Applications to Catalysis Workflows

Leandro Liborio¹, Patrick Austin², Alejandra González-Beltrán², Abraham Nieva de la Hidalga³ ¹Theoretical and Computational Physics Group, SCD, STFC ²Data and Software Engineering Group, SCD, STFC ³UK Catalysis Hub, Research Complex at Harwell

EGD meets ESG, October 5th 2023, Freiburg



Motivation

WP5 Objectives:

EuroScience Gateway

Develop customizable pilot workflows Onboard new communities Targeting Materials Science community



Enable researchers in the physical sciences to handle data more easily

WP4, Pathfinder 1:

...creating an integrated environment to support capture, analyse and reuse data. Prototype developed for XAS data



Catalysis use case



Nieva de la Hidalga, A, et al.: A workflow demonstrator for processing catalysis research data. Data Intelligence 4(2), 455-470 (2022). doi: 10.1162/dint_a_00143



Nieva de la Hidalga, A, et al.: A workflow demonstrator for processing catalysis research data. Data Intelligence 4(2), 455-470 (2022). doi: 10.1162/dint_a_00143

Current approach



https://bruceravel.github.io/demeter/



🔁 🖡 Eile Group Energy Mark Plot Freeze Merge Monitor Help Save A U I Au Chloride, scan 1 <untitled> 🗹 Au Chloride, scan 2 Main window Current group: Au Chloride, scan 3 Datatype: xmu 🛛 Freeze File /home/bruce/Data/ANL/20BM/2005.10/standards/aucl.0003 Element 79: Gold Edge L3
 Energy shift
 Importance Normalization and background removal parameters E0 11919.94738 🧿 Normalization order () 1 () 2 (3 Pre-edge range -150.000 (O) to -30.000 (O) 🗹 Flatten normalized data Normalization range 150.000 (o) to 869.893 (o) Edge step 0.600201
fix Rbkg 1.0 k-weight 2 Spline clamps low None V Spline range in k o to 15.955 o to 969.8772 0 high Strong V Spline range in E Standard None E k R q kq E k R q Forward Fourier transform parameters k-range 3.000 0 to 13.955 0 dk 1 window Hanning Plotting k-weights 0 0 1 0 2 0 3 0 kw arbitrary k-weight 0.5 phase correction Plot in energy Backward Fourier transform parameters 🖌 µ(E) Ο μ(E) 0 to 3 O dR 0.0 window Hanning ↓ R-range 1 ✓ Background pre-edge line **Plotting parameters** post-edae line Plot multiplier 1 y-axis offset 0 Normalized

Normalized Derivative Derivative 2nd derivative 🔲 2nd derivative Emin -200 Emax 800 Renamed aucl.0003 to Au Chloride, scan 3 2 :: Artemis [EXAFS data analysis] *<untitled>* LC X Eile Monitor Fit Plot Help GDS GDS Name Fit 2 Fit space: k
R
Q **⊖**Add OAdd Fit description Plot Hide "Pyrite FeS2" Show "FeS2.inp fit to Pyrite FeS2 Fit History 14 💋 Journal Showlog Plotted in Pyrite FeS2 in R space The second second 🛓 🔡 Artemis (Plot) 👝 🗡 1 [FeS2] S.1 Pyrite FeS2 k R q eS2] S.2) window eS2] S.3))) 🗹 Include path Plot after fit k-weight 0 0 1 0 2 0 3 0 kw 2 :: Artemis [] og] Fit - C × limits stack indic VPaths Name : Fit 1 (tlvww) : fit to Pyrite FeS2 Description Plot χ(R) Figure of merit Time of fit 2012-10-14T16:50:52 Demeter 0.9.13 with perl 5.014002 on linux Real Imag Magnitude Environment Interface Plot y(a) Artemis (Wx 0.9903) Prepared by bruce@clotho Magnitude

