Goldbeck Consulting Ltd
St John's Innovation Centre
Cowley Road
Cambridge CB4 0WS
United Kingdom

Goldbeck Consulting Ltd

The economic impact of molecular modelling

Impact of the field on research, industry and economic development.

Gerhard Goldbeck
5/1/2012

Contents

Executive Summary

The evidence for economic impact of molecular modelling of chemicals and materials is investigated, including the mechanisms by which impact is achieved and how it is measured.

Broadly following a model of transmission from the research base via industry to the consumer [1], the impact of modelling can be traced from the (a) authors of theories and models via (b) the users of modelling in science and engineering, and (c) the research and development staff that utilise the information in the development of new products that benefit (d) society at large.

The question is addressed whether and in what fields molecular modelling is applicable and where it is accepted as a mainstream tool that is useful, practical and accessible [2]. A number of technology trends have increased applicability and acceptance in recent years, including

- Much increased capabilities of hardware and software.
- A convergence of actual technology scales with the scales that can be simulated by molecular modelling as a result of nanotechnology.
- Improved know-how and a stronger focus in industry on cases where molecular simulation really works.

The acceptance level still varies depending on method and application area, with quantum chemistry methods having the highest level of acceptance, and fields with a strong overlap of requirements and capabilities of these methods such as electronics and catalysis reporting strong impact both anecdotally and as measured by the size of the modelling community and the number of patents. The picture is somewhat more mixed in areas such as polymers and chemical engineering that rely more heavily on classical and mesoscale simulation methods. One recent industry survey in the field of thermodynamics even concludes that "despite the academic success of molecular simulation techniques, the survey does not indicate great interest in it or its future development". On the other hand there are also reports of growing confidence that the applicability and impact of the technology is going to increase steadily.

A more quantitative approach to evaluating impact is attempted by considering available evidence throughout the expanding circles of influence from the model author to the end product consumer.

As indicators of the research base and its ability to transfer knowledge data about the number of publications, their growth and impact relative to other fields are discussed. Patents and the communities of users as well as interested 'consumers' of modelling results, as well as the size and growth of the software industry provide an insight into the transmission of impact further into industry and product development. The return on investment due to industrial R&D process improvements are a measure of the contribution to value creation, and justify determining the macroeconomic impact of modelling as a proportion of the impact of related disciplines such as chemistry and high performance computing. Finally the integration of molecular modelling with workflows for engineered and formulated products provides for a direct link to the end consumer.

Key evidence gathered in these areas includes:

- The number of publications in modelling and simulation has been growing more strongly than the science average and has a citation impact considerably above the average.

- There is preliminary evidence for a strong rise in the number of patents, also as a proportion to the number of patents with the respective fields.
- The number of people involved with modelling has been growing steadily for more than a decade. A large user community has developed which is different from the original developer community, and there are more people in managerial and director positions with a background in modelling.
- The software industry has emerged from a 'hype cycle' into a phase of sustained growth.
- There is solid evidence for R&D process improvements that can be achieved by using modelling, with a return of investment in the range of 3:1 to 9:1.
- On the basis of the above evidence of the strong contribution of modelling to academic research and to value creation in industry, the macroeconomic impact has been estimated on the basis of data for the contribution of chemistry research to the UK economy. The preliminary figures suggest a value add equivalent to 1% of GDP.
- The integration with engineering workflows is discussed in the context of studies on "Integrated Computational Materials Engineering". It is found that molecular modelling forms a small but very important part of workflows that have produced very considerable returns on investment.
- E-infrastructures such as high-throughput computing, materials informatics systems and high performance computing act as multipliers of impact. Molecular modelling is estimated to account for about 6% of the impact generated from high performance computing.

Finally, a number of existing barriers to impact are discussed including deficiencies in some of the methods that are being addressed, software interoperability usability and integration issues, the need for databases and informatics tools as well as further education and training.

Some reservations and shortcomings notwithstanding, this review found strong and in a number of cases even quantifiable evidence for the transmission of impact of modelling from the research based to economic benefits.

Introduction

The proverbial "afternoon in the library" that can be saved by "six months in the lab" has long been replaced by "an afternoon on the computer". This widely held view is based on many success stories of computer simulations and information systems, which reflect the inexorable rise of computing power and pervasiveness across many different sectors. According to a report by IDC [3], the top industry segments that purchased High Performance Computing (HPC) resources in 2009 included bio-sciences, computer-aided engineering, electronic design analysis, geo-sciences/engineering, weather and chemical engineering.

In engineering, computer simulations have become indispensable. A communication from the European Commission on ICT infrastructures for e-science [4] states that the production of complex artefacts such as aircraft, cars or personal appliances relies on complex modelling and simulation, and the cooperation of researchers and engineers. IDC [3] reports that from its roots in government and academic research, HPC-based modelling and simulation spread out into large industry starting in the late 1970s. Since then, HPC has enabled automakers around the world to reduce the time for developing new vehicle platforms from an average 60 months to 24 months.

Simulations based on molecular models first made an impact in the life science sector, and an associated software industry developed from the late 1970s. Today, molecular modelling is an essential part of the pharmaceutical discovery workflow. Outside of the life sciences sector, i.e. in applications such as chemical engineering, materials science and electronics, molecular modelling software and industrial applications took off in the 1990s and have since gained recognition for enabling key new insights into the behaviour of chemicals and materials. It has been adopted by many large companies in the chemicals, electronics, automotive and personal care sectors. Nevertheless, the question remains whether molecular modelling has become a mainstream tool in the way that process simulations or finite element modelling are [2].

While there are numerous studies that discuss potential and requirements for future development, much less attention has been given to reviewing the impact that has actually been obtained by modelling and simulation applications, and the mechanisms by which these benefits have been derived. This would seem especially important in the light of reservation and scepticism regarding the impact of molecular modelling. For example, a recent survey report on the "Industrial Requirements for Thermodynamics and Transport Properties" [5] found that "despite the academic success of molecular simulation techniques, the survey does not indicate great interest in it or its future development".

The aim of this report is to present current evidence about the impact of modelling, the mechanisms by which impact has been achieved as well as any metrics of impact and return on investment (ROI). Its findings are based on a review of publications, surveys and white papers, available statistics of publications and patents, as well as interviews with some key stakeholders.

Economic impact

Economic impact was traditionally associated with measures of economic growth and associated changes in business output, employment, income and wealth. Over time the definition broadened however to also include social and environment impacts and quality of life factors. The issue of

economic impact became central to the agenda of research funding in the United Kingdom following a report by the Research Council Economic Impact Group, chaired by Peter Warry, to the Director General of Science and innovation at the Department of Trade and Industry [1]. The report describes impact in terms of the aggregate improvements in welfare and enhanced economic growth. In particular "an action or activity has an economic impact when it affects the welfare of consumers, the profits of firms and/or the revenue of government. Economic impacts range from those that are readily quantifiable, in terms of greater wealth, cheaper prices and more revenue, to those less easily quantifiable, such as effects on the environment, public health and quality of life."

While it is difficult if not impossible to determine the impact of a research method such as molecular modelling on economic output, there may still be indications of the transfer of knowledge and benefits of utilising modelling in the development of products. In this context, the schematic shown in Figure 1 of the transmission mechanism of research base benefits to economic benefits from the Warry report is helpful. The current report is going to review evidence and indicators of these mechanisms and their impact from a range of sources.
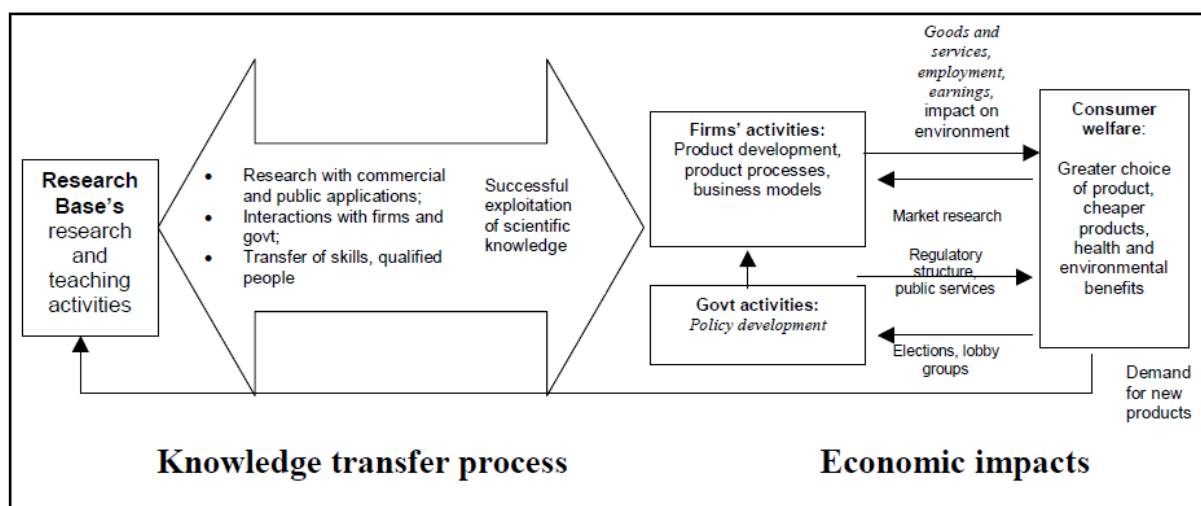

Figure 1: Transmission mechanism of research base benefits to economic benefits. From: Appendix A of [1].

Molecular modelling

In order to clarify the meaning of molecular modelling and simulation, a definition given by Maginn [2] is useful: "In the broadest sense, molecular modelling and simulation can be defined as the use of computational methods to describe the behaviour of matter at the atomistic or molecular level. There is a clear distinction between this and the familiar continuum-based modelling, in which atomic-level phenomena are neglected."

