(e) doctorant(e) pour réaliser ce travail. Toutes les

méthodes employées ci-dessus seront mises en place pour cette étude. Dans un premier temps, nous nous attacherons à comprendre l'effet de NaDES sur la structure et l'efficacité des enzymes isolées, puis nous tenterons de comprendre l'organisation de ces NaDES à proximité d'un métabolon ancré à la membrane. Les systèmes moléculaires pourraient atteindre plus de 600000 atomes à terme.

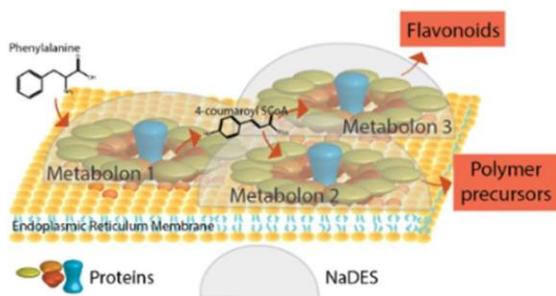


Fig 6 : implication des NaDES dans l'environnement des métabolons.

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'une enzyme ou d'un complexe multienzymatique 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 600 000 atomes pour des systèmes impliquant des complexes enzymatiques et la membrane. Nous étudions les changements de conformations qui ont lieu lors des processus de liaison d'un polyphénol avec une enzyme. Cela nécessite des simulations de dynamique moléculaire de plusieurs microsecondes. Pour cela nous utilisons les logiciels Amber et Gromacs. Amber est déjà installé sur les serveurs de l'Université, 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. Pour cela nous utilisons les logiciels Modeller et Autodock, également installés 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.

1 J. Diharce, S. Antonczak, How to model a metabolon, Recent Advances in Polyphenol Research, 2019, vol. 6, Chap 13. Ed. John Wiley& Sons.

2 J. Vaino, S. Mattila, S. M. Abdou, N. Sipari, T. H. Teeri, Petunia dihydroflavonol is only a few amino-acids away from producing orange pelargonidin-based anthocyanins, *Front. Plant Sci.*, 2023, 14:1227219. DOI 10.3389/fpls.2023.1227219.

3 K. Jorgensen, A.V. Rasmussen, M. Morant, A. H. Nielsen, N. Bjarnholt, M. Zagrobelny, S. Bak, B. L. Moller. Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Current Opinion in Plant Biology* 2005, 8:280–291. DOI 10.1016/j.pbi.2005.03.014.

4 S. Hilali,L. Van Gheluwe, M. Yagmur, L. Wils, M. Phelippe, B Clément-Larosière, B. Montigny, J. Jacquemin, E. Thiery, L. Boudesocque-Delaye. NaDES-based biorefinery of Spirulina (*Arthrospira platensis*): A new path for sustainable high value-added metabolites. *Separation and Purification Technology*, 2024, 329, 125123. <https://doi.org/10.1016/j.seppur.2023.125123>.

Deep Reinforcement Learning for Mass Spectrometry Data Acquisition in Metabolomics

Responsable scientifique : Louis-Félix Nothias

Laboratoire : ICN

Résultats obtenus grâce aux moyens informatiques attribués :

- L. Pradi, T. Jiang, M.Feraud, M. Bekbergenova, Y. Taghzouti and L. F. Nothias, An AI Pipeline for Scientific Literacy and Discovery: a Demonstration of Perspicacité-AI integration with Knowledge Graphs, ISWC2025
- L. Pradi, T. Jiang, M.Feraud, M. Bekbergenova, Y. Taghzouti and L. F. Nothias, Perspicacité-AI: An AI-Pipeline for Accelerating Scientific Education and Research. Metabolomics 2025, Jun 2025, Prague, Czech Republic. (hal-05232923)
- L. Pradi, T. Jiang, M.Feraud, M. Bekbergenova, Y. Taghzouti and L. F. Nothias, Perspicacité-AI. Metabolomics Symposium, Jul 2025, Nice, France. (hal-05233151)

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 :

1. Soutenance de thèse :

- Matej Hladiš, soutenue le 19/06/2024 et intitulée Graph neural networks and protein language models to reveal the combinatorial code of olfaction.
- Maxence Lalis, soutenue le 11/12/2024 et intitulée Numerical and analytical modeling of olfaction.

2. Publications dans des revues à comités de lecture :

- Combining machine learning and electrophysiology for insect odorant receptor studies. A. Comte, S. Fiorucci, E. Jacquin-Joly, in Olfactory receptors: methods and protocols, Methods Mol. Biol. (2025), accepted.
- Accelerating Ligand Discovery for Insect Odorant Receptors. A. Comte, M. Lalis, L. Brajon, R. Moracci, N. Montagné, J. Topin, E. Jacquin-Joly, S. Fiorucci. Int. J. Biol. Sci. 2025, 21(5), 2101-2117.
- Ligand discrimination in hOR1A1 based on the capacitive response. A. Lagunas, C. Belloir, M. Lalis, L. Briand, J. Topin, P. Gorostiza, J. Samitier. Biosens Bioelectron. 2025, 271, 117000.
- Study and Optimization of the Selectivity of an Odorant Binding Protein-based Bioelectronic Nose. M. El Kazzy, M. Lalis, C. Hurot, J.S. Weerakkody, R. Mathey, C. Saint-Pierre, A. Buhot, T. Livache, J. Topin, L. Moitrier, C. Belloir, L. Briand, Y. Hou. Biosens. Bioelectron 2025, 268, 116879.
- A Comprehensive Analysis of Olfactory Receptor Deorphanization. M. Lalis, M. Hladiš, S. Abi Khalil, C. Deroo, C. Marin, M. Bensaïf, N. Baldovini, L. Briand, S. Fiorucci, J. Topin. Chem. Senses 2024, 49, bjae037.

3. Conférences sur invitations :

- A Deep Metric Learning Approach for Olfactory Mixture Discriminability. M. Lalis, S. Fiorucci, J. Topin, M. Hladiš. RECOMB/ISCB Conference on Regulatory & Systems Genomics with DREAM Challenges (RSGDREAM 2024). Madison (Wisconsin, USA), 1-3 October 2024.
- Molecular basis of mammalian sweet taste detection. S. Fiorucci, L. Briand. 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024
- Olfaction under the lens of a computational microscope. J. Topin, J. Pacalon, M. Hladiš, C. Belloir, L. Briand, S. Fiorucci. 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024.

- Encoding Odor: Bridging Smell and Combinatorial code using M2OR database. M. Lalis, M. Hladiš, S. A. Khalil, L. Briand, S. Fiorucci, J. Topin, 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024.

4. Communications orales :

- How chirality affects the sense of smell ? A study by a combination of AI and sensory analysis. J. Topin, M. Hladiš, M. Lalis, S. A. Khalil, S. Fiorucci, Chirality, Kyoto (Japan) 2024.
- Navigating the unknown: unearthing novel ligand of insect odorant receptor through a predictive docking workflow. A. Comte, M. Lalis, E. Jacquin-Joly, S. Fiorucci. Club de Neurobiologie des Invertébrés, Gif-sur-Yvette, 30 et 31 mai 2024
- In silico approaches expand the chemical space of insect odorant receptors A. Comte, M. Lalis, G. Caballero-Vidal, N. Montagné, C. Meslin, S. Fiorucci, E. Jacquin-Joly. XXVII International Congress of Entomology (ICE2024); Dubai (Quatar), 25-30 August 2024.

5. Posters :

- In silico approaches expand the chemical space of insect odorant receptors, A. Comte, M. Lalis, G. Caballero-Vidal, N. Montagné, C. Meslin, S. Fiorucci, E. Jacquin-Joly, ICE 2024 Kyoto, 25-30 aout 2024
- Combining AI and mass spectrometry to analyze odorant molecules. J.B. Coffin, R. Krafft, S. Antoniotti, S. Fiorucci. 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024
- Navigating the unknown: unearthing novel ligand of insect odorant receptor through a predictive docking workflow. A. Comte, M. Lalis, E. Jacquin-Joly, S. Fiorucci. 19th International Symposium on Olfaction and Taste (ISOT), Reykjavik (Islande), 22 et 26 juin 2024

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).

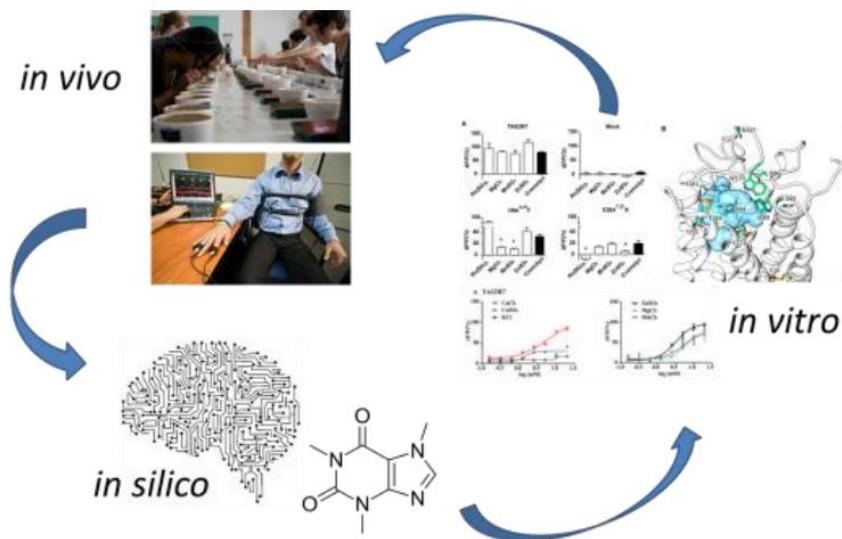


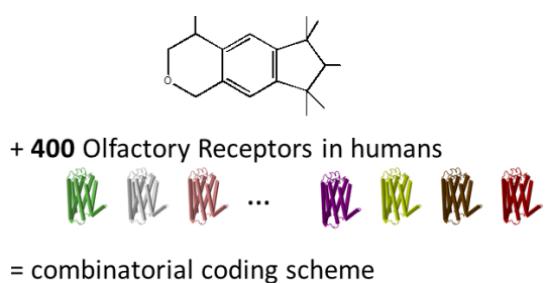
Figure 1 Synergie de l'approche pluridisciplinaire

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 savide) 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 infinie¹ (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

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

Figure 2 Codage combinatoire de la perception des odeurs



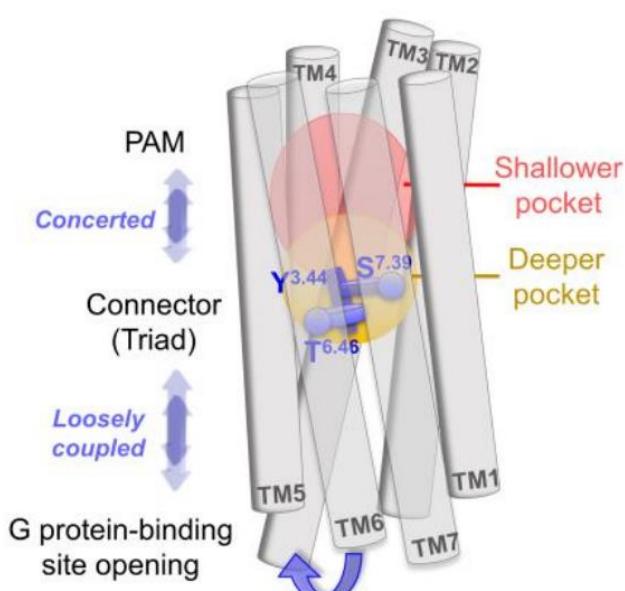
Liens chémo-génomiques de ces espaces de grande dimensionnalité et de mettre au point des modèles moléculaires et numériques capables de prédire de façon rationnelle des composés olfactifs et sapides nouveaux.

Faits marquants : Les travaux de l'équipe ont notamment permis de mieux comprendre les relations entre la structure d'un composé et son activité biologique² et permis d'élargir le spectre de reconnaissance de récepteurs olfactifs humains³, de poissons⁴ ou encore d'insectes⁵ mais également de récepteurs gustatifs humains^{6,7}.

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

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



² Bushdid et al. Science, 2014, 343, 1370-1372

³ Licon et al. Plos Comput. Biol. 2019, 15(4), e1006945

⁴ Cong et al. J. Biol. Chem., 2019, 294, 6762-6771

⁵ Caballero-Vidal et al. Sci. Rep. 2020, 10, 1655

⁶ Chéron et al. Food. Chem. 2017, 221, 1421-1425

⁷ Bouyssel et al. Food. Chem. 2020, 324, 126864

même en l'absence de ligand, i.e. état basal) 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^{8 9 10 11} mais également sur la famille des récepteurs gustatifs^{12 13} et plus largement sur les RCPG^{14 15 16}, 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

⁸ Yu et al. PNAS, 2015, 112(48), 14966-1497

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

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

¹¹ Bushdid et al. CMLS, 2019, 76(5), 995-1004

¹² Chéron et al. Chem. Senses, 2019, 44, 303-310

¹³ Wang et al. Chem. Senses, 2019, 44, 339-347

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

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

¹⁶ Cong et al. PCCP, 2018, 20, 24915-24920

parfums ou l'industrie agroalimentaire, particulièrement bien représentés dans la région PACA. Notre travail de recherche mobilise de nombreux partenaires au niveau local 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.

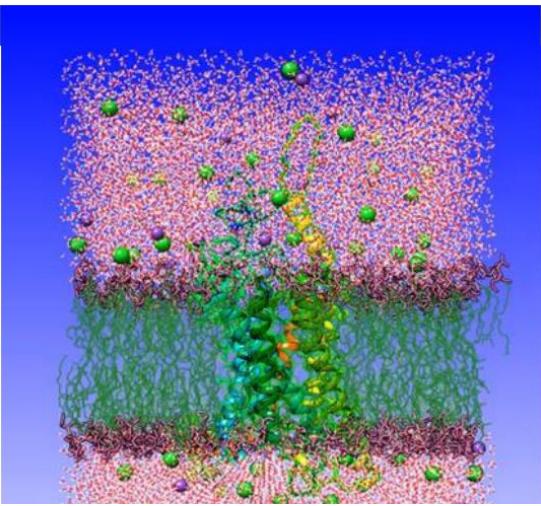


Figure 4 Boîte de simulation d'un RCPG dans environnement physiologique.

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 sapides 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.

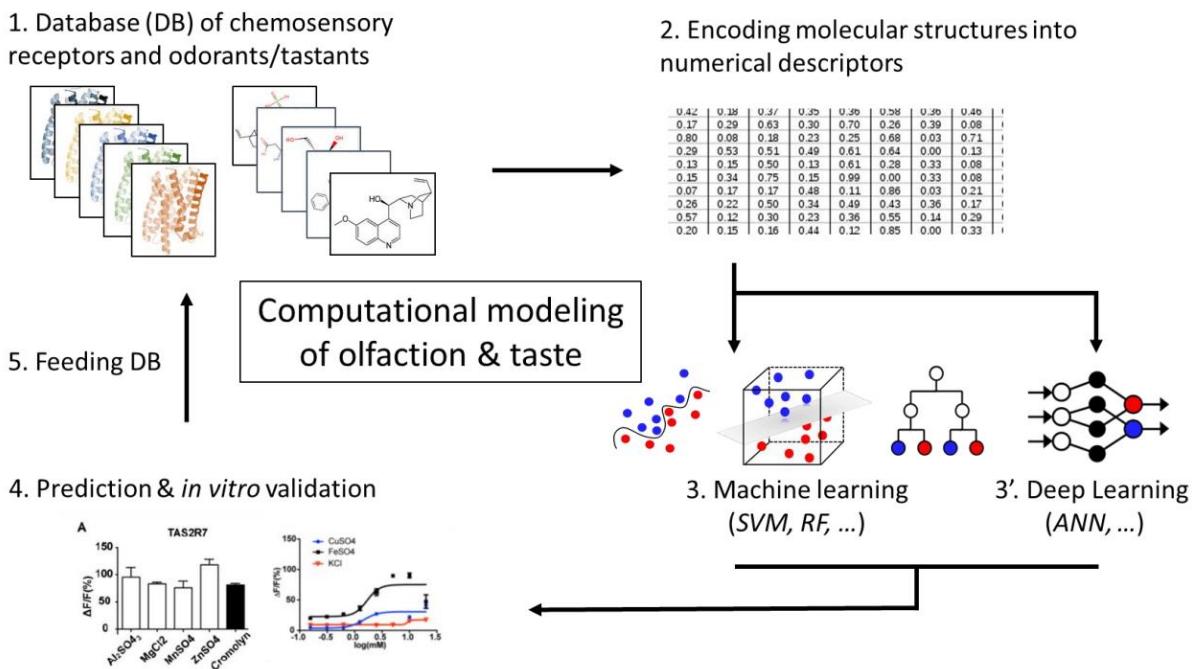


Figure 5 Différentes étapes de la construction du modèle numérique prédictif basé sur des approches d'apprentissage automatique.

Modélisation de molécules d'intérêt biologique, catalytique et éco-compatibles

Responsable scientifique : Fabien Fontaine-Vive

Laboratoire : ICN

Résultats obtenus grâce aux moyens informatiques attribués :

"A Remarkable Catalyst-Free Photochemical Alkene Hydrophosphination with Bis(trimethylsilyl)phosphonite", E. Bréger, J. Dussart-Gautheret, R. Taakili, F. Fontaine-Vive, C. C. Cummins, R. Benhida, *JACS Au*, 2025, 5, 6, 2861–2870

"Development of a reliable, sensitive, and convenient assay for the discovery of new eIF5A hypusination inhibitors", O. Benaceur, P. Ferreira Montenegro, M. Kahi, F. Fontaine-Vive, N. M. Mazure, M. Mehiri, F. Bost, P. Peraldi, *PLOS One*, 2025, 20(2), e0308049

"CuMg₂GaS₄: A novel photocatalyst with promising properties", Hocine H., Khelfaoui F., Al-Douri Y., Boudia K., Fontaine-Vive F. et al., *Multidiscipline Modeling in Materials and Structures*, 2025, 21 (3): 501–519

"Silver- and Gold-Catalyzed Divergent Cascade Cycloisomerization / [3 + 2] versus [2+2+1] Cycloaddition towards a Stereoselective Access to Heterohexacyclic Derivatives", E. Gentilini, A. Bouschon, V. Davenel, F. Fontaine-Vive, J.-M. Fourquez, V. Michelet, *Chemical Science*, 2025

"Ruthenium-Catalyzed Enyne Metathesis: an Entry to Functionalized Azaborine Heterocycles", Mao K., Fontaine-Vive F., Melot R, Michelet V., *Organic Letters*, 2025, 27, 12, 2952–2957

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'.

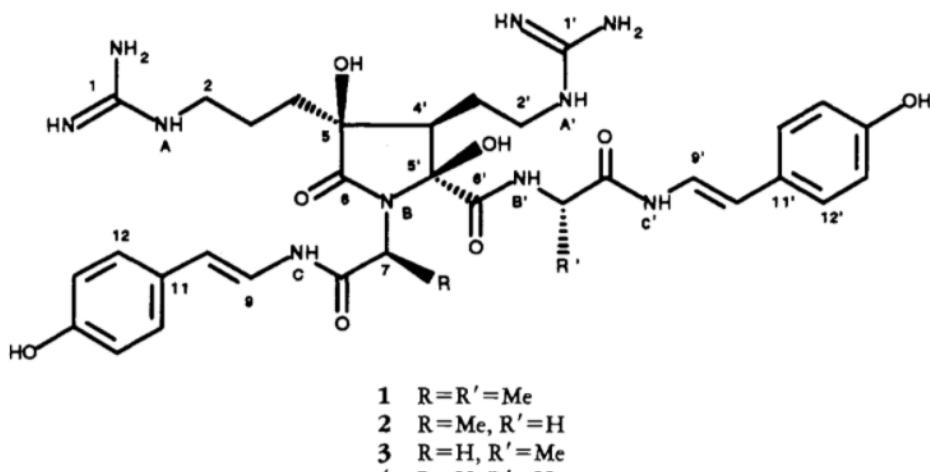


Fig. 1. Stéréochimie relative des molécules ancinopeptolides extraites d'éponges marines

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.

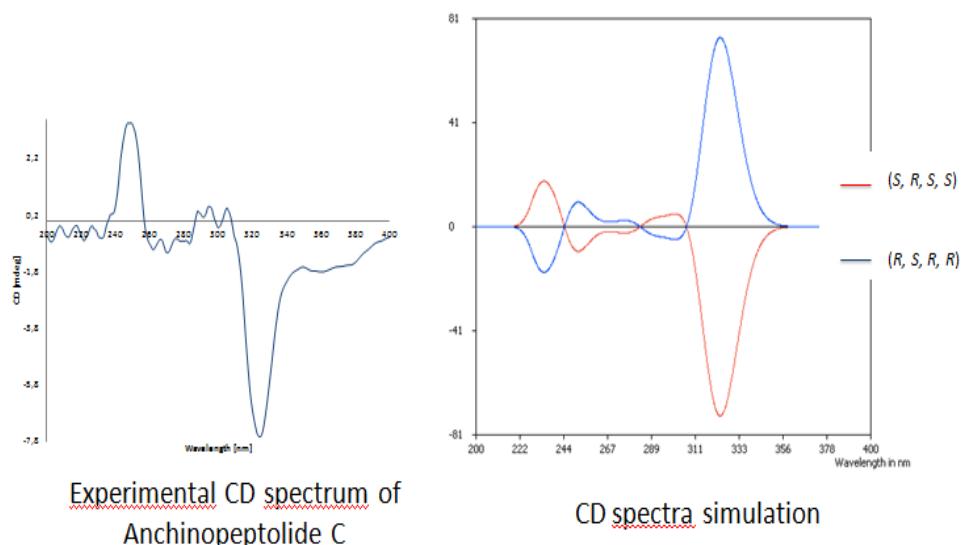


Fig. 2. Spectres ECD expérimental et calculé par TD-DFT de la molécule chirale ancinopeptolide C.

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^{17,18,19}.

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.

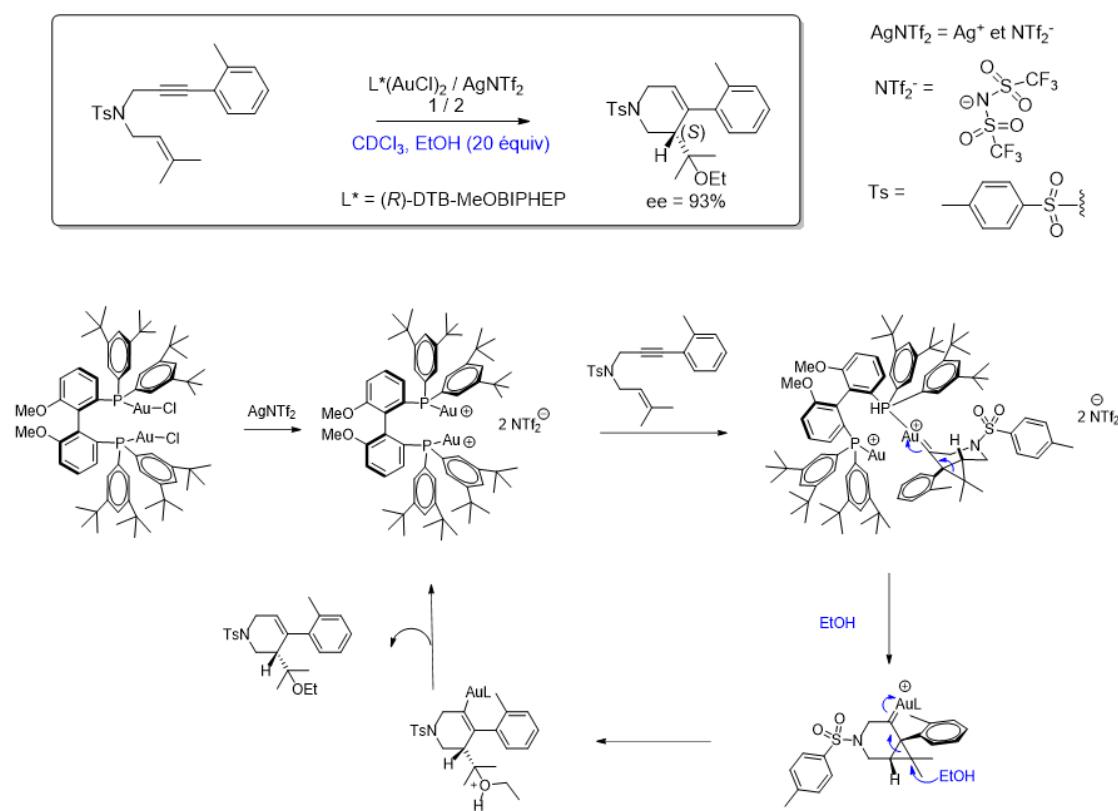


Fig. 3. Mécanisme de cyclisation asymétrique catalysée par des complexes organiques d'or

¹⁷ 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).

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

¹⁹ 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* 2014;80-PIL72.

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

Pour l'équipe « Matériaux et Polymères Eco-Compatible », les calculs DFT dans l'état solide avec le logiciel VASP nous permettront d'interpréter les spectres vibrationnels et de prédire la structure de polymères biosourcés destinés à remplacer les dérivés de produits pétroliers. Les changements conformationnels induits par des étirements pourront ainsi

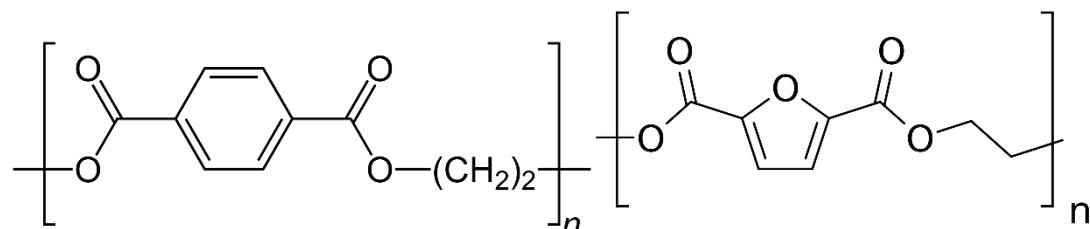


Fig. 4. Unités de répétition de PET (gauche) et PEF (droite)

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

²⁰ 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.

²¹ 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.

²² 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.

²³ 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.

²⁴ 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.

²⁵ 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.

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.

Inhibition of the drug efflux activity of Ptch1 to overcome chemotherapy resistance

Responsable scientifique : Stéphane Azoulay

Laboratoire : ICN

Description :

Guedu E.^{(1)*}, Azoulay S.⁽¹⁾, Mus-Veteau I.⁽²⁾

(1) Université Côte d'Azur, ICN UMR 7272, CNRS, Nice, France

(2) Université Côte d'Azur, IPMC, CNRS, Valbonne, France

Notre équipe a identifié le récepteur Hedgehog Patched comme une pompe d'efflux de médicaments qui participe à la résistance des cellules cancéreuses à la chimiothérapie [1]. Grâce à un programme de criblage, l'hydroquinone Panicein A (PAH) [2] a été identifié comme un inhibiteur de l'activité d'efflux de médicaments de Patched. La synthèse de la PAH nous a permis de confirmer qu'elle augmente l'effet cytotoxique de plusieurs agents chimiothérapeutiques sur des lignées cellulaires de mélanome *in vitro* et *in vivo*[3].

Une structure de patched en présence de cholestérol réalisé au cryo-EM a permis de mettre en évidence le site de fixation du mécanisme d'efflux de cholestérol. Cependant il n'existe pas d'information sur la structure de patched en présence d'un agent chimiothérapeutique ou d'un inhibiteur connu.

Différentes études *in silico* [4] ont donc été mené afin de déterminer la poche réactionnelle, puis une étude de la conformation des ligands a permis de mettre en évidence la conformation la plus favorisé énergétiquement. Il apparait donc que les composés hydroquinones actives possèdent une conformation similaire dite ouverte, dans laquelle les cycles aromatiques sont éloignés l'un de l'autre. Tandis que les structures quinones inactives possèdent une conformation fermée où les cycles aromatiques sont proches et présentent des interactions π-π.

Récemment, de nouvelles structures originales dérivées de la PAH ont été identifiées pour améliorer les propriétés physico-chimiques ainsi que l'activité de la molécule .

Des tests de log D, de stabilité plasmatique et de stabilité métabolique ont été réalisé sur trois composés, et des tests *in vivo* ont démarré sur la molécule regroupant les meilleures caractéristiques.

Dans l'optique de compléter l'étude de relation structure-activité, nous souhaitons réaliser une étude de la conformation la plus favorisé énergétiquement afin de déterminer si les meilleures molécules de cette nouvelle série présentent elles aussi une conformation ouverte.

- [1] M. Bidet, A. Tomico, P. Martin, H. Guizouarn, P. Mollat and I. Mus-Veteau, Mol. Cancer Res., 2012, 10, 1496–1508
- [2] S. Kovachka, G. Mallochi, S. Azoulay, I. Mus-Veteau, et al. European Journal of Medicinal Chemistry 236, 2022, 114306
- [3] L. Signetti, S. Azoulay, I. Mus-Veteau, et al. Cancers., 2020, 12, 1500
- [4] S. Kovachka, G. Mallochi, A. Vargiu, S. Azoulay, I. Mus-Veteau, et al. Phys. Chem. Chem. Phys., 2021, 23, 8013-8022

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 :

Un calcul sur le benchmark C₂H₄O utilisé internationalement, qui reprenait toutes les configurations de notre meilleur calcul à ce jour (qui je le rappelle n'est qu'à 1.10⁻³ cm⁻¹ du record mondial) plus ~5000 configurations supplémentaires a tourné 8j et 19h sur le noeud smp.

L'énergie obtenue est meilleure, mais seulement de 1.10⁻⁶ cm⁻¹. Autrement dit les ~5000 configurations ajoutées ne servent à rien. Mais chaque échec nous permet d'affiner notre stratégie, et nous pensons battre le record mondial de l'énergie de l'état fondamental la plus basse très prochainement.

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, C₆₀, 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 C₆₀ 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 C₆₀ 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-Chenaï, 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-Chenaï, 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).

pH-Dependent Adsorption of poly(Acrylic Acid) onto Amorphous TiO₂ 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 (TiO₂-NPs) have promising perspectives in a broad range of applications, such as nuclear science, the development of advanced reactors, the pigment industry, and nanomedicine. TiO₂-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 TiO₂-NPs with a stabilization agent and studying chemical behavior is very important.

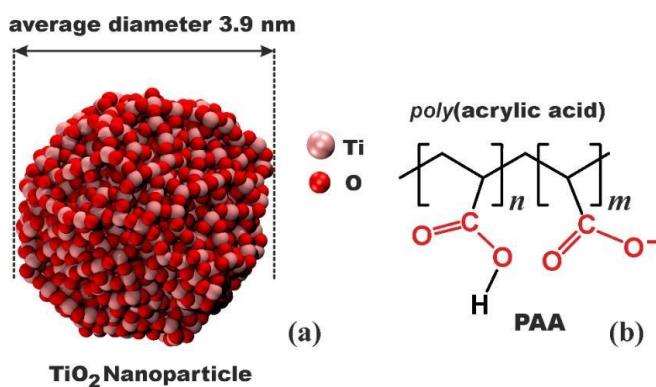


Figure 1. (a) Amorphous TiO₂ nanoparticle with the average diameter of 3.9 nm. (b) Molecular structure of poly(acrylic acid) (PAA) monomers in a protonated neutral (n) and dissociated charged (m) forms, respectively.

2. Project description

The project is devoted to studying the pH-dependent adsorption of polyacrylic acid onto amorphous TiO₂ 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 α 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 TiO₂-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 TiO₂ nanoparticles. All MD simulations and analyses were carried out using the GROMACS set of programs, version 2021.4.

