STudent REseArch Mobility Programme (STREAM)
Project proposal

Host University:
University of Cambridge

Field (drop-down list):
Natural sciences, mathematics and statistics

Specified field, subject:
Chemistry, Physics, Materials Science, Computer Science

Research project title:
Machine Learning in Chemistry and Materials Modelling

Possible starting month(s):

<table>
<thead>
<tr>
<th></th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>☜</td>
<td>☐</td>
<td>☜</td>
<td>☜</td>
<td>☜</td>
<td>☜</td>
<td>☜</td>
<td>☜</td>
<td>☜</td>
<td>☜</td>
<td>☜</td>
<td></td>
</tr>
</tbody>
</table>

Possible duration in months:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>☜</td>
<td>☞</td>
<td>☜</td>
<td>☜</td>
<td>☜</td>
<td>☜</td>
</tr>
</tbody>
</table>

Exact starting and end dates will be discussed between the supervisor and the student

Suitable for students in: ☐ Bachelor level ☒ Master level

Prerequisites:
(1) A university level course on chemistry or solid state physics
(2) Some experience in programming (C/C++ or Fortran or Python or Julia)

Restrictions:
NONE

Description (maximum 2,000 characters):
Research project may be adapted according to the student profile and the period/timeline

Machine learning and statistical learning approaches, coupled with continual increase in computer power, has recently enabled significant progress in materials modelling in a wide range of areas. Properties of materials can be trained on large databases and then subsequently predicted often with high accuracy. Propensity of binding of small drug-like molecules to target proteins can be predicted based on a very limited number of examples. For simple materials, the full quantum mechanical treatment of the coupled electron-ion system can be calculated, fed into a learning system, and subsequently simulated with an efficiency enhancement factor of a million or more.

This research project is aimed at extending these early efforts, many of which have originated in my research group, taking machine learning into new and important areas, such as dynamics of organic molecules, multi-component alloys, and metal oxides. There is scope for mathematical development, software engineering, atomistic simulation, data modelling and analysis depending on the background and interests of the student.
Faculty and/or Department: Engineering Laboratory

Contact person, including position: Professor Gábor Csányi

Nominations to be sent to: Lucy.Gager@admin.cam.ac.uk

Contact email: gc121@cam.ac.uk

Deadline for nomination to reach host university: 31st March

Notification of admission given by the end of: Within 3 weeks of the closing date for nominations

Additional information: