Welcome to Advanced Data Analysis (PHYS 605)

Prof. Claudia Gomes da Rocha

claudia.gomesdarocha@ucalgary.ca

Department of Physics and Astronomy
Faculty of Science, University of Calgary

UNIVERSITY OF CALGARY

The University of Calgary, located in the heart of Southern Alberta, both acknowledges and pays tribute to the traditional territories of the peoples of Treaty 7, which include the Blackfoot Confederacy (comprised of the Siksika, the Piikani, and the Kainai First Nations), the Tsuut'ina First Nation, and the Stoney Nakoda (including Chiniki, Bearspaw, and Goodstoney First Nations). The City of Calgary is also home to the Métis Nation of Alberta (Districts 5 and 6).

- **General information**

The content of this class is available online at D2L: **PHYS 605 L01 - (Fall 2024) – Advanced Data Analysis**

Check the course outline in the D2L course content for further details.

- **Course description (based on the calendar)**

A **hands-on course** that covers multiple methods of extraction of significant information from (experimental) data degraded by noise. These include parametric and non-parametric statistical methods; curve fitting; spectral analysis; filtering; sampling, convolution and deconvolution techniques.

Course Hours: H(3-0)

- **Textbook**

All materials will be available online at D2L: **PHYS 605 L01 - (Fall 2024) - Advanced Data Analysis**

Main Notes Source:
Datalogy: Measurement, Estimation, Transformation, Reduction, Analysis, Classification, and Modeling by Dr. Brian Jackel (UCalgary)

Other optional materials:
> > Alex Gezerlis, Numerical Methods in Physics with Python, Cambridge University Press, 2020.
>> C. A. Pruneau, Data Analysis Techniques for Physical Scientists, Cambridge University Press, 2022.
>> M. Bonamente, Statistics and Analysis of Scientific Data, Springer;
>> I. A. Rauf, Physics of Data Science and Machine Learning, CRC Press;
>> Cosma Shalizi, Advanced Data Analysis, draft available at
https://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/

There will be more materials being cited throughout the course…

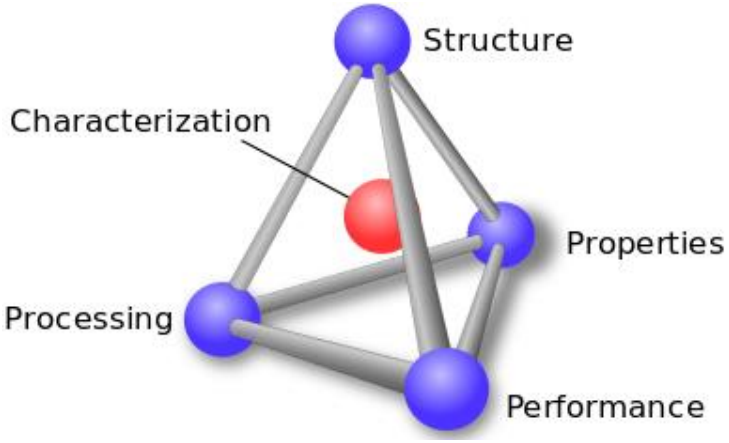
- **Course Syllabus (Fall 2024)**

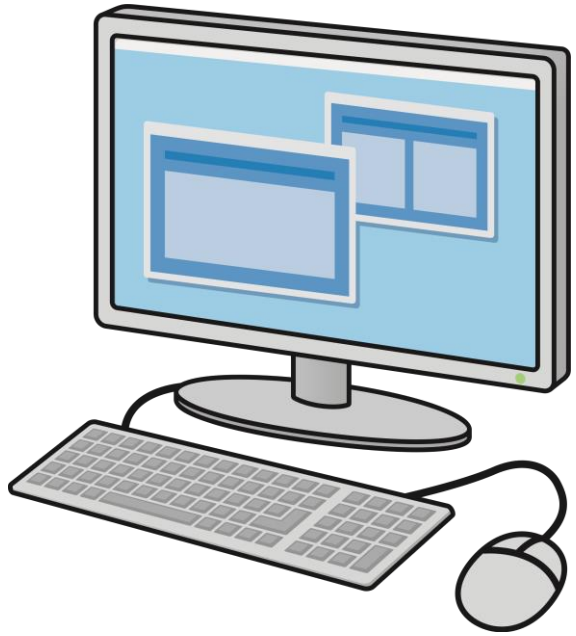
Probability, linear algebra, polynomials, least-squares, moments, covariance & correlation, information theory, linear inversion/retrieval, Fourier transform, Fast Fourier Transform (FFT), convolution, principal component analysis (PCA).

