# Material designing using computer simulation and machine learning

Jie Huang

December 22, 2021

## Summary of the Proposal

## Background

The rapid advancement of big data and artificial intelligence has resulted in new data-driven materials research and development, which has achieved substantial progress.[1] The amount of data being generated by these experiments and simulations has given rise to the fourth paradigm of science [2] over the last few years, which is data driven science, and it unifies the first three paradigms of theory, experiment, and computation/simulation. It is increasingly becoming popular in the field of materials science as well and has, in fact, led to the emergence of the new field of materials informatics. [3] Recent advances in experimental and computational methods are increasing the quantity and complexity of generated data. This massive amount of raw data needs to be stored and interpreted in order to advance the materials science field. Identifying correlations and patterns from large amounts of complex data is being performed by machine learning algorithms for decades. [4]

## Goal and Objectives

You insert a separate section here and write about the goal(s) of your research and the objectives that will meet the goal. Typically, the way to write this is something like, "The goal of this research is to ...", and then continue with something like, "Goal 1 will be met by achieving the following objectives ...", and so on. The goal is a broad based statement, and the objectives are very specific, achievable series of statements that will show how you will achieve the goal you set out.

## Population and Methods

This is going to be the third and final section of your proposal. In the people and methods section, you include the following items:

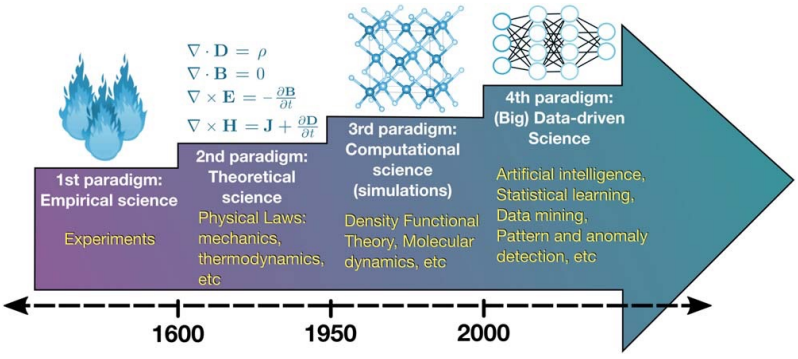


Figure 1: The four science paradigms: empirical, theoretical, computational, and data-driven. Each paradigm both benefits from and contributes to the others. Adapted from [3, 4]

- Write about the population that you will study
- If it is an observational epidemiological study, then write about the exposure variable you will study. Hopefully, your background section will already have covered the prevalence of the exposure. If it is an intervention research, you will write about the intervention that you want to test.
- You will describe in details about the comparison group. If your study is one where you will be testing hypotheses, then it is important that you write about the comparison groups. You will write about the prevalence and how you will obtain measurements about the exposure and comparison groups.
- You will write about the outcomes in details, and specifically about how you will measure the exposure/intervention and the health outcome you want to study
- You will describe in details the power and sample size for this study. You can use the OpenEpi webpage to calculate your sample size and power for your study
- You will write in details about how you will eliminate bias in the measurement of the different variables in your study
- You will need to write, once you obtain data from your participants, how you would propose to analyse such data. You need not write too much details here, as you have not yet collected any data but an indicative set of statements as to what you will do should be sufficient.

These are the three compulsory sections that you will need to include in your proposal and then submit using Learn. If you use this template on Overleaf, then you can generate a PDF of your paper by selecting the PDF symbol on the top of this window. Save the PDF in your hard drive and then upload that one copy of PDF to Learn. If you use this template on Word, then convert the Word document to PDF and upload the document through Learn. Your document must contain images, tables, lists. All facts that you write must be accompanied by appropriate citation and referencing. The referencing information must be simple (just a number in square brackets), and an alphabetical order of the references in the bottom of the document should be sufficient. If you want to use APA style of referencing, that is OK too. For example, I have cited here a secondary analysis of data from papers published for about 40 years on statistical inference. It was an interesting paper written by Stang et.al. [5] and published in 2016. In my case the paper is cited in square brackets like this: `[1]`, and a full citation of the paper is mentioned in the references section. If you want to do the same but use APA 6th Edition citation, this is fine as well (and is used at the University here). But do not be discouraged to use your own style, as long as the citation information and the reference is there, that should be OK.

If you want to write using this template on Overleaf, I have written a tutorial that you can use to learn more about how to write on Overleaf. Or watch their several videos on the site to learn more about this tool to write your paper.

# References


[1] Dezhen Xue Xue Jiang Huadong Fu Haiyou Huang Jianxin Xie, Yanjing Su. Machine learning for materials research and development. *Acta Metall Sin*, 57(11):1343, 2021.

[2] Tony Hey, Stewart Tansley, and Kristin Tolle. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, October 2009.

[3] Ankit Agrawal and Alok Choudhary. Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *APL Materials*, 4(5):053208, April 2016.

[4] Gabriel R Schleder, Antonio C M Padilha, Carlos Mera Acosta, Marcio Costa, and Adalberto Fazzio. From DFT to machine learning: recent approaches to materials science–a review. *Journal of Physics: Materials*, 2(3):032001, May 2019.

[5] Andreas Stang, Markus Deckert, Charles Poole, and Kenneth J Rothman. Statistical inference in abstracts of major medical and epidemiology journals 1975-2014: a systematic review. *European journal of epidemiology*, November 2016.