Within the area of molecular modelling one can broadly distinguish quantum mechanics based methods, classical simulations methods, and coarse-grained or mesoscale methods, as shown on the classic length and time scale scheme in Figure 2. A key concept arising from this view is that of multiscale modelling, i.e. many techniques are needed to cover the scales from the atomistic to the engineering scale, see e.g. Gubbins and Moore [6] for a discussion of the methods and their potential impact.
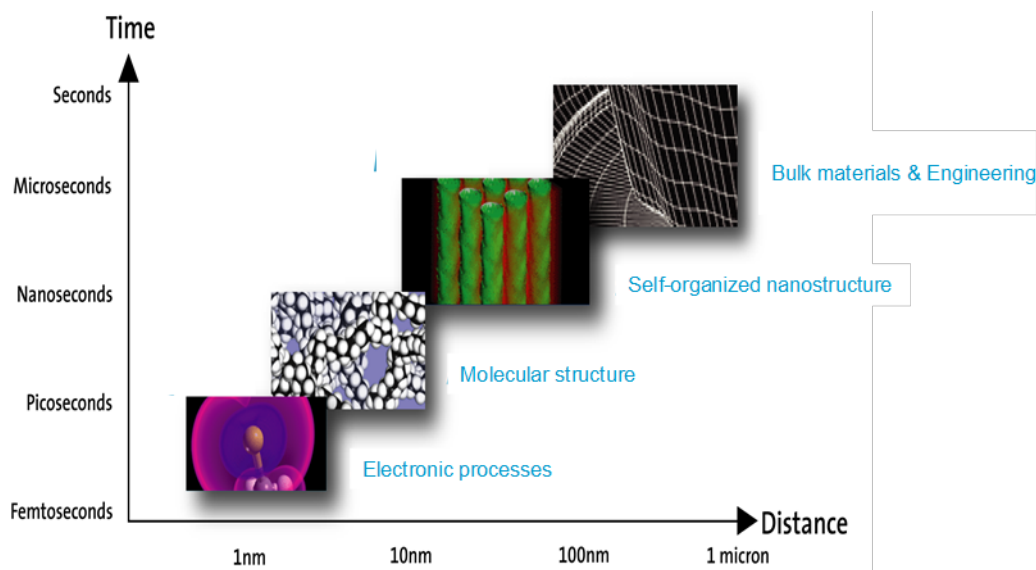
Figure 2: Multiscale modelling schematic. Courtesy of Accelrys.

Apart from the methods themselves, it is also useful to distinguish [2] between the 'discovery' and the 'data' mode. According to Maginn [2], the discovery mode consists of studies in which new phenomena are predicted that have not yet been observed experimentally, or explanations are sought for known phenomena that are not understood. Examples in the context of economic benefit were discussed in the report on the economic benefits of chemistry research to the UK [7]. They include the aerospace industry which relies on computational chemistry to better understand combustion and the impact of elevated temperatures on the stability of various components e.g. metal oxide surface coatings and catalysts. Another example given is nuclear power, where calculations of the atomic displacement caused by neutron impacts in the moderator material helped fill a major gap in the theory of radiation damage, which is crucial to determining the life expectancy of the moderator, a main determinant for reactor longevity. The potential economic impact is calculated by considering the potential costs of closing reactors unnecessarily early, which for the UK alone translates into losses running into billions of pounds, quite apart from potential issues arising from emission targets and energy supply shortages.

Data-driven simulations on the other hand consist of calculations where accurate property predictions are made with little or no input from experiment. Simulations can be used to interpolate between experimental data, extrapolate outside the range where data are available, or predict properties for compounds for which little or no data are available. Data-driven quantum chemical calculations in particular have an established return on investment in industry. The perhaps most widely quoted example is that measuring the heat of formation for a single molecule cost $100,000 in 2000, about 50 times the cost of an accurate quantum chemical calculation [2]. Data driven simulations also include the analytical simulations of specific experimental techniques such as x-ray diffraction, IR, Raman and NMR spectroscopy, enabling the interpretation and refinement of experimental data. In pharmaceutical crystallization for example, powder diffraction simulations are routinely used to help determine crystal structures by matching and refining models to experimental data.

Applicability and Acceptance

Molecular modelling can in principle impact on any sector in which product performance and innovation are based on controlling chemistry and the electronic and physical properties of materials. In practice, however, the importance of the atomistic and molecular scale and also the applicability and acceptance of different modelling techniques vary considerably. Maginn [2] asks the pertinent question whether molecular modelling has become a mainstream tool, defined as a method that is ''useful, practical, and accessible'' to a wide range of researchers, and finds that it depends on the method and application.

The most established methods are based on quantum chemistry and quantum physics, in particular density functional theory (DFT). They have achieved a very high level of acceptance as a result of the accuracy with which some key properties can be calculated and the strongly decreasing time and cost of such calculations over the last 15 years. Today, many useful materials properties can be computed with sufficient accuracy using well established DFT methods [8], including elastic constants, phonon dispersions and the related thermodynamic functions such as heat capacity, temperature dependent enthalpy and entropy, the coefficient of thermal expansion and diffusion coefficients. At the same time the cost of experiments typically has gone up, and the competitive advantage of exploring a wider range of systems can be achieved much more cost effectively by combining experiment and simulation. Also, the downscaling of CMOS technology for example has reached the scale at which it is accessible by ab initio methods. The structure and properties of new materials such as high-k dielectrics can be now determined quantitatively and not just qualitatively. As a result, atomistic modelling has become part of "the team", integrated with experimentation and engineering, as publications, patents and recruitments show [9][10][11]. Similarly, in heterogeneous catalysis the detail and accuracy with which reactions at surfaces can be calculated compares favourably with experiments [12].

However, despite developments such as linear scaling DFT [13] there are limitations to these methods in terms of length scale, time scale, and their ability to handle cases in which a large configuration space is a key factor. This includes many industrially important materials such as fluids and polymers, foods, paints, home and personal care formulation as well as alloys and ceramics.

Classical and mesoscale methods that are designed to deal with these issues, on the other hand, are still much less accepted. This is largely due to the fact that they have not yet reached the level of accuracy, applicability and validation of quantum chemistry methods [2]. This view was also expressed by Robert Meier from DSM [14]: "What can be achieved in solid state physics, e.g. semi-conductor modelling, is not necessarily comparable with modelling chemistry and polymers at the level of the accuracy and reliability of the computed data and models. First principles approaches are still impractical for application studies in (polymer) industry where predictive power is more crucial than understanding."

In 2001 an initiative was set up focussing on prediction of a variety of physical properties of great significance to the chemical industry [15]. The Industrial Fluid Properties Simulation Challenge [16] intended to drive improvements in the practice of classical molecular modelling, formalize methods for the evaluation and validation of simulation results with experimental data, and ensure the relevance of the academic community's simulation activities to industrial needs and requirements. The challenges have highlighted the promise and usefulness of molecular simulation for accurate

physical property prediction while also illustrating some of its limitations [17] regarding the relevance of simulation for property prediction. For example, despite the increase in computational power and algorithmic efficiency, fluid phase simulations of moderately-complex chemical species still require large computational resources and significant time investment in many cases. Also, the level of accuracy that can be expected for a given simulation is often not well understood, limiting the usefulness in many industrial applications.

On the other hand it has also been acknowledged [14] that rather than a 'brute force approach', a combination of clever selection of experimental data and a range of computational tools can make a very useful contribution to the design of a polymers, soft materials and chemical processes. The impact of such strategies is documented in examples from different industries. At Procter and Gamble [18] researchers determine how nanoscale structures impact the characteristics of the ingredients in their soaps, detergents, lotions and shampoos,, aiding the development of new products that meet tougher environmental and sustainability goals while retaining top performance. In the aerospace industry, Boeing has integrated molecular simulations into the materials design process [19]. For example, an atomistic level method to model thermoset polymer resins has been develop that allows for the quick screening of new resins and the exploration of structure/property relationships. The success of the approach led to the conclusion that "the future of aerospace materials development includes simulation tools."

In conclusion, acceptance of molecular modelling across a range of industrially relevant applications has reached a stage at which impact is being realised and confidence in the applicability is rising. Key factors for this development include [9]:

- Capabilities of hardware and software have reached a stage where cases with significant impact are being studied by a wider range of researchers.
- Some methods, in particular DFT, have reached a high level of acceptance due to the level of accuracy and reliability in data driven simulation.
- There is better know-how and a stronger focus in industry on cases where molecular simulation really works well, e.g. optic/magnetic properties for cases where the structure is well defined, i.e. avoiding cases where the configurational space is large.
- Strategies for applying molecular modelling in an impactful way in cases where the accuracy is not as high (e.g. in polymer/mesoscale modelling) have been developed.
- Nanotechnology has brought about a convergence of actual technology scales with the scales that can be simulated by molecular modelling [20].
- The property requirements on new materials and chemicals more often make atomic and nanoscale design a necessity.
- There is a new generation of technology managers with a deeper appreciation and more realistic judgement on what modelling can deliver. This is in contrast to the first "hype" phase of 10-20 years ago, when large, dedicated but often isolated computational groups were formed. Expectations were raised too high, which was followed by disappointment. Modelling is now part of "the team", integrated with experimentation and engineering.
- Experimentation that provides detailed insights is expensive. There is also some disillusionment with high throughput experimentation techniques that have delivered challenges due to the large amounts of data, but very little understanding.

Together, these developments provide the foundation for a deeper and stronger integration of molecular modelling into the R&D and even engineering workflow, as will be discussed later on.

Measuring impact

There is a widely shared view within the research community that modelling and simulation has a strong and growing impact [21]. A recent survey from the European FP7 project MULT-EU-SIM22 was reported [21] to find that 75 % of researchers see a high impact of modelling and simulation in their fields, and 70 % foresee a strong growth of these methods, with an impact far beyond the one currently achieved.

