Molecular Simulation and the Collaborative Computational Projects

The story of CCP5

Bill Smith[1], Martyn Guest[2], Ilian Todorov[1] and Paul Durham[1]

15 June 2020

Abstract

In the late 1970s, the embryonic UK research community in molecular simulation – physicists and physical chemists – organised itself around CCP5, one of a set of **C**ollaborative **C**omputational **P**rojects in different fields. CCP5 acted to develop and use the software required by an evolving and expanding scientific agenda, to exploit quickly and efficiently the revolution in computing hardware and to educate and nurture the careers of future generations of researchers in the field. This collaboration formally began in 1980, and is still fully active now, 40 years later. Today, molecular simulation techniques, many of them pioneered by CCP5, are now used very widely, including in several other CCPs in the UK's current family of Collaborative Computational Projects. This article tells the story of molecular simulation in the UK, with CCP5 itself at centre stage, using the written records in the CCP archives. The authors were, or are, all personally involved in this story.

1 Preamble

Recently the European Physical Journal has published a number of interesting articles on the historical development of molecular simulation [1], [2], [3]. As an additional contribution to this story, here we give an account of an important aspect of this development in the UK: the part played by the Collaborative Computational Projects – the CCPs. At centre stage in this story stands CCP5, the Collaborative Computational Project on Molecular Dynamics and Monte Carlo Simulations of Bulk Systems. CCP5 was founded in 1980 and is therefore just 40 years old; it is still going strong under a modified title of the Computer Simulation of Condensed Phases. Any 40th birthday deserves a little celebration; certainly, CCP5 has many achievements worth recording. However, there is a broader point that should be made regarding the overall CCP programme and CCP5 in particular. The key aspect of the initiative is that it was conceived as a *collective effort* by what was at the time a fledgling computational research community in the UK, and as it developed in scale and scope it has retained this collective, community ethos. We think it is important to highlight the role such social initiatives, as it were, can play in the intellectual development of a field of science and in

[1] STFC Daresbury Laboratory

[2] University of Cardiff and Daresbury Laboratory

dispersing new ideas and techniques into a wider scientific community. There are many aspects to this – scientific ambition, scientific politics, financial climate, the personalities of the major players – and the founders of the CCP5 Project were certainly visionary as well as opportunistic in a good way. Moreover, as time passed the basic constitutional and operational model of the CCPs proved to be highly robust, undoubtedly a key factor in CCP5's longevity. Equally, the story of CCP5 carries wider lessons applying to a greater or lesser degree to its sister projects, the other CCPs, in fields quite different from computational statistical mechanics.

In this article, therefore, we try to trace what happened in the CCP5 Project over its first 40 years, to describe its major achievements, to indicate who did what, and to offer some thoughts on the reasons for its success. In telling this story, we have taken as read the central importance of molecular simulation in modern science and we therefore did not feel it necessary to reiterate this in any detail in the present article. But as the CCP5 story unfolds, the scientific scope of its applications is truly impressive. As a matter of fact, in 2019 the molecular simulation methods originally championed by CCP5 are now used by several other CCPs (in molecular biology, condensed matter electronic structure, fluid dynamics etc).

A few remarks on the structure of this paper may be in order. Since CCP5 was not the only collaborative computational project established in the UK, we first need, in section 2, to set the scene as regards the CCP programme as a whole, so that the generic aspects of the CCP5 story can be seen more clearly. Next, from the various source documents[3] we have reconstructed as complete a chronology of CCP5 events as we can, from 1980 to the present time. This chronology is summarised in section 3. The specific technology developments, both in computing hardware and software, occurring at this time were absolutely central to the development of the CCPs; we cover these in section 4 paying particular attention to the question of how a collaborative computational community can effectively be supported technically to do its science. As a case study of how these issues played out in practice within CCP5, in section 5 we give a quite detailed account of the conception, development and support of the DL_POLY suite of parallel simulation codes. Section 6 is devoted to a particularly important communal activity of CCP5, its annual Summer School. Finally, in section 7 we give our reflections on how CCP5 operated to establish and nurture its community, on how well (or otherwise) it did it, and on how CCP5 stands now and in the future.

The authors of this paper are not professional historians; we are scientists who worked at Daresbury Laboratory on various aspects of CCP5 and other CCPs – scientific, technical and managerial. Clearly, we cannot avoid bringing our particular point of view to this account, but we have aimed to make it a fair and rounded one, and we have benefitted from the help and advice of many of the key players in the story. In fact, we hope that our own close involvement with this and other CCPs adds some value to this enterprise.

2 What are CCPs?

As the number in its title implies, CCP5 was actually one of a number of Collaborative Computational Projects founded around the late 1970s and early 1980s in the UK. By 1982 there were 9 CCPs covering the wide range of fields indicated in the following table.

[3] CCP5 has an extensive paper and digital archive [4] supported by Ilian Todorov at STFC Daresbury Laboratory.

Project	Title	Field
CCP1	Electron Correlation in Molecular Wavefunctions (1974)	Quantum chemistry
CCP2	Continuum States of Atoms and Molecules (1978)	Atomic and molecular physics
CCP3	Surface Science (1979)	Solid state physics
CCP4	Protein Crystallography (1979)	Molecular biology
CCP5	Molecular Dynamics and Monte Carlo Simulation of Macroscopic Systems (1979)	Computational statistical mechanics
CCP6	Heavy Particle Dynamics (1979)	Atomic physics
CCP7	Analysis of Astronomical Spectra (1980)	Astronomy
CCP8	Nuclear Structure Physics (1980)	Nuclear physics
CCP9	Electronic Structure of Solids (1981)	Solid state physics

Table 1. List of Early CCPs

The programme of CCPs was initially established by the Science Research Council (SRC), the UK government's main funding agency for scientific research at the time[4]. As well as funding university researchers by means of grants, the SRC also operated two large central laboratories: The Rutherford Appleton Laboratory (RAL) near Oxford and Daresbury Laboratory (DL) half way between Manchester and Liverpool. The SRC described the aims of the programme as follows [5], [6].

The major aim of these projects is to bring together scientists from several different universities and research groups to:-

i. *provide for the rapid interchange of information on theory, algorithms and computer codes;*
ii. *collect, maintain and develop relevant items of software;*
iii. *encourage basic research in the given areas by providing facilities for rapid computer implementation of new methods and techniques;*
iv. *assess and advise on associated computational needs;*
v. *disseminate information among University and other research groups by organising "symposia" or "workshops".*

To assist in this work the SRC is providing support in the following ways:

i. *support of staff from the Theory and Computational Science Division at Daresbury Laboratory;*
ii. *short term appointments of Senior Visiting Fellows;*
iii. *longer term Research Assistantships; [In practice, there is provision for one RA per project at Daresbury Laboratory or at one of the collaborating universities. The possibility of additional RAs working on specific topics of interest to a project at the collaborating universities is also envisaged.]*
iv. *support for university research workers needing to spend periods from a few days to a few months working on a project;*
v. *provision of travel and subsistence expenses for meetings of the working groups supervising the projects;*
vi. *allocation of computing time on SRC computers.*

We should perhaps pause briefly to reflect on this initial specification. In hindsight, it seems to us that the funding agency of the day, SRC, showed vision, ambition and generosity in accepting and implementing its academic advice to proceed with the project. That academic advice was strongly

[4]Over time, other Research Councils in the UK came into being, changed names (SRC became SERC, E standing for engineering), merged, de-merged and so on. Several became involved in the CCP portfolio: the Biology and Biological Sciences Research Council (BBSRC), the Natural Environment Research Council (NERC), the Medical Research Council (MRC) and especially The Engineering and Physical Sciences Research Council (EPSRC) – see section 7.5. The Daresbury and Rutherford Appleton Laboratories are now part of the Science and Technology Facilities Council (STFC).

driven by the quantum chemistry and atomic and molecular physics communities. Many of the essential elements later to form the basis of SERC's portfolio of CCPs, as we describe below, were put in place from the outset, including the central role of Daresbury Laboratory and its staff and management. Indeed, over time, the aims of the CCP programme have remained in essence unchanged, and although the practical ways in which the projects address these aims have certainly evolved to match circumstances, they can be crystallised into a "standard model" reading something like this:

*The CCPs bring together all the major UK research groups in a given field to pool their ideas and resources to tackle the development, maintenance, distribution and user-support of large scale scientific software. This is done by implementing **flagship code development projects** and by means of **networking activities**: curating libraries of code; organising training in the use of codes; holding meetings, workshops etc; inviting overseas researchers for lecture tours and collaborative visits; and issuing regular newsletters.*

Underlying all this was the realisation that the writing of a world-competitive scientific code was becoming a task too large for any research group to handle on its own. After all, the post-docs who are, in practice, the workhorses of every research group have a limited working lifetime (usually 3 years in one role). In computational research, this often meant that it was too much for even the most brilliant and effective post-doc to write a serious code, do some recognised research with it, find her next job and leave the code in a fit state to be taken up by a successor. The intellectual investment required to produce useful codes simply had to be protected beyond the canonical 3-year period. The collaborative, collective ethos exemplified by the CCP model was thus driven to a great degree by the increasing demands of remaining competitive in computational research. A cottage industry had to be replaced by a more coordinated collective effort. Of course, it is quite pleasant to work on one's own in a quiet cottage, and it is true that the CCP model, with its slight whiff of centralised planning [7], did not appeal to everyone. Nevertheless, the CCP model seems to have had, from its outset, a striking robustness, able to accommodate, for the most part, the internal tensions among participants that must inevitably occur from time to time. After all, CCP5 has been doing this for 40 years.

The typical way the Research Councils actually funded the CCPs evolved into the following process. Each Project was approved for a three-year period on a research grant containing

I. an administrative part - the costs of "networking" activities (eg meetings, newsletters, maintaining program libraries, travel and subsistence for collaborations, and so on);
I. a scientific part – a "flagship project" usually involving the development of a new code (or new functionality of an existing code) for new science. Such projects, intended to be able to be completed in three years, would thus result in new code for the program library, which would be maintained and made available for all;
II. a staffing part – the costs of the permanent staff at Daresbury associated with the both the running and the scientific/technical aspects of the Project, and the costs of any post-doctoral researchers needed to work on flagship projects. The latter were sometimes based at Daresbury, collaborating with the key University groups involved, and sometimes based within those groups themselves.

This model, or variations on it, has obviously suited the CCPs pretty well for 40 years. However, since the grant proposals are a mixture of administration and research, they do not necessarily fit comfortably into the peer review systems operated by funding agencies. This introduced certain tensions, potentially disadvantageous to the basic notion of a CCP; we will return to this point later in section 7.

All long-term programmes need periodic independent reviews to ensure their continued relevance and usefulness, and this has certainly been a prominent factor in the lives of the CCPs. The Research Councils tended to hold reviews conducted by international panels of scientists every 5-10 years, while the funding for the support provided by Daresbury was scrutinised annually and formally reviewed every 3 years or so. Finally, there were occasional high-profile reports commissioned by the Research Programmes to review their funding of entire fields; where relevant, the CCPs featured strongly in these exercises. Perhaps the best example of this was the report commissioned in 2003 by the EPSRC working in partnership with the German scientific research agency (DFG) chaired by Horst Simon [8]; we will return to this in section 7.

Figure 1: Phil Burke

What was the scientific context into which the CCPs were born? Perhaps we might identify the 1970s as the beginning of "computational science[5]", at least in the UK, as a respectable scientific activity, with its own intellectual and technical status. Nowadays, of course, it is obvious that computational methods are utterly indispensable to all fields of scientific research, both theoretical and experimental. Nevertheless, in the 1970s there was a feeling in some theory departments (again, at least in the UK) that there was something not quite respectable about computational work. "Brute force" numerical calculations was a phrase that one heard fairly regularly in physics departments, with the implication that gentlemen theorists did analytical work. One well-known theoretical physicist is reported to have referred to computational scientists as "hairy-handed mahouts operating their elephantine computers". But this kind of attitude was by no means universal and things were changing. In chemistry departments, quantum chemists were establishing a strong presence and doing hitherto impossible research. In the USA, the value of computational methods was already appreciated. By 1974, the Gordon Conference on the Liquid State had 5 or 6 computationally based papers [9]. Thus, while there was some resistance, it was clear that computational science was moving centre stage.

Given this kind of scientific atmosphere, it is interesting to look at the prevailing circumstances in the UK during the late 1970s that led to the establishment of the CCP programme. Phil Burke[6] gives an illuminating account of the background to the key decisions [6]; we summarise this as follows. The idea of the first project, CCP1, was born in 1973 at a Working Party chaired by Ron Mason at the Rutherford Appleton Laboratory (RAL). This group gave the proposed quantum chemistry project the intriguing name of "Meeting House", for reasons probably now lost in the mists of time. The Science Board of SRC considered this proposal in October 1973 and gave it a green light. The Science

[5]As opposed to "computer science" – something altogether different.

[6]Professor P G Burke FRS was a leading theoretician of atomic and molecular physics, with a chair at Queen's University Belfast. He made seminal contributions to the R-matrix theory of atomic collision processes. He was the founding director of the Theory and Computational Science Division at Daresbury Laboratory in 1977, and was the essential visionary of the CCP concept and its early implementations in several fields. With his quiet, thoughtful manner, he was extremely effective in establishing the basis for UK computational science and supercomputing strategy.

Board also suggested that this might be the first of a series of analogous projects on different computational topics. Thus, the notion of a programme or portfolio of CCPs was conceived very early on and the first project was approved as a pilot study. The Science Board asked Jack Howlett (Director of the Atlas Computer Laboratory at RAL) to establish a Steering Panel consisting of leading quantum chemists and atomic physicists: Brian Bransden (Durham), Phil Burke (Belfast), Roy McWeeny (Sheffield, chairman) and John Murrell (Sussex). At its first meeting in February 1974 this Steering Panel (which still exists after 46 years) agreed the aims and actions listed in the previous section. A couple of years later in July 1976, the Steering Panel approved the continuation of the first (pilot) project and recommended that "further projects of exceptional timeliness and promise should be initiated" [6]. The following year it was decided to co-locate CCP1 with the next project within the newly-formed **Theory and Computational Science Division**[7] at Daresbury, with Phil Burke as Division Head and John Pendry[8] as head of the theory group. As Phil Burke writes [6], this move "set the scene for a rapid increase in the CCP programme". It also established the key role of Daresbury as the cornerstone of the CCP programme.

Interestingly, as early as 1976 the Steering Panel encouraged the CCPs to form strong links with CECAM, the Centre Européen de Calcul Atomique et Moléculaire, founded in 1969 at Orsay by Carl Moser. The Panel thought that the UK computational community could benefit greatly from close contact with an organisation with similar interests and complementary objectives.

Thus, the initial CCP programme had many mothers and fathers, working in different environments, who came together at a time and place that now seems especially propitious.

Figure 2: John Pendry

Certainly, the early 1980s were exciting years at Daresbury. First, two new large-scale international accelerator-based experimental facilities were under construction. The Synchrotron Radiation Source (SRS) was the world's first dedicated source of synchrotron radiation, a facility which over its lifetime was used for atomic and molecular physics, solid state and surface physics and chemistry, structural molecular biology, materials science and much more. Experiments contributing to two Nobel Prizes were done on the SRS. Likewise, the Nuclear Structure Facility (NSF) was being built. This was a tandem Van de Graaff accelerator designed for the study of heavy ion collisions and the structure, shape and properties of nuclei. Among the motivations for the expansion of the CCP programme during this period, the need to "interact with and provide theoretical backing for the experimental work [on these

[7]Over the years, the Division grew in both scientific scope and computational expertise, and it changed its name within its parent organisations a number of times. It would be tedious to record all of this here. Hence, in this article, we simply refer to this unit as ***TCSE Department*** – Theory Computational Science and Engineering Department – a fictitious name conflating its various actual titles.

[8]Professor Sir John Pendry FRS joined the new TCS Division in 1977 as Head of the Theory Group at Daresbury. At Cambridge and then Bell Labs, he had pioneered the theory of LEED and EXAFS. When he came to Daresbury he drove theoretical support for the SRS experimental programme, making important contributions to the interpretation of ARPES and XANES spectra. He also established a number of the early CCPs, especially those with a solid state physics theme. The authors can testify to the stimulating atmosphere he brought to the Division in those days. In 1981 he left to take up a chair at Imperial College where he has had a glittering career in physics, notably in developing conformal optics using photonic materials.

facilities]" [6] was certainly important. The list of early CCPs given in Table 1 contains obvious clues to this link with experimental work. Second, the first Cray-1 computer in Europe arrived in Daresbury in the summer of 1979. This was the world's first vector supercomputer[9] - far more powerful than anything previously available to the UK research community. As Burke writes [6], this machine "was installed [at Daresbury] …. to support the work of the CCPs and other groups working in computational science …. and gave tremendous impetus to the work of the CCPs." To this statement, we might add that it was fairly clear that the research community needed expert help to implement its codes on this (at the time) novel architecture machine with anything like reasonable efficiency. This was an obvious role for the supporting CCP staff at Daresbury, and the idea of central support for the porting of code to novel architectures (especially parallel systems) and optimising code performance on them became a key specialty of the Daresbury support for the CCPs.

Quantum chemists (see [7]) were undoubtedly the most powerful initial driving force behind the idea of a CCP, closely followed by atomic and molecular physicists (particularly in Belfast, Durham etc). In solid state physics, people interested in electronic structure, both in bulk crystals (Balasz Gyorffy, Malcolm Stocks, Volker Heine et al) and surfaces (John Pendry and many experimental groups), had ideas for collaborative structures which quickly evolved into CCPs. In these fields, it was apparent that numerical calculation was essential to make progress in research and to avoid falling behind colleagues outside the UK.

For molecular simulation, the key leaders in the field who became involved in the establishment of the CCP5 Project included John Rowlinson[10], Konrad Singer[11], Jack Powles, Ian McDonald, Graham Hills, Julian Clarke, Stuart Pawley, John Finney, Richard Catlow, Ruth Lynden-Bell, Bill Mackrodt. It is interesting and impressive that CCP5 rapidly gained the involvement of many talented young

Figure 3: John Rowlinson (photograph courtesy of Vlad Sokhan)

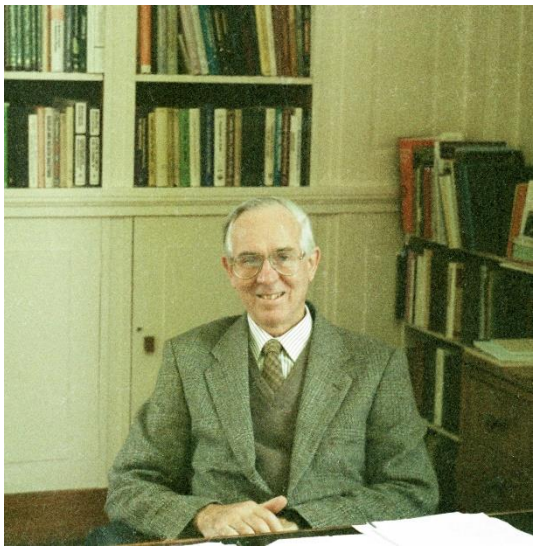
[9]Actually, the Daresbury Cray was, we believe, the very first Cray-1 (serial number 1), although Daresbury was not its first owner – see section 4.3.

[10]In the 1970s, Professor John Rowlinson was the leading theoretician of the liquid state. He played an essential role in the foundation of CCP5 and among his many achievements we note his nurturing of several key members of the Project in subsequent years.

[11]Konrad Singer, an outstanding pioneer of molecular simulation, was born in Bohemia in 1917. In his childhood he showed promise in science and music, but his studies at the University of Vienna were curtailed in 1938, when he fled the annexation of Austria to the United Kingdom. Thanks to generous sponsorship he was able to study for a doctorate at the University of Glasgow until 1941. He worked in industry until 1947 when he became a lecturer at Royal Holloway College, London. His initial research was in quantum chemistry and he was an early pioneer of the use of Gaussian functions in QC calculations. But in the early 1960s he began his research in molecular simulation with a pioneering Monte Carlo study of the thermodynamics of liquid argon. He then moved on to binary liquid mixtures and then to studies of transport properties. He follows these with the simulation of ionic liquids, which saw the first implementation of Ewald's method for calculating long ranged forces. Later in his career he began to apply quantum methods. Using a path integral approach, he successfully determined the thermodynamic properties of neon in the liquid and solid phases. The end of his career was devoted to solving the time dependent Schrödinger equation in condensed phase systems, using a variety of methods, in which he made novel contributions. He died in 2013.

simulators (Dominic Tildesley, Mike Allen, Paul Madden, David Adams, David Fincham, Nick Quirke) who went on to become the next generation of leaders in the field.

CCP5 also had strong support from leading non-UK figures from outset. Participants in the earliest activities of CCP5 included Herman Berendsen, Bruce Berne, Kurt Binder, Keith Gubbins, Gianni Jacucci, Bo Jonssen, Jean-Paul Ryckaert, Daan Frenkel, Giovanni Ciccotti, Aneesur Rahman, etc – an impressive list, many of them strongly associated with CECAM.



We have already noted in this section that the CCPs were encouraged to form close contacts with European computational science[12], mostly in the form of CECAM. In the USA, the National Resource for Computation in Chemistry (NRCC) [10] officially came into being on October 1, 1977, at Lawrence Berkeley Laboratory (LBL). With striking similarity to the CCP model, the concept of the NRCC evolved over many years and was the subject of many national committee and workshop studies. The National Science Foundation (NSF) and the U.S. Energy Research and Development Administration (ERDA) joined forces to sponsor the NRCC and to site it at LBL. However, unlike the CCPs, NRCC did not last for long. It was instructed to wind up its operations by Sept. 30, 1981, despite last-ditch efforts by many computational chemists to save it. The closing of NRCC ended the first effort of the US chemical community to create a centralized research facility where chemists could gain access to equipment too costly for individual laboratories to buy and maintain. In the case of NRCC, the equipment was to have been computers big and fast enough to do massive calculations on problems at the forefront of research, backed by full-time scientific staff for consultation, program development, and in-house research. The end came in July 1980, when an interagency committee of the National Science Foundation and Department of Energy, the funding agencies, voted to cease support. As late as September 1980, hope persisted that NRCC could be saved.
In view of the factors and circumstances outlined in this section, then, it seems reasonable to suggest that the foundation of the CCP programme was probably motivated by a happy combination of scientific altruism (the collaborative spirit) and self-interest (showing value for money, in order to get more money). Perhaps this is true of most programmes that turn out to be successful over a long period.

Over the lifetime of the Programme as a whole, computing technologies of all sorts have been completely revolutionised, funding organisations - both national and international - have come and gone, and individual CCPs have come and gone too. And yet, the essential CCP concept has proved to have a remarkable and perhaps surprising longevity. The story of CCP5 which follows illustrates how this concept worked out in practice within a particular field and community, one that was among the first CCPs to be established all those years ago and is still vigorously active now.

[12]In fact, with the encouragement of Phil Burke, many CCPs played leading roles in the formation of European collaborations in their own fields (eg CCP9 and the Psi-k Network, the European collaboration on the electronic structure of solids), but that is another story.

3 The CCP5 Chronology

We begin by giving a timeline for CCP5, its key people and events, so as to give the reader some rough-and-ready orientation for what follows.

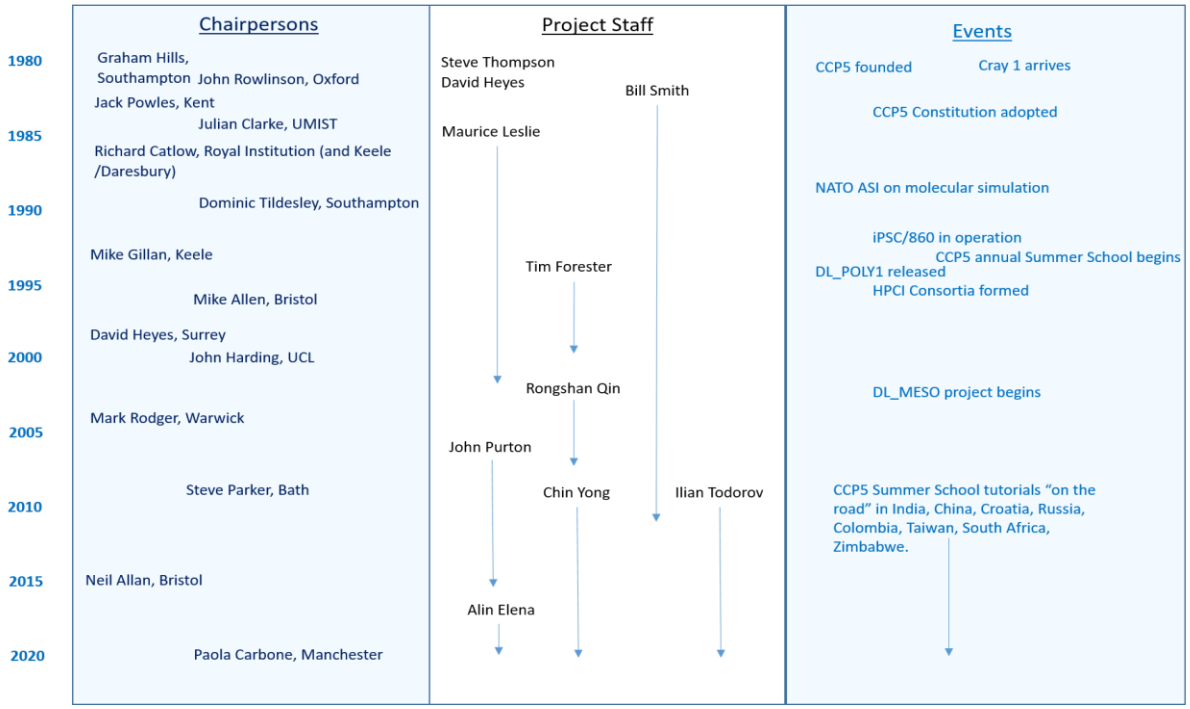


Figure 4: CCP5 timeline

This outline will be filled out in detail over the rest of this section. We note in passing that the project staff were all located at Daresbury laboratory (apart from Steve Thompson and his successor David Heyes, who were located at Royal Holloway College).

3.1 The genesis of the CCP5 idea

The origin of the fifth Collaborative Computational Project (CCP5) was a grant submission to the UK's Science Research Council (SRC) drafted by a community of theoreticians, experimentalists and computational scientists led by Graham Hills (Southampton), Konrad Singer (Royal Holloway), John Rowlinson (Oxford) and others in 1979, with the aim of fostering developments in the computational study of liquids, a field that seemed ready for radical advance by simulation methods. The success of the submission in 1980 established CCP5 as a project under the title: The Computer Simulation of Liquids – a title that belied its actual breadth of interests, since the Monte Carlo (MC) and molecular dynamics (MD) methods employed were equally applicable to all kinds of solids, liquids and interfaces. Indeed, it was not much later that solid state chemistry and materials science became a central theme in CCP5's work (section 3.2). Regarding this development, Richard Catlow[13]

[13]Professor C R A Catlow FRS has been a leading figure in the application of computational methods to problems in chemistry and chemical physics. He chaired CCP5 (see section 3.2) and served as the chairman of the overall CCP Steering Panel. He was the author of an influential review of UK research supercomputing strategy [53].

comments that “the key factor in the materials/solid state field was that the methodologies had reached the stage that they could directly interact with experiment and were being seen by experimentalists as valuable tools and not just theoretical methods that (at best) reproduced known experimental data.”

Dominic Tildesley[14] writes: “I recall the kick-off meeting at Royal Holloway College. I attended as a postdoc based in Oxford so I suspect this would have been 1980. Professors Graham Hills and Konrad Singer played a major role in developing the idea of an interest group that would be bound together by an extensive program library. Although Graham did not perform simulations himself, he was well-advised by his senior postdoc at Southampton, David Adams, who made strong, early contributions to the field. Right from the start, there were two quite separate themes in CCP5, one community performing energy minimisations and lattice dynamics on inorganic solids and another performing Monte Carlo and molecular dynamics on liquids at non-zero temperatures. These two groups ploughed separate furrows in the project for a long time under the care of the two Daresbury staff Maurice Leslie and Bill Smith with leadership from Richard Catlow and John Rowlinson (followed by Jack Powles). Graham Hills was an able politician and he needed to be. This type of project was rather alien to the SERC. They were not disposed to give out grants to groups and in this they were initially supported by a number of senior members of the statistical mechanics and thermodynamics community, who felt that their scarce grant money would disappear into a computational vortex. Nevertheless, good sense prevailed, and the venture got off to a flying start with two or three annual workshops to discuss detailed science and the underpinning computational methodology.”

3.2 The 1980s

That initial grant provided funds for a range of activities. First and foremost, was the fostering of a community of scientists with a common interest in simulation science and methodology, with associated network activities which included conferences, workshops, collaborations, a newsletter and a program library. The community was not to be restricted to the UK and participation by foreign scientists was encouraged. To drive these activities the project was awarded a PDRA to help build the community network, assist collaborations and develop software. In addition, computing time on SRC maintained computers was provided – in particular on the Cray 1s vector computer housed at Daresbury Laboratory. On the establishment of CCP5, Graham Hills was appointed as chair of the project. His appointment was short lived however, and he left after a few months to take up a vice chancellorship. His place in CCP5 was taken by John Rowlinson.

The project PDRA appointed when CCP5 began was Steve Thompson, who was installed at Royal Holloway under Konrad Singer. He commenced immediately with writing software for the CCP5 program library. This consisted of a series of molecular dynamic programs suitable for simulating small molecule systems ranging from rare gas systems (eg Ar, Kr), diatomics (eg HF, Cl_2), triatomics (eg CO_2, H_2O) and small organics (eg CH_4, CH_2Cl_2). These were relatively advanced programs for the

[14]Professor D J Tildesley was a PhD student of John Rowlinson in Oxford, and the co-author with Mike Allen of the well-known book on the computer simulation of liquids (see footnote 15). He has held chairs at Southampton and Imperial College, before moving to a vice-president position in Unilever Research and then became the Director of CECAM. He chaired CCP5 (see section 3.3), authored “A Strategic Vision for UK e-Infrastructure” for the UK government [53] and has been President of the Royal Society of Chemistry.

time and were designed specifically to exploit the Cray 1s at Daresbury. They enabled the calculation of thermodynamic and structural properties of liquid systems, including internal energy, enthalpy, temperature, pressure and radial distribution functions for comparison with experimental properties. These programs were to find wide application in the future project. Unfortunately, for CCP5, Thompson received a compelling job offer from Cornell University and departed soon after these programs were written. His place was quickly filled by David Heyes from the University of Amsterdam and about the same time, Bill Smith at Daresbury Laboratory was allocated to the project to provide additional support.

In March 1981, CCP5 held its first major conference on record, organised by John Finney and Julia Goodfellow at Birkbeck College London, and devoted to the topic of "Polarizability". This drew international participation and included several foreign speakers who would become regular CCP5 participants in later years including: Daan Frenkel, Herman Berendsen, Silvano Romano, David Beveridge, Mihaly Mezei and Paul Claverie. Notable speakers from the UK Included: Richard Catlow; John Murrell, Julian Clarke, Paul Madden, Roger Hockney, Maurice Dixon and Paul Barnes. The participants in the conference constituted the first draft of the CCP5 mailing list. At this conference an *ad hoc* Steering Committee Meeting (SCM) was called, to which the foreign scientists were also invited. At the SCM suggestions for future conferences and the structure of CCP5 were discussed. David Heyes became the project secretary and Bill Smith the program librarian and CCP5 newsletter editor, with David Heyes as deputy editor.

A report on the Birkbeck conference [11] and the associated abstracts formed the basis of the first ever CCP5 newsletter, which was printed on paper and circulated by mail from Daresbury Laboratory. The newsletter ran continuously every quarter year until 1996 and appeared sporadically after that. The newsletter was particularly important because it was the only place authors could put important technical details; the mainstream journals would not publish these [9]. Moreover, from the earliest days it had an international reach; the CCP5 newsletter managed to penetrate the Iron Curtain on the USSR side and was read with interest, for example, in the Ukraine [12].

