Online CryoEM Study Group

Geoffrey Woollard, Toronto, Canada

December 16, 2020

1 General Information

Feel free to share this document and direct people to sign up at `https://forms.gle/BUeUW14vV4pyQbDDA` so I have the emails in one place. Online meeting links are emailed to those on this list. **Please join the Slack group and ask questions there, rather than emailing me.**

1.1 Audience and Streams

Over 2/3 of the audience are grad students. There rest of the audience is a mix of long term staff (facility manager, seniour scientist, research associate, unspecified industry position, etc.), principal investigators (industry or academic), postdocs, and undergraduate. After carefully reading the responses, I think it would be best to have two streams each meeting. I have code named them the **blur** and **sharpen** stream, referring to map sharpening. I am counting on your sense of humour and humility for this playful naming schema! Perhaps its a bit more gentle than low and high resolution? Remember that both blurred and sharpened maps are important for being able to see what is going on and build a model in to the map. The goal of this study group is to have both intuition and be able to connect that back to the math, never missing the forest for the trees.

Blur stream: Beginners and intermediates with may have years of expertise doing sample prep, collecting data on the scope, making maps, and building models. However, many people with such expertise expressed a desire to go deeper into the fundamentals, and understand things better under the hood, and requested help to become more confident in the math. One person put it bluntly, "I am interested in being more than a button pusher."

Sharpen stream: This stream can be for people with a high degree of expertise that are looking to sharpen their skills, and interact with other experts. For example: methods developers with perhaps not that much experience processing datasets from start to finish, university instructors, facility managers who have a strong working knowledge but want to go deeper into the foundations and refresh, people with strong math and physics backgrounds that are fairly new but will catch on quickly, experts in crystallography that

are switching over to cryoEM but that have a strong theoretical basis in the underlying math already.

After reading through the responses a few times, I would estimate the ratio of the streams to be about 3/4 blur to 1/4 sharpen. In the first meeting you can choose your own stream, and go to breakout rooms with this in mind. I will make multiple breakout rooms so that there can be small group discussion. We can use the **white board and share screen features**. If you are an expert feel free to go to the blur stream and help out. If you judge yourself an intermediate-beginner, you can test the waters in the sharpen stream. The two stream format may evolve over time (e.g. three streams, a lecture with everyone together), but let's try out two streams and see how it goes.

1.2 Meeting Format

We will start with the "flipped classroom" and "breakout room discussions" to get things off the ground with minimal overhead. After some initial meetings we will try out a mix of the following formats listed below, with their surveyed popularity listed next to them. **If you would like to volunteer to give a lecture or be available for office hours then please volunteer yourself through the Slack channel.**

- Lecture with Q&A
- Flipped classroom: i.e. we have a syllabus with pre-reading (textbook chapter, review paper) that we go through before and then discuss online
- Breakout room discussion (small groups organized thematically)
- Office hours with an expert available to answer questions

2 Dates and Topics

Given the current distribution of our global audience (90%+ in North America and Europe), we will have the time at **9 AM PST, which is 12 PM EST, 5 PM GMT, 7 PM GMT+2, 1 AM next day GMT+8, 3 AM next day GMT+10**. I am open to proposals to occasionally have a different time to accommodate a global audience - especially if the presenter is from a less represented time zone.

Date	Time	Stream	Topic
Th 19 Nov 2020	9 AM PST	blur/sharpen	background math & defocus phase contrast
~~Th 26 Nov 2020~~			(cancelled for US Thanksgiving)
Th 3 Dec 2020	9 AM PST	blur	Fourier transform
Th 10 Dec 2020	9 AM PST	sharpen	Fourier transform
Th 17 Dec 2020	9 AM PST	blur	convolution, sampling, Nyquist
Th 7 Jan 2021	9 AM PST	sharpen	convolution, sampling, Nyquist
Th 14 Jan 2020	9 AM PST	blur	phase-contrast in the EM
Th 21 Jan 2021	9 AM PST	sharpen	phase-contrast in the EM
Th 28 Jan 2021	9 AM PST	blur	2D Expectation-maximization
Th 4 Feb 2021	9 AM PST	sharpen	2D Expectation-maximization
Th 11 Feb 2021	9 AM PST	blur	image formation (forward model)
Th 18 Feb 2021	9 AM PST	sharpen	image formation (forward model), multislice
Mar 2021	9 AM PST	...	Wah Chiu et al, lectures, ...
... 2021 ...	...	...	...