The question is whether and how these views can be substantiated and the economic impact measured. The rate of return to research and development spending in general has been investigated using a range of econometric models [22]. The review by Hall et al concludes that there are strongly positive returns to R&D investments by companies while the social returns are even higher, although variable and imprecisely measured in many cases. However, Hall's attempts to quantify the return on investment for research by the chemical industry [23] showed that the models used for calculating productivity from R&D don't work very well. Contributing to these challenges are a range of factors including the question how one depreciates knowledge, the heterogeneity of the firms doing R&D as well as the different types of R&D. Nevertheless it could be useful to consider in what way modelling affects some of the factors entering into the econometric models, for example the speed with which knowledge capital can be gained and is depreciated. Also, one could try to estimate the effect on certain types of R&D and their associated risks. For example, a higher return is generally reported on basic R&D as opposed to applied R&D. However, this is also associated with a higher risk factor of the long term commitment that this type of R&D requires. This finding would suggest that if a method such modelling reduces the risks and costs of early stage R&D it is expected to have significant returns.

Since the Warry report [1] impact has become a strategic goal of research councils such as the EPSRC in the UK. Professional bodies such as the Institute of Physics [24] and the Royal Society of Chemistry [7] commissioned reports on the impact of their subject areas on the UK economy. Similarly, there are a range of studies from the market research organisation IDC on the impact of high performance computing [25][26][3]. Since molecular modelling is a sub-discipline of these fields it may be possible to attribute part of the impact. These rather macro-economic measures of impact will be discussed later on.

There are also other, more specific indicators that together provide a useful picture of the impact of molecular modelling. These include the development and status of modelling relative to other areas of research and research tools, including publications and patents, the number of users and consumers of the methods, and the status of the related software industry. Also, there are some studies that estimated the return of investment of using modelling and simulation tools in industrial R&D processes [27][28].

Based on the concept of the transmission mechanisms in Figure 1, one can think of molecular modelling as an expanding circle of influence (Figure 3) starting from fundamental theories (such as the Kohn-Sham theory that was awarded the Nobel prize) which lead to models and software that enable the transition to a wider user community. The latter generates results that are of interest to

an even wider circle, which will be called the modelling 'consumers'[29]. Ultimately the outcomes support the development of products that benefit society at large.
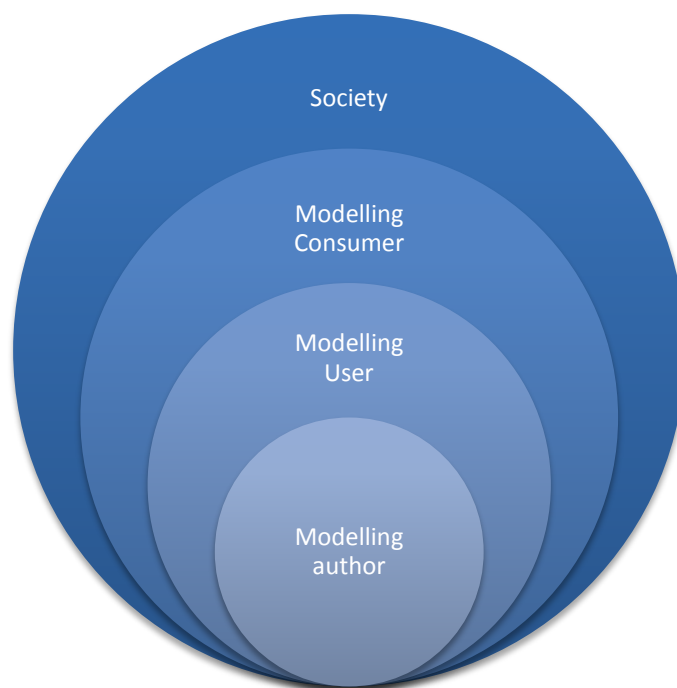
Figure 3: Expanding circles of influence of modelling, from the original author to the resulting products that impact society at large

Current impact of a technology such as molecular modelling can therefore in principle be determined by the size of these circles, i.e. how strong and beneficial the impact is within each of the communities, from science to society, and how far impact of the method is felt out. For example,

- What is the size of the modelling user community, and its impact on research processes and efficiency?
- Is there a significant community of 'modelling consumers' i.e. of researchers and engineers that can utilise the outcomes of modelling to enhance their activities, e.g. product design.
- Is there a measurable impact of the molecular modelling at society level? Most products we use in everyday life, from plastic containers to cars and aeroplanes were developed and designed using computational tools. While the impact of molecular modelling is clearly not at the same level, is it possible to determine at least qualitatively?

This report is going to summarize current knowledge about these levels of impact and consider the mechanisms and enabling factors of impact, including (see also Figure 4):

- Publications as a measure of the growth and spread of knowledge, know-how as well as an indicator of the growth and impact of modelling communities.
- People as a measure of the growth of the communities and their influence, as well as the direct economic benefit of their activity, from education to application.
- Patents as a measure of the potential impact on new products.

- Software and its market, as a measure of the growth of the user community as well as its 'customers', i.e. the modelling consumers. Also software as an industry has a direct economic impact.
- Impact on the ability of the R&D process to contribute to key indicators of value creation (R&D processes and performance): Benefits derived in industrial R&D are ultimately measured by value creation and impact on society. Examples include studies by IDC based on interviews of researchers, and a model derived on the basis of the impact of chemistry research on society.
- Impact on engineering and new product design: Considering the transition from fundamental research to engineering, the question is whether molecular modelling plays a role in improving product development at the engineering level.
- Exploiting e-infrastructures developments and their impact on growth and competitiveness: Much of the increased impact of molecular modelling is due to the rapid cost/performance improvements of hardware. Since this mechanism is shared with many other disciplines that benefit from HPC, independent studies on HPC provide a basis for estimating the impact of molecular modelling.
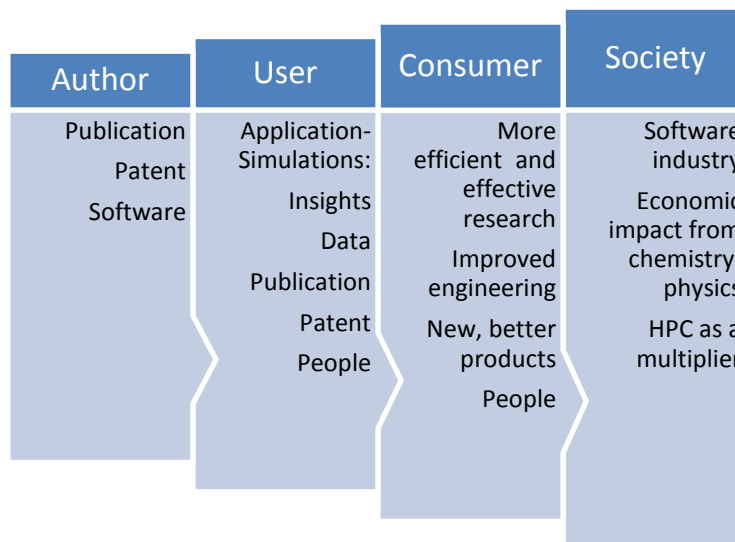

Author	User	Consumer	Society
Publication	Application-Simulations:	More efficient and effective research	Software industry
Patent	Insights	Improved engineering	Economic impact from chemistry, physics
Software	Data	New, better products	HPC as a multiplier
	Publication	People	
	Patent		
	People		

Figure 4: Impact indicators at the different stages of transmission.
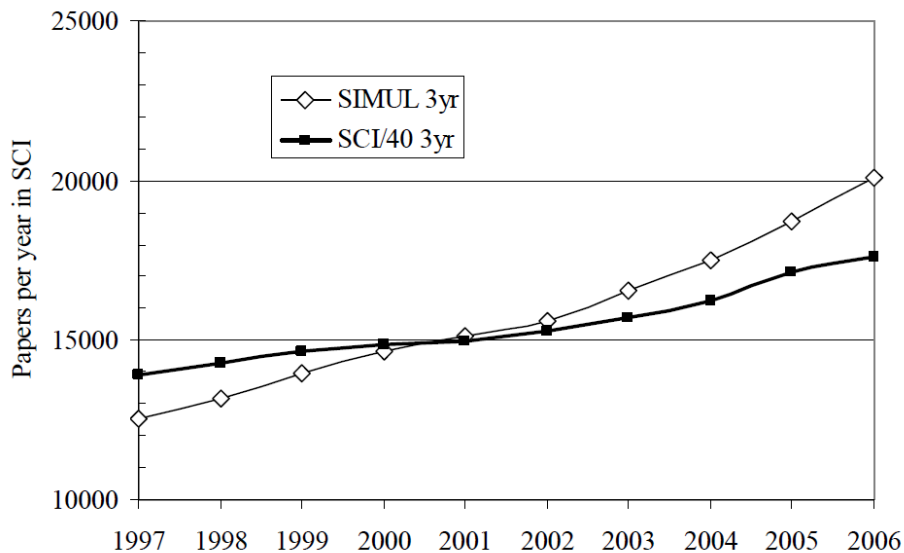
Publications

In the last few decades atomistic and molecular modelling has grown into a major academic activity. Its role and behaviour has changed gradually from being largely re-active to both theory and experiment to a much more pro-active role that is beginning to drive new developments. Simulation science has developed faster than many other areas, and is receiving a lot of attention.

The current section provides evidence of the importance of atomistic and molecular simulation in science as measured by publications, the number of researchers involved and their impact. In addition to their significance for academic research, these factors underpin economic impact resulting from new technology developments as for example measured by patent output, and by contributing experts to key industry sectors such as the chemicals and electronics industries.

The fact that the number of publications using atomistic and molecular modelling methods has been growing strongly can easily be demonstrated using "Google Scholar". Since Density Functional based modelling is arguably the most widely used technique it is a useful search term to provide a rough indication of the approximate number of publications in science and engineering journals growing from 3000 in 1995 to 10,000 in 2001 and 16,000 in 2006.

A more detailed study was carried out as part of the WTEC Panel Report on "International Assessment of Research and Development in Simulation-Based Engineering and Science"[30]. Its Appendix E contains a bibliometric analysis carried out by Grant Lewison of Evaluametrics Ltd. One of the stated aims of the analysis was to determine the outputs of simulation research in terms of both volume and impact [31]. Relevant publications were determined based on the Science Citation Index CD-ROMs and Web of Science. A relatively extensive filtering method was designed and papers were classified by research field.

