Burton Materials Data Lab
材料数据实验室

Group Guide
旅行指南

上海大学量子与分子结构国际研究中心
International Center of Quantum and Molecular Structures

Contents

1 INTRODUCTION

This group guide has two purposes, to help new members get started and to help existing members agree on the best way of working together. It is inspired by a similar guide that is available on line,[1] and many years working in different research groups all over the world.

The document is split into separate sections describing specific aspects of our professional lives at Shanghai University. It is arranged in the approximate order you might find the information useful. That is to say when you first arrive, Section 1: Shanghai will be probably of most relevance to you. Then understanding the group. Then conducting the work. Finally, creating outputs from your research.

I wrote this guide for someone with no experience working in research so I hope it can be useful to new arrivals at any level. It should give you an idea about the kind of expectations we have as a group and what you can expect from us in return. You definitely don't need to memorise anything in this document so don't worry. The guide will be constantly receiving updates, so please check for the most recent version. When you are finished reading it, please let us know if you have any suggested updates or changes.

The writing here is the level of English a scientist is expected to be able to understand in research so I deliberately do not try to make it easy. Please take the time to compare the English I use and the corresponding Chinese translation to improve your understanding as necessary. In this way, this document can also help establish a minimum level of English for all members and I will also use it to learn Chinese :)

[1] `https://hackingmaterials.lbl.gov/handbook.pdf`

2 SHANGHAI

Welcome to Shanghai! In this section I just write a brief guide to staying with us, understanding the university and finding your way around. However, the best way to find information will always be talking staff, students and me so feel free to ask :)

Shanghai University, (SHU, or 'Shangda'), is composed of 3 main campuses across shanghai. Baoshan Campus (which is where we are), Yanchang Campus and Jiading Campus. There are shuttle buses that can take you in between but be warned that these campuses are quite far apart and with traffic it can take quite a while. If you have classes or meetings at other campuses please be sure to leave plenty of time.

2.1 BAOSHAN CAMPUS

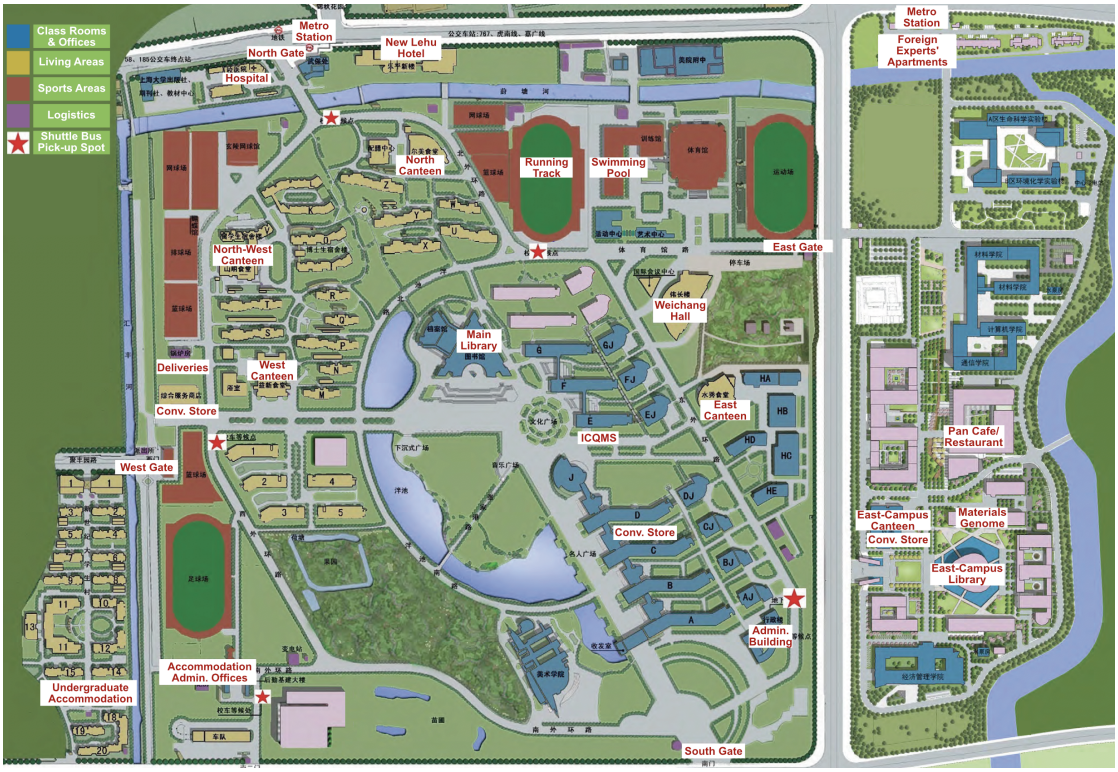
Baoshan campus has everything you need to have a pleasant stay in with us.

Note

Emergency service phone numbers in China are: a) Fire 119 b) Police 110 c) Ambulance 120

We are based in the International Centre of Quantum and Molecular Structures (ICQMS) in the department of Physics. Our offices are in building E of the main section of Baoshan campus (see campus map below). We also work quite closely with the Materials Genome Institute (MGI) in the east section of campus. Both ICQMS and MGI frequently host seminars and talks from domestic and international visitors so wherever you are based it's important you know these 2 locations.

The teaching buildings of the main campus are named in alphabetical order starting with A at the South gate to G by the main library. Each building also has a J annex to the east, which I'm told is short for 'joint'. These can be useful for orientating yourself if you get lost. Please



also take note that their are only 4 exits to a quite large campus. I once got stuck around the H-buildings thinking I could crossover the road directly to east campus but you can't.

There are 3 convenience stores around the 2 campuses (that I know of). Each are great for picking up snacks, drinks, pens, paper *etc.* but are not great for the kind of groceries you need for home living, for these kind of things there are alternatives just outside Baoshan campus, see below.

2.1.1 Campus Eating

There are quite a few canteens on campus, but the one closest to our office is the East Campus, just south of Weichang Hall and right in front of the eastern-most exit of our building.

The purchase of food is entirely cashless. If you want to buy something you have to use your university card (temporary or permanent university cards both work). In the canteen by our building you can add money to your card on the ground floor, to the left of the building as you look at the food counter.

Note

If you are vegetarian, vegan, have celiac or severe allergies the canteens might not be suitable for you. Talk to me and I'll help you find what you need.

The canteens are open 6am-6pm so try to plan your schedule accordingly. You wont find

food for a western-style breakfast here unfortunately but it's really good value for lunch and dinner. The North-east canteen is open until 8pm but the selection and quality is limited (I've heard that most students would opt to eat outside campus if they miss the 6pm canteen closures). All canteens are also open weekends if you are based around the university and want something quick and easy.

I recommend the east canteen for lunch, as they place plates of prepared food for you to take, which is both fast, easy and simple. However, for dinner I recommend the west or north canteens. If you go here and then up one floor, most of the counters have photographs of the dishes on offer and you can choose by simply pointing at the picture you'd like. These are then cooked fresh for you. Be aware that here you should walk directly to the till, place your order and THEN join the back of the queue. Of course if you already speak Chinese then you have no problem :p

Pan café is sign-posted as a book shop, but since the books are Chinese the most general use of the place is the food and drink. This is the only 'western' style place to eat on campus, serving pizzas, pastas, coffee and desserts. It's relatively expensive compared to the canteens but the quality is good and it is significantly less busy than any of the other places to eat during meal times.

2.1.2 University Library

The library does have some English language books available but they've proven quite difficult to find using the on-line catalogue. The best thing is to ask for help from a student here as they're usually familiar with the common textbooks.

2.2 Outside Baoshan Campus

The main area of non-university restaurants and shops is just outside of the west gate, in front of the undergraduate accommodation buildings.

Probably of most importance for non-Chinese members is the Walmart, on the right as you walk down the street from Westgate. This is a large supermarket that is great because it has signs in English for the aisles AND the products on the shelves. It can be surprisingly difficult to spot the difference between detergent and fabric softener, for example, when you cant read the Chinese!

Bicycles are a great way to get around campus and the outlying areas. I recommend not buying a bike because bicycles for public use are very common across Shanghai. There are a few different brands but by far the most common around the university is Mobike. Its a bicycle rental service that doesn't requiring docking bikes at specific locations, just lock the bike when you're done. This means you can hop on and off bikes wherever you please for a small fee and not have to worry

Note

Shanghai is relatively cool because of the Pacific breeze but the sun is still strong (we are at the same latitude as Egypt)! If you are easily burned be sure to keep sunscreen with you.

about forgetting where you put your bike or if it's in a safe location. If you are a foreigner you need to download the Mobike app and verify your ID with a passport, whereas locals can simply scan any bike with WeChat.