At this time (early 1980s), the CCP5 community already presented a wide range of interests. The properties of liquids and liquid mixtures were of interest to both theoreticians and experimentalists – particularly those employing neutron or x-ray scattering. There was also a wide interest in the thermodynamics and transport properties of liquids. Free energies were hard to obtain by simulation, but new methodologies were urgently sought. Transport properties like thermal conduction or diffusion were theoretically obtainable through either non-equilibrium simulations or Green-Kubo methods in equilibrium simulations, but the methodology was challenging. There were other interesting areas that were fraught with difficulty but promised great rewards, such as polymers, interfaces, liquid crystals, polar solids, biological molecules and quantum systems. Studies of these challenges were being taken up at the time and the need for better methods, software and extreme computer power were paramount. For the next 10 years, CCP5 was a major contributor to meeting these challenges. In the period 1980-1989, it organised no fewer than 22 conferences and drew participants from all corners of the world, many of whom were invited speakers sponsored by CCP5.

In 1981, John Rowlinson stepped down as CCP5 chair and Jack Powles (Kent) took the position. He continued to serve until 1983, which marked the end of the project's first grant period. Under Jack Powles's chairmanship several successful conferences were organised, the program library, which was now freely available on magnetic tape by request, was expanded (with new programs from Heyes, Smith and David Fincham) and the CCP5 newsletter continued to be published and was well

received. A particularly significant event that occurred during Jack Powles's term was the drafting of the CCP5 constitution, which has strongly influenced the character and function of CCP5 ever since.

At a Steering Committee meeting during a CCP5 conference at Southampton on 19th December 1983, Julian Clarke from UMIST proposed that CCP5 draft a formal constitution for deciding the membership of the CCP5 Steering Committee, as hitherto it had been *ad hoc* and ill-defined. Powles endorsed the proposal and in the following discussion it was agreed that:

1. The CCP5 chairman would hold the chair for three years.
2. Each chairman would be appointed by popular vote of the CCP5 community conducted by Daresbury Laboratory.
3. The administration of the project would be handled by an Executive Committee consisting of the CCP5 chairman, four members of the CCP5 community elected on a two-year cycle by popular vote, the project PDRA and the support staff at Daresbury Laboratory.
4. The role of the Steering Committee would henceforth be advisory. A Steering Committee meeting would take place once a year at the CCP5 Annual General Meeting where the chairman and support staff would report on the project and invite feedback and suggestions for future activities from the community.
5. Membership of the Steering Committee was informal and open to all. A list of members (in the form of an e-mail list) was to be maintained at Daresbury.
6. In addition to managing the project, the Executive Committee was to be responsible for writing a final project report to the funding body (SRC, SERC, EPSRC *etc.*) and drafting a case for renewing the project for a further three years.

This constitution has proved to be robust and, by and large, is still adhered to today. All subsequent CCP5 Chairmen and Executive Committee Members have been elected according to this prescription, ensuring that a continually changing stream of talent with new ideas has refreshed the project. Not all chairmen have adhered to a 3-year term, largely as a consequence of the variation in the lengths of the different renewal periods.

In 1983, the initial project came to the end of its period of grant support and it was necessary to make a new bid to the SRC for its renewal. A proposal was drafted and submitted, but through a misunderstanding, only a case for the network activity was made and the bid was promptly rejected. At about the same time, Richard Catlow was seeking to establish a CCP on the subject of "Polarizable Solids". Thus, when Julian Clarke was asked to draft a new scientific case for CCP5, the polarizable solids area was included. The new proposal highlighted several themes including liquids, polymers, ceramics, quantum simulations, silicates, phase transitions, potentials, interfaces and transport properties. This programme was duly accepted and CCP5 was reinstated under the title of "Computer Simulation of Condensed Phases". Unfortunately, however, the project was not assigned funding for a PDRA (an outcome that was often repeated in future renewal proposals) and the grant period was reduced to 2 years. The project thus lost the support of David Heyes. Nevertheless, it retained the support of Bill Smith and additionally gained Maurice Leslie (who was also based at Daresbury) to lend support to the ionic solids area, which depended on lattice modelling programs. Heyes meanwhile was awarded a Royal Society Fellowship at Royal Holloway College and continued to support CCP5 activities as an independent scientist. He wrote frequent newsletter articles and continued to contribute programs to the program library.

In Spring 1984, Julian Clarke became the first elected chairman of CCP5. He was also the first to run the project through an elected Executive Committee, the membership of which, beside the chairman, consisted of Paul Madden, Richard Catlow, Nick Quirke, Maurice Leslie and Bill Smith. Meeting at least twice a year, the committee proved to be an efficient method of managing the

project and facilitated a burst in network activity, by sponsoring a number of workshops that allowed in-depth discussion of specialist topics, in addition to the more traditional conferences. The Steering Committee of the project was re-cast as an open forum during one of the main conferences of the year, where opinions, suggestions and ideas were drawn from the international audience. By this method, as well as the CCP5 newsletter, the project maintained a dialogue with its beneficiaries.

During this grant period, four major conferences and seven workshops were organised by CCP5. It was also during this period that CCP5 initiated the (later common) practice of sponsoring an overseas expert to tour several universities of the UK to give lectures and engage in discussions with local researchers. This was distinct in intention from merely inviting a foreign speaker to a CCP5 conference or workshop, where deeper or more prolonged interaction was not possible. The first visitor of this kind was Bruce Berne from Columbia University, New York. An expert on many subjects in condensed phase simulation, he visited Oxford, Cambridge and Birkbeck College.

The 1980s were an exciting period for computer simulation science, driven in large part by several developments in computing technology (see section 4 for a detailed account of this). The decade started with the introduction of the first vector supercomputers (eg the Cray 1s at Daresbury and CDC Cyber 205 at Manchester). This rapidly transformed into the era of the personal computer, desktop computing and UNIX-based machines with fast RISC processors. The gains in performance permitted much faster and ever larger simulations, which greatly improved the sampling of phase space, leading in turn to increased statistical accuracy of all calculated properties. Running beside this were the advances in theory and methodology that enhanced the range of what simulation could achieve. Towards the end of the decade, another revolution in computer design brought the first practical parallel computers into the main stream, either as bespoke supercomputers in their own right, or simple clusters of ethernet connected desktop machines – the so-called Beowulf systems. CCP5 played a key role in enabling the simulation community to keep abreast of rapid changes.

In 1986 Richard Catlow was elected CCP5 chairman and was tasked with drafting another renewal proposal to keep CCP5 running from 1986-1989. The proposal document drew attention to the advances in computer technology and theoretical progress and emphasised the growing realism of simulation science. In addition to the established network activities of the project, its targeted areas of interest were molecular and macromolecular systems; polar solids, transport properties of fluids; and quantum mechanical systems. A request was made for a PDRA to undertake adaptation of CCP5's now extensive program library to the emerging computer architectures and to add additional software to address the project's targeted areas. Once again, the request for a PDRA was turned down, but the project was renewed for another two years and kept the existing support of Daresbury staff.

The period 1986-1989 continued in a vigorous fashion with CCP5 organising three conferences on the subjects of: “The Glass Transition”, “New Horizons in the Computer Simulation of Condensed Phases”, and “Industrial Applications of Molecular Simulations”, interspersed with workshops on transport processes, high temperature materials, porous media, and special purpose computers for condensed matter simulation. CCP5 also funded visits to the UK by two leading scientists from Europe: Michele Parrinello (Zürich) visited Oxford, Cambridge and Daresbury; and Herman Berendsen (Groningen) visited London, Cambridge and Daresbury.

In 1988, a proposal to renew CCP5 once again was drafted by Richard Catlow. The bid was for a rolling grant for the period 1988-92, which, in principle, would provide some stability in the continuance of the project by allowing it to run initially for 4 years and later to bid for an extension while the current grant was running. As usual, the proposal highlighted a range of themes.

Foremost was hardware exploitation - the development of algorithms for high-end computers, particularly the Cray 1s, CDC Cyber 205 and Cray XMP/48 that were available to the CCP5 community at the time, but also parallel machines, such as the Meiko M10 and FPS T/20 at Daresbury (see section 4.3). New methods for simulating complex, multicomponent systems were needed. The scientific targets included: the structures and dynamics of liquid crystals; emulsions; macromolecules (polymers and biomolecules); the transport properties of fluids and solids; the physical and chemical properties of materials; quantum methods; and industrial applications. Once again a PDRA was requested – this time to support the work on biological macromolecules. Alas, this was not awarded, but the project was funded sufficiently to run until 1990 and retained the support of Smith and Leslie at Daresbury Laboratory.

Arguably, the most outstanding event that occurred in the life of the renewed project was a NATO Advanced Study Institute (ASI) that took place at the University of Bath from the 5th-17th September 1988, and entitled "Computer Modelling of Fluids, Polymers and Solids". Funded by NATO (with additional sponsorship from CCP5), the ASI was organised by Richard Catlow, Steve Parker and Mike Allen[15] (who were all prominent members of CCP5) and derived much assistance from the CCP5 community (as lecturers and helpers). The format was a 12-day school in which graduate scientists were exposed to the cutting edge of simulation research through a comprehensive range of lectures, backed by practical sessions. Among the overseas lecturers were Charles Brooks (Carnegie-Mellon), Daan Frenkel (Utrecht), Denis Evans (ANU, Canberra), Tony Ladd (Lawrence-Livermore), Dennis Rapaport (Bar-Ilan), Jean-Paul Ryckaert (Brussels) and Joachim Sauer (Academy of Sciences, Berlin). The lectures ranged widely over computer hardware (vector and parallel processing), simulation algorithms (molecular dynamics, Monte Carlo, constraints, path integral methods, non-equilibrium methods), materials (polymers, alkanes, silicates, ionic solids, plastic crystals), interfaces and surfaces. A range of computer simulation exercises was prepared for the practical sessions by CCP5. The educational success of the ASI would later inspire the CCP5 'Methods in Molecular Simulation School', which the project runs annually in the UK.

Under the new grant CCP5 continued with its network activities. It organised three successful conferences under the titles: "Computer Modelling of New Materials" (with CCP9); "Grand Challenges in Molecular Simulation"; and "Novel methods in Molecular Simulation" and a study weekend entitled "Molecular Simulation and Protein Crystallography" (with CCP4). No fewer than five workshops were also held on the subjects: interionic potentials, chemical reactions, polymers, parallel molecular dynamics and minerals (with CECAM). A lively programme of visiting scientists included Kurt Binder (Mainz), Peter Wolynes (Rice), David Coker (Boston) and Dennis Rapaport (Bar-Ilan), all of whom visited several UK Universities, lecturing and sometimes taking part in workshops or research. CCP5 also sponsored collaborations between scientific groups. At least four such collaborations were assisted, in each case to enable travel between the UK and a foreign country.

In the period 1988-1990, parallel computers began to realise their potential as tools for simulation work. Prior to this, machines like the ICL Distributed Array Processor (DAP) had already proved their worth as engines for specialised spin systems (i.e. the Ising model), but the appearance of Transputer-based machines and the Intel Hypercube (one of which was based at Daresbury), offered

[15] Professor M P Allen is one of the pioneers of molecular simulation in the UK, author with Dominic Tildesley of "The Computer Simulation of Liquids" (Oxford Scientific Publications 1989). He has been strongly involved in CCP5 from the outset, chairing the Project as well as co-initiating the Summer School. He also was a coordinator of the Computational Science Initiative (see section 4.4) and also served as Chairman of the CCP Steering Panel, on which he was a forceful champion of the CCP concept and philosophy.

versatile platforms for more conventional kinds of simulation. Within CCP5, several UK groups (eg Daresbury, Edinburgh, Bristol, Southampton, Keele, Canterbury, UMIST and Queen Mary College) began actively to develop software for practical simulations. CCP5 provided an invaluable forum for these developments though its workshops, newsletter, and by enabling collaboration between the groups. As a result, practical parallel simulation algorithms began to emerge and new programs began to appear in the CCP5 program library. SOTON_PAR (developed by Mark Pinches and Dominic Tildesley at Southampton) was one of the earliest parallel domain decomposition programs. Dennis Rapaport independently wrote a similar program about the same time. The program MDMEGA (written by Bill Smith at Daresbury) took ideas from both SOTON_PAR and the Rapaport program to provide CCP5 with a basic parallel program for further development. This later formed the basis of the programs GBMEGA (for Gay-Berne molecular simulations) and DL_MESO (for dissipative particle dynamics). Andrew Raine (Keele) wrote SLS_PRO for protein simulations using systolic loop methods developed in collaboration with David Fincham and Bill Smith.

As these developments were taking place, Maurice Leslie worked with the part of the CCP5 community concerned with the solid state. The primary tool for this work was structural optimisation of crystalline and defective structures by energy minimisation. Leslie's main contribution was the development of three body force models which were incorporated into the crystal defect program CASCADE together with ancillary programs THBREL for perfect lattice properties, THBFIT for potential fitting, and PHONON for lattice dynamics. A key development was the use of THBFIT to model a potential for quartz using three body forces and shell model polarizability. This opened the field of silicate minerals and zeolite catalysts to further study. Other developments were the use of PHONON to simulate thermodynamic properties of solids. Also at this time a symmetry adapted program SYMLAT was written to simulate perfect lattice properties of solids, but only for two body forces. This program was used to calculate defect energies by the supercell method.

3.3 The 1990s

In 1990, Richard Catlow stepped down as CCP5 chairman and Dominic Tildesley from the University of Southampton was elected in his place. At this juncture, a renewal proposal was required to take the project onwards.

The submitted proposal featured an ambitious and diverse programme of research and software development arising from already active research programmes within the CCP5 community. Leading these was the exploitation of new hardware, which incorporated a diverse range of platforms: shared and distributed memory parallel machines, parallel machines with vector processors and mid-range scalar (RISC) computers. The dynamic load balancing of parallel algorithms and the exploitation of Particle-Particle/Particle mesh (PPPM) methods were seen as essential tools for future applications. The proposal identified the simulation of colloids and aggregates, rheological properties and phase equilibria via the Gibbs ensemble methods as being of acute interest. Likewise, the study of polymers and chain molecules, particularly biological macromolecules, was given a high profile. The development of bio-simulation software by a dedicated PDRA was identified as a key scientific requirement. Force field development and methods for free energy calculations were among the priorities. CCP5's interest in materials was also highlighted and ranged from liquid crystals through polar solids, silicate minerals and glasses to high T_c materials. Many of these topics were relevant to catalysis, for which studies of surfaces and interfaces and the application of Car-Parrinello methods together with quantum simulation methods were seen as

important activities. Interaction with industry (and indeed other CCP's) incorporating these research interests was an important objective.

This rich programme of research and development was duly funded for the period 1990-1993, though, as usual, without the requested PDRA, which meant that the development of a biosimulation program could not happen. However, the project did retain the Daresbury staff members Leslie and Smith, who continued to support the project at a 50% level. Under Tildesley's chairmanship, the themes of the renewal proposal were reflected in the scientific conferences and workshops. Under the new grant CCP5 organised (or sponsored) three meetings on computational algorithms and computer hardware, three on liquid crystals or molecular liquids, three on surfaces and interfaces, three on materials (including solid state ionics and polymers) as well as meetings on potentials, finite systems, the statistical mechanics of industrial chemicals and processes, and aqueous systems. In this period CCP5 sponsored several visiting lecturers to the UK including George Malenkov and Vladimir Filinov from Moscow, William Steele from Pennsylvania and Claudio Zannoni from Bologna. Other visitors were sponsored to take part in research projects, including K. Schroeder (Berlin) who collaborated with Maurice Leslie on Zeolite catalysis, Jacob Gavartin (Riga) who collaborated with Richard Catlow on diffusion in lithium ceramics, and M. Amini (Tehran), who collaborated with David Fincham and Benahm Vessal on silicate glasses. CCP5 also sponsored several UK research groups through provision of travel money. The CCP5 Executive Committee generally referred to this as "pump-priming" sponsorship, since it was intended to get collaborations started. An outstanding example of this kind was the liquid crystal group, consisting of Mike Allen (Bristol), Andrew Masters (Manchester), Mark Wilson (Lancaster) and George Jackson (Sheffield), that was sponsored by CCP5 and later became the Complex Fluids Consortium under the High Performance Computing Initiative (HPCI – see section 4.6).

During this period, the experimental parallel computing of the late 1980s evolved into practical software projects. A significant development at this time was the DL_POLY program [18], which in 1992 (when it first appeared) was one of the first general purpose parallel molecular dynamics simulation programs anywhere, and was originally written for CCP5 by Bill Smith at Daresbury Laboratory. The development of this program was much enhanced by grant funding, not as part of the CCP5 proposal itself but from the UK's Computational Science Initiative (CSI – see section 4.4) , which awarded a PDRA in the person of Tim Forester who, with Smith, developed DL_POLY into an advanced program with a wide range of capabilities. The award of a PDRA for this work was a rare and welcome outcome for CCP5 and went some way to offset the failure to secure a PDRA for the proposed bio-simulation program. Indeed, with additional support software developed by Forester, the DL_POLY program proved to be well capable of simulating biological systems and offered the interesting additional option of simulating the bio-inorganic interface. The program performed well on the parallel platforms of the time, showing good scaling with processor numbers and an ability to handle simulations of order 30,000 atoms. DL_POLY continues to be a major asset for CCP5 to this day, having undergone several major redevelopments to keep pace with advancing computer architecture and scientific demands (see sections 4 and 5).

In 1993, Dominic Tildesley stepped down as CCP5 chairman and Mike Gillan from the University of Keele was elected to the position. In the programme of activities offered in the project renewal proposal Gillan highlighted the scientific themes: Complex Fluids and Heterogeneous Systems (which included liquid crystals, Langmuir-Blodgett films and the ordering and phase equilibria of molecules adsorbed at surfaces); Molecular and Macromolecular Systems (including realistic modelling of inter-molecular interactions, either by multipolar electrostatics or Car-Parrinello simulation, drug/DNA interaction in fully aqueous solution and modelling water itself by means of new and accurate multipole potentials); Materials, in collaboration with sister projects CCP1 (quantum chemistry),

CCP3 (electronic properties of surfaces) and CCP9 (electronic properties of solids), particularly on complex systems such as zeolite catalysis and diffusion, fast ion conduction in β-aluminas, defect energies and ion migration by first principles methods; and the structure and formation of geological minerals); and lastly Methodology (including software development for national supercomputers, for example DL_POLY, and the application of first principles methods for accurate determination of the properties of ceramics). The role of the Daresbury-based staff (Tim Forester, Maurice Leslie and Bill Smith) in these projects was highlighted. In addition, Gillan pledged to continue the highly successful networking activities that CCP5 promoted: the conference and workshop programme; sponsoring of foreign visitors; pump-priming collaborations; outreach to industry; and maintenance of the program library and newsletter. On this occasion, there was no need to make a specific bid for PDRA support, since Tim Forester was already in place supporting the DL_POLY project with funding, as noted above, from another grant. This was a piece of luck for CCP5; Tim Forester made a large contribution to its activities.

The renewal of CCP5 for the period 1993-1996 was followed by a surge in progress in all areas. The DL_POLY program, much enhanced by Forester and Smith, was released as DL_POLY 2 in 1994 and was rapidly adopted by the CCP5 community and applied to a diversity of projects running on the Cray T3D supercomputer at Edinburgh. Some of the early projects included modelling the biological activity of valinomycin, structure and transport in silicate glasses, the phase properties of adsorbed molecules, the structure and properties of magnesium oxide and titanium dioxide microparticles; and the structure and stability of ice/hydrocarbon clathrates. In 1994 30 DL_POLY user licences were issued and by 1997 this had risen to 230. The utility of the program and its uptake by the community was perhaps influential in gaining an extension of Forester's appointment by two years through a bid to the Science Board Computing Committee.

Meanwhile, the scientific areas indicated in the renewal proposal proceeded apace. Investigations of polymers, ceramics, fluids, liquid crystals, interfaces and molecular systems continued. But in addition, strong interactions were formed between CCP5, CCP1, CCP3 and CCP9 in the materials modelling area, where Car-Parrinello, Tight Binding and Hartree-Fock methods were used. These first principles methods represented an expansion of the capabilities of the CCP5 community, which had formerly relied almost exclusively on classical Monte Carlo or molecular dynamics techniques, and showed the cross-over interest, dissemination of expertise and generally an overlap of computational practitioners, members of more than just one community. This undoubtedly reflects the quest for greater realism and multiscale view of computational science that was happening in atomistic modelling at the time.

During this period, CCP5 organised 10 workshops and conferences including "Modelling Localised States in Condensed Matter" (1995) with CCP1, "Recent Advances in Molecular Simulation" (1995) with CCP3 and the Royal Society of Chemistry (RSC), "How to derive the Interatomic Potentials needed for Simulation Studies" (1994) and "Advanced Computer Simulation of Materials" (1995). Sponsored foreign visitors included Tony Ladd (Florida), P. Vorontsov-Velyaminov (Moscow), Stefan Estreicher (Texas Tech.) and Martin Schoen (T.U. Berlin). Pump priming collaborations sponsored by CCP5 included: Liquid crystal simulations, a collaboration between the Universities of Bristol, Derby, Durham, Manchester, Sheffield, Sheffield Hallam, Southampton and Daresbury; and the Collaboration on distributed multipole analysis between the University of Liverpool, University College London and Daresbury.

A particularly important occurrence during the 1993-1996 period was the "Methods in Molecular Simulation Spring School", which took place at Southampton in April 1994, at Southampton (again) in March 1995 and Imperial College in March 1996. This was a one-week school, organised by CCP5

worthies Mike Allen, Julian Clarke and Dominic Tildesley (who were also the main course lecturers), and designed to teach the foundations of molecular simulation to postgraduate students. Taking example from the NATO ASI at Bath in 1988, the school was based on a series of lectures with integrated, practical, computer sessions. Additional guest lecturers provided talks on research topics. This was the start of a series of annual schools that fulfilled a valuable role for the following decade. The schools were independent from CCP5's official programme to begin with, though the project provided financial and other support (CCP5 software and personnel) in the manner of other workshops. Gradually however, the workshops became synonymous with CCP5 activities and were eventually (2002) absorbed into CCP5's official programme as its annual Summer School, to be run each year at a chosen UK University. Undoubtedly, the school has had a major impact on the education of postgraduate students, having trained an estimated 1500 students from the UK and elsewhere.

In 1996, Mike Allen from the University of Bristol was elected CCP5 chairman. His proposal for the renewal of CCP5 for the period 1996-1999 offered the continuance of CCP5's network activities (workshops, visitors, pump-priming collaborations, links to other CCPs etc.) and highlighted the development of mesoscopic simulation methods as a scientific theme for special attention. Such methods offered a bridge between microscopic and macroscopic computational modelling (i.e. from atoms to the continuum), and had the potential to tackle difficult areas such as grain boundary diffusion, crack propagation, dislocation dynamics, surface film growth, sintering of powders, polymer morphology control, liquid crystal morphology, colloids, surfactants and fluid interfaces. All of which were too large for conventional atomistic simulation and too complex in nature for macroscopic (eg Navier-Stokes) methods. In contrast, the mesoscale employed methods such as cellular automata, lattice Boltzmann and adaptive mesh schemes to model fluid flow incorporating complicated boundary conditions and interfaces.

The mesoscale area was already an active interest in CCP5, with groups at Bradford, Oxford, Cambridge, Sheffield Hallam and Edinburgh, and in industrial laboratories at Schlumberger, Shell and Unilever. These groups lacked a community in which ideas and methods could be developed and exchanged. CCP5 was positioned to provide such a community, to foster development of software tools in the mesoscopic area and additionally lend its expertise in parallel computing to provide enhanced solutions. As a focus of the activity, it was proposed that a PDRA be placed at the Materials Research Centre at University College London under Marshall Stoneham, to develop methods for generating regular and irregular grids for a range of lattice-based simulation problems in three dimensions, and then apply them to demonstration problems taken from the materials arena.

As things turned out, CCP5 was successfully renewed for the period 1996 to 1999, but once again only the network activities were funded. The proposed flagship project on mesoscale modelling had not been funded on the grounds that it was "too vague". Additional feedback from the grants committee commented that renewal proposals featuring a PDRA request should in future be sent to the relevant science programme. This comment flagged up a recurring problem with the CCP5 renewal proposals: within the grant-awarding structures of the SERC CCP5's requests for a PDRA to support a specific area of research did not sit well with its activity as a network project supporting a wide range of scientific interests. No doubt, this made it hard for grant awarding bodies to decide how, or even whether, a PDRA should be funded. We discuss this issue further in section 7.4.

Despite this disappointment, CCP5 energetically pursued its network programme. Between 1996 and 1999 it organised or sponsored eight meetings including: "Condensed Matter Simulation: Realising the Potential of the Computer" (1996), "Lattice Boltzmann Simulation Techniques" (1996),

"Molecular Modelling and Neutron Scattering Workshop: Techniques for Large-Scale Systems" (1997) (with the Institute of Physics, the Royal Society of Chemistry and ISIS - the Neutron Spallation Source at Rutherford Appleton Laboratory). It is noteworthy that several of these meetings were relevant to the mesoscale area, which had been the highlight of the renewal proposal. In addition, Spring Schools were organised in Bristol in March 1997 and March 1998 and in UMIST in Manchester in June 1999. The CCP5-sponsored visitors in this period were, Stefan Goedecke (MPI-Festkörperforschung, Stuttgart) in June 1996, Florian Müller-Plathe (MPI-Polymerforschung, Mainz) in June 1997, Dennis Rapaport (Bar-Ilan, Israel) in August 1997, Bruce Boghosian (Boston, USA) in 1997, Juerg Hutter (MPI-Festkörperforschung, Stuttgart) in April 1997 and Simone Melchionna (Rome) in June 1998. Five pump-priming collaborations were sponsored including: "Polarization model of zirconia" with Mike Finnis and Neil Marks (Belfast) and R. Wheatley (Nottingham); "Studies of calcite" with J.O. Titiloye (Abertay, Dundee) and Steve Parker (Bath); and "Structure of trihalides" with P. Salmon (UEA) and Paul Madden (Oxford).

In 1997 Tim Forester's contract ended and the DL_POLY project was left entirely in the hands of Bill Smith, who continued to add features to the program, most significantly long ranged (electrostatic) methods based on particle-mesh methods. In 1998, he began to develop DL_POLY 3, a radically different program from DL_POLY 2, exploiting domain decomposition as the primary parallel strategy. The aim was to make possible extremely large simulations of order 1 million atoms and beyond. However, without postdoctoral support, the necessary constant user support demanded by DL_POLY 2 made developments slow, and a viable program did not become available for several years. The program undoubtedly lost its strategic advantage as a result, though it remained popular. During this time period Maurice Leslie worked on the program DMAREL which modelled the inclusion of distributed multipolar electrostatic forces into perfect lattice simulation. The effect of polarization of the ions and anisotropic repulsions were also included.

The year 1998 saw the election of David Heyes (the former CCP5 PDRA) from the University of Surrey as CCP5 chairman. His first task was to draft the next CCP5 renewal proposal, which was due in 1999. In the proposal, CCP5 pledged to support a broad range of activity in particle simulations of condensed phases. In addition to its traditional role in the area of atomistic modelling, it would place special emphasis on three emerging areas requiring significant developments. The first of these was the incorporation of quantum mechanics in atomistic simulation codes, using semi-empirical *Tight Binding* or First Principles *Car-Parrinello* simulations, to generate the inter-atomic force field (required for dynamical simulations), "on the fly". There was quite close contact with CCP9 regarding the Car-Parrinello methods, given the central role of electronic structure calculations in them, but perhaps less by way of actual collaboration. To make progress in this area, CCP5 would collaborate with CCP1, CCP3 and CCP9. The second area was the adaptation of atomistic simulations methods to large-scale *mesoscopic* phenomena, such as metallic and van der Waals clusters, inorganic nanoclusters, adsorption in pores, indentation and surface rearrangement by nanoprobes, material fracture, diffusion at grain boundaries, liquid crystals, surfactants in micelles and at interfaces, polymer melts and glass formation, nucleation and crystallisation from solution. The key development required for such modelling was multi-timescale algorithms to handle fast and slow degrees of freedom. The final area emphasised was *coarse-grained* or *mesoscale s*imulations of single and multiphase systems on the length scale 10 nm to 100 μm, for example from liquid crystals, surfactants and colloids through to dry powders. The relevant techniques in this area were Brownian dynamics, Stokesian dynamics, lattice Boltzmann, dissipative particle dynamics, smoothed particle hydrodynamics and granular dynamics.

It is evident that some of these proposed areas had been brought forward from the previous renewal proposal, principally because they had continued to grow in importance. Significantly,

perhaps, in the light of previous proposals, no request for a supporting PDRA was made. Instead, named CCP5 collaborators and the supporting Daresbury staff (Bill Smith and Maurice Leslie) would undertake the work. Such a frugal proposal was undoubtedly attractive and was duly funded for three years (1999-2002).

3.4 The 2000s

In the year 2000, John Harding was elected to succeed David Heyes as CCP5 chairman for the remainder of the existing grant. During this period CCP5 continued to progress its activities with its usual energy. No fewer than 11 conferences and workshops took place in the 2000-2002 period with CCP5 as the organising body or as sponsor. These included: "Molecular Simulation in the 21st. Century" (Guildford, July 2000); "Statistical Mechanics and Molecular Simulation of Nucleation and Growth", (King's College, London July 2001, with the SIMU Network funded by the European Science Foundation); "Liquid and Liquid Interfaces", (Warwick September 2001) and "Advances in Simulation of Molecules and Materials", (Durham, September 2002). There were also workshops on long ranged forces, calcite minerals and interfaces and postgraduate workshops were organised in London and Bristol under the banners of the London Atomistic Simulations Group (LASG) and South West Computational Chemists (SWCC), all CCP5 sponsored.

During the 1999-2001 period, the CCP5 sponsored Spring Schools were managed by Professors Allen, Clarke and Tildesley. These took place at UMIST, Manchester (1999-2001) with additional sponsorship from the Statistical Mechanics Group of the Royal Society of Chemistry. In 2002 the original organisers stepped down and the responsibility for the future schools was taken up by the Summer School Working Group, a sub-branch of the CCP5 Executive Committee. The first Summer School organised by this group took place at King's College, London in July 2002, with additional sponsorship from SIMU.

The pump-priming programme for sponsoring collaborations in the CCP5 community supported eight projects during this period. Examples include: Daresbury Laboratory and Cambridge using path integral methods to look at quantum tunnelling in silicate glasses; Loughborough and Cambridge looking at carbon implantation by high energy ions; and a Birmingham and Bristol collaboration on simulated protein structure and folding dynamics. The overseas visitors sponsored by CCP5 in this period included Frederike Schmid (MPI-Polymerforschung, Mainz), S. Berry (Chicago), Mark Tuckerman (New York), John Tse (NRCC), Scott Auerbach (Massachusetts), Stefan Estreicher (Texas Tech), Jose Alejandre (Mexico), Billy Todd (Melbourne), Sharon Glotzer (Michigan) and John Kieffer (Michigan).

The software developments that took place at this time included major DL_POLY enhancements. A Java Graphical User Interface (GUI) was written by Bill Smith to facilitate use of the molecular dynamics program for research and tuition. DL_POLY 2 was adapted to run on parallel-vector computers in collaboration with Riken, Japan. Also, at this time a parallel dissipative particle dynamics program (DL_DPD) was written, based on the MD_MEGA program. Furthermore, the OXON tight binding program was incorporated into the CCP5 program library by Maurice Leslie. He also applied symmetry adaption to DMAREL giving a new program DMACRYS. This program allowed the calculation of zone centre phonon frequencies giving the possibility to calculate thermodynamic properties of the crystal. A program DL_MULTI [74] was developed from DL_POLY for the molecular dynamics of rigid molecules using distributed multipole analysis. With such a full programme of activities CCP5 was clearly a vital and stimulating project at that time.