Note that this study takes a relatively broad view of the field, defined as "the application of mathematical models using a computer to the study of the underlying physical and chemical processes, and prediction of the behaviour and properties of systems, including natural and artificial materials, flow in liquids and gases, energy at all scales including the cellular level, and biomedical sequelae." It will therefore include methods and applications beyond the molecular level, for example simulations based on continuum models. Nevertheless, it provides useful metrics about the field. For example, the study finds that simulation research represents (after correction for the lack of recall of the filter) about 5% of the papers in the SCI, but is growing much faster than science overall, the annual average percentage growth rate being 5.0% compared with 2.5% for all science (see Figure 5).

Figure 5: Number of papers per year (three-year running means) in simulation research (open diamonds) and in all science (solid squares). From [30] with permission from the author.

The total number of simulation publications based on the more comprehensive Web of Science (WoS) data reached 25,000 in 2004 and had grown to 30,000 by 2006. The impact of the publications was estimated based on three different "measures of esteem":

a. Percentage of reviews [32]. It provides a simple measure of research esteem of a field, institution or country relative to others.
b. Potential citation impact (PCI). The potential citation impact of each paper is defined as the five-year mean citation score of papers in the same journal and year.
c. Actual citation impact (ACI). The actual citation impact was determined from the WoS for papers from the leading countries for the year 2004, with citations counted over the four years, 2004-07.

These were proposed as useful indicators that, if they agree, provide confidence that the message they are conveying is reliable.

The finding regarding the reviews was that (up to 2006) simulation has a much smaller percentage (about half) than the mean for all science, and that in contrast to science overall, it had not been growing. This would indicate a relatively low impact into other areas as well as be typical of a field which is still quite young. There are of course no data beyond 2006, and it would be interesting to repeat the study with more recent publication data.

The citation impact factor on the other hand suggests a different picture. The worldwide potential citation impact factor (PCI) was about 7.8, and the actual citation impact was about 4.8, or about 60% of PCI. The study also provides a breakdown by field of research for the PCI, ranging from 4.5 for engineering, 7.5 for physics and 10.5 for chemistry to 13 for biomedical. The analysis by Lewison does not extend to a comparison of citation impact factors of simulation papers in different fields with that of the field average. A private communication from the author maintains that while it is quite likely that simulation papers in both chemistry and physics could be in higher impact journals, one would have to do a study tailored to this question to answer it.

In the absence of such a study, a preliminary figure can be obtained by applying the above 60% ratio to the potential forward impact PCI to obtain a hypothetical ACI of 4.5 for simulation papers in physics and 6.3 in chemistry. These values compare very favourably to the average impact factors of the respective fields. For example, as reported by Althouse et al [33], chemistry has an impact factor of 2.6, and physics 1.9. Hence the citation impact of modelling and simulation papers would seem to be more than twice that of their field average.

This finding is also underlined by the fact that as of April 2010, the ten most cited papers in the entire history of the journals of the American Physical Society all deal with computational electronic structure calculations of condensed matter!

The tremendous growth of publications using the Molecular Dynamics method is illustrated in Figure 6 [2]. The total number of publications in 2006 was about 5000, and had grown at a rate of about 5% over the past decade, consistent with the analysis by Lewison. It also confirms the above average impact of simulation papers. While in 2008 about 0.5% of the more than 1.4 million articles published in Science and Engineering used MD the percentage appearing in a very high impact journal such as Physical Review Letters was 3%.
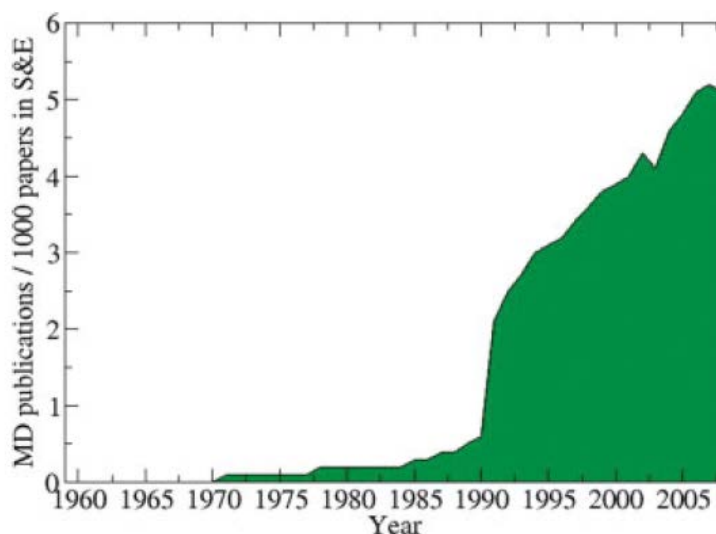
.

Figure 6: Fraction of articles in science and engineering that use molecular dynamics. The data were obtained from a search of WoS records for science and engineering journals. From [2] **with permission from John Wiley and Sons, © 2009 American Institute of Chemical Engineers.**

An bibliometric study of *ab initio* based publications output was carried out by members of the Psi-k network (www.psi-k.org) who reported [34–36] the results of a search of the ISI Web of Science for publications containing the keywords: 'ab initio' or 'first principle' or 'first principles' or 'density functional' in the title, abstract or topic-keywords sections. A recent update to these data is shown in Figure 7, confirming a steady annual increase by about 800 - 1000 to about 17 000 publications worldwide in 2011. This would in fact indicate a growth rate of about 8%, i.e. higher than even the 5% growth reported by Lewison for the simulation field on average.
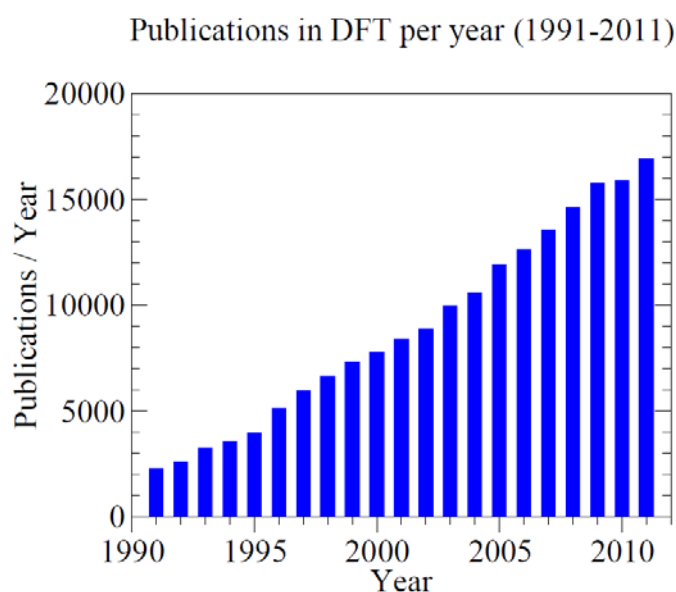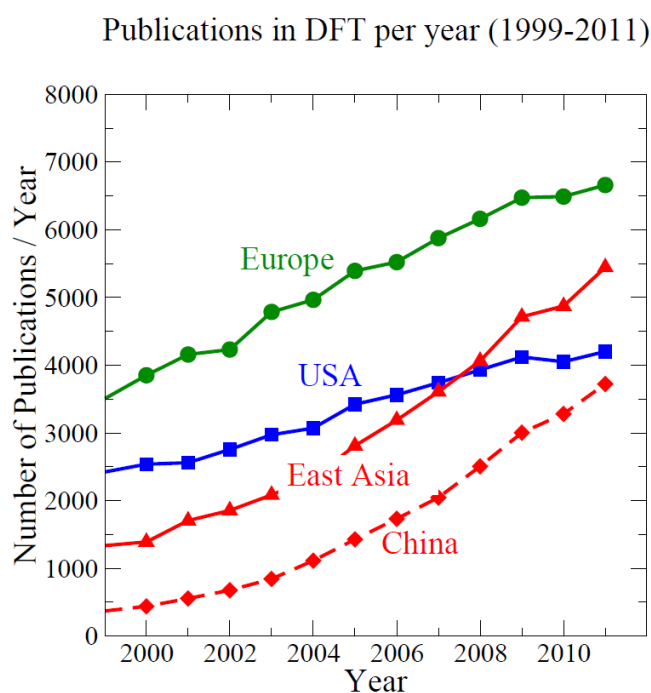

Figure 7: Publications using ab initio methods. Kindly provided by Peter H Dederichs and Phivos Mavropoulos, Forschungszentrum Jülich.

It was further argued that the total number of "ab initio publications" could be more than 20,000. An indication is that including further keywords such as 'local density approximation', already leads to an increase of 6-8 %, and there will also be many papers without obvious keywords in searched fields. As a consistency check, we have the following numbers for 2006:

- Total number of simulation based papers (WTEC report): 30,000 of which about 30-40% are in areas such as continuum engineering modelling, earth and space, clinical research etc, hence about 18-21,000 are likely to be in ab initio, atomistic and molecular modelling.
- Ab initio papers (psi-k study): 15,000
- Molecular Dynamics papers (Maginn): 5000

A regional breakdown of the publication data [35] (Figure 8) shows the strength of Europe in the field and the strong growth of China, which invests strongly in computer modelling as also evidenced by the growing strength of China in the HPC sector, a development which further corroborates the argument that the strong rise of scientific activity in this field is regarded as a key stage to economic impact.

Figure 8: Growth in number of publications using ab initio techniques by region. Kindly provided by Peter H Dederichs and Phivos Mavropoulos, Forschungszentrum Jülich.

In conclusion, we obtain a consistent picture of strong and growing impact of molecular modelling within science. The field is growing at rate significantly above average. It accounted for about 2% of science and engineering publications in 2006, but with a citation impact far above the field average.