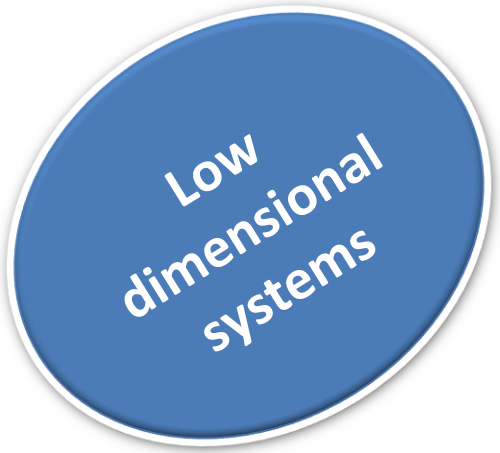
WARNING: Note that course content is subject to change since current computational procedures are highly dependent on hardware specifics, software versions and releases, updates, and successful compatibility tests.

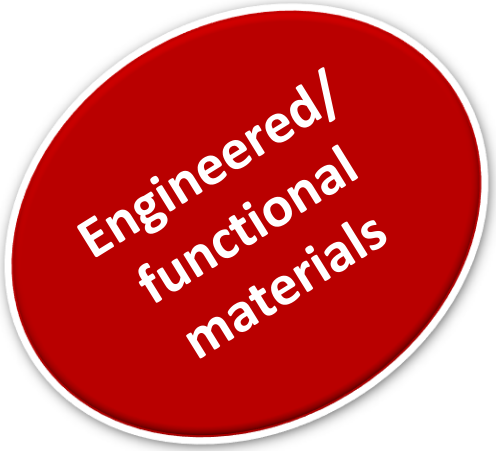
Complex (Nano) Materials Lab

- Theoretical condensed matter physics
- Complex systems, emergent phenomena
- Computational nanoscience
- Modelling Quantum Materials
- Brain-inspired Computing Systems









Complex (Nano) Materials Lab

Ghazaleh Gholizadeh
(MSc Student)

Marcus Kasdorf
(MSc Student)

Ciara Chisholm
(MSc Student, co-supervision with Dr. Brown)

Diego Simpson-Ochoa
(MSc Student, co-supervision with Dr. Nicola)

Gabe Komo
(Undergraduate final year physics thesis)

UNIVERSITY OF CALGARY
FACULTY OF SCIENCE
Institute for Quantum Science and Technology

Quantum Alberta

Complexity Science Group
Calgary, AB

Hotchkiss Brain Institute

NSERC CRSNG

Digital Research Alliance of Canada

MITACS Globalink

CMC MICROSYSTEMS

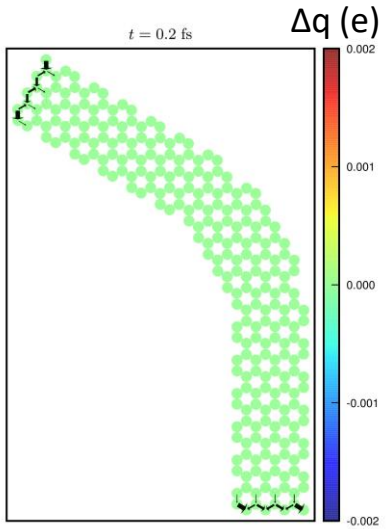
ALBERTA INNOVATES

Quantum City

University of Calgary

Complex (Nano) Materials Lab

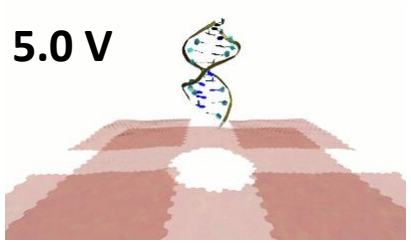
- Theoretical condensed matter physics
- Complex systems, emergent phenomena
- Computational nanoscience
- Modelling Quantum Materials
- Brain-inspired Computing Systems



