Queen Mary University of London

Data-Centric Engineering

Preparing Your Research Proposal

qmul.ac.uk/dce

Pre-application Workshop 2
Queen Mary University of London
13th April 2021

- What is a Research Proposal?
- What we are looking for
- Examples

Dr Eram Rizvi
Deputy Dean for Research

Engineering and Physical Sciences Research Council

Queen Mary University of London

What is a Research Proposal?

Research Proposal

- Opportunity to express yourself
- Explain what research excites you
- Communicate the project well / succinctly
- Illustrate the potential impact of the project

Research Proposal
300-700 words

Show us how you think
Show us how you communicate

Why is the research necessary?

Is there a demand?
Has the problem already been solved?

Who are the beneficiaries of the research?

industry sector
a specific company or organisation
is there a societal benefit?
will this work assist the UK economy?

You will not be bound to this project - need to find matching supervisor
You can contact potential supervisors now (see links in introduction slides)
We can assist in finding suitable supervisor
Projects will be adapted / refined with supervisor if you are successful

What is a Research Proposal?

Structure Of A Research Proposal

- A catchy title helps grab attention
- Give some background to the problem
- What is the challenge ? Why is this needed?
- What techniques / facilities will be used? Where will the data come from?
- Can you outline a work-plan for the research?
- Include a short bibliography (excluded from word count) not necessarily academic journal papers

https://www.phdassistance.com/blog/how-to-write-your-research-proposal/

What We are Looking For

Strictly 700 words maximum (excl. bibliography)
Describe 1 or 2 projects (max.)
Not much space to describe your proposal(s)

Below I list some things to consider
we do not expect all of these to be met in your proposal
consider what is most relevant to your topic

Proposal should show
- Clearly structured thought
- Ability to plan out a project - timeline and resources
- Informed background knowledge
- Match to your skills / competencies

Have you thought about the problem from more than one perspective?

Do you have the resources needed, how will you acquire them?

Are there ethical considerations, e.g. using personal data?

Are you framing the right research question?

Can you articulate why it is interesting?

In-depth academic knowledge of the stat-of-the-art is not necessary
…but if you have it then tell us about it!

Research Opportunities

Examples of Doctoral Research Projects on offer - please email supervisors for more information

Supervisor	School	Project Title
Dr. Chris Jones	Biological & Chemical Sciences	Developing of Novel Cannabinoids with Machine Learning
Prof. Christian Beck	M	g of power-grid frequency
Dr Akram Alomainy	Electronic	Development of Novel Data-Centric Techniques in emergent wireless
Dr Richard Cle		sis of software
Prof. Adrian Be		detectors
Prof. David Du		lihood Fits in Data-
Dr Lin Wang		Music Enhancement
Prof. Josh Reiss	Computer Science	Parametric Controls from Data Analytics
Dr Seth Zenz	Physics	Designing novel ultra-thin low mass curved silicon imagining sensors for X-ray diffraction and nuclear security
Dr Yannick Wurm	Biological & Chemical Sciences	A toolkit for pragmatic interrogation exploration & hypothesis testing of disconnected genomic data
Dr Anthony Phillips	Physics	Using machine-learning techniques to identify new perovskite materials with electrical properties that exceeding those of inorganic perovskites
Dr Jens-Dominik Mueller	Engineering	Adaptive multi-fidelity robust design optimisation driven by machine learning

https://www.qmul.ac.uk/dce/research/

Links to Available Doctoral Research Projects in our Schools

- School of Electronic Engineering and Computer Science
- School of Engineering and Materials Science
- School of Chemical and Biological Sciences
- School of Physics and Astronomy
- School of Mathematical Sciences