If you're looking to travel further afield, there are 2 metro stations right by the university- Shanghai University station and Nanchen Road station, besides Baoshan main and Baoshan east campus respectively. These are both on metro line 7, which travels right the way through Shanghai (see metro map).

Two or three stops to the north is a large and very beautiful park called Gucun Park if you want to pass a peaceful afternoon. Around March/April each year there is the Sakura, or cherry-blossom, festival that is definitely worth seeing.

A few stops south is the metro station Dachang Town, here is an IKEA store. Depending on how long you plan to stay this can be a very worthwhile trip.

While you can live perfectly well on campus, you'd be missing out if you didn't travel occasionally into central Shanghai. About 20-30 minutes south on line 7 (no changes) you can be in central Shanghai. Most often I find the Ji-ang Temple stop to be the best station to access the city but of course this depends on each individual trip.

It's important to note that the metro map I include here isn't the full version, just a clear image for the central stations. For example line 2 goes all the way to Pudong Airport, which

isn't shown in the figure.

2.3 Outside Shanghai

There are many obvious attractions in China outside of Shanghai, such as the Great Wall or the Forbidden Palace. I just take the time here to mention a few that are not so obvious and relatively nearby.

Suzhou is about 40 minutes away by fast train which is very reasonably priced if you book a ticket in advance (foreigners need their passports to pick up the ticket at the station). This means you can visit there just for the day without having to book a hotel.

Suzhou has a very modern, clean and wealthy city centre which isn't as crazy as Shanghai but, for me, the best things to do are more historical or cultural. These are the Lion Grove Garden and the Humble Administrator's Garden, which, in spring especially, are very beautiful places to visit.

If you are going for a longer stay, then 30 km south-east of the city centre of Suzhou is Zhouzhuang: a water village that has been called the Venice of the east.

Note

Pollution is an unfortunate aspect of life in China. The air in Shanghai is relatively clear but if you are travelling to somewhere like Beijing you may want to take precautions, especially if you suffer from a condition such as asthma.

Hangzhou is known as the city of heaven in China and is little bit further away than Suzhou but still only 1 hour by fast train.

Nanjing is also an important historical trade city whose name literally translates to 'South Capital' (whereas Beijing translates to 'North Capital').

Finally, if you are a fan of nature and the outdoors, there are many mountains to the west of Shanghai. Of these the Yellow Mountain (Huangshan) is perhaps the most famous.

3 GROUP

Welcome to Burton Materials Data Lab! We hope you have a productive and rewarding time with us.

I've worked in 9 different research groups in 6 different countries and I've found working culture and group synergy to be one of the most important factors behind successful research. Surprisingly though, it seems almost no one talks or writes about it. Perhaps most of this section will seem obvious to some people but since we are an international collective I think it's worth writing it down so we all understand each other.

3.1 YOUR COLLEAGUES

Ultimately, this group exists to produce internationally significant research. Even if you do not aspire to an academic career, the kind of outputs we aim to create will be of great benefit to you in the future. Almost all jobs will appreciate proof of the following:

- professional writing
- international languages
- public speaking
- numerical literacy
- data handling
- team work
- time management
- effective administration

As scientists we should also be open to new things, so if you are interested in alternate outputs for your time in our group please speak to me. For example, if you wish to create video's, websites, graphics, blogs, give public talks or any other form of science outreach I'd be happy to support you as much as I can. These can then be mentioned by me directly in a recommendation letter to future employers.

We should all commit to helping each other regardless of long term goals or how much time you intend to spend in the group. That way individuals benefit from evidence of professional development and the group benefits from being better established in the international research community. These effects enhance each other in turn, multiplying our collective success. [2]

Note

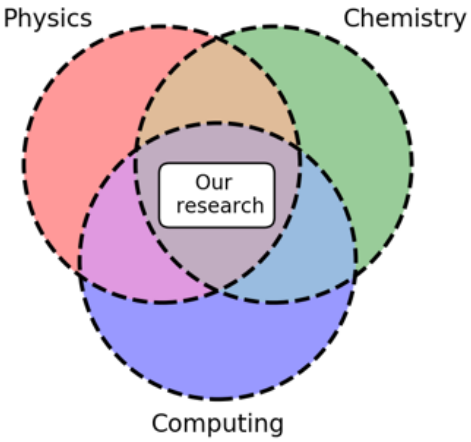
Margaret Thatcher studied chemistry at university and went on to become the first female prime-minister of the UK. The qualifications and skills you can obtain with us can be useful in almost any career path.

Everyone should be constantly learning as they work. Try to always have a notebook to hand so that you can doodle ideas or quickly jot down paper details whenever they may come up. The physics of the methods,

[2] V. Sekara, P. Deville, S. E. Ahnert, A.-L. Barabási, R. Sinatra, and S. Lehmann, "The chaperone effect in scientific publishing," PNAS, vol. 115, no. 50, pp. 12603–12607, 2018

the chemistry of the materials and the data science of the code packages all overlap in our area of research. No one person can understand all of these 3 areas, as they are individually developing rapidly in different directions, as shown in the Venn diagram below. It is important that we are honest with each other if we do not understand something, there is no shame in asking for help. However, adjusting numbers, altering figures or hiding errors will in no way be tolerated from anyone. If you are unsure, just ask :)

Please try not to be wasteful in your printing, drafting and note-taking. I won't stop you from doing whatever you feel is necessary, so this is something you have to judge for yourself. Even running unnecessary calculations prevents other jobs from running, costs us money and uses energy so take care.



Safety is incredibly important for a shared work space. Do not run wires across areas where people can trip and make sure to clean up any spills before someone can slip. If you must stay late at night, try not to be the only person in your office: if you fall and no one is around to help you, even a minor incident can become a major problem. There will be a minimum standard of cleanliness expected of everyone. We also share quite a lot of space with other groups so please be considerate. That said, I encourage you to decorate your desk and office space in whatever way makes you feel comfortable.

I want to specifically try and bring together students from different backgrounds but with complementary skill sets. This might mean you don't feel that you have a lot in common with other group members. For this reason, I will regularly suggest we go for dinner or perform an activity that is not work related. This will help people discover mutual interests from different perspectives and get to know each other better on a personal level.

Finally, always try to be kind to each other. Bullying, harassment or treating anyone differently because of their gender, origin or appearance will not be tolerated. Every member will treat every other member equally, with respect and professionalism. This extends to any guests, visitors or support staff you encounter. If you have an issue with another individual, first try explaining to them that you are not comfortable with their behaviour. If the issue persists, please bring it to my attention.

Note

Richard Feynman (Nobel prize in physics) famously said "if you think you understand quantum mechanics, you don't understand quantum mechanics." QM is just one of many areas we need to be familiar with, so don't feel bad if you don't understand everything.

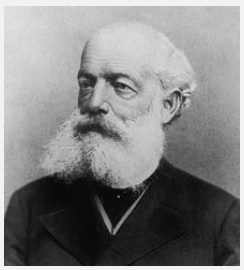
3.2 YOURSELF

Research has to be challenging. If you think a project is easy that's because you are not pushing yourself. Science exists as a series of increasingly difficult models, from the cartoons you were taught as a child right up to the physical reality that is impossible to understand. Where you fall on this scale depends on your level of training but if you are finding research easy you need to consider moving to a higher, more sophisticated level. Conversely, if you are finding research too difficult you can move one stage back until you feel you are ready to move forward again.

However, just because research must be difficult doesn't mean it has to make you miserable. I had a fantastic time during my early research career and I hope I can help you to do so too. Happy people are more productive and creative,[3,4,5,6,7,8] which are the 2 most important skills in research. Happy people are also nicer to be around, increasing the chance of having interesting discussions and establishing collaborations. So it is really to everyone's benefit that you are happy in your research!

Happiness is greater in more scenic environments,[9] and we are lucky to have a beautiful campus right outside. In England it is so rarely hot and sunny that people don't think twice about leaving the office when the weather is nice. As long as people can reach you, *via* WeChat for example, and we have no meetings scheduled then feel free to enjoy the beautiful scenery around Baoshan.

Note



In 1865 August Kekulé famously deduced the structure of Benzene based on a dream of monkeys dancing together. Sleep and rest have long been associated with effective problem solving!

I will always try to offer a selection of projects from which you can choose the one you would prefer to undertake. Think carefully about the one that motivates you most on a personal level. Things won't always go perfectly to plan, so believing in what you're trying to accomplish in the project can really help you stay motivated. I can think of almost no scientific reason why a project would need to be abandoned. Even if someone else publishes your exact project

[3] A. M. Isen, K. A. Daubman, and G. P. Nowicki, "Positive affect facilitates creative problem solving," Journal of Personality and Social Psychology, vol. 6, no. 52, pp. 1122–1131, 1987.