Patents

While publications are an established measure of research output, their citation impact factors provide only a limited measure of their impact to the wider economy. Patents on the other hand relate directly to economic impact and wealth creation, and are one of the factors used to determine the return-on-investment of industrial research. Some companies measure the 'profitability' of their research by the returns in terms of patents as well as government co-sponsored collaborations etc [10].

While there is no published study of patent output in the modelling field, a simple Patentscope (http://patentscope.wipo.int) search for patents mentioning "Density Functional" shows an increase from about 30 a decade ago to about 80 in the late 2000s and 150 in 2011 (Figure 9). Of course it remains to be seen whether the 2011 number was a one-off. If this trend is confirmed, it would be strong an indicator of transmission across the expanding circles of impact. While publications started growing strongly during the 1990s (Figure 7) patents followed with a strong growth during the 2000s, as indicated for example by the ratio of "density functional" patents to publications, which increased from about 0.4% in 2002 to 1% in 2011.
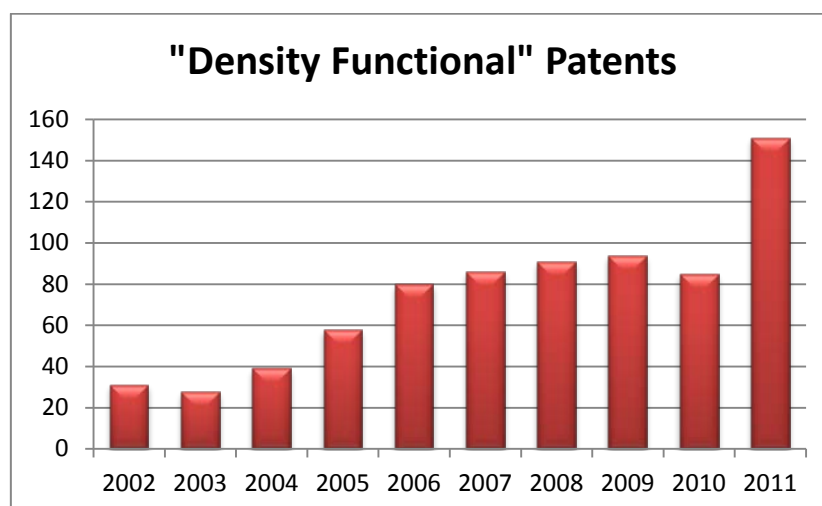

Figure 9: Results of Patentscope search for patents that include the term "density functional" in any field.

There are indications that this trend will continue and in fact accelerate. For example, according to Gerbrand Ceder from MIT [37], who develops rapid computational search and exploration technology, the company Pellion Technologies has patented more insertion cathodes for magnesium batteries in the last 18 months than had been invented in the last 25 years.

It is worth noting which fields the majority of these patents are in (Table 1). The single most important area is electronics (including semiconductors and other electric elements), following by organic chemistry and other areas of chemistry such as dyes and adhesives.

Another indicator is the relation of the number of these patents to those in a field as a whole. Preliminary data based on Patentscope searches show an increase from about 0.1% to about 0.5% over the last ten years.

Table 1: Fields with the largest numbers of density functional related patents published during 2002-2011. IPC is the International Patent Classification identifier.

IPC	Description	Percentage
H01	Basic electric elements	26%
C07	Organic chemistry	21%
C …	Other areas of chemistry	24%
B01	Physical or chemical processes	10%
A	Health, Medical	6%
G06	Computing, Calculating, Counting	5%

Possible reasons for the relatively large share of electronics related patents are that

- Electronic structure simulations are highly relevant to the field.
- Patenting is a key mechanism used in the industry not only to secure IP, but also as a trading commodity between companies.
- There are examples of start-up companies that contributed a significant number to the patents. For example, of a total of 128 patents in IPC H01L (Semiconductor devices) nearly a third were from the company Mears (http://www.mearstechnologies.com/).

In addition to these figures, some specific cases below provide an illustration of the role modelling and simulation plays.

A patent by Bayer Materials Science [38] describes a selection method for additives in photopolymer formulations for producing holographic media. Polymer based holographic materials can be used as relatively low cost, very high density optical storage media, and if manufactured as films, they can be used to add holograms to ID cards and other authentication tags to make them extremely secure at relatively low cost.

The method uses a series of established molecular modelling techniques in order to calculate two physical properties that, according to the invention, can be used as a selection criterion for suitable additives. The properties are the refractive index and the volatility of the compounds. Calculations are compared with a series of experimental data to demonstrate the selectivity of the computational method. A broad claim is therefore established that any compound determined to be suitable according to this model will be covered by the patent.

A similar approach to patents was also described in an interview with Dr Mike Makowski from PPG [39], who maintained that modelling results were used for supporting patent applications based on first principles that have allowed a broader coverage of claims.

A recent example from the electronics sector demonstrates that first principles modelling now also make a direct impact on claims about manufacturing methods. The patent [40] is about a semiconductor manufacturing method including a silicide layer which reduces interface resistance. A key aspect of the claim includes the formation of an impurity region, and a particular distribution of impurities. The link between the process method and the distribution is examined by using first-principles calculation method. Formation energies for all relevant impurities introduced or substituted in the various crystal structures are calculated. The results can indeed rationalise the observations and reduced resistance and hence underpin the claim.

In conclusion, simple patent searches indicate a strong growth in molecular modelling based patents during the last ten years. Specific cases demonstrate impact mechanisms for very different applications and methods. It would hence be worthwhile to carry out a more in depths analysis of patents in the field and their impact.

People

Given the academic strength and growth of the field itself, there should be significant growth in economic impact due to the number of people active in the field, be it as authors, users or consumers of modelling technology and outcomes. Interviews with representatives of modelling software companies [29][9] indicate that there has been strong growth in the number of people trained in modelling and simulation, as well as the number of computational materials and molecular modelling positions in academia.

There is also some quantitative evidence for the increase in the community of researchers involved in modelling, whether as authors, users or 'consumers' of molecular modelling. Based on the Web of Science data on the number of publication in the field of *ab initio* simulations, the number of authors of papers in the year 2008 was determined by Dederichs et al at Research Center Jülich [34]. Here 'author' was defined as a person with a given surname and first initial, and multiple entries with that combination were counted only once. Since there could be more than one person with the same surname and initial, and of course papers not captured by the keywords, the figure is a lower bound of the actual number of authors. On the other hand all co-authors on a paper are counted. The results are that in 2008 there were more than 22,000 authors on ab initio simulation papers, of which 11,000 were in the European Union: 11 000, 5 600 in the USA, and 5700 in East Asia (including China, Japan, Korea, Taiwan and Singapore). To put the figure into perspective, the total number of chemists in the US is estimated to be about 80,000 according to a Wolfram Alpha search.

Not surprisingly, the number of academic positions in the field is also growing. It has been suggested [9] that the number of positions advertised seems to keep rising, an observation which is corroborated by counting the jobs advertised on Europe's largest network, psi-k (Figure 10).
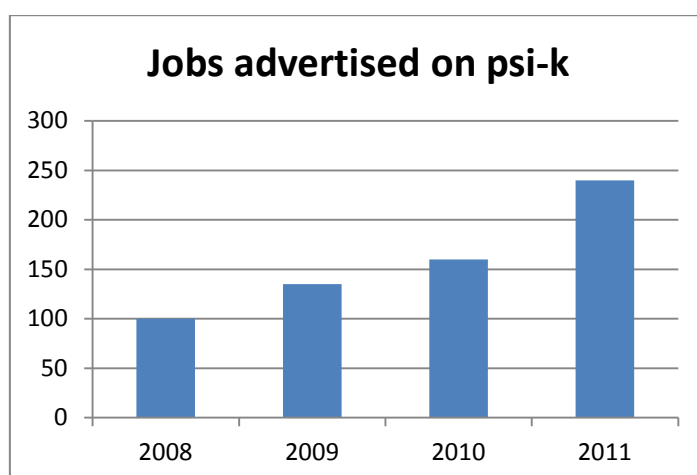

Figure 10: Jobs advertised on psi-K network

Further evidence of the impact beyond the field is appointments of "modelling academics" to director positions of renowned institutions. Examples include Prof Peter Gumbsch, Director of Fraunhofer Institute for Mechanics of Materials IWM and recipient of the Leibnitz prize, Germany's

highest science award and the recent appointment of Prof Claudia Felsner as Director of the Max Planck Institute for Chemical Physics of Solids. Both appointments are to positions that would traditionally be held by 'experimentalists'.