Research Proposal Examples

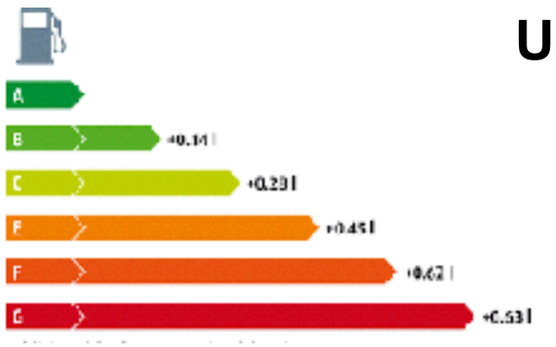
Some examples given here - looking at titles only for this exercise

Research into new black holes

exciting topic - Nobel Prize for physics 2020!
vague title - what will you investigate?
not data-centric engineering
do you have skills in general relativity?
will industry partners be willing to participate? probably not…
not applied research

Building a large-scale distributed database for financial assets

no research element - application of existing technology
where is the challenge? This is a solved problem



Using machine learning to increase vehicle fuel efficiency

good topic
potential interest from many industry organisations
clearly a data-engineering problem
you will need a background in vehicle design engineering

Research Proposal Examples

Using machine-learning to identify new perovskite materials with novel electrical properties

new application of machine learning
potential benefit to industries using perovskites (e.g. renewable solar energy)
merges chemical / materials engineering with data science approach

Designing novel ultra-thin curved silicon sensors for nuclear security

interesting data-centric engineering topic
requires software modelling skills
methodology is not clear from title
target industry is clear

Building a model of urban air pollution

hot topic currently
potential interest from many industry and governments (incl. internationally)
clearly a data-engineering problem: use input data to generate software model

Final Tip

Good exercises to prepare the proposal
Outline with bullet-point headings
Re-order bullet points for logical flow
Assign rough word counts to each paragraph in order of importance
Draft each paragraph
….
Forget about the draft for 1-2 days
Return to it with fresh eyes and re-order sections if needed
Cut the word-count by 20% or 30% but without losing any content
…
Repeat if you have time

This process helps you focus on what is important
Your final draft will be much ‘tighter’

in the calibration of the electron and HFS energy. This is determined by varying θ_e and γ_h by the angular measurement uncertainty.

5.3 Neutral Current Measurement Procedure

Inelastic ep interactions are required to have a well reconstructed interaction vertex to suppress beam induced background events. High Q^2 neutral current events are selected by requiring each event to have a compact and isolated cluster in the electromagnetic part of the LAr calorimeter[4]. The scattered lepton is identified as the cluster of highest transverse momentum. In the central detector region, $\theta \geq 30°$, the cluster must be associated to a CTD track. Forward going leptons with $\theta < 30°$ traverse the region between the FTD and CTD where an increased amount of dead material causes electrons to shower. Since in this kinematic region the scattered lepton has high energy and the contribution from photoproduction background is very small, no tracker information is required to be associated with the lepton for $\theta < 30°$.

Energy-momentum conservation requires the variable $E - P_z$ summed over all final state particles (including the electron) to be approximately equal to twice the initial electron beam energy. Restricting $E - P_z$ to be greater than 35 GeV considerably reduces the photoproduction background and the radiative processes in which the scattered lepton or bremsstrahlung photons escape undetected in the lepton beam direction.

The photoproduction background increases rapidly with decreasing electron energy, therefore the analysis is separated into two distinct regions: the *nominal* analysis ($y_e \leq 0.63$ for $Q_e^2 < 890\,\mathrm{GeV}^2$ and $y_e < 0.93$ for $Q_e^2 > 890\,\mathrm{GeV}^2$) for which the minimum electron energy is $11\,\mathrm{GeV}$ and the *high y* analysis ($0.63 < y_e < 0.9$ and $56 < Q_e^2 < 890\,\mathrm{GeV}^2$) for which the minimum electron energy is $5\,\mathrm{GeV}$. The techniques employed to contend with background in each analysis are described below.

5.3.1 Nominal Analysis

For the *nominal* analysis the small photoproduction background contribution is statistically subtracted using the background simulation. The normalisation of the background simulation is checked using a sample of data events in which a true scattered lepton is observed in the electron tagger which, however, has low acceptance.