[4] F. G. Ashby, A.M. Isen, and A. U. Turken, "A neuropsychological theory of positive affect and its influence on cognition," Psychological Review, vol. 3, no. 106, pp. 529–550, 1999.

[5] B. L. Fredrickson, "The broaden-and-build theory of positive emotions," Philos. Trans. R. Soc. Lond., vol. 359, no. 1449, pp. 1367–1378, 2004.

[6] A. K. Anderson, P. E.Wais, and J.D. E. Gabrieli, "Emotion enhances remembrance of neutral events past," Proceedings of the National Academy of Sciences, vol. 103, no. 5, pp. 1599–1604, 2006.

[7] G. Rowe, J. B.Hirsh, and A. K. Anderson, "Positive affect increases the breadth of attentional selection.," PNAS, vol. 1, no. 104, pp. 383–388, 2007.

[8] A. J. Oswald, E. Proto, and D. Sgroi, "Happiness and productivity," IZA Discussion, no. 4645, p. 3, 2009.

[9] C. I. Seresinhe, T. Preis, G. MacKerron, and H. S. Moat, "Happiness is Greater in More Scenic Locations," Scientific Reports, vol. 9, no. 1, p. 4498, 2019.

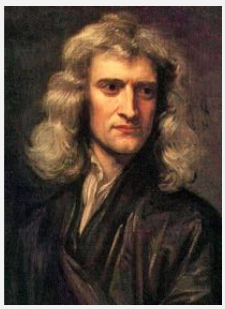
before you do, it can still be published and is valid as a body of work itself. If you are really having trouble please talk to me, but don't just try to switch projects because it's not as much fun as you first thought.

Manage your time and health properly. I guarantee you will be more productive working for 4 hours after an 8-hour sleep than you will working for the same 12 hours with no sleep, and other scientists see the same.[10] Do not think I will be impressed to find you in the office late at night or asleep under your desk in the morning (yes, I have had this happen). I once spent three days converting data tables and then on the third day thought up a line of code that could do the same job in less than 1 minute and contain no human errors. Having the patience, foresight and creativity that comes with being well rested can save a lot of time in the long-run.

There also is the danger of undoing good work that has already been done. I once deleted the entire headnode of a super computer, that hadn't been backed-up for 3 days. I promise you that every single user of that computer would prefer I had stayed in bed that day rather than go to work!

It has been shown that researchers had very high incidences of poor mental health compared to other professions. [11] The Institute Of Physics (IOP) has published a 40-page guide to building and maintaining well-being for researchers in physics,[12] so while I don't discuss self-care any further in this guide, there is a lot more to this than what I've written here. I will make the IOP guide available along with this one.

Note



Isaac Newton discovered gravity after an apple fell on his head. He was sat outside, in a garden, under a tree. Your frame of mind in conducting research is more important than sitting in the office all day.

Increasingly, companies in Europe, America and Australasia are moving towards a 4 day week.[13,14,15] I'm happy working 5 days of the week, but do you really need to work all 7? I won't stop you, but I encourage you to think more broadly about what 'work' means. If you go home to visit your parents and read papers on the way there, practice English via an app in your room, and then write a report on the train coming back, were you not working that day? As far as I am concerned I will not expect to see or hear from you on Saturday or Sunday. If you feel

[10] https://www.sciencemag.org/careers/2019/04/academia-hard-work-expected-taking-break-effort-well-spent-too

[11] S. Guthrie *et al.* "Understanding mental health in the research environment: A Rapid Evidence Assessment", Santa Monica, Calif.: RAND Corporation, RR-2022-RS, 2017

[12] S. Shinton, "Resilience Toolkit: A Physicist's Guide to Building and Maintaining Wellbeing", Institute of Physics, 2019.

[13] `https://www.theguardian.com/world/2019/mar/12/string-of-british-firms-switch-over-to-four-day-working-week?CMP=Share_iOSApp_Other`

[14] `https://www.nytimes.com/2018/07/19/world/asia/four-day-workweek-new-zealand.html`

[15] `https://www.huffingtonpost.com/entry/four-day-working-week-overwork-life-balance_us_5c360351e4b0f5aba7da3d5a`

you must work that's fine, but otherwise I hope you can enjoy these days in whatever way suits you.

Google doesn't implement a 4 day work week but they have a policy that is similar. It is called the 20 % time rule, and you can hear the former CEO of Google talk about it yourself.[16] This is a rule that employees should spend 1/5th of their time (i.e. one day out of 5) working on something unrelated to their specific job. Back when this idea first became famous a lot of people thought it was stupid and wouldn't last, however the company has gone from strength to strength and, even now, still win the title of most innovative company. [17]

If you are ever sick, please do NOT come to the office. I know some people think that this is a great display of commitment to their work, but I really don't. I don't want your illness and nor does anybody else. If you need time off, take time off. Similarly, if you are not sick but need to stay home because of, for example, a sports injury or pregnancy, it is possible for you to accomplish some work while outside of the office. We can arrange Skype meetings or alternatives so that progress in your research can still be made. However, overall you will be most productive at the office and, if you are able to do so, please try to be around campus between 9am-5pm.

Note

I've been told that my ideas are western but Confucius said 'To put the world in order... we must first cultivate our personal life.' Taking care of yourself makes sense, no matter where you are from.

Finally, money is important to everyone but scientists rarely talk about it. There is unfortunately very little money in science for students but I strongly believe that science should not just be available to students from wealthy families. If you need to get a job talk to me, I will not consider it a demerit. I already said I won't expect to see you outside of a 9am-5pm, Monday to Friday. In theory this leaves time for you to work evenings and weekends. I advise working a second job no more than 3 or 4 days each week, but this decision ultimately must be made by you.

There are opportunities available in the university and even our research centre where you can earn a little extra money. I think this is the most ideal way to do some work on the side, as it won't distract you too much from your research and you don't have to spend time travelling between different jobs *etc.* The best way is to talk to me and we can try and organise something based on your specific requirements.

3.3 Your Supervisor (Me)

As far as you are concerned my only job is to support you; so you never have any reason to lie to me or hide from me. 99 % of the time our relationship will only be work related but that doesn't mean you can't talk to me about other issues in your life. If your house burns down in

[16] `https://mashable.com/2018/05/11/google-20-percent-rule/#0uoJ09zf5kqU`
[17] `https://which-50.com/google-not-apple-named-bcgs-most-innovative-company-in-2019/`

the middle of the night and you have no-where to go you absolutely *should* contact me. Just because it is not work related doesn't mean I can't help you.

In the past I have worked with people who had suffered brain damage, were HIV positive, and had lost small children. No matter what your issue you will not surprise me BUT please understand that I am not a trained therapist. I will support you as much as I can but most of the time this can only extend to providing you free time to seek alternate help or to help yourself.

I will correct your spoken English, especially if there is scope for confusion in what you said. My first week in Shanghai I was asked if I would like to buy some beard (I can grow my own face hair thank you!) but they were really asking if I would like to buy some BREAD. I've also been asked if I was hungry for a snake (SNACK), if I went there by sheep (SHIP) or if I liked to eat stick (STEAK). This is certainly never intended to humiliate you, but small corrections over the years spent studying a masters or PhD can add up to a professional level spoken English by the time you graduate.

I've tried to encourage you to be creative but this doesn't help us if we never discuss your thoughts. Everyone should feel free to approach me with project ideas whenever they are ready. In principle coming up with a project is easy. You can simply follow the framework: Component -> Class -> Application -> Method. Any combination of suitable words or phrases can be thrown together in this way to make a reasonable sounding project, for example "Aluminium alloys for the construction industry by structure prediction" or "Titanium perovskites for water-splitting by machine learning." etc etc. However, just because it has a sciency title doesn't mean it's a project worth pursuing!

Note

Stephen Hawking was diagnosed with Motor-Neuron Disease at 21. Most people might have thought this the end of his work but he went on to have a long and productive career because he received the right support.

In Europe it is possible to obtain a PhD having published no papers and so they are more able to take on 'risky' projects. Here, we do not have this luxury so we have to be extra careful when thinking about starting new projects. While I encourage everyone in the group to cultivate project ideas please don't feel dejected if we don't decide to pursue it. Before a project can start there needs to be consideration of the literature and activity in the field to know which are the most feasible. Things to consider include: do the tools already exist to perform these projects or would we need to create our own? Would this study be of interest to the community even if the results are negative? Is there someone we know who we could ask to collaborate with us? *etc.*

It's an unfortunate truth that I will need more help from you than Chinese supervisors will need from their students. Even small tasks can be quite difficult in another language. It also means that you cannot always rely on me to warn you of arbitrary things like paperwork dead-

lines. I already heard of one student missing out on a PhD scholarship because their (foreign) supervisor didn't know about an early deadline. If you ever hear me say something that you think is wrong you MUST correct me. This applies for science as well as administration.