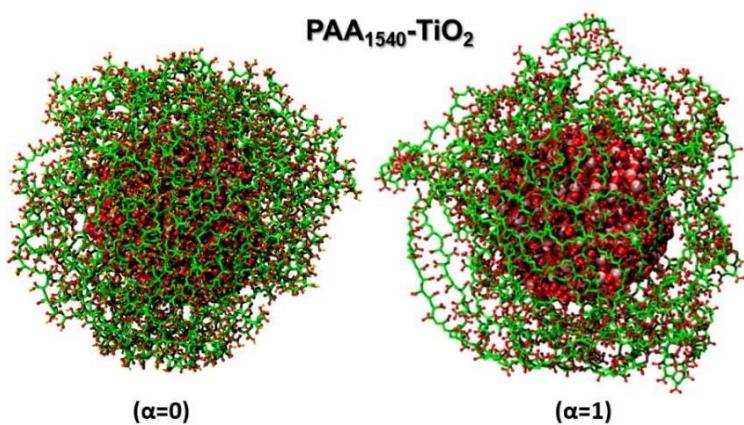


Figure 2. MD simulations of PAA1540 - TiO₂ nanoparticles at the two degrees of PAA ionization $\alpha = 0$ and 1, respectively.

Our MD simulations revealed that polymer adsorption onto amorphous TiO₂ 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 TiO₂ nanoparticles opens up the opportunity for their use as “smart” TiO₂-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 :

Accepted for publication:

Dikarieva K., Koverga V., Kalugin O. Local structure and Li-ion transport mechanism in LiFSI/DME/BTFE electrolyte revealed by molecular dynamics simulation. Kharkov University Bulletin; Chemical Series, No 45(68) 2025.

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 LiPF₆ in dimethyl carbonate / ethylene carbonate mixture (1:1), (ii) the solutions of SBPBF₄ in acetonitrile, and (iii) the mixtures of room-temperature ionic liquids (ILs) C₄mimX (X= BF₄⁻, PF₆⁻, TFO⁻, TFSI⁻) with dipolar aprotic solvents such as acetonitrile, γ-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.

Adsorption Modes of Main Protease of Coronavirus SARS-COV-2 onto Gold Nanoparticles of Different Shapes and Sizes

Responsable scientifique : Oleg N. Kalugin, Alexander Kyrychenko

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

Résultats obtenus grâce aux moyens informatiques attribués :

Kyrychenko A., Prud M. V., Lohachova K. O., Prezhdo O. V., Kalugin O. N. Native folding stability of the main protease of coronavirus SARS-CoV-2 on a gold surface. *J. Phys. Chem. Lett.* 2025, 43 (16), 11222-11229. doi: 10.1021/acs.jpclett.5c02743

Description :

Introduction

The outbreak of the novel coronavirus disease, COVID-19, caused by the new coronavirus 2019-nCoV, assigned as Severe Acute Respiratory Syndrome-related coronavirus (SARS-CoV-2), represents a pandemic threat to worldwide public health. Extraordinary efforts have been undertaken to suggest the treatment of SARS-CoV-2 infection, spanning various approaches ranging from traditional medicinal chemistry, and drug repurposing, up to numerous computational methods.

The main peptidase (M^{pro}) from the coronavirus SARS-CoV-2 is one of the most promising molecular targets for the development of anti-SARS agents. Recently, gold nanoparticles (AuNPs) have been reported as a promising alternative to available strategies based on traditional medicinal chemistry and drug repurposing.¹ It has recently been reported that inorganic nanoparticle (NPs) can interact strongly with viruses due to morphological similarities and size domain of SARS-CoV-2.¹⁻³ M^{pro} binding to gold nanoparticles leads to the formation of the protein corona, which, however, may alter the protein structure and enzymatic activity. Despite biological importance of this phenomena, our understanding of the molecular aspects M^{pro} -gold interactions remains incomplete.

Project description

The project is devoted to studying the binding interactions of the main protease (M^{pro}) of virus SARS-CoV-2 with bare gold nanoparticles of different sizes and shapes by all-atom molecular dynamics (MD) simulations ([Figures 1-2](#)). Crystal structures of M^{pro} (PDB code: 6LU7)⁴ was obtained from the Protein Data Bank. Protonation states for all titratable

residues were set at their default state at pH = 7. The ribbon representation of M^{pro} and its catalytic dyad residues His41 and Cys145 are shown as stick models in Figure 1. The CHARMM27 force field⁵ and TIP3P model⁶ were used to treat the protein and explicit water molecules, respectively.

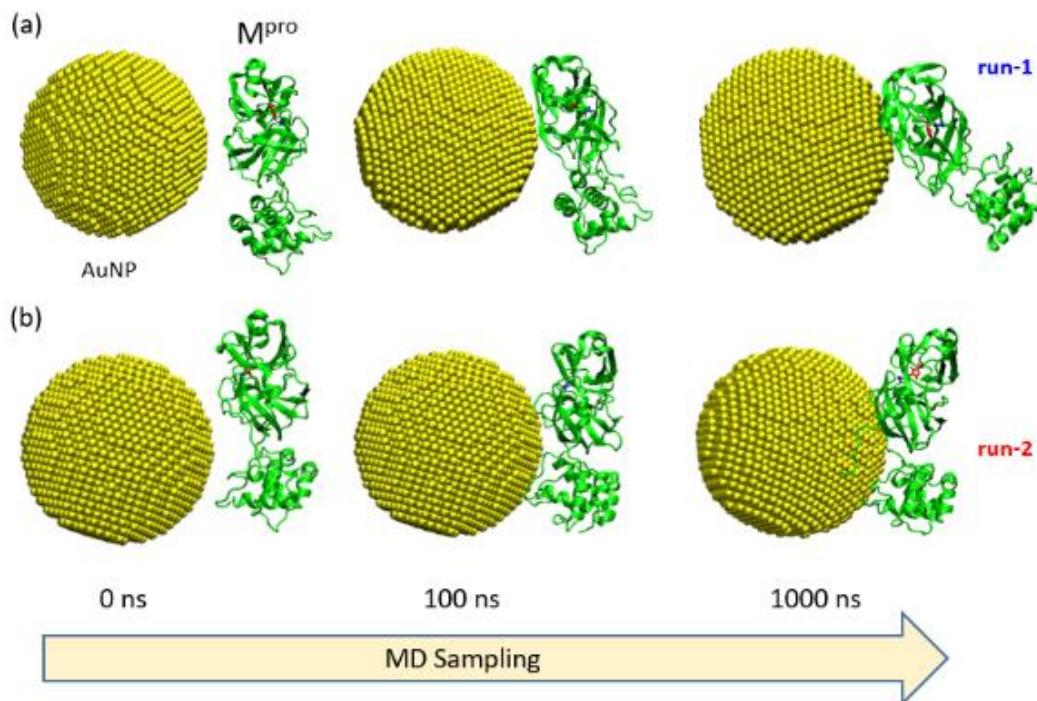


Figure 1. The simulation results of the interaction of M^{pro} with the gold nanoparticle (AuNP₆₆₃₂). Typical snapshots of the initial and adsorbed structures of M^{pro} onto AuNP for two independent MD runs, referred to as run-1/run-2, respectively. The protein is shown in green. Water molecules and ions are omitted for clarity.

To model AuNPs, the repulsion and dispersion terms of nonbonded interactions between Au atoms were computed by using the Lennard-Jones (LJ) 12-6 pairwise potential energy function with $\sigma=0.2629$ nm and $\epsilon=22.13$ kJ/mol, respectively. The performance of these 12-6 LJ pair potentials has recently been validated by modeling the adsorption behavior of small ligand and peptides onto gold nanoparticles.⁷ The Lorentz-Berthelot combination rules were used to calculate mixed nonbonded interactions between the protein, gold, and water atoms. All MD simulations and analysis were carried out by using the GROMACS set of programs, version 2021.4.

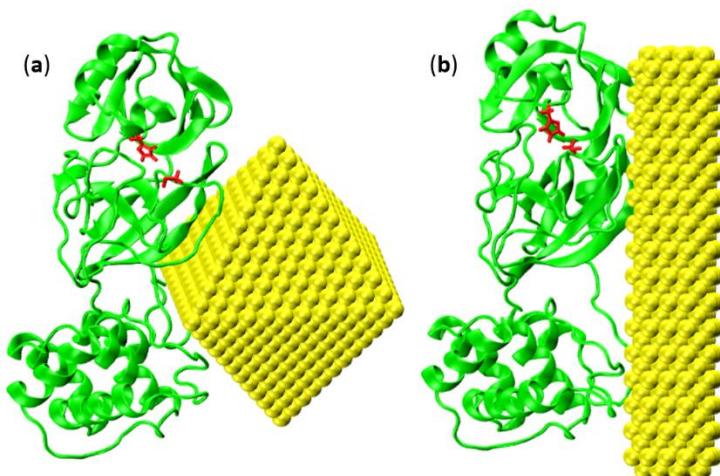


Figure 2. Structure of M^{pro} (PDB 6LU7) adsorbed onto AuNPs of different shapes: (a) bipyramidal (AuNB₁₁₅₆) and (b) rod (AuNR₆₀₀)

Preliminary results of this project suggest that M^{pro} reveals a high binding affinity towards all studied AuNPs ([Figures 1-2](#)). Importantly, inhibitory effects of AuNPs were observed as local protein refolding and the loss of the native spatial structure of its active site residues His41 and Cys145 accompanied the M^{pro} binding onto small sharp-edged AuNPs ([Figure 2](#)). Our findings point out that the local misfolding of M^{pro} on an AuNP surface may lead to partial or complete loss of its catalytic function, inhibiting the replication of the virus SARS-CoV-2

References:

1. Medhi, R.; Srinoi, P.; Ngo, N.; Tran, H.-V.; Lee, T. R. Nanoparticle-Based Strategies to Combat Covid-19. *ACS Applied Nano Materials* 2020, 3(9), 8557-8580.
2. Chen, L.; Liang, J. An Overview of Functional Nanoparticles as Novel Emerging Antiviral Therapeutic Agents. *Materials Science and Engineering: C* 2020, 112110924.
3. Mallakpour, S.; Azadi, E.; Hussain, C. M. The Latest Strategies in the Fight against the Covid-19 Pandemic: The Role of Metal and Metal Oxide Nanoparticles. *New Journal of Chemistry* 2021, 45(14), 6167-6179.
4. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from Sars-Cov-2 and Discovery of Its Inhibitors. *Nature* 2020, 582(7811), 289-293.
5. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell Jr, A. D. Charmm General Force Field: A

-
- Force Field for Drug-Like Molecules Compatible with the Charmm All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* 2010, 31(4), 671-690.
6. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* 1983, 79(2), 926-935.
7. Slavgorodska, M. V.; Gurova, Y. O.; Kyrychenko, A. Γ -Cyclodextrin as a Capping Agent for Gold Nanoparticles. *Computational and Theoretical Chemistry* 2021, 1194, 113060.

Development of MD simulation model for DRAK1 kinase

Responsable scientifique : Cyril Ronco

Laboratoire : ICN

Description :

Serine/threonine kinase 17A, also known as DRAK1 (death-associated protein kinase-related apoptosis-inducing protein kinase 1), is a member of the death-associated protein kinase (DAPK) family and is part of the "dark kinase." The current understanding of DRAK1's cellular functions and its role in disease processes is quite limited. Nonetheless, recent studies have linked DRAK1 to the development of several cancers and other pathogenic phenomena such as ROS production and inflammation. However, very few selective inhibitors targeting DRAK1 have yet been developed. Our group has recently identified a new family of compounds presenting nanomolar inhibition of DRAK1. These compounds are usually available as two regioisomers, one giving the activity, the other inactive. We have synthesized and evaluated 60+ analogues of two different models: 1/ a kinase binding assay on isolated recombinant DRAK1, 2/ antiproliferative cancer cell assay (XTT). These tests allowed to decipher structure-activity relationships and identify a lead compound. Docking experiments using Drak1 (PDB: 7QUF) have been performed for a number of analogues; however, the binding poses and energies obtained are very similar between the active and the inactive regioisomers. Therefore, simple docking experiments are not sufficient to explain the difference in activity observed in the biological tests.

The main goal of this master project is to develop a more precise model that can explain the regioselective attachment of the created lead compounds in inhibiting DRAK1. Several physicochemical parameters will be examined, such as the influence of the water molecule network and the changes in the total charge due to protonation/deprotonation of the ligand molecule within the enzyme's active site. The project will also include all-atom molecular dynamic simulations in a water solution to explore the potential to discover new zones within the active site, which may not be accessible with the fixed-receptor docking model.

Practically, the MS project of Vira Morozova will include the following successive tasks:

- Build and validate a proper MD force field based on the X-ray structure of DRAK1 kinase (PDB: 7QUF) [1].
- Validate the MD setup by reproducing the structure of the ligand-receptor complex for

the reference compound (CK-156) [2].

- Apply rigid and flexible molecular docking to generate initial sets of ligand-receptor structures for further MD evaluations.

- Use the developed MD model to understand the molecular-level mechanism of regioselective specificity, in which one regiosomer is active and the other inactive.

[1] <https://www.rcsb.org/structure/7QUF> (The STK17A (DRAK1) Kinase Domain Bound to CK156).

[2] Kurz C. G., Preuss F., Tjaden A., Cusack M., Amrhein J. A., Chatterjee D., Mathea S., Berger L. M., Berger B.-T., Krämer A., Weller M., Weiss T., Müller S., Knapp S., Hanke T. Illuminating the dark: Highly selective inhibition of serine/threonine kinase 17a with pyrazolo[1,5-a]pyrimidine-based macrocycles. *J. Med. Chem.* 2022, 65, 7799-7817.

Synthesis and characterization of anisotropic polymers, based on natural chromophores

Responsable scientifique : Alice Mija

Laboratoire : ICN

Description :

This PhD is in co-direction between Kharkiv University, Ukraine (supervised by prof. Oleksandr Roshal) and Institute de Chimie de Nice of University Cote d'Azur (with co-supervision of prof. Alice Mija). It is targeted to disclosing the potential of some natural flavonoids and their synthetic analogous to be applied as Non-Linear Optical polymer materials.

Non-linear optical materials are widely applied in electronics, namely they are a key component of the optical signal modulator, a central element of the optoelectronic communication systems.

Generally used inorganic NLO crystals, despite huge advantage in chemical and thermal stability over organic NLO materials, lose them in performance. Yet, for organic NLO materials their thermal and especially directional stability, which is crucial for this effect, are not finally resolved. Thus, we are concentrating on enhancing performance of sustainable organic NLO materials based on flavonols.

The topic of the PhD is combining ideas by prof. Roshal, who previously studied aligned in electric field polymer films fabricated of epoxidized flavonol (quercetin) and a skill of the Cote d'Azur University chemists who intensively study bio-based epoxy polymers including flavonols cores, quercetin etc.

Although NLO performance of the film made of bare quercetin was found not to be very high, they effectively immobilized it inside the polymer structure making it a part of polymer molecule. My propositions were to enhance push-pull phenomena in flavonol monomers by chemical modification, similarly to some previously studied, but non-polymerizable flavones or similar compounds. Thus, we hope to obtain more effective NLO-chromophores themselves, that can additionally be reliably included into a polymer chain and then poled in electric field.

To justify my considerations on the ways of chemical modifications and their effect on the electronic and polarizability properties of the molecules, and to have initial data to

compare then with the measurements of those properties, we need to perform DFT calculations for each of those molecular structures.

At the beginning, we started our DFT calculations on the cluster of Ukrainian-American Laboratory of Computational Chemistry (UALCC, Kharkiv, Ukraine). Unfortunately, due to the Russian-Ukrainian war and destruction of major electrical power plants in Kharkiv in January 2023, this cluster is not available since.

Therefore, now we are looking for a cluster to finalize our calculations in France, to complete the calculated data for our molecular array.

Physique non linéaire, fluides complexes

EUR Spectrum

Performance of a high-order Kokkos accelerated fluid solver

Responsable scientifique : Holger Homann

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Publication acceptée dans Computer Physics Communications :

Filipp Sporykhin, Holger Homann

Performance of a high-order MPI-Kokkos accelerated fluid solver

Description du projet :

Hydrodynamical turbulence is a phenomenon that affects many domains of the physical world, from the Earth's atmosphere to the interactions between solar plasma winds and bodies in the solar system. Due to the instability of solutions for the NavierStokes equations, turbulence has become a widely studied problem, both empirically, and numerically. When it comes to numerical studies, due to the cascading nature of turbulence, very high numerical resolutions are required in order to simulate the smaller scales. High numerical resolutions imply very long computation times, even with high order and efficient schemes. Rather than using processors with higher clock speeds, today's solution is to parallelise the computation over multiple devices. Our research project consists in parallelising a 6th order Discontinuous Galerkin-Runge Kutta hydrodynamical solver and comparing the performance on varying architectures. This is achieved with the help of Kokkos, a C++ abstraction layer in the form of a template library that is used to compile code to architecture specific backends such as CUDA or OpenMP. The main goal is exploring the capabilities of Kokkos' vectorisation and comparing performance between GPU and CPU code execution with respect to the dimensionality of the hydrodynamical problem. The problems covered are 2D advection (4 variables), isothermal gas (5 variables), and viscous isothermal gas (7 variables).

The lack of concrete analytical and experimental results in literature begs us to investigate the maximum speed-up factor between multicore CPU and GPU execution. In the case of finding significant results, the cost of GPU implementation is to be studied for both price and power usage. If the aforementioned are relatively low in comparison to traditional

CPU computation, an important message in code optimisation can be delivered to the scientific community

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 :

Articles publiés (2024-2025)

- Michel Orsi, Laurent Lobry, Elisabeth Lemaire, and François Peters. Mass and momentum balance during particle migration in the pressure-driven flow of frictional non-brownian suspensions. *Journal of Fluid Mechanics*, 998 :A16, 2024.
- Michel Orsi, Laurent Lobry, Elisabeth Lemaire, François Peters; The effect of coupled friction and adhesion on the rheology of non-Brownian dense suspensions. *J. Rheol.* 1 July 2025; 69 (4): 515–539

Communications orales 2024-2025

- Michel Orsi, Elisabeth Lemaire, Laurent Lobry, and François Peters. Detailed mass and momentum balances during particle migration in the pressure-driven channel flow of frictional non-brownian suspensions. In 12th International Conference on Multiphase flow, Toulouse, France, May 2025.
- Michel Orsi, Elisabeth Lemaire, Laurent Lobry, and François Peters. Rheology of non-brownian suspensions of adhesive and frictional particles via particle-resolved numerical simulations. In AERC 2025 - The Annual European Rheology Conference., Lyon, France, 2025.

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.²⁶ ²⁷ ²⁸ ²⁹

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

²⁶ 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.

²⁷ 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.

²⁸ 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.

²⁹ 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.

méthode des domaines fictifs à cet environnement, avec en particulier la parallélisation 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 années³⁰, 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

³⁰ 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.

grandeur locales au sein de la suspension (fraction volumique locale, taux de 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.³¹ ³²

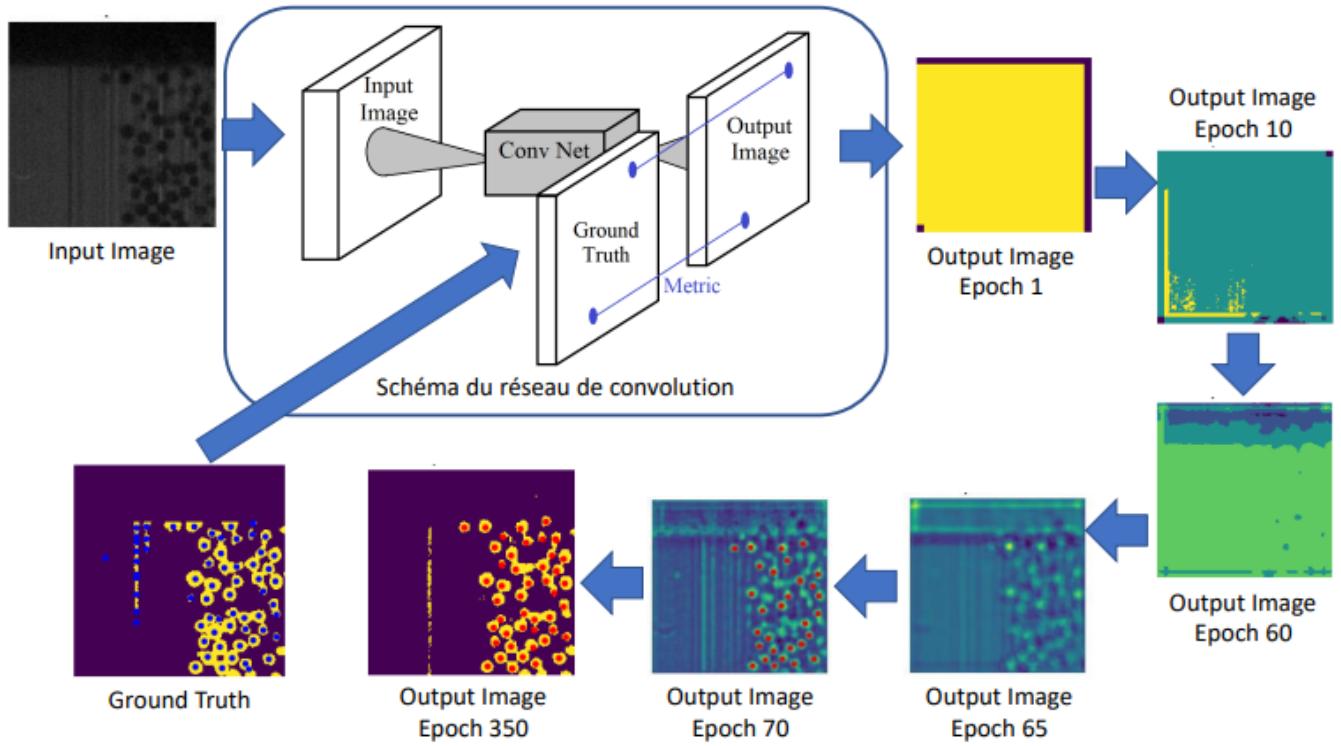
Ceci se fait au prix d'expériences complexes et d'un traitement de données assez lourd. En effet les suspensions utilisées doivent être rendues transparentes en adaptant l'indice des particules à celui du fluide, qui est lui-même rendu visible par l'adjonction d'un colorant fluorescent activé par une nappe laser, dont l'image est enregistré par une caméra. L'épaisseur de la nappe laser étant plus fine que la 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 convolutif³³ (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

³¹ 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.

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

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

réseau sont subtilement modifiés par un processus de « descente de gradient ». Ce 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.

Microstructure 3D de suspensions de particules molles - Application au développement d'un appareil médical de détection de la drépanocytose

Responsable scientifique : Frédéric Blanc

Laboratoire : INPHYNI

Description du projet :

Grâce à un dispositif expérimental que nous avons développé à l’Institut de Physique de Nice, nous produisons des images tridimensionnelles de suspensions concentrées de sphères molles. Dans ces images, les sphères apparaissent comme des objets lumineux en suspension dans un liquide sombre. À l’aide d’un algorithme de segmentation automatique basé sur des réseaux de neurones convolutifs (CNN), nous détectons avec précision la forme et la position de chaque particule, ce qui nous permet de caractériser finement la microstructure tridimensionnelle de la suspension.

Cette microstructure sert ensuite à calibrer une expérience de rétrodiffusion par ondes ultra sonores, menée par une collègue du LOMA à Marseille, dans une suspension identique de sphères molles. L’objectif est de comparer les signaux de rétrodiffusion obtenus avec ceux issus de suspensions de sphères rigides, pour lesquelles nous disposons déjà de données de référence.

Cette étude constitue la première étape vers le développement d’un dispositif médical innovant permettant d’accéder à la microstructure du sang par rétrodiffusion d’ondes acoustiques. À terme, ce dispositif pourrait offrir une mesure directe de la rigidité des globules rouges, ouvrant ainsi la voie à une méthode de détection précoce des crises de drépanocytose, pathologie dans laquelle l’élasticité des globules rouges est fortement altérée.

D’un point de vue technique, l’algorithme de détection automatique repose sur un apprentissage supervisé à partir d’un jeu d’images annotées, dans lequel les particules ont été identifiées manuellement – ces annotations servent de « vérité terrain ». Ce processus d’entraînement, essentiel pour optimiser les performances du réseau neuronal, est particulièrement coûteux en temps de calcul. Le recours à un cluster de GPU tel que

Azzurra constitue alors un atout majeur pour accélérer cette phase et permettre un apprentissage efficace.

Ce projet est financé par l'ANR HEMO.

Wave turbulence simulations and theory

Responsable scientifique : Sergey Nazarenko

Laboratoire : INPHYNI

Résultats obtenus grâce aux moyens informatiques attribués :

- Rapport de stage de Jorge Carreno Donoso "Dynamics and statistics of quantized vortices in a disk". Master OAM: Fundamental physics and applications (2025).
- Strong and weak wave turbulence regimes in Bose-Einstein condensates Ying Zhu, Giorgio Krstulovic and Sergey Nazarenko <https://arxiv.org/pdf/2411.19812.pdf>

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 :

Labarre, V., Krstulovic, G., & Nazarenko, S. (2025). Wave-kinetic dynamics of forced-dissipated turbulent internal gravity waves. *Physical Review Letters*, 135(1), 014101.

Description du projet en pages suivantes.

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 atteint 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 par 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 utilisées 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 finies 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 poloidale, 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 complémer 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 ρ 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 ρ' 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, ν est la viscosité cinématique, et κ 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

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

tandis que leur vitesse de groupe est

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

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 inertuelle 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 **Fluiddyn** [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 sauvegardés au format hdf5. La librairie hdf5/1.12.0-mpi est utilisée pour gérer l'enregistrement parallélisé des données de sortie.

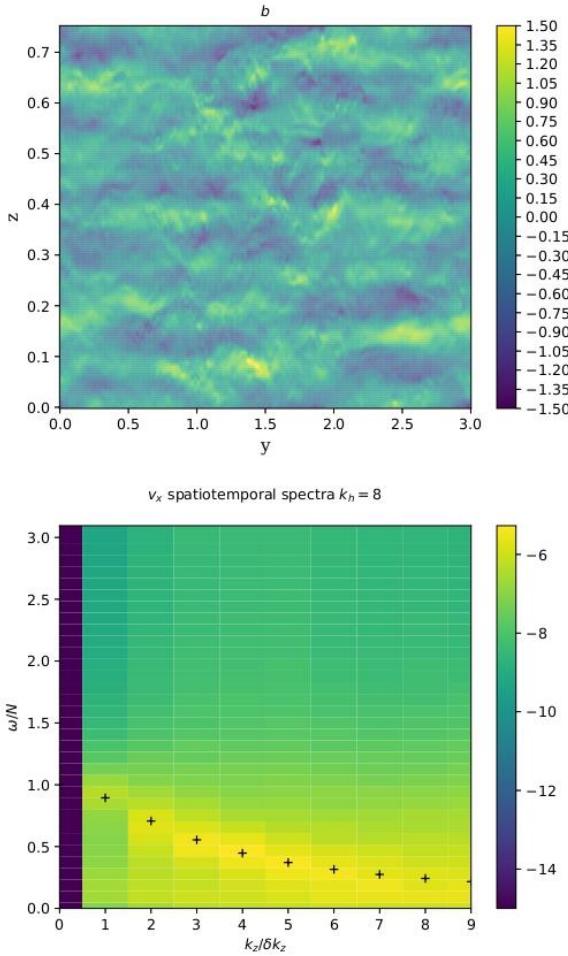


FIGURE 1 – Coupe verticale du champ de flottabilité (haut) et spectre spatiotemporel $k_h - \omega$ (bas) pour une simulation avec $N = 40$, $R_b = 40$ et la projection poloidale réalisée sur Jean-Zay. Les croix représentent la relation de dispersion théorique pour les ondes de gravité.

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'enregistrer 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.

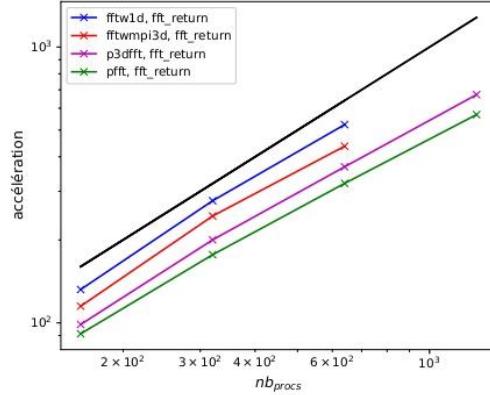


FIGURE 2 – Facteur d'accélération en fonction du nombre de processeurs pour réaliser les transformées de Fourier rapides par Fluidfft . Le test a été réalisé sur la machine Jean-Zay pour une grille de $1280 \times 1280 \times 640$ points et 4 librairies : fftw1d, fftw3d, pfft, et p3dfft.

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 (pfft et p3dfft) sont plus lentes à nombre de processeurs égal, mais permettent d'utiliser plus de processeurs. La librairie p3dfft est plus rapide que pfft, 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.

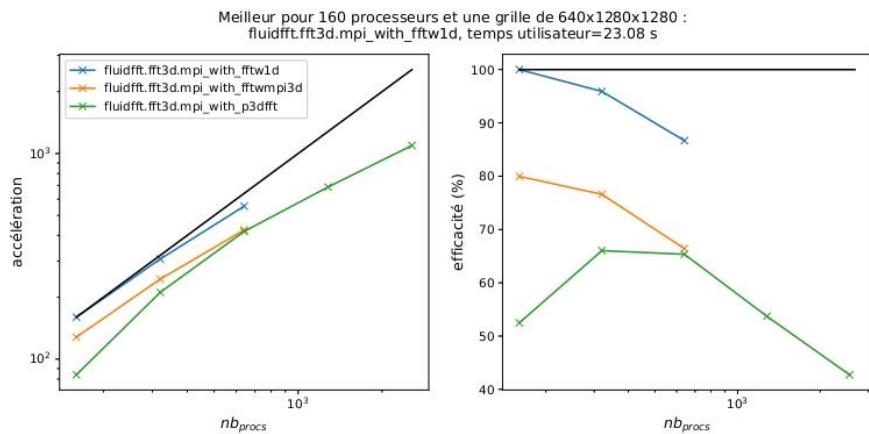


FIGURE 3 – Propriétés de parallélisation du solveur ns3d.strat de Fluidsim. Le test a été réalisé sur la machine Jean-Zay pour une grille de $1280 \times 1280 \times 640$ points et 100 pas de temps. À gauche : accélération relative du temps d'exécution t par rapport au cas de référence (1 noeud = 40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

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.
- [Mohanan et al., 2019a] Mohanan, 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.
- [Mohanan et al., 2019b] Mohanan, 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.

Wave turbulence in geophysics: internal, inertial and water surface waves

Responsable scientifique : Sergey Nazarenko

Laboratoire : INPHYNI

Description :

This project aims at understanding the propagation and interactions of weakly nonlinear waves in geophysical media. The main question to understand is how well the non-linear wave systems are described by the respective wave kinetic equations (WKEs). The secondary, but equally important questions are how to correct WKE to adopt it to cases when the idealised theory fails, and how to predict transport processes caused by wave turbulence. The scientific problem will be addressed using the wave turbulence theory (in both idealised and non-idealised settings) and studying solutions of the associated wave kinetic equations and comparing them with corresponding predictions of Direct Numerical Simulations(DNSs). To complement theoretical predictions, numerical simulations of the wave kinetic equation and the original dynamical equation describing the whole physics (DNS). Based on the data arising from numerics, a study will be undertaken on how to improve the WKE description in non-idealised (realistic) situations and how to predict the transport properties.