3 Organizational Team

I would appreciate help

- **Giving lectures. You can send a proposal to me including topic, format, time, date.**
- **Being the leader of a breakout room to answer questions. You could list your areas of expertise, so people would know to come to you for those things. A sort of 'office hours'.**
- Developing the syllabus, and suggesting (or making) learning content.
- Formulating challenge problems to test comprehension, build intuition, and conceptual understanding.
- Zoom co-host and admin help.

4 Slack

We will use the Slack channel 'cryoem_study_group' for asynchronous chat. The link to join is `https://join.slack.com/t/cryoemstudygroup/shared_invite/zt-j66wuws3-~UcfsdmtQow~7qYC~iJu_g`. Please join the Slack group and ask questions there, rather than emailing me, as I may not respond!

5 Syllabus

We will start off drawing heavily from the content developed by Dr Frederick Sigworth, `https://cryoemprinciples.yale.edu/video-lectures`. If you are having problems with links, then try viewing his content on YouTube.

Thanks to the interactive learning material developed by Arjen Jakobi, for a course on High-Resolution Imaging at TUDelft: "The practicals are computational assignments in the form of interactive Jupyter notebooks hosted in a virtual learning environment. These notebooks contain code that can be executed to perform certain tasks or visualise results; you do not need any active knowledge of coding to work through the notebook." For the curious, the code that generates the visualizations is available on the repository.

If you haven't already gone through Grant Jensen's popular online course 'Getting Started in Cryo-EM' , now would be a good time to do so. I think enough people have gone through this on their own that we will mainly draw from other material.

5.1 19 Nov 2020

5.1.1 Blur

–Pre-reading

- Complex numbers and the complex exponential (10 min)
- Review of complex numbers (3 pages)
- Defocus phase contrast (35 min)

–Questions

- What is physically happening to the sample when the electron is detected at small diffraction angles vs large diffraction angles? What else can happen to the electron?

5.1.2 Sharpen

–Pre-reading

- Defocus phase contrast (35 min)
- 6.2 The Wave Equation for Fast Electrons in 'Advanced Computing in Electron Microscopy', Kirkland (2000), pp. 156-159.

–Questions

- In Sigworth's derivation of $|\Psi|^2$ in the 'Defocus phase contrast' video, he made various assumptions such as small theta, small epsilon. Under what extreme conditions would they break down. Do these occur in cryoEM? In other experimental regimes besides what is typical in cryoEM?
- Around the 23-25 min mark of the 'Defocus phase contrast' video the CTF seems to oscillate to zeros. Is this observed in practice? Why or why not?
- Work through the derivation with pencil and eraser, justifying each step as best you can. come with your questions to the group study.
- Biological samples are made of atoms that give faint contrast, when compared to samples with higher atomic numbers. Where does the atomic number of the sample come into the equations presented here?

5.2 3 Dec 2020 - Blur

–Pre-reading

- The Fourier transform in one dimension (35 min)
- The 1D Fourier Transform (7 pages)
- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks: Introduction, Fourier Series, Frequency Spectrum, 2-D Fourier Analysis. See the links at the bottom. For this week, work through notebooks "ip_basics_part1", "ip_basics_part2", "ip_basics_part3".

–Questions

- Prove the linearity, scale, shift, convolution properties of the FT in 1D.
- Equation 4 in The 1D Fourier Transform expresses how "a narrower function of x transforms into a broader function of u". Can you think of some examples of this in practice? Hint: what happens when you change magnification?
- Assume you have a 128 Åbox size, with pixel size of 1 Å. What is the spacing spacing between bins in Fourier space, if there are 64 bins in the negative direction and 64 bins in the positive direction? What length ranges (in units of Å) do the first few and last few frequency bins correspond to? How many frequency bins are between 10 and 5 Å, versus 5 and 2.5 Å?
  Hint: $\langle ..., [0, 1/128), [1/128, 2/128), ..., [\frac{128/2-2}{128}, \frac{128/2-1}{128}), [\frac{128/2-1}{128}, \frac{128/2}{128})\rangle$.