Note

Cecilia Payne was the first person to discover that stars are made of H_2 but was discouraged when nobody believed her. She described her results as spurious at the time but regretted it later saying: "If you are sure of the facts, you should defend your position."

I know some of concepts I discuss are not standard ideas for a Chinese research group and a lot of them can be interpreted as sheer laziness from an outside perspective. If any member of staff or other students give you problems because you don't spend 12 hours each day in the office, tell them to bring their issue to me and not complain to you. As long as I'm happy with your output and we're meeting the requirements of the university, it doesn't matter what others think.

Finally, I will keep a calendar that is up to date, and available for you to see. This means you can know when I'm busy and when I'm not - so if I have an upcoming work trip, for example, you might want to speak to me before I leave. This is better than me announcing where I'm going to be all the time. You are free to add your own work-related events in to the calendar so that I know when I can find you too, but this is not required. Furthermore, if you see that I am or someone else is participating in something you might be interested in, then you can always ask to join too.

4 Work

Research and studying are both necessary in academia; studying is time spent acquiring new knowledge, (*i.e.* reading papers) while research is time spent applying that knowledge to create new results (*i.e.* running calculations). If you only do research and never study, you will not succeed academically because you won't know how your results fit with modern technology and society. Similarly, if you only study and never do research you will also not succeed because you will never contribute to the community. The balance between these two things ultimately depends on the person and the project, so resist the urge to compare yourself to others.

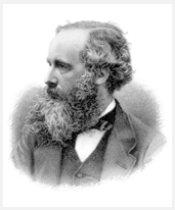
One should start a project by studying what has already been done, so I begin this section with a guide to understanding the literature. Then I discuss some basic details for performing research with software that is already available. Finally, I discuss automating analyses or processes, that is to say making software to perform tasks that pertain to your research specifically.

4.1 Understanding Existing Research

4.1.1 Papers

Research papers are published in journals. Usually, different journals cover different topics. However, there are different journals that cover the same topic, and these usually fight fiercely between each other to be considered 'better'. These days, how good a journal can be considered is usually measured with an impact factor, which is updated every year. It's simply a metric derived from the average number of citations of papers published by this journal. It should in no way be considered as a indication of trustworthiness for a study. There are bad papers in very highly esteemed journals and good papers in relatively obscure journals. Our training as scientists always takes priority over such simple metrics.

Note



Around 150 years ago James Clerk Maxwell said "It is of great advantage to the student of any subject to read the original memoires of that subject." As true then as it is today!

Journals can also have subcategories of themselves. For example the Journal of Materials Chemistry is split into A, B and C. These are not grades of quality or merit, simply distinct topics. Here, the Journal of Materials Chemistry A is for research in materials for energy and sustainability; B is for research related to biology and medicine; and C is for research related to optical, magnetic and electronic materials.

There are also journals that aim to cover ALL topics in science. You may of heard of the journals Nature and Science. These journals concern themselves with 'breakthroughs' in research, and are usually written and read very differently to more specific journal articles.

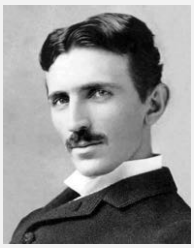
Each journal can have different types of publication:

- Research article. This is the standard publication format for a research project. it is the primary vehicle by which results are disseminated.
- Review. This is a publication in which one or a few authors try to summarise progress in an entire field of research.
- Perspective. This is a publication in which an author talks about their own research, usually over a long time frame.
- Editorial. This is an endorsement of another research groups publication in a relatively casual format. These exist because research papers can be quite difficult to understand or follow if you are not immediately familiar with a research area.
- Commentary. A commentary is an opinion on a topic or area of research. They are written at a level accessible to readers who are not researchers themselves but who are interested in research.

There is an enormous difference between these types of paper. You need to understand these differences so that you can better judge the literature. For example, if you find a research article that makes a conclusion and a commentary that states the opposite conclusion, this is not a 50:50 split in the literature. The article should be given more weight in your consideration because it is a study, whereas the commentary is an opinion.

Each Journal is made by a publisher and each publisher has several journals. This can be a little confusing because, for example, the journal Nature is published by the publisher Nature. Some people have strong feelings about the distinct publishing groups because some of them are not-for-profit and some of them are very much for profit.

Note



Nikola Tesla once said 'I was sorry to witness [Thomas Edison work] knowing that a little theory and calculation would have saved him ninety per cent of his labor.' Our research is not about writing papers, it is about quickening the process of discovery for society!

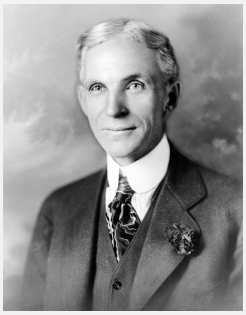
Understanding the difference between journals and publishers, and the types of articles on-line, empowers you to better judge the field. It is an unfortunate fact that people do try to misrepresent this information. The most common misrepresentation of this information I've seen is for the journal Scientific Reports, published by Nature. This journal was created to accept any research paper that is scientifically valid, no matter perceived importance. This means that there is little or no barrier for publication to this journal. The Journal Nature published by Nature on the other hand only accepts what is perceived to be the most exciting and impactful research in the world. As you can see, these two journals by the same publisher have an enormous difference in impact on average. However, more times than I can count, I've seen people reference work published in Scientific Reports and try to imply it is published in Nature. If I ever see you do this I will be super unhappy.

4.1.2 Preprints

A problem that existed for a long time in science is that it is possible for someone else to publish a result before you, even though your paper was submitted first. The peer-review and editorial process can be quite arbitrary, sometimes getting published quickly, other times taking months.

A solution to this came in the form of a website called arXiv (pronounced "archive") which acted as a place to upload your work before submitting it to a journal. Such a thing is called a pre-print and has no pay-wall or access restrictions for potential readers. This allows scientists to lay claim to a result or discovery as soon as a coherent paper has been put together rather than when a journal deems your work worthy of publication.

Note



Henry Ford was one of the most important people in the development of mass production, but he didn't patent a single machine, tool or process! While we may try to create patents for some projects, do not feel left out if there are not similar opportunities in your work. All research has merit.

The use of pre-prints divided the research community originally with most journals refusing to publish an article already on arXiv. Increasingly however they are being used, with new repositories such as ChemArXiv (chemistry archive) and BioArXiv (biology archive) appearing recently. Their use has even increased to the point that some people claim that journal publications are no longer necessary.

4.1.3 Patents

A patent is legal ownership of an idea or technology. If you discover something first and patent it, companies/universities must seek your permission before using your discovery.

However, during the process of a patent application, clerks will search through all previous discoveries and publications to check if you really are the first to make this discovery. If you have ever revealed information publicly about your invention, it can then never subsequently be patented. This includes everything from a presentation at a big conferences, right down to a tweet or facebook post.

Our group will always strive to make our work available to everyone where-ever possible. For example, this guide can be read by anyone on line. However, when it comes to scientific research it is better to keep that information within the group until the results are under peer-review or already published. Even if you have no plan to file for a patent, a journal usually won't consider publishing works already available elsewhere. Please do not post your results to social media and certainly never post somebody else's information on line.

As part of a patent application it must be proven to exist. For we theorists, that means we must collaborate with experimentalists to prove our predictions. No country will let you patent

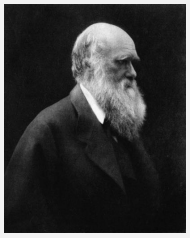
something based on a DFT prediction, unfortunately for us! Luckily there are a lot of amazing experimental scientists who are happy to collaborate, especially when it can lead to a patent.

4.1.4 Leaders in the Field

An important thing to know for a successful time working in research is who are the leaders of the field. These are the people doing good work we should follow outside of Shanghai. If you go to a conference these are the names you'll hear people speak about and it can help a lot to have a general understanding of what they do in the field. These days there are also tools like Twitter and Google alerts that allow you to keep in touch with or follow the work of these groups all year round.

In a lot of cases it's not possible to compete with these leaders directly because they have longer academic experience and therefore deeper knowledge. They are also likely to have larger groups, or more money, or alternate expertise or a combination of all of these. This does not mean they are enemies or that we have anything to worry about. There is an infinite amount of research to be done, we just want to make sure we're using our time effectively.