Spontaneous stochasticity in surface quasi-geostrophic turbulence

Responsable scientifique : Jérémie Bec

Laboratoire : Inria

Résultats obtenus grâce aux moyens informatiques attribués :

- Nicolas Valade, Jeremie Bec, Simon Thalabard - Surface quasigeostrophic turbulence: The refined study of an active scalar - soumis à Journal of Fluid Mechanics, 2025.
- Thèse de doctorat de Nicolas Valade. Turbulence quasi-géostrophique de surface. 2025, Université Côte d'Azur. <https://theses.fr/s345052>

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

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 notamment 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 :

Preprint: Dynamics of transport by helical edge states

Luis Alberto Razo López, Pierre Wulles, Geoffroy J. Aubry, Sergey E. Skipetrov, Fabrice Mortessagne

<https://arxiv.org/abs/2502.20918>

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>

Hydraulique urbaine

EUR Spectrum

Modélisations thermomécaniques de l'état de contraintes au sein du Bassin parisien

Responsables scientifiques : Carole Petit

Laboratoire : Géoazur

Description en pages suivantes.

Modélisations thermomécaniques de l'état de contraintes au sein du Bassin parisien

Les couches géologiques du bassin de Paris sont étudiées par l'ANDRA pour le stockage potentiel de déchets nucléaires. Le projet Cigéo (site de Bure) se localise en partie orientale du Bassin parisien à l'aplomb d'une pile sédimentaire mésozoïque (252 à 65 millions d'années) de quelque 2000 m d'épaisseur reposant sur le socle continental plus ancien (hercynien). Cette région est caractérisée par l'absence de sismicité (Figure 1) ainsi que de tout indice de déformation récente sur les failles majeures et ce plus probablement depuis les 20 derniers millions d'années. Par opposition, les massifs des Vosges, du Massif Central et du Massif Armoricain voisins montrent une sismicité non négligeable, même si de faible magnitude. On suspecte un effet de cohésion au niveau les bassins sédimentaires augmentant localement la résistance à la fracture (Petit et al., 2019), mais il n'est pas possible d'extrapoler ces effets sur le long-terme. Plus localement, la couche hôte des argilites du Callovo-Oxfordien (Jurassique moyen-supérieur) devant accueillir le stockage à 500 m de profondeur est caractérisée par une surpression et un tenseur des contraintes anisotrope.

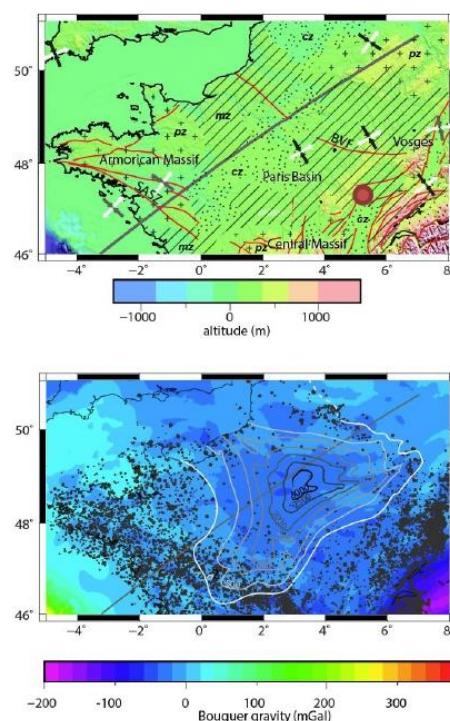


Figure 1. Haut : topographie, géologie et champ de contraintes du Bassin de Paris (le site Cigéo est indiqué par le rond rouge. Les failles sont indiquées par des traits rouges. Bas : anomalie gravimétrique, épaisseurs sédimentaires (courbes) et sismicité (points noirs). Le transect en gris est l'objet de l'étude publiée en 2019 (Petit et al., 2019).

La problématique est de comprendre comment des déformations, même très lentes à l'échelle des observations humaines (Masson et al., 2019, Figure 2) pourraient affecter sur le long terme l'état de la croûte et de la pile sédimentaire au niveau du bassin de Paris et plus précisément du site de Bure.

Afin d'apporter un éclairage sur les variations spatiales de l'état de contraintes de la croûte, l'ANDRA a défini comme sujet prioritaire (<https://www.andra.fr/sites/default/files/2024-02/AAP%20Th%C3%A8ses%20Andra%202024%20-%20descriptif%20des%20th%C3%A8mes.pdf>, thème 2) pour une bourse de thèse la modélisation des processus physiques à différentes échelles d'intérêt : de la lithosphère à la pile sédimentaire, de l'échelle locale à l'échelle régionale, des temps courts (sismicité) aux temps longs (million d'années). Pour ce faire, une approche multi-échelle sera mise en œuvre reposant sur des modélisations

thermomécaniques. Au travers des études de sensibilité, l'étudiant devra caractériser l'influence (1) de la rhéologie, de la cohésion et du couplage mécanique au sein de la pile sédimentaire, (2) des variations spatiales du flux de chaleur, et (3) des paramètres contrôlant le flambage lithosphérique et crustal sur le comportement mécanique de la lithosphère et – en particulier- de la croûte supérieure. Les conditions aux limites des modèles feront le lien entre la compression alpine, les contraintes venant de la marge Atlantique et la déformation du bassin.

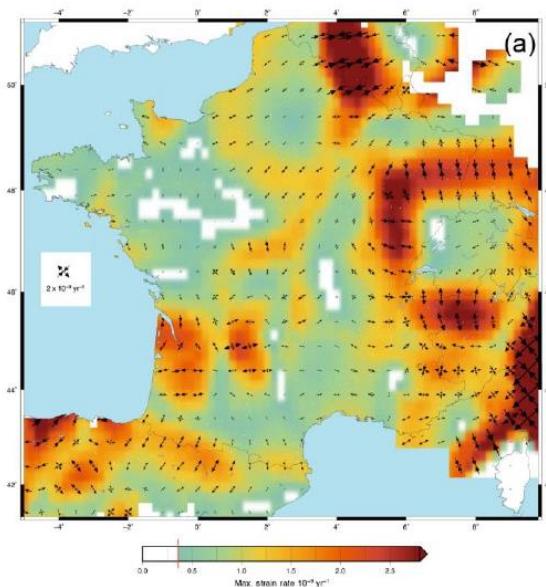


Figure 2. Carte des taux et du régime de déformation en France métropolitaine (Masson et al., 2019).

La présente demande concerne une étude préliminaire qui sera menée par un étudiant de M2 (Nicolas Lapicque, Master 2 « 3G ») encadré par moi-même (Carole Petit) dans le cadre de son stage de recherche.

Après avoir intégré les données disponibles dans la littérature, nous allons réaliser une série de modélisations thermo-mécaniques 3D à l'échelle du Bassin de Paris pour tester l'influence des conditions aux limites (vitesse des plaques tectoniques) et initiales (état thermique et rhéologie de la lithosphère) sur la déformation du bassin de Paris au cours des 1 à 10 millions d'années qui viennent.

Le modèle est basé sur le code de modélisation en différences finies LaMEM (<https://github.com/UniMainzGeo/LaMEM>, Kaus and Popov, 2014) qui permet de calculer la déformation de matériaux visco-élasto-plastiques avec des rhéologies visqueuses non-linéaires (caractéristiques des matériaux terrestres) et prenant en compte le couplage avec la température et les processus d'érosion-sédimentation.

Références :

Kaus, B., and Popov, A., LaMEM: a Massively Parallel Staggered-Grid Finite-Difference Code for Thermo-Mechanical Modeling of Lithospheric Deformation with Visco-Elasto-Plastic Rheologies, AGU Fall Meeting Abstracts, 2014.

Masson, C., Mazzotti, S., Vernant, P., and Doerflinger, E., Extracting small deformation beyond individual station precision from dense Global Navigation Satellite System (GNSS) networks in France and western Europe, *Solid Earth*, 10, 1905–1920, 2019.

Petit C, de Barros L, Duclaux G, Mazabraud Y. Why Are There No Earthquakes in the Intracratonic Paris Basin? Insights from Flexural Models. *Geosciences*. 2019; 9(12):502. <https://doi.org/10.3390/geosciences9120502>

High-resolution topography upscaling for overland flows

Responsables scientifiques : Konstantin Brenner

Laboratoire : LJAD

Description :

This application is made as a part of the ANR project Top-up: High-resolution topography upscaling for overland flows. The project develops numerical methods for the simulation of floods in urban environments, taking into account high-precision geometric data. The data is public and is provided by Metropole Cote d'Azur. The ANR project started in 2021 and will end in early 2025.

Thanks to the PhD of Miranda Boutilier funded by the project (defence planned on October 4th) we have been able to design a family of efficient numerical algorithms of Multi-scale and Domain decomposition type. Those methods allow for a significant acceleration in the solution of the algebraic nonlinear and linear systems resulting from the numerical discretisation of the flood models. Our current effort is focused on parallel implementation of the algorithms from the project's PhD. The implementation is done in python with mpi4py and rely on a research engineer contract founded by ANR. The bulk of the development has been done and we were able to perform the rough validate the code using the cluster of JAD laboratory. At this point we are in need of performing rigorous scalability tests for which the infrastructure of LJAD's cluster is not sufficient. The main issue is that we can not be reserved for a given application.

Modélisation de l'hydrologie d'un bassin versant avec une approche à base physique : le cas du bassin de la Brague

Responsables scientifiques : Olivier Delestre

Laboratoire : LJAD

Résultats obtenus grâce aux moyens informatiques attribués :

Maëline Baillet, Matteo Geraci, Tiberio Potenza, Théo Salinas, Mohammed Assaba, Florian Cordier, Pierfranco Costabile, Olivier Delestre, Margherita Lombardo and Roger Moussa. Novel methodology for strategic mesh refinement in two-dimensional (2D) rainfall-runoff simulations. SIMHYDRO 2025

Description :

Le bassin de la Brague est régulièrement impacté par des précipitations intenses donnant lieu à des crues voire des débordements au niveau de la partie aval de la Brague. Les événements les plus récents sont ceux de 2011, 2015 et 2019 qui ont généré beaucoup de dégâts et des pertes en vies. Il est nécessaire de réaliser des études hydrologiques et hydrauliques de ce bassin afin de comprendre son fonctionnement, de réaliser des cartes d'inondation et de réfléchir à des aménagements. De plus il y a une seule station hydrométrique au sein de ce bassin. La modélisation peut donc également aider à avoir une idée assez précise des débits transitant à différents endroits du bassin versant (sur la Brague et ses affluents).

Souvent un modèle type réservoir est utilisé afin de réaliser l'étude hydrologique d'un bassin versant, puis les hydrogrammes générés sont utilisés en entrée d'un modèle hydraulique afin de simuler l'emprise de l'inondation. L'objectif ici est de pratiquer une approche originale en utilisant un outil de modélisation hydraulique (TELEMAC-2D) afin de réaliser la simulation de la réponse hydrologique du bassin versant de la Brague pour différents événements (2011, 2015 et 2019). L'intérêt particulier porte sur la capacité à s'appuyer sur un maillage assez fin de la zone aval permettant d'obtenir des résultats de simulation donnant des emprises des zones inondées dans la partie aval les plus proches possibles des observations (il y a des données notamment pour les événements de 2011 et de 2015). Pour ce faire, il est nécessaire de solliciter des ressources de calcul performantes, qui nous poussent à solliciter des méthodes HPC. Lors de l'événement de 2015, les buses se trouvant sous l'autoroute ont subi des embâcles jouant le rôle de « barrage », il pourra

donc être envisagé de prendre en compte ces structures hydrauliques au mieux. Enfin, il y a des karsts répartis sur le domaine, différentes configurations seront envisagées afin de représenter au mieux ces singularités. Il s'agit d'un sujet qui permet de développer les compétences en modélisation hydrologique et hydraulique, utiles pour un futur ingénieur en bureau d'études. Le sujet est également exploratoire et fera l'objet d'une première valorisation dans le cadre de la conférence Simhydro2025 et un projet de valorisation au format article dans une revue scientifique à comité de relecture international.

Dynamique des fluides

EUR Spectrum

Dynamique de polymères en turbulence et turbulence élastique

Responsable scientifique : Dario Vincenzi

Laboratoire : LJAD

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

Description :

Earthquakes are a fundamental component of plate tectonics, one of the main processes that shape the surface of our planet, and one 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 :

Jourdon, A., May, D. A., Hayek, J. N., & Gabriel, A.-A. (2025). 3D reconstruction of complex fault systems from volumetric geodynamic shear zones using medial axis transform. *Geochemistry, Geophysics, Geosystems*, 26, e2025GC012169. <https://doi.org/10.1029/2025GC012169>

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 raccorçissement 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*³⁴ ; *Duclaux et al., 2019*³⁵) 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 décrochevauchantes sismogènes observées en surface dans les zones de convergence obliques.

³⁴ *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.*

³⁵ *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.*

Mathématiques, informatique

EUR DS4H, Spectrum

Modélisation des Interactions entre Croissance Tumorale et Réponse Immunitaire

Responsable scientifique : Thierry Goudon

Laboratoire : LJAD

Résultats obtenus grâce aux moyens informatiques attribués :

- A mixture-like model for tumor-immune system interactions. C. Tayou Fotso, S. Girel, F. Anjuere, V. M. Braud, T. Goudon, F. Hubert. J. Theor. Biol., 111738, 2024.
- Simulation of mixture flows by the Discrete Duality Finite Volume method, C. Tayou Fotso, T. Goudon, F. Hubert, submitted
- Tumor growth and immune response : on the impact of the space-structuration of the tumor microenvironment, C. Tayou Fotso, F. Anjuere, V. Braud, T. Goudon and F. Hubert, submitted.
- Modeling of the interactions between tumor growth and immune response, C. Tayou Fotso, PhD thesis, Univ. Côte d'Azur, july 2025, <https://hal.science/tel-05322358v1>

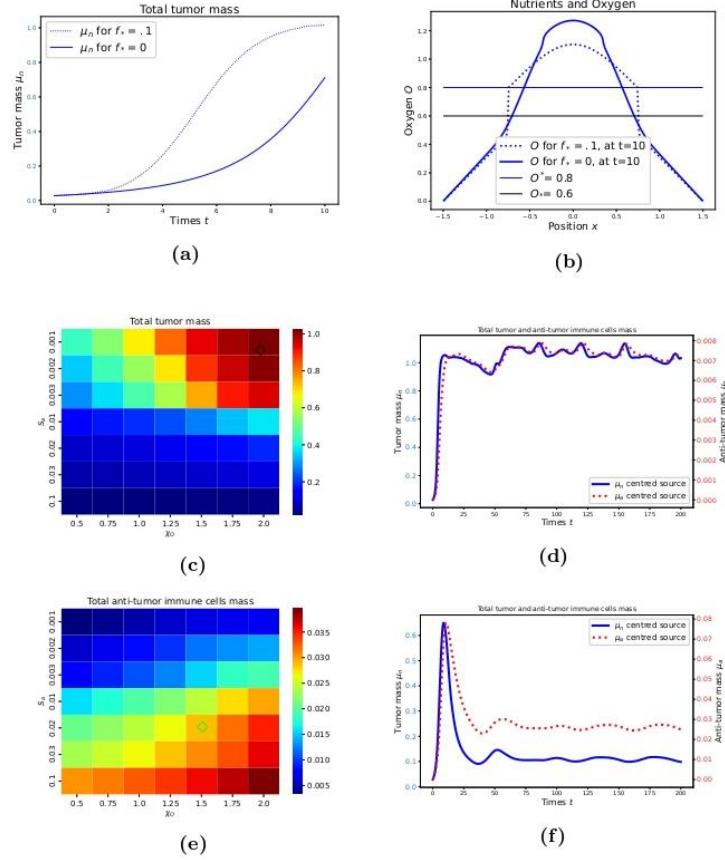
Description en pages suivantes.

Summary of research work:

*'Modélisation des Interactions entre
Croissance Tumorale et Réponse Immunitaire'*

This project is devoted to the mathematical model based on mixture theory intended to describe the tumor-immune system interactions within the tumor microenvironment. The equations account for the geometry of the tumor expansion, and the displacement of the immune cells, driven by diffusion and chemotactic mechanisms, which can have anti- and protumoral effects, and the environment. They also take into account the constraints in terms of nutrient and oxygen supply. The project is based on a programme of numerical simulations and comparison with biological data derived from an innovative mass cytometry imaging technique providing an exhaustive mapping of the tumor microenvironment of human cutaneous squamous cell carcinomas. Many mathematical models have been introduced so far with quite elaborate descriptions of the tumor-immune interactions based on systems of ordinary differential equations, see for instance [3]. However, the spatial repartition of the tumor and immune cells seems to play a crucial role [4] and cannot be easily taken into account by such ODE systems. For instance, this observation has motivated the work detailed in [1, 2]. Therefore, we adopt in [5] a different viewpoint and we are going to set up a continuous model based on partial differential equations, describing the tumor growth and the immune response by considering the cell densities, and unknowns that depend on both the time t and the position x . To this end, we shall describe the tumor and immune cell interactions by means of fluid mechanic principles, appealing to methods of mixture theory [7] and their application to biological flows [6].

Taking into account access to nutrients and oxygen, therefore the equations are established through the combination of mechanical and biological effects that govern the competition for space and resources between different species of cells [5]. Our aim is to contribute to our understanding of the spatial structuring of the tumour microenvironment and, ultimately, to predict recurrence and the effect of immunotherapy treatments, as a complement to conventional approaches using surgery or chemotherapy. We carried out a preliminary study in [5] that involved discussing the influence of modelling assumptions and attempting to prioritise the effects of biological and mechanical mechanisms through simulations, in order to assess whether or not the model is capable of reproducing the equilibrium and escape phases, and to shed some light on the critical phenomena



that shape the space organization of the tumor microenvironment. Our main initial conclusions are as follows. Firstly, oxygen/nutrient supply is essential in shaping tumour development (see Figures 1a and 1b). Secondly, the different versions of the model are capable of reproducing the "3E" phases ; in particular, equilibrium phases with a residual tumour remaining under the control of the immune system can be observed (see Figures 1e and 1f). Thirdly, taking account of the protumoral effects of the immune response radically alters the dynamics and could lead to unlimited tumour expansion (see Figures 1c and 1d). We then extended our geometry to 2D by developing a discrete duality finite volume method (DDFV) to numerically solve the designed model. The results are currently being studied in order to further our discussions based on clinical data obtained.

Références

- [1] L. Almeida, C. Audebert, E. Leschiera, and T. Lorenzi. A hybrid discretecontinuum modelling approach to explore the impact of T-cell infiltration on anti-tumour immune response. *Bull. Math. Biol.*, 84 :141, 2022.
- [2] K. Atsou, F. Anjuère, V. M. Braud, and T. Goudon. A size and space structured model describing interactions of tumor cells with immune cells reveals cancer persistent equilibrium states in tumorigenesis. *J. Theor. Biol.*, 490 :110163, 2020.
- [3] L. G. de Pillis, A. E. Radunskaya, and C. L. Wiseman. A validated mathematical model of cell-mediated immune response to tumor growth. *Cancer Res.*, 65(17) :7950–7958, 2005.
- [4] C. Draghi, L. Viger, F. Denis, and C. Letellier. How the growth rate of host cells affects cancer risk in a deterministic way. *Chaos*, 27 :093101, 2017.
- [5] Christian Tayou Fotso, Simon Girel, Fabienne Anjuère, Véronique M Braud, Florence Hubert, and Thierry Goudon. A mixture-like model for tumor-immune system interactions. *Journal of Theoretical Biology*, 581 :111738, 2024.
- [6] B. Polizzi, O. Bernard, and M. Ribot. A time-space model for the growth of microalgae biofilms for biofuel production. *J. Theor. Biol.*, 432 :55–79, 2017.
- [7] K. R. Rajagopal and L. Tao. *Mechanics of Mixtures*. World Scientific, 1995.

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 LMSADE (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.

Modèles de plongement de graphes de connaissances : injection et découverte de connaissances symboliques

Responsable scientifique : Catherine Faron

Laboratoire : I3S

Description :

Cette thèse vise à étudier les interactions entre modèles de plongement de graphes de connaissances (GCs) et les connaissances symboliques représentées au sein de ces graphes. Ces dernières peuvent être représentées de diverses manières : via des ontologies (OWL ou RDFS), des contraintes SHACL, des règles SWRL, etc. Une récente ligne de recherche s'intéresse à l'injection de ces connaissances au sein des modèles de plongement de graphes. Néanmoins, ces travaux concernent majoritairement les ontologies OWL ou RDFS. De plus, une perspective complémentaire et peu considérée réside dans la découverte de connaissances émergeant dans les espaces de plongement. L'enjeu de ces travaux de thèse est donc de combiner injection et découverte de connaissances symboliques au sein des modèles de plongement afin de créer une boucle vertueuse entre graphes de connaissances et modèles de plongement dans une perspective d'IA neuro-symbolique.

Nous proposons dans le cadre de cette thèse (1) d'étudier l'injection de différentes formes de connaissances dans les modèles de plongement et (2) d'étudier la fouille des espaces de plongement résultant pour découvrir de nouvelles connaissances. Nous considérerons les tâches d'amélioration de GCs (par ex., prédition de liens, classification de liens, alignement d'entités) et les modèles de plongement développés à cet effet.

- Injection de connaissances. Il sera nécessaire d'examiner les différentes structures symboliques injectées dans les modèles de plongement, les approches utilisées pour une telle injection, leurs avantages, et leurs inconvénients. Cette étude permettra de considérer d'autres types de connaissances symboliques actuellement non prises en compte, ou de proposer de nouvelles approches pour réaliser leur injection.

- Découverte de connaissances. Il faudra étudier les différents modèles et espaces de plongement disponibles afin de proposer des méthodes de fouille de ces espaces. Ces méthodes pourront être construites par symétrique avec les méthodes d'injection. Par exemple, une méthode injectant des concepts en les représentant comme des régions dans un espace et leur subsomption par l'inclusion des régions pourra mener au développement d'une approche de fouille basée sur la détection d'inclusion de régions. Ces méthodes pourront dans un premier temps être spécifiques à chaque espace, puis ensuite généralisables à une ou plusieurs familles d'espaces.

Il sera possible d'étudier d'autres tâches d'amélioration des GCs (par exemple, l'élagage, la détection de la dérive sémantique, la confrontation de connaissances) ou d'autres métadonnées à injecter (par exemple, l'incertitude, la validité temporelle).

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

Résultats obtenus grâce aux moyens informatiques attribués :

Celian Ringwald, Fabien Gandon, Catherine Faron, Franck Michel, Hanna Abi Akl. Kastor: Fine-tuned Small Language Models for Shape-based Active Relation Extraction. Extended Semantic Web Conference 2025, Jun 2025, Portoroz, France. (hal-05078493)

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

Assessing argument sufficiency with Large Language Models and logic

Responsable scientifique : Serena Villata

Laboratoire : I3S

Résultats obtenus grâce aux moyens informatiques attribués :

Rapport de stage de Loup Doinel «Apprentissage automatique appliqué au raisonnement argumentatif. », GRENOBLE INP Esisar (2025).

Description :

The project is to explore different strategies for Argument Analysis based on Large Language Models (LLMs). In particular, we initially focus on the tasks of Argument Sufficiency Assessment, which is about analysing how well an argument's premises supports its claim. The first research question is about evaluating LLMs on this task. In particular, this involves testing a fine-tuning set-up with some LLMs of reasonable size (small Llama3 models, mistral models, ...) in order to evaluate their performance on an argument sufficiency assessment dataset (based on real essays) and compare it with state-of-the-art methods. Moreover, since this task is about a logical analysis of the arguments, we will evaluate how using the LLMs to express these arguments in logic could help for the assessment, in terms of performance, explainability and bias reduction. This involves the tasks of translation from natural language to logic, generation of missing commonsense knowledge, feeding the generated formulas to a logic solver, generation of explanations, and likely more. Various strategies for natural language to logic translation should be explored, including fine-tuning and possibly others (Reinforcement Learning with Human Feedback, Tree-of-Thought prompting, etc...).

The goal of this project is thus to help to provide answers to various research questions:

- How well can LLMs perform argument quality assessment? sufficiency assessment in particular?
- How well can they generate logic following given constraints based on real natural language arguments?
- Are the logic expressions they generate faithful to the text semantics?
- Which prompting/ fine-tuning strategies can give the best results?

- Can we design a framework for argument quality assessment involving an intermediary step of logic translation and making a final decision based on logic solvers (which would allow better interpretability and control of bias)?
- Which logic formalism (First Order, Second Order, Fuzzy Logic, etc...) is the best suited for these tasks?

This work could have multiple real world applications, in particular improving tools for efficient and transparent prevention of disinformation and fallacies in online information streams.

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

Responsable scientifique : Serena Villata

Laboratoire : INRIA, I3S, CNRS, UniCA, 3IA

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 :

Nous avons obtenu en 2025 avec le code SDM (voir sa page web) une nouvelle collaboration (SWAM) avec le Chili.

Deux projets sont en cours

- Calibration under uncertainty of turbulent closure with Boussinesq approximation (stratified atmosphere)
- CDF Calcul for brine discharge dispersion in coastal marine environment.

Ce dernier projet en phase de test va demander des calculs bientôt plus lourds sur Azzurra.

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'appuyant 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 :

Article publié

Singlan, Nina, Fadi Abou Choucha, and Claude Pasquier. "A new Similarity Based Adapted Louvain Algorithm (SIMBA) for active module identification in p-value attributed biological networks." *Scientific Reports* 15.1 (2025): 11360.

HAL : hal-05020436 v1

DOI : 10.1038/s41598-025-95749-6

Articles soumis ou en préparation

"MEAST: a Novel Framework for Analyzing Gene Set Activity and Stability in Transcriptomics" par Fadi Abou Choucha, and Claude Pasquier, soumis au journal *Scientific Reports*

"Efficient Mining of Closed Contiguous Sequences through Target Motif Extension" par Claude Pasquier, Winona Pasquier, Célia da Costa Pereira et Andrea Tettamanzi, soumis à la conférence ICDE

"Contiguous targeted sequence mining in attributed trees", par Claude Pasquier, Batoul Hayda, Nanée Chahinian, Célia da Costa Pereira et Andrea Tettamanzi (en phase de rédaction)

Manuscrit de thèse de Nina Singlan intitulé "Détection de communautés évolutives dans des graphes pondérés par les nœuds", en phase finale de rédaction avec une soutenance prévue en décembre 2025

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 canalaire 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 cancérogènes 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 :

- M. Elsawy, A. Gobé, G. Leroy, S. Lanteri and C. Scheid. Advanced computational design of complex nanostructured photonic devices using high order discontinuous Galerkin methods and statistical learning global optimization. Proceedings of the Emerging Technologies in Computational Sciences for Industry, Sustainability and Innovation - Math to Product 2023 Conference. Lecture Notes in Computational Science and Engineering, Springer, Volume 146, to appear (2025)
- M. Elsawy, A. Gobé, G. Leroy and S. Lanteri. Inverse Design of Photonic Devices with Statistical Learning-Based Global Optimization Algorithms. 46th PIERS (Photonics and Electromagnetics Research Symposium), Abu Dhabi, UAE, 4-8 May 2025
- Bellouch, M. Elsawy, S. Lanteri, E. Vandelle and T.Q.V. Hoang. Nonlocal metasurfaces based on advanced Bayesian learning for satellite communications. Under review in Communications Engineering, Springer-Nature (2025). Preprint on HAL: <https://hal.science/hal-04850081v1>

Description :

Atlantis (<http://www-sop.Inria.fr/atlantis/>) est une équipe-projet commune entre 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 les interactions lumière-matière à l'échelle du nanomètre.

Le domaine scientifique associé est la nanophotonique. 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. Dans ce contexte, on étudie différentes problématiques en collaboration avec des

chercheurs physiciens de laboratoires du CNRS (notamment ceux du périmètre UniCA, i.e., le CRHEA et INPHYNI) et des partenaires industriels (notamment, STMicroelectronics à Crolles, Thales R&T à Palaiseau et la startup Solnil à Marseille).

Nos contributions méthodologiques sont déployées dans la suite logicielle DIOGENeS (<http://diogenes.inria.fr>). dédiée à la nanophotonique computationnelle.

DIOGENeS intègre plusieurs solveurs numériques haute fidélité basés sur la méthode Galerkin discontinue, ainsi que des workflows d'optimisation numérique pour la conception de nanostructures et dispositifs photoniques guidée par des objectifs de fonctionnalisation et performance spécifiques aux applications envisagées.

Plus récemment, nos recherches méthodologiques se sont aussi orientées vers la réduction de modèle pour mettre au point des solveurs numériques rapides.

Différentes approches sont étudiées qui font toutes appel à l'apprentissage machine par réseaux de neurones.

Un point sur les membres d'Atlantis qui ont des comptes sur Azzurra :

Claire Scheid : MCF UCA, membre du LJAD et membre d'Atlantis; elle utilisera Azzurra plus régulièrement dans les 3 prochaines années avec le démarrage du projet ANR SWAG-P

Mahmoud Elsawy : ISFP (Inria Starting Faculty Position); il utilise très régulièrement Azzurra pour sa recherche sur la conception de métasurfaces actives et passives

Alexis Gobé (IR Inria, membre du SED du centre) et Guillaume Leroy (ingénieur R&D en CDD) : ils utilisent Azzurra principalement pour le portage et l'optimisation d'outils numériques de la suite DIOGENeS;

Enzo Isnard : doctorant Cifre avec Thales R&T (en seconde année de thèse); sa thèse porte sur la conception de métasurfaces dans des systèmes optiques complexes

Alexandre Pugin : actuellement en stage pré-doctoral, et en thèse à partir d'octobre 2024; sa thèse porte sur l'étude du concept de PINN (Physics-Informed Neural Network) pour l'électromagnétisme

Thibault Laufroy : doctorant mais qui depuis quelques mois en CDI chez Air France (il soutiendra sa thèse fin 2024); sa thèse porte sur la modélisation numérique pour la thermoplasmonique

Projets en cours qui impliquent des simulations sur Azzurra:

DGA RAPID Aerocom (du 1/1/2023 au 30/9/2025) avec Thales Research & Technology à Palaiseau et la société NANOE à Ballainvilliers.

Titre complet: AntennE ultRa cOmpacte à balayage mécanique, fabriquée par impression 3D de matériaux composite haute permittivité, pour l'application satCOM.

Ce projet vise à la conception d'antennes basées sur des métadéflecteurs et leur fabrication par impression 3D.

Notre contribution concerne les aspects de conception numérique en exploitant les outils de la suite DIOGENeS.

L'essentiel de mon utilisation d'Azzurra depuis janvier 2023 est en lien avec ce projet. Une première publication est en préparation.

ANR SWEET (<https://anr.fr/Project-ANR-22-CE24-0005>, du 1/1/2023 au 31/12/2026) piloté par le LAAS, avec le CRHEA et la société NAPA Technologies.

Titre complet : Sub-WaveLength Electro-optic sysTems

Ce projet porte sur la conception de systèmes optiques basés sur des métasurfaces actives.

Mahmoud Elsawy pilote notre contribution à ce projet et il utilise Azzurra (notamment) pour les études réalisées dans le cadre de ce projet.

ANR SWAG-P (<https://anr.fr/Project-ANR-23-CE09-0014>, du 1/10/2023 au 31/12/2027) piloté par l'Institut Pascal, avec l'IM2NP, le CINaM et l'ICR.

Titre complet : Sensing With A Gap-Plasmon

Ce projet porte sur la conception de résonateurs à gap-plasmon pour la détection de substances chimiques. Une nouvelle doctorante devrait arriver en novembre 2024 pour travailler avec Claire Scheid sur ce projet.

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).