5.3 10 Dec 2020 - Sharpen

–Pre-reading

- The Fourier transform in two and three dimensions (43 min)
- 2D and 3D Fourier transforms (9 pages)
- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks: Frequency Spectrum, 2-D Fourier Analysis. See the links at the bottom. For this week, play with notebooks "ip_basics_part4", "ip_basics_part5".
- Optional
  - Explanatory Material: "Tutorials and introductions to Fourier transforms and FFTs, in no particular order." `fftw.org`
  - Roberta Piroddi and Maria Petrou. Non-Uniform Fourier Transform: A Tutorial
  - `https://en.wikipedia.org/wiki/Non-uniform_discrete_Fourier_transform`

–Questions

- Sample a simple function (e.g. exp, sin, gaussian) in 1D or 2D. Then use some library to compute the FFT. Then compute the DFT in your own implementation. How close is the error? Work our the solution analytically for the continuous case. What should the answer be at some discrete points according to the continuous case, and how is it different from what the FFT gave? What is the typical floating point error? What was the speedup of the FFT, and how does this compare to the theoretical limit?
- In practice, how long are FFTs taking (2D and 3D)? How does this compare with other computational bottlenecks like disk I/O and interpolation?
- What are some numerical issues or bottlenecks that can arise when using an implementation of the FFT in practice? What are some ways to overcome them?

5.4 17 Dec 2020 - Blur

–Pre-reading

- Fourier transform: convolution, sampling and Nyquist (37 min)
- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks: 2-D Fourier Analysis, Convolutions. See the links at the bottom. For this week, work through notebook "ip_basics_part6".

–Questions

- Solve the "homework problem" (without peeking at the answer!) in the first part of Fourier transform: convolution, sampling and Nyquist. If $g(x) = e^{-\pi(x-2)^2} - e^{-\pi(x+2)^2}$, what is $G(u)$?
- By the Convolution theorem $f(x) = g(x) * h(x) \Rightarrow F(u) = G(u)H(u)$. Since $F(u) = G(u)H(u) = H(u)G(u)$ we should have $f(x) = g(x) * h(x) = h(x) * g(x)$. Show explicitly the commutativity of the convolution operator: i.e., $g(x) * h(x) = \int dt\ g(t)h(x-t) = \int dt\ h(t)g(x-t) = h(x) * g(x)$
- Look at the animation in the notebook "ip_basics_part6". Write down the mathematical object that is being visualized, including integration bounds. What does one snapshot of the animation represent?
- When does convolution (physically or computationally) happen in cryoEM happen? Be specific when connecting things back to the math, e.g. what is being convolved, what is it's functional form, and how many dimensions are involved?
- When we convolve with a *broad/wide* gaussian function, what is happening in Fourier space? How does this relate to Fourier filtering? Hint, use the convolution theorem.

5.5 7 Jan 2021 - Sharpen

–Pre-reading

- Fourier transform: convolution, sampling and Nyquist (37 min)
- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks. For this week, work through notebook "ip_basics_part6".