Real

Imag. Contact Plot fit Plot bkg 3 Plot window Plot residual Radial distance (Å) Plot running R-factor : 7.2382812 ependent points 63007 -6 61324 unber of variables 15909.175891 kmin 0 kmax 15 hi-square Meduced chi-square M-factor 4912.8456310 rmin (rmax 6 actor : 0.0200580 surement uncertainty (k) : 0.0003983 itting k weight: ▼ 1 ▼ 2 ▼ 3 other 0.5 surement uncertainty (R) : 0.0009550 qmin 0 qmax 15 mber of data sets ther parameters lotting list ✓ Include in fit ✓ Plot after fit Fit bar ✓ Data: Pyrite FeS2 cito 0 ted "Pyrite FeS2" as the magnitude and real # +/- 0.08853672 # +/- 1.45351925 # +/- 0.01253567 # +/- 0.00157563 0.60191655 = 1.90186138 = -0.00575161 0.00200663 ons between variables delr & enot ss & amp other correlations below 0.4 --> 0.7947 --> 0.7926 Freeze d Clear Save next plot to a file Print Preview - Print Sclose 😸 Save Save

XAS in Galaxy



https://xraypy.github.io/xraylarch/



Larch Athena:

- Processing and Normalization
- Cropping energy range
- Outputs project file and plots
- Larch FEFF:
 - Load from cif and converts to FEFF input file (or loads FEFF input)
 - Outputs zipped directory of paths
 - Larch Select Paths:
 - Selects which paths from Larch FEFF to use
 - Defines parameters for these paths
- Larch Artemis:
 - Performs fitting on FEFF paths
 - Outputs report on fitting and plots

Also implementing utility tools and tools for other XAS techniques (not shown on diagram).

Process and Normalise XAS with Larch Athena in Galaxy

📮 Galaxy μSR	😤 Workflow Visualize Shared Data - Help - User - 🌲 🕿 🏢			Using 44.1	мв
Tools 🌣 -	Larch Athena generate Athena projects from XAFS data (Galaxy Version 0.9.66+galaxy0)	•	History	+ ≓	•
search tools X	XAFS data file	- 1	search datasets	*	×
1. Upload Data	C () (5: test.xmu.txt -	2	PSDI_XAS		
Tools Under Development	Normalised X-ray Absorption Fine Structure (XAFS) data, generated by DAWN.	_	S 16 7 kB	00 15	0
PyMuonSuite AIRSS UEP Optimise run UEP optimisation	Plot graph ONO				\$
PyMuonSuite AIRSS Cluster run clustering for optimised structures	Whether to plot the pre/post edge fitting and the normalised xµ data.		This history is	empty.	
MuSpinSim Combine combine datafiles generated from MuSpinSim	✓ Execute		data from an extern	al source.	
MuSpinSim Generator Generate MuSpinSim config from a structure file	Using Larch, create an Athena project file from the input X-ray Absorption Fine Structure (XAFS) data file. This renames the xmu column of the data to mu. Optionally, plot the xµ data along with pre and post edge fitting lines for visual inspection.				
Larch FEFF generate FEFF paths from XAFS data	Citations: 🏢				
Larch Select Paths select FEFF paths for XAFS data	- Newville, M. (2013). Larch: An Analysis Package for XAFS and Related Spectroscopies. Journal of Physics: Conference Series, 430, 012007. https://doi.org /10.1088/1742-6596/430/1/012007 🖆				
Larch Athena generate Athena projects from XAFS data	- Ravel, B., & Newville, M. (2005). \lessi\greaterATHENA\less/i\greater,\lessi\greaterARTEMIS\less/i\greater,\lessi greater,\lessi\greater,\grea				
Larch Artemis generate Artemis projects from XAFS data	Requirements: ?				
MuSpinSim Configure define simulation parameters	- xraydb (Version 4.4.7) - sqlalchemy (Version 1.4.46)				
Collection Operations	- matplotlib (Version 3.5.2)				
WORKFLOWS	MIT License Z				
All workflows	Creators:				
https://muongalaxy-dev.esc.rl.ac.uk	A Patrick Austin 🐨 🗠				>