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

Responsable scientifique : Serena Villata

Laboratoire : INRIA, I3S, CNRS, UniCA, 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.

Compression des réseaux de neurones

Responsable scientifique : Thanh Phuong Nguyen

Laboratoire : I3S

Description en pages suivantes.

FOAM: eFfective cOmpression of trAnsformer Models

Thanh Phuong Nguyen^{a,*}

^a*I3S, UMR 7271, France*

Keywords: network compression, tensor decompositions, pruning

1. Related work for network compression

In this section, we offer an overview of deep neural networks (DNNs) compression, beginning with a brief review of commonly used compression techniques for DNNs. Subsequently, we detail some particular perspectives for compression of transformer models with a focus on low-rank approximations.

1.1. Introduction to network compression

In recent years, DNNs have achieved remarkable progress across a range of fields, including computer vision [29], natural language processing [27], and signal processing [2], to mention a few. Recently, a transformer model which is a Sequence-to-Sequence neural net transforming a given sequence of elements, such as the sequence of words in a sentence, into another sequence is particularly exploited in Large Language Models (LLM) leading to the huge success of ChatGPT. Despite their effectiveness, the large number of parameters in DNNs leads to considerable processing time and memory demands [1]. These high computational costs present a significant challenge for deploying DNNs on edge devices with constrained resources, such as memory, computation, and energy [11]. For example, in the Internet of Things (IoT) applications, where edge devices like sensors and embedded systems are widely used, these constraints are particularly critical [3]. Real-world IoT applications—such as autonomous vehicles and environmental monitoring—demand both high accuracy and efficient resource utilization, including real-time inference and a compact memory footprint [25]. The computational complexity of DNNs often makes them impractical for these edge devices [12]. To address this challenge, researchers have developed various DNN compression techniques, such as pruning [1], quantization [11], knowledge distillation [12], neural architecture search (NAS) [8], and low-rank approximations [4].

Pruning reduces model size and computational complexity by eliminating unnecessary, redundant or less important parameters. Pruning can be examined across five critical aspects: what to prune, when to prune, how to prune, how often to prune, and how to configure pruning. First, pruning components can be unstructured, targeting individual weights, or structured, focusing on entire neurons, filters, or channels; semi-structured pruning offers a compromise by removing groups of parameters while retaining some structures. Second, pruning can occur at different stages: before training, during training, or after training. There are also run-time pruning methods that dynamically adapt the model based on the input. Third, the criteria for pruning, which determines how to prune, include magnitude-based methods (removing small weights), norm-based approaches (using norms of filters), similarity-based (removing similar units), sensitivity and saliency analysis (evaluating the impact of each component), and loss-based approaches (measuring how pruning affects the loss function). Fourth, pruning can be performed in a one-shot manner, where a single pruning step is applied, or iteratively, where the model is gradually pruned over multiple steps. Finally, the pruning configuration can be manual, with hyperparameters set by the user, or automatic, leveraging algorithms that adjust pruning strategies during optimization. We refer readers to [1] for a comprehensive review on pruning.

Quantization reduces the bit-width of weights and activations from high-precision 32-bit floating-point representations to 16-bit, 8-bit, or even lower. There are different types of quantization, such as uniform quantization, where

*Corresponding author

Email address: thanh-phuong.nguyen@univ-cotedazur.fr (Thanh Phuong Nguyen)

all values are evenly spaced, non-uniform quantization, which assigns more precision to critical parameters, and learned quantization, where the quantizer is learnable and is often compatible with bitwise operations to accelerate the inference phase. The granularity of quantization can vary, including layerwise, groupwise, channelwise, and sub-channelwise approaches. Quantization can be applied at different stages: during training, namely quantization-aware training (QAT), or after training, namely post-training quantization (PTQ). QAT incorporates the effects of quantization during model optimization, helping to reduce the accuracy loss typically seen with lower-bit representations. In contrast, PTQ is simpler and faster to implement but may result in greater performance degradation. Additionally, mixed-precision quantization can be used, where different layers or components are quantized to varying bit-widths based on their sensitivity to precision loss. For a comprehensive review of DNN quantization, we refer readers to [11]. Knowledge distillation transfers knowledge from a large, complex neural network (the teacher model) to a smaller, more efficient network (the student model), intending to maintain the performance of the teacher while significantly reducing the student model's computational and memory demands. During training, the student model learns not only from the ground truth labels but also from the predictions of the teacher, which provide more nuanced information about the output distribution, including the relative probabilities of incorrect classes. This enables the student to better capture the teacher's generalization patterns. Distillation strategies are primarily divided into two categories: offline distillation, where the teacher model is pre-trained and remains fixed, and online distillation, where both the teacher and student are trained simultaneously. Readers interested in a detailed survey can refer to [12].

NAS is an automated technique for designing optimal neural network architectures, aiming to outperform manually crafted models in both accuracy and efficiency. NAS explores a vast search space of possible architectures and selects the best-performing designs based on predefined objectives such as accuracy, latency, or model size. Typically, NAS frameworks consist of three components: the search space, which defines the set of candidate architectures; the search strategy, which determines how to explore the search space (e.g., random search, reinforcement learning, evolutionary algorithms); and the evaluation method, which measures the performance of candidate architectures, often using proxy tasks or performance predictors to reduce computational costs. A review of NAS is available in [8].

Finally, a brief review of low-rank approximation for transformer compression is addressed in Section 2.3.1

1.2. Brief review of transformer models

A Transformer block [27] consists of two main components: an attention layer and a feed-forward network (FFN). In the attention layer, there are four weight matrices, which are used for linear projections: query (\mathbf{W}_Q), key (\mathbf{W}_K), value (\mathbf{W}_V), and output (\mathbf{W}_O), each of size $d \times d$, where d represents the dimension of an input embedding vector (i.e., the number of features). For an input sequence of length n (i.e., n tokens), the input to the attention layer, \mathbf{X} , is an $n \times d$ matrix. Each of the query, key, and value matrices, \mathbf{X}_Q , \mathbf{X}_K , and \mathbf{X}_V , respectively, is computed by multiplying the input matrix with their corresponding weight matrices: $\mathbf{X}_Q = \mathbf{X} \times \mathbf{W}_Q$, $\mathbf{X}_K = \mathbf{X} \times \mathbf{W}_K$, and $\mathbf{X}_V = \mathbf{X} \times \mathbf{W}_V$. After these projections, the attention score, \mathbf{X}_S , is computed as softmax($\mathbf{X}_Q \times \mathbf{X}_K^T$). Finally, the output of the attention layer, \mathbf{X}_O , is computed by multiplying the attention value ($\mathbf{X}_S \times \mathbf{X}_V$) with the output weight matrix \mathbf{W}_O , and adding a residual connection to the input: $\mathbf{X}_O = \mathbf{X}_S \times \mathbf{X}_V \times \mathbf{W}_O + \mathbf{X}$. The FFN processes the output of the attention layer, \mathbf{X}_O . It consists of two FC layers, with corresponding weight matrices \mathbf{W}_{FC1} and \mathbf{W}_{FC2} , of dimensions $d \times h$ and $h \times d$, respectively. The output of the FFN is computed as ReLU($\mathbf{X}_O \times \mathbf{W}_{FC1}$) $\times \mathbf{W}_{FC2} + \mathbf{X}_O$. Figure 1 presents the architecture of a transformer model while Figure 2, which is reproduced from Sapna Limbu, highlights how to calculate multi-head self attention in a layer.

2. Proposal for compression of transformer models

In this project, we would like to tackle the following challenges which make the huge computation complexity and memory foot-print of transformer models : matrix multiplication in self attention modules and Feed-Forward Network blocks.

2.1. Compression of Self Attention modules

A transformer model is a sequence-to-sequence deep model taking a sequence of embedding items and outputs another sequence of items. The items can be words of sentence in Natural Language Processing tasks or can be image patches in Computer Vision tasks. The main computing component of attention module consists of the following steps :

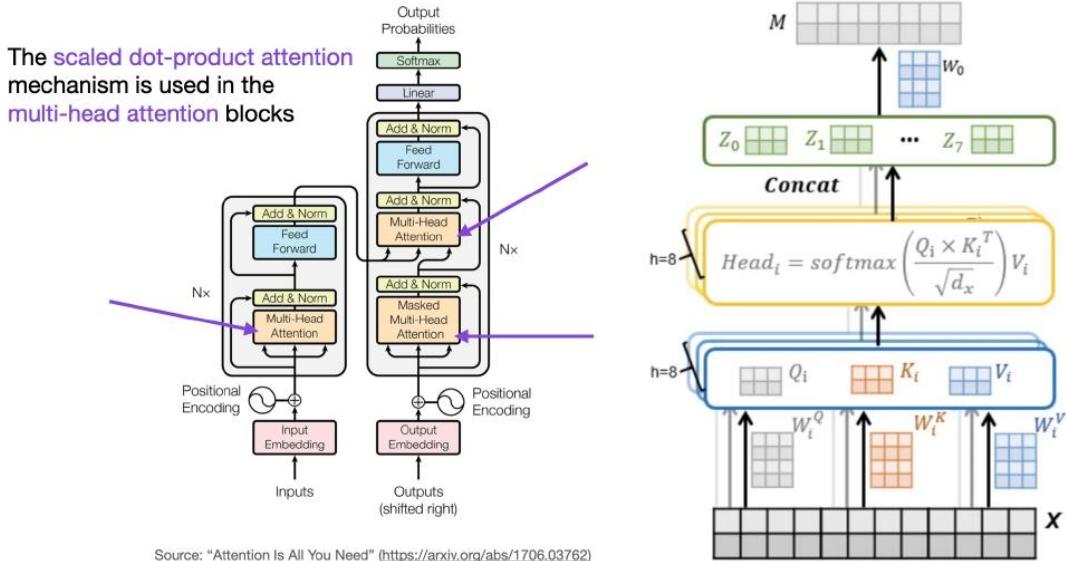


Figure 1: Architecture of a transformer model.

Figure 2: Multi-head attention calculation.

- The first step is to create three vectors from each of the encoder's input vectors x_i (i.e word) of dimension d , where $1 \leq i \leq n$. So for each word, we create a Query vector, a Key vector, and a Value vector. These vectors are created by multiplying the embedding input vector by three matrices that we trained during the training process (i.e. query (\mathbf{W}_Q), key (\mathbf{W}_K), value (\mathbf{W}_V)). If we represent the sequence of n item by matrix X of size $n \times d$ where each row i represents each input vector x_i , this step can be summarized by three operators of matrix multiplication : $Q = X \times \mathbf{W}_Q$, $K = X \times \mathbf{W}_K$, and $V = X \times \mathbf{W}_V$.
- The second step in calculating self-attention is to calculate a score of self similarity. For example, let us calculate the self-attention for the first item, we need to score each item of the input sequence of items against this item. The score determines how much focus to place on other parts of the input sequence as we encode a item at a certain position. The score is calculated by taking the dot product of the query vector with the key vector of the respective item we're scoring. So if we're processing the self-attention for the item in position 1, the first score would be the dot product of q_1 and k_1 . The second score would be the dot product of q_1 and k_2 . Finally, these scores are used to weight the values matrix by using softmax function. This step can be resumed by the following operator of matrix multiplication.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

It could be seen that all two main above operators of Self Attention Module are relied on matrix multiplication operators. Our proposal is hence interested in fast and memory-efficient matrix multiplication. The following perspectives can be then considered.

- Fast matrix computation algorithms: Our purpose is to introduce new effective algorithms for matrix multiplication or to improve existing ones such as [20], [22], [9], [7].
- Effective matrix multiplication based on quantization: Reducing the number of bits for representation of a weight allows to accelerate matrix multiplication while also reducing the model size [6]. In addition, a multiplication of two floating point numbers consumes much more necessary energy and computational cost than that of two integers. We try to introduce novel quantization approaches reducing effectively both memory foot-print

- and computation cost related to matrix multiplication in self attention sub-layer.
- Matrix decomposition: Tensor decomposition techniques allow to decompose higher-rank tensors into lower-rank representations in order to simplify memory foot-print for compressing model size and to accelerate inference phase by reducing computational complexity. We will detail the perspectives in this direction in Section 2.3.2.

2.2. Compression of Feed-Forward Network blocks

Feed-Forward Network block is a sub-unit within each encoder and decoder layer in the transformer model. It takes the output from the self-attention sub-layer capturing relationships between different parts of the input sequence and then transforms it. It takes into account two main following tasks. First, the FFN, which consists of two linear transformations separated by an activation function (typically ReLU), assumes a non-linearity transformation. Second, the FFN provides each token with a richer, more refined representation, which helps the model capture intricate features. Since the FFN is position-wise, it refines each token’s representation independently, complementing the context-based processing done by the attention mechanism.

Because the FFN can be seen as Fully Connected Layers (i.e. CNN layers of kernel 1×1 in CNN architecture, our recent advances [23, 24] in CNN compression can be good starting points to develop their direct perspectives for reducing model size and computational complexity of FFN blocks. Some further developments based on low-rank representation will be detailed in Section 2.3.2. In addition, our perspectives in this direction are relied on unstructured pruning approach to suppress unnecessary or less important weights. On the other hand, typical techniques [31, 5] for compression of Fully Connected Layer can also be investigated for this purpose.

2.3. Low-rank approximations for compression of transformer model

2.3.1. Related work

The weights of DNNs, also called parameters, are inherently structured as matrices or higher-order tensors. The primary operations in DNNs—matrix multiplication and convolution—are applied to tensor data as it moves through the network, accounting for a principal portion of the computational load. As a result, low-rank approximations emerge as an effective approach for reducing both the number of parameters and the computational complexity of DNNs. This technique leverages the low-rank nature of the model’s weights, factorizing them into smaller matrices or tensors using either matrix or tensor decompositions. In the following, we review decomposition methods tailored to Transformer architecture.

Since matrix multiplication is the primary operation in transformer blocks, SVD naturally emerges as a popular method to factorize and compress weight matrices [21, 16, 14, 13, 10, 26]. On the GLUE benchmark dataset, Noach *et al.* [21] applied this technique to compress the BERT-base model by first decomposing the weight matrices into smaller components and then used feature distillation to recover performance. Fisher-weighted SVD [16] enhanced this by leveraging Fisher information to prioritize the parameters critical for model predictions, thereby minimizing the performance drop caused by truncating the decomposed matrices. Hajimolahoseini *et al.* [14] compared two strategies—layer-by-layer and progressive decomposition—on pre-trained transformer-based models like ViT on CIFAR and GPT-2 on WikiText-103. Guo *et al.* [13] introduced PELA, a method that enhances the parameter efficiency of pre-trained vision transformer models by integrating SVD with feature distillation and weight perturbation. More recently, Gao *et al.* [10] introduced an adaptive rank selection approach to compress BERT and LLaMA-7B on GLUE. Furthermore, LASER [26] selectively factorized the weight matrices of the GPT-J model on the CounterFact dataset, demonstrating performance gains.

Recently, there has been a growing trend toward applying higher-order decomposition methods, such as BTD and TTD, by tensorizing weight matrices in transformer models into higher-order tensors for more effective compression [19, 15, 30, 17, 18, 28]. Ma *et al.* [19] introduced a novel self-attention model that combines BTD with parameter sharing, achieving strong performance on three language modeling tasks (PTB, WikiText-103, One-Billion) and a neural machine translation task (WMT-2016 English-German). Hrinchuk *et al.* [15] used TTD to compress the embedding layers of transformers across four tasks: sentiment analysis, neural machine translation, language modeling, and click-through rate prediction. For vision tasks such as object classification and detection, Zhen *et al.* [30] reshaped the weight matrices of FC layers and then adopted TTD on them, compressing Detection Transformers (DETR) on

ImageNet and COCO. In another effort, FacT [17] tensorized the weights of vision transformer models into a third-order tensor, which was then decomposed into lightweight factors for parameter-efficient transfer learning. In T3SRS, Li *et al.* [18] introduced a tensor train layer that decomposes the original weight matrix into smaller tensors, effectively compressing sequential recommender systems. More recently, Yang *et al.* [28] introduced LoRETTA, a parameter-efficient framework that employs TTD to significantly reduce trainable parameters, demonstrating strong performance on BERT-family and LLaMA-2 models.

2.3.2. Some perspectives in low-rank representation

We will take into account low-rank representation in different components of a transformer model. In the following, we detail some perspectives for each component.

1. **Effective representation of token embeddings:** We are interested in an effective representation of token embeddings by low-rank representation. This allows to save considerably data volume on disk.
2. **Weight matrix decompositions:** Weight matrix factorization allows to reduce the model size and computation complexity since it replaces the original weight matrix with the product of smaller matrices.
3. **Multi-head attention decomposition:** It should be noted that the attention calculation over the entire sentence (i.e. sequence of items) consists of a large number of redundant computations. A classic solution to resolve this issue is to address attention calculation in smaller groups by either binning them using spatial locality, magnitude-based locality, or an adaptive attention span. In addition, because the outputs are calculated independently, local attention methods also enable a higher degree of parallel processing and individual representations can be saved during inference for multiple uses. This allows to improve the runtime memory costs and execution speed but not the model size since the multi-head self-attention layer does not contain weights. Our perspective is to enhance the attention decomposition by reducing its complexity. This can be done jointly with weight matrix decomposition in the above step.
4. **FFN compression based on low-rank representation:** A Fully-Connected layer can be seen as a 4-order tensor. Therefore, tensor decomposition techniques can be used to obtain effective low-rank approximation of this layer. This allows to reduce the memory foot-print and computational complexity of FFN consisting of 2 FC layers. In this direction, we investigate several typical techniques such as Tucker decomposition, CPD, and so on for this purpose.

References

- [1] Cheng, H., Zhang, M., Shi, J.Q., 2024. A survey on deep neural network pruning: Taxonomy, comparison, analysis, and recommendations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [2] Cheng, Z., Li, B., Fan, Y., Bao, Y., 2020. A novel rank selection scheme in tensor ring decomposition based on reinforcement learning for deep neural networks, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3292–3296.
- [3] Dai, C., Liu, X., Cheng, H., Yang, L.T., Deen, M.J., 2021. Compressing deep model with pruning and tucker decomposition for smart embedded systems. *IEEE Internet of Things Journal* 9, 14490–14500.
- [4] Dai, W., Fan, J., Miao, Y., Hwang, K., 2023. Deep learning model compression with rank reduction in tensor decomposition. *IEEE Transactions on Neural Networks and Learning Systems* , 1–14.
- [5] Dara Nagaraju, N.C., 2023. Compressing fully connected layers of deep neural networks using permuted features. *IET Computers & Digital Techniques*.
- [6] Dettmers, T., Lewis, M., Belkada, Y., Zettlemoyer, L., 2022. Llm.int80: 8-bit matrix multiplication for transformers at scale. *CoRR* abs/2208.07339. URL: <https://doi.org/10.48550/arXiv.2208.07339>, doi:10.48550/ARXIV.2208.07339, arXiv:2208.07339.
- [7] Drevet, C., Islam, M.N., Schost, É., 2011. Optimization techniques for small matrix multiplication. *Theor. Comput. Sci.* 412, 2219–2236. URL: <https://doi.org/10.1016/j.tcs.2010.12.012>, doi:10.1016/J.TCS.2010.12.012.
- [8] Elskens, T., Metzen, J.H., Hutter, F., 2019. Neural architecture search: A survey. *Journal of Machine Learning Research* 20, 1–21.
- [9] Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera-Paredes, B., Barekatain, M., Novikov, A., Ruiz, F.J.R., Schriftwieser, J., Swirszcz, G., Silver, D., Hassabis, D., Kohli, P., 2022. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* 640, 47–53.
- [10] Gao, S., Hua, T., Hsu, Y.C., Shen, Y., Jin, H., 2024. Adaptive rank selections for low-rank approximation of language models, in: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 227–241.
- [11] Gholami, A., Kim, S., Dong, Z., Yao, Z., Mahoney, M.W., Keutzer, K., 2022. A survey of quantization methods for efficient neural network inference, in: *Low-Power Computer Vision*. Chapman and Hall/CRC, pp. 291–326.

- [12] Gou, J., Yu, B., Maybank, S.J., Tao, D., 2021. Knowledge distillation: A survey. *International Journal of Computer Vision* 129, 1789–1819.
- [13] Guo, Y., Wang, G., Kankanhalli, M., 2024. Pela: Learning parameter-efficient models with low-rank approximation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15699–15709.
- [14] Hajimolahoseini, H., Ahmed, W., Rezagholizadeh, M., Partovinia, V., Liu, Y., 2022. Strategies for applying low rank decomposition to transformer-based models, in: 36th Conference on Neural Information Processing Systems (NeurIPS2022), pp. 1–6.
- [15] Hrinchuk, O., Khrulkov, V., Mirvakhabova, L., Orlova, E., Osledets, I., 2019. Tensorized embedding layers, in: Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 4847–4860.
- [16] Hsu, Y.C., Hua, T., Chang, S., Lou, Q., Shen, Y., Jin, H., 2022. Language model compression with weighted low-rank factorization, in: International Conference on Learning Representations, pp. 1–11.
- [17] Jie, S., Deng, Z.H., 2023. Fact: Factor-tuning for lightweight adaptation on vision transformer, in: Proceedings of the AAAI conference on artificial intelligence, pp. 1060–1068.
- [18] Li, H., Zhao, J., Huo, H., Fang, S., Chen, J., Yao, L., Hua, Y., 2024. T3rs: Tensor train transformer for compressing sequential recommender systems. *Expert Systems with Applications* 238, 122260.
- [19] Ma, X., Zhang, P., Zhang, S., Duan, N., Hou, Y., Zhou, M., Song, D., 2019. A tensorized transformer for language modeling. *Advances in neural information processing systems* 32.
- [20] Martel, M., 2022. Compressed matrix computations, in: IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2022, Vancouver, WA, USA, December 6-9, 2022, IEEE. pp. 68–76. URL: <https://doi.org/10.1109/BDCAT56447.2022.00016>.
- [21] Noach, M.B., Goldberg, Y., 2020. Compressing pre-trained language models by matrix decomposition, in: Proceedings of the 10th International Joint Conference on Natural Language Processing, pp. 884–889.
- [22] Pagh, R., 2013. Compressed matrix multiplication. *ACM Trans. Comput. Theory* 5, 9:1–9:17. URL: <https://doi.org/10.1145/2493252.2493254>, doi:10.1145/2493252.2493254.
- [23] Pham, V.T., Zniyed, Y., Nguyen, T.P., 2024a. Efficient tensor decomposition-based filter pruning. *Neural Networks* , 106393.
- [24] Pham, V.T., Zniyed, Y., Nguyen, T.P., 2024b. Enhanced network compression through tensor decompositions and pruning. *IEEE Transactions on Neural Networks and Learning Systems* , 1–13.
- [25] Phan, A.H., Ermilov, D., Kozyrskiy, N., Vorona, I., Sobolev, K., Cichocki, A., 2024. How to train your unstable looped tensor network. *IEEE Journal of Selected Topics in Signal Processing* , 1–10.
- [26] Sharma, P., Ash, J.T., Misra, D., 2024. The truth is in there: Improving reasoning in language models with layer-selective rank reduction, in: The Twelfth International Conference on Learning Representations, pp. 1–12.
- [27] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I., 2017. Attention is all you need, in: *Advances in Neural Information Processing Systems*, pp. 1–11.
- [28] Yang, Y., Zhou, J., Wong, N., Zhang, Z., 2024. LoRETTA: Low-rank economic tensor-train adaptation for ultra-low-parameter fine-tuning of large language models, in: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pp. 3161–3176.
- [29] Zhang, X., Zou, J., He, K., Sun, J., 2016. Accelerating very deep convolutional networks for classification and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 1943–1955.
- [30] Zhen, P., Gao, Z., Hou, T., Cheng, Y., Chen, H.B., 2022. Deeply tensor compressed transformers for end-to-end object detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 4716–4724.
- [31] Zhou, S., Wu, J., 2015. Compression of fully-connected layer in neural network by kronecker product. CoRR abs/1507.05775. URL: <http://arxiv.org/abs/1507.05775>, arXiv:1507.05775.

ANR TRACTIVE: AI-driven video interpretation of character representation in films

Responsable scientifique : Lucile Sassatelli

Laboratoire : Laboratoire I3S - Université Côte d'Azur, CNRS

Résultats obtenus grâce aux moyens informatiques attribués :

- [Tores et al., 2025] Tores, J., Ancarani, E., Sun, R., Sassatelli, L., Wu, H.-Y., & Precioso, F. (2025). Improving concept-based models for challenging video interpretation. In Proc. ACM Int. Conf. Multimedia, Dublin, Ireland.
- [Ancarani et al., 2025] Elisa Ancarani, Julie Tores, Rémy Sun, Lucile Sassatelli, Hui-Yin Wu, and Frederic Precioso. 2025. Leveraging Concept Annotations for Trustworthy Multimodal Video Interpretation through Modality Specialization. In Proceedings of the 1st International Workshop on Cognition-oriented Multimodal Affective and Empathetic Computing (CogMAEC '25). Association for Computing Machinery, New York, NY, USA, 11–19. <https://doi.org/10.1145/3746277.3760415>

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.

AVEC-SOUTIEN : un Assistant Virtuel Et Contextuel – Soutien Opérationnel aux Utilisateurs pour les Travailleurs en Intervention, Exploitation et Navigation

Responsable scientifique : Gaetan Rey

Laboratoire : I3S UMR 7271 Université Côte d'Azur CNRS

Description :

Contexte applicatif

Les travailleurs mobiles – techniciens de maintenance, équipes de sécurité civile, opérateurs industriels – évoluent dans des environnements où **l'accès instantané à l'information critique et la capacité à agir efficacement sont vitaux**. Les interfaces actuelles (tableaux de bord, moteurs de recherche documentaires, assistants vocaux généralistes) présentent des limites importantes :

- charge cognitive élevée en situation de stress [8],
- absence de prise en compte du **contexte dynamique** (localisation, état de l'utilisateur, capteurs temps réel) [3],
- dépendance au cloud et consommation énergétique incompatible avec la mobilité embarquée[7].

Les avancées récentes en **modèles de langage (LLMs)** et en **génération augmentée par récupération (RAG)** ouvrent une voie nouvelle, mais restent peu adaptées aux environnements contraints et critiques. Notre projet **Genius** propose une solution innovante :

- s'appuyer sur des **Small Language Models (SLMs, 1B–3B paramètres)** optimisés pour **l'efficacité énergétique [2]**,
- intégrer une **modélisation multi-niveaux du contexte utilisateur** (statique, dynamique, cognitif, organisationnel) [1] [6],
- combiner **RAG et outils externes** pour exploiter conjointement connaissances froides (procédures, manuels, cartographies) et données chaudes (capteurs, localisation, événements) [5].

Objectifs scientifiques et retombées

L'objectif est de concevoir un assistant conversationnel embarqué, capable de fournir « le bon savoir au bon moment », avec une empreinte énergétique maîtrisée et une fiabilité suffisante pour des environnements critiques.

Les retombées attendues sont :

- **scientifiques** : nouvelles méthodes de modélisation contextuelle et d'intégration SLMs + RAG,
- **technologiques** : prototype embarqué validé en scénarios réels (maintenance, urgence, campus intelligent),
- **environnementales** : sobriété computationnelle et réduction de l'empreinte énergétique des IA,
- **sociétales** : amélioration de la sécurité, de l'efficacité et des conditions de travail des personnels mobiles.

Travaux scientifiques en cours

Dans le cadre de travaux préparatoires, nous avons initié une première étape vers la conception d'assistants virtuels intelligents frugaux, destinés à des agents mobiles. Nous avons élaboré des jeux de données originaux en français, intégrant à la fois des contextes statiques (infrastructures, outils) et dynamiques (position, environnement), et entraîné des modèles de langage de petite taille (1 à 3 milliards de paramètres). Ces expérimentations, menées notamment sur le contrôle de l'éclairage d'un campus universitaire, ont montré que même des modèles réduits, après ajustement, peuvent atteindre une bonne précision, détecter les commandes erronées et s'adapter à des contextes nouveaux, tout en conservant des performances compatibles avec des environnements contraints. Ce travail préparatoire a déjà donné lieu à une publication scientifique ([*Intelligent Virtual Assistant for Mobile Workers: Towards Hybrid, Frugal and Contextualized Solutions*](#)) et constitue une base solide pour le développement de solutions hybrides, robustes et économies dans le cadre du projet proposé.

Nos besoins

Le développement d'un assistant contextuel sobre exige une phase initiale de calcul intensif pour :

1. **Préparer et affiner des SLMs (1B–3B paramètres)** via des techniques de distillation [6], **quantification [2]** et **fine-tuning (LoRA, Adapters) [4]**.

2. **Constituer et vectoriser de larges bases documentaires** (procédures, cartes, données techniques) pour le RAG.
3. **Évaluer systématiquement** différents compromis performance / efficacité énergétique.

Bien que le produit final doive être **embarqué sur des dispositifs légers**, la RCD impose un accès à du matériel de pointe (GPU H100) pour :

- **entraîner / adapter des SLMs :**
 - fine-tuning d'un modèle 1,3B ≈ 30–60 GPU-heures (H100) selon corpus d'apprentissage et la taille du modèle,
 - quantisation, pruning et autres traitements : 5 – 10 GPU-heures.
 - évaluation du modèle entraîné : 20 – 30 heures.
 - expérimentation multi-variants (≈ 20 runs sur des modèles différents et des jeux de données différents) → ordre de grandeur : 2 000 GPU-heures,
- **prétraitement et vectorisation de données** : plusieurs millions de passages **documentaires, estimés à ≈ 200 GPU-heures,**
- **évaluations massives** (robustesse, explicabilité, consommation) nécessitant ≈ 300 GPU- heures supplémentaires.

Au total, nous estimons un besoin de l'ordre de **2 000–2 500 GPU-heures H100** pour couvrir les phases de recherche, optimisation et validation. Cette ressource est indispensable pour garantir que les modèles obtenus soient **compacts, fiables et sobres** une fois déployés localement.

Références

- [1] Baldauf, M., Dustdar, S., C Rosenberg, F. (2007). A survey on context-aware systems. *International Journal of ad Hoc and ubiquitous Computing*, 2(4), 263-277.
- [2] Dettmers, T., Lewis, M., Shleifer, S., C Zettlemoyer, L. 8-bit Optimizers via Block-wise Quantization. In *International Conference on Learning Representations*.
- [3] Dubiel, M., Halvey, M., CAzzopardi, L. (2018). A survey investigating usage of virtual personal assistants. *arXiv preprint arXiv:1807.04COC*.
- [4] Huan, M., CShun, J. (2025). Fine-tuning transformers efficiently: A survey on LoRA and its **impact**.
- [5] Kerdvibulvech, C., C Jiang, X. (2025, May). Generative AI in Human-Computer Interaction: Enhancing User Interaction, Emotional Recognition, and Ethical Considerations. In *International Conference on Human-Computer Interaction* (pp. 62-71). Cham: Springer
- [6] Rey, G. (2006). Méthode pour la modélisation du contexte d'interaction. *INGENIERIE DES SYSTEMES D'INFORMATION*, 11(5), 141.
- [7] Schmidhuber, J., Schlägl, S., C Ploder, C. (2021, September). Cognitive load and productivity

-
- implications in human-chatbot interaction. In *2021 IEEE 2nd international conference on human-machine systems (ICHMS)* (pp. 1-6). IEEE.
- [8] Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press.

Responsable scientifique : Marc Antonini

Laboratoire : I3S UMR 7271 Universite Cote d'Azur and CNRS

Description :

La mémoire de l'humanité repose sur notre capacité à gérer des quantités de données de plus en plus importantes. En 2040, on estime ainsi à plusieurs milliers de zettaoctets (1021 octets) la quantité de données qu'il faudra conserver, sur des périodes allant de quelques années à plusieurs siècles.