The comparison of e^-p data and the simulation is shown in figure 4(a) for the scattered lepton energy spectrum and polar angle, and the distribution of $E - P_z$, which are all used in the kinematic reconstruction of x and Q^2 using the $e\Sigma$-method. The corresponding distributions for e^+p data and simulation are shown in figure 4(b). In the figure the R and L data are combined and the simulation is normalised to the luminosity of the data, as is also done for all later performance figures. All distributions are described well by the simulation aside from a small difference in normalisation which is discussed in section 7.2 where the data are compared to the NLO QCD fit.

[4]Small local detector regions are disregarded in the analysis where the cluster of the scattered electron is not fully contained e.g. intermediate space between stacks, or where the trigger is not fully efficient.

For the NC analysis in the region $y < 0.19$ the noise component has an increasing influence in the transverse momentum balance $P_{T,h}/P_{T,e}$ through its effect on $P_{T,h}$. The event kinematics reconstructed with the $e\Sigma$-method in which the HFS is formed from hadronic jets only, limits the noise contribution and substantially improves the $P_{T,h}/P_{T,e}$ description. The jets are found with the longitudinally invariant k_T jet algorithm [59, 60] as implemented in FastJet [68, 69] with radius parameter $R = 1.0$ and are required to have transverse momenta $P_{T,\mathrm{jet}} > 2\,\mathrm{GeV}$. In figure 5(a) the quality of the simulation and its description of the e^-p data for $y_e < 0.19$ can be seen for the distributions of the $P_{T,h}/P_{T,e}$, γ_h, and $E - P_z$ where all HFS quantities are obtained using the vector sum of jet four-momenta. Distributions for the e^+p sample are also shown in figure 5(b). Overall both sets of distributions are well described in shape by the simulation.

At low y, the forward going hadronic final state particles can undergo interactions with material of the beam pipe. In some cases the products of these secondary interactions are incorrectly assigned as originating from the primary vertex, producing a bias in the determination of the primary interaction vertex position. Such cases can be identified and corrected by considering a vertex position calculated using a standalone reconstruction of the track associated with the electron cluster [65, 67].

For the nominal analysis the photoproduction background contribution is low, and this allows the electron candidate track verification for $\theta \geq 30°$ to be supplemented with an alternative method which increases the acceptance for NC events with no CTD track associated to the electron cluster, the track being achieved by searching for hits in the CIP located on the line from the vertex to the electron cluster.