Path Generation in FeS₂ using Larch FEFF in Galaxy

			/	0.000004000	201120.00		-				
Т	ools	· ·	7434.1940	0.89577218	304310.00			History	+	= -	
	0015	14	7437.7090	0.89366937	304338.00			Thistory			
			7441,2440	0.89233424	304342.00						
S	earch tools	×	7444.7980	0.89109983	304576.00			search datasets		* ×	
			7448.3710	0.88991984	304397.00						
	+ Unload Data		7451.9620	0.89021992	304464.00			Uppamod history			
	S Opioad Data		7455.5740	0.88787942	304333.00			offinamed history		-	
			7459.2040	0.88208656	304131.00						
G	et Data		7462.8530	0.87531841	304375.00			71 6 1.0		0.1	a
			7400.3210	0.07040240	204040 00			2 / 1.0 KB		¥2	N
SF	PIN DYNAMICS		7473 9150	0.85785170	304975 00						-
			7477.6400	0.85054938	305708.00			S2 57			*
M	uSpinSim		7481.3850	0.84453901	306082.00						1
			7485.1480	0.83863201	305961.00			2 : test.xmu	0	-	
FI	ND MUON STOPPING SITES		7488.9310	0.83371670	306201.00						
	ED Math ad		7492.7330	0.82904991	306134.00			1 : 1564889 cif	0	11	
U	EP Method		7496.5540	0.82459446	306039.00			1.1501005.00			
-			7500.3940	0.82079844	305977.00						
0	ther Methods		7504.2530	0.81677576	305156.00						
			7508.1310	0.81178911	306748.00						
0	THER TOOLS		7512.0280	0.80667240	306054.00						
D	14		7515.9440	0.80196850	305949.00						
P	MuonSuite		7519.8790	0.79709432	305906.00						
			7523.8340	0.79034294	205933.00						
FI	le Conversion		7531 8000	0.77566164	305816 00						
			7535.8120	0.77172412	305895.00						
To	ools Under Development		7539,8420	0.76918155	305839.00						
			7543.8920	0.76588805	305840.00						
C	ollection Operations		7547.9610	0.76149338	305863.00						
			7552.0490	0.75668300	305740.00						
W	ORKFLOWS		7556.1560	0.75093523	305560.00						
	and all access		7560.2820	0.74539760	305668.00						
A	worknows		7564.4270	0.74088376	305785.00						
			7568.5910	0.73831944	305684.00						
			7572.7740	0.73529922	305582.00		2				
			7576.9770	0.73187551	305523.00						
			7585 4300	0.72922440	305547 00						
			7589 6980	0.72359612	305208 00						
			7593,9770	0.71807714	306038.00						
			7598,2750	0.71252104	305315.00						
			7602.5920	0.70549579	305573.00						
			7606.9280	0.69954396	305378.00						
			7611.2830	0.69265029	305425.00						
			7615.6570	0.68689223	305488.00						
			7620.0500	0.68092659	305297.00						
			7624.4620	0.67499882	305001.00						
			7628.8940	0.66889046	305246.00						
			7633.3440	0.66341014	305391.00						
			7642 2020	0.05/98296	305277.00						
			7646 8100	0.03332747	305125.00						
			7651.3360	0.64433064	305107.00						
			7655.8820	0.64130624	304968.00						
			7660.4470	0.63726602	303292.00						
			7665,0300	0.63282645	306701.00	01/21					~