Dans ce contexte, les technologies actuelles de stockage sont insuffisantes pour garantir la future conservation d'une telle masse de données. Leur stockage et archivage dans des centres de données ne sera pas soutenable, en termes de ressources d'alimentation et de gestion de l'obsolescence du matériel. L'alternative la plus prometteuses est fournie par un stockage de l'information sous forme d'ADN, stable sur de très longues périodes avec des conditions simples de mise en œuvre.

L'objectif à 5 ans du PEPR , et plus particulièrement des 4 projets ciblés qui le composent, est de déployer les recherches et infrastructures nécessaires pour accélérer d'un facteur 100 la vitesse actuelle du cycle d'écriture/lecture d'ADN, tout en réduisant son coût d'un facteur équivalent.

Magnétohydrodynamique

EUR Spectrum

Responsable scientifique : Florence Marcotte

Laboratoire : LJAD

Résultats obtenus grâce aux moyens informatiques attribués :

Article de journal :

A robust, discrete-gradient procedure for optimisation with time-dependent PDE and norm constraints. Mannix, P.M., Skene, C.S., Auroux, D. & Marcotte, F. SMAI Journal of Computational Mathematics 10, pp 1-28 (2024).

Présentations en conférence : Les Houches 2025 et ICTW Durham 2025.

Description :

Modeling magnetic field generation by dynamo instability in stellar objects is a long-standing challenge with far-reaching implications for stellar evolution theory. Underlying motivations are exemplified by the need to understand stellar spin-down and accretion rates in protostellar discs, which are known to be dynamically impacted by magnetic fields. The interest sparked by recurring discrepancies between predictive evolution models and rapidly-progressing observations drives the current research into the characterization of dynamo mechanisms in stellar objects.

This important challenge cannot be solved analytically due to the strong nonlinearities of the magnetohydrodynamics (MHD) equations. Solving it therefore requires the development of innovative numerical approaches. In many astrophysical flows, infinitesimal magnetic seeds cannot be amplified by the flow, whereas finite-amplitude magnetic seeds with a favourable spatial structure can drive, through the Lorentz force nonlinear feedback, the very flow motions on which they subsequently feed by subcritical dynamo instability. This situation is particularly relevant for radiative stellar layers or for the innermost regions of protostellar discs, where the history of perturbations can thus define the magnetic fate of the object. Yet, classical stability methods fail to systematically characterize subcritical dynamo solutions and identify their critical dynamo seeds. The CIRCE project will tackle this theoretical obstacle by developing the recent mathematical tools of nonlinear stability analysis, based on adjoint-based optimal control, for MHD flows. The aim of CIRCE is to identify the least-energy perturbations that can trigger subcritical dynamos and transition to MHD turbulence in models of (a) radiative zones and (b) protostellar discs, and to predict how the resulting transitions determine rotational dynamics and accretion rates.

Effet Dynamo et Instabilités Magnétohydrodynamiques

Responsable scientifique : Yannick Ponty

Laboratoire : Lagrange / OCA, CNRS

Résultats obtenus grâce aux moyens informatiques attribués :

Yannick Ponty, Helene Politano, Annick Pouquet. Intermittency assessed through a model of kurtosis-skewness relation in MHD in fast dynamo regimes,

Journal of Plasma Physics. 91(2):E69 2025 <https://doi.org/10.1017/S0022377825000169>

Pouquet, A., Marino, R., Politano, H., Ponty, Y., and Rosenberg, D.: Intermittency in fluid and MHD turbulence analyzed through the prism of moment scaling predictions of multifractal models, Nonlin. Processes Geophys., 32, 243–259, 2025 <https://doi.org/10.5194/npg-32-243-2025>

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.

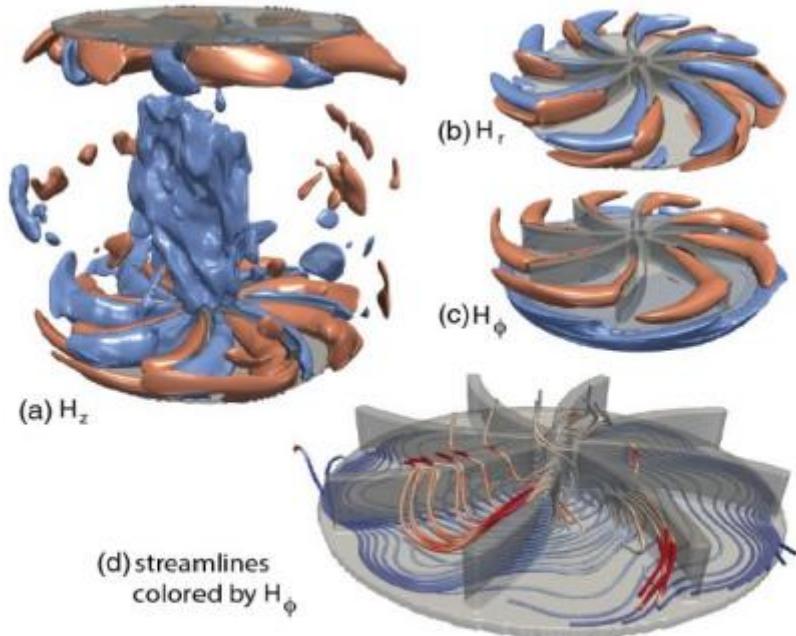


FIG. 4. (a) H_z , (b) H_r , (c) H_ϕ isoamplitude contours (red, positive value; blue, opposite negative value), and (d) magnetic streamlines colored by the H_ϕ value. $Re = 1500$, $Rm = 500$, and $\mu_r = 14$.

Fig1 : Visualisation des composantes du champ magnétique H_z , H_r , H_ϕ 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. Lett³⁶

Pour 2021 et ...

³⁶ 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).

- 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 $\theta = 0.23$

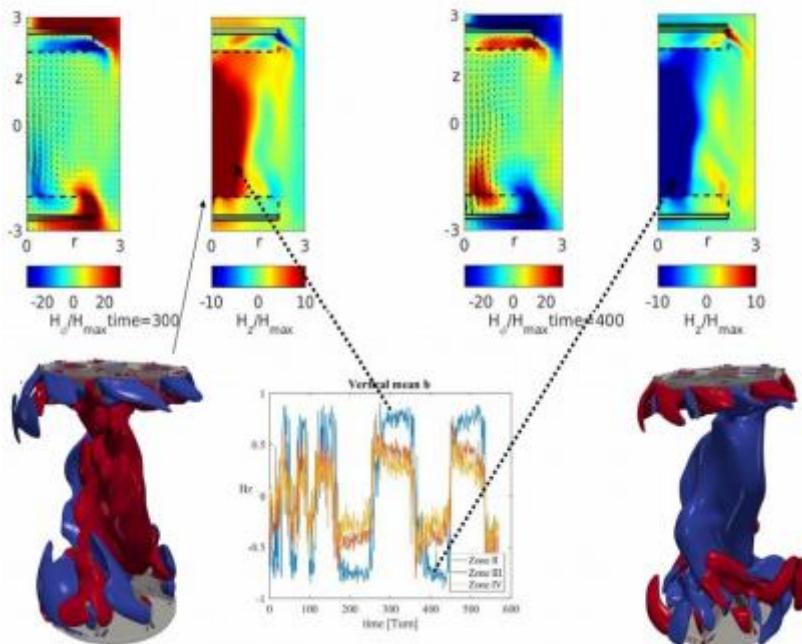


Fig 2: Inversion du champ magnétique dans une simulation où les deux disques ne tournent pas à la même vitesse : $\theta = (\Omega_2 - \Omega_1)/(\Omega_2 + \Omega_1)$.

II) Effet dynamo de fluctuation

{ 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 Kazantsev³⁷ (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

³⁷ Kazantsev, AP (1968) Sov. Phys. JETP 26, 1031

Kazantsev continue à revêtir une grande importance dans la théorie de la dynamo^{38 39}, 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 additif⁴⁰ 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 analytiques⁴¹.

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)}.

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

³⁸ Rincon, F (2019) J. Plasma Phys. 85, 205850401

³⁹ D. Vincenzi, The Kraichnan–Kazantsev Dynamo J. Stat. Phys. 106, 1073-1091 (2002)

⁴⁰ A A Schekochihin et al 2007 New J. Phys. 9 300

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

d'évolution, conditions de bord, etc) à l'aide de multiplicateurs de Lagrange⁴². 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 stratifié⁴³. L'application de la méthode DAL pour la première fois à des écoulements instationnaires governeds 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 DRESDYN⁴⁴) ou à Madison (projet MPDX⁴⁵). Ce projet utilisera le code Cubby qui permet facilement d'implémenter de nouvelles physiques ou des systèmes d'équations. 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

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

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

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

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

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.

Astrophysique

EUR Spectrum

EXOWINDS: High-resolution spectra of hot exoplanets

Responsable scientifique : Vivien Parmentier

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Adrien Simonnin. Thèse de Doctorat: *Exploration de la diversité physique des exoplanètes et des naines brunes à travers la spectroscopie infrarouge à très haute résolution au VLT et à l'ELT.* Ecole Doctorale Sciences Fondamentales et appliquées (EDSFA). Laboratoire Lagrange.

Analyse des données exowinds de TOI-1518b : Simonnin, A; Parmentier, V. et al. Time-resolved absorption of six chemical species with MAROON-X points to a strong drag in the ultra-hot Jupiter TOI-1518 b *Astronomy & Astrophysics*, Volume 698, id.A314, 24 pp. <https://hal.science/hal-05134227v1>

Modèles de circulation de WASP-80b. Présentation à la conférence Exoclimes VII Combining JWST data and General Circulation Models for a 3D view of the warm Jupiter WASP-80b (<https://exoclimes.org/program.html>), publication en préparation

Analyse des données exowinds de KELT-20b: Rapport de stage de Valentin De Lia, master MAUCA (2025).

Etude de l'atmosphère de WASP-121b et comparaison aux données JWST: rapport de Stage Pablo Drake, master MASS (2025).

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 :

- Papier en preparation sur les modeles MHD d'une etoile AGB (Chiavassa et al.)
- Poster dans la Conference "Why Galaxies Care About AGB Stars" November 17–21, 2025 in Santiago, Chile. Evolution vers des modeles binaires d'une etoile AGB. Tomassini et al.
- Problemes (resolus) pour le passage de ifort vers ifx
- Debut des tests pour la migration d'une partie du code vers le GPU.

Description du projet en page suivante.

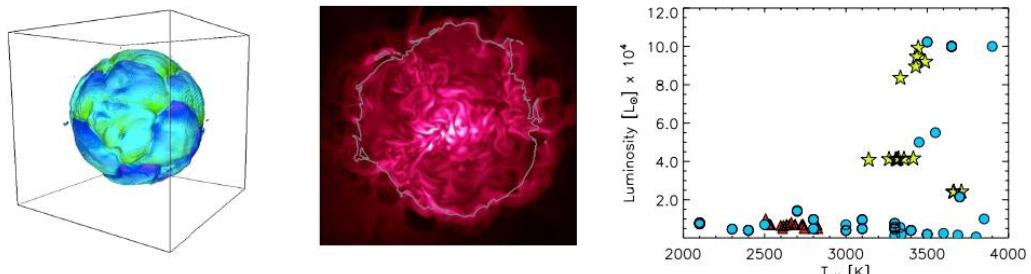
Red supergiant and AGB atmosphere and winds

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.



Couplage turbulent gaz-grains solides dans les disques protoplanétaires

Responsable scientifique : Héloïse Méheut

Laboratoire : Lagrange / OCA

Résultats obtenus grâce aux moyens informatiques attribués :

- Aurélien Crida, Clément Baruteau, Jean-François Gonzalez, Frédéric Masset, Paul Segretain, et al.. The reflex instability: exponential growth of a large-scale $m = 1$ mode in astrophysical discs. *Open Journal of Astrophysics*, 2025, 8, [10.33232/001c.143230](https://doi.org/10.33232/001c.143230). [hal-05222976](https://hal.archives-ouvertes.fr/hal-05222976)
- H. Meheut, F. Gerosa, J. Bec. A unique solution to overcome the barriers to planetesimal formation at low dust-to-gas ratio. *Semaine de l'astrophysique française*, 2025, Toulouse, France. [hal-05343429](https://hal.archives-ouvertes.fr/hal-05343429)

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 cisailé. 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 :

En 2024-2025, grâce à l'utilisation d'azzurra, nous avons abouti à la première sélection d'astéroïdes de la ceinture principale perturbant les orbites des planètes sur la base d'algorithme de machine-learning. Nous avons utilisé l'algorithme XGBoost d'arbres de décision pour classer les astéroïdes en fonction de leur perturbations principalement sur l'orbite de Mars. Nous avons ainsi profité du nombre de perturbateurs à considérer dans le modèle dynamique de Mars et réduit de 40% la sur-paramétrisation du modèle. Un papier au comité de lecture a été soumis à publication (V. Mariani, A. Flenga, Z. Murray et al. 2025, Boosting decision trees for Main Belt Asteroid selection in planetary ephemerides: an alternative model). Nous avons aussi travaillé sur une façon plus synthétique ces mêmes itérations. Z. Murray a développé une nouvelle modélisation en cours de tests.

3 présentations au symposium 401 de l'IAU seront faites pour présenter ces différents résultats : (Fienga et al. Updates on the INPOP ephemerides), (Murray, A PCA based effective model for asteroid accelerations), (V. Mariani et al. Boosting decision trees for Main Belt Asteroid selection in planetary ephemerides)

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 (Sambridge, 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 a 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 theses 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.

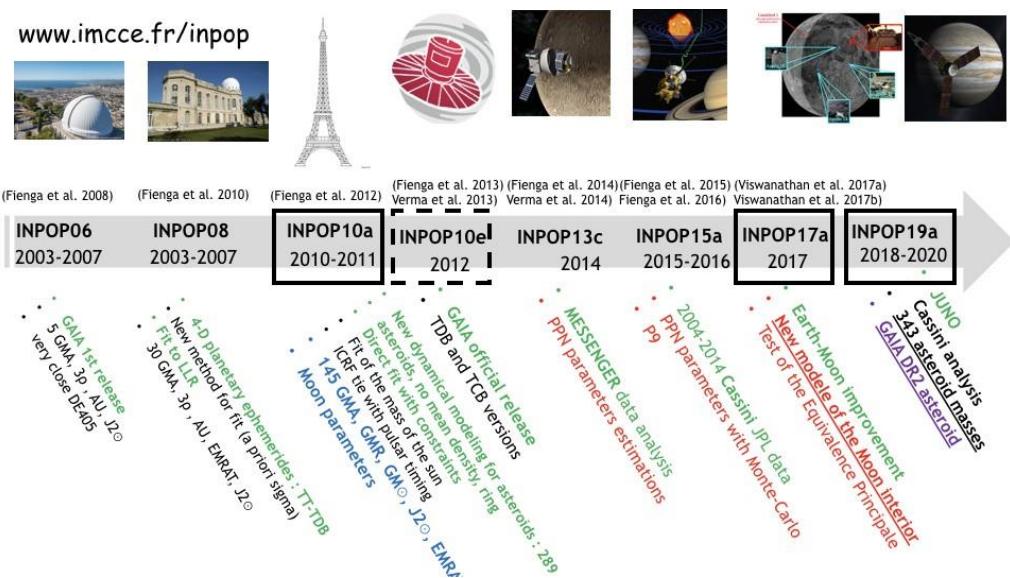


Figure 1: INPOP planetary, lunar and asteroid ephemerides Evolution

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.
- 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 (χ^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 β and γ in the frame of the Bepi-Colombo simulations. This work is currently under development. Results have already been obtained with a frequentist method (χ^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

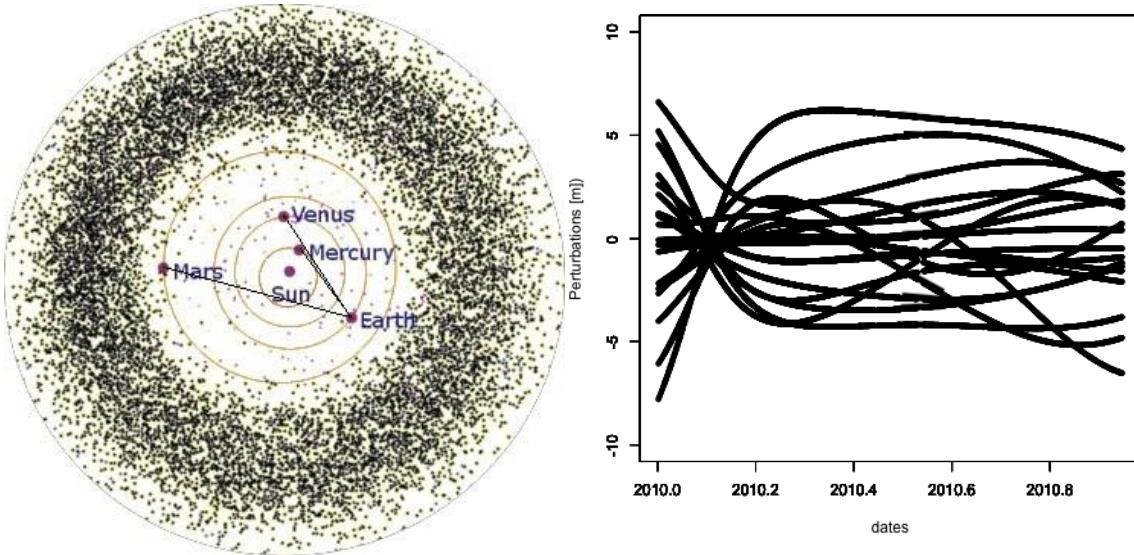


Figure 2: Main belt perturbations on inner planer orbits. On the right-hand side, is plotted an example of perturbations induced by 20 asteroids on the Earth-Mars distances over 1 year.

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 models 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 γ and about a factor 4 on PPN β thanks to the combination of Bepi-Colombo radio-science measurements and to derivative-free techniques.

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 χ^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. Cambridge, 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. Cambridge, Geophysical Journal International **138**, 727 (1999b).
- P. Kuchynka and W. M. Folkner, Icarus **222**, 243 (2013).

Détection d'exoplanètes par imagerie directe exploitant le machine learning

Responsable scientifique : David Mary

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Rapport de stage de Théo Llopis. Deep Learning pour la Détection d'Exoterres.

Ecole : ENSTA Bretagne (2025)

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 indécelables 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 :

Guinard, A., Fienga, A., Mémin, A., Ganino, C. 2024. Coupled tidal tomography and thermal constraints for probing Mars viscosity profile. AGU Fall Meeting Abstracts.

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/Géoazur 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

Résultats obtenus grâce aux moyens informatiques attribués :

Publication avec comité de lecture: Planetesimal gravitational collapse in a gaseous environment: Thermal and dynamic evolution Segretain, P.; Méheut, H.; Moreira, M.; Lesur, G.; Robert, C. & Mauxion, J., A&A, EDP Sciences, 2024, 692, A118

Présentation en atelier scientifique ou conférence: Idefix User days (Grenoble), Biphasic meeting (Nice), Les disques et la formation planétaire à l'ère du JWST et de l'ELT (Roscoff), Conférence de la Société Française d'Astronomie et d'Astrophysique (Marseille) [Titre:Planetesimal gravitational collapse in a gaseous environment, Auteurs: Segretain, P.; Méheut, H.; Moreira, M.]

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 du 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 2×10^7 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>.

Premier inventaire robuste de trous noirs supermassifs : l'apprentissage machine au service des missions spatiales Gaia et Euclid

Responsable scientifique : Eric Slezak

Laboratoire : Lagrange / OCA

Résultats obtenus grâce aux moyens informatiques attribués :

Le premier modèle d'apprentissage profond du projet est presque achevé : construction d'un autoencodeur variationnel, et association avec un "Normalizing Flow".

Aucune publication scientifique dans un journal n'a eu lieu pour le moment (publication prévue à l'automne 2025). Les résultats obtenus au cours de l'année ont été présentés deux fois :

- Mahe Cloé, Eric Slezak - Séminaire Institut Interdisciplinaire Intelligence Artificielle (3IA) le 16 mai 2025 - "First robust census of binary supermassive black holes : Deep Learning Techniques applied to the Euclid space mission."
- Mahe Cloé - Colloque des doctorants de deuxième année de l'Ecole Doctorale SFA, le 20 mai 2025 - Poster "First robust census of binary supermassive black holes : Deep Learning Techniques applied to the Euclid space mission."

Description :

Afin de poser des contraintes observationnelles sur la croissance des trous noirs supermassifs au centre des galaxies, il s'agit d'être d'abord en mesure de reconnaître les couples de noyaux actifs attendus dans le cadre du scenario hiérarchique de formation des structures. Très peu de ces couples sont connus aujourd'hui. Notre stratégie consiste donc à construire un algorithme d'identification base sur de l'apprentissage automatique entraîné par conséquent sur des données issues de simulations. Cet échantillon d'apprentissage doit être construit conditionnellement aux paramètres de la paire sur la base des quelques dizaines de réalisations présentes dans la simulation Horizon-AGN, cela par le biais d'un autoencodeur variationnel. La reconnaissance automatique de ces objets dans les images du relevé spatial Euclid sera-t-elle faite au moyen d'un réseau de neurones convolutif, avec là encore une étape d'apprentissage précédée d'une adaptation de domaine. Toutes ces étapes nécessitent de réaliser plusieurs séries d'apprentissage profond sur des milliers ou millions d'exemples. Afin d'être en mesure de pouvoir réaliser les tests préparatoires, le conditionnement des algorithmes et leur exécution dans des temps raisonnables, l'expérience montre la nécessité de pouvoir s'appuyer sur des cartes GPU. C'est la condition sine qua non pour que notre projet soit réalisable.

SAXO plus

Responsable scientifique : Mamadou N'Diaye

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Premiers résultats de simulations pour évaluation de la performance de correction de SAXO+ avec un analyseur de type Zernike.

Description :

SPHERE est une imageur d'exoplanètes sur le Very Large Telescope au Chili et initialement développé par un consortium d'instituts européens. Il s'appuie sur le module SAXO pour corriger les effets de la turbulence atmosphérique sur les images des étoiles observées. Ce module est constitué d'une analyseur de surface d'onde pour mesurer les erreurs induites par l'atmosphère, un miroir déformable pour corriger ces erreurs et un calculateur en temps réel pour convertir les mesures de l'analyseur en commande sur le miroir déformable.

Malgré l'excellente correction observée avec une fréquence de correction de 1kHz, elle s'avère encore insuffisante pour observer les compagnons planétaires les plus proches de leur étoile. Pour étendre l'espace de découverte des exoplanètes, le consortium développe actuellement une jouvence instrumentale avec un module SAXO+, un étage de correction supplémentaire à SAXO et capable de tourner à 3kHz pour réduire les résidus de turbulence atmosphérique.

L'objectif de ce travail est d'évaluer la performance de correction de SAXO+ pour différents cas scientifiques et des conditions d'observations. En particulier, notre groupe cherche à estimer et valider la pertinence de SAXO+ avec un analyseur pour lequel nous disposons d'une expertise unique en France en collaboration avec nos collègues du Laboratoire d'Astrophysique de Marseille.

Ces simulations requièrent l'utilisation de scripts développés au Centre de Recherche en Astrophysique de Lyon (CRAL) et le package COMPASS développé par le LESIA à l'Observatoire de Paris. COMPASS s'appuie sur des GPU NVIDIA pour la réalisation des simulations.

Exo-Andes

Responsable scientifique : Mamadou N'Diaye

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Premiers simulations d'estimation du nombre d'exoplanètes par détection et caractérisation d'exoplanètes avec ELT/ANDES.

Description :

L'ère de la caractérisation des exoplanètes a déjà commencé il y a deux décennies avec la caractérisation atmosphérique des Jupiters chauds et fortement irradiés à l'aide d'observations de transit (en transmission et en émission). De telles observations ont été rapportées pour plus de 30 exoplanètes à ce jour, y compris des Jupiters chauds anciens (0,5-10 Gyr), des Neptunes chauds et même des super-Terres. Plus récemment, des observations avec des spectrographes à haute résolution spectrale ($R \lambda = 100\,000$, ci-après appelée spectroscopie à haute dispersion, HDS) ont montré que les caractéristiques spectrales des atmosphères planétaires peuvent être distinguées des raies telluriques et stellaires en utilisant la variation de la vitesse radiale de l'exoplanète (en transit ou non). L'atténuation des raies moléculaires (en transmission, en émission mais aussi en réflexion) permet aux astronomes de rechercher des signatures moléculaires et de cartographier leur distribution spatiale en utilisant des techniques d'imagerie Doppler. Ces observations ont permis d'étudier la composition, les structures spatiales et la dynamique de l'atmosphère exoplanétaire afin d'explorer les processus d'inversion thermique observés pour les Jupiters chauds, de mélange vertical, de circulation globale et d'évaporation.

Pour les exoplanètes jeunes (1-100 Myr) qui peuvent être résolues spatialement, les techniques d'imagerie à haut contraste (HCl) (optique adaptative extrême, coronographie, spectrographes à champ intégral...) fournissent actuellement des spectres de haute fidélité de Jupiters massifs brillants principalement à une basse résolution spectrale ($R \lambda = 100-1000$). Ils sont typiquement constitués de dizaines à milliers de points de données sur une large gamme de longueurs d'onde (0,5-5 μ m) et peuvent être acquis en quelques heures d'observation. Ils offrent la possibilité d'explorer les propriétés des jeunes Jupiters ainsi que déterminer leur composition et la présence de nuages. Elles donnent finalement une vision très complémentaire des atmosphères

exoplanétaires par rapport aux études de transit et de HDS des exoplanètes plus anciennes.

Dans la perspective de l'Extremely Large Telescope (ELT) <https://elt.eso.org>, l'un des objectifs ultimes est de caractériser l'atmosphère des planètes terrestres tempérées et de détecter les premières biosignatures. La combinaison des techniques HDS et HCI offre une approche très prometteuse. Cette synergie de techniques est actuellement prévue pour divers instruments tels que HARMONI, METIS, ANDES et PCS. Ces installations mettront en œuvre différents niveaux de spécifications instrumentales qui ne répondent pas toujours à toutes les exigences en matière d'imagerie et de caractérisation des planètes terrestres.

Dans ce contexte, un des aspects clé concerne l'étude des performances de détection et de caractérisation de SCAO-IFU, le futur mode spectro-imageur d'ANDES (<https://elt.eso.org/instrument/ANDES/>) à l'ELT qui combinera les techniques innovantes de HCI et HDS. Notre objectif sera d'explorer en particulier les performances de détection du mode HCI pour ANDES, y compris la nécessité éventuelle d'un mode coronographique, de déterminer le nombre d'exoplanètes qui seront détectées en lumière émise et potentiellement réfléchie, et d'étudier les performances de caractérisation en termes de détection des signatures atomiques et moléculaires liées aux caractéristiques de l'atmosphère de l'exoplanète. Cette étude sera réalisée avec le consortium ANDES et en synergie avec d'autres études sur les instruments ELT (par exemple HARMONI, METIS et MICADO) et le JWST, afin d'identifier les principaux cas scientifiques de l'instrument ANDES pour la caractérisation des atmosphères exoplanétaires.

Nos travaux consisteront à réaliser des simulations de bout-en-bout d'observations synthétiques de l'ELT-ANDES, en incluant les résidus de turbulence d'optique adaptative, la coronographie pour l'atténuation du signal stellaire, et le spectrographe à intégral de champ pour l'obtention de spectres d'atmosphères exoplanétaires. Ces jeux de données nécessitent l'utilisation d'un cluster pour mener une large exploration de l'espace des paramètres afin d'étudier différentes conditions d'observations et différentes planètes dans le but de déterminer les capacités de détection et caractérisation d'exoplanètes d'ANDES.

Evolution primitive des familles d'astéroïdes

Responsable scientifique : Paolo Tanga

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Article en journal avec comité de lecture de rang A: Balossi, R., P. Tanga, M. Delbo, A. Cellino, and F. Spoto 2025. An ancient L-type family associated with (460) Scania in the middle main belt as revealed by Gaia DR3 spectra. *Astronomy and Astrophysics* 702, A239. <https://doi.org/10.1051/0004-6361/202555527>

Présentation à conférence Internationale : Balossi, R. and P. Tanga 2025. Post-impact self-gardening: a poorly known mechanism with consequences on the properties of asteroid surfaces. EPSC-DPS Joint Meeting 2025 (EPSC-DPS2025 EPSC-DPS2025-486. <https://doi.org/10.5194/epsc-dps2025-486>

Description :

Les familles d'astéroïdes de la ceinture principale, source d'astéroïdes géocroiseurs (NEA), sont nées de la destruction catastrophique de corps parents préexistants à la suite d'impacts à grande vitesse. Des groupes de fragments à faible vitesse relative se sont ré-accumulés sous l'effet de la gravité pour former les membres observables de la famille. Malgré le consensus autour de ce mécanisme, l'une des plus grandes questions non résolues concerne l'homogénéité spectrale de la famille résultante. Comme les corps parents étaient probablement différenciés à l'intérieur, certaines familles devraient présenter des fragments de composition différente. Cependant, une diversité spectrale n'est soupçonnée que dans très peu de cas.

Nous voulons tester une nouvelle hypothèse que nous appelons 'auto-bombardement', 'précoce' ou 'post- impacte' des surfaces d'astéroïdes. Après un impact générant une famille, les membres se dispersent autour de l'orbite du corps parent. Cependant, ils reviennent à une distribution plus compacte dans l'espace à des intervalles correspondant à leur période orbitale moyenne. Ces retours périodiques à une configuration dense devraient se traduire par une augmentation temporaire des collisions mutuelles. Ce n'est qu'après plusieurs centaines d'orbites que la dispersion (conséquence de la randomisation des angles orbitaux due aux perturbations gravitationnelles) est suffisamment importante pour réduire la probabilité de collision à un niveau négligeable.

Alors que la phase de distribution aléatoire des orbites a été modélisée statistiquement (Dell'Oro et al. 2001), la phase initiale de collisionnalité élevée nécessite des simulations numériques de l'évolution dynamique. Cette exploration n'a jamais été réalisée auparavant et constitue le cœur de la recherche proposée. S'il est confirmé que des fragments de nature différente entrent en collision et se contaminent efficacement, il pourrait en résulter une homogénéisation générale des surfaces des membres de la famille.

We study the early evolution of collisional families by a N-body code optimized for this goal.

GaiaMoon - looking for asteroid satellites in Gaia astrometry

Responsable scientifique : Paolo Tanga

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Presentation à symposium international: Liu, Z., D. Hestroffer, J. Desmars, R. Lallemand, P. David, and L. Liberato 2025. Binary system dynamics and physical property analysis using Gaia astrometry. EPSC-DPS Joint Meeting 2025 (EPSC-DPS2025 EPSC-DPS2025-1274. <https://doi.org/10.5194/epsc-dps2025-1274>

Presentation à symposium international: Liberato, L., D. Mary, P. Tanga, J. Desmars, R. Lallemand, Z. Liu, and D. Hestroffer 2025. Updated method for binary asteroid search in Gaia FPR. EPSC-DPS Joint Meeting 2025 (EPSC-DPS2025 EPSC-DPS2025-814. <https://doi.org/10.5194/epsc-dps2025-814>

Article en journal à comité de lecture: Minker, K., and 22 colleagues 2025. A dynamical dichotomy in large binary asteroids. *Astronomy and Astrophysics* 701, A42. <https://doi.org/10.1051/0004-6361/202554973>

Description :

Les astéroïdes sont considérés comme des vestiges du processus de formation du système solaire. Ils nous renseignent sur l'environnement dans lequel les planètes, y compris notre propre Terre, ont évolué de manière dynamique et ont été façonnés par des phénomènes tels que l'accrétion initiale et la destruction par les collisions mutuelles.