A key mechanism that has led to this impact on an increasing number and range of people has been the increasing accessibility of tools to general users rather than just modelling authors and experts. The development of commercial software tools contributed since the 1990s. More recently, much improved open source/open access tools have led to what can only be described as an explosion of a user community, different from the original developer community. An organisation which made a huge difference in this regard is nanoHUB (www.nanohub.org)[41]. NanoHUB is an online resource for nanoscience and nanotechnology which offers a wide range of simulation tools as well as teaching and education resources around these technologies. The rapid growth of the user community resulting from the adoption of on-line resources in 2005 is shown in Figure 11.
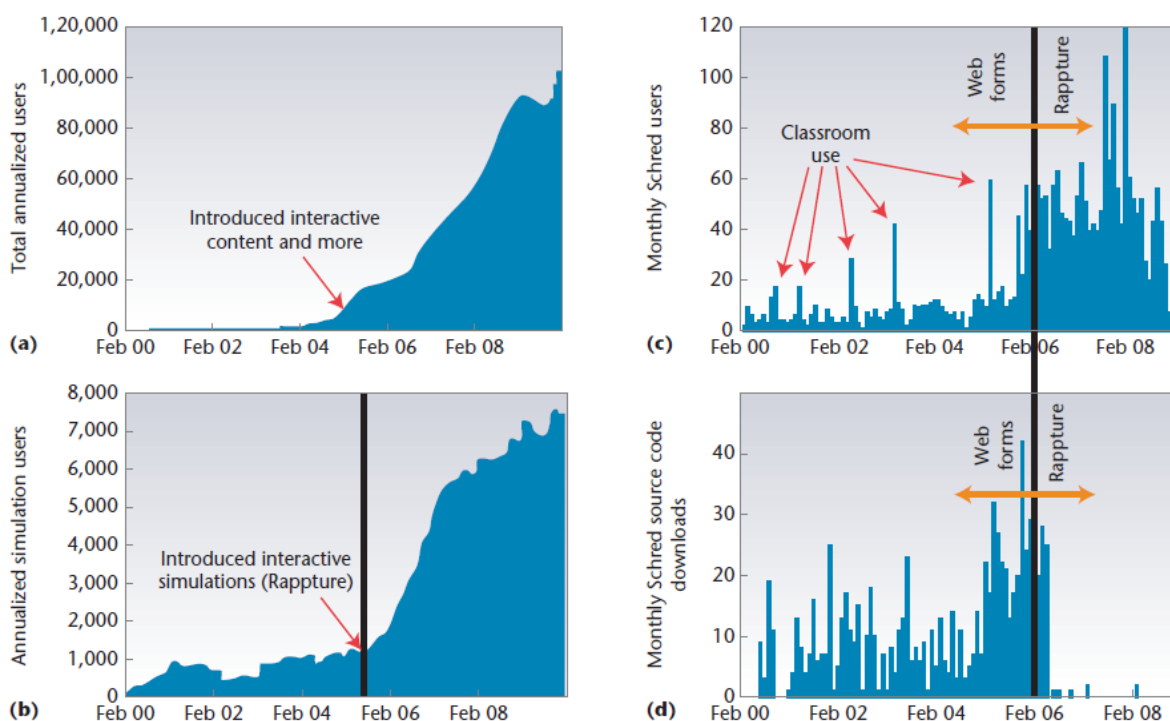
Figure 11: The impact of introducing new concepts, such as nanoHUB online simulation, and user-friendly GUIs to replace Web forms on nanoHUB.org. (a) Total number of annual users over time. This includes simulation and other users. Expanding nanoHUB beyond online simulation via dissemination of interactive research seminars, full classes, and tutorials increased the total number of users to more than 100,000. (b) Total annualized growth in simulation users. As these graphs show, the significant increase in simulation tool users beginning in June 2005 coincided with the deployment of interactive tools with friendly GUIs to replace the traditional Web forms. The introduction of a GUI for the popular Schred tool exemplifies this trend, showing (c) a dramatic increase in users per month and (d) an equally dramatic decrease in source code downloads. Overall, nanoHUB.org user numbers increased by factors of four to five times their pre-GUI levels, while source code downloads all but vanished. Reproduced from [41], with permission © IEEE 2010.

There are also impressive statistics which demonstrate the transmission of impact onto a much larger population, i.e. the 'consumers' of modelling. According to nanoHUB director Gerhard Klimeck [42], the simulation field even ten years ago was largely limited by the fact that one had to be a simulation code developer to be a user, at least in the nano-electronics field which the nanoHUB largely represents. Making tools available to a larger community started to change that,

with a step change in usage resulting from the introduction of interactive online content and simulation tools in 2005 [41]. Today, nanoHUB has more than 11,000 users of simulation tools and more than 100,000 (!) registered site users [42]. According to Klimeck, the key to achieving this impact has been that the software has been made easily accessible, and is made accessible in the form of many very specific tools. Currently there are more than 230 such tools available. Tools can easily be built and adapted for a specific task, and also such tool versions are given Digital Object Identifiers (DOI) so that they can more easily be quoted. Apart from research, there is substantial use of these tools in education, as shown by the user log patterns. Current estimates are that about 14,000 students across close to 800 courses use nanoHub. Due to the flexible nature of the setup, the median time from first tool publication to first education use is less than six months, i.e. less than the introduction of a textbook.

It was also investigated whether the transitioning of the tools from the developers to the users and consumer had any negative effect in the quality of the resulting publications. A Nanohub study [42] confirmed that it is possible to do quality research based on tools written by others. There are 850 papers which refer back to Nanohub tools which in turn have more than 8000 secondary citations. This translates into an impressive H-index for the nanoHUB of h=40 (http://en.wikipedia.org/wiki/H-index), which in any case demonstrates impact.

In industry, anecdotal evidence suggests that there is also a steady increase in the number of users, mainly resulting from a widening of the usage to a larger range of companies beyond the first tier in each industry. Surveys from nanoHUB indicate that about 8% of its users are from industry [42], which would mean about 800 people in the actual user category. To put this number into perspective, amongst the top 1400 companies worldwide by R&D spending there are about 650 in relevant industries such as chemicals and semiconductors [43], which seems to suggest that a large fraction now has some modelling and simulation interest.

According to Erich Wimmer from Materials Design [9], there is also more impact resulting from a new generation of technology managers with a deeper appreciation and more realistic judgement on what modelling can deliver. This is in contrast to the first "hype" phase of 10-20 years ago, when large, dedicated but often isolated computational groups were formed. Expectations were raised too high, and large modelling groups were financed on the basis of medium term speculative expectations rather than direct impact on current R&D projects. What followed was a period during which computational science groups in industry were with a few exceptions all but disbanded. Today, molecular modelling is regarded as one of a number of activities that have to justify value by contributing directly to R&D projects. Conversations with a number of modellers in industry suggest that demand for their services is very high. This anecdotal evidence is backed up by a recent survey [29] which indicates that there is widespread interest in industry in molecular modelling outcomes, with the number of interested consumers about 3-10 times larger than the number of users.

Software industry

The development of the computational chemistry software industry up to 2001 was documented by Richon [44], [45]. The first companies formed in the late 1970s, followed by a rapid growth in the 1980s, a flattening off in the 1990s and some consolidation in the early 2000s around the time of the burst of the dot-com bubble. It is estimated that there are about 30 companies in the field today.

While there are just a few dominant players in the market, the software infrastructures developed in the last 10 years support a business model of niche providers that link as Independent Software Vendors (ISVs) into the larger framework providers.

Outside of life sciences, the first companies formed in the 1980. Initial growth led to consolidation in the 1990s down to just one dominant player (MSI) and a few smaller companies. Following this initial period of rapid growth and a subsequent period of consolidation and stagnation, there has been more gradual, sustained growth in recent years (Figure 12).
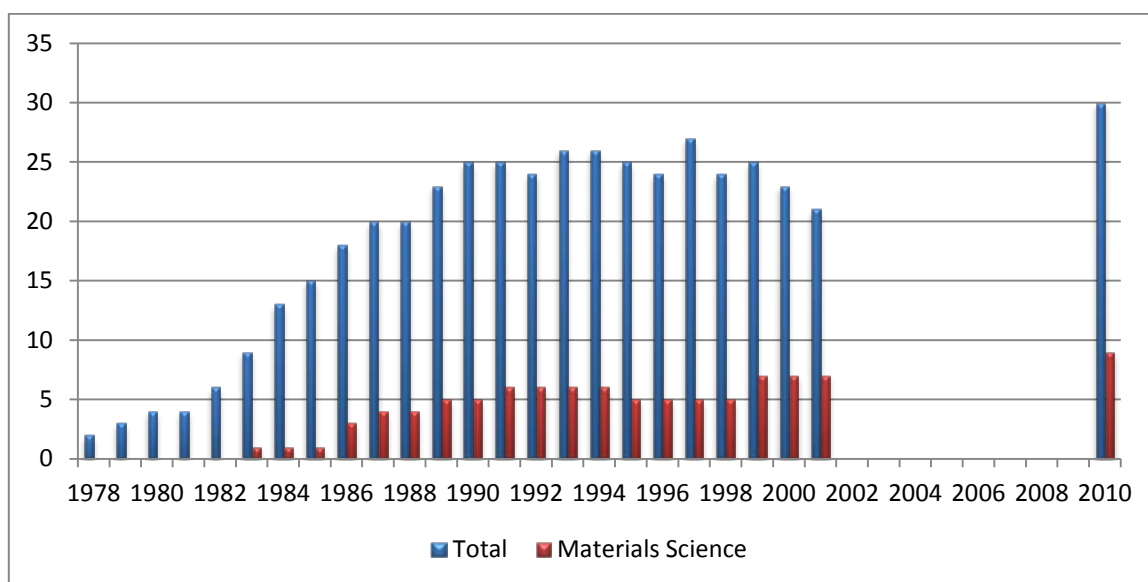
Figure 12: Number of software companies, including data from [44] **up to 2001.**

The size of the software market outside of the life sciences field today is estimated to be in the region of $50m, served by one dominant player (Accelrys), a number of smaller companies (e.g. Materials Design, CULGI, Scienomics, SCM, COSMOlogic) and a range of academic codes sold for relatively small fees.