In 2002 CCP5 was due for renewal. Under John Harding's direction a proposal was drafted that highlighted mesoscale simulation as the main theme. A new multifaceted program to be called DL_MESO was proposed, which would incorporate a range of mesoscale methodologies including Dissipative Particle Dynamics), Lattice Boltzmann and Smoothed Particle Hydrodynamics and incorporating a wide range of boundary models. The plan was to have a graphical front end to access its many features and would be run on a range of parallel architectures, including the so-called GRID, composed of computers distributed over the internet. Associated with this development would be a programme of conferences and workshops promoting the uptake of mesoscale methods by the CCP5 community and training novices in the use of the new program. To develop the program and liaise with project members, a PDRA was required. The PDRA would be based at Daresbury Laboratory and the scientific and technical agenda would be set by a Management Committee consisting of Christopher Care (Sheffield Hallam), Peter Coveney (Queen Mary, London), Patrick Warren (Unilever), David Emerson (Daresbury, CCP12), David Heyes (Surrey), Sebastian Reich (Imperial), Bill Smith (Daresbury) and others. In addition to this flagship project CCP5 would continue its traditional role of supporting atomistic simulation through its network activities and the existing program library. Somewhat against the trend of previous occasions, the bid for the flagship project was wholly successful and CCP5 was fully funded until 2006.

The mesoscale project commenced with the appointment of Rongshan Qin as the PDRA to develop the package DL_MESO. His first task was to write a Lattice Boltzmann program using C++ and the supporting graphical user interface (GUI), which was written in Java. Using the GUI, a model system could be constructed, simulated, and analysed. The package, with documentation, was made available to the academic community and quickly found a substantial user base. By 2006 the package had 70 registered academic users and had been taken up by 2 commercial companies. At a later stage a version of Smith's DL_DPD program was incorporated to provide a particle based methodology. Qin undertook a number of scientific studies and collaborations with groups at UCL, Daresbury (CCP12), Southampton, and Brunel. Qin also carried out his own studies on a diverse range of topics including: phase separation in binary fluid mixtures; the effect of flow on the kinetics of the solid-liquid interface; and the effect of electro-pulsing on cell membranes.

A major boost to the development of DL_POLY_3 occurred through the eMinerals project (2002-2008, funded by the Natural Environment Research Council, NERC, as part of the UK eScience programme), managed by Martin Dove at Cambridge. The project adopted DL_POLY_3 into its high performance computation programme, as well as associate it with its workflow and data curation programmes that were part of eMaterials, an STFC funded eScience programme. This led in 2003 to Ilian Todorov being seconded from Cambridge to Daresbury Laboratory for the duration of the project, where he became the main software developer and project lead for DL_POLY 3 under Bill Smith's mentorship and guidance. This accelerated progress significantly and the first public version (DL_POLY 3.1) appeared in April 2003. A notable feature of this version was the fully distributed Smooth Particle Mesh Ewald (SPME) for electrostatics evaluations implementation that incorporated a distributed 3D Fast Fourier Transform developed by Ian Bush at Daresbury, which exhibited impressive scaling on large scale parallel computers, such as the 2560 processor HPCx (see section 4). For the life of the eMinerals project Todorov functioned as the chief developer and principal custodian of the DL_POLY 3 package.

While the mesoscale project and software development were progressing, CCP5's network programme continued with a series of conferences: "Advances in Simulations of Molecules and Materials", at Durham, September 2002; "Simulation of Materials - Techniques and Applications", at Cardiff, September 2003; "New Perspectives in Modelling and Simulation: Theory and Applications", at Sheffield, September 2004; and "Computational Techniques and Applications to Materials

Modelling", at Keele, August 2005. The six workshops organised during the 2002-2006 period included: the Joint CCP3/CCP5 Workshop on "Nanoscale Modelling", at Cambridge January 2003 (with 3 other organisations); "Modelling of Materials: Atomistic and Ab Initio Approaches", at Oxford, April 2003 (with RSC); "Complex Inorganic Materials: Crystal Growth"; and "Proton Conduction in Diverse Media", at Cambridge, April 2005 (with EPSRC). In addition to these traditional workshops, several more informal postgraduate workshops were sponsored: the South West Computational Chemists (4), the Northern Atomistic Simulation Group (2) and the London Atomistic Simulation Group (1).

CCP5's overseas visitors program flourished. The scientists CCP5 invited to the UK during the 2002-2006 period were: P. Kusalik (Dalhousie); C. Chakravarty (Indian Institute of Technology-Delhi); B. Pedro Uberuaga (Los Alamos National Laboratory); J Shelley (Schroedinger Inc., Portland); Peter Cummings (Vanderbilt University and Oak Ridge); Denis Rapaport (Bar-Ilan); and F. Illas (University of Barcelona). These visitors followed the usual programme of travelling between universities giving lectures and engaging in discussions with local scientists.

Also, between 2002 and 2006 CCP5 "pump primed" seven scientific collaborations with travel funding including: "Calculation of the solid-liquid interfacial free energy of a water model" between King's College London and the University of Leicester; "Computational studies of oxyfluoride superconductors" between the universities of Surrey and Hull; "Phase transitions and thermodynamic stability of phases of solid water, including ices and gas hydrates" between Heriot-Watt University and the University of Warwick; and "The effect of salt solution on mineral growth and dissolution" between the universities of Bath and Cambridge.

In 2004 John Harding had stepped down as CCP5 chairman and in his place Mark Rodger[16] from the University of Warwick was elected. Harding's term had been notably successful and progress continued with the new chairman until the grant period ended in 2006. However, things became difficult when CCP5 sought renewal beyond 2006. Surviving documentation shows that at least four attempts were made to continue the project after 2006 and it was in fact not properly funded again until 2009!

The first of a sequence of unsuccessful renewal attempts was made in the spring of 2006. Mark Rodger and John Harding put together the proposal's flagship project, which was concerned with extending the time scales accessible to molecular simulation using the methods of *hyperdynamics* due to Voter (i.e. temperature accelerated dynamics, or TAD, and bias potential dynamics, or BPD) and *metadynamics* due to Laio and Parrinello. A new parallel and grid exploitable program was proposed: DL_HYPER, to incorporate and combine these techniques to study a range of scientific topics with long natural time scales. Proposed topics of study included the immobilisation of radioactive waste, alkali diffusion in glasses, drug molecule polymorphism and crystal nucleation in aqueous solution. In addition, the renewal proposal sought to preserve CCP5's network activity of annual meetings, local and international workshops, visitor programme, pump priming of promising collaborations, Summer Schools and software training. Once again this bountiful offering suffered as previous renewal proposals had done. Despite being strongly approved by referees, the flagship project was not funded. But unlike previous occasions, the network programme was also rejected.

[16] Mark Roger died in March 2017 after a short illness. In his role at Warwick, Roger's principal research interests were in molecular simulation, where he was active both in developing methods and applying them to important biological and industrial problems.

When delved into by Mark Rodger, the explanation for this outcome appeared to be that changes in the funding arrangements for the CCPs had resulted in the bid going before a peer-review panel at EPSRC that, unlike the previous funding body, had responsibility for other scientific areas beyond computational science alone and was less disposed to funding such activity. Furthermore, it appeared that a proposal with both scientific and network components did not sit well with the review panel's funding brief. Mark Rodger was therefore advised to resubmit the proposal without the flagship project and with an assurance that the review panel would be properly advised beforehand about the strategic role CCP5 played in the simulation community.

Following this advice, a grant submission was made in September 2006. The new proposal was for a network project without a scientific flagship, but instead emphasised support for a number of scientific areas. These included long time scale phenomena, bio-materials and bio-mineralization, granular materials, bilayer interactions, free energies and equilibria, and mesoscale methods. These areas would be vigorously supported by the CCP5 network programme, which was detailed in full. Despite the huge changes made, the proposal failed once again to acquire funding. Feedback from the referees strongly suggested that they had not been adequately briefed as promised and as a result had found it hard to respond favourably to a network project that was in competition with scientific research projects. Issues even arose concerning the "full economic cost" of attendance at CCP5 administrative meetings by committee members, though the frequency of meetings and the numbers attending were arguably appropriate to the level of activity managed. Rodger responded vigorously to the referee's comments and highlighted some evident misapprehensions of CCP5's role, but there was no other option than to try another submission.

A third renewal proposal was put together for submission to EPSRC in March 2007. This was also cast as a network project, with all the usual activities of CCP5, and which identified four principal scientific themes key to the future progress and reputation of molecular simulation in the UK, these being: methods to extend the time scales of molecular simulation; simulations of nano-particles and materials with complex nano- and micro-structures; exploitation of hardware developments, particularly Hector and HPCx (see section 4); and development of an atomistic understanding of new and re-emergent energy technologies. It is not clear from surviving documentation if this proposal was ever submitted, but it was not successful if it was. However, a fourth proposal was submitted in September 2008, which carried over the scientific themes from the March 2007 version. No doubt to Mark Rodger's relief, barring a few questions raised about proposed travel costs which required further clarification, this proposal was accepted and the CCP5 network was funded for the period 2009-2012.

While the saga of CCP5's renewal rolled on from 2006 to 2009 it would be reasonable to expect that the project fell into inertia. However, to the contrary, CCP5 continued to function without direct funding. This was possible for a number of reasons. The support staff at Daresbury Laboratory were able to continue their work since they were permanently in place and funded through other programmes. In addition, the director of TCSE Department at Daresbury, Paul Durham, allocated a small but adequate fund to enable the Executive Committee to continue meeting. Planned conferences were able to go ahead as self-financing operations. In particular, the CCP5 Summer Schools were able to continue running annually with substantial support from Marie Curie Actions and the participation fee charged to the students. These were held at Cardiff (17-25 July 2006), Sheffield (8-17 July 2007) and Sheffield (6-15 July 2008). Also, Annual General Meetings were held on: "Multiscale Modelling", 29-30 August 2007 (Cambridge); and "Surfaces and Interfaces", 8-10 September 2008 (London). Evidently, even without proper funding, CCP5 retained enough coherence and energy to function for 2 years in a constructive manner, and also remained confident that funding would return (as it did; EPSRC were actually rather encouraging in their dealings with

CCP5). This is testimony to Mark Rodger's determination and drive at a difficult time. Nevertheless, this activity represented a stripped-down operation, as funding could not fully sponsor all of CCP5's usual activities. It was inevitable that pump-priming of collaborations and invitations to overseas scientists to lecture in the UK were affected. A further unfortunate consequence of CCP5's informal existence during this period, was that its activities were not properly documented since, apart from the Summer School sponsors, there was no official requirement for any account to be produced. The restoration of these "missing items" from CCP5's portfolio of activities followed immediately when the new grant came through.

3.5 The 2010s

The securing of CCP5's grant for the period 2009-2012 meant that a full programme of activity could proceed as planned. Mark Rodger called a CCP5 Executive Committee meeting in March 2009 and began organising future workshops, the Summer School, overseas visitors and the Annual General Meeting for 2009. The work of the Daresbury staff was also discussed and members were brought up to date on the software developments that were under way, in particular the DL_POLY_4 and DL_MESO packages, which were restructured to make them conform more closely with software standards and more easily adapted to different machine architectures. At this meeting Rodger also announced his intention to step down as chairman. In the subsequent election Steve Parker from the University of Bath became chairman.

The Daresbury staff linked to CCP5 at this stage consisted of Bill Smith (who was due to retire in 2011), John Purton (who had replaced Maurice Leslie in 2008) and Chin Yong, who together provided the core support, each at a 50% level. Additionally, Ilian Todorov (leaving Cambridge for Daresbury Laboratory in March 2007) and Michael Seaton (joining Daresbury Laboratory in 2008 straight after his PhD at Manchester University), were a part of the High-End Computing Architecture team, tasked with assisting development of the DL_POLY and DL_MESO codes and Laurence Ellison was a PDRA collaborating with Todorov and Smith on DL_POLY via an EPSRC funding to explore parallelisation paradigms for particle based dynamics (led by Smith and Roger 2007-2009). CCP5 undoubtedly gained much synergy from this arrangement of personnel. The range of simulation software the Daresbury staff were developing at this stage were DL_POLY, DL_MESO, DL_MONTE, DL_FIELD, DL_ANALYSER and ChemShell.

One of the earliest issues Parker had to deal with was the drafting of a reply to the request of the funding agency, the EPSRC, for a 'Statement of Need for the CCPs' in 2011. The purpose of this was to assist the EPSRC in assessing priorities for funding in its overall programme. At this stage the CCP's had been running for 30 years, so it was reasonable to query their continued existence. Replies were required to focus on a number of areas: the science areas covered by the CCP; an explanation of the current requirement for the CCP; a summary of current support; the requirement for the support; the justification of the support and the beneficiaries of the CCP. It was clear that the request was intended to probe deeply into the purpose and value of every CCP and evaluate its place in UK science. A serious and well-presented case was demanded if a CCP was to continue. While the task was certainly not easy, Parker was able to call on CCP5's leading lights for supporting evidence and a robust reply was drafted. Gratifyingly, EPSRC's response to CCP5's submission was positive. The review panel noted CCP5's well developed network activities, its impact on UK science, its track record in leading science, its openness to newcomers, its progressiveness and its strong engagement with industry. On this basis the future of CCP5 was assured. Overall, the responses to the EPSRC's request led to a clearer assessment of their worth in the EPSRC programme. One impact of this was to increase the CCP stable, in particular by creating the new CCPs: CCP-NMR and CCP-BIO, both of which shared interests with

CCP5 with respect to molecular simulation. CCP5 was to hold several joint meetings with both projects in the following years.

In 2011 CCP5 made a successful renewal bid to the EPSRC to continue CCP5's now familiar network activities (2012-2015) but also to fund a PDRA on Multiscale Modelling. The aim was to address an important challenge faced by the soft matter community, which was to understand and predict the phase behaviour and stability for systems that were too large or too slowly evolving for atomistic MD to be practical, such as polymers, liquid crystals and biomolecules. The intention was to introduce coarse graining into the DL_POLY code, which would allow the package to reach longer time and larger length scales. The project recruited Dr Andrey Brukhno[17], who started in Jan 2013 and began to introduce systematic coarse-graining (SCG) methodologies in DL_POLY. He developed new code to link to the Mainz open-source Versatile Object-oriented Toolkit for Coarse-graining Applications (VOTCA[18]) code which incorporated several well-established SCG techniques such as Iterative Boltzmann Inversion (IBI), Inverse Monte Carlo (IMC), Force Matching (FM), Relative Entropy (RE) and Thermodynamic Force (TF) methods. The work resulted in a program, DL_CGMAP, that allowed DL_POLY users to benefit from the SCG functionality of VOTCA. DL_CGMAP had two parts, a stand-alone analysis tool for automatically mapping between the full-atom and coarse-grained representations in DL_POLY format. Along with creating a template for the corresponding CG force-field file (topology), distribution functions necessary for obtaining effective CG potentials were calculated, thereby facilitating preparation of a CG simulation. Additionally, this same utility allowed (previously unavailable) analysis of DL_POLY simulation results based on grouping of atoms into larger entities. The second part was a fully featured VOTCA style (C++ based) interface between DL_POLY and VOTCA. The work was carried out in collaboration with the VOTCA development team and the interface routines were incorporated into later releases of the VOTCA package, so that researchers using DL_POLY would be able to employ all the SCG methodology available in VOTCA, including any future developments.

In 2013, the EPSRC issued a call for grant proposals under the banner 'Widening Participation for the EPSRC-funded CCPs' to which Parker, with other members of CCP5, responded. Support for two projects was requested and ultimately funded. The first was 'Engaging with the Maths Community', which sought to improve contact between mathematicians and the molecular simulation communities. Funding of this project led to a series of intensive joint workshops with CCP5 and MDNet (a mathematics project concerned with simulation algorithms). The first of which, on the subject of 'Mathematical Challenges in Molecular Dynamics', took place in Warwick 2-5 April 2013. The second project was 'Engaging with Industry', which aimed to foster interaction between industry and academia through funding summer research experience for undergraduate students. Under this project 5 pairs of students from university and industrial laboratories worked on topics of industrial relevance. Five industrial companies were involved, which included Unilever, Sharp Laboratories Europe Ltd., Accelrys Inc., Nestle and the National Nuclear Corporation.

CCP5 also responded positively to other EPSRC calls. A grant submission was made under the 'Software Infrastructure' call on 'Sustainable Software for New Paradigms in Condensed Phase Modelling', which was not funded. Another grant submission entitled 'CCP5 Flagship DL_MONTE:

[17] It is worth noting that Brukhno worked with Todorov to implement potential distribution functions and associated analysis within DL_POLY 4.

[18] http://www.votca.org

Monte Carlo Simulation of Condensed Phases' which aimed to accelerate development of the CCP5 Monte Carlo program, was sent to the Software for the Future II call and was successful.

In the middle period of the grant period CCP5 was required to submit to the EPSRC a 'Mid-Term Review' document outlining the progress the project was making. In the normal run of things this was an unusual request, but as with the previous Statement of Need (see above) Parker was able to respond with a fulsome document detailing the wide programme of activity that CCP5 was undertaking (which is recounted here). All the usual network activities were in full flow and the flagship project on multiscale modelling was making excellent progress. The support given by Daresbury staff and the wide range of software they developed and maintained for CCP5 was also highlighted. An additional document provided a number of 'Case Studies' exemplifying the use of the CCP5 programs. These included: a collaborative study with Liverpool University on superoxide dismutase (DL_POLY, DL_FIELD and DL_ANALYSER); a collaborative project with Unilever entitled 'Computer Aided Formulation' (DL_MESO); and a collaborative project with Liverpool and Bath universities on the 'Uptake and Diffusion of Gases in Inert Matrices' (DL_MONTE). If the EPSRC needed reassurance of the value of CCP5, the mid-term report surely provided it. No doubt this was re-enforced in an interview with the EPSRC attended by CCP5 representatives Steve Parker, David Quigley and Andrey Brukhno.

Throughout Steve Parker's chairmanship, which lasted from 2010 - 2015, CCP5 continued with its programme of network activity. The Annual General Meeting, that CCP5 had held regularly throughout its existence continued to be held. Parker himself organised the 2011 AGM at Bath on 12th-14th September on the topic of 'Recent advances in condensed phase simulations'. The Summer Schools also thrived and were held at Queen's University Belfast (2010-2011), Cardiff University (2012), and Manchester University (2013-2015), always with a strong and enthusiastic attendance. As always there was an abundance of workshops organised by CCP5 members themselves with financial support from the Executive Committee and sometimes support from other organisations in joint meetings, such as those organised in collaboration with CCP-BIO, CCP-NMR and MDNet. Other workshop examples include: 'Adsorption in polymer materials' (Manchester 2012); 'Modelling of Li-Battery Materials', (Huddersfield, April 2013); 'Non-equilibrium molecular dynamics' (Imperial College, June 2013); 'Applying quantum mechanics to alloy design for nuclear reactor steels' (Edinburgh, June 2013). Regional workshops were also sponsored, such as those arranged by the Northern Atomistic Simulation Group and the South West Computational Chemists. Also of note were the training workshops organised by the Daresbury team, covering the suite of software packages developed at Daresbury. These meetings took place at Daresbury, London and Warwick. Indeed, the workshops could take place at any university requesting one. At least six collaborations were sponsored by CCP5 in this period through pump-priming. The collaboration between Ruslan Davidchack (Leicester), Michael Tretyakov (Nottingham) and Thomas Ouldridge (Oxford) on the development of new Langevin and Brownian dynamics equations and numerical algorithms for constant temperature simulation of rigid bodies and between P. Martin and Robert Cywinski (Huddersfield) and Steve Parker (Bath) on atomistic simulations to investigate the high temperatures and pressure structural stability of uranium doped thoria, are typical examples. Finally the programme of inviting visiting lecturers from overseas was fully populated with visits by Xavier Daura (Universitat Autònoma de Barcelona, October 2010), Glen Martyna (IBM-Watson, November 2011), Peter Bolhuis (Amsterdam, Jan 2012), Gregory Voth (Chicago, March 2012), Coray Colina (Penn. State University, May 2012), Sebastien Kerisit (Pacific North-west National Lab., September 2012), Kristen Fichthorn (P.S.U),May 2013) and Julian Gale (Curtin University, April 2014).

Steve Parker stepped down as Chairman in 2015 after Neil Allan from the University of Bristol was elected as his successor in May 2015. His term began in January 2015 and the CCP5 new grant

started in April 2015, following an EPSRC grant proposal by Neil Allan in November 2014. From Parker, Allan inherited a thriving and healthy CCP. CCP5 had also been successful with its flagship proposal to EPSRC (led by Parker and Wilding at the University of Bath and Purton at Daresbury Laboratory) to develop DL_MONTE[19] into a world-class Monte Carlo code. The flagship project inherited the MCX[20] software know-how from Neil Allan's group (Mikhail Lavrentiev, Ilian Todorov, John Purton, Gustavo Barrera) and started officially in April 2015 with Andrey Brukhno as the project PDRA at Daresbury Laboratory and two other PDRAs due to start in July. Later, Vladimir Sokhan and Ivan Scivetti would join CCP5 in January 2017 and March 2017 respectively, after the departure of Laurence Ellison in April 2016.

An important issue for CCP5, as well as for all STFC supported CCPs, at this time was the rebuilding of the CCP5 website. Since 2014 it had been apparent that the old web infrastructure needed upgrading and its underlying server infrastructure had become antiquated. Unfortunately, since the website provided the gateway to much of CCP5's activity, the rebuild was to take longer than expected. Thanks principally to Chin Yong at Daresbury, the new CCP5 website eventually came online in 2017 with a new content management system hosted on a virtual Apache server based at RAL.

Another early issue concerned the CCP5 visitor program. In the past CCP5 funded the international travel of visitors, while the host universities paid the UK travel costs, but it was now evident that UK rail travel often exceeded the cost of international flights, so some addition assistance from CCP5 was demanded.

At a CCP Steering Panel meeting in June 2017 Allen reported that CCP5 had a UK membership exceeding 450 and an international membership exceeding 1000. The UK membership included academic staff from 35 different UK universities and a minimum of 18 other UK industrial, charitable or government organisations. CCP5's software development and support included five active projects: DL_POLY, DL_MESO, DL_MONTE, DL_FIELD and ChemShell, which together accounted for over 4,000 active licence holders worldwide and over 500 google scholar citations in 2016. These programs, which enable molecular modelling from the quantum regime to the mesoscale, were collectively known as the 'DL_Software' suite.

DL_Software was the vehicle of a number of scientific collaborations between Daresbury and the universities. One example collaboration involved Bill Smith, Ilian Todorov and Chin Yong at Daresbury with John Satherley and David Cooper at Liverpool University on the atomistic interaction behaviour of small gas molecules in organic binary solvents, which provided for a PhD student. Elsewhere, the versatility of DL_Software was demonstrated in the ADDoPT project WP4 (led by Kevin Roberts of University of Leeds) through a series of presentations and meetings. This led to the decision to include DL_Software as the principal software infrastructure for molecular simulations of organic crystal dissolution. In another example, the value of DL_Software for educational purposes was shown by the adoption of DL_POLY 4 by Kostya Trachenko for his masters course at Queen Mary College.

---

[19] DL_MONTE: a multipurpose code for Monte Carlo simulation, A. V. Brukhno, J. Grant, T. L. Underwood, K. Stratford, S. C. Parker, J. A. Purton, N. B. Wilding, Molecular Simulation, doi:10.1080/08927022.2019.1569760, Simulation of mineral solid solutions at zero and high pressure using lattice statics, lattice dynamics and Monte Carlo methods, I.T. Todorov, N.L. Allan, M.Yu. Lavrentiev, C.L. Freeman, C.E. Mohn and J.A. Purton, *J. Phys.: Condens. Matter*, **16**, S2751- S2770 (2004) doi:10.1088/0953-8984/16/27/011.

[20] Ab Initio Calculation of Phase Diagrams of Oxides, M. Yu. Lavrentiev, N. L. Allan, G. D. Barrera, and J. A. Purton, J. Phys. Chem. B 2001, 105, 3594-3599.

The years 2015-2020 saw frequent advances in the DL_Software programs, resulting in new capability and new releases for all programs based not only on CCP5 funding but also on a number of other EPSRC funded grants and contributory effort from the Material Chemistry Consortium. Some of the enhancements included the following. DL_POLY: electrostatic multipoles and dipole polarisation (2016, EPSRC-NSF collaborative funding to Todorov); an interface to PLUMED (2016, EPSRC funding to Todorov) – a meta-dynamics plugin; umbrella sampling (2017); a two temperature thermostat model for radiation damage studies (2018, EPSRC funding to Todorov and Trachenko); major code restructuring to facilitate future releases (2018); Empirical Valence Bond and Forward Flux Sampling (2020 – the 2016 CCP5 flagship to Alla, Todorov and Elena); Python 3 graphical utilities (2019); and a Jupyter notebooks infrastructure, to help create workflows for data analysis pipelines. DL_MESO: SPME technology for DPD simulations of charge-based interactions (2017); and a post-simulation Widom's insertion facility for chemical potential calculations (2017). DL_MONTE: Free energy difference methods, lattice switch Monte Carlo, user manual and improved usability (2016); new potential models (force-fields); trajectory output formats compatible with DL_POLY and VMD software; and an optimized Ewald summation method, phase switch and planar pore constraint "slit" (2017); DL_FIELD: new solvent models and force fields and introduction of the Standard DL_F Notation for CVFF (2016); and multiple force fields (CHARMM variants and CGenFF) for multiple organics and hybrid organic/inorganic systems (2016). ChemShell: Py-ChemShell (v19.0 – EPSRC funding to Keal, Catlow and Woodley), a Python-based redevelopment of Tcl-ChemShell (2019).

Staff issues and turnover among the Daresbury CCP5 support staff was a particularly challenging issue during Neil Allan's chairmanship. The unfortunate ill health of the CCP5 secretary John Purton in mid-2018 led to the immediate, temporary appointment of Alin Elena (then a Hartree Centre funded PDRA at Daresbury) as the acting Secretary. He was set to work on the CCP5 Summer School website, which required immediate attention for an upcoming School. Due to this intervention and his wider support of CCP5 activities, Elena was later confirmed as the full secretary of CCP5. Meanwhile the effort required to develop and maintain DL_Software suite as well as train and advise users was considerable and drove a constant need for support staff. The provision of permanent Daresbury staff to CCP5 under the Service Level Agreement at this time was 3.4 FTEs *per annum* (ranging from 2 to 4.5). Essential additional support was provided by fixed term employees on various grants and it was essential that such support be maintained at the necessary level.

In May 2017 Ilian Todorov and Michael Seaton sought and obtained H2020 funding for 2 PDRAs, over 3 years, to work on a software interoperability project, VIMMP, in the area of mesoscale and molecular modelling. The project involved a large number of EU institutions and commercial organisations – industries as well as software houses. On the UK side it included Unilever, IBM UK and the University of Manchester. In March 2017 extra development for DL_POLY_4 was awarded to Daresbury and the University of Bristol by the EPSRC eInfrastructure call of July 2016. This resulted in the appointment of Jim Madge at STFC (from Durham University, November 2017) and Alex Buccheri at Bristol (from Oxford, February 2018) to extend the capability of DL_POLY beyond the classical MD with Forward Flux Sampling and DFTB force-field integration respectively. In 2018 Alin Elena, Ilian Todorov and Kostya Tratchenko were successful in securing a grant with Queen Mary College from the EPSRC Impact fund to develop new analysis tools for DL_POLY_4. A new EPSRC proposal was submitted in July 2018 for furthering the capabilities and special applications of DL_MONTE in partnership with the universities of Bath and Lancaster. Unfortunately, like the 2012 second-stage application for DL_POLY, this proposal was not funded.

One of the aims of CCP5 in the period 2015-2020 was 'Outreach' - an intention to push awareness of CCP5 and its expertise out beyond the confines of the simulation research community. One of the mechanisms by which this was accomplished was the Summer Bursary scheme for collaborative

summer projects, which provided funds for students, with the support of a UK academic supervisor, either to work with industry or a European academic group or alternatively, in liaison with Daresbury, provide teaching material for undergraduates using DL_Software, or to provide outreach material using computer simulation with the software. The awards began with two students in 2015, rising to eight in 2017 and 2018, and five in 2019. Applications for the awards were generally oversubscribed and often involved industrial SMEs. Reports on the awards appeared on the CCP5 website. In a similar vein in 2018, CCP5 devised an Outreach Award, which invited contributions to the website's Outreach pages in the form of an original simulation image and supporting text suitable for pre-university students. A small monetary prize was awarded for the best three contributions.

Outreach also took the form of conferences and workshops. The work of CCP5 was highlighted in the meeting: Simulations for the Experimentalist and the Industrialist, which was held at the Diamond Light Source, Oxford on 15-16 November 2016 and similarly in November 2018. In May 2018 the whole CCP5 team gave a public lecture at Daresbury "The Power of Computational Science – "What matters about matter: Artificial design and virtual experiments". In a different arena, Ilian Todorov was invited at an invited speaker at EMMC-CSA Workshop on Industrial impact of materials modelling in Torino, July 2019. Ivan Scivetti was an invited participant in the National Physical Laboratory workshop, Computational Methods for 2D Materials Simulations, 30th July 2018. Scivetti gave invited talks at Universidad Autonoma de Madrid in May 2019. Ilian Todorov and Michael Seaton were invited speakers at the HPC 2019 conference in Borovets, Bulgaria in September 2019. Alin Elena was invited to the Electronic Structure Library Hackathon in August 2019, a CECAM-ECAM sponsored event. Andrey Brukhno gave a talk at a CECAM workshop in Liverpool (July 2019) and John Purton at the University of Bath Symposium on Energy Materials. Alin Elena was a science judge at the Big Bang Science festival in Liverpool and also a volunteer for two STEM events at the Manchester Science and Industry Museum on materials recycling and at MakeFest. Ilian Todorov gave two presentations at schools in Liverpool and Sandymoor. Arguably the most impressive examples of international outreach by CCP5 were the CCP5/CCPBioSim Molecular Simulation and Software Training Schools – Materials and Biomolecules, which took place at the Universidad del Norte in Barranquilla, Colombia, 25-29 June 2018 and again at the Universidad Catolica del Norte, at Antofagasta in Chile on 13-17 May 2019. This international series of events started in October 2015 with the four day Hartree Centre School on Modelling Applications led by Ilian Todorov at the National Super Computing Centre at Guangzhou (China, then holding the largest HPC in the world Milky Way 2) funded by BBC-China.

CCP5's steady provision of software workshops and training events was maintained and a number of them (~5 *per annum*) took place in the period 2015-2019. These took the form of introductory lectures and hands-on practical sessions using the programs in the DL_Software suite. These workshops were most often hosted at Daresbury Laboratory, but were otherwise held at universities in the UK. Also, by special invitation of the National Centre for Supercomputer Applications in Bulgaria, a training workshop was held at Bulgarian Academy of Sciences in Sofia under the PRACE initiative. Furthermore, during this period, a DL_POLY Developers Meeting was initiated, at which future requirements and developments of DL_POLY could be discussed and planned. Two such meetings took place at Daresbury, in November 2017 and November 2018 and one in QMUL in November 2019.

The Annual General Meetings and workshops remained key activities of CCP5 during Neil Allan's chairmanship. AGMs took place at the universities of Lancaster (2015), Harper Adams (2016), Strathclyde (2017), Manchester (2018) and the London School of Economics (2019). The 2020 AGM is planned for Huddersfield and will mark the 40th anniversary of CCP5. The AGMs were generally attended by both academic and industrial researchers and were often partly sponsored by other

organisations, such as the Royal Society of Chemistry. In addition, CCP5 either organised, jointly organised or gave support to, approximately 40 workshops relevant to molecular simulation in the period 2015-2019, not including the workshops based on DL_Software, nor indeed the Summer Schools – an impressive record. The range of topics covered was wide and included: adsorption at surfaces; industrial applications; mesoscale systems; ferroelectric materials; high energy materials; computational methods; nucleation and growth; radiation damage in solids – and others. Many of these workshops were collaborations with other organisations, including the RSC, CECAM, various universities and the other CCPs – in particular CCP9 and CCP-Bio. Some were regional workshops providing a forum for postgraduate researchers.