Néanmoins, les propriétés fondamentales telles que la forme, la masse et la densité ne sont connues que pour une petite fraction du million d'astéroïdes connus. Notre projet s'attaque directement à ce problème pour produire un nombre inédit de propriétés d'astéroïdes en découvrant et en caractérisant une nouvelle population d'astéroïdes binaires. Les astéroïdes binaires, de par leur nature, ont la particularité de révéler des détails sur leurs propriétés physiques et dynamiques, qui ne sont pas accessibles à distance pour d'autres objets.

Dans le cadre du projet collaboratif GaiaMoons, nous souhaitons vivement exploiter au mieux la publication du Data Release 3 (13 juin 2022) de la mission Gaia de l'ESA, comprenant l'observation de plus de 150 000 objets du système solaire et du Data Release 4 suivant (2025) contenant toutes les données de la mission nominale (350.000 astéroïdes).

Notre code prend en entrée les positions des astéroïdes obtenues par Gaia (en pratique, les résidus de leur ajustement à une orbite), et recherche des variations périodiques par le biais d'une recherche périodique optimisée. Pour chaque astéroïde, il effectue une série de tests statistiques de Monte-Carlo, qui prennent beaucoup de temps, afin d'évaluer la fiabilité du signal. Cette approche nous a déjà permis de trouver plus de 300 binaires candidates (notre premier article: Liberato et al. 2024, *Astronomy & Astrophysics*, 688, A50). Cet article exploite le DR3 de Gaia (2018). Nous voulons maintenant passer à l'exploitation de la « Focused Product Release » publiée l'année dernière. Cela nous permettra d'explorer un échantillon d'astéroïdes deux fois plus grand, avec une série temporelle beaucoup plus longue (un autre facteur 2 plus long).

Notre approche statistique s'est également beaucoup améliorée et nous attendons une meilleure version de nos découvertes potentielles de satellites. Pour être efficace, nous devons répartir les calculs sur plusieurs cœurs, ce qui nous permet de traiter plusieurs objets en même temps.

Ring Formation Induced by the Presence of a Giant Planet

Responsable scientifique : Elena Lega

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Lega, E., Morbidelli, A., Masset, F. and Béthune W. Exploring the formation of multiple dust-trapping rings in the inner Solar System. A&A, 702, A45 (2025)

<https://doi.org/10.1051/0004-6361/202555284>

Description en pages suivantes.

Research project: Ring Formation Induced by the Presence of a Giant Planet

Rings and gap in low viscosity disks

Observations of protoplanetary disks have revealed a multitude of rings and gaps (ALMA and DSHARP surveys). Planets interact with the gaseous disk by launching spiral density waves which may steepen into shocks, depositing angular momentum and driving the opening of one or more gaps. The radial density distribution will consist of one or more pressure maxima close to such gaps. The pressure maxima can trap dust and form the bright rings that are observed in continuum emission. This process is particularly efficient in low viscosity disks (Bae et al. 2017) although it is very sensitive to the thermodynamics (Ziampras et al. 2023).

It is therefore important to understand the conditions under which a single planet can open multiple rings and how long such rings can survive in a disk. We recall that in the literature it is common to associate the presence of a planet to each gap observed in disks; however Tzouvanou et al. (2023) have recently studied the dynamical stability of systems hosting multiple planets at the location of the observed gaps and concluded that not every gap can host a planet.

In the specific case of our Solar System we know that Non Carbonaceous chondrules and planetesimals of the inner disk have formed when the disk was 2 Million years old. Therefore, at that time there was millimeter dust and such dust could not come by radial drift from the outer disk because the isotopic signature of chondrules and planetesimals would have been in agreement with that of the outer disk (Carbonaceous Chondrites). Therefore we need to find a mechanism to keep the dust for a few Million years in the inner disk. The presence of rings offer a nice mechanism, and the presence of a giant planet of the mass of Jupiter explains the net separation between the inner and the outer disk.

We aim at studying rings formed by a single Jupiter-mass planet and their dependence on the thermodynamics properties, in particular on the cooling time T_{cooling} , on both pure hydrodynamic and MHD models. For the MHD model we have promising preliminary results in the case of a short cooling time $T_{\text{cooling}} = 1/\Omega_K$ (Fig.1). We will continue the above simulation until the gap is stationary and repeat this exploration for $T_{\text{cooling}} = [10,100]/\Omega_K$. We will do the same for a pure hydro-dynamical model for the same set of cooling times.

In the pure hydro-dynamical case we will also study smaller masses from Saturn down to super-Earth (20 Earth masses) as an interesting study for exo-planetary systems.

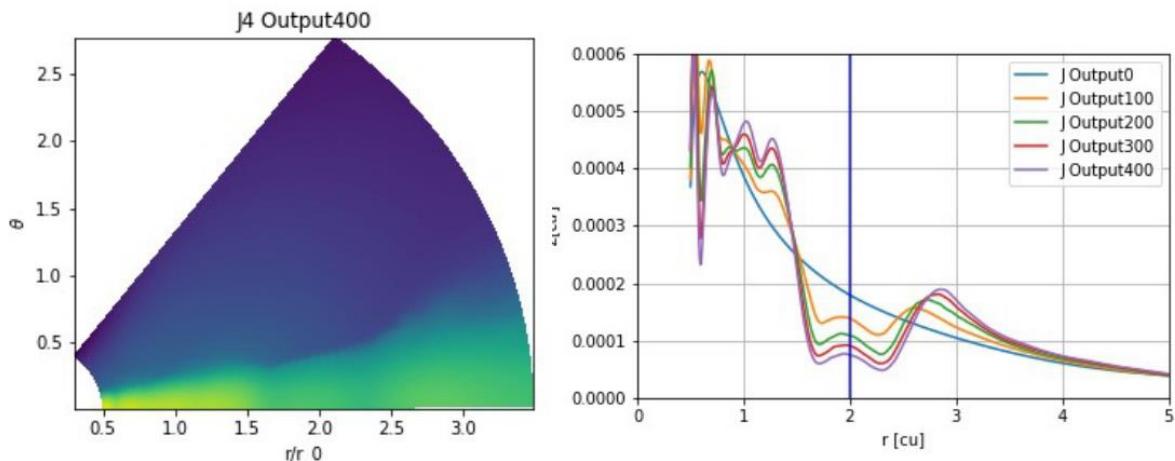


Fig.1 MHD simulation in a disk dominated by Ohmic diffusion with embedded a single giant planet in a prescription with cooling time $T_{\text{cooling}} = 1/\Omega_K$

Before the production stage which will take place on national computer centers we would like to run low resolution cases on Azzurra. These cases will be part of a MASTER project for 2 students of the local master MAUCA.

Resolution	Code	GPU	Estimated time	Cooling time
800x384x60	Fargo3D (MHD)	4	2000h	1, 10,100
800x384x60	FargOCA (HD)	1	2000h	1,10,100 (Jupiter, Saturn, super-Earth)

Table 1 Parameters and estimated computational time

This project is part of a global numerical project funded by the ERC “A holistic approach to understand Earth formation” (P.I. A. Morbidelli) and will be done in collaboration with A. Morbidelli and F. Masset. On this project we will also have 2 students of the local master MAUCA.

Numerical Method :

1) The FARGOCA and FARGO3D codes

The code FARGOCA (FARGO with Colatitude Added) (Lega et al. 2014) and Fargo3D codes (Benitez-Llambay and Masset 2016) are both based on the FARGO code (Masset 2000) extended to 3 dimensions with respectively the additional introduction of an energy equation to provide a realistic modeling of radiative effects (Kley et al. 2009) and the introduction of dust dynamics as well as of magnetic field treatment including non ideal effects (Stone and Norman 1992b).

The protoplanetary disk is treated as a three dimensional non self-gravitating gas whose motion is described by the Navier-Stokes equations.

We use spherical coordinates (r, θ, ϕ) where r is the radial distance from the origin, θ the polar angle measured from the z-axis and ϕ is the azimuthal coordinate starting from the x-axis. The midplane of the disk is at the equator : $\theta = \pi/2$ and the origin of the coordinates is centered on the star (non inertial reference frame). The problem is periodic in azimuth, different possibilities are modeled for the treatment of the boundary conditions in r and θ . The simulation domain is either a full annulus of the protoplanetary disk or a sector of azimuthal width $\Delta\phi < 2\pi$ (see Szulagyi et al. 2016). Modeling a sector allows to study the accretion of gaseous envelopes on planets held on fixed orbits with resolution comparable to the one obtainable with nested meshes. Suited boundary conditions in azimuth are implemented for the sector case. High resolution can be also achieved on both the full annulus or the sector using a nonuniform grid geometry (Fung et al. 2017, Lambrechts and Lega 2017).

We work in a coordinate system which can be fixed or rotating at an arbitrary frequency or co-rotating with one planet. The gravitational influence of the planets on the disk is modeled through their gravitational potential.

2) Implementation

The Navier-Stokes fluid equations are solved using finite-differences with a time-explicit-implicit multistep procedure on a staggered mesh.

Precisely, two main steps are introduced called « source » and « transport ».

In the source step velocities are updated according to the pressure, the potential and the viscosity (see Tassoul 1978 for the viscous stress tensor). The updated velocities are used in the transport step where fluid advection is computed using its integral expression according to Van Leer (1977) method. Advection is computed in the standard method or using the FARGO algorithm (« Fast Advection in Rotating Gaseous Objects », Masset 2000). The FARGO algorithm allows to use a computation timestep larger than the one provided by the usual CFL (Stone et Norman 1992a) condition. The gain in computational time is of about an order of magnitude in case of a Jupiter planet embedded in a standard protoplanetary disk.

As the frame centered on the star is not inertial, indirect forces are added to account for the acceleration of the star around the center of mass of the system ; a special attention has been recently paid to which acceleration should be applied to which part of the system (Crida et al. 2023, in preparation).

Regarding the energy equation (Mihalas, D. and Weibel-Mihalas, B. 1984), we first update the energy by explicit integration of the compressional heating term and, in a separate step, we integrate the viscous heating and the radiative diffusion terms. In this second step we follow the backward Euler

method, which is an unconditionally stable implicit method solved with a standard Successive Overrelaxation Reduction solver. Energy advection is the last step and it follows the same procedure introduced for the density and the velocity. The code can also take into account the energy irradiated by the star, the energy equation is written in this case following a two-temperature approach (Bitsch et al. 2013, Bitsch et al. 2014,a, Bitsch et al. 2014,b, Lega et al. 2015).

Planets are massive points, their orbits is determined with a fifth order Runge-Kutta integrator. Planets feel the gravitational influence of the star, of the disk as well as their mutual gravitational interaction. The planetary influence on the disk is obtained by computing the gravitational potential modeled with a cubic function suited to 3D disks (Kley et al. 2009).

MHD equations are solved in fargo3D according to the prescription in Stone and Norman 1992b, namely by using a method of constrained transport to ensure the divergence free for the magnetic field. The FARGOCA code implements hybrid parallelism where each MPI process uses OpenMP or GPU to take advantage of multi-core architecture. Communication between MPI processes occurs once per timestep and is applied to 5 ghostcells.

The first version of the code, has been running on ADA (IDRIS) than on Occigen (CINES) since 2014 ; it was written in C and didn't require external libraries. The code has been constantly modified since 2013 mainly by adding new physical modeling or by modifying the computational grid or the computational domain in order to gain resolution. We have recently re-factorized the code (mainly done by A. Miniussi and E. Lega) in order to improve code maintainability, reliability and performance as a small user community is growing around the code. We have had the first FARGOCA user meeting in October 2018, and the second meeting in May 2020 on zoom platform (first period of confinement in 2020).

The new version of the code is written in C++, it depends on a C++ compiler with OpenMP support, Boost libraries, MPI environment and rely on Cmake for installation and unit testing. This version of the code has been regularly used for production on Adastra (CINES) as well as Jean-Zay (IDRIS).

Development, testing and low resolution runs are regularly done on our local *mésocentre de calcul SIGAMM* which is currently not working and will not be accessible for a long period (3 months or more).

3) Adapting the code to GPU

3.1 FargOCA code

We have started in October 2021 to introduce the Kokkos library in order to adapt our code to hybrid CPU/GPU platform. After participation to the IDRIS-GPU hackathon in May 2022 we have started using this new version of the code for production purposes on Jean-Zay V100 and on 2024 on Adastra AMD MI250 accelerators.

3.2 Fargo3D code

The fargo3D code is written in C and does not requires any external library. Introducing a set of very simple rules, the authors were developed a python script that translates the C routines into CUDA kernels. Therefore, the code implements hybrid parallelism where each MPI process uses 1 NVIDIA GPU.

4) Bibliography

- Bae J., Zhu Z., Hartmann L., 2017, ApJ, 850, 201. doi:10.3847/1538-4357/aa9705
Benítez-Llambay P., Masset F.~S., 2016, ApJS, 223, 11. doi:10.3847/0067-0049/223/1/11
Bitsch B., Crida A., Morbidelli A., Kley W., Dobbs-Dixon I., A&A, 549, A124, 2013.
Bitsch, B., Morbidelli A., Lega, E., Crida, A., A&A 564, A135, 2014a.
Bitsch, B., Morbidelli A., Lega, E., Kretke, K. Crida, A., A&A, 570, A75, 2014b.
Kley, W., Bitsch, B. and Klahr H. , A&A, 506, 971-997, 2009.
Lambrechts M. et Lega, E., A&A 606, A146, 2017
Lega, E., Crida, A., Bitsch, B., and Morbidelli, A., MNRAS, 440, 683, 2014.
Lega, E., Morbidelli, A., Bitsch, B., Crida, A. and Szulagyi, J., MNRAS 452, 1717–1726 ,2015.
Masset, F.S. Astrophys. Suppl. Series, 141, 165-173, 2000.
Mihalas, D. and Weibel-Mihalas, B. Dover Publications, 1984.
Stone J.M., Norman M.~L., 1992a, ApJS, 80, 753. doi:10.1086/191680
Stone J.M., Norman M.~L., 1992b, ApJS, 80, 791. doi:10.1086/191681
Tassoul J.L., Theory of rotating stars, Princeton University Press, (1978).
Tzouvanou A., Bitsch B., Pichierri G., 2023, A&A, 677, A82. doi:10.1051/0004-6361/202347264
Van Leer, B., J. of Computational Physics, 23, 276, (1977).
Ziampras A., Nelson R.~P., Rafikov R.~R., 2023, MNRAS, 524, 3930. doi:10.1093/mnras/stad1973

Secular evolution of planetary rings

Responsable scientifique : Patrick Michel

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Nous avons compilé et testé avec succès notre code GPU. Maintenant que le cluster est de nouveau opérationnel, nous prévoyons d'utiliser les GPU mis à niveau pour réaliser des simulations de perturbations catastrophiques d'astéroïdes afin de reproduire des familles d'astéroïdes et des satellites, en soutien à la mission Lucy de la NASA vers les astéroïdes troyens de Jupiter dont nous sommes membres. Par ailleurs, nous avons presque terminé certaines simulations modélisant Phobos, la lune la plus interne de Mars. Cet article sera soumis prochainement, mais sa publication pourrait ne pas avoir lieu avant 2026.

Description en pages suivantes.

The radiative effects on the planetary ring

— why inner planets barely possess dusty rings

Motivations: Studying planetary rings provides insights into fundamental processes such as planetary formation, dynamics, and evolution. As natural laboratories, rings allow us to observe gravitational interactions, collision dynamics, and the formation of moons. They also serve as indicators of environmental conditions around their host planets, which is crucial for understanding planetary history and potential habitability in the Solar System.

All outer planets beyond the asteroid belt have rings. Saturn rings are the most massive in the Solar system with a total estimated mass of about 10^{-8} of the Saturn mass. The rings of Jupiter, Uranus and Neptune are more tenuous with the mass of $\sim 10^{-15}$, 10^{-12} and 10^{-14} of their host planets.

However, none of the inner Solar System planets have rings, and the reasons for this remain an open question.

Problem Statement: The absence of rings around inner planets could be due to two factors: either rings cannot form in these environments, or they only exist for very short periods. This leads us to explore the origins and evolution of planetary rings.

The origins of rings can vary significantly across different planetary systems. Proposed mechanisms include: (1) primordial rings formed during the planet's accretion process; (2) tidal fragments of moons or external bodies; and (3) collisional fragments from the planet or its moons. The tenuous rings of Jupiter, Uranus, and Neptune are likely the result of erosion or destruction of moons by meteoroid bombardment. The origin of Saturn's massive rings is more complex and could involve any of the aforementioned scenarios. In fact, these potential origins also apply to inner planets. For instance, Mars has moons like Phobos, which is near the Roche limit and could potentially generate rings through tidal disruption or collisions. The proximity of Mars and Earth to the asteroid belt increases the likelihood of high-eccentricity asteroids causing tidal disruptions or impacts, potentially leading to ring formation. Similar collisions could occur on Venus and Mercury. Therefore, the conditions for ring formation around inner planets can sometimes be met, suggesting that the mere possibility of ring formation does not explain the absence of rings in the inner Solar System.

The evolution of planetary ring systems is primarily governed by viscous evolution, interactions with satellites, and external forces. We are particularly interested in the lifetime of rings under these influences. Viscous evolution involves the exchange of angular momentum between the faster rotating inner regions and the slower outer regions, causing the rings to spread out and dissipate. Gravitational interactions with moons shape ring structures and can cause moons to migrate outward due to back-reaction forces. These mechanisms are similar for both inner and outer planets. External forces, such as magnetic and Poynting-Robertson (P-R) drag, are more significant for inner planets but typically only affect very small particles (e.g., micron-sized). **Consequently, the current theoretical framework does not provide a definitive explanation for the absence of rings around inner planets.**

Proposed Solution: To address the absence of rings around inner planets, we need to explore new mechanisms that behave differently for inner planets compared to outer ones. We propose that **the Yarkovsky-Schach-Zhou (YSZ) effect**, a thermal effect recently discovered in binary asteroids (Zhou et al., 2024), could provide insights into this problem.

The YSZ effect causes particles experiencing periodic eclipses from the planet to migrate towards a synchronous orbit, where their spin period matches their orbital period. Mutual collisions among particles can disrupt synchronous rotation, activating the YSZ effect. The timescale for this effect is given by Zhou et al (2024):

$$\tau_{\text{YSZ}} \simeq 40 \text{ kyr} \left(\frac{r_p}{5000 \text{ km}} \right) \left(\frac{r_s}{1 \text{ cm}} \right) \left(\frac{a/r_p}{2.5} \right)^{1/2} \left(\frac{0.005}{f_{Y,0}} \right) \left(\frac{a_h}{1 \text{ au}} \right)^2. \quad (1)$$

Here r_p is the planet radius, r_s is the particle radius, a is the semi-major axis of the particle, $f_{Y,0}$ is the Yarkovsky coefficient and a_h is the heliocentric semi-major axis. This timescale is considerably short compared to the lifetime of Saturn's ring (e.g. ~ a few Gyrs, Salmon et al 2010). For approximation, the timescale for viscous evolution is given by Crida & Charnel (2012):

$$\tau_{\text{vis}} \sim 3 \text{ Gyr} \left(\frac{a/r_p}{2.5} \right)^{3/2} \left(\frac{M_p/M_{\text{disk}}}{10^7} \right)^2 \quad (2)$$

without considering the interaction with moons or other effects.

A direct comparison between Eqs. (1) and (2) already indicates that YSZ effect could play a very important role during the lifetime of an ordinary planetary ring. However, we should not simply conclude that YSZ effect deplete the ring on the timescale of Eq. (1), since a few factors should be considered in details, including the optical depth that affects the illumination of particles, the rotation states modified by frequent mutual orbit, the obliquity of the ring that affects the shadow region, and so on. To better understand this effect on the planetary rings, my future plan is scheduled as shown in Fig. 1 and listed as below:

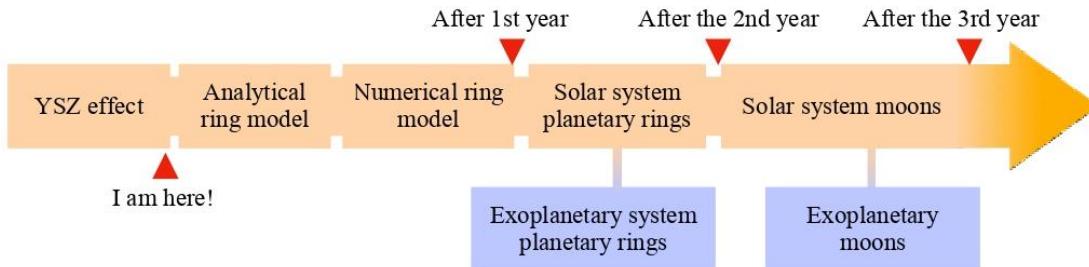


Fig. 1. Flow chart of my research plan.

1. A novel planetary ring dynamics theory involving the YSZ effect

Analytical Model Development. Derive equations of motion for ring particles that incorporate the Yarkovsky effect along with gravitational and non-gravitational forces. The model will include parameters such as particle size, spin rate, thermal properties, and solar incidence angle.

Numerical Simulations. Develop a numerical simulation tool to model the long-term orbital evolution of ring particles under the combined influence of gravitational forces and the YSZ effect. Use N-body simulations to study short-term particle interactions and refine the long-term orbital evolution model. The computer cluster will be used for the simulation to investigate the spin distribution of ring particles.

2. Application to planetary rings in Solar system and exo-planetary systems

Solar system. Apply the model to study the lifetime of planetary rings around inner planets. By combining this analysis with the collisional history of small bodies, we may answer why inner planets rarely possess dusty rings. In addition, we may also constrain the dynamical environment of the inner Solar system during history.

Exoplanetary system. Some exoplanets are suspected to have ring systems based on their light curves (Teachey, 2024). Applying our model to these systems will provide valuable insights into planetary evolution and the structure of planets and small bodies in the exoplanetary systems .

3. Application to origin of moons

Planetary moons. The formation of planetary moons are thought to be influenced significantly by planetary disks, with the size and spatial distribution of moons dependent on the disk's lifetime (Crida & Charnoz, 2012). Our study on dusty planetary rings will contribute to understanding formation and evolution of planetary moons.

Space mission on Phobos. Specifically, we will investigate the ring system around Mars, which may be related to the origin of Phobos, a Martian moon targeted by JAXA's Martian Moons eXploration (MMX) mission. This research could generate significant interest in planetary science.

References

- Crida, A., & Charnoz, S. (2012). Formation of regular satellites from ancient massive rings in the solar system. *Science*, 338(6111), 1196-1199.
- Salmon, J., Charnoz, S., Crida, A., & Brahic, A. (2010). Long-term and large-scale viscous evolution of dense planetary rings. *Icarus*, 209(2), 771-785.
- Teachey, A. (2024). Detecting and Characterizing Exomoons and Exorings (Handbook of Exoplanets. *arXiv preprint arXiv:2401.13293*.
- Zhou, W. H., Vokrouhlický, D., Kanamaru, M., Agrusa, H., Pravec, P., Delbo, M., & Michel, P. (2024). The Yarkovsky effect on the long-term evolution of binary asteroids. *The Astrophysical Journal Letters*, 968(1), L3.

The Formation of Small Satellites in the Solar System

Responsable scientifique : Patrick Michel

Laboratoire : Lagrange

Description :

We propose two associated projects related to the formation of small satellites in the Solar System. The first is on the formation of small binary asteroids, such as Dimorphos, the smaller component of the Didymos binary asteroid and target of NASA's DART and ESA's Hera missions. The second project is on the formation of Phobos and Deimos, Mars's two asteroid-like moons from a giant impact, and on the interaction of the wheel of a rover on Phobos's regolith.

Based on images taken by the DART spacecraft, it was determined that Dimorphos is likely a gravitational aggregate (also called rubble pile), dominated by meter-sized boulders with a flattened oblate spheroid-like shape. We aim to explain the shape of Dimorphos via gravitational reaccumulation of loose material shed from the equator of the fast-rotating primary. This material accumulates into satellites at a close distance from the primary, where tidal and centrifugal forces are strong enough to "flatten" the body. While there is an extensive body of literature on binary asteroid formation, there are zero studies that self-consistently model the coupled dynamical and shape evolution of the secondary, making this the first every study to reproduce the shape of Dimorphos, with important implications on the understanding of asteroid binaries and for the development, operations and data interpretation of the ESA Hera mission that will visit the system in late 2026.

The second project will constrain the formation of Phobos and Deimos to inform JAXA's MMX mission which will return a sample from Phobos to determine whether the Martian moons are captured asteroids or remnants from a giant impact. In the giant impact scenario, there are two issues with existing studies. First, most simulations neglect Mars's material strength which can have a strong effect on mass ejected into a circum-Mars disk. Second, most studies model the gravitational accumulation of the Martian moons using analytic methods or simplified numerical models due to the significant computational costs. We plan to improve both issues, possibly leading to a major breakthrough. First, we

will perform smoothed particle hydrodynamics (SPH) simulations that include Mars's material strength to quantify the importance of this parameter.

Second, we will use a newly improved N-body code to directly simulate the gravitational accumulation of satellites from the debris disk. We will also support the development and data interpretation of the rover onboard MMX by modeling the wheel's interaction with Phobos' surface.

Study of the circumstellar environments of B[e] star using ray-tracing techniques

Responsable scientifique : Gilles Niccolini

Laboratoire : Lagrange

Résultats obtenus grâce aux moyens informatiques attribués :

Article sous le point d'être soumis : « The circumstellar environment of FS CMa. Modelling spectroscopic, VLTI-PIONER and MATISSE data with ray-tracing », J. Perdigon et al.

Description du projet en pages suivantes.

Project description

This project is conducted within the framework of the MASSIF¹ ANR, a collaboration between the Observatoire de la Côte d’Azur, the Université de Toulouse, and the Observatoire de Paris-Meudon. The group aims to enhance our understanding of massive stars by coupling state-of-the-art two-dimensional simulations of stellar interiors, including rotation (Toulouse), pulsations (Paris-Meudon), and mass loss, with radiative transfer models of their atmospheres and environments (Nice).

The study aims to characterise the circumstellar environment of the B[e] star FS CMa (HD 45677). The classification "B[e]" originates from Allen et al. (1976) and designates B-type stars that show forbidden emission lines in their spectrum, as well as a significant near- or mid-infrared excess due to circumstellar material. In reality, this spectral classification encompasses stars that are highly heterogeneous, with very different types and evolutionary stages. FS CMa is a prototype star for B[e]-type stars (Miroshnichenko, 2007). It is thus an important star to study, as it sets the standard physical/observational criteria for this specific group.

The large number of spectroscopic and interferometric observations now available allows us to study the close surroundings of FS CMa. We performed a multi-wavelength study using interferometric data (squared visibilities and closure phases) from the VLTI-PIONIER (Le Bouquin et al., 2011), GRAVITY (GRAVITY Collaboration et al., 2017), and MATISSE (Lopez et al., 2022) instruments (H, K, and L-N bands, respectively), along with a wide range of spectroscopic data spanning from 0.3 to 100, μm . From previous studies, we suspect that the star possesses, in addition to a disc of dust, a compact gaseous disc around the central star (Hofmann et al., 2022), which this study aims to characterise.

To model the circumstellar environment, we used an improved version of the ray-tracing module developed in Perdigon et al. (2024). The tool is a stand-alone Python package, publicly available at ². A short description of how the code works is given in Fig. 1.

The model consists of a two-component disc: a gas disc in the very-near environment of the central star, and a dust disc separated from the central regions by a free cavity. We computed images using the ray-tracing model presented above. We employed a Bayesian approach coupled with Markov Chain Monte Carlo (MCMC) methods (Foreman-Mackey et al., 2013) to statistically characterise the parameters of the model. Some preliminary results are shown in Figs. 2 and 3. These results are very promising (paper in preparation), and we aim to confirm them by performing additional CPU-intensive simulations that will benefit from the use of the AZURRA facilities.

References

- Allen, D. A. and J. P. Swings (Mar. 1976). “The spectra of peculiar Be star with infrared excesses.” In: *Astron. Astrophys.* 47, pp. 293–302.
- Foreman-Mackey, Daniel et al. (Mar. 2013). “emcee: The MCMC Hammer”. In: *PASP* 125.925, p. 306. DOI: [10.1086/670067](https://doi.org/10.1086/670067). arXiv: [1202.3665 \[astro-ph.IM\]](https://arxiv.org/abs/1202.3665).
- GRAVITY Collaboration et al. (June 2017). “First light for GRAVITY: Phase referencing optical interferometry for the Very Large Telescope Interferometer”. In: *Astron. Astrophys.* 602, A94, A94. DOI: [10.1051/0004-6361/201730838](https://doi.org/10.1051/0004-6361/201730838). arXiv: [1705.02345 \[astro-ph.IM\]](https://arxiv.org/abs/1705.02345).

¹<https://www.anr-massif.fr/>

²<https://gitlab.oca.eu/jperdigon/raytracingimage>

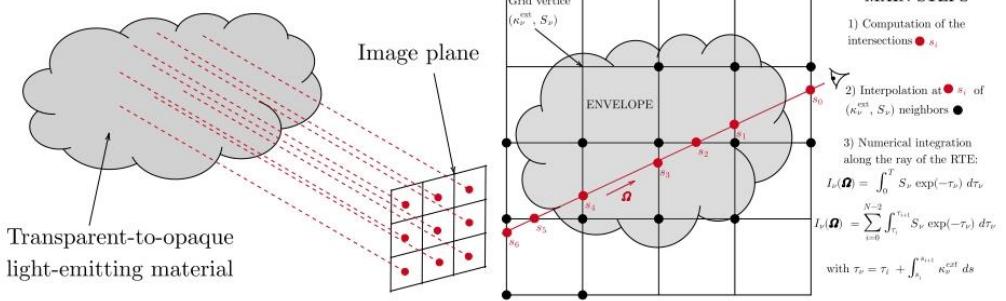


Figure 1: Left: Principle of ray tracing. An image plane is divided into pixels, where the intensity is evaluated along rays intersecting the material. Right: Illustration of a particular ray. The ray intersects the numerical grid, where the properties of the medium are defined at the vertices. The values at the intersections (red dots) are interpolated from the adjacent grid vertices (black dots). The intensity I_{ν} is subsequently computed by integrating the radiative transfer equation along the ray.

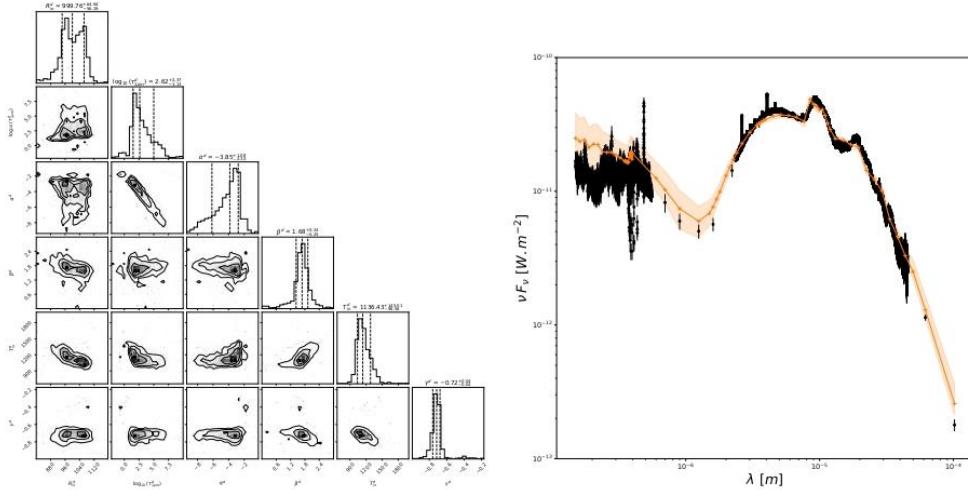


Figure 2: Left: Posterior probability distribution for the parameters of the dust disc. Note that the model has 14 parameters in total, which also include the gas disc parameters. Right: Spectral energy distribution of the data (black dots) and of the posterior distribution (orange-shaded area). The shaded area delimits the 0.16 and 0.84 percentiles of the posterior distribution samples.

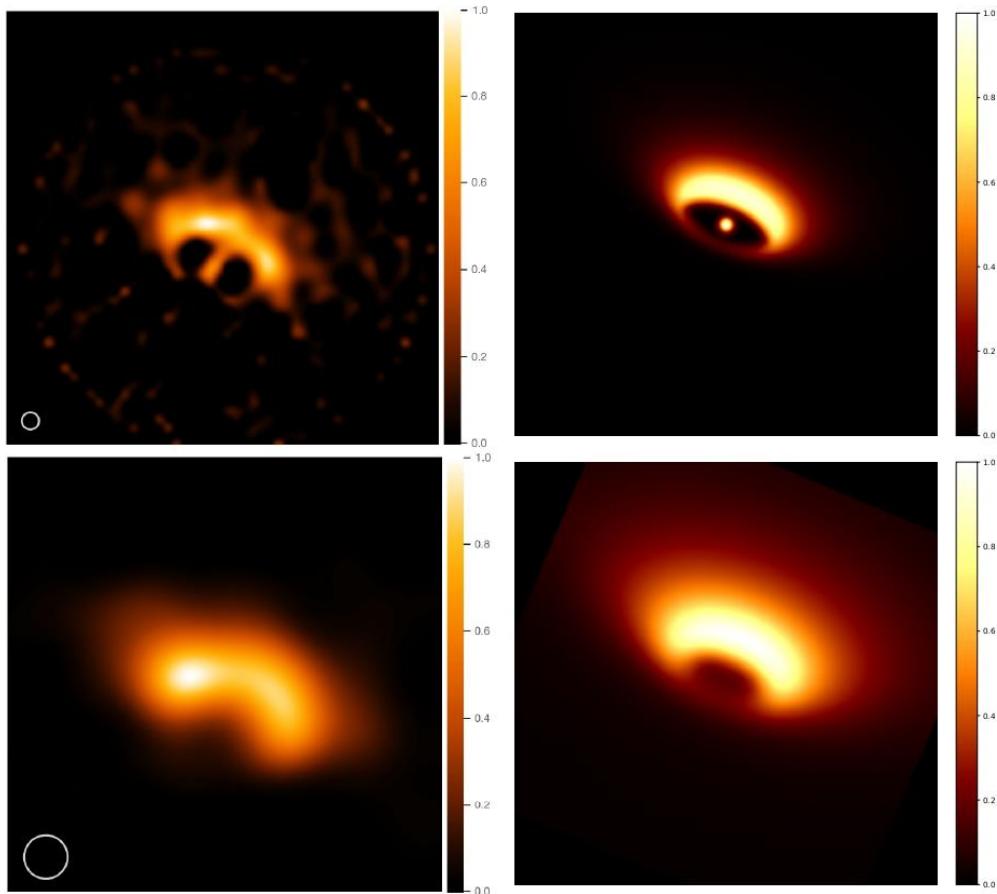


Figure 3: L- ($3.4 - 3.8 \mu\text{m}$) and N- ($8.6 - 9 \mu\text{m}$) band images (top and bottom, respectively) of the inner regions ($64 \times 64 \text{ mas}^2$) of FS CMa. Left: Reconstructed images from Hofmann et al. (2022). Right: Images of the model computed with the "best-fit" parameters from Fig. 2.

-
- Hofmann, K. -H. et al. (Feb. 2022). “VLTI-MATISSE L- and N-band aperture-synthesis imaging of the unclassified B[e] star FS Canis Majoris”. In: *Astron. Astrophys.* 658, A81, A81. doi: [10.1051/0004-6361/202141601](https://doi.org/10.1051/0004-6361/202141601). arXiv: [2111.12458 \[astro-ph.SR\]](https://arxiv.org/abs/2111.12458).
- Le Bouquin, J. -B. et al. (Nov. 2011). “PIONIER: a 4-telescope visitor instrument at VLTI”. In: *Astron. Astrophys.* 535, A67, A67. doi: [10.1051/0004-6361/201117586](https://doi.org/10.1051/0004-6361/201117586). arXiv: [1109.1918 \[astro-ph.IM\]](https://arxiv.org/abs/1109.1918).
- Lopez, B. et al. (Mar. 2022). “MATISSE, the VLTI mid-infrared imaging spectro-interferometer”. In: *Astron. Astrophys.* 659, A192, A192. doi: [10.1051/0004-6361/202141785](https://doi.org/10.1051/0004-6361/202141785). arXiv: [2110.15556 \[astro-ph.IM\]](https://arxiv.org/abs/2110.15556).
- Miroshnichenko, A. S. (Sept. 2007). “Toward Understanding the B[e] Phenomenon. I. Definition of the Galactic FS CMa Stars”. In: *ApJ* 667.1, pp. 497–504. doi: [10.1086/520798](https://doi.org/10.1086/520798).
- Perdigon, J., G. Niccolini, and M. Faurobert (Oct. 2024). “Discontinuous Galerkin finite element method for the continuum radiative transfer problem inside axis-symmetric circumstellar envelopes”. In: *Astron. Astrophys.* 690, A192, A192. doi: [10.1051/0004-6361/202244322](https://doi.org/10.1051/0004-6361/202244322). arXiv: [2408.14934 \[astro-ph.SR\]](https://arxiv.org/abs/2408.14934).

Simulations numériques pour l'entraînement d'une nouvelle méthode d'identification de familles d'astéroïdes

Responsable scientifique : Benoit Carry

Laboratoire : Lagrange

Description :

Contexte

Les astéroïdes entre Mars et Jupiter, dits de la ceinture principale, sont les vestiges des planétésimaux qui se sont accrétés pour former les planètes. Leurs distributions orbitale, en taille, et compositionnelle portent la trace des différentes étapes de la formation de notre système solaire, de l'accrétion des solides dans le disque aux migrations planétaires (DeMeo & Carry 2014).

Néanmoins, 4 milliard d'années d'évolution dynamique et collisionnelle ont imposé leur marque, brouillant ces informations. En particulier, les astéroïdes se détruisent mutuellement lors de collisions catastrophiques, éjectant de nombreux fragments. On appelle les groupes formés par ces fragments les familles d'astéroïdes. Identifier correctement ces familles est critique pour remonter le temps et déterminer la distribution originelle des astéroïdes dans la ceinture principale.

Deux méthodes sont actuellement utilisées dans la communauté pour classer les astéroïdes au sein des familles. Plus de 200 familles ont été identifiées, par ces méthodes, permettant d'étudier de nombreux aspects de l'évolution des astéroïdes. Toutefois, ces deux méthodes ont plusieurs limites Dermott et al. (2018). Depuis Mars 2025, nous, Vincent Vandewalle (UniCA/INRIA) et Benoit Carry (UniCA/OCA/Lagrange), co-encadrerons le doctorat de Maya Guy dont le but est la mise au point d'une méthode moderne de clustering, un modèle de machine learning non supervisé, pour identifier les familles d'astéroïdes. Ce doctorat, à l'intersection de deux disciplines : astrophysique et statistiques, fait suite à un stage de M2 de Maya Guy de six mois en 2024.

Besoins de simulation numérique sur Azzurra

Nous avons mis au point un algorithme de clustering probabiliste (dans une approche basée sur modèles, Bouveyron et al. 2019), dont les premiers résultats sont très

prometteurs et que Maya Guy a déjà présenté en conférence internationale (Guy et al. 2025).

Avant de publier la méthode ainsi que les premiers résultats sur la ceinture principale d'astéroïdes, il est fondamental de tester la méthode sur une population synthétique simulée d'astéroïdes contenant des familles afin de quantifier son efficacité. Nous avons alors utilisé les simulations réalisées par Deienno et al. (2021), par ailleurs utilisées dans d'autres études (e.g., Marshall-Lee et al. 2026). Néanmoins Maya Guy a identifié plusieurs problèmes fondamentaux de ces simulations, qui les rendent inaptes à l'usage (posant question sur les résultats des études publiées se basant sur elles).

Nous demandons un accès à Azzurra pour réaliser une simulation numérique de l'évolution d'une population d'astéroïdes sur des centaines de millions d'années, afin de construire un ensemble de données synthétiques de référence qui servira de benchmark. Cette simulation, portant sur un grand nombre de particules et un grand nombre de pas temporels, requiert l'utilisation massive du calcul parallèle pour être réalisée en un temps raisonnable, en phase avec le travail de thèse. La réalisation de cette simulation est cruciale pour le doctorat de Maya Guy afin de démontrer la pertinence et l'efficacité de sa méthode auprès de la communauté.

Nous mettrons les données simulée à disposition de la communauté (principe FAIR) pour aider les futures études dans le domaine (ce qui n'était pas le cas des simulations de Deienno et al. (2021) que nous avons du contacter pour obtenir les données, qui étaient loin d'être proprement documentées).

Finalement, dans un deuxième temps, Azzurra sera ensuite utilisé pour l'entraînement du modèle.

Références

Bouveyron, C., Celeux, G., Murphy, T. B., & Raftery, A. E. 2019, Model-based clustering and classification for data science : with applications in R, Vol. 50 (Cambridge University Press)

Deienno, R., Walsh, K. J., & Delbo, M. 2021, Icarus, 357, 114218

DeMeo, F. E. & Carry, B. 2014, Nature, 505, 629

Biologie, santé

EUR Life

Leveraging sex differences to decode the core principles governing brain cell numbers

Responsable scientifique : Bruno Hudry

Laboratoire : IBV

Description :

How are cognitive capacity and brain functions determined? Cognitive capacity relies on neuron (and glial cell) numbers, connections, and axonal conduction velocity. Recent studies on neuron numbers have revealed a variety of neuronal scaling rules (allometric relationship between brain mass and neuron numbers) but have yet to identify unifying principles. These studies have mainly focused on differences between species and have overlooked the potential influence of biological sex. Sexual identity, as a conserved intraspecific variation, could offer insights into the biological rules governing neuronal scaling. Sex differences in brain organization and cerebral functions have been documented. For example, a recent study in chickens described intriguing sex differences in neuronal scaling, with females showing higher neuronal densities. Yet, the mechanisms, functional implications, and underlying reasons for these sex disparities remain unknown.

Another study using a genetically tractable model system, *Drosophila melanogaster* larvae, has recently implemented a new approach allowing precise and automated counting of mature neurons and glial cells in the entire brain. Unexpectedly, the authors have revealed sex differences in number of mature neurons and glial cells. Females have 10% more neurons (900 cells), while males exhibit 4% additional glial cells. This indicates that there is a different ratio of mature neurons over glial cells according to sex, already at this juvenile stage. These findings show distinct cellular scaling rules apply in males and females and that sex differences extend to fundamental features of brain organization. I propose to leverage the sex differences in larval fly brains to understand how brain cell number is determined and how it impacts cerebral activities. This project aims to better understand how biological sex influences cellular scaling in the brain. The study of this question and its molecular mechanisms have been hindered due to various technical issues and limitations. However, we can now overcome these challenges using *Drosophila melanogaster*. The powerful genetic tools available, combined with high-throughput analyses, make this model organism ideal for deciphering the cellular and molecular mechanisms generating different neurons over glial cell ratios. For example, genetic

manipulations allow for conditional and bi-directional cell-type-specific sex reversals. The existing literature questions how biological sex can influence cellular scaling in the brain. This proposal will integrate diverse experimental approaches and levels of analysis to address the following objectives:

- (1) Determine the cell types on which sex acts to control brain cells number
- (2) Establish the critical period when the sex determination cascade regulates cellular scaling in the brain
- (3) Decipher the molecular mechanisms underlying brain cellular scaling according to sex and investigate its functional significance for brain functions

Evolutionary studies have assessed neuron numbers across species and discovered that different cellular scaling rules apply in the brain. Most of evolutionary studies on brain cellular scaling have focused on interspecies differences, whereas a few intraspecies studies have reported opposing findings among mice and reptiles. However, most have disregarded the potential effect of biological sex. This project aims to fill this void in our current knowledge and explain how the number of brain cell subtypes is determined, whether biological sex is involved and the mechanisms underlying such control. I propose to use biological sex as a conserved intraspecies variation to study cellular scaling in the brain with an unprecedented level of refinement. The use of *Drosophila melanogaster* will allow the precise characterization of cellular scaling mechanism in the brain and further develop the field of neurodevelopment by demonstrating how cellular sex orchestrates brain cell number homeostasis. The powerful genetic tools available in *Drosophila* will allow me to perform conditional and bidirectional sex reversal experiments in specific brain cell types and understand how different cell type ratios can give rise to functional brains in males and females.

Transcriptomics has been used for three decades as a powerful tool to analyze gene expression among dissected tissues or whole organisms, therefore providing information on gene activity and, ultimately, on its function. In this project, I will use the newly developed enhanced crosslinking combined immunoprecipitation and sequencing (eCLIP-seq), to identify new targets downstream of the sex chromosomes. This new technique will allow me to identify with an unsurpassed level of sophistication, new molecular actors of the sexual differentiation cascade active in the brain.

Overall, this research has the potential to significantly advance our understanding of how biological sex influences brain cell numbers by uncovering new genes, pathways, and concepts shaping brain sexual differentiation.

Objective 1. Where does the sex act to control brain cell numbers?

Drosophila females are 30% larger but display 10% more neurons than males⁷. This suggests that neuronal scaling is not linear to body size dimorphism. I'm interested in understanding how biological sex determines this neuronal scaling to produce a functional brain in males and females. To answer this question, I will conduct bi-directional cell-type specific sex reversals and measure cell numbers in intact, whole brains using a light sheet microscope available in the host institute. I will use genetic reporters consisting of cell-type-specific Gal4 combined with nuclear fluorescent reporters to aid in nucleus counting. This will allow for high throughput and precise quantitation with cellular resolution. The Hudry lab has pioneered work on cellular sex, and many genetic tools enabling sex reversals are already established , .

First, I will confirm the sex differences in total neurons and glial cells number, using the pan-neural nSynaptobrevin-Gal4 and the pan-glial Repo-Gal4. nSyb encodes an essential presynaptic vSNARE factor, while Repo is a transcription factor with glial-specific expression. Additionally, I will investigate whether specific cell subtypes are differentially affected by biological sex using Gal4 lines expressed in dopaminergic, glutamatergic, or cholinergic neurons. Furthermore, I will quantify glial subtypes by examining cortex glia, astrocytes, and unsheathing glial cells. These experiments will provide insights into the scaling rules of all brain cell populations between males and females. Next, I will manipulate the sexual identity of all mature neurons/glial cells by altering the expression levels of female-specific splicing factors called Sex-lethal (Sxl) and Transformer (TraF) followed by cell counting to analyze the effects. These two splicing factors, which are only expressed in XX cells, have been found to regulate all the observed sex differences downstream of the sex chromosomes in flies (XX versus XY). In addition, thanks to the homology-assisted CRISPR knock-in (HACK) method , I will substitute the Gal4 sequence with the Sxl/TraF cDNA for specific neuronal and glial cell subtype Gal4 lines (stated above) to feminize particular brain cell types. Many such lines are already available in Hudry's lab. These experiments will determine whether sex acts in all or specific brain cells to control sex differences in brain cellular scaling.

Therefore, those experiments will provide a brain-wide map of cells where the sex acts to drive sex differences in brain cell numbers, which, to our knowledge, has not been done in any animal.

Objective 2. When is the sex determination cascade regulating sex-specific cellular scaling in the brain?

In the following phases, I will concentrate on one or two types of cells that exhibit more notable differences between the sexes to thoroughly understand the molecular mechanisms driving sexual dimorphism in cell numbers. I will determine the critical developmental period when the sex differentiation cascade influences sex-specific neuronal scaling by employing conditional genetic manipulations. Initially, I will use the cis-allele-switching strategy to express thermo-sensitive (TS) knock-in alleles of the cofactor of the sex determinant, Transformer 2[TS], specifically in the subpopulations of interest. In addition, I will reduce the activity of the sex determinants at various developmental stages (embryonic, larval stage L1, L2, L3) using RNAi lines under the control of specific Gal4s combined with a temperature-sensitive Gal80 line (Gal80[TS]) and count the cells number in the selected subpopulations. Notably, Gal80 functions as a repressor of Gal4 in both yeast and Drosophila. Besides, I will examine whether these sex- and cell-type-specific differences in cell numbers are present and sustained in the adult brain.

Objective 3. What is the molecular mechanism underlying sex-specific neuronal scaling and its functional significance?

With a defined cellular context and developmental window, I will delve into the molecular events downstream of the sex determinants leading to brain cells number divergence according to sex. In flies, the splicing factors Sxl and TraF are the key players in the sex determination cascade. Their only known direct targets, doublesex and fruitless, are expressed only in a few cells in the larval brain , suggesting the involvement of new, uncharacterised targets in neuronal scaling. To explore this, I will employ the eCLIP-seq^{10,11} method on whole brain tissue from larvae expressing FLAG-tagged Sxl/TraF specifically in the selected cell types. eCLIP-seq allows transcriptome-wide detection of RNA-binding proteins' binding sites through immunoprecipitation. The functional importance of the newly identified actors of brain sexual differentiation will be tested through knockdown experiments followed by cell counting. CRISPR mutants will be

generated for the most promising candidates to back up the obtained results. In the remote eventuality that eCLIP-seq does not provide any exploitable results, I will perform a biased RNAi-based screen on female larvae. I will search for genes whose inhibition in the selected brain cells phenocopies *Sxl/TraF* silencing and induces "male brain cells subtype number". The candidates tested will be defined through a transcriptomic approach comparing gene expression between wild type males, females and sex-reversed individuals.

Finally, I will manipulate sex-specific neuronal scaling in the selected brain cells and investigate its functional significance for brain functions. The functional readouts will depend on the cell-type under investigation. For example, if motor neurons or dopaminergic neurons display the more significant sexual dimorphism in number, then, I will test locomotor behaviour or goal-directed behaviours respectively, first in wild type male and female larvae and subsequently in larvae overexpressing/silencing the sex determinants.

Egg laying in worms

Responsable scientifique : Christian Braendle

Laboratoire : IBV

Description :

Genetic assimilation describes the evolutionary process during which an environmentally induced (conditional) phenotype becomes a genetically controlled (constitutive) phenotype. This important concept remains controversial because the molecular mechanisms of genetic assimilation are rarely understood. Our project seeks to characterize the molecular genetic underpinnings of genetic assimilation in nematode egg retention. This conditional phenotype depends on the environmental modulation of egg-laying behaviour through neuromodulatory changes in the underlying neural circuit.

Our past collaborative research has established that natural and experimental populations of the nematode *Caenorhabditis elegans* show great variation in the environmental sensitivity of egg-laying behaviour, and in extreme cases, display genetic assimilation in egg retention. The existence of this natural variation represents a uniquely well-suited experimental paradigm to characterize the molecular basis of genetic assimilation. Leveraging natural diversity by employing a powerful multi-parental experimental mapping population and a global panel of wild strains, we will conduct unbiased whole-genome screens to uncover the molecular mechanisms governing genetic assimilation.

Our first aim involves conducting genome-wide association (GWA) mapping to identify the Quantitative Trait Loci (QTL) that modulate environmental sensitivity of egg retention and thought to underlie genetic assimilation. Our second aim will employ CRISPR Cas9 gene editing to pinpoint the molecular nature of QTL and explore its functional consequences on egg retention and environmental sensitivity in egg-laying behaviour. In our third final aim, we will evaluate the significance of molecularly validated QTL during the process of genetic assimilation across 100 generations of experimental evolution. Through the integration of these complementary approaches, we will contribute to a deeper understanding of the molecular architecture of genetic assimilation and generate fundamental insights into how environmental and genetic information can act interchangeably to regulate gene activity and phenotype expression.

This project is based on extensive past research, novel unpublished data and unique resources developed by the consortium, uniting two project partners with highly complementary expertise in molecular, developmental and evolutionary quantitative genetics.

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

Responsable scientifique : François Hug

Laboratoire : LAMHESS

Résultats obtenus grâce aux moyens informatiques attribués :

- Dernoncourt, F., Avrillon, S., Logtens, T., Cattagni, T., Farina, D. and Hug, F. (2025), Flexible control of motor units: is the multidimensionality of motor unit manifolds a sufficient condition?. *J Physiol.*, 603: 2349-2368. <https://doi.org/10.1113/JP287857>
- Hug F, Dernoncourt F, Avrillon S, Thorstensen J, Besomi M, van den Hoorn W, Tucker K. Non-homogeneous distribution of inhibitory inputs among motor units in response to nociceptive stimulation at moderate contraction intensity. *J Physiol.* 2025 Jun;603(11):3445-3461. doi: 10.1113/JP288504. Epub 2025 May 31. PMID: 40449022; PMCID: PMC12148210.

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)

Résultats obtenus grâce aux moyens informatiques attribués :

Présentation orale congrès JFBTM 2025, Mai 2025, Reims:

V. Rezapova, M Materozzi, C Girolet, A Wakkach, M Artyomov, C Blin-Wakkach. Dynamics in osteoclasts: a single cell perspective on fusion and fission.

Articles en préparation:

1- V. Rezapova, M Artyomov, C. Blin-Wakkach. Single-cell RNA sequencing data analysis of osteoclasts as multinucleated cells.

2- V. Rezapova, M Materozzi, C Girolet, A Wakkach, M Artyomov, C Blin-Wakkach. Dynamics in osteoclasts: a single cell perspective on fusion and fission.

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 majeure¹. 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 innée¹⁻³. Comme d'autres cellules myéloïdes (telles que les macrophages), les OCLs ont des origines multiples, selon le stade de développement où le contexte pathologique^{1,4}. 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^{1,4,5}. 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^{1–3}. 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 α -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 Cx3cr1. *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.

Génomique fonctionnelle de l'horloge circadienne

Responsable scientifique : Franck Delaunay

Laboratoire : IBV

Résultats obtenus grâce aux moyens informatiques attribués :

- Rapport de stage de Lucie Bolelli « Analyse expérimentale des oscillations biologiques de 12 heures ». Master1 Sciences de la Vie ; Parcours Bioinformatique et Biologie Computationnelle (2025)
- Inès Krawczyk et al., Poster meeting ICON "A systems biology approach to analyze the synchronization of cellular circadian clocks " 2025
- Inès Krawczyk et al, Poster meeting DYNABIO "A systems biology approach to analyze intercellular coupling between peripheral circadian clocks" 2026
- Franck Delaunay, Projet CSI MECACLOCK et projet ANR InSync, analyse de séries temporelles dans les cellules MMHD3 (données en cours d'analyse)

Description :

Thématique de l'équipe

L'horloge circadienne des mammifères synchronise la plupart des processus biologiques avec le cycle lumière/obscurité de la Terre pour assurer une coordination temporelle optimale de notre physiologie et de notre comportement. Le mécanisme qui sous-tend les rythmes circadiens est une horloge moléculaire présente dans pratiquement toutes les cellules et régie par un oscillateur moléculaire qui coordonne des processus moléculaires et cellulaires clés dans tous les organes. Des résultats de plus en plus nombreux indiquent que l'altération des rythmes circadiens a des conséquences négatives sur la santé. La compréhension des mécanismes qui sous-tendent ce lien entre la perturbation ou la désynchronisation des rythmes circadiens et la physiopathologie est essentielle compte tenu de la forte proportion de la population exposée au travail posté, à la lumière nocturne ou au décalage horaire chronique. Notre équipe de recherche fondamentale s'appuie sur ses travaux antérieurs portant sur le couplage entre les oscillateurs circadien et du cycle cellulaire et sur le rôle de l'horloge circadienne dans le métabolisme hépatique pour développer actuellement plusieurs projets complémentaires visant à mieux comprendre le rôle critique des horloges circadiennes périphériques dans la santé et la maladie. Le premier projet vise à comprendre comment les horloges des cellules d'un même organe se synchronisent. Le

deuxième projet a pour objectif d'analyser le rôle d'un facteur de transcription circadien dans la physiologie musculaire et l'homéostasie métabolique systémique.

Un projet plus récent concerne les oscillations qui ont des périodicités harmoniques de la période circadienne (12 h) et qui pour l'instant restent très mal caractérisées.

Génomique fonctionnelle de l'horloge circadienne et bioinformatique

Pour tous les projets de l'équipe, nous utilisons des approches de génomique fonctionnelle qui consistent dans la grande majorité des cas à analyser l'expression des gènes à l'échelle du génome entier (20-25 000 gènes) au cours du temps et dans différent organes ou des cellules cultivée *in vitro* chez la souris. Nous produisons pour cela des séries temporelles de plusieurs jours avec une résolution temporelle de 1,5-3 heures. Avec cette stratégie, nous cherchons par exemple à déterminer l'impact du contexte cellulaire (type d'organe), d'une mutation génétique (délétion de gène circadien), d'une conditions expérimentale spécifique (régime riche en sucres, activité physique) sur le transcriptome rythmique (perte ou gain d'amplitude des oscillations, déphasage). Les jeux de données obtenus sont des ensembles de fichiers de séquence générés par les séquenceurs à haut débit et nécessitant des moyens informatiques relativement conséquent en particulier pour les premières étapes de l'analyse.

C'est ce besoin qui va croissant qui nous motive à déposer ce projet pour accéder au HPC Azzurra. Une fois ces traitements réalisés, des pipelines bioinformatiques moins gourmands peuvent être utilisés au sein de la plateforme bioinformatique de l'iBV pour réaliser des analyse d'expression différentielle, détecter des gènes rythmiques ou visualiser les données par exemple.

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 :

Participation à la conférence IABM 2025 en mars 2025 à Nice pour une présentation poster

Participation à la conférence CAOS 2025 à Davos en juin 2025 pour une présentation poster et une présentation podium avec l'obtention du prix de la meilleure présentation technique de poster.

Auteur : Aguilar Vidal Théo, Robin Cremese, Jean-Baptiste Masson, Remy Winter, Isa Costantini, Thibault Poujade, Marc-Olivier Gauci

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.

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

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 :

Antoine Collin. Annotation cellulaire automatique pour la construction d'un atlas cellulaire. Médecine humaine et pathologie. Université Côte d'Azur, 2024. Français. [\(NNT : 2024COAZ6039\)](#). [\(tel-04991529\)](#)

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

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.

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 events¹. 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 phenotypes². Hybrids can display superior phenotypes when compared to their founding parental lineages. This hybrid vigour has been widely exploited to improve agricultural crops³. 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 genes⁴. 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 regulation⁵. 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 backgrounds⁶. This is unexpected, given the signatures of past genomic instability that typifies hybrids⁶. 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 stable⁶. 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

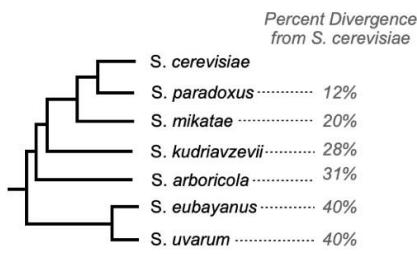


Figure 1: Schematic representation of the phylogeny of the genus *Saccharomyces*. The percentage divergence from *S. cerevisiae* is indicated on the right.

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

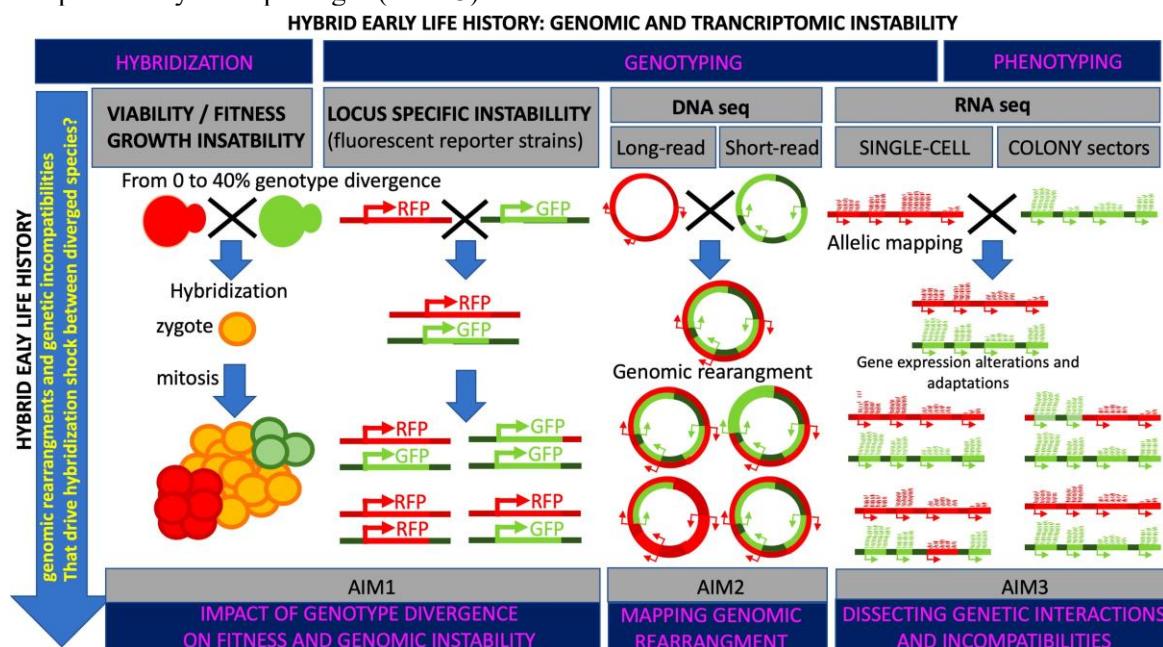
by mass-mating and marker selections.

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.

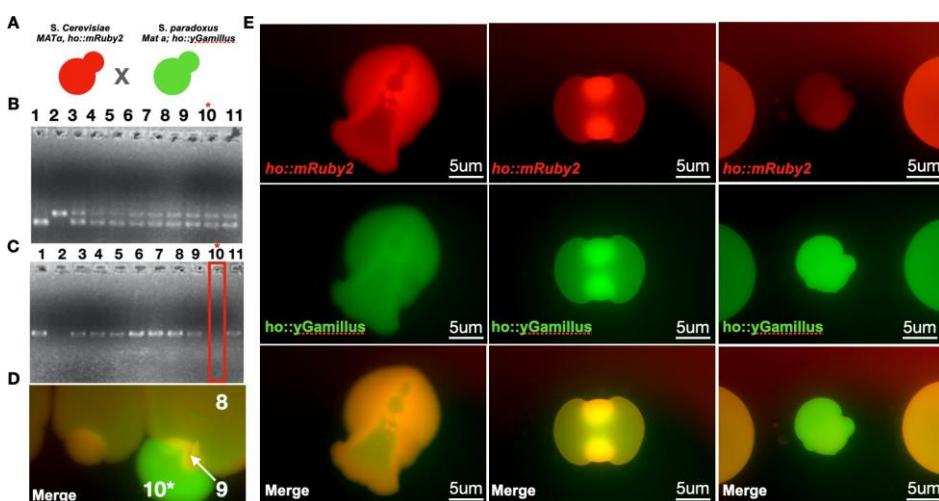
Saccharomyces yeasts are an ideal model system to mechanistically study hybrids as they can form despite extreme sequence divergence⁴. *Saccharomyces* hybrids between diverged parents naturally occur and have numerous industrial applications⁷⁸. Hybrids can be generated and maintained in the lab, evolved 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 hybrids⁹. 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)⁹¹⁰¹¹¹². 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 lab¹³¹⁴¹⁵. 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 fluorescence instability phenotypes can be observed at the RNA level. Preliminary data indicates that there is indeed genome instability in a r l y *Saccharomyces* interspecies hybrids relative to diploid parental strains (fig 2). A s h y b r i d



incompatibilities can result in ploidy changes I will also assay for ploidy. Contour-clamped homogenous electric field (CHEF) gel electrophoresis¹⁶, 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 ofzygotes to identify and re-array their resulting daughter cells¹⁷¹⁸. 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 hybrids⁶⁹¹⁹. **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 yeast²⁰. 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

Résultats obtenus grâce aux moyens informatiques attribués :

Compte utilisé pour assistance au développement et test de workflows ou des travaux ponctuels pour des projets de l'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.