Transient currents in nanoswitches

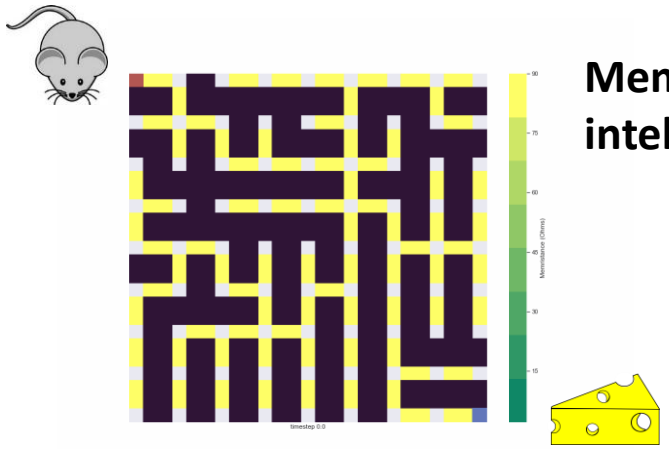
Nanoscale **7**, 8627 (2015)
Nanoscale **11**, 12296 (2019)

Space-time mapping of currents at the nanoscale

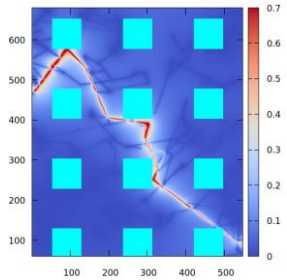


Conductive sensors

J. Phys. Condens. Matter **28**, 235001 (2016)
PCCP **16**, 3558 (2014)
Nano Lett. **13**, 1969 (2013)

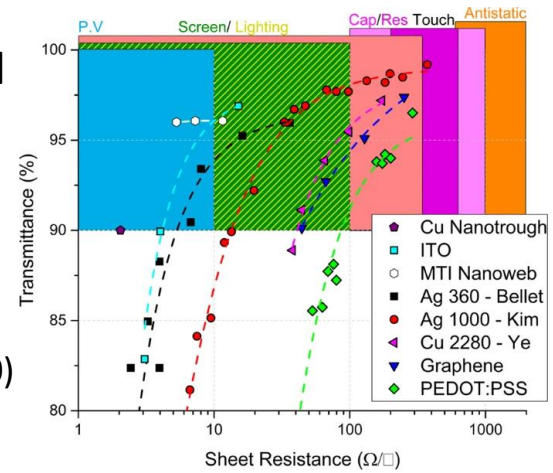


Memristive systems and intelligent materials



Transparent conductors and flexible display simulations

Nanoscale **13**, 15369 (2021)
Sci. Rep. **9**, 11550 (2019)
Appl. Phys. Lett. **116**, 251902 (2020)



Partenaire régional de l'
Alliance de recherche numérique du Canada

A regional partner of the
Digital Research Alliance of Canada

https://docs.alliancecan.ca/wiki/Technical_documentation

https://rcs.ucalgary.ca/ARC_Cluster_Guide

UNIVERSITY OF CALGARY

List of compute clusters

Name and link	Type	Sub-systems	Status
Béluga	General-purpose	• beluga-compute • beluga-gpu • beluga-storage	In production
Cedar	General-purpose	• cedar-compute • cedar-gpu • cedar-storage	In production
Graham	General-purpose	• graham-compute • graham-gpu • graham-storage	In production
Narval	General-purpose	• narval-compute • narval-gpu • narval-storage	In production
Niagara	Large parallel	• niagara-compute • niagara-storage • hpss-storage	In production