One notable conference was the special meeting to celebrate the 25th anniversary of the DL_POLY project was held at Chicheley Hall in Milton Keynes, on 2-3 November 2017. At this meeting the original and current developers, supporters and principal users of the software were present. The common feature of all the presented lectures was the application of DL_POLY software. The event was sponsored by STFC, CCP5, HEC-MCC, EMCS and the journal Molecular Simulation. A special issue of Molecular Simulation based on the meeting was commissioned.

The other notable annual events organised by CCP5 (with additional sponsorship from CECAM) were the Summer Schools. These proceeded with the usual efficiency, and some changes in personnel and content, throughout the period 2015-2019. They were staged at Manchester (2015), Lancaster (2017-2018) and Durham (2019-2020). The courses were always well attended, with an average of 70 students participating. A new feature was introduced in 2018: a programming course on Python and FORTRAN, which was run before the main school for students less familiar with practical programming. Otherwise, the courses followed the well-established pattern of five days of lectures and practicals on basic simulation methods followed by three further days of advanced topics: Biosimulation, Mesoscale methods and First Principles simulation, interspersed with specialist seminars from international experts. No doubt, founder of the School, Mike Allen, when attending in 2019 found the original spirit of the School very much alive.

CCP5 also sponsored several overseas visitors to tour UK universities in this period. These included: Ignacio Pagonabarraga, from the University of Barcelona, in 2016; Mark Tuckerman, from the University of New York, in 2017; Martin Schoen, from Technische Universität Berlin, in 2017; and Fernando Escobedo, from Cornell University, in 2018; Billy Todd, from Swinburne University, in 2018; and Roberto Car, from Princeton University, in 2019.

In recognition of the software creation work of CCP5, in September 2016 Ilian Todorov stood for election at the Research Software Engineering (RSE) Association and become an executive committee member until the end of his term in September 2019. The Association started in 2012 after a move for recognition by software creators in academia at the collaborative workshop of the EPSRC's Software Sustainability Institute in March 2012 in Oxford[21]. During the initial two-year term the association committee worked extensively to professionalise and transform into a Society and generated a new constitution and relevant community codes. These were finalised in March 2019 and in June 2019 the executive committee members became the first trustees of this new Society for Research Software Engineering [https://society-rse.org/]. During his term on the committee Todorov used the opportunity to spread the initiative to the Netherlands giving a talk "The brief

[21] "The Research Software Engineer", Rob Baxter, Neil Chue-Hong, Dirk Gorissen, James Hetherington, Ilian Todorov, Digital Research 2012, Oxford - http://digital-research-2012.oerc.ox.ac.uk/papers/the-research-software-engineer

history of the UK RSE" at the First Meetup of the Netherlands Research Software Engineer Community, eScience Centre in Utrecht, 20 September 2018.

In March 2018 Ilian Todorov was awarded a title Professor at Queen Mary University of London given the long standing collaboration with Kostya Trachenko on radiation damage studies following their original eMinerals fellowship led by Martin Dove.

In collaboration with Imperial College London, CCP5 supported a one month student placement of Masahiro Takenaga at Daresbury Laboratory in February 2018. The Master of Research (MRes) student from Patricia Hunt's group, who was an experimentalist at Cannon Inc (Japan), came to learn the science and art of molecular modelling under the supervision of Ilian Todorov and Vlad Sokhan. Takenaga was awarded a first prize for his talk 'Investigation of viscosity in ionic liquids' at a Molecular Science & Engineering MRes Symposium in September 2018 based on his work at Daresbury Laboratory. Similarly, in March 2019 Daniele Visco arrived from Patricia Hunt's group at Imperial College London for a month student placement at Daresbury Laboratory. Supervised by CCP5 Vlad Sokhan, Tom Keal and Ilian Todorov, the student became versatile practitioner in both QM and MM modelling tools in CCP5's arsenal. He won the 'Best Poster' price for his multiscale study of ionic solvents at the MReS poster show at Imperial College London in September 2019.

The end of Neil Allan's era as CCP5 Chair was signalled in June 2019, when a new election was held. The new Chair was Paola Carbone from the University of Manchester, who took over from Neil Allan 1st January 2020. Carbone's election marked a significant moment in CCP5's History – the first woman elected as Chair. Following a comprehensive survey of CCP5 members and with the help of all CCP5 involved staff at Daresbury Laboratory, Neil Allan and Paola Carbone wrote a renewal application for CCP5 for submission in October 2019. The proposal highlighted the importance of CCP5 in code development and promoting new methodology and its relevance to major research groups in the UK academia and industry. The support and training given by Daresbury staff (under the SLA) for these codes was essential. CCP5's role in disseminating scientific knowledge though its network activities consisting of conferences, workshops, Summer Schools and sponsorship of new collaborations and visiting speakers, was a key part. The aims for the renewed grant were to widen participation and outreach and extend the capabilities of the DL_Software suite to meet new scientific challenges – aims that are consistent with CCP5's historical function. There could be no doubt from this proposal, that CCP5 played an important strategic role in the molecular simulation community of the UK and arguably, in the wider scientific world. Nevertheless, the proposal was rejected and CCP5 was forced to enter its fifth decade with its future in doubt. History has shown that CCP5 has met similar crises before and has enough resourcefulness and support to prevail. However, at the time of writing the matter is not resolved.

4 Hardware, Software and Community Support

4.1 The research computing ecology

In order to do computational science, you need computers. This trite statement becomes rather less trivial if we ask what kind of computers computational scientists need. In basic terms, we can distinguish the following types of computers:

> **Tier 4**: desktop machines for writing code, documentation, dissemination, communications etc.

Tier 3: local machines housed in research groups or departments for the development of large-scale codes, sizeable production runs, fancy visualisations, etc. Recent developments have seen such systems housed in a built-for-purpose University data Centre to optimise power consumption and cooling requirements.

Tier 2: regional machines funded and operated by groups of institutions to give access to higher levels of compute power.

Tier 1: national/international machines (supercomputers) for the most challenging production runs, often (but not always) where the strongest scientific interest lies.

plus the local and wide area networking needed to link together all of the above. While Tier-1 and

Figure 5: The research computing "pyramid"

Tier-2 systems are typically general purpose, there is an increasing trend towards thematic centres focusing on specific application areas.

No doubt the boundaries between these categories are fuzzy (and time-dependent), but the essential point remains clear. In any event, we need a hierarchical computing infrastructure – once drawn as a pyramid, now usually described by the "tier" terminology indicated in the list and figure (taken from EPSRC's 2017 roadmap) above. In fact, what we really need is to work within a vibrant and flexible computing ecology[22]. There has to be a certain balance between the different tiers, a balance whose dynamic is determined by the demographics and scientific progress of the relevant research fields. Perhaps even more importantly, there must be a balance struck between the hardware infrastructure and the needs of software development, usage and maintenance.

In what ways do these levels or tiers of resource differ, apart from their obvious compute speeds? First, the machines generally have different ***architectures***. Over time, we have gone from scalar mainframes, through vector processors, commodity processors, moderately parallel systems (SIMD, MIMD, special processor-based, commodity processor-based etc) to massively parallel systems. Perhaps in the future we will have practical quantum computers. Such radical changes (generally surpassing Moore's Law) raise serious questions for computational scientists. Codes need to be ported between very different architectures. Proliferating versions need to be controlled. Performance tuning becomes important, especially for the high-end systems. And all this work must be repeated frequently if the research community is to take advantage of rapidly evolving architectures.

[22]We use the term "ecology", because it evokes a complex system in which the component parts evolve under their mutual interactions and external forces. That seems to fit research computing quite well.

Second, the ***costs*** of the different tiers are very different. The capital cost of a desktop machine is about $\$10^3$; that of a massively parallel national supercomputer might be around $\$10^8$. To this must be added, at least for the upper tiers, the running costs (electrical power, the data centre, support staff, etc). This means that the process of acquiring different types of machine is completely different; desktop systems are bought from petty cash, tier 1 supercomputers require full-scale international procurement procedures.

Third, very different ***modes of access*** apply to the different tiers. To access your desktop machine, you sit down at your desk. To access your departmental machine, you ask your head of department. To access a regional system, you probably have to write a formal proposal for peer review. Getting time on a national supercomputer typically involves a heavy-weight peer review process, possibly including the formation of research consortia and other sociological aspects (user groups, etc). Formalised reporting of the outcomes of research to the funding agencies is normally a requirement. These differences are clearly driven by the costs of the systems. One might note that for the most powerful machines a key objective is capability computing – what can you calculate on this machine that you can't on any other machine? For the lower tiers, more emphasis is generally given to efficiency of service – are you using the full capacity of the machine effectively?

Scientists doing their research projects do not naturally interest themselves in most of these issues. But they will get interested if it will give them time on the systems they need for their problem. For computational scientists engaged in writing, developing and supporting real large-scale codes across multiple platforms, a further issue becomes important – the distinction between development runs and production runs. In preparing a code to run efficiently (or at all) on some novel machine, it is very helpful, perhaps essential, to have informal access, as it were, to a quite powerful machine of appropriate architecture. A highly bureaucratic system of access is not what code development needs, and, by definition, porting/debugging/testing work generally does not count as efficient usage of the highest end machines.

Within the UK, over the lifetime of the CCPs, research computing resources have changed radically, as computing technologies have changed and as computational science itself has developed in response to new experiments, theories, algorithms and, ultimately, new codes. These changes in the computing infrastructure have mainly taken the form of initiatives conducted by the funding agencies (the Research Councils), often motivated by the perceived to balance and re-balance the various elements of the research computing ecology. For our purposes, this raises some interesting questions:

- How far does a collaboration such as a CCP assist the access of its members to the computing resources they need, and can it influence the procurement of such resources over time?
- How did these issues play out in the specific case of CCP5?

In this section, we outline the story of research computing provision in the UK over the last 40 years, in the light of the questions above. We also take the opportunity to comment on the increasing influence of software engineering techniques in computational science, and on the challenges of collaboration between computational scientists and computer scientists. We believe that the collaborative development and support of large-scale scientific software is, and always has been, a key aim of the CCPs – resulting when successful in true ***community codes***. The technical aspects of producing such software are at least as important as hardware technologies. Note: community codes live much longer than any hardware platform. A good part of our story actually relates to the CCPs as a whole, if not to any form of collaborative computational research. Although we note the

involvement of CCP5 with the many machines and initiatives in our account, we reserve for section 5 a detailed case study of how all this played out in a key CCP5 project: the DL_POLY code.

What follows in this section is essentially a precis of a more detailed technical account of the research computing infrastructure in the UK over the last 40 years. We hope to publish a longer paper on this elsewhere.

4.2 Research computing in the UK

The broadening impact of computational science, spearheaded by the CCPs across successive generations of computing, has been one of the major features of scientific computing in the UK throughout the past four decades. We here provide a necessarily brief overview of successive generation of hardware systems (see Table 2) that became increasingly available to CCP5, and to the CCP community in general. These inevitably start with the first vector computer, the Cray-1, moving through the attached processor era typified by the FPS 164 and 264, to the evolution of parallel processing and high-end computing (HEC) systems – from the iPSC-2, iPSC-860, and the early Beowulf clusters to the later generations of commodity clusters and Peta-scale high-end computing systems.

While this overview is at best limited in scope, we focus specifically on two of the initiatives that shaped a distributed computing landscape for the UK scientific computing community,

- The Computational Science Initiative (CSI) that led in turn to the Distributed Computing (DisCo) Programme at Daresbury.
- The High Performance Computing Initiative (HPCI)

A common thread throughout lies in the mutual interaction of hardware and software systems, with examples from CCP5, along with the critical role played by development systems in ensuring the successful exploitation of the more expensive production platforms.

4.3 Early vector and parallel systems

4.3.1 Vector computing

The first Cray-1 supercomputer in the UK was installed at Daresbury in November 1977. This was on loan from Cray Inc. and was one of the first Cray vector supercomputers outside the USA. Indeed, it was actually Cray serial number 1 (SN1), which had been installed at Los Alamos National Laboratory for a six-month trial in March 1976. Over the next two years, it was upgraded to a Cray-1A and then Cray-1S/500 which finally became a 1S/1000 with 1 MWord (64-bit) memory.

Successive Generations of Science-driven HPC Systems exploited by the CCP Communities.

Vendor	Description	Systems	Dates	Location
Cray	Vector	Cray-1; Cray-1A; Cray-1S/500; Cray-1S/1000	1977-1983	DL
		Cray-1S/1000 + Cray-1B	1983-1989	ULCC
NAS	Scalar	AS/7000	1981-1987	DL
FPS	Scalar + vector	164/MAX	1984-1988	DL
FPS	Scalar + vector	264	1987-1991	DL
CDC	Vector	Cyber 205	1984-?	MCC
Convex	Vector / Parallel	C220 – 2 × 200 MHz CPU	1988-1994	DL
Cray	Vector	X-MP/28	1989-?	ULCC

Cray	Vector	X-MP/416; Cray Y-MP 81/8128	1987-1994	RAL
Meiko	Parallel	M10; 13 × T800 transputers	1989-1993	DL
FPS	Parallel	T20; 16 × T414 transputers with additional vector processing chips	1986-1988	DL
Intel	Parallel	iPSC/2 - 32 SX/VX nodes	1988-1993	DL
Intel	Parallel	iPSC/860 – 64 Intel i860 nodes	1990-1994	DL
Cray	Parallel	T3D; 512 processor DEC "alpha" processors	1994-1997	EPCC
Cray	Parallel	T3E-900; 256 "alpha" 450 MHz processors	1997-1998	EPCC
Cray and SGI	Parallel	T3E-1200E; 816 processors Newton - an SGI Altix 3700 system with 512 Itanium2 processors.	1998-2006	MCC
IBM	Parallel	**HPCX**: 2002 - 1280 x 1.3 GHz POWER4 CPUs; 2007 – 2560 x 1.5 GHz POWER5 CPUs.	2002-2010	DL
Cray	Parallel	**HECToR:** A multi-phase Cray deployment based on AMD cores: **Phase 1;** Cray XT4**,** 11,000 AMD cores **Phase 2a;** upgrade to 22,000 AMD cores **Phase 2b;** Cray XT6, 44,544 AMD cores **Phase 3;** Cray XE6, 90,112 AMD cores, with Cray Gemini interconnect.	2007-2014	EPCC
Cray	Parallel	**Archer:** Cray XC30, 4920 nodes, each node containing two 2.7 GHz, 12-core Ivy Bridge processors and Cray Aries interconnect.	2014 - 2019	EPCC

The Cray was to provide the opportunity for significant advances in those scientific disciplines that were severely constrained by a lack of computer power. We have already noted in section 2 the connection between the arrival of the Cray, the initial CCP concept and the formation of TCSE Department at Daresbury. In fact, the Cray remained at Daresbury for four years, during which time over 400 individual grants selected for their potential for breaking new and worthwhile scientific ground were awarded for Cray use [13]. More complex studies than previously possible were undertaken in protein crystallography, atomic physics and aerodynamics. The enhanced computing capabilities advanced research in astrophysics, nuclear theory and theoretical chemistry. Users of the system were drawn from many Universities and research groups throughout the UK.

The Cray system was demonstrated to be of tremendous benefit in physics and chemistry applications, many from the CCPs, with help from staff at Daresbury tuning codes to run on it. The availability of the Cray in both a production, and perhaps more importantly, a development environment enabled the UK to maintain or establish a lead in areas such as atomic physics, the theory of the electronic structure of molecules, solid state and surface physics [15].

Following this period of concentrated use by a relatively select community, the Cray system was bought outright for the UK and moved to the University of London Computer Centre in May 1983 where it was to be controlled by an Amdahl V8 front-end machine while satisfying a more general computing need for an unrestricted community of users. A full service for the 738 active users continued with an additional second hand Cray-1 (a Cray-1B) in 1986 as recommended in the Forty Report on "Future Facilities for Advanced Research Computing" [16]. The two Crays ran until 1989 when they were replaced by a Cray X-MP/28.

Unfortunately, severe limitations of the links to London made it difficult for the academic community, and TCSE Department staff at Daresbury, to develop work on the Cray after its transfer from the laboratory. Furthermore, production work could only be carried out with significant delays. A full service on the Cyber 205 at the Manchester Regional Centre commenced in May 1984.

Experience suggested that many of the problems in accessing the Cray were evident in trying to use the Cyber 205, with good networking facilities seemingly far into the future.

With the delay in the implementation of the recommendations of the Forty report [16], one community of users, principally those involved in the CCPs, acted to fund an FPS-164 facility to fill at least part of the vacuum. Attached to the AS-7000 at Daresbury, the FPS-164 provided "pseudo" vector processing, upgraded by the addition of MAX Boards for matrix operations. Benchmark tests demonstrated that code vectorised for the Cray-1 ran effectively on the 64-bit processor with the minimum of effort. A proposal for the acquisition was accepted by Science Board, with the machine delivered to Daresbury in October 1984, and subsequent memory (1.5 MWord) and disk (1.8GBytes) led to a final configuration well suited to large-scale scientific computations.

The more general purpose FPS-264 performed at 38 Mop/s but did not have MAX boards. These FPS array processors were in use between 1987 and 1991 and many of the codes in common use by the CCP community were migrated to the systems [17]. The strength of the FPS product lay in the final CPU performance obtained, with an 8-code benchmark revealing factors of between 10-25% of Cray 1S-performance over a wide range of disciplines at a tiny fraction of the cost.

4.3.2 Parallel Computing

In the late 1980's, early 1990's two important developments in computing occurred. At the *high-cost* end of the scale, supercomputers became parallel computers. The fast (specialist) processors and the expensive vector-computers of a few years earlier, largely gave way to systems that combined extremely large numbers of processors with fast inter-processor communications. At the *low-cost* end of the scale, cheap PC processors started to dominate the market. This led to the growth of distributed computing, with clusters of individual PCs linked with slow (but very cheap) communications such as simple ethernet. For both types of computer system, effective parallel simulation techniques were essential if these machines were to be used for the macromolecular simulations developed by CCP5.

It was shown that simple replicated data methods could be used to carry out molecular dynamics effectively, without the need for major changes from the approach used in scalar codes. Domain decomposition methods were then introduced as a path toward reducing inter-processor communication costs further to produce truly scalable simulation algorithms. Indeed, this was the path followed by CCP5 with the evolution of DL_POLY (see section 5), from the replicated data implementation of DL_POLY 2 [19] to the distributed data architecture of DL_POLY 3 and DL_POLY 4 [20], [21].

The transputer, which came on to the market in 1985, was programmed using a special parallel language called Occam. The Meiko M10 (with 13 T800 transputers) was the first explicitly parallel computer at Daresbury, but required applications to be significantly re-written. The FPS T20 (with 16 × T414 transputers and additional vector processing chips) operated under Ultrix on a microVax front end, with access to JANET (the UK's Joint Academic Network) and with the Occam-2 compiler for FPS developed at Daresbury. There is a point well worth making regarding the relatively minor impact of these transputer-based systems within the CCPs – simply put, the need to adhere to standards. Wandering off into new languages or unproven, unsupported, long-term software development environments that offer potential performance improvements has historically provided a minefield of unexpected problems and time-delays. Indeed, such a path introduces a high level of risk that has repeatedly failed to deliver the promised outcomes.

Following this work on early transputer systems, parallel processing at Daresbury as a service focused around the Intel iPSC/2 [22], [23] and iPSC/860 hypercube computers [24] with respectively 32 and 64 processors. Each node of the iPSC/860 consisted of a high performance Intel 64 bit i860 micro-processing chip, memory and an internal network interface to send data to other nodes. The initial configuration of the iPSC/2 comprised 32 SX/VX nodes, provided by the SERC in 1988. A further 32 SX nodes and concurrent I/O system with 1.5 GBytes of disc were funded by a collaboration between TCSE Department and the UK chemical company ICI plc (see section 7.4). The system became expensive to maintain as a service in 1993, although it was still being used in 1994 for system code development.

The Intel iPSC/860 consisted of 64x Intel i860 nodes. Initially 32 nodes were bought with SERC funding and money from the trade-in of ICI's iPSC/2 nodes. The system was installed in June 1990, and upgraded to 64 nodes by the Research Councils in 1993, following the first successful year of a national peer reviewed service. Thus from 1990 to 1993 the Intel iPSC/860 evolved from an initial parallel development system to a resource acting as the focus of a National Supercomputing Service. The iPSC/860 system had a peak performance of around 2.5 Gop/s, equivalent to the Cray Y-MP/8I but at considerably lower capital cost. Some 100 users were registered on the Intel with 140 projects having made use of the system, many of which were surveyed in “Parallel Supercomputing 1992” and “Parallel Supercomputing 1993” [25]. We note that the 1992 edition reports on 21 iPSC/860 projects in Molecular Dynamics – the highest number of any of the fields identified. The associated newsletter, “Parallel News”, was circulated to around 1,000 people, providing up to date information on the service, together with a variety of articles, programming hints etc.

When first installed, the Intel iPSC/860, much like its predecessor the iPSC/2, was very much an experimental machine. Indeed, it was purchased to further investigate the feasibility of parallel computing, but proved to be a much-valued resource by researchers and remained in great demand, despite the introduction of alternative facilities at other UK sites. The rapid development and/or the conversion of a number of scientific applications – not least DL_POLY – to run on the machine proved that parallel computing was a viable technique and was by no means as difficult as some were claiming.

4.4 Local Computing Resources

Along with the radical developments in high performance computing outlined above, it was becoming clear in this period that the research community also need access to local resources, ideally under their own control. The ***Computational Science Initiative (CSI)*** of SERC was intended to meet the needs of the community for local (distributed) high performance computing.

The birth of the Initiative came with the increasing recognition that scientific computation now provided an essential tool for fundamental science and engineering design that complemented the established disciplines of experimental and theoretical investigation. The case for providing local computing facilities was overwhelming. This had been recognised in the Forty Report [16], which recommended the allocation of funds for a National Facility for Advanced Research Computing that would comprise:

1. A central installation of the most powerful supercomputer available,
2. An enhancement of the JANET network to ensure communications for remote users of this facility,

3. A distributed system of other forms of advanced research computing, including special purpose machines and powerful graphics workstations to enhance local resources in selected university and research council sites, and
4. A national organisation of advanced research computing to ensure the effective use of these resources.

Subsequently the Cray X-MP/48 was installed at the Rutherford Appleton Laboratory (RAL) in 1987, with major extensions to JANET. The distributed component of the Forty recommendations, namely a rolling programme of provision of £10M for capital equipment and £6.5M for staff and recurrent costs over the five-year period, 1987-1991 remained, however, unfunded and one could say that the Computational Science Initiative was conceived in the atmosphere of frustration that stemmed from this. The case for the Initiative was prepared by a multidisciplinary panel chaired by Professor David Wallace (Edinburgh). The panel argued that the Forty recommendations were insufficient for science needs. It proposed a 5-year programme, which would require about £21M to provide host and attached processors at some 24 sites, approximately 125 workstations and personnel support. The SERC approved the Initiative in principle, and provided £1M pump priming for the 1986/87 academic session, with a further £1M in the 1987/88 session and £2M in the 1988/ 89 session.

The Initiative had several formal objectives calling for the provision and support of local computing resources in co-ordination with the national supercomputing facilities. Significantly, these objectives included the following:

> "To fully integrate the activities of the CCPs into the Initiative through liaison with the Daresbury Laboratory."

Initially, CSI funding covered only equipment and maintenance costs, but subsequently included staff support (systems or applications programmers). SERC appointed two coordinators (see section 4.2.6) who worked together with successful applicants in the procurement of appropriate hardware. The initial CSI Coordinators were John Inglesfield (Head of TCSE Department) and Martyn Guest (Head of the Advanced Research Computing Group within the Division). Later, David Fincham replaced John Inglesfield as Coordinator. Mike Allen became co-coordinator in 1990-92, replacing Fincham. Both Fincham and, especially, Allen were closely involved in CCP5, the latter serving as Chair (see section 3).

The equipment provided may be broadly classified into the following categories:

- Minisupercomputers, typified by the Convex C210, Alliant FX/40, FPS M64/60. Within the numerically intensive modelling work associated with many of the awards, these machines typically produced 20% single processor Cray X-MP performance.

- Networks of graphics workstations, typified by the products from SUN and Orion. The lower end offerings, such as the SUN-3 were often involved in hosting transputer-based products, or were being used for the pre- and post-analysis of results from central facilities. Several microVAX based networks were also being exploited in similar fashion.

- The higher end workstations, such as the SPARC-based SUN-4 with graphics accelerators, were associated with more compute intensive activities such as computational fluid dynamics.

- Processor array machines of the DAP architecture, which in applications with a significant integer and logical component could exceed X-MP performance.

- High performance chips, in particular the T-800 transputer, as building blocks for flexible parallel arrays such as the Meiko Computing Surface.

- Superworkstations, such as the Ardent Titan and Silicon Graphics machines for those applications where real-time visualisation proved computationally demanding.

What did the Initiative achieve [26]? First, in reacting rapidly to the availability of the new generation of super workstations typified by the Ardent Titan and Silicon Graphics Power Series, the Initiative boosted the use of graphics in scientific computation. Second, it stimulated software development through the provision of local resources. All of the funded groups pointed to the invaluable role that the local facility played in major code development programs that led to new science. Moreover, it enhanced the effectiveness and use of central facilities through the provision of powerful workstations attached to a networked minicomputer eg for file handling, database work etc, acting to reduce overheads at the central facility. Third, it prompted the exploration of novel architectures by means of the Initiative's Parallel Computer User Group. Coordinated by the Advanced Research Computing Group at Daresbury, activities included a series of meetings and workshops dedicated to various aspects of parallelism. This link with the Laboratory further ensured that the activities of the CCPs were successfully integrated into the Initiative. For example, the DAP system at Queen Mary College was used by Dominic Tildesley and David Fincham, working with Nick Quirke, on a project to predict azeotropic behaviour in liquid mixtures [9]. CCP5 contacts and meetings interested other groups in using DAPs for simulations [27].

Members of the expanding CCP5 community benefited from the CSI, through a variety of related projects and activities, many of which were focused in the area of molecular biology and bio-molecular simulations. eg, *Macromolecular Dynamics in Live Cells* (D.M. Shotton, University of Oxford) and *Analytical & Numerical solution of Problems in Plasma Physics and Molecular Biology* (D.N.J. White; University of Glasgow, "*Force Fields and Protein Model Building*"). Many of the funded proposals might best be characterised as traditional compute intensive projects in molecular simulation eg, *Supercomputer for Simulations in Chemistry* (G.R. Luckhurst, D.J. Tildesley and N.G.J. Richards, University of Southampton); *Computer Liquids* (J.G. Powles and W.A.B. Evans, University of Kent); *Computer Simulation of Inhomogeneous Systems* (M.P. Allen, University of Bristol; "*Simple Fluids and wettin*g", "*Liquid Crystals*") and *Molecular Modelling Studies in Chemistry* (J.H.R. Clarke, University of Manchester; "*Computer Simulations of Condensed Phases*"). Algorithmic developments featured in the work by Smith, Fincham & Raine, "*Dynamical systems, Systolic Loops*" (as part of the *Edinburgh Concurrent Supercomputer* from D.J. Wallace, University of Edinburgh), while others focused on the development of techniques in molecular graphics eg *Database-backed Stereo Molecular Graphics and Software for Protein studies* (I. Haneef, University of Leeds, "*Molecular Dynamics of an RNA stem-loop structure*") and Graphics Enhancement for Transputer Array (C.R.A. Catlow, D. Fincham and W. Fuller, University of Keele).

Dominic Tildesley [9] comments: "I am convinced that the work of [the CCPs] led straightforwardly to the development of the SERC's computational science initiative (1994-1998). This allowed for the distribution of high-performance computers (eg Convex, Alliant, parallel clusters) to individual research groups to enable them to run codes developed by the CCP's without recourse to fastest national supercomputers. Although this model of the Departmental or Group computer has now been largely superseded, it played an important role in developing this type of research in the UK."

4.5 Distributed and Commodity-based Computing

The CSI had successfully extended the reach of computational science from the national, centralised supercomputer systems of the early 1980s to the distributed, local level of individual workers, departmental groups and local consortia. However, the unrelenting pace of technology change continued, and with it the need for a framework to support those researchers who were looking to

upgrade and enhance local systems. This led to the introduction of the ***Distributed Computing Programme (DisCo)*** at Daresbury that built on the CSI coordination effort provided by Daresbury staff. This provision ensured the close involvement of the CCPs given their contacts with members of TCSE Department.

Moving from the late 90's into the new millennium saw the choice of supercomputing technology broaden as a function of commodity market competition, technology evolution, historical hardware and software legacies, and leadership choices within industry. Computer vendors, driven by developments such as the US Department of Energy's Accelerated Strategic Computing Initiative (ASCI [28]), aggressively pushed the performance levels of parallel supercomputers higher and higher. It was clear that new architectures capable of PFlop performance could not be built using the same technology in TFlop computers – they would require too much space and consume too much power. As regards the more important technology directions for the CCPs, we highlight the following three developments.

First, **Custom and Commodity Clusters**: The evolution of cluster technologies proceeded on two fronts: in the first development, beginning in the late nineties, IBM, Compaq and SGI - among others - began creating *proprietary clusters* using their shared-memory servers and custom-designed or semi-commodity networks. For example, the IBM approach had been to use their own custom multi-level switch fabrics to interconnect shared-memory nodes based on their Power workstation processors. These nodes had 32 processors, as in the case of HPCx, the tera-scale national supercomputing system housed at Daresbury (see section 4.6). At the same time, true *commodity clusters* were being built and deployed based on uni-processor or dual-processor nodes utilizing Compaq (DEC) Alpha or Intel X-86 processors. These clusters used mainly semi-commodity interconnects from Myricom; but smaller examples were sometimes based on gigabit (or slower) Ethernet switch fabrics. In all cases, the system software was built around the Linux open-source operating system. A number of Terascale commodity clusters had been installed, including Intel x86-based clusters at Los Alamos and Lawrence Livermore National Laboratories. Furthermore, commodity clusters based on open source software (i.e. Beowulf clusters) had demonstrated 2× to 10× more cost effectiveness than clusters based on proprietary solutions. None of these clusters - custom or commodity - had system balance between computation and communications that was competitive with that found on true massively parallel supercomputers such as the Cray T3E. Although they were capable, for many important classes of applications, of achieving high parallel efficiency on a thousand processors or more, they were less suitable for large-scale user services.

Second, **Open Source Solutions:** A notable trend saw open source developments yielding significant enhancements to the state-of-the-art in operating systems (eg Linux OS) and tools (eg Apache). Multiple accretions of open software resulted in profitable enterprises that combined these tools into single offerings with support. In addition, there existed many efforts to build clustering tools that extended the desktop environment to medium and high performance computing. The potential deployment of open source solutions in satisfying requirements of high-end, technical computing was more an open question. What was clear was that the then current trends in developing and implementing ultra-scale computers fell well below the requirements capable of addressing many scientific challenges.

Third, **Beowulf Computers:** In line with the discussion above, the first commodity cluster at Daresbury, then known as a Beowulf (from Norse mythology), was built in 1994 and used as a test and development system until 1998. This was built from off-the-shelf PCs linked together with ethernet - a total of 32 × 450 MHz Pentium III processors. The PCs were connected by dual fast Ethernet switches – 2 × Extreme Summit48 – one network for IP traffic (eg NFS) and the other for MPI message passing. Some of the applications running on these systems and the national Cray T3E

supercomputing facility in Edinburgh, many arising from the work of the CCPs, were showcased at the HPCI Conference in 1998 [29].

We now turn to the "DisCo" project at Daresbury. Support for the distributed computing community involved TCSE Department staff effort in assessing commodity processor based cluster hardware, providing system management support and in the development, porting and optimising of scientific applications codes onto these platforms. Additionally, new technologies with the potential to provide cost-effective alternatives for both capacity and capability high performance computing, such as cellular architectures and field-programmable gate arrays, were assessed. Resources were allocated to assist University researchers with system evaluation, benchmarking, training, documentation, workshops etc., and in the development and distribution of tools for performance evaluation and network-based system and data/file system management.