Des résultats produits pour partie sur Azzurra depuis 2 ans font l'objet d'un article en cours de rédaction, dans le cadre du projet suivant :

The vital role of Greb1L during axial-progenitor maturation through the modulation of TGF β signalling and fatty-acid oxidation

Motamedi L.-C. Weerasinghe-Arachchige , L. Martin, Y. Neirijnck, A. Perea-Gomez, M. Rettel, F. Stein, F. Jian , C. Jeanpierre, A. Schedl

GREB1L is a haploinsufficient gene with clinical mutations identified in patients suffering from a range of congenital abnormalities such as renal agenesis, Müllerian agenesis, and earing impairments. The wide range of developmental defects observed in Greb1L knock

out (KO) embryos indicate that the gene has a vital role during early embryonic development. In this study, we characterise the dynamics of Greb1l expression and its role in mouse gastrula development.

GREB1L is widely expressed in many progenitor tissues of early mouse gastrulae but becomes restricted to the caudal mesoderm and neural progenitors at later stages. Furthermore, Greb1l knock out (KO) embryos develop an axial truncation phenotype at E8.5, which becomes pronounced at E9.

Proteomic analysis revealed KO embryos express significantly lower levels of proteins involved in anterior-posterior patterning, while proteins that play active roles in cell adhesion, cytoskeletal organisation, cell stemness, lipid metabolism, and hexose metabolism, were highly expressed. These results suggest that GREB1L may play a vital role in embryonic patterning by altering the metabolic state of cells.

scRNaseq analyses revealed a divergence in gene expression of several developmental clusters in KO embryos, including the caudal lateral epiblast (CLE), the caudal neuroepithelium, the paraxial mesoderm, and the lateral and intermediate mesoderm. In agreement with our proteomic analysis, transcripts of several genes involved in fatty-acid metabolism were enriched in several of these clusters. The highest level of transcriptional divergence was found in the CLE, a structure that harbours the earliest progenitors for caudal elongation of embryos at this stage. Interestingly, CLE cells in KO embryos exhibited an abnormally high expression of pluripotency markers, such as Oct4 (Pou5f1), and expressed high levels of Fst, a modulator of the TGF β signalling pathway. Further analysis revealed a specific sub-cluster in the CLE of Greb1l KO embryos that was characterised by a strong TGF β /BMP signature. Taken together, these results demonstrate a vital role for Greb1l in modulating TGF β signalling and cellular metabolism during early embryogenesis, thus permitting the maturation of the caudal epiblast into axial progenitors.

Epigenetic signaling in blood cancer

Responsable scientifique : Maria del Pilar Dominguez

Laboratoire : EPIgénétique du Cancer et iMMunothérapies (CEPIMMY). Centre Méditerranéen de Médecine Moléculaire (C3M)/ Inserm U1065

Résultats obtenus grâce aux moyens informatiques attribués :

The results obtained through the Azzurra HPC have been included in a manuscript currently under review:

Calaud, C., Serrano-Lopez, E., Cluse, L. A., Todorovski, I., Mardivirin, T., Teater, M., Maarad, A., Tawfic, N., Rutaquio, S., Fraser, P. J., Guiraud, K., Yoannidis, D., Kerreneur, E., Jacquel, A., Melnick, A. M., Johnstone, R. W., & Dominguez, P. M. (2025). Enhanced efficacy of a specific HDAC3 inhibitor in combination with 5-Azacitidine against diffuse large B-cell lymphoma. bioRxiv, 2025.04.21.648509. <https://doi.org/10.1101/2025.04.21.648509>

And were presented in the 2025 FASEB Hematological Malignancies conference (Southbridge MA, USA):

ENHANCED EFFICACY OF A SPECIFIC HDAC3 INHIBITOR IN COMBINATION WITH 5-AZACITIDINE AGAINST DIFFUSE LARGE B-CELL LYMPHOMA

Emilio M. Serrano-Lopez, Calixte Calaud, Leonie A. Cluse, Izabela Todorovski, Thomas Mardivirin, Matt Teater, Abdelillah Maarad, Noura Tawfic, Sophia Rutaquio, Peter J. Fraser, Killian Guiraud, David Yoannidis, Tim Semple, Emeline Kerreneur, Arnaud Jacquel, Ari M Melnick, Ricky W. Johnstone, Pilar M. Dominguez

Description :

Our project focuses on diffuse large B-cell lymphoma (DLBCL), an aggressive form of lymphoma. Despite progress in the treatment of DLBCL, approximately 40% of patients relapse or do not respond to the chemotherapy, which usually has fatal consequences. DLBCLs are characterized by profound alterations in epigenetic mechanisms (i.e. DNA methylation and chemical modifications of histones), which correlate with poor outcomes for patients. Our results show that combining epigenetic therapies is more effective than the treatment with each single agent, with higher induction of cell death and reduced proliferation in combo-treated lymphoma cells compared to cells treated with each epigenetic drug as monotherapy.

To investigate the molecular mechanisms that mediate the anti-tumor effects of the combinatorial therapy we have used NGS sequencing approaches, including RNA sequencing (RNA-seq) for transcription, chromatin immunoprecipitation followed by sequencing (ChIP-seq) to measure histone acetylation and reduced representation bisulfite sequencing (RRBS) to analyze DNA methylation. The proposed project involves the analysis of these individual datasets using established pipelines with the final goal to perform an integrative analysis to identify genes regulated by the epigenetic therapy in lymphoma cells (epi-genes).

The results of this project will reveal how the treatment modulates the epigenome and transcriptome of lymphoma cells, identifying the pathways targeted by the epigenetic therapy and uncovering new targets that will be further exploited therapeutically.

Dissect the effects of arginine deprivation on melanoma aggressiveness

Responsable scientifique : Michaël Cerezo

Laboratoire : C3M INSERM U1065

Description :

Melanoma is the most aggressive form of skin cancer and highly resistant to therapy. We recently uncovered a very specific role of arginine among all the metabolites for which melanoma cells and stromal cells compete to sustain their homeostasis. We have demonstrated that arginine deprivation induces downregulation of DNA methyltransferase 1 (DNMT1) and the development of resistance to targeted therapies as well as immunotherapy.

The goal of our project is to understand how arginine deprivation drives epigenetic changes and supports treatment resistance. To achieve this, we have set out the following specific aims:

- 1) Integrate Reduced Representation Bisulfite Sequencing (RRBS) and RNA-seq to identify the consequences of decreased DNMT1 on the DNA methylation profile and the transcriptomic consequences induced by arginine deprivation.
- 2) Identify a specific transcriptomic signature of arginine deprivation in tumor cells, and determine how this transcriptomic reprogramming contributes to melanoma resistance, using public transcriptomic datasets of melanoma patients treated with different therapies.
- 3) Determine whether a specific transcriptomic signature of arginine deprivation can be used as an early predictive biomarker to identify a subpopulation susceptible to developing resistance to a specific treatment.

Intégration de données multi-omiques pour l'identification de signatures pertinentes pour la stratification et la prise en charge thérapeutique de patients atteints de cancers pulmonaires

Responsable scientifique : Baharia MOGRABI

Laboratoire : IHU-Respirera

Description :

Ce projet vise à réaliser une analyse intégrative multi-omique de cohortes de patients atteints de NSCLC, combinant données transcriptomiques, altérations génomiques et informations cliniques. L'objectif est d'identifier des signatures multi-omiques permettant de distinguer des sous-groupes de patients, puis d'en évaluer la pertinence clinique en vue d'orienter les décisions thérapeutiques.

Les données proviennent de larges cohortes publiques TCGA-LUAD et SU2C-MARK (environ 500 et 150 échantillons respectivement), de jeux de données publiques plus restreints (lignées cellulaires, séquençage sur cellule unique), et de données internes de séquençage ciblé. Le traitement et l'intégration de ces données volumineuses et complexes nécessitent l'accès à des ressources HPC afin de garantir des analyses efficaces, pertinentes et reproductibles.

Les analyses s'appuieront sur des pipelines standardisés et personnalisés pour réaliser des tâches intensives en calcul, telles que l'alignement des lectures, le comptage et le traitement des altérations génomiques. Elles nécessiteront ensuite des environnements interactifs pour l'exploration et la représentation des données via des workflows statistiques dédiés, incluant des méthodes d'apprentissage automatique.

Droit, économie, géographie

EUR ELMI, LexSociété, Odyssee

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 :

Auteurs: Veronika Czellar et Engin lyidogan

Titre: Multifractal Cryptocurrencies (en révision dans une revue de rang A)

Lien SSRN au papier: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5194867

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

Résultats obtenus grâce aux moyens informatiques attribués :

Colin Brisson, Ayoub Kahfy, Marc Bui, Frédéric Constant, « Named Entity Recognition in Context: Edit_Dunhuang team Technical Report for Evahan2025 NER Competition », Proceedings of the Second Workshop on Ancient Language Processing, pp 176–181.

Ancien modèle en ligne (https://github.com/colibrisson/CHAT_models/releases). Un nouveau modèle basé sur une architecture transformer va être rendu public dans les semaines à venir.

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.

BATI – Modeling Soil Sealing in France

Responsable scientifique : Dennis Fox

Laboratoire : UMR ESPACE 7300 CNRS

Résultats obtenus grâce aux moyens informatiques attribués :

Kporha V.K., Fox D.M, Banitalebi M., Bouroubi Y., Fournier R, 2025 (accepted). Comparing Daily and 8-Day MODIS Land Surface Temperature Data for Urban Heat Island Assessment Using Random Forest Modeling in Data-Limited Regions. *Remote Sensing Applications: Society and Environment*.

Kporha, Fox & Banitalebi: UMR ESPACE CNRS, Université Côte d'Azur

Bouroubi & Fournier: Université de Sherbrooke (Québec)

Description du projet en pages suivantes.

Description de projet

Titre du projet :	BATI
Nom et Prénom du porteur :	Dennis FOX (PR UCA)
Adresse mail du porteur :	dennis.fox@univ-cotedazur.fr
Laboratoire / établissement :	UMR ESPACE CNRS / UCA EUR ODYSSEE
Autres équipes / laboratoires impliqués	UPR CERDAFF / UCA, EUR Droit et Science Politique
Durée d'accueil souhaitée (maximum 1 an)	6 mois
Période envisagée pour l'accueil :	6 Janvier – 5 juillet 2025

1. Contexte scientifique et objectifs à court terme (1 page max)

Il est estimé que la part de la population mondiale habitant en milieu urbain augmentera de 55% aujourd’hui à 68% en 2050 (United Nations, 2019) ; ceci représente près de 2 milliards de personnes supplémentaires vivant en ville (Angel et al., 2011 ; United Nations, 2019). Typiquement, la croissance urbaine se fait aux dépens des sols agricoles (en priorité) ou naturels (EEA-FOEN, 2016). Afin de lutter contre l’étalement urbain et l’artificialisation des sols en France, l’état a progressivement mis en place une législation qui incite à augmenter la densité de l’habitat en ville : loi SRU de 2000, loi Alur de 2014, loi dite « climat et résilience » de 2021. L’objectif du gouvernement à terme est d’atteindre « Zéro Artificialisation Nette » (ZAN) (Fosse et al., 2019). La protection des sols agricoles ainsi que de la biodiversité est une préoccupation majeure de l’Europe (EEA-FOEN, 2016), l’Amérique du Nord et ailleurs dans le monde.

Si la densification des villes est une stratégie essentielle à la protection des sols agricoles et des milieux naturels, les effets négatifs de l'imperméabilisation des sols n'ont pas été pris en compte dans les politiques actuelles. Plusieurs études ont montré que l'imperméabilisation des sols augmente le risque de ruissellement ainsi que l'exposition aux inondations (Konrad and Booth, 2002 ; Braud et al., 2013 ; Miller et al., 2014 ; Nguyen et al., 2021 ; Fox et al., 2024b). A l'échelle mondiale, la valeur des infrastructures et le nombre de personnes exposées aux inondations seraient en hausse et cette tendance pourrait se renforcer sous la double contrainte de la densification des villes et l'augmentation du nombre d'événements pluvieux extrêmes dans un contexte de changement climatique. Une deuxième conséquence de l'imperméabilisation des sols est l'intensification des îlots de chaleur urbains (ICU) (Imhoff et al., 2010 ; Rahaman et al., 2022).

La France possède incontestablement les meilleures données géoréférencées de bâti au monde. Plus de 50 millions de bâtiments sont numérisés sous forme de polygones vectoriels dans la base de données BD-TOPO de l'IGN, exploitables gratuitement dans les Systèmes d'Information Géographique (SIG). De plus, ces données ont une durée temporelle unique, remontant à environ 2008. Nous avons commencé à exploiter ces données pour élaborer une méthode de calcul de l'imperméabilisation des sols originale (Fox et al., 2024a) en partenariat avec l'Université de Sherbrooke au Québec, ainsi que pour le risque d'inondation (Fox et al., 2024b).

De nombreuses études ont cartographié l'évolution de l'occupation du sol à partir d'imagerie satellitaire. Quelques études, comme les nôtres (Fox et al., 2024a, 2024b), ont utilisé les building « footprints », mais elles sont relativement rares car c'est une forme de donnée relativement peu disponible mais en croissance exponentielle. Aucune étude n'a combiné les deux approches, occupation du sol et building footprints, dans une seule analyse ; aucune étude n'a encore lié l'évolution temporelle des surfaces totales des bâtiments (*footprint * nombre d'étages*) à l'évolution de l'occupation du sol, et encore moins à l'évolution future dans un contexte où la limite urbaine serait fixe dans le cadre d'une loi de type ZAN.

La cartographie multi-temporelle des bâtiments a de nombreuses applications : calcul du taux d'imperméabilisation des sols à l'aide d'IA, impact de l'imperméabilisation des sols sur les ICU, évolution temporelle de l'exposition aux risques d'inondation, calcul des besoins en surface bâtie réels et relation au gel de l'étalement urbain.

Dans le court terme, les données de bâti serviront dans le contexte d'une thèse co-encadrée avec l'Université de Sherbrooke (début 1/10/2024) sur les ICU et l'exposition aux inondations. Nous avons aussi, avec Anne Rainaud (Droit), débuté un projet sur l'expansion urbaine à l'échelle nationale par le biais d'un projet financé par l'Académie-5 sur la perte des sols agricoles à l'urbanisation. Un premier stage de M1 a été réalisé ce printemps sous la codirection de Dennis Fox et d'Anne Rainaud. Le traitement de données dans le cadre de cet AAP est essentiel à la suite du projet.

2. Description du besoin en support ingénieur (1 page max)

Si notre approche est innovante et pertinente, elle rencontre 3 problèmes que nous essayons de régler peu à peu, mais lentement et sans toutes les compétences nécessaires : 1- la qualité des données, 2- le temps de calcul, et 3- les approches de modélisation. Chacun est décrit ci-dessous.

1) La qualité des données : les premières couches vectorielles de bâtiments datent d'environ 2002-2005 avec des mises à jour quasi-annuelles. Elles sont plus fiables à partir de 2008 (Version 2 de la BD-TOPO) mais elles nécessitent, néanmoins, des corrections. La technique de numérisation des bâtiments a évolué avec le temps : au début, ils ont été numérisés à partir de photographies aériennes, progressivement ils ont été basés sur le cadastre. Le même bâtiment n'a donc pas le même *footprint* à travers le temps. De plus, les premières couches rassemblent des blocs de bâtiments, parfois avec des cours intérieurs intégrées dans le bloc, tandis que les couches successives se sont affinées pour éliminer l'intégration de ces surfaces dans le « bâti ». Il y a donc 2 types d'erreurs – 1- la précision du positionnement des bâtiments et leurs contours dans l'espace, et 2- la superficie des bâtiments. Ces erreurs sont illustrées dans la Figure 1.

Pour certaines années, l'évolution du bâti montre une perte de surface entre deux années successives, ce qui est théoriquement possible par la destruction de bâtiments existants, mais largement artificiel dans notre cas. Il faut développer une approche qui partira de la couche de référence de 2023 pour corriger les polygones des années précédentes en combinant des critères de proximité et de forme géométrique. L'ingénier peut automatiser les modifications dans un premier temps, puis un stagiaire pourrait faire un contrôle visuel, notamment sur les bâtiments détruits, par la suite.



Figure 1 : l'image de gauche montre la configuration du bâti en 2010. A droite, en 2023, des blocs de cours intérieurs ont été éliminés de la surface du bâtiment et la géométrie des bâtiments a été affinée (voir bâtiment en bas à gauche, par exemple).

2) Le temps de calcul : jusqu'à aujourd'hui, nous avons extrait les superficies de bâtiments à l'aide d'un programme python qui tourne sous ArcGIS et qui calcule la surface au sein d'un polygone urbain de 100 m x 100 m. Nous avons 211 polygones urbains, d'une superficie allant de 60 km² à 470 km² (moyenne de 268.5 km²); chacun est découpé en carrés de 100 m de côté pour un total de plus de 5.6 millions de carrés. Le nombre de bâtiments par an dans ces polygones urbains varie de 8 millions en 2008 à 20 millions en 2023. Chaque extraction de la superficie du bâti pour une année prend environ 21 jours sur un desktop classique. Il nous faut donc une puissance de calcul bien supérieure à ce que nous utilisons aujourd'hui. Nous avons donc besoin du mésocentre Azzurra afin de progresser plus rapidement. Si nous voulions faire tous les bâtiments de France, cela reviendrait à plus de 200 jours de calcul par année que pour l'extraction du bâti, plus pour le calcul des surfaces réellement imperméabilisées.

Pour information, Microsoft vient de donner accès à une couche vectorielle de *building footprints* avec près de 130 millions de bâtiments pour les USA : <https://opengislab.com/blog/tag/Microsoft+Bing+Building+Footprints>.

Nous avons également identifié 17 villes avec des données vectorielles de même nature en Europe. Notre ambition est donc d'être parmi les premiers à exploiter ces données.

3) La modélisation : enfin, nous avons un troisième domaine où un ingénieur serait utile. L'étude sur la modélisation de l'imperméabilisation du sol publié (Fox et al., 2024a) lors du colloque *International Conference of Environmental Remote Sensing and GIS* (Zagreb, juillet 2024) a nécessité différents modèles de régression et traitement d'IA pour prédire le taux d'imperméabilisation à partir de la couche vectorielle du bâti. Nous avons utilisé les approches mathématiques qui nous semblaient les plus pertinentes, mais l'apport d'un modélisateur nous permettrait d'aller plus loin dans cette deuxième phase où nous comptions élaborer nos modèles à partir des données françaises, européennes (17 villes) et américaines (citées ci-dessus).

L'ingénieur de la MSI serait sous la supervision de Dennis FOX et travaillerait en collaboration étroite avec Mostafa Banitalebi, IE de l'UMR ESPACE. Il/elle serait installé.e dans les locaux de l'IMREDD. Pour le volet protection des sols agricoles, Anne Rainaud participerait également à l'encadrement.

Références :

- Angel, S., Parent, J., Civco, D.L., Blei, A., Potere, D., 2011. The dimensions of global urban expansion: Estimates and projections for all countries, 2000-2050. *Prog Plann* 75, 53–107. <https://doi.org/10.1016/j.progress.2011.04.001>
- Braud, I., Breil, P., Thollet, F., Lagouy, M., Branger, F., Jacqueminet, C., Kermadi, S., Michel, K., 2013. Evidence of the impact of urbanization on the hydrological regime of a medium-sized periurban catchment in France. *J Hydrol (Amst)* 485, 5–23. <https://doi.org/10.1016/j.jhydrol.2012.04.049>
- EEA-FOEN, 2016. Urban Sprawl in Europe, Joint EEA-FOEN report no. 11/2016.
- Fosse J., Belaunde J., Dégremont M. et Grémillet A., 2019. Objectif « Zéro Artificielle Nette » : quel leviers pour protéger les sols ? Rapport au Ministre de la Transition écologique et solidaire, au ministre de la Cohésion des territoires et des Relations avec les collectivités territoriales et au ministre chargé de la Ville et du logement.
- Fox D.M., Banitalebi M., Fournier R., Bouroubi Y., 2024a. Quantifying total imperviousness from building footprint area and very high resolution air photographs. *International Conference on Environmental Remote Sensing and GIS (ICERS)*, 11-12 July, Zagreb, Croatia. Available here : <https://zenodo.org/records/11584934>
- Fox D.M., Banitalebi M., Rainaud A., 2024b. Building footprint layers show that flooding risk increased more due to greater building exposure than to greater peak discharge with urbanisation in SE France. *Journal of Hydrology, Regional Studies*.
- Available here: <https://www.sciencedirect.com/science/article/pii/S2214581824002313>
- Imhoff, M. L., Zhang, P., Wolfe, R. E., & Bounoua, L. (2010). Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sensing of Environment Volume* , 114(3), 504–513.
- Konrad P., and Booth D.B., 2002. Hydrologic Trends Associated with Urban Development for Selected Streams in the Puget Sound Basin, Western Washington. U.S. Geological Survey Water-Resources Investigations Report 02-4040.
- Miller J.D., Kim H., Kjeldsen T.R., Packman J., Grebby S. Dearden R., 2014. Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *Journal of Hydrology*, 515, 59-70.
- Nguyen, H.D.; Fox, D.; Dang, D.K.; Pham, L.T.; Du, Q.V.V.; Nguyen, T.H.T.; Dang, T.N.; ran, V.T.; Vu, P.L.; Nguyen, Q.-H.; et al.. Predicting Future Urban Flood Risk Using Land Change and Hydraulic Modeling in a River Watershed in the Central Province of Vietnam. *Remote Sens.* 2021, 13, 262. <https://doi.org/10.3390/rs13020262>.
- Rahaman Z.A., Al Kafy A., Saha M., Rahim A.A., Almulhim A.I., Rahaman S.N., Fattah M.A., Rahman M.T., Al- Faisal K.S.A., Al Rakib A., 2022. Assessing the impacts of vegetation cover loss on surface temperature, urban heat island and carbon emission in Penang city, Malaysia. *Building and Environment*, 222. <https://doi.org/10.1016/j.buildenv.2022.109335>
- United Nations, Department of Economic and Social Affairs, Population Division, 2019. World Urbanization Prospects The 2018 Revision.

OPAL

Azzurra fait partie de l'infrastructure OPAL, dont une présentation peut être consultée [sur cette page](#).

Au cours de l'année 2025, **217** utilisateurs ont pu bénéficier d'une accréditation OPAL **par l'université**, parmi lesquels :

- **143** utilisateurs disposant d'un compte sur Azzurra pour un projet en cours de validité.
- **74** utilisateurs sans compte sur Azzurra, dans le but de calculer sur un autre cluster OPAL (tous sur NEF), décrits en page suivante.

43 personnels accrédités par une autre entité ont pu disposer d'un compte sur Azzurra.

L'accréditation OPAL donnée par l'université 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

Small Language Model

Responsable scientifique : Frederic Precioso

Laboratoire : I3S Maasai - Equipe Commune INRIA-CNRS-UniCA

Plateforme OPAL : NEF (Inria)

Nombre d'accrédités : 4

Description :

Context:

With the emergence of LLM models, many NLP tasks have been revolutionized. However the access to a proprietary external LLM is often a problem for sensitive domains. If you have data, internal to your company, you may not want to disclose them to a third party. Thus how can you benefit from the power of LLMs for your company without any risk: building your own Small LM (SLM) may be the answer.

This strategy raises then many questions on the amount of data you will need to provide to train your model, on the diversity and purity of this data, on the GPU resources required to train such a model even if you target a Small LM, and on the control you can have on the hallucinations and the biases your model may present finally.

FATES properties (Fairness, Accountability, Transparency, Ethical, Sustainable/Safety/Security) are targets for companies which use Machine Learning models for their business. However, many models available on the shelf do not consider these FATES properties and enforcing them after building the model remains a challenge. This is even more critical for Large Language Models.

Objectives:

In this project, we will explore all the techniques available to evaluate how reasonable it is to hope for training our own SLM. We will try to identify how much data we should have to train, how precise the answers will be, how high the risk of hallucinations is, how easy it is to control biases, and if at the end it is not simpler and more efficient to use the current state-of-the-art technique in this case which is distillation of a larger model.

Throughout the project we will evaluate the performances of our models on standard benchmarks and on more dedicated ones in order to evaluate if focusing on a specific and restricted professional domain helps to reach satisfying performances with less data and less computation. We will focus for this last aspect on targeted medical domains where the vocabulary and the downstream tasks are specific.

We will also investigate the different available benchmarks for FATES properties and develop a benchmarking platform to evaluate some of the recent LLMs for these properties. In particular, if it is easy to find Fairness or Sustainability benchmarks, it is more challenging to find some for Accountability, Transparency, or Ethics.

We will integrate some of the mitigation techniques we will have found in the literature to show their effect on pretrained LLMs, and our own SLM.

References:

The Ultra-Scale Playbook: Training LLMs on GPU Clusters, Huggingface,
https://huggingface.co/spaces/nanotron/ultrascale-playbook/resolve/main/The_Ultra-Scale_Playbook_Training_LLMs_on_GPU_Clusters.pdf

Allal, L. B., Lozhkov, A., Bakouch, E., Blázquez, G. M., Penedo, G., Tunstall, L., ... & Wolf, T. (2025). SmolLM2: When Smol Goes Big--Data-Centric Training of a Small Language Model. arXiv preprint <https://arxiv.org/abs/2502.02737>

Kristina Gligorić, Tijana Zrnic, Cinoo Lee, Emmanuel J. Candès, and Dan Jurafsky. 2025. Can Unconfident LLM Annotations Be Used for Confident Conclusions? To appear, NAACL 2025.
<https://arxiv.org/pdf/2408.15204>

Researchers at UC Berkeley have open sourced Sky-TI, a 'reasoning' AI model that can be trained for less than \$450. | TechCrunch

AI researchers at Stanford and the University of Washington were able to train an AI "reasoning" model for under \$50 in cloud compute credits, according. | Researchers created an open rival to OpenAI's o1 'reasoning' model for under \$50 | TechCrunch

Fan, Z., Chen, R., Hu, T., & Liu, Z. (2024). FairMT-Bench: Benchmarking Fairness for Multi-turn Dialogue in Conversational LLMs. arXiv preprint arXiv:2410.19317.

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., ... & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. Computational Linguistics, 50(3), 1097-1179.

Wang, A., Hertzmann, A., & Russakovsky, O. (2024). Benchmark suites instead of leaderboards for evaluating AI fairness. Paterns, 5(11).

Estimer la proximité thématique des chercheurs

Responsable scientifique : Frédéric Giroire

Laboratoire : Projet COATI Inria / I3S / Université Côte d'Azur / CNRS

Plateforme OPAL : NEF (Inria)

Nombre d'accrédités : 2

Description :

Ce projet se place dans le contexte de notre étude de l'évolution de la productivité et des collaborations des chercheurs, et de comment elles sont impactées par des financements qu'ils soient nationaux ou internationaux. Dans ce cadre, nous avons recueilli l'ensemble des publications (avec au moins un auteur français) de SCOPUS. Nous avons développé différents algorithmes et métriques pour évaluer la proximité entre les chercheurs, en utilisant les journaux et conférences dans lesquels ils publient. Les données ont déjà été fortement consolidées, mais il reste des ambiguïtés (notamment dues aux noms des journaux et des conférences). De plus, il faudrait évaluer et comparer les différentes mesures de proximité afin de proposer un outil en ligne, permettant aux chercheurs de se situer vis-à-vis des disciplines et/ou de leurs collègues.

Ce travail sera supervisé par Frédéric Giroire (CNRS/I3S) et Nicolas Nisse (Inria/I3S) et est en collaboration avec Michele Pezzoni (UniCA/GREDEG).

Plus précisément, les étudiants devront réaliser les tâches suivantes :

Outil en ligne pour estimer la distance entre deux chercheurs

Projection en 2D des vecteurs représentant les chercheurs dans l'espace des publications.

Options : publications principales, période de publications

Consolidation de la liste des conférences et journaux de toutes les disciplines

Outil : Conception d'un modèle IA de classification des noms de conférences.

Exploration des composantes principales ou clusters des publications d'un chercheur.

Etudier la relation / corrélation entre les différentes métriques

Evaluer les domaines de recherche des chercheurs en utilisant de grands modèles de langage (LLM)

NeuroMod project

Responsable scientifique : Gilles Scarella

Laboratoire : LJAD

Plateforme OPAL : NEF (Inria)

Nombre d'accrédités : 9

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>

ANR TRACTIVE: AI-driven video interpretation of character representation in films

Responsable scientifique : Lucile Sassatelli

Laboratoire : I3S

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.

Langage Python pour l'Intelligence Artificielle

Responsable pédagogique : Chuan Xu

Plateforme OPAL : NEF (Inria)

Nombre d'accrédités : 22

Description :

L3 IA : SLEI505 - ECUE IA : Langage Python pour l'IA

Ce cours offre une introduction approfondie à l'utilisation du langage Python dans le domaine de l'Intelligence Artificielle (IA). Les étudiants apprendront à concevoir, implémenter et entraîner des modèles d'IA en utilisant des bibliothèques populaires telles que TensorFlow, PyTorch et Scikit-learn.

Le cours couvre les principes fondamentaux de l'apprentissage automatique, de la manipulation des données à l'entraînement de modèles, en passant par l'évaluation des performances. Un accent particulier sera mis sur la pratique, avec des exercices et projets réguliers pour appliquer les concepts abordés en classe.

Les étudiants auront également la possibilité de travailler sur le cluster NEF de l'Inria, un environnement de calcul haute performance, pour entraîner leurs modèles sur des ensembles de données de grande échelle. Cette expérience unique leur permettra de se familiariser avec les infrastructures distribuées et d'acquérir des compétences en calcul parallèle, cruciales dans le domaine de l'IA moderne.

Apprentissage Automatique Distribué (Distributed Machine Learning)

Responsable pédagogique : Chuan Xu

Plateforme OPAL : NEF (Inria)

Nombre d'accrédités : 16

Description :

C'est un cours pour Master I Informatique qui explore les concepts, les algorithmes et les technologies liés à l'apprentissage automatique dans des environnements distribués. Les étudiants apprendront à concevoir, implémenter et optimiser des algorithmes d'apprentissage automatique adaptés aux systèmes distribués, en prenant en compte les défis tels que la latence réseau, la répartition des données, et la synchronisation entre les nœuds. Un compte sur un cluster NEF est essentiel pour tout projet ou cours portant sur l'apprentissage automatique distribué, car il permet de bénéficier des ressources nécessaires pour traiter de grandes quantités de données, d'optimiser les algorithmes distribués, et d'effectuer des tests dans un environnement réaliste.

M2 DSAI

Responsable pédagogique : Michel RIVEILL

Plateforme OPAL : NEF (Inria)

Nombre d'accrédités : 11

Description :

Dans le cadre de la seconde année du MSc DSAI (formation dédiée à l'Intelligence Artificielle et aux Traitements de Données), nous souhaiterions que nos étudiants puissent accéder à NEF (Inria est d'accord) affinée pouvoir entraîner des modèles de Deep Learning de taille significative dans le cadre de certains cours mais aussi lors du research project qui démarre le 1er novembre et quitte fin mars.