There is plenty of evidence that the molecular modelling software industry has been following the typical technology "hype cycle" according to Gartner, shown in Figure 13.
http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

Figure 13: Gartner Hype Cycle [Ref: Jeremy Kemp at en.wikipedia, see http://en.wikipedia.org/wiki/Hype_cycle]

The Gartner Hype Cycle methodology has been applied to many technology developments such as automotive electronics, cloud computing and software as a service with the aim of providing decision support for management regarding dep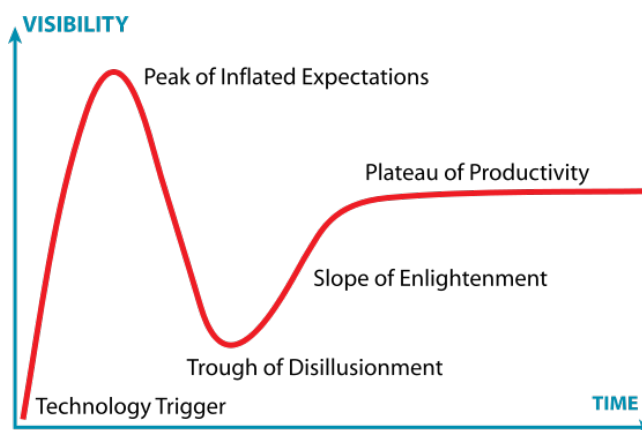loyment of the technology. The graph in Figure 13 shows the five key phases of a technology's life cycle, and it is argued that modelling and simulation is currently in the slope of enlightenment phase. The Technology Trigger phase was in the late 1980s, characterized by the confluence of relatively accessible hardware including graphics workstations, availability of algorithms such as semi-empirical methods, molecular dynamics and DFT codes, and GUIs provided by a range of software providers. Case studies (catalysis, corrosion inhibition, polymers) and first success stories triggered significant industrial interest. Major industry established sizeable computational modelling groups, and revenues, share values and staff sizes were based more on future expectations than actual revenue. As a result, modelling companies grew very substantially up to "The Peak of Inflated Expectations" which was reached in the late 1990s. What followed were the dot-com bubble and the Trough of Disillusionment for modelling and simulation. Industry scrutinized the actual impact on R&D success and found it to be wanting. There are a number of reasons, including the fact that modelling groups were not integrated into the R&D process, as the expectation had been that virtual R&D can somehow produce results in competition with or at least separate to experimental research. Also, methods, computational power as well as experience were still quite limited. At the same time, the academic community continued to grow, as both the algorithms and computing power led to more and more possibilities and interesting new science. In particular, Density Functional Theory based methods really started to come into their own with increased computing power, and algorithm developments that together made really relevant systems accessible. Expertise started to become more widespread, and software companies consolidated their business models. Large players reduced staffing to more sustainable levels, and growth for the sector as a whole, while small, was still found by focussing on growth markets such as Japan, a wider use in the academic sector, and providing consulting and contract research to the electronic industry that started to require nanoscale insights for further CMOS development. There are indications that the sector is now on the Slope of Enlightenment. Software companies are reporting stronger growth even in a period of general economic strain. Patent numbers are rising. Integration of the technology into the R&D workflow is now accepted by an increasing number of companies and those early adopters that kept an investment in the technology are reaping substantial benefits.

Nevertheless there are reports [30] pointing out that "the viability of commercial ventures for chemicals and materials modelling has long been a subject of debate, largely driven by questions about whether the worldwide commercial market for chemicals and materials modelling is currently large enough to sustain multiple commercial ventures."

The software infrastructures and business models for software licensing have been evolving as well, largely reflecting the growing impact of the modelling tools to a wider audience. There is a trend towards more open framework technologies that make integration and deployment of tools much easier. At the same time licensing models have been changing towards multi-user, site or open licenses as previous individual user based license models no longer reflect the usage requirements.

Hence despite a number of years of market stagnation, the impact has typically increased resulting from a combination of more powerful tools, lower price per user and better integration with experimental efforts rather than funding largely separate modelling groups. There are signs now that

the market is also moving forward on the basis of these new business and licensing models, indicating that we entered into a new period of market growth.

R&D process improvement

The evidence on the software market and the number of users across academia and industry suggests that modelling and simulation has become a widely accepted and utilised technique in R&D. As outlined above, molecular modelling in industry is no longer a long term strategic pursuit. If there was no demand or contribution to research projects, the activity would be abandoned. While this 'internal market' provides some mechanism effectively to determine value, it is useful to understand in a bit more detail the mechanisms by which modelling impacts R&D processes and outcomes.

In general terms R&D processes can be considered as the foundation of the technology value pyramid Figure 14 that culminates in value creation [46].
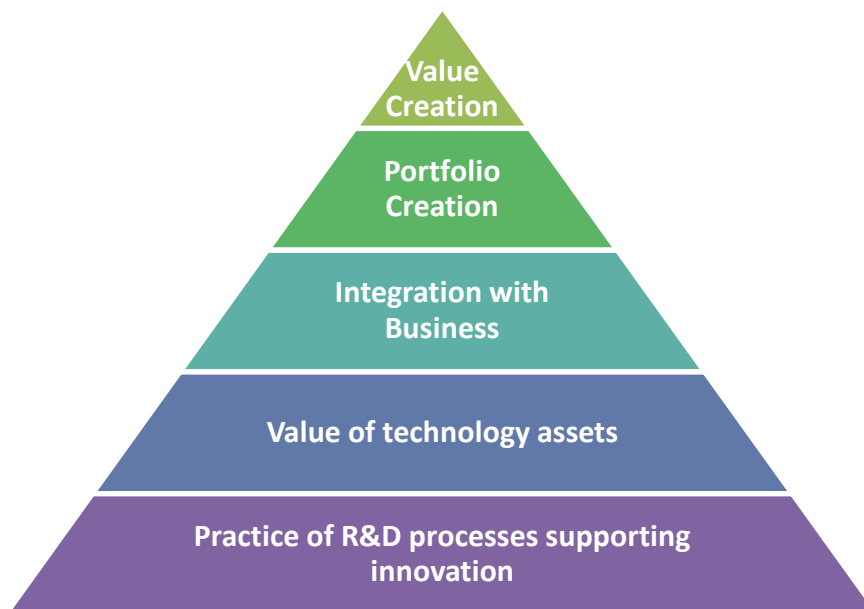
Figure 14: Value creation pyramid adapted from [46]**.**

Parish [46] maintains that one can use metrics to judge the value of research activities that are directly connected to business needs, which we assume to be the case for modelling in industry. In particular three key metrics of value creation are discussed, namely, the new-sales ratio, cost savings, and the present value of the product pipeline. All three are shown to satisfy the criteria of being credible, relevant and not overly complex.

As the pyramid shows, the value creation rests on the practices employed in R&D processes. As pointed out by Hall et al [22], R&D can increase productivity by improving the quality or reducing the average production costs of existing goods or simply by widening the spectrum of final goods or intermediate inputs available. As a consequence, we may observe profit increases, price reductions, and factor reallocations as well as firm entry and exit.

The link of these mechanisms and metrics to modelling and simulation has been established in a couple of studies by the market research organization IDC [27][28]. Based on interviews with researchers and managers in a range of companies, IDC identified significant benefits for the R&D process supporting innovation, due to

- More efficient experimentation
- Broader Exploration and Deeper Understanding
- Saving a Product Development Project and/or Accelerated Product Development
- Improved safety testing and hazard avoidance

These attributes can in fact be related to the value creation metrics of new-sales, cost savings and product pipeline value. More efficient experimentation not only reduces cost, but also accelerates the development. A broader exploration and deeper understanding is more likely to lead to innovative products that will impact the new-sales ratio. Accelerating or even saving product developments clearly contributes directly to all three metrics.

The IDC findings were based on a number of interviews with researchers in a range of companies, and were confirmed in two separate studies, one covering the chemicals industry and the other the pharmaceutical development field. IDC developed a Return on Investment calculation based on the cost of setting up and maintaining a modelling resource and the different benefit derived from the impacts of modelling (see also the Appendix: Metrics and ROI calculations).

Costs considered include software licenses, computational resource, training, IT support and labour. Three different levels of investments are considered. The low end level is based on existing staff without specialist training using standard computing equipment, whereas at the high end specialist staff is employed and equipped with powerful computing equipment. Cost data for the study were taken from typical software license cost, and market rates of hardware at the time the studies were carried out.

Benefit scenarios and values were estimated on the basis of a number of interviews carried out, as well as industry typical data. For example, the cost of an experiment was assumed to be in the range of $500 to $30,000, and a project was assumed to consist of about 10 experiments. Based on the interviews it was concluded that modelling leads to a reduction in the number of experiments required. The level of reduction as well as the number of projects that could be covered depended on the level of investment in modelling, with the modelling and simulation specialist working on 18 projects in a year, and achieving a 35% reduction in the number of experiments in each of the projects. While these numbers seem quite optimistic, they were backed up by interviews. It is obvious, however, that this level of impact requires not only excellent software and hardware but also highly skilled people. The ROI from this activity alone was found to be a factor of 2.3, and given that the software and hardware costs have come down significantly since the study was carried out, the benefit is likely to be higher today.

The impact of broader exploration and deeper understanding is assessed by the fraction of projects that yield product improvements which in turn secure additional market share. While there is evidence provided from interviews which has also been confirmed by this study, quantification rests on too many assumptions to be useful in a general sense. IDC determined an ROI for this activity of 1.5 for the specialist user case.

One of the most cited impacts of modelling in industrial R&D scenarios is that it helps to come up with solutions when a project gets stuck. This may be due to the method being very suitable to coming up with explanations for failure and potential alternatives, but probably also due to the fact that traditionally modelling help has been requested more often when things go wrong. To quote Dr Richard Gilbert, formerly Principal Scientist at e2V Biosensors [47]: The use of modelling solved a problem that had been present for about one year" and "Materials modelling, when used to solve a problem with an existing product, saved over £500k in development costs. This work found a solution in less than two weeks, so the cost of the software was re-couped in about one week.

The quantification of the ROI again rests strongly on the assumptions about the potential product revenue and resulting losses due to delays. IDC assumes that a typical 6 month delay of introducing a new product costs of $100k per month in R&D to solve the problem and leads to lost revenue of $1m a month, totalling $6.5m. The interviews suggested that modelling was key to such a product save once every few years, hence it was assumed that about 1 in 80 modelling projects made that type of contribution, which for the experienced modeller still results in an ROI factor 4.

Similar scenarios were outlined in a second study [28] for the pharmaceutical development sector, which translated into a cumulative ROI of between 3 and 10.

The studies also discuss the barriers that remain to a wider realization of the benefits, and it is interesting to consider to what extent these still apply today. The challenges included the initial investments, hiring or training to the high level of expertise required and lack of integration due to a combination of low acceptance and lack of management support. It is fair to say that the situation has improved markedly since the publication of the study in 2004, with investment cost considerably lower, and a new generation of researchers and managers with a much better understanding of the methods and their integration into the workflow. Nevertheless, molecular modelling is by no means a commodity, and the broad picture remains that considerable and sustained investment is usually required to make an impact.

Contribution to chemistry research impact

The IDC studies have demonstrated that wherever modelling and simulation is applied in chemicals research it makes a positive contribution to the efficiency of R&D processes and their effectiveness in creating value. The findings are also backed up by interviews conducted for this project that lead to the conclusion that there is strong demand for the services of modellers in companies today, and that the market for modelling software tools is growing.