The DisCo team generated benchmark timings for a number of key CCP Applications, including DL_POLY, on a wide variety of computer systems [30]. These benchmarks were presented at the series of annual Daresbury Machine Evaluation Workshops (1989-2014 [eg [31]), and at the successor Computing Insight (CIUK) conferences from 2015 [32]. A tremendously wide range of processors, file systems and network interconnects were evaluated in this way (see, for example, [31] and [32]). As for the software systems used, most of the commodity clusters used successive generations of Intel compilers along with Intel MPI, although a range of MPI libraries were used – OpenMPI, MPICH, MVAPICH and MVAPICH2. We believe that this effort was among the most comprehensive reviews of distributed computing (indeed, commodity-based) technologies in scientific research world-wide.

The DisCo programme thus provided a unique opportunity to track the performance of a number of CCP applications over the past two decades, an activity that has continued beyond the formal end of the DisCo programme at Daresbury in 2014 [30].

4.6 The High Performance Computing Initiative

Coinciding with the launch of the restructured UK Research Councils in 1994, the new Engineering and Physical Sciences Research Council (EPSRC) became responsible for national HPC provision for the whole UK research community. EPSRC established an Advisory Board for HPC to advise on all aspects of the national service, including the operation of existing services, the future programme, the continuing investment in HPC facilities (including distributed workstations) and the relationship of the national service to other national and international programmes. EPSRC's mission for HPC facilities provision was:

To promote and support world-class basic, strategic and applied research and related postgraduate training by the provision of HPC computing facilities for the Research Councils thereby meeting the needs of their industrial and academic communities and enhancing the UK's industrial competitiveness and quality of life.

In order to promote the national coordination of all activities in the HPC arena, the ***High Performance Computing Initiative (HPCI)*** was established to help computational scientists and engineers of all the UK Research Councils. The primary aims of the Initiative were:

- To achieve and maintain for the UK academic research community a competitive international position in basic and applied computational science and engineering.
- To provide staff to support consortia in porting and developing code for massively parallel processors, including the new Cray T3D at Edinburgh.

- To develop long-term co-ordinated programmes of applications which will effectively exploit current and future generations of parallel machines.

Consortia bids from all research areas were reviewed by the appropriate Research Councils, and an Initiative Panel was established to consider direct staff support for these projects. In addition, EPSRC established three HPCI Centres, at Edinburgh, Southampton and Daresbury, to assist in parallel implementation, optimisation and applications development. The remit of the HPCI Centres was to provide wide-ranging support, specifically for chosen Scientific Consortia; new applications; new and advanced programming environments; matching of systems/methods to applications; access to software; access to novel HPC hardware; contact with industry; publishing information; and education and training.

Initially, 80% of the Daresbury Centre's effort was directed to work of the HPC Consortia. There were twelve consortia formally associated with the Centre, of which five were *directly-supported* and seven *indirectly-supported*. The consortia supported directly by the Daresbury Centre were "Computational Combustion for Engineering applications", "Chemical Reactions and Energy Transfer Processes", "Large Eddy Simulation of Complex Engineering Flows at high Reynold Numbers", "UK Car-Parrinello" and "Macromolecular Modelling". The connection between these consortia and the overall CCP portfolio of the time is clear. The Daresbury Centre also interacted with a further 7 consortia: "External Aerodynamics"; "Sea Shelf Hydrodynamic Modelling"; "Transition Modelling"; "Micromagnetism"; "Atomic Multi-photon Processes"; "Fundamental Electron-collision Processes"; and "Ab initio simulation of Covalent materials". Regular updates on these activities were presented in the "HPC Profile" newsletter [33].

The procurement strategy associated with the HPCI resulted in an overlapping succession of high-end national supercomputing services to be used mainly by the consortia, assisted by the Centres. In early 1994 the Research Councils purchased a large Cray T3D parallel computer at Edinburgh, containing 512 DEC Alpha processors as a national High Performance Computing facility. The original T3D was supplemented in 1997 by a Cray T3E. In 1998, a Cray T3E at Manchester came into service, forming part of the CSAR service (Computer Services for Academic Research) based on "Newton" (an SGI Altix 3700 system with 512 Itanium2 processors) and "Green" (an SGI Origin 3800 system with 512 MIPS processors). Supercomputing returned to Daresbury in 2002 with the arrival of the HPCx system that was operated by the HPCx Consortium (UoE HPCX Ltd.), comprising EPCC at the University of Edinburgh and TCSE Department at Daresbury, with IBM as the system supplier. Installed initially with 1,280 cores of IBM's pSeries 690 Turbo 1.3GHz power4 processor, the system was upgraded in 2004 to 1,600 cores of eServer pSeries 690 (1.7 GHz Power4+ processor), and again in 2005 to 1,536 cores of IBM's eServer pSeries p5 575 1.5 GHz processor. The final upgrade took place in 2005 to yield a system with 2,560 cores. The resulting HPCx service provided 160 nodes for compute jobs for users, giving a peak computational power of 15.3 Tflops.

Successive replacements of the HPCx service have seen both the HECToR (2007-2013) and Archer Service (2013-2020) located at EPCC, driven by multi-Phase systems from Cray. The HECToR system comprised a multi-phase Cray deployment based on AMD cores, with the initial 11,000 core Cray XT4 system installed in 2007 evolving through Phase2a, Phase2b to the final Phase 3 service, a Cray XE6 system with 90,112 AMD cores featuring the Cray Gemini interconnect. The HECToR-replacement system, Archer, commenced service in late 2013 with a Cray XC30 system comprising 3,008 nodes, each node containing two 2.7 GHz, 12-core Ivy Bridge processors, with the Cray Aries interconnect. The system was upgraded to its current and final configuration in November 2014, comprising 4,920 nodes and 118,080 Ivy Bridge cores.

4.7 Community Codes - the Software Challenge

So much for the hardware infrastructure. Now we ask: to what degree has the software used by the CCPs kept pace with the underlying changes in the technology outlined in sections 4.2-4.6 above? As we demonstrated in section 2, a key role of the CCPs has always been to address the joint requirements of performance and portability in the software packages of their associated user communities.

Standards in hardware (eg IEEE 754 [34]) and software have helped to make moving software from one machine to another less painful than it once was, both in terms of it running at all and in terms of its running "correctly". Language [35], [36] and library [37] standards have also served to make life considerably easier for those developing applications. Despite the success of standardization, however, achieving performance portability is still beyond the reach of many, and platform-specific optimizations tend to result in unmaintainable (or very difficult to maintain) code. Despite this movement towards uniformity, CPU architectures have diverged in other ways, and recent newcomers to the scientific computing world have called for the adoption of new approaches and techniques. Those working on software for these systems now have to consider variation in clock speed between individual cores within a chip, as well as differing clock speeds between "identical" chips that are still deemed to be within specification [38].

Before turning to the specific developments associated with CCP5's flagship code, DL_POLY, in the next section, here we take a closer look at just what is involved in the key CCP activity of developing, maintaining and supporting a code, and consider further the fundamental requirements of a successful HPC ***community code***. This is quite a long section. We think that there are several critical aspects of the realities of computational work in scientific research that are seldom articulated, although probably widely felt, particularly in the context of a community code. We will look at six key issues.

First, we consider the development, maintenance and support of a scientific code. It must be recognised that enormous variation exists across actual computational scientists, especially in their attitudes to software engineering; generalization about this "HPC community" is risky. Nevertheless, several common traits characterise scientific software built for High-Performance Computing (HPC) machines.

1. Many developers receive their software training from other scientists. While the scientists have often been writing software for many years, they generally lack *formal* training software engineering, especially in managing multi-person development teams and complex software artefacts.
2. Many of the codes are not originally designed to be large. They start small and then grow based on their scientific success.
3. A substantial fraction of development teams use their own code (or code developed as part of their research group). For these reasons (and many others), development practices in this community are quite different from those in more "traditional" software engineering.

The focus within many of the CCPs has been on codes that require non-trivial communication among the individual processors throughout the execution, codes that have been written to explicitly harness the parallelism of HPC systems. While many parallel programming models exist, the dominant model is MPI, a message-passing library where the programmer explicitly specifies all communication. FORTRAN remains widely used for developing new HPC software, as is C/C++. It is not uncommon for a single system to incorporate multiple programming languages, with many projects using dynamic languages such as Python to couple different modules written in a mix of FORTRAN, C, and C++.

From the launch in 2004 of DARPA's High Productivity Computing Systems (HPCS) project, the emphasis shifted from *execution time* to *time-to-solution*, which incorporates both development and execution time. Clearly, the impact of different parallel programming models became evident with this change in emphasis. The following table shows some of the many attributes that vary across the HPC community.

Attribute	Values	Description
Team Size	*Individual*	One developer, sometimes called the "lone researcher".
	Large	"Community codes", multiple groups, possibly geographically distributed.
Code Life	*Short*	Code that is executed few times (eg a code from the intelligence community) may trade-off less time in development (spending less time on performance and Code Life portability) for more time in execution
	Long	Code that is executed many times (eg a physics simulation) will likely spend more time in development (to increase portability and performance) and amortize that time over many executions
Users	*Internal*	Used only by developers
	External	Used by other groups within the organization (eg at U.S. Government Labs) or sold commercially (eg Gaussian, VASP)
	Both	"Community codes" are used both internally and externally. Version control is more complex in this case because both a development and a release version must be maintained

Second, we look at how important it is for a code to run quickly. Scientific researchers are focused on producing publishable results: writing codes that perform efficiently on HPC systems is a means to an end, not an end to itself. This point may sound obvious, but it is overlooked by many in the HPC provider community. If scientists can execute their computational simulation using the time and resources they are allocated on the HPC system, they see no need or benefit in spending time optimizing the performance. The need for optimization is only seen when the simulation cannot be completed at the desired fidelity with the allocated resources. When optimization is necessary, it is often broad-based, including not only traditional computer science notions of code tuning and algorithm modification, but also re-thinking the underlying mathematical approximations and potentially making fundamental changes to the computation. Thus, technologies that focus only on code tuning are of somewhat limited utility to this community. Computational scientists do not view performance gains in the same way as computer scientists. For example, one of the Daresbury team (trained in computer science) improved the performance of a code by more than a factor of two. He expected this improvement would save computing time. Instead, when he informed the computational scientist, the reaction was that they could now use the saved time to add more function, i.e. to get a higher fidelity approximation of the problem being solved.

Scientists have to balance performance and development effort. They have a marked preference for technologies that allow the scientist to control the performance to the level needed for their science, even by sacrificing abstraction and ease of programming. Hence the extensive use of C and FORTRAN, which offer more predictable performance and less abstraction than higher-level programming languages. Conversely, scientists are not driven entirely by performance. They will not make significant maintainability sacrifices to obtain modest performance improvements. Because the codes have to run on multiple current and future HPC systems, portability is a major concern. Codes need to run efficiently on multiple machines. Application scientists are not interested in performing machine-specific performance tuning because this effort will be lost when ported to the next platform. In addition, source code changes that improve performance typically

make code more difficult to understand, creating a disincentive to do certain kinds of performance improvements.

Third, we can ask: How do we know if a scientific code produces "good" answers? Simulation software commonly produces an approximation to a set of equations that cannot be solved exactly. One can think of this development as a two-step process: translating the problem to an algorithm and translating the algorithm to code. These approximations (mapping problem to algorithm) can be evaluated qualitatively based on possessing desirable properties (eg, stability) and ensuring that various conservation laws hold (eg that energy is conserved). The required precision of the approximation depends on the nature of the phenomenon being simulated. For example, new problems can arise when approximations of different aspects of a system are integrated. Suddenly, an approximation that was perfectly adequate for standalone usage may not be of sufficient quality for the integrated simulation. Identifying and evaluating the quality of an algorithm is a very challenging task. So is the validation of simulation codes. In principle, a code can be validated by comparing the simulation output with the results of a physical experiment. In practice, since simulations are written for domains where experiments are prohibitively expensive or impossible, validation is very difficult. Entire scientific programs, costing hundreds of millions of dollars per year for many years, have been built around experimental validation of large codes.

Fourth, scientists have a somewhat cynical view of new computing technologies. The history of HPC is littered with new technologies that promised to increase scientific productivity but are no longer available. Some of this scepticism is due to the long life of HPC codes; it is not unknown for a code to have a 30-year life cycle. Because of this long life cycle, scientists will only embrace a new technology if they believe it will survive for the long term. This explains the widespread popularity of MPI, despite constant grumbling about its difficulty. A common strategy used by scientific programmers is to develop code in such a way that different technologies can be plugged in to be tested. For example, when MPI was new in the 1990s, many groups were cautious about its long-term prospects and added it to their code alongside existing message passing libraries. As MPI became widely used and trusted, these older libraries were retired. Similar patterns have been observed with solver libraries, I/O libraries, and tracing tools. The new languages developed in the DARPA HPCS projects were intended to extend the frontiers of what is currently possible in today's machines – but just how attractive is the prospect of migrating to these languages for the end practitioner? For those working on very large codes and running on very large machines, the thought of porting to a new language is filled with great trepidation simply because of the time they have already invested in their codes and their need for longevity.

Fifth, there is the inconvenience of having to use shared equipment. Because of the cost, complexity and size of HPC systems, they are typically located at HPC centres and shared among user groups, with batch-scheduling to coordinate executions. Users submit their jobs to a queue with a request for number of processors and maximum execution time. Since these systems are shared resources, scientists are in many instances physically remote from the computers they use. Historically, useful tools that were designed to be interactive become unusably slow and were soon discarded, because they were not designed to take into account the long latency times of remote connections. This situation has improved significantly with improved networking speeds. However, unfortunately for scientists, using an HPC system still typically means interacting with the batch queue. Debugging batch-scheduled jobs is also tedious because the queue wait increases the turnaround time. Some systems provide "interactive" nodes that allow users to run smaller jobs without entering the batch queue. Unfortunately, some defects only manifest themselves when running with large numbers of processors. The problem of the queue may be exacerbated by policies adopted by HPC centres which use *system utilization* as a productivity metric. Since utilization varies inversely with

availability, policies that favour maximizing utilization will have longer waits. There can be little doubt that the availability of local development resources eg the iPSC/2 (see section 4.3.2) really accelerated the development of the DL_POLY parallel codes. Trying to develop parallel codes – particularly at scale – has repeatedly proven a challenge given the fundamental requirement of rapid turnaround of short jobs is not well matched to the typical scheduling policies of production-class supercomputers.

Sixth, we address the problem of effective inter-working between computational scientists and software engineers. It is a brutal fact that software engineering technologies that do not take into account the constraints of research scientists fail and are not adopted. Software engineers collaborating with scientists should understand that the resistance to adoption of unfamiliar technologies is based on real experiences. Here are some examples.

Object-oriented languages: While object-oriented technologies are firmly entrenched in the software engineering community, in the HPC community, C, FORTRAN are still in widespread use. Usage of C++ is on the increase, with Java and Python usage also on the increase. Historically, we see that FORTRAN-like MATLAB has seen widespread adoption among scientists, although not necessarily in the HPC community. Until recently, Object Orientation has not been a good fit for scientific HPC, even though some concepts have been adopted.

Frameworks*:* Frameworks provide a higher-level of abstraction to the programmer, but at a cost of adopting the framework's perspective on how the code should be structured. An example of HPC frameworks is the Common Component Architecture (CCA) for implementing component-based HPC software. Scientists using such frameworks are in the minority – instead the tendency is to implement their own abstraction levels on top of MPI to hide low-level details, and develop their own component architecture to couple their subsystems together. There is the perception that such frameworks force the scientists to adapt their problem to the interface supported by the framework. Scientists may feel that it would take more effort to fit their problem into one of these frameworks than to build their own framework on top of lower-level abstraction such as MPI. One significant barrier to the use of many frameworks is that they cannot be integrated incrementally. As noted earlier, a common risk mitigation strategy is to allow competing technologies to co-exist within a code while under evaluation. However, the nature of many frameworks makes this impossible.

Integrated Development Environments*:* There is still little use made of integrated development environments (IDEs) such as Eclipse because they do not fit well into the typical workflow of a scientist running a code on an HPC system. Historically, IDEs had no facilities for submitting jobs to remote HPC queues, and often lacked debugging and profiling support for parallel machines. Several efforts to provide this functionality [39], [40], [41] in Eclipse were through the Parallel Tools Platform (PTP) project. Furthermore, while HPC languages such as FORTRAN and C/C++ are supported in Eclipse, they proved second-class citizens in the Eclipse ecosystem which is focused on Java-related technologies. It is an open question whether these technologies will be adopted by the larger HPC system community.

On the other hand, scientists *do* embrace some software engineering techniques and concepts, *when they are a good fit*. Almost every multi-developer project uses a version control system such as CVS, Subversion or git to coordinate changes; the CCP community set up a facility of this kind called CCP Forge [42]. The need for this became particularly clear when CCP5 spawned HPCI Consortia. There is also extensive use of regression testing methods, including tests across platforms and compilers, and in reusing externally developed libraries such as preconditioners, solvers, adaptive mesh refinement support, and parallel I/O libraries. On multi-physics applications which involved integration of multiple models maintained by independent groups, considerable effort has

been devoted to software architecture for integrating these components, including use of object-oriented concepts. It is fair to say that the general use of software engineering methods is now quite widespread in computational science and engineering as a whole, and the CCPs are no exception.

Finally, we return to the essential idea and role of community codes, clearly of special relevance to CCPs. Community codes are characterised in the following table:

Attribute	Values	Description
Team size	*Large*	"Community codes", multiple groups, possibly geographically distributed.
Code life	*Long*	Code that is executed many times (eg a physics simulation) will likely spend more time in development (to increase portability and performance) and amortize that time over many executions.
Users	*Internal & External*	Community codes are used both internally (only by developers) and externally (by other groups within the organization or academic community, or sold commercially (eg Gaussian, VASP). Version control is more complex in this case because both a development and a release version must be maintained.

Why are community codes so important?

- Scientists can focus on developing for their algorithmic needs instead of becoming bogged down by the associated infrastructural developments.
- Graduate students do not start developing codes from scratch. Typically students will look at the available public codes and converge on the ones that most meet their needs, considering the effort of customization for their purposes before selecting the code of choice, and building upon it as they need. It is important to remember that they still need to understand the components developed by others that they are using, but don't actually have to develop everything themselves. This is particularly true of detailed infrastructure or solvers that are too well understood to have any research component, but are time consuming to implement correctly.
- Researchers can build upon the work of others and get further faster, instead of reinventing the wheel. Community codes promote code component re-use, thus avoiding the need for the developer to become an expert in every numerical technique.
- Community codes lead to more reliable results because of more stress tested code, with different users using the code in different ways and stressing it in different ways. Enough eyes looking at the code will find any errors faster. In contrast new implementations typically take several years to iron out the bugs and deficiencies.
- Open-source science results in more reproducible results, adding to the credibility of the work undertaken.

There is a strategic aspect of this too. Community codes can help to give a voice to, and increase the impact of, a research community on strategic issues such as funding policy for national or regional computing infrastructure. For example, on the one hand, CCP5 codes are extensively used in performance and acceptance tests for UK high-end computing procurements, as is also the case for other CCPs. On the other hand, leading members of CCP5 have played a very prominent role in developing UK government policy for research computing [54].

The essential message of this section is that research scientists and computer scientists must work together to support computational research, but this is not easy! It is obvious that computer scientists have much to offer but can only be effective if they really understand what research

scientists are trying to get from their computing. DL_POLY, CCP5's main community code, clearly exemplifies the aspects pointed to above. In the next section, we turn to a detailed history of this project.

5 The DL_POLY Project

5.1 Origins

The origin of the DL_POLY molecular dynamics package lay in the developments in parallel computing at Daresbury Laboratory during the mid to late 1980s. In a cause championed by Martyn Guest in the Theory and Computational Science Department, the Laboratory began to invest in parallel computers (see section 4.3), in particular a 32 node Intel iPSC/2, which was later expanded to 64 nodes. To this was later added a 128-node Intel iPSC/860 with Intel 860 numerical processors. These were advanced *Multiple Instruction / Multiple Data* (MIMD) machines, with an efficient, bespoke, hypercube communications network that was accessible through simple Fortran-callable subroutines, the inner workings of which the user was spared knowing. There was also a Meiko M10 Computing Surface, which was based on the UK-devised Transputer, offering parallel processing via the Occam program language. These machines were ideal platforms for developing and testing parallel algorithms and, subsequently, large scale parallel programs. As these machines came online, CCP5 members Bill Smith and David Fincham began to explore their capabilities and devise algorithms. They were soon joined by other CCP5 members, including Andrew Raine (from Keele), Mark Pinches and Dominic Tildesley (from Southampton) and Dennis Rapaport (from Bar-Ilan in Israel), who visited the Laboratory to try out some of his own programs. Some of the parallel techniques that were being developed at this time include: replicated data, systolic loops and domain decomposition (as underpinning parallelisation strategies), methods for parallel electrostatic calculations (principally the Ewald sum) and methods of data management (for stopping and restarting simulations in a non-disruptive manner). In these studies, simple mathematical models of the algorithms were developed to help understand and predict the performance and scaling.

The DL_POLY package itself was conceived by Bill Smith who, in March 1992, decided to strike out in a novel direction. He noticed that there were no general purpose parallel molecular dynamics programs available anywhere and particularly to the CCP5 community. It was thus apparent that the project would derive a considerable advantage from having such a code. He therefore commenced to write one, based on the expertise he had acquired while working on the Daresbury machines with his collaborators. The intention was that the program would be capable of simulating as wide a range of atomistic systems as possible, from monatomic molecules to polymers, ionic materials in solids, solutions and at interfaces, polymers in bulk and solution and biological systems. The program would be easily portable between computing platforms, so that exploratory studies on single processor mainframes and desktop computers could be scaled up to run on parallel computers without a change in software or operating procedure. It would have a comprehensive internal error checking of user input and, as far as possible, shield the user from the issues of parallel processing, such as load balancing and message passing. The potential userbase of the package was estimated to be ~35 groups in the U.K.

The first major decision was the choice of replicated data as parallelisation strategy for the program. Experience had shown that this offered the simplest and quickest route to a viable program, and it was needed quickly if CCP5 was to gain maximum benefit. Nevertheless, where possible, the

program would resort to domain decomposition and even systolic loops if these strategies offered a practical solution.

Unsure if his idea would actually work, Smith initially developed the program in secret, but once he had practical solutions for the more difficult issues, including distributing the molecular topology over the processors and devising a parallel version of the SHAKE algorithm for rigid bonds, he let it be known what he was doing – to the general approval from CCP5. The product was a Fortran 77 program he called DL_POLY 1, which was a practical program already capable of a wide range of applications (indeed, equivalent to the combined capability of a large fraction of the programs in the CCP5 program library), but falling short of the overall vision. However, it provided a good development platform, and in 1992 he approached Dominic Tildesley, the CCP5 chairman, and with his participation a formal CCP5 bid was made to the EPSRC's Science Board Computing Committee (SBCC) for PDRA support to develop DL_POLY further. A positive outcome to this seemed improbable, given CCP5's record of obtaining PDRAs at renewals, but a grant was nevertheless awarded and Tim Forester was appointed in January 1993 as an additional developer for two years. The grant was later extended a further 2 years by the Science and Materials Computing Committee (SMCC) in 1995.

In the beginning, a number of tasks faced Forester and Smith. The existing program required documentation and the methodology had to be published. The program needed to be tested thoroughly and a suite of test cases devised to give potential users a body of examples to work from. These tasks were completed quickly, after an intensive effort, and the fully documented program (with a comprehensive user manual) was publicly released as DL_POLY 1.1 in October 1994. This was already a sophisticated program capable of accommodating many different molecular species in the same simulation and offering a wide selection of interatomic and intramolecular force options. The processing of user input data handled all the internal bookkeeping and force exclusion rules and performed a detailed error check followed by advice to the user for correcting the input. The calculation of pair forces was facilitated by an efficient self-updating Verlet neighbour list that was fully distributed over the processors. In 1994 this version was an early installation on the 256 processor Cray T3D at Edinburgh. Adaptations for the T3D required replacement of the Intel message passing routines with portable Parallel Virtual Machine (PVM) equivalents. (DL_POLY 1 in fact exceeded expectations on the Edinburgh T3D, which allowed efficient memory-to-memory message passing that the program was able to exploit to its advantage.)

5.2 DL_POLY 2 – Evolution

Over the following months, the package was enhanced and extended considerably. The range of options for electrostatic forces was expanded to include the reaction field and neutral group methods favoured by the bio-simulation community. Integration algorithms for different ensembles (NVT, NPT etc) were introduced and a multiple time step algorithm to improve efficiency in large scale simulations. Rigid molecules (requiring rotational dynamics) for small molecule simulations extended the range of molecular models. Three-body forces for silicates such as zeolites and the Sutton-Chen potential for metals catered for those interested in materials. An external (field) force facility opened the possibility of non-equilibrium systems. In addition, tools for generating lattice structures, building and simulating biological systems and for analysing simulation results were added. These improvements were consolidated into the package that became DL_POLY 2, which was managed using a CVS repository. The new package was released as version 2.1 in February 1996.

One of the early difficulties using version 2.1 was that it was hard for the user to estimate the required sizes for the internal data arrays, which had to be specified at compile time. To overcome this a utility program (*parset*) was developed to scan the user input data beforehand and produce a file that could be incorporated into the DL_POLY compilation as a Fortran block data file to define the arrays dimensions. This proved to be extremely valuable to novice users.

Updates of the DL_POLY 2 package were frequent in the years 1996-1997, as new features were added and bugs fixed. Each of these updates introduced new functionality and extended the scientific applicability. Inversion and other four-body forces were added in 1996. There were also algorithmic improvements extending the range of computers on which the code could run. In 1997 the *parset* utility was incorporated into the main program together with Fortran 90 dynamic array allocation, which allowed the required arrays to be allocated on the fly, thus making the compiled program a universal executable. The same year saw the appearance of a Graphical User Interface (GUI) for DL_POLY, based on the Cerius 2^{TM} commercial graphics program, which allowed the user to build test systems, run simulations and analyse output data. By the end of 1998, there had been 535 downloads of the full package.

Management of the package at this time was a full-time job consisting of adding new features, reporting and fixing bugs, releasing new versions, writing documentation and reports, answering user queries and giving lectures and training to users. Underpinning all program releases was a user licence drafted by Daresbury Laboratory, which permitted free use of the code for academic research but retained commercial rights for the Laboratory. The release of the new versions required the users to first complete a registration form, after which they were permitted to download an encrypted, UNIX tar file, for which a decryption code was supplied. This rather elaborate procedure served to prevent unauthorised use of the package by commercial companies, while also enabling the accumulation of user statistics for assessment purposes. This procedure was used until 2006, when it was changed to an online registration, in which the user "clicked-to-accept" the licence terms. The package was provided in source form (in FORTRAN) so the user would be able to add features of their own design before compilation. Such use was entirely within the spirit of a CCP5 program. News of bugs, fixes and upgrades were sent at intervals to all registered users by e-mail.

An early disappointment for the authors of DL_POLY was the lack of uptake by simulators of biological systems. This was undoubtedly due in part to the rather limited range of support tools for analysis and system setup provided – though these did seem to be underestimated, given that a few users were able to undertake quite serious studies with what was available. It also ignored the advantages the package offered which included the fact that it was one of very few programs available that could both exploit parallel computers and also offer, in the Ewald sum, a fundamentally superior model of electrostatic interactions than the truncated or otherwise modified electrostatic models used by bio-simulators at the time. In time, bio-simulation packages adopted both parallelism and methods based on the Ewald sum, but failure to use DL_POLY ahead of these developments was arguably an unnecessary delay.

In January 1997, Tim Forester's contract ended. The duty of maintaining, distributing and developing DL_POLY fell to Bill Smith alone. In 1998, he added the Smoothed Particle Mesh Ewald (SPME) method, which exploited a 3D Fast Fourier Transform (FFT) to accelerate the calculation of electrostatic forces. In 2000 he added the Hautman-Klein Ewald sum method for interfaces in ionic systems and in 2003, the Tersoff potential for modelling covalently bonded carbon systems. A development of particular significance in this period was that in 1998 Smith began to develop

DL_POLY 3, a radically different molecular dynamics program from DL_POLY 2 that would become the principal version of DL_POLY in the future (see below).

The development of the program DL_POLY 3 was made necessary by the rapid advancement of large-scale parallel computers that had taken place over the previous decade. DL_POLY 2 was designed for computers with no more than 100 processors and the replicated data design scaled well on such machines. However, by 1998 computers with thousands of processors were emerging and rival programs like LAMMPS, developed at Sandia National Laboratory, were already employing domain decomposition to meet the challenges these represented. Experience at Daresbury suggested that such a program would be capable of handling systems composed of millions of atoms and run on thousands of processors. So it was time for DL_POLY to develop in the same direction.

Though superficially resembling DL_POLY 2 in terms of its input, output and general user experience, DL_POLY 3 was radically different in its internal structure. It was based on the CCP5 program MDMEGA, which employed a linked-cell version of domain decomposition due collectively to Pinches, Tildesley and Rapaport. However, MDMEGA handled only Lennard-Jones atoms and was relatively simple in construction. DL_POLY 3 on the other hand would handle polyatomic molecules that potentially could straddle several domains. Though there was much that could be adapted from DL_POLY 2, some difficult and novel features were required. Over the period 1998-2000, Smith was able to devise solutions for major issues and incorporate them into a *pre-release* version of DL_POLY 3. This did not boast the full range of features of DL_POLY 2 (for example, the code could not handle rigid bodies), but it was a major advance nevertheless.

At this juncture, Smith sought additional help. This came in the form of the eMinerals Project (2002-2008), which was a NERC-sponsored e-Science project headed by Martin Dove at Cambridge University. Dove was keen to include large-scale molecular simulation as a component of the project and invited Smith to join the team. In January 2003, this brought Ilian Todorov to Daresbury as an assistant developer of DL_POLY 3. Todorov quickly learned the intricacies of parallel programming, domain decomposition and its associated contiguity issues. He then proceeded to test the program thoroughly. A number of bugs were discovered and rectified, but the basic program structure and underlying concepts proved to be sound. A body of test cases was gathered and the package was fully documented and released as the first public version (DL_POLY 3.1) in May 2003. A notable feature of this version was the fully distributed SPME implementation that incorporated a distributed 3D FFT developed by Ian Bush at Daresbury [49] [50], which exhibited impressive scaling on large-scale parallel computers, such as the 2560 processor HPCx at Daresbury (see Table 2). Thereafter, DL_POLY 3 became CCP5's "Flagship" molecular dynamics package.

As DL_POLY 3 was being developed, its predecessor DL_POLY 2 continued to be a popular program in the CCP5 community. Bill Smith therefore continued to manage it and add new features to it. In 2001 a new GUI was written by Smith that cast the features of the older Cerius 2™ based GUI into the Java language. Though the graphics of this were decidedly inferior to the commercial offering, the Java GUI offered advantages in cost (it was free) and portability (it was able to run on any platform). It therefore became an invaluable tool for training workshops. In 2002, the entire suite of integration algorithms was recast into the velocity Verlet form, which made it easier to add new *symplectic* methods in the future. The DL_POLY 2 source code was restructured using Fortran 90 modules, commencing in 2007, in a change intended to facilitate future software extensions. This change paid off in some later collaborations. A collaboration with John Harding at the University of Sheffield, and his postgraduate student Duncan Harris, led in 2008 to the incorporation of a module for long timescale dynamics in the solid state, which included Temperature Accelerated Dynamics (TAD) and Hyperdynamics. A collaboration with Ross Brown, Patrice Bordat and postgraduate

student Pierre-André Casade at the University of Pau in 2009 added a module for handling issues of solvation, in particular, energy decomposition, free energy differences and solvent effects in spectroscopy.