Path Selection in FeS2 using Larch Select Paths in Galaxy





Multiple scattering of emitted photoelectron



Extended Analysis in FeS₂ using Larch Artemis in Galaxy

Tools 🏠 🔹	Larch Artemis generate Artemis projects from XAFS data (Galaxy Version 0.9.66+galaxy0)	☆ •	History +	+ =
search tools X	Execution mode		search datasets	* 7
1 Upload Data	O Parallel		Unnamed history	
JEP Method	Whether to execute Athena projects in parallel or in series.		🛢 251 kB	Q 10
Other Methods	Athena project file		•	
THER TOOLS	D D 8: Athena project of test.xmu	B	10 : Flattened plot of test.x	0/1
yMuonSuite	Normalised X-ray Absorption Fine Structure (XAFS) data, in Athena project format. If a collection of files is provided, these will be submitted and executed in parallel.		mu	
ile Conversion	FEFF paths file		9 : Edge fitting of test.xmu	0/1
ools Under Development	D D 4: FEFF paths of 1564889.cif		۲	
PyMuonSuite AIRSS UEP Optimise run UEP optimisation	Zipped directory of the FEFF paths.		8 : Athena project of test.x	• / 1
PyMuonSuite AIRSS Cluster run clustering for optimised structures	Image: Comparison of the second se		7 : Selected paths of CSV su	0/1
MuSpinSim Combine combine datafiles generated from MuSpinSim	File defining the fitting parameters as a 'guess' (to be varied in the fit), 'def' (defined by an expression evaluated throughout fitting) or 'set' (evaluated at the start of fitting, then left unchanged).		mmary of 1564889.cif	0.45
MuSpinSim Generator Generate MuSpinSim config from a structure	D D 7: Selected paths of CSV summary of 1564889.cif		mary of 1564889.cif	
file	File defining the scattering paths.		5 : FEFF input of 1564889.ci	0/1
Larch FEFF generate FEFF paths from XAFS data	Fitting Variables	۲	Ť	
Larch Select Paths select FEFF paths for XAFS data	Plot graph		4 : FEFF paths of 1564889.ci	0 / 1
Larch Athena generate Athena projects from XAFS data	No Whether to plot the data.		3 : CSV summary of 156488 9.cif	0/1
Larch Artemis generate Artemis projects from XAFS data	✓ Execute		2 : test.xmu	• / •
MuSpinSim Configure define simulation parameters	Using Larch, perform fitting on an Athena project file, originally from the input X-ray Absorption Fine Structure (XAFS) data file.		1 : 1564889.cif	0/1
Collection Operations	Optionally, plot the xµ data along with RMR and ChiKR plots for visual inspection of the fit.			
VORKFLOWS				
Il workflows	Citations: 🗈			
			× III	





Saving Full EXAFS Workflow for FeS₂ in Galaxy

Tools	XAS Workflow Adjustable	/ 13 16 💠 🕨
search tools	×	Name
Inputs	D 3: input.xmu O ★ 9: Larch Athena O ★	XAS Workflow Adjustable
Get Data	output (input) XAFS data file	Version
SPIN DYNAMICS	□ 1: structure.cif	4: Apr 6th 2023, 12 steps 🔹
MuSpinSim	output (input)	Annotation
FIND MUON STOPPING SITES	✓ 10: Larch FEFF 🗘 ×	
UEP Method	✓ 2: Absorbing	11.
Other Methods	atom Padiur Plating atom Plating Absorbing atom Plating Absorbing atom Plating Absorbing atom Atomic	These notes will be visible when this
OTHER TOOLS	Absorbing atom (text)	workflow is viewed.
PyMuonSuite	Paths file PEFF paths file	Specify a license for this workflow.
File Conversion		Creator
Tools Under Development	Radius (float)	Add a new creator - either a person or an organization.
Collection Operations	kmin	Tags
Workflows	🖍 5: kmin 🗘 🗙 kmax	۲
	kmin (float)	Apply tags to make it easy to search for and find items with the same tag.
	1 max	
	✓ 6: kmax Ø × ✓ fit_report (txt)	
	kmax (float)	
	//// Chikr (png)	
	rmin (float)	
	rmax (float)	
	- 100% + System Settings	
(>

- The tasks that we included in Galaxy are:
 - Processing and Normalization of Raw Data
 - Extended XAFS Analysis of Data
- Larch Athena: Galaxy tool for Processing and Normalization of raw data.
- Larch FEFF: Galaxy tool for the generation of paths in a material using FEFF.
- Larch Select Paths: Galaxy tool for the selection of paths generated using Larch FEFF.
- Larch Artemis: Galaxy tool for extended analysis of EXAFS data.



Reproducing results

The highly surprising behaviour of diphosphine ligands in iron-catalysed Negishi cross-coupling

Antonis M. Messinis, Stephen L. J. Luckham, Peter P. Wells, Diego Gianolio, Emma K. Gibson, Harry M. O'Brien, Hazel A. Sparkes, Sean A. Davis, June Callison, David Elorriaga, Oscar Hernandez-Fajardo & Robin B. Bedford

Nature Catalysis 2, 123–133 (2019) Cite this article

3395 Accesses | 26 Citations | 12 Altmetric | Metrics

Iron Negishi mechanistic NatCat 2018

Data supporting Nature Catalysis paper

Complete download (zip, 1.5 GiB)