Cedar

Storage [edit]

Home space 526TB total volume	• Location of home directories. • Each home directory has a small fixed quota. • Not allocated via RAS or RAC. Larger requests go to Project space. • Has daily backup
Scratch space 5.4PB total volume Parallel high-performance filesystem	• For active or temporary (`/scratch`) storage. • Not allocated. • Large fixed quota per user. • Inactive data will be purged.
Project space 23PB total volume External persistent storage	• Not designed for parallel I/O workloads. Use Scratch space instead. • Large adjustable quota per project. • Has daily backup.

Node characteristics [edit]

Cedar has a total of 94,528 CPU cores for computation, and 1352 GPU devices; note that Turbo Boost is deactivated for the ensemble of Cedar nodes.

nodes	cores	available memory	CPU	storage	GPU
576	32	125G or 128000M	2 x Intel E5-2683 v4 Broadwell @ 2.1Ghz	2 x 480G SSD	-
96	32	250G or 257000M	2 x Intel E5-2683 v4 Broadwell @ 2.1Ghz	2 x 480G SSD	-
24	32	502G or 515000M	2 x Intel E5-2683 v4 Broadwell @ 2.1Ghz	2 x 480G SSD	-
24	32	1510G or 1547000M	2 x Intel E5-2683 v4 Broadwell @ 2.1Ghz	2 x 480G SSD	-
4	32	3022G or 3095000M	4 x Intel E7-4809 v4 Broadwell @ 2.1Ghz	2 x 480G SSD	-
114	24	125G or 128000M	2 x Intel E5-2650 v4 Broadwell @ 2.2GHz	1 x 800G SSD	4 x NVIDIA P100 Pascal (12G HBM2 memory)
32	24	250G or 257000M	2 x Intel E5-2650 v4 Broadwell @ 2.2GHz	1 x 800G SSD	4 x NVIDIA P100 Pascal (16G HBM2 memory)
192	32	187G or 192000M	2 x Intel Silver 4216 Cascade Lake @ 2.1GHz	1 x 480G SSD	4 x NVIDIA V100 Volta (32G HBM2 memory)
640	48	187G or 192000M	2 x Intel Platinum 8160F Skylake @ 2.1Ghz	2 x 480G SSD	-
768	48	187G or 192000M	2 x Intel Platinum 8260 Cascade Lake @ 2.4Ghz	2 x 480G SSD	-

- **How to get help?**

Office Hours: TUESDAYS, 9 am - 10 am, via Zoom ONLY (starting on September 17)

The Zoom link for office hours is available in our course D2L. You can only access our Zoom sessions with your UCalgary account!

Office hours are dedicated to students who have questions about our course content, general software settings, and assignments. Please, only attend if you have questions. Do not come to the office hours to just listen to the questions of other students.

For more specific software (or even hardware) inquiries, please, contact the UCalgary IT. We do not have the capacity to debug all programming implementations given the size of the class.

Students are responsible for setting up their own computers and coding platforms to complete the assignments and exams applied in this course.

The name of the course is ADVANCED Data Analysis, therefore an advanced background knowledge in programming, plotting, mathematical and statistical techniques is expected.

This is NOT an introductory computational course!

There will be NO introduction to coding!

- **How to get help?**

D2L Q&A Forum (main communication venue)

For off-class questions, we will be monitoring the D2L Q&A forum space which can be found by clicking on 'Communication' --> 'Discussions' in the top navigation bar. This D2L Q&A is an ungraded (off-class) open discussion forum in which students can post anonymously if they wish and users can provide attachments together with their inquiries. Note that I will only answer questions related to the course content, assignments, and other course materials posted in the D2L forums. Please, reserve email contact for more urgent matters or individual reasons only.

In this forum space, students can also help each other during the regular assignments! Only during individual exam settings, the forum will be temporarily closed.