The 2009 release of DL_POLY 2.20 marked the last release of DL_POLY 2. In preparation for the retirement of Bill Smith in 2011, when support of the package would not be maintained, the package was re-launched under a Berkeley Software Distribution (BSD) licence in December 2010 as DL_POLY Classic 1.1 [19]. The open licence permitted modification and re-distribution of the package by third parties but was warranty free and retained the copyright of the developers. Thereafter the package was free to anyone to develop further. A second release of the package occurred early in 2011 and included a Metadynamics module for accelerated dynamics and free energy calculations, developed by David Quigley and Mark Rodger at the University of Warwick. Apart from bug fixes and minor changes the package was not much further enhanced until March 2017 when Smith (now in retirement) added a Path Integral Molecular Dynamics (PIMD) module for semi-classical quantum simulations.

5.3 DL_POLY 3 – Shift in size and demand

For the life of the eMinerals project Todorov functioned as the chief developer and custodian of the DL_POLY 3 package, which underwent many modifications and established itself as the CCP5's leading molecular dynamics program [20], [21]. The eMinerals project started in 2002 and demanded great *flexibility, portability and usability* from the package. It had to be able to handle multimillion atoms simulations routinely with scalable performance on leading large-scale computers and at the same time satisfy the consortium's ambitions for wide ranging atomistic studies. The code had to compile seamlessly across a variety of operating systems, such as Windows, OSX and Unix/Linux, in both parallel and serial mode while retaining explicit MPI, PVM and Intel proprietary inter-processor communication libraries in order to utilise the emerging Condor pool systems in a high throughput manner. The program had to cross-compile with consortium libraries for CML/XML annotation and be able to read and write CML annotated files for inter-operability with other CML enabled codes such as SIESTA and GULP. The CML annotation was also to serve enhanced usability purposes such as producing structured output, easily convertible to DHTML, in which data are automatically depicted as graphs and tables with explicit information about the methods, scales, physical units and functionality used in the simulation.

In early 2003 Todorov began to address the consortium's requirements by incremental development of DL_POLY 3 while simultaneously porting essential features from DL_POLY 2 to help the larger community of MD practitioners. The first and most prominent feature added for the eMinerals project was the variable time step algorithm, which dynamically adjusted the time-step size according to a system's dynamics and which was needed for the popular radiation damage cascade studies of glass and ceramic matrices containing many hundreds of thousands of atoms. Further features to help with these non-equilibrium events included particle density variation and atom displacement counters to account for the highly inhomogeneous particle distributions due to the amorphisation processes in glasses.

By the end of 2003 it was clear that despite the great advancements in DL_POLY 3 through which the materials modelling communities had already greatly benefited [56], further progress would require significant restructuring of the code to ensure its sustainability. The way ahead was an investment in modernisation, which would include significant changes to code's file and memory infrastructures. This meant a full rewrite of all algorithms, not only with HPC efficiency in mind but also a higher

degree of abstraction of methods and routines, with better coupling together as part of one package. In the following year Todorov wrote from scratch a new modularised code, with a syntax strictly conforming to Fortran 90 standards. Major design changes were made. These included: new routine naming conventions and allocating of algorithms to elements of the molecular system (eg configuration) and the dynamics methods (eg ensembles); explicit memory allocation and deallocation for program features (eg minimisation) or elements (eg interaction) in order to trim memory usage and leakage; abstraction and restriction of the inter-processor communication to MPI as well as removing all pre-processing features and declaring explicit routine dependencies in makefiles. Following the best software engineering practices emerging at the time, a strict style of coding was followed with strict quality and standards control, helped by using NAGWare Fortran95 compiler and FORCHECK standards tester. It was intended that the new code would facilitate the adoption of the program by new users and assist the transition in between DL_POLY 2 and DL_POLY 3.

The potential of the new DL_POLY 3 code was soon revealed in December 2004, when during testing of the new HPCx upgrade, Todorov and Ian Bush were able to run successfully a 300 million atom system of NaCl on 1024 physical cores. The model system took an hour to load up with an MD time step taking approximately a minute with 80% this time spent on the evaluation of the long-range electrostatic interactions. This test revealed that the software engineering efforts were well spent and paying off greatly in terms of HPC performance and scalability. However, it also showed that the major bottlenecks were the FFT communications, needed for the long-ranged part of the electrostatics evaluation, and the I/O handling – which required an hour to read from disk and over 10 hours to write to disk.

In April 2005 the re-engineered DL_POLY 3 made a major impact. It represented a successful collaboration between Todorov, Bush and Smith, and included many classic DL_POLY 2 features and highly developed and verified electrostatics routines [50]. The release embodied a highly-structured synthesis of original DL_POLY concepts with a much improved HPC performance and a much leaner memory demand. This release also coincided with the automation of the project's user registration via a web-form and an on-line distribution via an FTP server, as well as a UBBthreads community forum facility (due to Richard Wain) as the new collaboration platform. Last, but not least, the project's scalable performance to multimillion atom molecular systems on HPCx (with assistance from Richard Blake) together with research results from radiation damage studies in ceramics were highlighted internationally by Todorov at the ECI Conference Nonstoichiometric Compounds at Kauaii, USA (3-8 April 2005).

These events paved the way for DL_POLY to emerge as a world-leading project and brought an unprecedented increase in the userbase with more than 500 new users joining in the next few years. By the end of 2005 another major release added two new features - a conjugate gradient method for solving electrostatic polarisation via the classical Drude model, and a method for unscrambling the order of atoms in output files (resulting from the mechanics of the domain decomposition algorithm) for visualisation purposes. The new release also brought further performance improvements and new algorithms such as Brownian dynamics ensembles, defects detection and pseudo boundary thermostats for radiation damage studies.

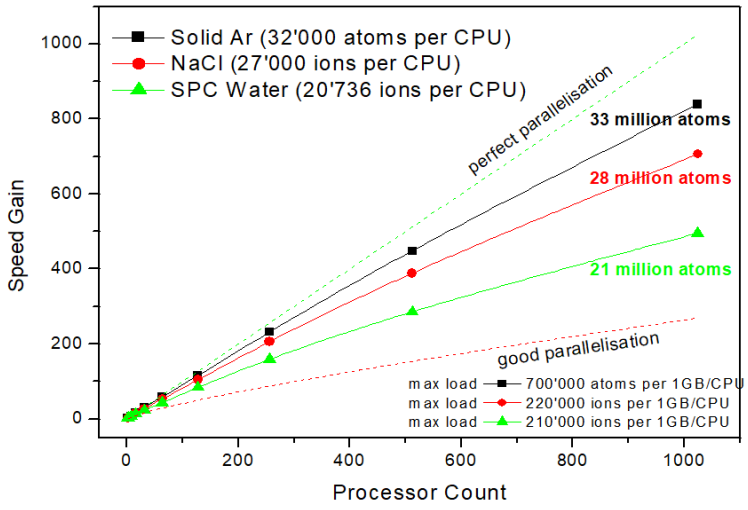


Figure 6: Examples of DL_POLY scaling

The tempo of development accelerated further in 2006 bringing significant improvements in performance with a multitude of new features - new metal and Tersoff potentials, volumetric model expansion, sub-celling of the Verlet neighbour-list and the Martina/Tuckerman and Klein ensembles as improvements of the of Nose-Hoover method. During this year two target requirements of the eMinerals project were also delivered: (i) the independence from MPI libraries, enabling the code to run as a serial application on any operating system with a Fortran90 compatible compiler and (ii) the CML enabling of the I/O, providing both browsable output and also interoperability between other CML enabled codes, such as GULP and SIESTA, so that output produced by one code could be read as input by another. At this juncture DL_POLY users could run both versions of DL_POLY on the same model systems to compare and contrast parallel performance and scalability and also verify results between the two when run in a serial mode. The CML feature and the standardized Fortran of DL_POLY 3 coupled nicely with the GRID computing technologies [59] of the time and led to exemplary research studies utilising university Condor pools composed of Windows and Apple workstations [60].

In 2007 Todorov formally left Cambridge University and joined the staff of CCLRC Daresbury Laboratory in the Advanced Research Computing Group, where he could work more closely with Smith and Bush, to advance DL_POLY HPC capabilities. He remained focused on the eMinerals community especially in the area of large-scale molecular dynamics simulation of radiation damage, which was growing in popularity and complexity [61]. By this time, the access of DL_POLY users to hundreds of CPU cores was becoming commonplace. However, the demand for larger simulations was made especially challenging by the advancements in HPC hardware which saw increases of CPU core density, increased clock rate, higher interconnect speed bandwidth and reduced latency. The two slowest components of the HPC hardware technologies, I/O and communication, were also the two biggest bottlenecks in MD software. In particular, severe demands on access to and from disk were problematic.

Up to the end of 2007 all access to disk in DL_POLY was done in a master-slave manner whereby one nominated processor read and wrote the disk files while all the remaining processors received from or sent data to the master. Between 2008 and 2010 Todorov, Porter and Bush explored various alternatives for addressing this bottleneck [62] and, additionally, the issue of scrambling the printing order of atoms in domain decomposition due to their diffusion across space and domains.

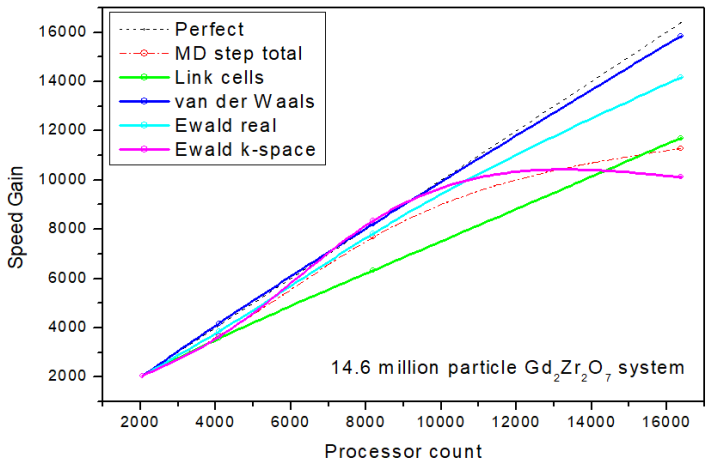


Figure 7: Scaling of algorithms in DL_POLY

In 2008 Todorov was elected to the CCP5 Executive committee while also actively participating in the High-End Computing Consortium for Materials Chemistry (MCC, led by Richard Catlow at UCL) as part of his eMinerals project function. These communities were to become the driver of the DL_POLY project in the future. They had demands for new methodological advancements in both HPC and materials modelling to meet the growing use of computer simulation inn postgraduate research in the areas of condensed matter. Todorov's involvement in these communities presented an opportunity to promote DL_POLY and also establish regular training workshop events in the UK (two to three times a year). Opportunities abroad were also enabled by ongoing collaborations such as those with Ross Brown and Patrice Bordat at the University of Pau, funded by the Franco-British Council. These collaborations also furthered the demand for new developments in both DL_POLY projects. This was met by Todorov and Smith with a string of proposals to EPSRC's managed distributed Computational Science and Engineering (dCSE) calls as part of the HECToR national HPC service between 2007 and 2012. They were able to obtain 5 years of additional development time from 2008 to fund the testing of the HPC methodology and collaborative development efforts by Ian Bush (then at NAG Ltd) and Laurence Ellison at Daresbury Laboratory. The work funded by these sources was reported at the CUG2010 Meeting [65]. The same funding stream enabled the prototyping, testing and implementation of ideas such as multi-core optimisation, rigid-body dynamics, openMP sub-parallelisation and the relaxation of the Ewald summation restriction to 2^N processor counts, among others.

5.4 DL_POLY 4 – New beginning

Between 2008 and 2010 new features in DL_POLY 3 continued incrementally with the addition of new bonded and many body potentials, new ensemble variations (NVT Andersen, NPT and NsT Langevin) and extensions for soft-matter systems with interfaces (NPnAT and NPnγT). However, Todorov's major effort was in the introduction of rigid body dynamics into DL_POLY 3. This was a mammoth task, as it required the domain decomposition framework to deal with objects that could simultaneously exist on more than one domain. It also doubled the code size through the addition of new dynamics routines for molecular rotation and code structures to handle rigid body objects [65]. During this period, another big development was driven by the emergence of HPC platforms hosting multi-core technologies and the appearance of cheap compute accelerators (made popular by PC gaming). NVIDIA, the company dominating the gaming market, began promoting their GPGPU

cards for evaluation purposes in academic institutions. These were supplemented with their brand new CUDA model toolkit that was available free to research software developers. In 2009 Todorov successfully collaborated with Christos Kartsaklis at the Irish Centre for High-End Computing (ICHEC) Dublin, in exploring the pros and cons of the GPU as a co-processor accelerator. This aimed at a GPU port of DL_POLY 3 using CUDA and openMP on top of the original code. This task was major and very disruptive to the DL_POLY 3 code and infrastructure. By the end of 2009 the developmental effort had produced substantial structural changes to the code which demanded a better management approach and a more modern repository for future collaborative developments. The decision was made to move development to CCPForge (see section 4.7) and then advance the code to a major new version. In 2010 all work after the DL_POLY 3.10 release was relaunched as a new code: DL_POLY 4. To the general user, this appeared like the old DL_POLY 3 but was a very different code internally. It arrived in September 2010 packed with new rigid body dynamics, parallel input and output, a relaxation of the restriction to 2^N processors for charged systems, and a new CUDA port.

In 2010 Todorov moved from the Advanced Research Computing Group into the Computational Chemistry Group and soon became the sole custodian of the DL_POLY project after Bill Smith retired in 2011. His role in the group was jointly funded by CCP5 and MCC both of which projects had great need of his training programmes and his collaborative research support. He continued to advance the DL_POLY project by exploring different modes of funding such as EPSRC grants, dCSE work, PRACE projects, EC R&I FP7 and H2020 grants, while keeping engaged with the community networks such as CCP5, MCC, mdNET and EMMC. The mdNET network, which served collaborating scientists across different disciplines with interest in molecular dynamics algorithms, led to Todorov's exposure to the work of Benedict Leimkuhler from Edinburgh University and to the incorporation of a truly ergodic thermostat (the Gentle Stochastic thermostat) into DL_POLY_4.

In late 2010, Todorov and Kostya Trachenko from Queen Mary University of London successfully applied to the very first EPSRC "Software for the Future" call. This funded Michael Seaton between 2011 and 2013 at STFC to prototype a parallel implementation of the Two Temperature Model (TTM) within the DL_POLY 4 code. The development of the TTM prototype was quick and enable high-energy cascade simulations in metals on a grand scale [66] with cells containing hundreds of millions of atoms run on tens of hundreds of processing units on the UK's HPC service, HECToR. Unfortunately, despite this great success and the productive collaboration on this project, EPSRC decided not to fund the project into a second stage. It took over four years before this prototype matured into to a full release ready development in 2018 with support from the Computing Consortium for Materials Chemistry, Manchester Computer Centre.

In 2012 general interest in metal alloys led to the introduction of more advanced many-body interactions (extended Embedded Atom Models) into the project, in collaboration with Ruslan Davidchak at Leicester University. Seaton, who was also working on the DL_POLY sister code, DL_MESO, a Dissipative Particle Dynamics (DPD) and Lattice Boltzmann (LB) program, for CCP5, also brought new ideas and developments into the project. In 2013 these resulted in DL_POLY 3 acquiring velocity auto-correlation facility, new symplectic DPD thermostats, a three-seeded random number generator to remove processor count driven stochasticity for ensembles, and functionality with stochastic components [67].

In 2011 Todorov joined a European Union FP7 call in a proposal to demonstrate the power of "Multiscale Materials Modelling on HPC". This enabled Laurence Ellison to re-join the DL_POLY team in 2012 and demonstrate how DL_POLY_4 can be encapsulated in a GridBean of the Unicore framework and used seamlessly in Unicore workflows in both academic (Panepistemio Patron,

Greece) and industrial (SONY Deucheland Gmbh) environments. This work culminated in 2013 with DL_POLY training and Unicore GridBean demonstration workshops to both industrial and academic modellers at Bologna (Italy) and Jülich FZ (Germany) [68]. This grant also led to establishing a Karlsruhe Institute of Technology spin-out, Nanomatch GmbH [nanomatch.com], which preserved some of the original ideas of the project and is still a successful software as a service consultancy.

In 2013 EPSRC and the US National Science Foundation issued a joint call for collaboration on materials science algorithms and frameworks. Todorov responded with an intention to revive the earlier work of Smith in extending the Ewald summation to systems containing higher order electrostatic poles beyond simple charges, and of Leslie in the practical implementation of multipoles in DL_MULTI [69] (a DL_POLY 2 spin-off, which was able to handle molecular systems with electrostatic interactions up to hexa-decimal order). The proposal aimed at a cross-dissemination of technology between two trans-Atlantic projects - the UK's DL_POLY with HPC capability based on domain decomposition technology and the USA's TINKER, in the forefront of multipolar MD, but still employing the inferior replicated data parallelisation. For the DL_POLY project this was an excellent opportunity to recast the multipole technology in domain decomposition form. The project was joined by Henry Boateng, an expert in multipolar methods, in February 2014. Only a year in, Boateng and Todorov were able to expand and generalise the Ewald Summation to an arbitrary order of permanent multipolar interactions [70]. The methodology was also successfully implemented in DL_POLY in a general manner up to hexa-decimal order (and optimised for dipoles and quadrupoles) reusing and extending the DL_POLY Smoothed Particle Mesh Ewald routines. Further functionality such as 14-7 van der Waals interactions and coupled Drude polarisation for permanent dipoles was also introduced to bring DL_POLY and TINKER closer in capability.

Between 2009 and 2013 the DL_POLY project achieved its largest user adoption as indicated by the statistics in section [5.6]. This led to numerous invitations to Smith and Todorov for training workshops and schools worldwide. Smith led the first international workshop in Capetown, South Africa, in 2009 and then the next one in Zagreb, Croatia in 2010. Todorov followed this up with workshops in Shanghai, China and in Taipei, Taiwan in 2010. In 2012, the DL_POLY workshop made an extended appearance as a three-day DL_Software training session at the PRACE autumn school in Sofia, Bulgaria. The training included DL_FIELD and DL_MESO day components (led by Michael Seaton and Chin Yong). In October 2013 a two-day DL_Software training was given by Seaton and Todorov on the back of a conference on HPC Algorithms in St. Petersburg, Russia.

In 2014 Todorov, in collaboration with Elad Tadmore and James Elliott at University of Minnesota (USA), leading the openKIM project [73], enabled DL_POLY to use DOI published model interactions in the KIM database. At the same time Andrey Brukhno joined the project to work on coarse grained methodology on an EPSRC e-Infrastructure funding to CCP5 (headed by then CCP5 Chair, Steve Parker). The work by Brukhno and Todorov extended the ability of DL_POLY beyond pair distribution functions and included a consistent approach for probability distribution calculations and analytics to all intra-molecular interactions in a model. Numerous other features and algorithmic changes also entered the project - generalised Tersoff interactions, many-body potentials for metal oxides, etc, putting the code framework under stress again. The collaborative effort with Dublin on the CUDA port had already finished by 2012 and was no longer fitting to the current state of the program. While the port had to be discontinued in 2014, the collaboration, then led by Alin Elena at Dublin (2013-2015), did not come to a full stop and indeed brought further opportunities. These included the 2014 openMP testing as part of Intel Parallel Computing Centre at Dublin and also a two-day DL_Software training workshop at TNI in Cork, which also included DL_FIELD and DL_MESO components.

5.5 Current and Future Directions

In 2015 CCP5 was funded generously and aimed at delivering on all components of the DL_Software project. The DL_POLY project in particular was expanding in two very different directions. Inclusion of coarse-grained mesoscale dynamics via Gay-Berne potentials and of reactivity via the Empirical Valence Bond model (EVB) of Arieh Warshel. Todorov was faced with recruitment difficulties due to Henry Boateng and Laurence Ellison leaving for new appointments elsewhere. Meanwhile Daresbury Laboratory's Hartree Centre was growing fast and was soon to become a department on its own. The first opportunity to satisfy both the DL_POLY and the Hartree Centre strategic partners' projects was the EPSRC funded "Energy Materials: Computational Solutions" (EMCS) consortium [71], led by Saiful Islam and Steve Parker at the University of Bath. This project was the vehicle that enabled Alin Elena to join the Daresbury team in July 2015 as part of the Hartree Centre contribution to EMCS. While Elena attended to the diverse HPC and software needs of the consortium, he quickly provided for the further integration and portability of DL_POLY by integrating PLUMED [72] functionality in the project and employing an advanced Cmake strategy for the various compilation scenarios. Having a head-start from previous work on the DL_POLY project, Elena established a range of computer science practices to the future benefit of the project. These were embodied in a framework of Continuous Integration and Testing (CIT), which was later incorporated as part of a GitLab server to succeed the ccpForge repository.

In October 2015, the Hartree Centre chose DL_Software for a Hartree Centre international school led by Todorov abroad. Funded by the British Broadcasting Corporation, the school was hosted by the Guangzhou National Supercomputer Centre in China, which held Tianhe-2 (Milkyway-2) - the largest and most powerful supercomputer in the world between 2013 and 2016. More than 200 Chinese researchers applied for the 50 places available on this school. The DL_Software international activities were resumed in June 2018 with the first CCP5 school in South America at Universidad del Norte in Barranquilla, Colombia, led by then CCP5 Chair, Neil Allan. This was followed in November 2018 by a PRACE Winter School, led by Todorov, at the Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, in Sofia, Bulgaria. In May 2019, the CCP5 Chair, Neil Allan, led a second South American school at Universidad Católica del Norte, Antofagasta, Chile.

In 2015 and 2016 CCP5, and the DL_POLY project in particular, were under great pressure to recruit. This was driven by the departure of Laurence Ellison in early 2016 and the redeployment of Andrey Brukhno on to the DL_MONTE project (the EPSRC funded CCP5 flagship, led by Steve Parker, Nigel Wilding and John Purton) in 2015. Furthermore, the ambitions of Neil Allan, Alin Elena and Ilian Todorov to extend the project beyond classical MD were also successfully funded by a 2016 e-Infrastructure EPSRC call. Hence, the project team rapidly expanded with four more people joining in 2017. Three were based in the STFC – Vlad Sokhan to work on incorporating Gay-Berne dynamics, Ivan Scivetti to implement Empirical Valence Bond methodology, and Jim Madge to work on Forward Flux Sampling techniques. Finally, Alex Buccheri started at University of Bristol to enable DFTB driven MD.

In 2017 Ilian Todorov organised a 25th anniversary celebration of the DL_POLY project at Royal Society's Chicheley Hall where all project contributors and long-term collaborators met. In 2018, due to the large size of the project and the diversity of new development directions, it was decided that the code needed major restructuring. Alin Elena was appointed to lead this effort, assisted by Ivan Scivetti and Jim Madge. Todorov continued with the support and development of the existing DL_POLY 4, pushing to one large release in September 2018 with many new features and improvements contributed by Elena, Seaton, Todorov and other collaborators. Andrey Brukhno who

worked on DL_MONTE joined the DL_POLY efforts returning back to his previous systematic coarse-graining (SCG) work to implement mixed inter-scale Hamiltonian dynamics where the atomistic and CG system can drive each other.

5.6 Growth, Impact and Industry

There are many metrics to describe the evolution of the DL_POLY project over the last three decades. The most useful perhaps, is the growth of the DL_POLY userbase, presented in Figure 8, which shows the growth of the registered user email list per annum.

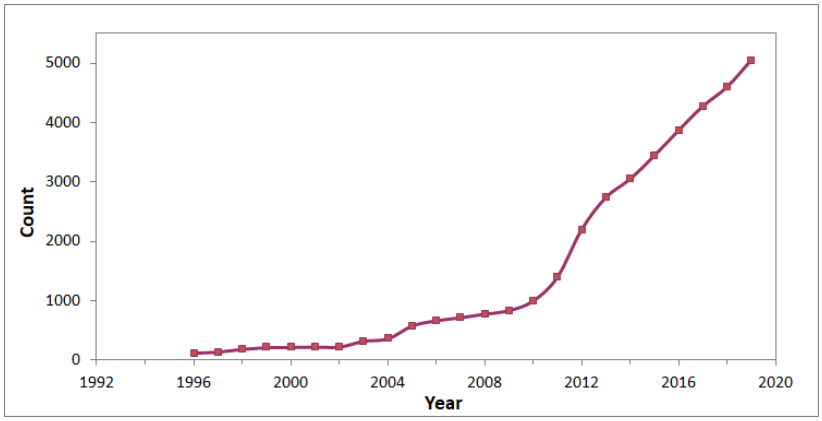


Figure 8: DL_POLY Userbase Growth per annum

The growth in the userbase from a modest 30 in 1996 to 5000 in 2018 tells the main story, but there are additional features along the way. The rise in uptake in 2005 is partly due to a change in user registration and code distribution procedures that happened in that year. Paper licensing and source code delivery by post(!) changed to a web-registration and FTP downloads. The acceleration in project uptake from 2009 was assisted by three factors: (i) the addition of new features as DL_POLY 3 was extended; (ii) the increase of training and workshop events by a factor 3 to 4; (iii) the expansion of the training to cover extra components such as DL_MESO, DL_FIELD and ChemShell; and (iv) the introduction of training events hosted outside the UK.

Another measure of the project's value is acknowledgement by the user community when the project is cited in published work. This is an evolving picture revealed by the Google Scholar returns as presented in Figure 9.

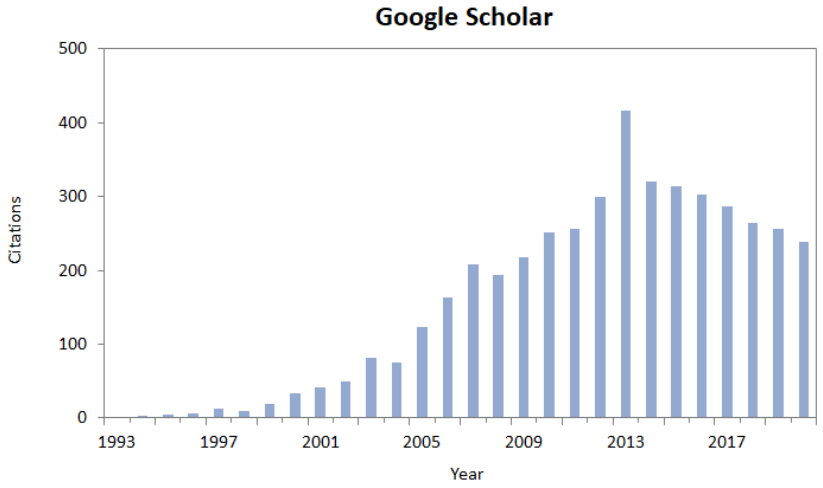


Figure 9: DL_POLY citations as measured by Google Scholar

Figure 9 shows clearly that DL_POLY saw impressive growth in user application from the earliest years to 2013, when the number of citations peaked at 400. The apparent decline that follows requires some explanation (though without detailed survey this is likely to be speculative). One possible factor is the growth of competitive packages such as NAMD [43], AMBER [44], CHARMM [45], GROMACS [46] and LAMMPS [47]. These packages are all widely used and supported, with NAMD and LAMMPS being the most similar in capability to DL_POLY. Since NAMD and LAMMPS are packages produced in the USA the above trend may simply reflect a choice based on national preference. Another possibility is that the scope of molecular simulation has widened to the extent that no single package, however broad in scope, can hope to meet every new need. There is also a price to be paid for extended capability in the form of complexity and diminished ease of use. Simpler and more specialised codes are likely to be leaner and faster and so more appealing to cutting edge researchers. However, it seems likely that, provided DL_POLY continues to be developed and supported by UK based computational communities, the apparent decline in its use will be arrested. Its target userbase is, after all, the CCP5 community and the code is tailored to fit the project's needs.

Another useful metric for the adoption of DL_POLY is presented in Figure 10. It shows the change in the relative demographics of the userbase over time.

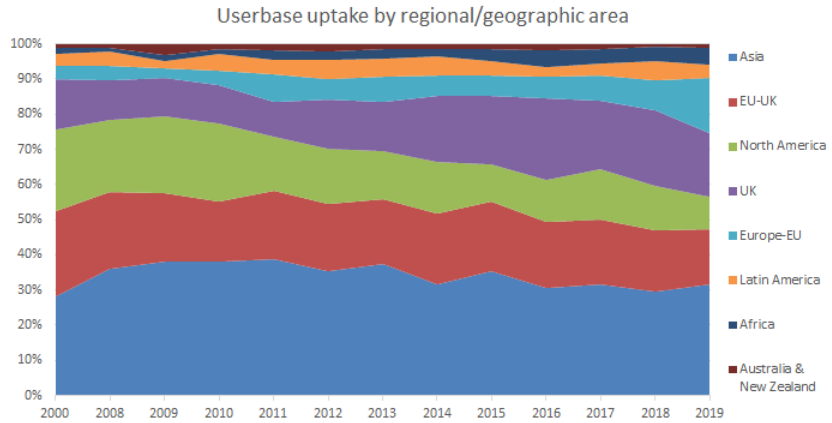


Figure 10: DL_POLY users by region

Figure 10 employs the following notation: **EU-UK** - the European Economic Area based users, excluding those in the UK; **Europe-EU** – users based at the remainder of the European states that are not part of European Economic Area (Russia, Turkey, Albania, etc.); **Latin America** – South America, Mexico and Panama, etc. It is worth noting that the data presented for year 2000 is the average of the project since 1994 and 2007. Demographically, the largest number of new users adopting the project comes from Asia, although that number has always been smaller than that of North America and Europe combined. Interestingly, while the relative share of adoption from North America has stayed the same, the shares of the UK and those of the rest of the EU states seems to balance each other with the former steadily increasing and the latter decreasing over time.

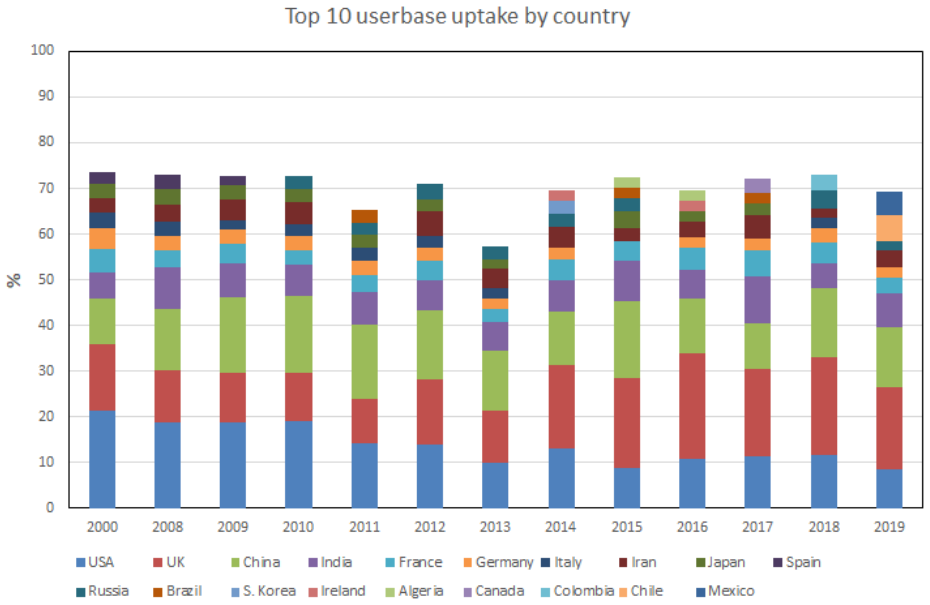


Figure 11: DL_POLY users by country

Drilling down this demographic information by country, Figure 11, reveals that only 5 of the top 10 have persisted over the last two decades – USA, UK, China, India and France, with a second tier of states failing to make it in top 10 just a few times – Germany, Italy, Iran, Japan and Russia.

The analysis of the userbase uptake by scientific domain is also quite striking, Figure 12. The aggregated information in year 2000 (from 1994 to 2007), before the sustained userbase reached the size of 1000, only acknowledges an equal split between chemistry and physics with small contributions from the mechanics and software domains. Post 2010, it becomes obvious that the relative share of chemistry persists, while that of physics seems to have split and diversified into engineering (possibly claiming some from the chemistry area too) and materials. 2010 also marks the emergence of a bio-oriented community, which is targeted by the DL_FIELD project. From 2016 onwards we also see a small but clear adoption by the pharma and geology communities.