Note



"It is the long history of humankind (and animal kind, too) that those who learned to collaborate and improvise most effectively have prevailed." – Charles Darwin

- Materials Project/Pymatgen:
  - Gerbrand Ceder, University of California, Berkeley, USA.
  - Kristin Persson, Lawrence Berkeley National Laboratory, USA.
  - Anubhav Jain, Lawrence Berkeley National Laboratory, USA.
- Materials Theory:
  - Chris van der Waal, University of California, Santa Barbara, USA.
  - Fumyasu Oba and Yu Kumagai, Tokyo Institute of Technology, Japan.
- Experimentalists:
  - Hideo Hosono, Tokyo Institute of Technology, Japan.
  - Andriy Zakutayev, National Renewable Energy Laboratory, USA.
- Materials Modelling:
  - Suhai Wei, Fudan University, China.
  - Alex Zunger, University of Colorado, USA.
  - Aron Walsh, Imperial College London, UK.

  - David Scanlon, University College London, UK.

- VASP:
  - Georg Kresse, University of Vienna, Austria.

- FHI-AIMS
  - Volker Blum, Duke University, USA.
  - Matthias Scheffler. Fritz-Haber-Institut der Max-Planck-Gesellschaft, Germany.

This is not everyone doing good work in materials science. You will find that depending on your specific topic you will hear certain names repeatedly and see them recurring on your favourite papers. There is no such thing as absolute science, and the human element is important for all research. Learn these names, contact people, speak to colleagues and you will find your own sense of community in research and establish your own identity within that community.

4.2 Performing New Research

Everybody has a research project that is unique to them. I don't try and cover the science in this guide, there are plenty of textbooks available for that. What I want to describe here is a framework that can allow anyone to perform their own research, even as it changes over the course of a project.

Virtually everything we do will take place on a computer, even reading papers is almost exclusively done electronically. I will try hard to maintain a computer structure that allows all of us to create accurate and reliable results. It's important that we keep on top of the tools that we use, which other tools they interact with and what they require to work so that we don't get swamped with competing codes from all over the internet. This framework will naturally change over time and if you have any suggestions for change you are more than welcome to make them.

Note

Oliver Heaviside was a physicist who worked with concepts before they were formally proven. When asked why he said, "Shall I refuse my dinner because I do not fully understand the process of digestion?" You too shouldn't wait to understand all aspects of VASP before you run calculations.

Having a recommended framework of computer codes offers several benefits. Firstly, we can develop a deeper understanding of an individual tool as a group, rather than try to keep track of multiple implementations of the same theory in similar codes. Secondly, it may be necessary to write codes that wrap around specific tools, either to process the output or generate input files etc. These wrappers can be more easily shared, developed and used by the whole group if there are fewer to work with. Finally, it is inevitable that we will have young students start with

us who are keen to contribute to our research. If we have specified tools for common outputs, then even the newest students can begin creating impactful research that is consistent with previously generated results without necessarily understanding all of the theory behind it (of course, everyone is strongly encouraged to study any methods they use as their research progresses).

The one drawback that comes with relying on only one code for each output/property is that if there is a bug in that implementation then we may not be aware of it because we are not comparing across multiple codes. We must always be on the lookout for strange and unusual behaviour from our simulations, which is where our understanding of materials science becomes crucial.

4.2.1 Your Computer

You are welcome to install or access any software you feel will support your research and learning.

4.2.1.1 Slack

I know that WeChat is ubiquitous in China, and of course you are welcome to use this for your personal communication as much as you like. However, for discussions and/or questions related to science and research please use Slack. Slack is a communication package designed for teams working in different but related areas. In private communications you are of course welcome to use whatever language you like, but in the open channels please use English only (this is a great place to practice your written English in an informal setting). Using English in the channels allows me to keep track of progress in the group and any international visitors or future group members can look back on past discussions and understand how issues were resolved or how things work. For exactly this same reason, please use Slack whenever seeking advice or help rather than sending an e-mail, as these are much more difficult to share, and engaging with the entire group is more likely to yield the best advice.

4.2.1.2 LaTeX

The end goal of any project is a paper, report, or document that describes your research. It is likely that when you start in the group you may only be familiar with Microsoft Word. This is totally understandable and widely used the world over. You can continue to use this for as long as necessary, but it is strongly encouraged that you begin to teach yourself LaTex to make professional PDF documents as soon as possible, especially, if you are writing a PhD thesis. I wrote my PhD thesis

Note

William Herschel was a musician who's hobby was astronomy. He discovery several moons and a new planet! He had no supervisor or deadlines to force him to do good work but he's still remembered to this day as an eminent scientist.

not that long ago and almost everyone in my graduate school who chose to write their thesis in Word encountered problems eventually. Word is fine when your documents are a few pages long but when you are looking at making 150+ pages including, cross-referenced figures, tables, citations, equations and all the rest, it quickly becomes unmanageable.

LaTeX also allows for the use of templates and standard formats. This means we can create and maintain group templates for masters and PhD theses and other document types. Of course, you are welcome to modify your own document as you wish, but for the most part, having a minimum standard of quality and clarity will significantly simplify the process of thesis writing, and help minimize stress towards the end of your project.

4.2.1.3 ANACONDA

With the kind of experimental coding we will be doing you don't want to be breaking any core functions of your computer (think of all your precious data!).

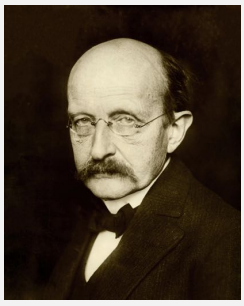
So the best thing to do is to use a package manager/virtual environment package. This will allow you to make virtual computers within your actual computer (called virtual environments). This is useful because you can have several configurations (for example one virtual environment for Python2 and one for python3) and if you break anything it doesn't harm your computer operating system. Conda is the recommended package for doing this.

Anaconda is a tool to create virtual environments, which are like smaller computers within your computer. The reason this is useful is that it can create an area that is standard for programs, whether you are using windows, apple or linux machine. Because of this a lot of research code writers develop their codes for use within anaconda and in some cases even use anaconda as a platform for distribution. It's also useful for us to work inside virtual environments because if you make a mistake and break a code or cause a software error, we can simply reset the environment. If you do something damaging and it's not inside an environment you can cause serious harm to your computer and lose work.

4.2.2 THE CENTRAL COMPUTER

Calculations can require a long time and a lot of computing power. For this reason we have a very powerful central computer that is shared within ICQMS. This computer, also called a cluster, is optimised for running these difficult calculations all day every day. It's possible to find software to interact with this computer in a similar way to your own computers, but it's faster, cheaper, easier and more powerful to use terminal languages. I strongly encourage you to start learning these methods of interacting with computers as soon as possible as you will need

Note



Max Planck (Physics Nobel Prize) once stated "We see in all modern scientific advances that the solution of one problem only unveils the mystery of another". At some point we have to stop and say, ok this level of theory is enough to describe what we seek.

them to start running calculations. As your skills develop they will also allow you to write your own codes to automatically perform jobs (see next section).

4.2.2.1 VASP VASP is the central engine of our research. It is the tool we use to turn atomic structure into materials properties. It is not the only package that can do this and it is certainly not perfect (I wouldn't even say it is the best available). The reason I, and many others, choose to use VASP for research is that is fast and easy to use, this means the barrier to entry is low for students and researchers outside of quantum physics.

Other *ab initio* packages like VASP include but are not limited to Quantum Espresso, CASTEP, ABINIT and GPAW.

When you are a little more familiar with coding, there are ways to generate the files for VASP automatically. In the beginning its best for you to play around with the files yourself so that you understand better the machinery of your research. The absolute minimum you need to know to run a calculation is that there are 3 files you must prepare INCAR, POSCAR and POTCAR. These files are so named because a long time ago computers used to be given information on punch-cards. So INCAR- comes from INput CARd, POSCAR comes from POSitions CARd and POTCAR is POTential energy CARd.

A basic INCAR contains the option flags for the calculation and will look like this:

INCAR

```
ISIF = 3              #Allows structure to relax to its lowest energy

IBRION = 2            #Sets method to optimise relaxation

NSW = 100             #The maximum number of ionic steps.

ISMEAR = -5           #Sets treatment for partial occupancies orbitals
```

VASP will run a calculation with this INCAR but you should know there are more than 300 options that can be set in this file! VASP has an online wiki that explains these features in more detail and is updated with the software. So for further details I'll refer you to the website. [18]

POSCAR is the file containing the atomic positions and is split in to three sections. First the repeating unit cell is defined as a matrix, then the number of atoms is specified, finally the 3-D positions are listed individually as coordinates. A file will look like this (the primitice cell of silicon):

[18] https://cms.mpi.univie.ac.at/wiki/index.php/The_VASP_Manual

POSCAR

```
Si2
1.0
        3.8669745922          0.0000000000          0.0000000000
        1.9334872961          3.3488982326          0.0000000000
        1.9334872961          1.1162994109          3.1573715331
   Si
    2
Direct
     0.750000000          0.750000000          0.750000000
     0.500000000          0.500000000          0.500000000
```

The line that says 'Direct' can either be this, which means the coordinates are relative to the unit cell or 'Cartesian' which means the coordinates are defined absolutely.