–Questions

- Around the 16 min mark in Fourier transform: convolution, sampling and Nyquist, we see that the de-convoluted (recovered) signal blows up around the origin, where there is a delta function. From the math, why exactly did this happen? Is there a way to over come this blowing up effect?
- Let $f_a = e^{-ax^2}$. Show that $f_1 * (f_2 * f_3) = (f_1 * f_2) * f_3$. Note that $\int dx\ e^{-ax^2} = \sqrt{\pi/a}; (a > 0)$
- The notebook "ip_basics_part6" has an interactive where the size of the kernel (Sobel, Gaussian, etc) can be changed. Convoluting with larger kernels could be a more expensive computation, but we could speed it up by doing this in Fourier space by the convolution theorem. If we take the FT of the image and kernel, there sizes to not match, so how can we multiply them element wise? How would you match the pixel sizes and do things in Fourier space to achieve the same result as the real space convolution?
- In the last part of "ip_basics_part6", the Fourier filtering versus the convolution filtering appear the same for Guassian blur, but different for box blur, and very different for sharpen. What might be causing this? Vary the kernel size and notice the different run times. Is doing things in Fourier space always quicker?
- How can we increase the Nyquist frequency during data collection to get higher resolution information? What is the trade off? What should guide our choice of an optimal Nyquist, given our particular microscope and specimen?
- Code up a simple example illustrating the convolution theorem, where you also actually do the convolution in real space. How close is it to the answer where you did the multiplication in Fourier space? What was the speedup? Now try speeding up the calculation by of the convolution in real space by making the convolution kernel smaller. How good of an approximation is this? Use a 2D projection of a 3D density map and a meaningful kernel (e.g. low pass filter) to build your intuition in a useful way.

5.6 14 Jan 2020 - Blur

–Pre-reading

- Phase-contrast imaging in the EM' (10 pages)
- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks: CTF. See the links at the bottom.

–Questions

- See the 'signal star' image on panel (b). What would this image look like if the CTF was applied (i.e. through convolution)?
- As defined in equation 11, what happens when the B factor is negative versus positive?
- In practice, on the scope, the CTF is easier to see in an FFT if you are at high mag. Why? Can you relate this back to the math? How is magnification playing into the equation?
- As defined in equation 11 of the 'Phase-contrast imaging in the EM' pdf, what physical behaviour is involved in this B factor? What other factors of B are there in cryoEM and what physical behaviour is involved in those B factors? Why are they all called B-factors?

5.7 21 Jan 2021 - Sharpen

–Pre-reading

- Philippsen et al. 2007. "The contrast-imaging function for tilted specimens" https://www.sciencedirect.com/science/article/pii/S0304399106001525
- Voortman et al. 2011. "A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM" https://www.sciencedirect.com/science/article/pii/S0304399111000878
- Voortman et al. 2012. "Fast, spatially varying CTF correction in TEM" https://pubmed.ncbi.nlm.nih.gov/22728402/

5.8 28 Jan 2021 - Blur

–Pre-reading

- Nelson, P. C. (2019). Chapter 12 : Single Particle Reconstruction in Cryo-electron Microscopy. In Physical Models of Living Systems (pp. 305?325).
- Interactive coding notebook 'Practical 2 - 2D/3D reconstruction' in High-resolution imaging course at UTDelft.

5.9 4 Feb 2021 - Sharpen

–Pre-reading

- Nelson, P. C. (2019). Chapter 12 : Single Particle Reconstruction in Cryo-electron Microscopy. In Physical Models of Living Systems (pp. 305?325).
- Interactive coding notebook 'Practical 2 - 2D/3D reconstruction' in the High-resolution imaging course at TUDelft.

5.10 11 Feb 2021 - Blur

–Pre-reading

- NCCAT SPA short course 2020, Lecture 4: Algorithms and foundational math Part I & 2, Fred Sigworth (1:18:10 - 1:28:46, 10 min)

5.11 18 Feb 2021 - Sharpen

–Pre-reading

- Section '2. Theory' in Vulović, M., Ravelli, R. B. G., van Vliet, L. J., Koster, A. J., Lazić, I., Löcken, U., ? Rieger, B. (2013). Image formation modeling in cryo-electron microscopy. Journal of Structural Biology, 183(1), 19?32. `http://doi.org/10.1016/j.jsb.2013.05.008`
  - Supplementary material associated with the article (18 pages)
- 6.4 The Multislice Solution in 'Advanced Computing in Electron Microscopy', Kirkland (2000), pp. 162-165

5.12 March 2021

By this time the anticipated textbook by Wah Chiu, Robert Glaeser, and Eva Nogales will hopefully be out, and we can see how it looks. We could also plan some lectures.