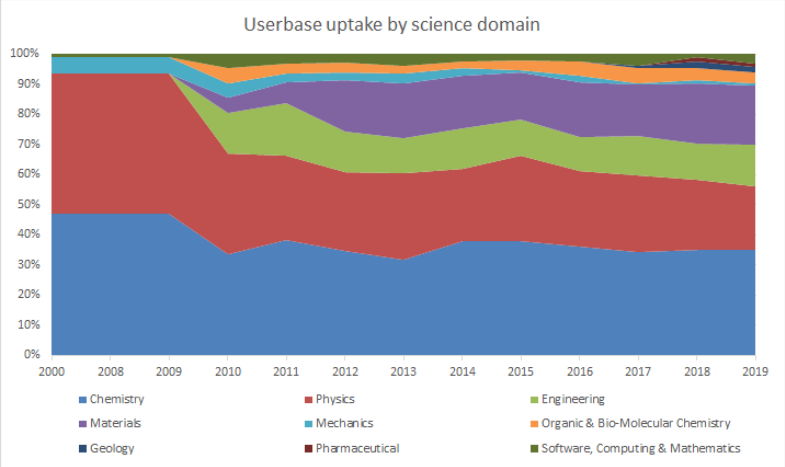


Figure 12: DL_POLY usage by domain

Commercial use and exploitation of academic software is difficult to measure exactly. The commercial proliferation of DL_POLY is not easy to highlight due to various factors. Some of these can be summarised as: (i) the 'free to academia' spirit of the project which has led to companies preferring to fund projects in academia; (ii) the rising competition from USA and Swedish MD projects; as well as (iii) the lack of momentum for strategic commercialisation with external partners by CCLRC/STFC. Despite this, since 1999 Daresbury Laboratory has issued over 23 commercial licences for DL_POLY to industrial and government regulated companies in 9 countries – UK, Norway, India, USA, France, China, Japan, Germany and South Africa – in industrial areas as diverse as petrochemical, chemical, marine biology, glasses, metal alloys and consumer and food safety chemistry (soft matter). Some companies such as ICI, Unilever and IBM have also entered strategic collaborations with STFC and used the program for their own research in a collaborative manner with STFC. There has also been an additional range of companies with interest in the HPC aspects of DL_POLY which acquired free of charge licences for testing and software engineering purposes in order to tender to HPC services procurement calls by UK and French academic establishments. Since early 2000s this list includes companies such as IBM, Cray, SGI, NEC, AMD, Intel, Fujitsu, Hitachi, Bull, Dell, Atos, OCF, NAG, NVIDIA, AWE, ClearSpeed, etc.

The actual impact to industry and their usage of the project cannot solely be judged by the aforementioned licences alone. As demonstrated above DL_POLY has been used in postdoctoral and graduate studies in academia funded by industrial partners. Companies such as Unilever, Astra Zeneca, AWE, GSK, Johnson Matthey, Syngenta, Tata, BASF, Sony and Samsung have funded such projects with partnering universities and institutes since the mid-1990s. Bill Smith and Maurice Leslie were in collaboration with ICI and later Unilever and AZ through CCP5 academic colleagues such as Neil Allan (ICI and University of Bristol) and Dominic Tildesley (University of Southampton and then Unilever), who have had careers in both industrial and academic settings. It is worth noting that many of the big CCPs – 4, 5, BioSim – as well as their HEC equivalents, also have an industrial member on their executive committees, whose function is to facilitate industrially funded projects in academia that are carried out with software packages under the CCPs flagships.

Finally, other funding avenues such as UK's BEIS (Hartree Centre) and FP7/H2020 (by the EC's R&I at STFC) calls have also provided research and technology development projects in direct collaboration with industrial partners such as IBM, Unilever, GSK, Novidec, Infineum, DOW, Syngenta, GSK, Astra Zeneca, BMS, Pfizer, Johnson Matthey, Sony, NPL and NNL. Also, hardware vendors have funded a number of small, short-term projects to explore optimisation and acceleration strategies within DL_POLY 4 – IBM (2007-2010), Cray (Cray Centre of Excellence - 2009-12), Fujitsu (2013), Intel (Intel

Parallel Computing Centres at Dublin and STFC's Hartree Centre - 2014-2018) and NVIDIA (Dublin - 2009-2014, STFC 2014-2018). While in most such cases these projects were a learning exercise, one example stands out – the Dublin -STFC collaboration funded by NVIDIA and leading to DL_POLY CUDA+openMP port, which was community released in 2010 and supported until 2014.

6 The CCP5 Summer School

We noted above that the CCP5 Summer School[23] has been an outstanding feature of its programme over many years. A great many students have experienced the School; in the early years participation was around 30-40 per School, and now the figure is at the 70 – 80 level. We are grateful to Mike Allen and Dominic Tildesley, two of the main instigators, for this account of the origins of the School.

Mike Allen writes: "Following the 1988 NATO ASI in Bath, Richard Catlow organised a "debriefing" discussion over dinner during which the organisers reviewed how things had gone. I believe that the idea for a regular summer school came out of that. The practical exercises at the ASI had been very wide-ranging, and were graded by degree of difficulty, from introductory exercises (some of which merely required pen and paper) to small projects suitable for a team. They seemed to be much appreciated by the participants at the ASI, and it was agreed that a regular event, incorporating lectures and hands-on workshops, with a small team of lecturers, would be valuable. It took a few years to get this off the ground, under the CCP5 umbrella, but without financial support. It soon became clear that there was a demand within the UK for such a school, and eventually we were able to open it up to the international community, with support from CCP5, the ESF programme SIMU, and later, CECAM and other sources."

Dominic Tildesley writes: "As Mike has noted, one of the important developments was the CCP5 Spring and later Summer Schools. I was not involved in the NATO meeting in Bath where these were first discussed but I think, we ran the first school in Southampton in 1994. Subsequently, the schools were run in Bristol, Imperial College, and Manchester and the last one that I was personally involved in was Manchester in 2001 (well after I had started working for Unilever). I know they have gone from strength to strength and are still running in 2018 (in affiliation with the CECAM programme). We evolved a good set of basic lectures covering all the fundamentals of MD and MC and in the afternoons, we ran hands on practical sessions which provided many of the students with their first coding experience. Mike put an enormous effort in to developing a set of robust problems and exercises which should surely be part of the archive[24]. At the end of each day, we had a guest speaker on a topical modelling theme. These seminars went down really well with the students and we also used the visits as an excuse to continue the scientific discussion over some memorable dinners. Over the years, I would estimate that this course will have trained between 700 and 1000 young scientists in this important area. If we had a list of the participants, I am sure that we would

[23]We note that the Amsterdam Summer School shares many characteristics with the CCP5 School, and there has been much cross-participation over the years.

[24]They are indeed preserved in the CCP5 archive [4].

identify a good many of the current thought-leaders in the UK computational research both in industry and academy."

These comments give a good flavour of a research family in action; something for which CCP5 clearly feels proud. The following paragraphs give the full story of the Summer School as it developed. This shows, among other things, just how much effort was put into it by the leaders of the UK simulation community.

The "Methods in Molecular Simulation" School was started by Mike Allen (Bristol), Julian Clarke (UMIST) and Dominic Tildesley (Southampton), who were all key participants in, and sometime chairmen of, CCP5. The first School was launched on 11th April 1994 at the University of Southampton, with Tildesley as local organiser. The course was intended for postgraduate students requiring a knowledge of molecular simulation for their research. The School ran over five days and covered basic and advanced molecular dynamics and Monte Carlo augmented by lectures on molecular modelling and data visualisation. The principal course lecturers were the three organisers and the supplementary speakers were Stephen Warde (Molecular Simulations inc) and Tim Forester (Daresbury Laboratory). Following the example of the NATO ASI course on "Computer Modelling of Fluids, Polymers and Solids" at Bath in 1988, which was said to be the inspiration for the School[25], the afternoons were given over to practical computational exercises, designed to give students hands-on experience of simulation. Each day was rounded off with a guest seminar, which at this School were given by Julia Goodfellow (Birkbeck), Geoffrey Luckhurst (Southampton), Paul Madden (Oxford) and Mark Rodger (Reading). According to a questionnaire returned by students, the School was well received. It also emerged that the students attending fell into two distinct categories: those who were active programmers and those who were purely users of established packages. This initiated a modification of the computational exercises for future workshops, so that these different needs were catered for.

The School became an annual event. Between the years 1994 and 2001, Schools were held at Southampton (2), Imperial College (1), Bristol University (2) and UMIST (3). As with the first School, the course covered the basics of molecular simulation methods (molecular dynamics, Monte Carlo, potential energy functions etc.) supplemented by more specific topics such as polymers, *ab initio* methods, non-equilibrium methods, quantum effects, catalysis, surfaces and interfaces, liquid crystals, phase transitions etc. which varied from year to year. Time was also given to computational matters, in particular molecular graphics, parallel computing (the DL_POLY parallel simulation package was emerging at this time) and basic programming. The course lectures were generally presented in morning sessions with the afternoons given over to the practical computational exercises. The educational material was mostly devised and produced by Mike Allen, Julian Clarke and Dominic Tildesley, with additional input from Tim Forester and Bill Smith. Each day a research lecture on an advanced topic lecture was given by a guest speaker drawn mostly, but not exclusively, from a UK University. Some well-known speakers from these early days include Steve Parker, Mike Gillan, Richard Catlow, Paul Madden, Julia Goodfellow, George Jackson, Mike Finnis, John Harding, Mark Rodger, Julia Yeomans, Nigel Wilding, Leslie Woodcock, Michiel Sprik and Berend Smit. Some of these speakers also occasionally lectured in the basic areas. Other events that took place during the School included the poster session (where students could present their own research), tutorials, a "round table" (in which students could ask wide ranging questions) and social events intended to foster a communal spirit among the participants.

[25] The practical sessions at the NATO ASI are described in a 100-page appendix to the proceedings [51].

In the early period (1994-2001) the School was resourced in a variety of ways. To begin with the registration fees from the students (which numbered about 30) constituted the major part of the funding. There was additional sponsorship from the host Universities, principally in the form of facilities (lecture theatres, computing resources, secretarial assistance etc) and sometimes sponsorship from commercial ventures, including the Numerical Algorithms Group (NAG), Molecular Simulations and Biosym. The School also obtained sponsorship from the Statistical Mechanics and Thermodynamics Group of the Royal Society of Chemistry and from the ESF Programme SIMU. CCP5 also provided support, though in the very beginning the School was a self-financing venture. Tim Forester and Bill Smith participated under the Daresbury CCP5 support arrangement and contributed secretarial and lecturing support along with the DL_POLY software. From 1997 however CCP5 also provided direct financial support and thereafter made provision for this funding in its renewal proposals, beginning with the bid in 1999.

As the School evolved, the augmentation of the basic course with specialised topics developed in 1998 at Bristol into advanced sub-courses within the framework of the School. The advanced courses offered a degree of choice. In the first choice students could choose between *First order Phase Transitions* or *An Introduction to Tight Binding* and in the second between *Configuration-biased Monte Carlo techniques* or *O(N) Methods*. The courses were given by Marjolein Dijkstra (Bristol) and Andrew Horsfield (Oxford). The inclusion of these courses meant the School extended from five to six working days. The extra day was accommodated by running the course from a Wednesday through to the following Wednesday and included a rest day on the Sunday. Advanced sub-courses and an extended programme thereafter became a permanent feature of all subsequent Schools.

The year 1999 marks the point at which the School switched from a Spring to a Summer event. The programme for the first Summer School, held at UMIST in Manchester over ten days (21-30 June), was the most ambitious to date. The basic course provided five lectures on molecular dynamics (speakers: J Clarke, D Tildesley and B Smith), four on statistical mechanics (speaker: M Allen), five on Monte Carlo (speakers: D Tildesley, J Clarke and M Allen), two on potentials (speaker: D Tildesley) and three on applications (speakers: D Tildesley, J Clarke and M Allen) - all supported by eleven practical sessions. In addition, there were seven research seminars from guest speakers (Peter Bolhuis, Richard Catlow, Ian Hillier, Paul Madden, Keith Refson, Nigel Wilding and Mark Wilson), an advanced course of three lectures on *Complex Fluids* given by Berend Smit (Amsterdam) and a day-long advanced course on *Mesoscale Simulation* given jointly by Julia Yeomans (Oxford) and Neil Spenley and Maarten Hagen (Unilever). Two tutorial sessions were provided and two poster sessions. The full programme included a free Saturday afternoon and Sunday[26] for the students to relax.

[26]A "Conference Trip" was organised for Sunday, the details of which were not recorded.

Figure 13: The CCP5 Summer School in Sheffield, 2009

At the second UMIST Summer School in 2000 (19-27 June), the basic course was only a slight variation on the previous year. There was only one potentials-related lecture - on long ranged forces (given by Michiel Sprik) and one applications lecture - on industrial applications (given by Dominic Tildesley). The guest lecturers speaking on research topics remained as an essential component. Only one tutorial session was offered, but two new lectures on "special topics" (quantum Monte Carlo and rare event simulation) were presented by Mike Allen. The advanced courses were condensed into two days offering three choices running concurrently: *Advanced Monte Carlo* (speaker: Nigel Wilding (Edinburgh)); *Advanced Molecular Dynamics* (speakers: Pietro Ballone (Messina)); and *Dissipative Particle Dynamics* (Speakers: Andrew Masters (Manchester), Wouter den Otter (UMIST), Patrick Warren and Neil Spenley (both Unilever)). Each advanced course consisted of four lectures and six practical sessions. The advanced courses started on the Sunday at midday and ended on the following Tuesday at midday, making the School nine days in duration. The format of this School was repeated again at UMIST in the following year (2001) and it set the pattern for the School for a long time after. By this time, the attendance at the School was 81 students.

The year 2002 marked a significant moment in the history of the Summer School. Its three originators, Mike Allen, Dominic Tildesley and Julian Clarke handed over the running of the School to the CCP5 Executive Committee. Dominic Tildesley was by then VP Chief Scientist for the Home and Personal Care Business in Unilever, and time constraints made it impossible for him to continue with the School. This may have been the decisive factor influencing his colleagues, but whatever the reason, it was clear that the School was now a successful and much valued asset of the molecular simulation community, which CCP5 was keen should continue. John Harding, the chairman of CCP5 at that time immediately set up a new group he called the Summer School Working Group (SSWoG) to take on the job of continuing the School (statement of intent: *to train the next generation of researchers in state-of-the-art molecular simulation techniques and applications*). This group was to meet three or four times each year to plan the annual event and report back to the Executive Committee. The original members of the SSWoG were: John Harding (Sheffield), Bill Smith (Daresbury), Jamshed Anwar (King's College London), David Heyes (Guildford), Philip Lindan (Kent) and Keith Refson (Rutherford-Appleton) and they first met on 8 May 2001. Between May 2001 and April 2002, the SSWoG met six times to plan for the 2002 School, which would be held at King's College London.

Having inherited a successful operation, the SSWoG intended to adhere to existing practice and only gradually introduce changes. Much of the material of previous Schools was copied wholesale, in particular, the associated website and the exercises for the practical workshops. Additional material was also available from the DL_POLY training exercises. The website was installed at Daresbury and

made accessible via the CCP5 website. In succeeding years, new practical exercises would gradually replace the originals.

The 2002 Summer School was announced in December 2001 and was set to run from July 8 to July 16. CCP5 covered the costs of staging the event, while SIMU provided travel funds to assist attendance by European students. (SIMU also provided bursaries in the 2003 School.) As with previous Schools there would be a set of basic lectures and a choice of advanced topics for the final weekend. The local organiser for the event was Jamshed Anwar.

Following the announcement, applications to take part substantially exceeded the 60 places available and a means of selecting the students deemed most suitable was drawn up. Postgraduate students with at least one year of research experience and could demonstrate a real need for subject matter offered were considered most suitable. A demonstrable enthusiasm to attend (evidenced by a willingness to give a seminar or present a poster) was also a factor, as was the whole hearted backing of the student's supervisor. Students were thus selected on purely academic criteria and not restricted on the basis of nationality or gender.

The lectures for the basic course were as follows:

Topic	Lectures	Speaker
Why simulation?	1	John Harding
Molecular Mechanics	1	David Heyes
Basic Statistical Mechanics	1	David Heyes
Statistical Mechanics	2	David Heyes
Molecular Dynamics	2	Bill Smith
Advanced Molecular Dynamics	1	Keith Refson
Monte Carlo Methods	2	Nigel Wilding
Advanced Monte Carlo	1	Nigel Wilding
Quasi Harmonic oscillator/solid state thermodynamics	1	Philip Lindan
Programming and Program Design	1	Jamshed Anwar
Surfaces and Interfaces	1	John Harding
Free Energy Methods	2	Jamshed Anwar
Long Ranged Forces	1	Bill Smith
Potentials	1	John Harding

These lectures were supplemented by appropriate practical sessions held in the afternoons. The advanced courses, each of which consisted of four lectures and supporting practical exercises, were: First Principles Simulation (Keith Refson, Philip Lindan); Simulation of Organic and Bio Molecules (Xavier Daura); and Mesoscale Methods (Peter Coveney, Jonathan Chin).

Each day of the School a plenary (research) lecture was presented. The speakers at this School were Mike Allen (Bristol), Julian Clarke (UMIST), David Adolf (Leeds), Paul Madden (Oxford), Dominic Tildesley (Unilever) and Richard Catlow (Royal Institution).

An innovation at the 2002 School was the presentation of short talks by four selected students on the subject of their research. A prize (a book) was awarded to the student who was considered by the organisers to have made the best presentation. A similar award was made to the student presenting the best poster in the poster session. Feedback from a student questionnaire distributed by John Harding confirmed the impression that, as a learning exercise, the 2002 School had been a popular success. Less popular however, was the accommodation allotted to the students, which proved to be much too far from the entertainment centres of London!

Having successfully run their first Summer School, the SSWoG continued in a similar vein for several successive years at King's College (2003), Cardiff (2004, 2005, 2006) and Sheffield (2007, 2008, 2009). The main changes that took place were primarily permutations of the course lecturers and membership of the SSWoG. The structure of the School however remained recognisably the same. Karl Travis (Sheffield) took the place of David Heyes from 2003 and David Willock joined from 2004. Rongshan Qin (Daresbury) took over the advanced Mesoscale Method course in 2005 and was himself replaced in 2006 by Ian Halliday (Sheffield Hallam) in 2006. Also in 2006, Mark Rodger (Warwick) took over the advanced molecular dynamics lectures, Philip Camp (Edinburgh) took on the Monte Carlo lectures and Keith Refson became sole lecturer on the advanced First Principles Simulation course. After the 2009 School, Bill Smith stepped down and was replaced by John Purton (Daresbury).

After the School of 2003 the SIMU support ended. However an application was made to Marie Curie Actions for support of the Schools from 2004. This bid was successful and provided an annual budget of 85,000 Euros, which allowed the students to participate for a reduced registration fee of £30. This however rose to £70 in the following year, as it became apparent that too low a fee encouraged late cancellations. Throughout this period, CCP5 continued to contribute £7,500 in support. The Marie Curie Actions support came to an end after the 2007 School and, at this juncture, CCP5 was unfunded. The registration fee for 2008 promptly jumped up to £400. Remarkably, the School remained oversubscribed. However, from 2009 on the School has obtained support from the CECAM organisation in Lausanne.

The 2010 Summer School took place at Queen's University, Belfast 18-27 July. Marco del Popolo (Belfast) joined SSWoG as the local organiser and also a course lecturer. The other lecturers at this time were John Harding, Mark Rodger, John Purton, Philip Camp and David Willock. The advanced courses were given by Keith Refson (First Principles), Xavier Daura (Biomolecular) and Ian Halliday (Mesoscale). The Plenary speakers included Xavier Periole (Groningen), Christian Holm (Stuttgart) and Michiel Sprik (Cambridge) among others. With the exception of Mark Rodger, who was replaced by Steve Parker and a returned Jamshed Anwar (after an absence of one year), the same course lecturers also took part in the 2011 School, which was also at Belfast (17-26 July). At this event the plenary speakers included Julia Yeomans (Oxford), Rebecca Wade (Heidelberg), David Coker (UC Dublin) and David Quigley (Warwick).

In July 2012 the School returned to Cardiff, where David Willock was the local organiser and the programme followed that of 2011. After ten years Xavier Daura stepped down and was replaced as the advanced course lecturer on the simulation of bio- and organic molecules by Jamshed Anwar (Bradford) and Rebecca Notman (Warwick). After the School ended, Philip Camp stepped down as the lecturer on Monte Carlo methods.

The following three years from 2013 to 2015 the Summer School was held at Manchester in the month of July, with Paola Carbone acting as the local organiser. The course lecturers by this stage were John Harding, Jamshed Anwar, Keith Refson and L. Lue, with the advanced courses being given by Jamshed Anwar with Rebecca Notman (Biosimulation), Ian Halliday with Michael Seaton (Mesoscale simulation) and Keith Refson with Leonardo Bernasconi (First Principles simulation). In 2015 the biosimulation advanced course was presented by Richard Henchman (Manchester) and Syma Khalid (Southampton).

Figure 14: Daan Frenkel gives an award

In the three years from 2013 to 2015 the CCP5 Summer School was held at Manchester and from 2016 to 2018 it was at Lancaster. All of these Schools followed similar programmes: lectures on basic methodology in molecular dynamics and Monte Carlo and advanced lectures on the simulation of bio-molecules, first principles simulation and mesoscale simulations – all backed by challenging practical sessions interspersed with research seminars delivered by plenary speakers. The thrust and motivation of the School remained the same as it was when first established by Allen, Clarke and Tildesley in 1994. It is a successful formula that retains its impact and appeal.

This brings us to the present day. The success of the Summer School is hard to deny and is due to many factors. The financial support of various organisations, beginning with CCP5 and extending to SIMU, Marie Curie Actions and CECAM, has been extremely valuable, as has the generosity, in terms of resources committed, of the universities where the Schools have been staged. As welcome as these contributions have been, however, they are not the main reason for the School's success. A much more important factor has been its wholehearted adoption by the molecular simulation community. There is an enthusiasm for the School that is very evident to those who run it. It is seen in the willing participation of the lecturers who have given their valuable time and hard effort to keeping the courses relevant and accessible. Some have worked for the School for a surprising number of years and there are undoubtedly many who would take their places if the opportunity arose. Then there are the many plenary lecturers who have responded positively to invitations to speak at what is very far removed from a prestigious conference. This is all capped by the enthusiastic participation of the now many hundreds of students who have appreciated what the opportunity to attend meant for their future careers. The shared experience and evaluation of all who participate in, and afterwards promote, the School is a true measure of its worth, and a real testament to the value of the CCP5 enterprise.

7 Reflections on CCP5

In the preceding sections we have given a factual account of what CCP5 did during its 40 years of existence. It now seems appropriate to reflect on this history and draw out what we think are some of the more significant themes.

7.1 Building a community

What *is* a research community, as distinct from a disconnected collection of individual researchers or groups in a given field? In our opinion, a research community is one in which the members know each other personally, collaborate with each other when appropriate on specific research projects, jointly nurture the careers of young researchers within the community and self-organise to act together to promote and advance their field. No doubt there are other important characteristics, but, broadly speaking, this kind of community exhibits a "family feeling", a sense of belonging and

mutual support. CCP5 put quite a lot of effort into developing and sustaining this community ethos, as have other CCPs, and this must go a long way to explaining its longevity.

It seems clear that building this kind of community depends chiefly on the personalities and attitudes of the main players. CCP5 has always been blessed to have many leading figure who acted naturally and instinctively in ways to support and exemplify this ethos. Given that critical factor, there are many mechanisms that an organisation like CCP5 can put into place to provide an infrastructure with which its community can flourish. Most of these mechanisms have been mentioned in the CCP5 chronology of section 3, but it is worth drawing out the key features here.

From the outset, CCP5 has: issued newsletters; maintained a program library; conducted various outreach activities with schools and the general public; and issued via its extensive email list information on activities and forthcoming events. But a real research community has to *meet*, regularly and in person. Thus the rather full programme of workshops, national and international conferences has undoubtedly been vital in gluing the community together. CCP5 developed an especially effective method of education and social networking in the form of its annual Spring/Summer School – we gave a detailed history of this in section 6. Finally, at any given time, an active research community contains individual research collaborations. Through its programme of pump-priming support, CCP5 acted to make such collaborations easier to initiate, and the many visits by distinguished visiting scientists was undoubtedly an additional stimulus.

7.2 The role of Daresbury Laboratory as a host institution

CCP5, along with most of the CCPs, was based in the TCSE Department at Daresbury. The CCP5 story gives a feel for how this operated and why it worked well. We in TCSE Department did the administrative tasks – the financial processes (budgeting and accounting), payment of expenses etc. We also employed the permanent staff and often the PDRAs working on development projects. This brought a certain stability to the Project, a resilience to the many changes in the funding agencies (the people involved and the rules under which they operated) and in the Project itself (its evolving scientific priorities and research community). If, as did indeed happen with CCP5, grant proposals were rejected (or required to be resubmitted in a different form), Directors of TCSE Department could contrive, in one way or another, to keep things going until we were past the rough spots, discreetly bending the rules if necessary. In particular, the permanent support staff provided essential continuity – intellectual, scientific and technical – and this undoubtedly nurtured the Project over its 40 years. A key aspect of this role is this: scientifically strong permanent staff such as those who worked on CCP5 attract good young researchers, post-docs etc to work with and learn from them. This is vital to the success of any scientific activity, and in addition leads to all sorts of benefits later as the careers of those young scientists develop. From this, it is clear that the CCPs were probably the single most important part of TCSE Department's overall programme, being essential to the Department's expansion throughout the 1990s and 2000s. It was, therefore, very much in the interests of Directors of TCSE Department to make a success of the CCPs.

Likewise, the senior managers of Daresbury Laboratory provided strong backing to the CCPs. Laboratory Directors such as Alan Leadbetter correctly identified computational science as a growth area for Daresbury, bringing to our remit a new community of very high quality and numerous researchers based in the Universities. No wonder, then, that they were strongly motivated to support the CCP programme.

There is also perhaps a point to be made about the ethos of working in a "national laboratory" at least in the UK. Institutions like Daresbury exist, in the UK public research system, to support the

university-based scientific community to build and exploit research facilities of a scale or scope that requires collaboration across the whole UK. We are used to constructing large experimental facilities – particle accelerators, synchrotron radiation and laser sources, neutron sources, etc – in collaboration with university-based scientists, and to hosting such scientists while they perform their experiments. The CCPs are not, of course, in the same physical category as accelerator systems, but the same spirit of collaboration and equal treatment of all partners is entirely appropriate and indeed necessary. Science is a highly competitive business, and university research groups are both collaborators and competitors with each other (for grants as well as glory). Programmes such as the CCPs must be operated in an even-handed way and this is often best orchestrated by an institution outside the university system it must support. As long-time employees of Daresbury Laboratory, we can testify to the reality of this collaborative ethos among all its staff – managerial, scientific, technical and administrative staff. A special word about the administrative staff is in order here. In our experience, these staff (finance people, Human Resources people, site administrators, personal assistants etc) hardly ever have any scientific background but nevertheless acquire a sense of being involved in something significant, pioneering and worthwhile - something to be proud of. This is much more valuable, in our opinion, than research scientists usually realise; helpful and enthusiastic administrative staff are truly worth their weight in gold.

7.3 What went well

First, the science went well; throughout its existence, CCP5 has practiced and enabled a very wide variety of high quality science. This article is not the place to document the entire scientific record of CCP5, but we can at least point to a number of special journal editions devoted to CCP5 and its codes (see Figure 15). Molecular simulation is now a widespread and essential component of research in many, many fields (and several CCPs); CCP5 can claim to have played an important role in bringing this about, especially in the UK. Collaborations, including cross-membership, with other CCPs were quite common. Several joint meetings with CCP4 and 9; proposals on materials with CCP1, 3 and 9; work with CCP12 on DL_MESO. Moreover, by staying at the forefront of developments in computing technology, CCP5 ensured that its members could run its codes efficiently on the most powerful computers (and influenced the provision of such equipment in the UK).

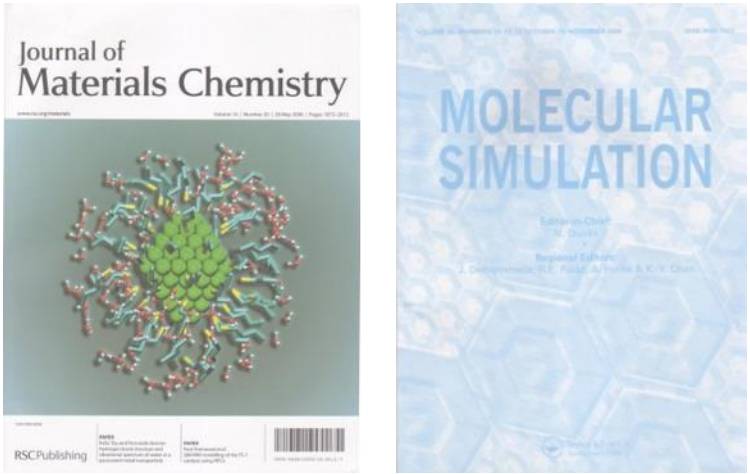


Figure 15: Special journal editions devoted to CCP5

Second, CCP5 really did build and nurture a research community, as we commented in section 7.1, the Newsletter and the Summer School being striking examples of this. The CCP5 newsletter was produced as hard-copy regularly (initially quarterly) from 1981 to 1995 and sporadically after that until its last issue in 2002. It provided CCP5 members with announcements and registration forms pertaining to forthcoming CCP5 conferences, workshops and schools and featured informal articles on scientific and computational matters contributed by members. Many of the articles submitted were speculative or novel and had an impact that exceeded expectation for an informal circular. It did much to propagate new and better methodology among its international readership and without doubt did much to promote CCP5's international reputation. Its impact in the early days was substantial and drew many contributions. It was therefore a major factor in establishing CCP5 as a global project. Its slow demise was due in part to its cost (reprographics and postage costs were substantial) and partly because its original pool of enthusiastic contributors dwindled with time. Its primary functions are now, of course, managed through the CCP5 website. Dominic Tildesley comments: "the *CCP5 information quarterly* was regularly published. This enabled the discussion and documentation of simulation methodology which did not appear in the contemporary scientific journals. These pamphlets were informal and yet highly informative. They were a model for the development of journals such as Molecular Simulation which insisted on the details and in many ways inspired the approach that Mike and I took in writing the first edition of *Computer Simulation of Liquids*."

Third, CCP5 has been strongly international in character from its birth. This seems to have been a natural inclination of the founders of the Project, but over the years it has become ever clearer that internationalism has been a key element of CCP5's success. Links with other organisations such CECAM were quickly initiated, and today are very strong. Indeed, in its own early days in Orsay, CECAM played a key role in the intellectual development of molecular simulation, as CECAM's archives demonstrate [55]. There has always been close collaboration between the leaders of CECAM and CCP5, particularly regarding the CCP5 visitor programme and the Summer School[27]. CCP5's overseas visitor programme has indeed been a key element in its internationalist outlook, contributing greatly to the profile of the Project outside the UK. Vlad Sokhan, now at Daresbury, has marvellous stories [12] about how the CCP5 Newsletter and CCP5 visiting scientists influenced him in his early days as a researcher in the mid-1980s at the Kiev Institute of Theoretical Physics, Ukraine (then USSR). The Institute fund its way on to the CCP5 mailing list (those were the days of perestroika in the USSR), and as soon as the Newsletter arrived it was immediately taken and circulated among Vlad and his colleagues. Although they could not access the FTP site, codes and papers in the Newsletter were avidly read and discussed in the group. The Newsletter was literally read until "it had holes in it". Vlad comments: "In fact, I am indebted to CCP5 for changing my life – it was due to visits first of Keith Gubbins in 1987 and later, in 1988 by John Rowlinson that broke a window (I could not say open – the effect was dramatic) in the post-cold-war wall encasing us and initiated the contacts that eventually led me to the UK." Today, CCP5's outreach activities are also fully international; members of the project (including those in the materials community) have taken versions of the Summer School on the road to India, China, Croatia, Russia, Colombia, Chile, Taiwan, South Africa and Zimbabwe. This is surely something of which CCP5 can be proud.