- **How to get help?**

CONTACT EMAIL

Students can reach out to the instructor via email. Please, reserve email contact for more urgent matters or individual reasons only. To better manage the volume of emails, I kindly ask you to add the following keywords to the subject of your email as in the examples below:

P605F24: <brief statement about the subject or matter to be discussed>

Sign your email with <first name>, <last name>, and UCID. Make sure to use your @ucalgary.ca e-mail address in all official communication.

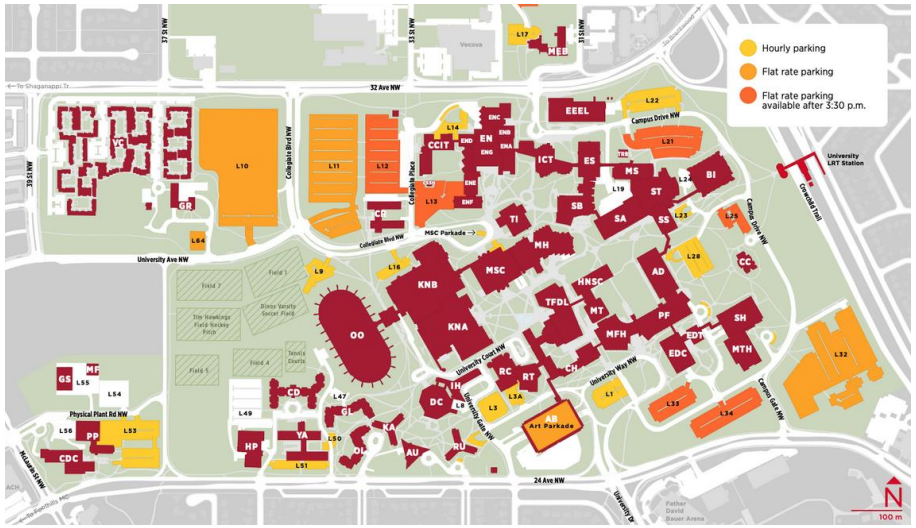
If you decide not to follow this recommendation, there will be a risk your email will be lost in my extensive inbox.

• Attending our classes

LEC 1 MWF 12:00 pm - 12:50 pm

MWF = Mondays, Wednesdays, Fridays

Classes will be delivered in person in SA 107 **and will NOT be recorded**.



EMERGENCY LECTURE ZOOM LINK:

An emergency lecture Zoom link may be created **ONLY** to be used in case I am unavailable to teach in person, but I am still sufficiently able to deliver a class.

You will receive an email prior to class in case I need to pivot to online delivery. Note that these (potential) online classes will NOT be recorded to alleviate our Zoom storage resources.

- **Computer/study room (ST 25/26)**

Other study/working spaces available are located in ST 25/26.

Note that I do not administer the computers in these rooms. For any issues with these spaces (and computers), please contact PHAS Tech team at

phas.edulabs@ucalgary.ca

Peter Gimby: pgimby@phas.ucalgary.ca

- **Course procedure and grading (check course outline)**

Lectures (Mondays/Wednesdays/Fridays)

- Mostly focused on course content delivery, but some will be dedicated to computational lab demos.

4 Regular Assignments (every few weeks, online submission, graded, 30% weighting)

- Several problems scattered throughout the term and based on the content covered in the lectures. Many of these problems will involve coding.

Midterm exam (06 November, in-class, graded, 35% weighting)

- Further information about midterm exam content will be given in class or posted on D2L.

Final project (due 13 December at 8 pm, online submission, graded, 35% weighting)

- Report exploring some data analysis techniques and applications (can be related to your thesis research).
- Further information about the final project will be given in class or posted on D2L.