Finally, a caculation needs a POTCAR file. VASP comes with a POTCAR file for each element and you just need to bring the files for the elements of your system together for the calculation. Be careful because the order of the elements needs to match the order you specify them in the POSCAR. I heard of one student who found out at the end of her project that the POTCAR was in the incorrect order, making all of her conclusions incorrect.

These three files need to be in a folder together, there can be other files in there too but i recommend each folders be kept clear of anything except the relevant calculation files.

4.2.3 External Resources

The Material's Project is a database of many thousands of materials that have already been simulated with VASP. Even if you are not looking to directly interact with the Materials Project it has become an important point of reference for scientists in our field.

4.2.4 Summary

I've talked about the bare-minimum in this guide but you should know there are a lot more tools out there that we use as a community and many of these interact with or even depend on each other. The best pictographic representation of a overall infrastructure for the group that I've been able to come up with is shown in Figure 4.1. You will probably only use part of this work flow so don't worry about keeping on top of everything.

4.3 Automating Research

In section 2 I talked about taking care of yourself so that you can be rested, creative and alert. It is in coding that I've found this most critical. In what we do there is an inevitable degree of waiting, as calculations run or codes compile or updates are installed. Take this time to begin thinking about how you interact with the computer could be automated by the computer.

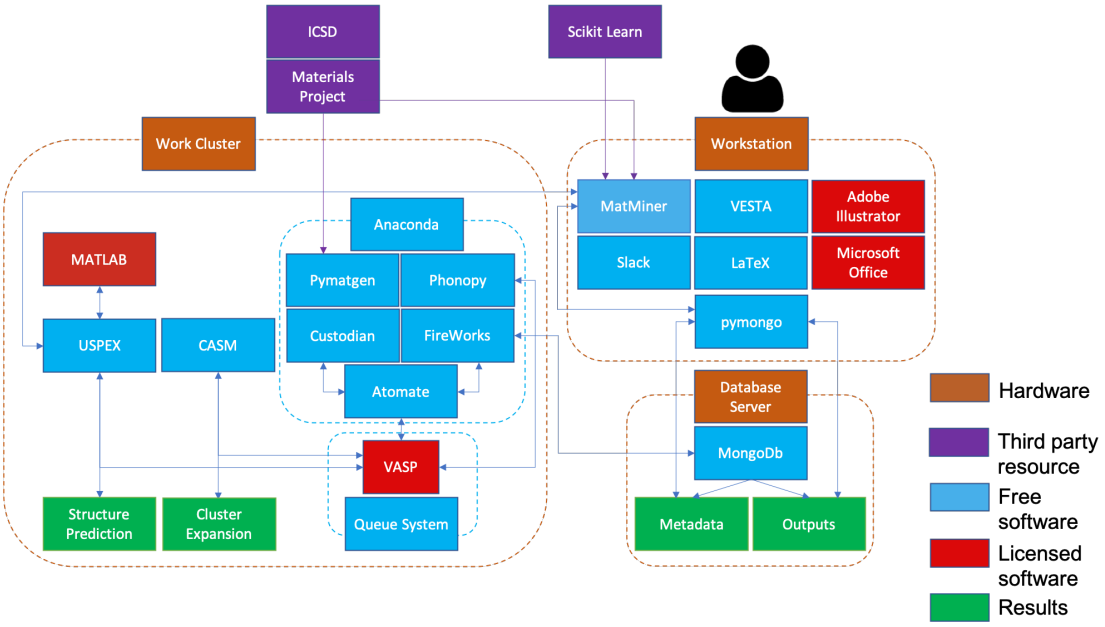


Figure 4.1: A schematic of the ideal software infrastructure. Each box indicates a code, package or computer and the arrows indicate with which other code they interact.

There are 3 forms of language you are going to need to learn to be successful in materials modelling. This can seem overwhelming if you have spent your education focusing on the science and not on the computing side, but be patient and you will begin to unlock the potential of computers.

Firstly, BASH is the language that you use in the terminal to navigate, move, remove, rename, copy, paste files *etc.*

Secondly, VI or VIM is a powerful text editor that comes pre-installed, and can be opened within terminal after navigating to the suitable locations with BASH.

Finally, Python is a powerful, human-readable, free and very popular programming language. Most of the tools we do use will be written in python.

The good news is that these 3 languages cover the majority of the tools we use now or ever will in the future. The reason why we need to learn all 3 is that PYTHON files, are edited in VI after finding them with BASH.

Note

I'm not the first to see the importance of creativity in research. Aida Lovelace said 'Imagination is the Discovering Faculty, pre-eminently. It is that which penetrates into the unseen worlds around us, the worlds of Science'

4.3.1 BASH

Firstly, whenever you open a terminal, you have to be located somewhere i.e. inside a folder, also called a directory. We usually think of folders as being arranged vertically with subdirectory being below or deeper; and superdirectory begin above or higher the folder in which you are located.

Starting with the basics you almost always use a command with a target for that command. For example 'cd Documents' means change directory to the folder called Documents.

Commands like cd also have more advanced functionalities built into them that can be accessed usually by either typing '−' followed by a letter or '−−' followed by a word. To find out what these are you can type 'man' followed by the tool in your terminal (man is short for manual), for example "man cd". The most common option of this kind is probably '-r' where r stands for recursive. if you want to copy a file you type 'cp file-name new-file-name'. If you want to copy a folder however you use the recursive version: 'cp -r folder-name new-folder-name'.

There is grep, which searches for a string in a file. For example grep toten OUTCAR will find the value of total energy in the output file from VASP by matching the string 'toten'.

A useful example of this is the find command. This searches for file by name if you use "find . -name". This has saved me countless times when I've forgotten where I saved a file.

Finally, as you become more confident you can start to combine tools together using the pipe '|' symbol. For example you can use find to retrieve a file and pipe to grep to then search that file.

Here are just of few useful snippets of code that I've kept handy for a long time now:

Count the number of times <string> occurs in a file:

```
grep '<string>' wc -l <FILENAME>
```

Find the 10 largest files in all directories beneath current location:

```
find . -type f -ls | sort -k7 -r | head -n 10
```

Delete all WaVECAR files in all directories beneath current location:

```
find . -name "WAVECAR" -exec rm {} \;
```

Look for <string> in all files in all directories beneath current location:

```
grep -rnw './' -e '<string>'
```

These relatively simple commands can simply be copy and pasted from here when you need them.

It is possible to have codes that run without your input at all. A cron-tab is

Now you should know enough to navigate a computer and manipulate files and data with bash. I won't go any further into coding with this guide but there are plenty of resources

available that can teach you way better than I can. I recommend you transition to python for learning more advanced concepts, the website code-academy is one that I think is quite universal for students in the field.

Once you're good enough to write your own codes you can have them set-up calculations for you, monitor jobs in the queue or even clean your hard-drive. In the past I've had several codes operating at the same time that complimented each other.

While it is possible to things like try and solve equations in bash, I recommend you use it only for moving around and copying/pasting/moving/renaming files *etc*.

4.3.2 Vi

Vi is a command like cd, but when you use it to target a file you open it and can edit the text.

Vi has two modes insertion mode and command mode. It begins in command mode, but you can trigger insertion mode by pressing 'i'. [ESC] returns the editor back to command mode, but if you are a beginner I suggest only using command mode to save ':w', save and quit ':wq', or quit without saving ':q!' (executed as soon as you press the return key).

In insertion mode you can move with the arrow keys, copy and paste, delete and type similar to how you would in any other document editor you may have used. Once you're comfortable with this you can start using command mode to do powerful things. For example ":sort" sorts all lines of a file alphabetically and ":sort u" sorts all lines *and* removes any duplicate lines.

There are a LOT of powerful things you can do with Vi, I wont discuss any more here but trust me when I say that vi is worth the time and effort it takes to learn how to use it.

4.3.3 Python

python is also a command that you can use from the terminal. If you type python by itself the computer will begin a session in which python is now the language rather than BASH. If you type python and then a file written in python, the computer will execute that file.

4.3.4 Github

Scientists recognise that there is great benefit to opening their work to the community *via* the internet, as I said earlier, you should never try to do everything yourself. However, you can't simply allow anyone to edit your codes as they please or they could introduce errors, change the way you wanted it to work or break the code entirely.

Github is a popular software package and website whose sole purpose is version control to address this need directly. You upload your code to Github and people can 'pull' the code to their computer and 'push' changes for you to consider. If you don't like the changes you can reject them, but if you do approve you can accept.

You can even use Github just by yourself. Sometimes even you can break your own code trying to incorporate changes (this happens to me lot!). If you do, you can use Github to revert to a previous version of the code with no harm done.