Fourth, CCP5 came into being at a time when computational science was about to undergo an explosion in interest precipitated by the promise of new, relatively inexpensive and powerful

[27]The CCP5 Summer School has recently been co-funded by CECAM, along with the Amsterdam School in molecular simulation.

computer architectures. This opened up the promise of completely new approaches to chemists, physicists, biologists, materials scientists and engineers. However, alongside the rising interest there was comparatively little expertise in software development and exploitation. The CCP5 program library (in common with other CCP program libraries) helped to overcome this deficit by offering free and expertly developed software to get newcomers started. The range of software that rapidly filled the library (allied to the associated articles in the CCP5 newsletter and CCP5 workshops) opened the door to many new participants. Dominic Tildesley comments: "the establishment of a strong program library which was extensive, well-documented and where the FORTRAN code was available to the end users was particularly powerful. Although Bill [Smith] and Maurice [Leslie] took the lead, the community helped to develop the code and to debug it through applications. It was free to academics and essentially free to industry. This helped the field to grow sustainably in both environments in ways that the commercial codes (Molsim, Biosym etc) could not."

While the CCP5 program library has been a valuable resource to the molecular simulation, by the early 1990's it was apparent that a radically new software approach was required for state-of-the-art simulation, which was inherently parallel in nature (see section 4.5). It was also clear that molecular simulation was moving away from simple chemical and physical systems to systems of great complexity, including biological systems *in solution*, catalytic systems (eg zeolites), polymers, liquid crystals, nucleic acids and so on. It made little sense when also faced with the architecture of parallel machines to offer a suite of single-application programs. A demand for a general purpose program that could perform all the likely simulations then nascent was therefore evident. This led to the development of first program with the DL_ prefix: DL_POLY_1.0. We have told the story of the DL_POLY program suite in some detail in section 5. As a prime example of a community code (see sections 4.7 and 5), it is undoubtedly one of the key achievements of CCP5 over its lifetime, exemplifying the close connection the Project maintained between its scientific goals and the burgeoning developments in hardware and software technologies. There are great simulation codes elsewhere, accessible remotely, but the critical issue for users is *support* and support must be local. This is what DL_POLY achieves for CCP5.

Though not immediately apparent, it is probable that CCP5 played a major part in the creation of a number of spin-off projects, some of which included major computational consortia. This partly resulted from CCP5 "pump priming" collaborations in the community through provision of travel funding for meetings between potential collaborators, with the intention that they make subsequent bids for research funding, which included bids for consortia. In this way and in other less direct ways (eg workshops) CCP5 was a catalyst for future research programmes [52]. Examples include: Complex Fluids Consortium (Mike Allen et al); Bio-Inorganic Interface Consortium (John Harding et al); Radiation Damage/ Immobilisation Consortium (Richard Catlow).

Fifth, that CCP5 adopted a constitution defining the role of the chairman, Executive Committee and Steering Committee has been a major factor in its longevity and its success as a resource for the simulation community. Its particular strength was that it was democratic in principle. The chairman and the Executive Committee were elected by a poll of the members of the community (as defined by its membership/mailing list). They served a fixed term (at least formally, though in the case of the chairman the term may be longer than the nominal three years if circumstances required it – such as a failure to renew the project on schedule). The Steering Committee was an annual gathering of CCP5 participants at a special session of the Annual General Meeting, which was otherwise a scientific meeting with a programme of lectures. The chairman and Daresbury staff presented an account of work done on CCP5's behalf and comments, ideas, suggestions and requests were invited. Significantly, overseas members of CCP5 were able to offer constructive comment. That the officers of CCP5 were so accountable to the community meant that project

could not degenerate into a private project at a single university and was therefore free of conflict over its resources.

Finally, we would argue that the most significant success of CCP5 was, and is, the quality of the people who worked in it in the Universities and at Daresbury. This may have been good judgment or good luck; probably a combination. In any case, the essential point is that a project like CCP5 needs to be run by quite special people. The Simon report [8] put this perfectly:

"The research community needs national, institutionalized user support, where the practitioners have a long-term career path, which is not the case in a typical university environment. This support must be provided by well-rounded professionals who can speak to the users as scientific peers, but who also have excellent understanding of algorithms, code optimization, and software engineering, as well as a mindset geared to helping others rather than a focus on independent scientific research. Such people can be a source of cross-fertilization between different research groups, recognizing that the expertise of one group can help another. …. [TCSE Department] at Daresbury appears to play this role already, and the Panel often heard testimony of users' satisfaction with the support they received from Daresbury."

"Well-rounded professionals" is an apt description of the people who worked in CCP5. They attracted and mentored several young scientists who became well-rounded themselves. Finally, over its 40 years of existence, CCP5 has touched and influenced the scientific careers of many researchers who, together with their students and co-workers, are now at the heart of molecular simulation and computational statistical mechanics in the UK. CCP5 gave such people not only the scientific and technical education and tools needed for their research, but personal contacts with others in their field; it gave them their personal scientific network.

Ultimately, though, the only way that a project like CCP5 could have survived and prospered for so long is through the continuing commitment of its members in the academic research community. It is impressive to see how CCP5 established itself as a community in the true sense, nurturing its members and acting together to reach common scientific goals. This is the essential collaborative ethos for which CCP5 should be celebrated.

7.4 And not so well

It is evident from the CCP5 chronology in section 3 that the Project has a poor record of securing funding for so-called flagship projects at the time of its funding renewal. Indeed, it succeeded in obtaining funding for only a single flagship project, DL_MESO[28]. This meant that the project was often denied a primary scientific focus for research and development (with an associated PDRA) to run parallel with its highly successful network activities. On its inception in 1980, the project was awarded a PDRA for the purpose of developing a suite of simulation programs for the Cray 1s supercomputer at Daresbury. But on subsequent bids to renew the project it failed many times to obtain similar support for new scientific areas.

Thus, although CCP5 always had a very strong programme of scientific, technical and community code developments, this was done more or less in spite of the CCP funding system rather than because of it. This is certainly not true of all CCPs, many of which were able to conduct plenty of flagship projects funded by the CCP grant system. Of course, peer review is, to a degree, a lottery.

[28]DL_POLY eventually received grant funding, but this was a one-off research proposal, not part of the regular CCP5 renewal process (see section 5).

Perhaps CCP5 simply did not write strong enough proposals, although it is difficult to believe this. It is true that any flagship project dedicated to a single scientific topic might appear unconnected with and superfluous to CCP5's flourishing and wide ranging network programme. In a competitive funding environment, the temptation for referees to cut off an expensive appendage must have been great. However, in our opinion it is likely that intrinsic tensions in the peer review of CCP grant proposals had something to do with CCP5's lack of success here. The story of Mark Rodger's heroic attempts to make a successful renewal proposal in the 2006-2009 period (see section 3.4) illustrates many of the factors involved.

Because CCPs are supposed to include all the major UK research groups in their field, CCP proposals are *community proposals*. The UK scientists most suited to assess such proposals are likely to be already involved in them and therefore naturally excluded as referees. To those of us who saw many CCP proposals over the years, it was apparent that many of the chosen referees did not know anything at all about the CCPs; what they aimed to do, how they worked and so on. This could result in a level of misunderstanding that was sometimes almost comical. The point is that a CCP is an *infrastructural component* of its given research field. It certainly is not a stand-alone individual research project of the kind that standard peer review systems are set up to assess. Scientific excellence is always required, of course, but, compared to a normal 3-year research project, in a CCP that excellence is different in nature, in its timescale and in the way it manifests itself. One could argue (and did so, from time to time) that standard grant peer review was not the right way for Research Councils to assess and fund CCPs, but this argument was always somewhat counter-cultural and did not gain support from the funding agencies, perhaps understandably. Nevertheless, to mitigate these difficulties, Daresbury and EPSRC colleagues produced an explanatory briefing note (reproduced in Appendix A which for some years was included in the instructions to referees reviewing CCP proposals. This helped, until the "system" forgot about it.

By and large, successive CCP5 Executive Committees took the axing of the flagship project at face value (i.e. as failing on scientific merit), but arguably failed to evolve an approach that integrated the flagship and network projects more closely. Perhaps this also points to something of a failure to *market* CCP5 more generally. We could have paid more attention to raising the profile of the Project more effectively among our scientific peers, the research funding community, industrial research and, indeed, the general public. After all, we had a good story to tell, but we didn't tell it very often. This might or might not have solved the flagship problem, but it surely would have been of benefit to CCP5 in all sorts of ways.

We noted in section 4.6 that in the 1990s, the Research Councils set up Consortia to ensure the efficient use of the succession of HPC systems procured under their high-end computing strategy. CCP5 spawned a number of these Consortia, in the sense that CCP5 members were involved in and often led Consortium proposals and operations. In the subsequent years, these Consortia were the most important means by which researchers could access the most powerful computing resources. This, in turn, gave rise to an increasing need for the use of the kinds of software engineering methodologies we discussed in section 4.7. As Mike Allen comments: "When the HPCI [high performance computing initiative - see section 4.6] consortium got going, it became clear that none of the groups had enough experience of software development in the context of a large collaboration: version control, branching, merging etc. CCP5, or the CCPs in general, might have done a bit more, a bit earlier, to encourage us to learn more formal software development methods." Moreover, in the HPCI the CCPs *responded* to a strategy adopted by their funding agencies, the Research Councils. This perhaps raises the question: could the CCPs, especially CCP5, have asserted their undoubted scientific and technical authority so as to *shape Research Council strategy*, not merely react to it? The influence of CCP5 was no doubt there in the strategy, but

mainly through individuals rather than CCP5 as an institution in its own right. There is clearly an intrinsic tension between CCPs and their funders. On the one hand, a group of scientists associate autonomously together to represent and advance their own field, and therefore naturally include all (or nearly all) the major groups and leaders in that field. They presumably feel justified in claiming a particular expertise and knowledge in their own domain. On the other, the Researcher Council funders, who have the responsibility of allocating public money effectively, naturally regard such research associations simply as grant-holders with no particular authority in policy issues. The system must try to find a balance between these points of view, and it did try. But perhaps CCP5 and the other CCPs underplayed their hand in strategic matters, generally.

Notwithstanding its problems with peer review and profile, CCP5 has survived the many Research Council reviews intact. As managers, we sometimes complain of "lack of ownership" by Research Council programme managers, fading institutional memory of what the CCPs are supposed to be about, and so on. But a bottom-up, community-based structure has advantages in terms of autonomy and identity. Certainly, CCP5 kept going despite funding knockbacks – this is undoubtedly a resilient and committed community. We believe that it should be on the record that the relationship of CCP5 with its Research Council funding agencies has, over 40 years, been a strong and productive one, and this is a credit to both sides.

7.5 The present state of CCP5

Table 2 shows what the overall CCP Programme looks like now. Note first that the overall scale of the CCP programme has remained essentially the same for 40 years, with defunct projects being replaced by new ones, and that CCP5 is one of the few individual projects to have survived intact for all that time. Note second that molecular simulation is now practiced in several of the CCPs: in molecular biology, crystallography, fluid dynamics, the electronic structure of solids, and so on. This reflects the development of molecular simulation world-wide over the last 40 years, of course. CCP5 itself has "given birth" to some of these UK developments, as specific interest groups within its own community have grown and adopted their own particular approaches – biomolecular simulation might be an example of this.

- CCP4 : S/W for macromolecular x-ray crystallography (David Brown)
- CCP5 : Computer simulation of condensed phases (Neil Allan)
- CCP9 : Computational electronic structure of condensed matter (Mike Payne)
- CCP-Mag : Computational magnetism (Julie Staunton)
- CCP-NC : NMR crystallography (Jonathan Yates)
- CCPQ : Quantum dynamics in ATMOP (Graham Worth)
- CCP-Plasma : Computational plasma science (Tony Arber)
- CCPi : Tomographic imaging (Phillip Withers)
- CCP PET-MR : Synergistic PET-MR reconstruction (Kris Thielemans)
- CCPBioSim : Biomolecular simulation at the life sciences interface (Adrian Mulholland)
- CCP-EM : Electron cryo-microscopy (Peter Rosenthal)
- CCPEngSci : Umbrella organisation for computational engineering consortia (Mark Savill)

Funded by:

EPSRC Engineering and Physical Sciences Research Council

BBSRC bioscience for the future

MRC Medical Research Council

Table 2: the current list of CCPs

Ideas such as the famous Car-Parrinello method now common in CCP9, on the other hand, have, in one sense, merged the techniques of molecular simulation with the quantum mechanics so as to

allow the atoms in electronic structure calculations to be moved. Another way to put this is that the forces in the molecular simulation are determined *ab initio* from electronic structure calculations "on the fly". There is now such a high degree of cross-over between the traditionally separate communities of statistical mechanics and electronic structure that conferences on condensed matter inevitably involve both disciplines. The current CCP portfolio reflects this while maintaining the continuing identities of CCP5 and CCP9. In engineering, specifically computational fluid dynamics, much interest currently focuses on flows through device structures of micro- and nano-scale, where the traditional continuum approximations are not necessarily good. Approaches based on a more atomistic view – both coarse-graining techniques of various kinds and fully atom-resolved molecular dynamics – are now being implemented in an engineering context. This requires codes to handle complex geometries and boundary conditions, non-equilibrium (even reacting) flows and so on. Of course, all this recognises the obvious point the scope and scale of molecular simulation has been transformed over the lifetime of CCP5; in the present discussion we are merely noting how this transformation continues to play out within the CCPs.

Regarding the present through the prism of CCP5's history, Mike Allen writes: "Rejoining the roster of lecturers at the CCP5 Summer School in 2019, it was good to see that the basic structure we established 25 years ago is still effective. Naturally, the material has evolved: the practical workshops now involve less low-level coding, the advanced topics have changed to reflect current research activity, and each generation of lecturers has fine-tuned the content. All these components are carefully coordinated by the CCP5 team, and supported by Daresbury staff. Clearly, there is still a desire, and a need, for students and researchers to understand what is going on behind a molecular simulation program, and the course continues to be oversubscribed. More generally, CCP5 still plays a key role at the heart of the simulation community, through its organisation of meetings, provision of codes, and the mailing list."

In a similar vein, Dominic Tildesley comments: "I think the CCPs are in a strong position to influence the industrial uptake of simulation and modelling. This requires support in the provision and use of leading-edge machines (as exemplified ion the Hartree Centre at Daresbury). This central provision at a reasonable cost is welcomed by industry who are not interested in the ownership of big machines. Give industry open and free access to the software library and expect them to make contributions of code and examples to the community. They will probably seek support from a software vendor for their major codes as they are applied more widely in organisations. Finally, and most importantly, make sure that the CCPs continue to offer consultancy and direction on the use of modelling in industry through special events, invitations to CCP conferences, and confidential evaluations of company modelling strategies. I believe that an important future role for the CCP5 will be as the custodian of a number of honest and inspiring case studies of the use of molecular simulation in pharmaceuticals, fast moving consumer goods, chemicals and the aero-space and automotive industries."

Somewhat ironically, in a very recent review by the Research Councils, funding for the next few years of CCP5 was not approved. As our chronicle shows, this is not the first time something like this has happened, and CCP5 not only survived, but subsequently regained funding and prospered. We are certain that this will happen again. The Project must continue to evolve, scientifically and technically, but we believe that the molecular simulation community in the UK is strong and mutually supportive, in the way that the CCP concept embodies. CCP5 will, no doubt, decide its route into the future by the essentially democratic methods it has developed over the last 40 years: the Project's collaborative ethos is, we hope, future-proof.

Coda

What would have happened to the molecular simulation or computational statistical mechanics field in the UK if CCP5 had never existed? "The rich would have got richer and the poor would have got poorer" (MFG); a few of the most powerful groups would have prospered, but smaller ones would not have survived. Arguably, there would have been no UK research *community* in this field. A collaboration like CCP5 obviously needs an infrastructure of capital equipment and facilities to do its work. But such collaborations are themselves a kind of infrastructure, a *human and intellectual infrastructure* within which their field can flourish. This is perhaps the most important element of CCP5 as it has evolved, the factor that has ensured its vigorous longevity.

Acknowledgements

We thank Professors M P Allen, C R A Catlow, D J Tildesley and J E Inglesfield for helpful comments and insights, and Dr M Leslie for material on lattice energy simulations. We are grateful to Professor Giovanni Ciccotti for his valuable advice and constant encouragement.

Dedication

We dedicate this article to the memory of Professor P G Burke FRS, who died in June 2019. Phil has the right to be regarded as the father of the Collaborative Computational Projects.

References

[1] Michel Mareschal, Eur. Phys. J H https://doi.org/10.1140/epjh/e2018-90006-7
[2] Giovanni Battimelli and Giovanni Ciccotti, Eur. Phys. J H …
[3] J-P Hansen and D Levesque, Eur. Phys. J H …
[4] The CCP5 Archive, available at the CCP5 website –
[5] "CCP News: a news bulletin of the Collaborative Computational Projects", distributed by Daresbury Laboratory between May 1979 and ….
[6] P G Burke, Computer Physics Communications, 114 (1998) xii-xvii
[7] S J Smith and B T Sutcliffe, Reviews in Computational Chemistry, **10**, 271 (1997)
[8] "International Review of Research Using HPC in the UK", Horst Simon (Chair), EPSRC and DFG (2005)
[9] D J Tildesley, private communication (2019).
[10] Lester W.A. (1980) "Recent Developments in Computational Chemistry in the U.S.: the NRCC (National Resource for Computation in Chemistry)" In: Bargon J. (eds) Computational Methods in Chemistry. The IBM Research Symposia Series. Springer, Boston, MA.
[11] J.M. Goodfellow, Review of CCP5 Polarizability Meeting, CCP5 Information Quarterly No.1 p4 June 1981, Daresbury Laboratory.
[12]V Sokhan, private communication (2019).
[13] P.G. Burke, B.W. Davies and D.P. Edwards (eds.) Some Research Applications on the Cray-1 Computer at the Daresbury Laboratory 1979-81 (Daresbury Laboratory, 1982)
[14] B.W. Davies Central Computing Committee Review Working Party Report. Appendix C: a next generation Supercomputer for Academic Research. (CRWP, May 1984)
[15] D.J. Taylor, J.F.L. Hopkinson and C.C.T. Henfrey The Cray-1s and the Cray Service provided by the SERC at the Daresbury Laboratory Computer, Physics Communications 26:3-4 (1982) 259-265, DOI:10.1016/0010-4655(82)90115-1
[16] A.J. Forty et al. (eds.) Future Facilities for Advanced Research Computing, A report of a joint working party of ABRC, UGC and the Computer Board (SERC, June 1985) ISBN 0-901660-73-6
[17] Computational Physics and Chemistry on an FPS-164 / MAX – Experience within SERC, Martyn F. Guest, TCS Division, Daresbury, CCP/86/1, 1986
[18]W. Smith and T.R. Forester, DL_POLY_2.0 A general Purpose parallel Molecular Simulation Package. *J. Molecular Graphics*, 14 (1996), 136.
[19] The DL_POLY Classic Molecular Simulation Package, W. Smith, STFC Daresbury Laboratory,
[20]Todorov, I. & Smith, W. (2004). DL_POLY_3: The CCP5 national UK code for molecular-dynamics simulations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Science*, *362*(1822), 1835–1852.
[21] Todorov, I., Smith, W., Trachenko, K., & Dove, M. (2006). DL_POLY_3: New dimensions in molecular dynamics simulations via massive parallelism. *Journal of Materials Chemistry*, *20*, 1911–1918.
[22] The Intel iPSC/2, An Introductory Guide to the Hypercube, Revision A., R.J. Allan and M. F. Guest, Advanced Research Computing Group, Daresbury Laboratory, 1992.
[23] The Intel iPSC/2, Guide to Hypercube Applications, Revision B., R.J. Allan and M. F. Guest, Advanced Research Computing Group, Daresbury Laboratory, 1992.
[24] Introduction to the INTEL iPSC/860 and iPSC/2 Supercomputers, Version 1.0, R.J. Allan and M. F. Guest, Advanced Research Computing Group, Daresbury Laboratory, November 1990.
[25] Parallel Supercomputing 1992, Parallel Supercomputing 1993 (both published by Daresbury Laboratory, SERC).
[26] Computational Science Initiative; A progress Report, July 1989, D. Fincham and M.F. Guest, Daresbury Laboratory, SERC.
[27] M P Allen, private communication (2019).
[28] ASCI ref

[29] HPCI Conference 1998. EPSRC and the three national HPCI Centres jointly organised a conference in Manchester from 12-14 January 1998 to mark the end of the first 3 years of HPCI.

[30] M.F. Guest, A.M. Elena and A. B. G Chalk, "DL_POLY – A Performance Overview: Analysing, Understanding and Exploiting available HPC Technology", Molecular Simulation (Taylor & Francis) special issue - "DL_POLY: Twenty-five years of molecular dynamics evolution" (2019) (in press).

[31] 23rd Daresbury Machine Evaluation Workshop, 27-28 November 2012, https://eventbooking.stfc.ac.uk/uploads/mew23flyer.pdf, 27-28 November 2012

[32] Computing Insight UK 2017, "Joining Up the UK e-Infrastructure", Manchester Central, 12-13 December 2017,

[33] HPC Profile, Editor R.J. Allan, Issue 1, March 1995.

[34] W. W. GROUP FOR FLOATING-POINT ARITHMETIC, 754 – standard for floating-point arithmetic. https://standards.ieee.org/develop/project/754. html, 1985.

[35] ISO/IEC 9899:2011 programming languages – C. 2011

[36] ISO/IEC 1539-1:2010 information technology – programming languages – Fortran – part 1: Base language. , 2010.

[37] T. M. FORUM, Message passing interface forum. http://www.mpi-forum. org/, Mar. 2009.

[38] H.P. Logan Sankaran, *Hpc clusters: Best practices and performance study*. November 2016

[39] A Computational Science IDE for HPC Systems: Design and Applications, David E. Hudak, Neil Ludban, Ashok Krishnamurthy, Vijay Gadepally, Siddharth Samsi, John Nehrbass, International Journal of Parallel Programming, February 2009, Volume 37, Issue 1, pp 91–105.

[40] Developing a Computational Science IDE for HPC Systems, David E. Hudak, Neil Ludban, Vijay Gadepally, Ashok Krishnamurthy, Third International Workshop on Software Engineering for High Performance Computing Applications (SE-HPC '07), DOI: .

[41] SkIE: A heterogeneous environment for HPC applications, B. Baccia, M. Danelutto, S. Pelagatti, M. Vanneschi, Parallel Computing, **25**, Issues 13–14, December 1999, pp. 1827-1852.

[42] **CCPForge** was the software development repository with extra tools to facilitate collaborative development across the UK's CCP communities. It allowed experts in all strands of computational research to come and work together on common projects and promoted cross-dissemination of best software practices. The Joint Information Systems Committee (JISC) and EPSRC both funded the CCPForge project. The STFC's Scientific Computing Department developed and maintained CCPForge via EPSRCstrategic eInfrastructure fund https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/L000342/1.

[43] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. "Scalable molecular dynamics with NAMD", *Journal of Computational Chemistry*, **26** (2205) 1781-1802.

[44] D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods. "The Amber biomolecular simulation programs." *J. Computat. Chem.* **26** (2005) 1668-1688

[45] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations", *J.Comp.Chem.* **4** (1983) 187.

[46] Lindahl, E., Hess, B. & van der Spoel, "D. GROMACS 3.0: a package for molecular simulation and trajectory analysis", *J Mol Model* **7** (2001) 306–317, DOI: 10.1007/s008940100045

[47] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics" *J Comp Phys*, **117** (1995) 1-19.
[48] A Smooth Particle Mesh Ewald Method, U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Physics*, **103** (1995) 8577-8593.
[49] The design and implementation of FFTW3, M, Frigo, S.G. Johnson, Proceedings of the IEEE. **93** (2): 216–231.
[50] A DAFT DL_POLY distributed memory adaptation of the Smoothed Particle Mesh Ewald method, I.J. Bush, I.T. Todorov, *Comp. Phy. Commun.* **175** (2006) 323.
[51] "Computer Modelling of Fluids Polymers and Solids", edited by CRA Catlow, SC Parker and MP Allen, Kluwer 1990.
[52] CCP5 Information Quarterly No. 43 February 1995, Daresbury Laboratory.
[53] "Research Requirements for High Performance Computing", CR A Catlow (chair) (SERC 1992); "A Strategic Vision for UK e-Infrastructure", D J Tildesley (chair) (Department for Business, Innovation and Science, 2011)
[54]DL_POLY: Application to Molecular Simulation, W. Smith, C.W. Yong and P.M. Rodger, Molecular Simulation **28** (2002) 385
[55]https://www.cecam.org/
[56] DL_POLY_3: the CCP5 national UK code for molecular-dynamics simulations, I.T. Todorov and W. Smith, Phil. Trans. R. Soc. Lond. A, **362**, 1835-1851 (2004)
[57] A DAFT DL_POLY Distributed Memory Adaptation of the Smoothed Particle Mesh Ewald Method, I.J. Bush, I.T. Todorov and W. Smith, Comp. Phys. Commun., **175**, 323-329 (2006)
[58] Application and Uses of CML within the eMinerals project, Toby O.H. White, Peter Murray-Rust, Phil A. Couch, Rik P. Tyer, Richard P. Bruin, Ilian T. Todorov, Dan J. Wilson, Martin T. Dove, Kat F. Austen, Steve C. Parker, Proceedings of the UK e-Science All Hands Meeting (2006), EPSEC Sept 2006 ISBN 1-9553988-0-0, page 606-613, (http://www.allhands.org.uk/2006/proceedings/
[59] Molecular Dynamics in a grid computing environment: experiences using DL_POLY_3 within the eMinerals eScience project, M.T. Dove, L. A. Sullivan, A.M. Walker, R.P. Bruin, T.O.H. White, K. Trachenko, P. Murray-Rust, I.T. Todorov, R.P. Tyer, P.A. Couch, K. Kleese van Dam and W. Smith, Mol. Sim., **32**, 945-952 (2006)
[60] The origin of the compressibility anomaly in amorphous silica: a molecular dynamics study, Andrew M Walker, Lucy A Sullivan, Kostya Trachenko, Richard P Bruin, Toby OH White, Martin T Dove, Richard P Tyer, Ilian T Todorov and Stephen A Wells, J. Phys.: Condens. Matter **19**, 275210-275219 (2007)
[61] Use of Massively Parallel Molecular Dynamics Simulations for Radiation Damage in Pyrochlores, Ilian T. Todorov, Neil L. Allan, John A. Purton, Martin T. Dove and W. Smith, J. Mater. Sc., **42** (6)**,** 1920-1930 (2007)
[62] DL POLY 3 I/O: Analysis, Alternatives and Future Strategies, I.T. Todorov, I.J. Bush, A.R. Porter, Proceedings of Performance Scientific Computing (International Networking for Young Scientists) (February 2008, Lithuania),
[63] R. Čiegis, D. Henty, B. Kågström, J. Žilinskas (Eds.)(2009) Parallel Scientific Computing and Optimization. Springer Optimization and Its Applications, ISSN 1931-6828, Vol. 27, Springer, ISBN 978-0-387-09706-0, doi:10.1007/978-0-387-09707-7
[64] Optimisation of the I/O for Distributed Data Molecular Dynamics Applications, Ian Bush, Ilian Todorov and William Smith, CUG2010 Meeting proceedings
[65] Rigid Body Molecular Dynamics within the Domain Decomposition Framework of DL_POLY_4, Ilian Todorov, Laurence Ellison, and William Smith, Parallel Computing

Technologies: Lecture Notes in Computer Science (V. Malyshkin (Ed.)), **7979**, pp 429-435, doi:10.1007/978-3-642-39958-9_40, Springer-Verlag Berlin Heidelberg 2013

[66] Electronic effects in high-energy radiation damage in iron, E. Zarkadoula, S.L. Daraszewicz, D.M. Duffy, M.A. Seaton, I.T. Todorov, K. Nordlund, M.T. Dove and K. Trachenko, J. Phys.: Condens. Matter **26** (2014) 085401 (8pp), doi:10.1088/0953-8984/26/8/08540

[67] Efficient Domain Decomposition of Dissipative Particle Dynamics via Choice of Pseudorandom Number Generator, Michael Seaton, Ilian Todorov and Yaser Afshar, Parallel Computing Technologies: Lecture Notes in Computer Science (V. Malyshkin (Ed.)), **7979**, pp 250-257, doi:10.1007/978-3-642-39958-9_23, Springer-Verlag Berlin Heidelberg 2013

[68] Introduction to Modelling, Scalability and Workflows with DL POLY, Ilian T. Todorov, Laurence J. Ellison, Michael A. Seaton, William Smith, "Multiscale Modelling Methods for Applications in Material Science" CECAM Lecture Notes, IAS Series **19** (2013), FZ Julich, ISBN: 978-3-89336-899-0

[69] DL_MULTI—A molecular dynamics program to use distributed multipole electrostatic models to simulate the dynamics of organic crystals, M. Leslie, Molecular Physics, Volume 106, 2008 - Issue 12-13, Pages 1567-1578, 2008, doi:Published online: 01 Dec 2010

[70] Arbitrary order permanent Cartesian multipolar electrostatic interactions, H. A. Boateng and I. T. Todorov, The Journal of Chemical Physics **142**, (2015) 034117; doi: 10.1063/1.4905952

[71] https://researchportal.bath.ac.uk/en/projects/energy-materials-computational-solutions

[72] G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi. PLUMED2: New feathers for an old bird, Comp. Phys. Comm. 185, 604 (2014), preprint available as arXiv:1310.0980

[73] https://openkim.org/

[74] Applications Of Dl_poly And Dl_multi To Organic Molecular Crystals, S. L. Price ,S. Hamad,A. Torrisi,P. G. Karamertzanis,M. Leslie &C. R. A. Catlow, Molecular Simulation, Volume 32, 2006 - Issue 12-13, Pages 985-997, 2006, doi: 10.1080/08927020600880810

Appendix A Briefing Note for Referees of CCP Proposals

The following note was drafted in 2000 for the EPSRC to send to peer reviewers of CCP grant proposals. This was done for some years following, and did help referees to distinguish CCP proposals from standard research grant proposals. In spite of this, misunderstandings were not entirely eliminated.

"This note gives some information on the general characteristics and ways of working of the CCPs. It will be seen that these Projects are rather different from most proposals made to responsive grant lines, and referees are asked to bear this in mind when assessing them.

[There followed an outline of the CCP Standard Model, described in section 2, which we do not repeat here.]

The flagship' projects represent innovative and challenging software developments at the leading edge of the relevant area of science or engineering. They are normally undertaken for a period of around 3 years and funded via a research grant which may support a PDRA associated with the project. After the completion of a flagship project, the resulting software will normally become part of the code library, and the CCP will maintain, distribute and develop it according to the on-going demand within the research programmes of members and users. However, for several projects, especially those most closely aligned to experimental research (eg CCP4, CCP11, CCP13, CCP14), greater emphasis is given to the collation, standardisation, maintenance and distribution of existing data analysis codes.

The foregoing activities are funded by research grants. CCPs may also benefit from support by staff at CLRC's Daresbury Laboratory, funded via a Service Level Agreement with the EPSRC. Such staff provide expert technical and administrative back-up, perform many of the functions outlined above, and are frequently involved in the flagship projects in a supervisory role.

Thus, the activities of the CCPs contain a strong element of support for a wide and coordinated computational community. Emphasis is given to the quality, utility, longevity and exploitation of codes generated in flagship projects. It is these infrastructural characteristics which distinguish CCP proposals from most responsive mode research proposals on specific scientific problems – rather, a CCP aims to enable and underpin high quality UK computational research in its field. [Our italics]"