Creator(s)	Robin Bedford, Antonios Messinis
Publication date	08 Jan 2019
Language	eng
Publisher	University of Bristol
Licence	Non-Commercial Government Licence for public sector information
DOI	10.5523/bris.1kp2f62x3klb02mfz2qymcmxmx
Citation	Robin Bedford, Antonios Messinis (2019): Iron Negishi mechanistic NatCat 2018. https://doi.org/10.5523 /bris.1kp2f62x3klb02mfz2qymcmxmx
Total size	1.5 GiB



Benefits of Galaxy

- Embed parameters used in the calculation in workflow invocations
- Potential to integrate with existing or new resources:
 - Data storage at the synchrotron
 - Compute resources already managed by STFC/UKRI/PSDI
- Apply workflow to thousands of measurements of the same sample
 - ...without requiring scientists to write their own Perl scripts

Workflow Type	Average Process	ing Times		Estimated		
	Normalise	Fit	Per Group	(3790 groups)		
Manual Novice	3 minutes	21 minutes	24 minutes	~ 63 Days		
Manual Expert	2 minutes	8 minutes	10 minutes	~ 26 Days		
Scripted Demeter	7.68 Seconds	13.56 Seconds	21.24 Seconds	~ 23 Hours		

Challenges

- Parameters sometimes missing from published papers
- Building on a Python library not executables
 - Have to decide on tool scope and write own Python scripts to execute
- Large parameter space requiring expert knowledge
- Difficulty in embedding interactivity to allow "trial and error"
 - Should we be considering Jupyter/interactive tools?

😼 🕂 Athena [XAS data	a processing] 🗸 🔨 🔨
<u>File Group Energy Mark Plot Freeze Merge Monitor</u>	Help
* <untitled></untitled>	Save A U I 🗹 Au Chloride, scan 1
Main window	Au Chloride, scan 2
	Au Chloride, scan 3
Current group: Au Chloride, scan 3 Dataty	rpe: xmu 🔲 Freeze
File /home/bruce/Data/ANL/20BM/2005.10/standards/aucl.00	003
Element 79: Gold V Edge L3 V Energy shift 0	Importance 1
Normalization and background removal paramete	rs
E0 11919.94738 💿 Normalization order	1 0 2 • 3
Pre-edge range -150.000 (O) to -30.000 (O) 🗹 Flatter	n normalized data
Normalization range 150.000 (a) to 869.893 (b) Edge step	0.600201) 🗆 fix
Rbkg 1.0 0 k-weight 2 0 Spline c	lamps Ione
Spline range in E 0 0 to 969.8772 0 high S	itrong
Standard None Energy-dependent n	ormalization E k R q kc
Forward Fourier transform parameters	E k R q
k-range 3.000 (0) to 13.955 (0) dk 1 window Ha	Plotting k-weights
arbitrary k-weight 0.5	
Backward Fourier transform parameters	
R-range 1 \odot to 3 \odot dR 0.0 window Han	ning V (E) ○ µ(E)
Plotting parameters	pre-edge line
Plot multiplier 1 y-axis offset 0	Normalized Normalized
	Derivative Derivative
	🗖 2nd derivative 🔲 2nd derivative
	Emin -200 Emax 800
Renamed aucl.0003 to Au Chloride, scan 3	

Conclusions

Galaxy can reproduce the tasks associated to the:

- 1) Processing and Normalization of Raw Data
- 2) Extended XAFS Analysis of Data.
- We have created 5 Galaxy tools for this:
 - Larch Athena: Galaxy tool for Processing and Normalization of raw data.
 - Larch FEFF: Galaxy tool for the generation of paths in a material using FEFF.
 - Larch Select Paths: Galaxy tool for the selection of paths generated using Larch FEFF.
 - Larch Artemis: Galaxy tool for extended analysis of EXAFS data.
- The full workflow can be extracted from a history of individual tool executions and then be saved, reused and shared or exported as an RO-Crate object
- Galaxy provides a single interface for all tools, and can submit jobs to HPC resources without users needing to worry about the details (for example writing Slurm submission scripts)
- We need to refine these tools by interacting with Catalysis scientists.



The Galaxy Platform: Applications to Catalysis Workflows

Theoretical and Computational Physics Group - SCD



LIBORIO

Data & Software Engineering Group SCD



GONZALEZ-BELTRAN



AUSTIN

UK Catalysis Hub



NIEVA DE LA HIDALGA









Software Sustainability Institute

Scie Faci COEOSC EuroScienceGateway