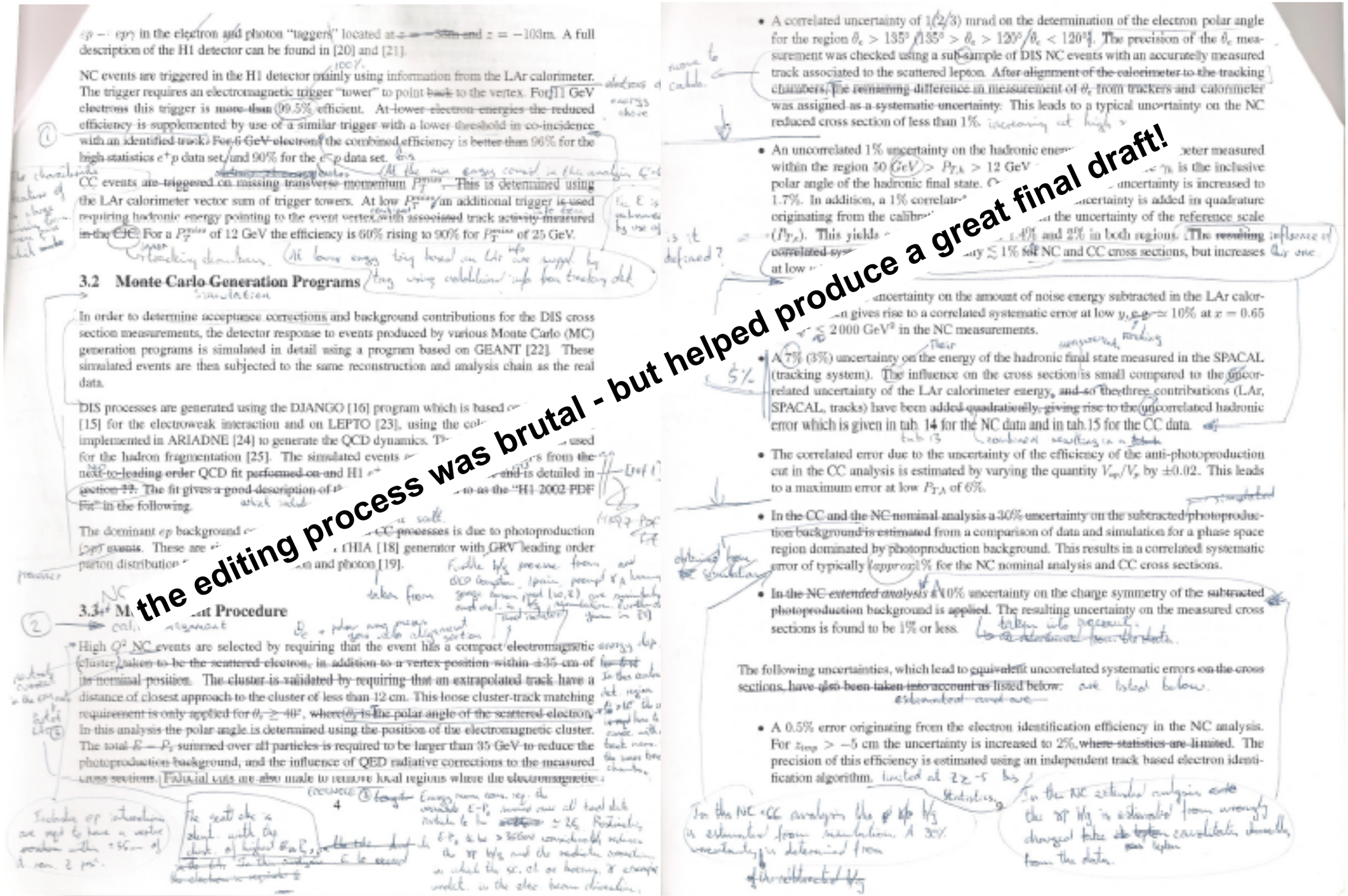
The optimised treatment of the vertex determination and verification of the electron cluster with the tracker information improves the reliability of the vertex position determination and increases the efficiency of the procedure to 99.5%.

5.3.2 High y Analysis

In the *high y* region the neutral current analysis is extended to lower energies of the scattered electron, $E'_e > 5\,\mathrm{GeV}$. At low energies photoproduction background contributions arise due to $\pi^0 \to \gamma\gamma$ decays and charged hadrons being mis-identified as electron candidates. Part of this background is suppressed by requiring a well measured track linked to the calorimeter cluster. The track is furthermore required to have the same charge as the beam lepton. The remaining background in the correctly charged sample is estimated from the number of data events in which the detected lepton has opposite charge to the beam lepton. A charge asymmetry can arise due to the different detector response to particles compared to anti-particles, in particular p and $\bar{p}$ [70, 75]. By taking into account the charge asymmetry between negative and positive background, the background estimate is statistically subtracted from the correctly charged sample. The charge asymmetry between fake lepton candidates in the e^+p and e^-p data sets is determined by measuring the ratio of wrongly charged fake scattered lepton candidates in e^+p and e^-p scattering, taking into account the difference in luminosity. The asymmetry is found to be 1.03 ± 0.05. This is cross checked using a sample of photoproduction events in which the scattered electron is detected in the electron tagger. Further details are given in [67, 71].

The e-method using scattered lepton variables alone has the highest precision in this region of phase space and is used to reconstruct the event kinematics.

Final Tip



the editing process was brutal - but helped produce a great final draft!