Github can get a little complicated as codes can be branched and merged. One code can have several branches, each with different versions to the main branch. This kind of complexity is usually only encountered for very popular or ambitious codes however, so you don't need to worry about such things in the beginning.

Note

Marie Curie was Polish but worked in France. She said, 'science is essentially international, and it is only through lack of historical sense that national qualities have been attributed to it'. The internet is a powerful resource to connect us to the globe - we must use it.

But this does highlight some of the best practices for coding. If you have 3 things you want to fix in a code, don't try and fix them all at the same time. Fix one, then push to git. Then fix the second, check the first still works, then push to git. Finally fix the third, check the first and second and push to git.

I also find commands such as 'git status' to be useful. It lets you see the difference between the files on your computer and the files of the last upload to Github.

4.3.5 Data Management

Once you are automating processes with coding, it is easy to generate massive amounts of data over the course of a project. It is important that we manage our outputs properly or we will be out of storage space in no time. This can be as simple as NOT outputting wavefunctions and charge densities from calculations where they are not needed, but certainly we must be careful to not delete data we rely upon in drawing later conclusions.

I think its important to understand the difference between data and metadata. 'Data' are the results from your research, but metadata is the pertinent information related to these results. For example, the results file is data, whereas the location of the results file is metadata. Metadata includes but is not limited to: date, file size, file ownership and permissions *etc.*

As an example of this I once found that the total energy for a calculation was slightly different depending on the number of cores over which I parallelised the calculation. In theory this should have made no difference, and I use it as an example because almost no one records the number of cores over which a calculation is run. In the end it turned out to be such a small difference as to not affect the results of the study overall, but we should still try to record as much information about our work as possible so that we can trace back any unexpected issues.

My suggestion for dealing with this is as follows. Each potential research paper should be a folder on your computer. Any result that you present in the paper should have the corresponding calculation from which that result came in said directory. Once the paper is published (AFTER all of the peer review etc), compress the directory with all of the relevant calculations in it (and perhaps a little plain text explanation) with the 7-zip method. This is a very thorough compression that takes a long time but allows for a large reduction in size.

We'll have a central, backed-up storage for keeping these, and then you can clear out the calculations for that project and free up some space. This kind of data preservation is not only good practice but usually a required condition for research funding so please take care with this.

5 Outputs

This section is dedicated to creating impactful research outputs from your understanding of the field and new results. In theory, there is no limit to the ways you can share good science with society, but there are unfortunately only a few ways you can do so and receive official recognition.

Note

Margaret Hamilton standing next to the software she and her team at MIT produced for NASA's Apollo Moon missions. This is an amazing feat of coordination, teamwork and data management... not to mention science!

5.1 Articles

The formats of research articles vary depending on the journal we will submit to. It can be frustrating because if the paper is rejected it might need entirely reformatting before submitting to the next journal. However, regardless of which journal we submit to, you should have a master-draft article containing the following sections: Abstract, Introduction, Methods, Results, Discussion, Conclusion.

I recommend that you start collecting pertinent data from the very beginning of the project. The first few weeks of any new project should be spent reading the literature. Similarly, any document should begin with a literature review. Take the opportunity to note your thoughts and impressions as you read in a LaTeX document. Do you spot any conflicts in the literature, are there any ideas that everyone seems to agree upon, what is the range of values you see reported for the same materials property? I find a useful exercise is to see how far back you can trace the topic of your study. How have things changed since then? Before you know it, you will have a Literature Review on your hands.

Any time a calculation finishes, put the result of that calculation in a document along with the method you used and where you stored it. I guarantee there will be times where you forget from where you got a specific number. Under no circumstances can we publish results that we can't trace back to the completed calculation files. This may mean running calculations again if they are misplaced.

Keep the calculation details by them self so that they can be added to as progress is made. The details should be a stand-alone, factual record of inputs and outputs. The inputs will form the methods section and the outputs will form the results section.

Note

You are always free to challenge or contradict me. J. Oppenheimer (father of the atomic bomb) once said "there are children playing in the street who could solve some of my top problems in physics". Fresh insight is invaluable in research!

For a methods section there are also a few other rules you should follow. If you make any modifications to anything they must be described. If you use any software you should use the version number or release and if you write any wrappers or stand-alone codes yourself they should be made available along with the paper. Anybody should be able to sit down with your paper, assemble the tools you used, follow your method and get the same results without having to access another paper.

There is no room for opinion in either of the methods or results sections, but as we start to see many different results side-by-side we can start to draw on our training as scientists to infer underlying physical phenomenom. This is what the discussion section is for and should include observations such as do all the results from the same method follow an obvious trend? Why do you think changing that one flag in the input file changed the result the way it did? Are there other materials that behave similar to this one?

Do not worry too much about your English. One of the benefits of a British supervisor is I can take care of that but its good for you to practice by yourself. No matter how bad you think a draft might be, do not EVER pay someone else to write an article or essay for you.

It's very important that everyone receives credit for work they have done. If someone contributes to a project such that their result is included and they are the best person to explain it, they need to be an author. If their result or idea is included but no explanation is necessary, then it might be acceptable to simply credit them in acknowledgements. I've encountered a lot of strange ideas about authorship in the past but ultimately there is no diminished value of your work from sharing your name with others on a paper. Do NOT try to do all the work yourself to keep all the credit for a paper! I promise you your work will be of higher quality overall by involving specialists where needed. Finally, it's not possible to go back and add someone's name to a paper after publication and missing someone off a paper they worked on will generate an incredible amount of ill-will for the group from important collaborators and colleagues.

Respect in the community may not seem immediately useful. But I know a scientist who, at the time of writing, has five publications in Nature sister journals that are all reviews, commentaries or perspectives entirely because their opinion is respected and their expertise recognised internationally. There's no way a group can elevate to this kind of status if they are known for shoddy work, falsified results or stealing credit!

5.1.1 Citing works

Pre-prints CAN be cited in your research but the exact wording or format of the document may have changed by the time it is published to a journal. The arXiv website will say if I paper was

published after being uploaded there and if possible you should cite the journal publication rather than the preprint.

I encourage students to cite text books as well as journals.

The term DOI is an acronym of Digital Object Identifier. This is a tag or code that should always be used rather than a website url. These days not just papers but also preprints, presentations and computer codes can have DOI's. Figshare, Zenodo and the preprint servers provide DOIs for citations. Usually when a code is released it is accompanied by a paper describing how it works. Most often this paper is the thing the authors would like you to cite if you use their code.

Copying from other published items (including your own!) is illegal. It is called plagiarism in English. While this means you can't copy sections or paragraphs i actively encourage you to copy lines from papers as long as you reference them properly. If you do not make any change to a line from another work it must be included as a quote, otherwise, as a simple citation is fine. Now more than ever its super easy to find if something has been copied from published work (even just using Baidu or Google) so don't waste our time with this.

5.1.2 Submission

Each article is written and submitted with a cover letter explaining why the study deserves to be published and why it was submitted to that journal in particular. The paper will then either be rejected based on the editors opinion or sent for peer-review.

Note

J. K. Rowling's 'Harry Potter' manuscript was rejected from the first 12 publishers to which she applied. Now it is a franchise worth billions! It can take a lot of time and patience to get published no matter how good your work.

Peer-review is the process by which other scientists judge your work anonymously. If they consider your work is scientifically sound and of interest to the community, they will recommend it be published. The paper is usually sent to 2 or 3 scientists and then the editor will make a decision based on the reviews.

This can be quite a difficult process because most of the reviews will come back with criticisms of the work. While it can be a little demoralising to receive such kind of feedback, take heart that incorporating such constructive criticism from outside the group into the body of the work invariable makes for a stronger study overall.

Usually a paper is submitted to the highest impact factor journal that seems plausible at first, and will be rejected with criticisms. These criticisms improve the paper, which will be submitted again, but this time to a lower impact journal and so on and so on. Overall these effects combat each other to create a natural balance where the work gets better even as it moves down the list of 'best' journals.

No matter what do not give up on a project. As long as your results are scientifically valid (which they will be for us to get to the stage of submitting for publication) then they will be

accepted somewhere.

5.1.3 Open Access

Most papers must be purchased to be read. If you have ever tried to access a paper outside of the university network you may have been asked to pay to download a pdf.

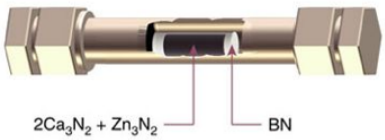
If a paper is made open-access, the authors pay so that everyone can read the paper for free. This massively increases the impact of the study because more people are likely to read it and cite it.

We should always try to make papers open access, but in reality it is too expensive for a group to pay for all of their papers to be published this way.