- Each graded component is worth a total of 10 points.
- Do not forget to add your name and your UCID to your homeworks.
- If other materials (in addition to those provided in our course D2L) are used in your solutions, cite them in your homework.
- Some graded components in this course are open-book. You can consult any book and online materials, but you must write up your solutions on your own, in your own words, and with your own understanding. **DO NOT VERBATIM COPY!**

Visit Student Success Centre – Academic Integrity
https://www.ucalgary.ca/student-services/student-success/learning/academic-integrity

To ensure a fair and consistent learning experience for all students, the use of advanced **AI tools such as ChatGPT or Dall-E 2 is strictly prohibited for all academic (written/coding/creative/etc.) work**, assignments, and assessments in this course. Each student is expected to complete all tasks without substantive assistance from others, including AI tools.

- **If a regular assignment is submitted past its official due date, a 1-penalty point (out of 10) will be applied.** For example, if an assignment is due on 30 September but submission occurred anywhere between 01-07 October, a 1-point (out of 10) will be deducted from the assignment's final grade. ONLY 1 WEEK EXTENSION WILL BE GRANTED FOR EACH ASSIGNMENT. If an assignment is not submitted 7 days past its official due date, a grade of zero will be assigned.

- NOTE: this point deduction flexibility is **ONLY APPLIED TO THE REGULAR ASSIGNMENTS**. The due date of our **final project is SET and NO SUBMISSIONS will be accepted after its due date.**

- **Attention to extension requests!** These will not be considered simply based on "not enough time to complete" or "too busy". No extension requests will be considered after the assignment deadline.

More information about our assignments and exams will be given in due course.

Our first regular assignment will be released on September 16 in class and its due date is September 30 at 11:59 pm.

Delayed submissions of assignment #1 will be accepted until October 07 at 11:59 pm with 1 point (out of 10) deduction from the final mark. No submissions for assignment #1 will be accepted after October 07.

Final project (due 13 December at 8 pm, online submission, graded, 35% weighting)

Throughout the course, students will decide on their preliminary research topics that need to contain a data analysis component/technique covered in the course. Students will submit (on October 31 by 11:59 pm) a 1-page report describing the chosen topic and potential data analysis techniques required for the project. Students will then proceed to work on their research projects until the final report due date. Students will give a 3-minute presentation about their research projects (and achieved data analysis milestones) in class on December 02 and 04. Finally, a final written report is to be submitted by December 13, 8 pm.

Final project (due 13 December at 8 pm, online submission, graded, 35% weighting)

This means you can start working on your final project now. IMPORTANT: since the final project is a long-term assessment, NO EXTENSIONS on the due date for the oral presentations nor the final report will be granted. You have the opportunity to start planning on your final report today, therefore, do not let this become a "last minute" item.

Once I announce something publicly in a class, I cannot turn it back because now, everyone is brainstorming on their schedules and planning their time management schemes to meet the deadline. I cannot offer to one student what I cannot offer to the others.

Some decisions I can make/consider on a case-by-case basis, however, many others will need to be made in the context of a group (because you are in a group).

Regular Assignments Calendar

Assignment #1:
Start = Monday, September 16
Due = Monday, September 30

Assignment #2:
Start = Monday, September 30
Due = Tuesday, October 15

Assignment #3:
Start = Tuesday, October 15
Due = Tuesday, October 29

Assignment #4:
Start = Tuesday, October 29
Due = Tuesday, November 19

You can collaborate with peers in the regular assignments only. But note that the final work has to be your own! Make sure the collaboration is well-balanced!

- **Internet and Electronic Device Information**

➢ Students having difficulties with their own laptops, please, contact me in advance. However, note that students are responsible for the maintenance of their own devices and installation of the required software. Contact ucalgary IT for technical issues beyond our class scope.

➢ **BACKUP YOUR DATA!** Use (free) cloud storage services (Dropbox, Google Drive, One Drive, etc.) and/or your own local storage devices (flash drives, external hard drives). Do not risk losing your report's data due to computer malfunctioning! You are responsible for your own backups! Assignment or project extension requests due to not-backing-up data will not be considered.

In exceptional cases where I may consider extensions, I will request to see partial work done. Our assignments are long-term (not timed) graded components; therefore, some work is expected to be achieved as time goes by.