5.2 Images

Images are of increasing importance in science. I encourage you to think of them as an output in their own right because they can be used in the research article (previous section) or presentation (next section) as well as many other outlets such as a press release, blog post, lectures, tweets *etc*.

Once a paper is published you can't then go back and add a better image into that study later. In my first ever paper is an ugly and highly embarrassing Microsoft Excel plot which will haunt me forever :'(



There are many software packages for making images, some of which are very expensive. However, I'd like to draw attention to the image on this page that was drawn by a collaborator of mine for a previous study. This image was made using the shape tool in Microsoft Powerpoint, which, I think, is a very impressive use of non-specialist software.

There are two types of image file, raster or vector.

Raster images use many coloured pixels or individual building blocks to form a complete image. JPEGs and PNGs are common raster image types. Almost all of the photos found on the internet are raster images because the file size is smaller. However, that means they cant be dramatically resized without compromising their resolution. When stretched to fit a space they weren't designed to fill, their pixels become visibly grainy and the image distorts.

Vector images, alternatively, allow for more flexibility. They are constructed using mathematical formulas rather than individual coloured blocks and have the common file types EPS, SVG and AI. Because vector-based images are not made up of a specific number of dots, they can be scaled to a larger size and not lose any image quality. But, most importantly, the image is made up of objects that exist in layers. So in a vector graphic of a graph with 2 lines, if plotted properly, you can open the graph and delete one of the lines without affecting the other.

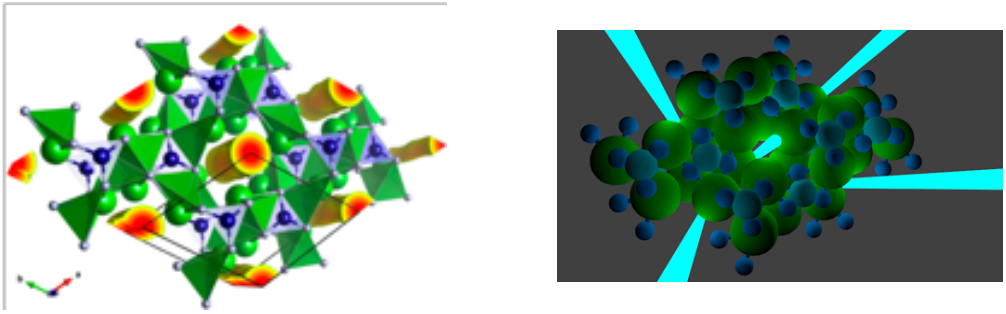


Figure 5.1: Images taken from a recent paper of mine showing the same electron density in a crystal structure. The image on the left was made with VESTA and the image on the right by Blender.

5.2.1 VESTA

Vesta is free software that I've used since the first year of my PhD. It is used to visualise 3-dimensional objects such as crystal structures and charge density among others. It also allows you to modify or expand structures in a way that you can check by eye if you it worked properly or might have some unexpected effects because of symmetry or something.

5.2.2 BLENDER

While blender is easy, clear and quick, the images you can create are fairly standard in the field. If you want to do something more sophisticated you will need to transition to a more powerful but more complicated software.

To begin with you have to load a specific addon called `href="https://wiki.blender.org/index.php/Extensions:2.6/Py/Scripts/Import-Export/PDB">AtomicBlender`, so that blender will understand chemical structures.

After opening Blender, click File -> User preferences, then the Add-ons tab. Here you have to tick Import-Export: Atomic Blender - PDB

From now on you can load protein data bank (pdb) files; into blender (it seems like recently you can also load XYZ files with atomic blender but I'll need to check that). Vesta allows you to export pdb files by default, so there is a pathway to load structures into professional 3D imaging software without having to modify anything by hand.

5.3 PRESENTATIONS

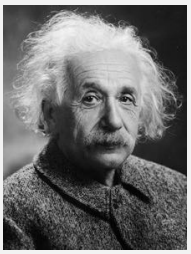
Every slide should have a title. Several slides can have the same title but make sure there aren't any slides with no title. I find that text on a presentation is only useful if you refer to it specifically. I don't know anyone who can read and listen to two different sentences at the same time, and of course you want to make sure people are listening to you.

Most importantly relax. I've seen amazing projects be completely messed up by a speaker because they were afraid they would mess it up... isn't that silly? Even if somebody asks you

the most difficult question you ever heard, you can respond that you'll look into it, or refer them to your supervisor (me).

Depending on the audience depends on the type of presentation. Unless it is a presentation for a job interview, there's no need to try and show off.

Note



Sometimes students make things sound complicated to seem smart but I promise you this impresses no-one. Einstein was credited as saying 'If you can't explain it to a six year old, you don't understand it yourself'. Keep things simple!

I think most importantly, you should not need to memorise anything. The content of the presentation should be instructive for the audience but also be a prompt for you. If you are worried you will forget to say something, put that in the presentation specifically and this will remind you to talk about it. If it doesn't matter whether you remember to say it or not, it doesn't need to be on the slide at all.

Some people disagree but I find animations are a great way to break a slide into sections. They also make sure the audience are focusing on the part you are talking about. If you have 3 images on a slide, the audience will look at all 3 of them as soon as they appear. If you use animations you can talk about the first, then make appear the next one, talk about that, then make the final one appear and conclude the slide.

Again some people disagree but I think it's better if your talk is SHORTER than the allocated time slot rather than longer. Almost always the most important part of your talk is at the end, you will first introduce the topic, explain what you did and then explain the results. I have seen talks when someone spent 15 minutes introducing the topic and methods, but then was forced to stop before they could talk about their results. This of course renders the whole point of presenting completely obsolete.

5.3.1 Data repositories

Increasingly, it is either expected or demanded that the data used in a study be made available alongside a paper publication. There are lots of different opinions on the best way to do this so I'll simply mention one.

Here is one that was made available via Github by the British Scientist Ben Morgan:
`https://github.com/bjmorgan/data_NEB_spinel`

It clearly displays the title of the study, the authors, links to the original paper and has a full explanation of the data.

Perhaps most importantly the python codes necessary to recreate the figures that they publish are there. You can download this repository and play with the notebook to re-plot or modify these figures yourself.

Notice that, everything is available and free, even the paper is open-access. This massively increases the impact of the study.

5.4 CODES

5.4.1 LICENSES

All codes are distributed with a license. If you upload a code to github or similar platform they will prompt you to add your own license or generate one for you.

There are many many different licenses but the 2 most important types we need to understand are the GNU General Public Licence (GPL) and the MIT license.

The biggest difference is probably that GPL license impacts "derivative works", but the MIT license does not. If you use a code with an MIT license, you are obligated to provide attribution with your code or binary (e.g. say "this project uses code that is MIT licensed" – with a copy of the license and copyright of the author of the open source code). In the case of the GPL license, you have the additional requirement of making your source code available.

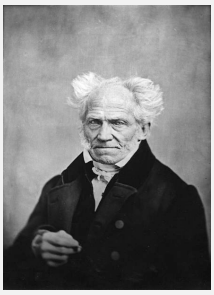
Let's say you create project which contains your code incorporating another code that has a GPL licence. If you distribute the source or binary then you must make it available, because if your code is a derivative work, then your code is also licensed under GPL.

This works for others who may use your code in building their own tool later. If they distributes that code, then it must be licensed under the GPL and the source code has to be made available. This is a property that some people call "hereditary" or "viral".

Alternatively if you make a derivative code using an tool which is distributed under an MIT license. If you distribute it then you have to indicate that part of it that contains MIT licensed code (but you don't have to say how much of it does, where it is used or how it is used). You don't have to make your code available and you can license it any way you choose. You can license it under MIT as well if you want, or under GPL instead. You could even sell a code based on MIT licensed tools under a commercial license but you couldn't do that for a GPL license.

Each license has advantages and disadvantages. In the case of GPL you can be assured that any creative contribution made to your code will be available to the world. In contrast, the creator of a code distributed with an MIT license will never know how many people are creating codes based on it or even making money off of it. But then again, you might want to retain rights in case there is a potential for profit in the future.

Note



The philosopher Arthur Schopenhauer said 'Every man mistakes the limit of his vision for the limits of the world'. There are more opportunities in society than you can possibly imagine! Spend some time to explore jobs and careers outside of academia.

5.5 Job Applications

Ultimately, no one can stay in a research group forever. Different job's in different fields will require their own evidence of experience and talent. I am happy to help you as much as I can to prepare these documents.

Every job application should involve an up-to-date curriculum vitae (CV) and cover letter, at the very least. A recent version of my own CV will be always available online for you to find and I can show you previous cover letters that I have written. You can copy these directly but I advise against it, everyone's CV should be as unique as their life experience and every job application requires a different cover letter. For these reasons I don't go any further in to details here, but feel free to ask me any time.