It is therefore reasonable to assume that at least a proportionate share of the value created by chemicals and materials research can be attributed to modelling and simulation methods. In the absence of studies determining the proportion of research that is based on modelling and simulation, a rough estimate can be based on the number of modellers relative to the number of chemists in total. In the US there are about 80,000 chemists [source: WolframAlpha], and from the *ab initio* publications study [35] we know that there are about 5-6000 people in the US involved in modelling. Although not all modellers are chemists, it would seem reasonable to assume 4000 chemists to be modellers, i.e. 5% of the total. Another 'ball-park' figure can be derived from the proportion of the relevant UK research council EPSRC spending on computational and theoretical

chemistry. Based on the number and value of current grants (excluding training grants) on 01 April 2011, the spending £22.3m, equivalent to 3% of total (http://www.epsrc.ac.uk/ourportfolio/researchareas/Pages/comptheochem.aspx)

The value creation of chemicals research in the UK has been analysed by Oxford Economics [7]. The dependence of industry sectors on chemicals research was investigated and the respective proportion of the gross value added (GVA) associated with chemicals research determined as a result. The report finds that the chemicals industry itself creates a GVA of £17.1 bn, equivalent to 1.4% of GDP, of which about half is related to the pharmaceutical industry. In addition, there are indirect and induced value added mechanisms due to supply chains, the spending of those employed in the sector etc, which leads to a total of £36.5bn, equivalent to 3.1% of GDP and affecting 824,000 jobs. It is argued that 100% of this value creation can be related to chemicals research. In addition, 15 major downstream industries that rely to some extent on innovation from chemicals research are taken into account. Each is given a weighting, considering the inputs of products from the chemicals industry as well as their importance and any chemistry related R&D conducted internally within the industry. For example, it is concluded that some sectors such as aerospace and automotive depend fully on chemistry research, whereas electronics is highly dependent and the energy and construction sectors are moderately dependent. The combined contribution from the downstream industries is calculated as a GVA of £222bn, equivalent to 18% of GDP, and 5.2m jobs, Hence, in total the UK's 'upstream' chemicals industry and 'downstream' chemistry-using sectors contributed £258 billion in value-added in 2007, equivalent to 21% of UK GDP, and supported over 6 million UK jobs.

If we assume modelling and simulation to make a 5% contribution (as given by the proportion of people involved in modelling), this would mean that for the UK the GVA contribution is close to £13bn, equivalent to about 1% of GDP, and supporting 300,000 jobs. Similar calculations could be made on the basis of GVA figures of physics to the UK economy for example [24].

Integration with engineering

Background

The Oxford Economics study [7] demonstrated the importance of including the downstream industries such as automotive, aerospace and construction in the analysis of the benefits of chemistry based research. In fact the ability of the downstream industries to utilise the possibilities arising from nano- and information technologies depends strongly on a tighter integration of know-how and methods across the sectors, from chemistry to engineering. Currently, engineering design cycles and the development of new materials that enter into the products are decoupled activities and develop on different timescale. For example, while the development of a new airliner takes of the order of 5 years, it takes 15-20 years for new materials to be adopted in the industry. This creates a drag on the development of new products, and threatens to undermine competitiveness [48]. This argument is also underpinned by a study which attempts to quantify the role of materials innovation in overall technological development for a number of applications [49]. Magee concludes that the contribution of materials to technological development has been rising, and is nowhere less than 20%. A reasonably firm quantitative estimate for the IT sector over the past 40 years is that about two-thirds of the overall progress is due to materials, and there are strong indications for a contribution of more than 80% in the case of energy storage.

According to a 2004 report on "Retooling manufacturing" [50] computational materials science has started to play a key role in the last two decades in this context, enabling a more integral link between materials, design, and manufacturing. Molecular simulations tools are included in a potential new framework for engineering.

In 2006, a report on simulation based engineering science (SBES) [51] stated that: SBES will have a long-term impact on materials innovation. Three attributes of SBES in particular are mentioned that lead to this conclusion:

- Exceptional Bandwidth: The conceptual basis of materials modelling and simulation encompasses all of the physical sciences. It makes no distinction between what belongs to physics versus chemistry versus engineering and so on. This universality of SBES technology represents a scientific bandwidth that is at least as broad as the entire range of multiscale applications in science and engineering. In materials modelling and simulation, as in SBES more generally, traditional disciplinary barriers vanish; all that matters is "the need to know."
- Elimination of Empiricism: A virtue of multiscale modelling is that the results from both modelling and simulation are conceptually and operationally quantifiable. Consequently, empirical assumptions can be systematically replaced by physically-based descriptions. Quantifiability allows researchers to scrutinize and upgrade any portion of a model and simulation in a controlled manner. They can thus probe a complex phenomenon detail by detail.
- Visualization of Phenomena: The numerical outputs from a simulation are generally data on the degrees of freedom characterizing the model. The availability of this kind of data lends itself not only to direct animation, but also to the visualization of the properties under analysis, properties that would not be accessible to experimental observation. In microscopy, for example, researchers can obtain structural information but usually without the energetics. Through simulation, however, they can have both. The same may be said of data on deformation mechanisms and reaction pathways.

The above developments finally led to the concept of Integrated Computational Materials Engineering (ICME) [48], which promises to "reinsert materials into the design and manufacturing process optimization loop". In order to achieve this ambitious goal, the ICME report calls for the development of integrated computational materials engineering defined as the integration of materials information, captured in computational tools, with engineering product performance analysis and manufacturing-process simulation.

Similarly, the WTEC report [30] concludes that computational materials science and engineering is changing how new materials are discovered, developed and applied. It maintains that we are in an extraordinary period in which the convergence of simulation and experiment at the nanoscale is creating new opportunities for materials simulation, both in terms of new targets for study and in terms of opportunities for validation.

Integration concepts

The divide between chemistry based, molecular simulations and continuum engineering models has of course long been recognised and many integration approaches have been investigated. For a detailed discussion of these, see for example the review article by Fish [52], who distinguishes between information-passing and concurrent integration of scale. In the concurrent methods the discrete and continuum scales are modelled simultaneously, i.e. in a truly integrated fashion. In information-passing schemes (which are also sometimes referred to as hierarchical multiscale models) only the gross response from the discrete (e.g. atomistic) scale model is transferred into the continuum scale model. In any case, the focus is on providing eventually an engineering solution based on multiscale integration, as shown in the by now classic scheme in Figure 2.

However, despite many years of research and many publications about multiscale simulations, it appears that these methods are not that widely accepted or used in industrial or engineering applications. For example, polymer mesoscale simulations were positioned in the 1990s as providing the missing link between atomistic and macroscale properties. This led to considerable investment of time and effort into methods to determine input parameters from atomistic simulations on the one hand, and further processing the resulting structures using continuum models. While there have been some promising results, the user base for mesoscale methods remained quite small. Also there is at least anecdotal evidence that most successful industrial applications of mesoscale simulations do not rely on a multiscale approach, but are useful since a key property needs to be calculated at that scale. The question is therefore not one of multiscale, but one of relevance of method to a particular phenomenon.

The ICME report [48] in fact makes a similar argument. The section "Integration tools, the technological "I" in ICME" contains a key statement regarding the integration concepts, therefore quoted here in full: Much of the work that could be deemed computational materials science entails performing calculations in each of these regimes by passing information from one regime-specific tool to another, linking the phenomena across the scales (see Figure 2). While this concept is often useful for defining a modelling strategy, its importance is sometimes overemphasized. Developing and linking models across length scales is **not** required for a workable ICME tool set.

To many in the field, the last statement runs contrary to the concepts and objectives that have been pursued for many years. However, the report contrasts the above with developing models as an engineering activity. While the proper matching of relevant scales to models is important (as in the mesoscale example above), and may require expert assessment, the focus is really on (a) fully assessing the influence of manufacturing processing on materials properties, and (b) bringing together knowledge from disparate sources and domains. Hence an ICME integration scenario looks more like shown in Figure 15.
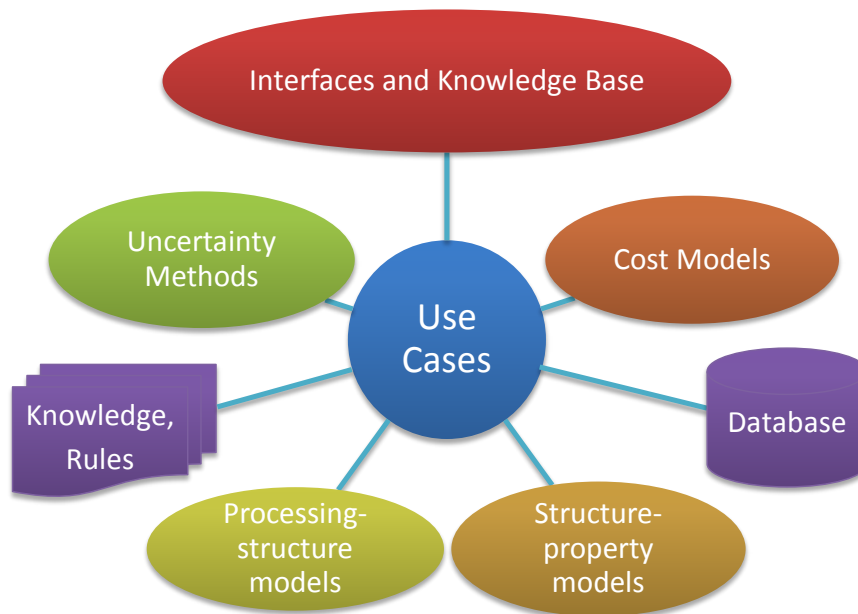
Figure 15: Schematic representation of Integrated Computational Materials Engineering. Adapted from Figure 1-1 of [48]**.**

Another way of looking at the ICME integration is that it is based on the classical process-structure-property concept, and is set up in a way that the state-of-the-art in terms of knowledge and models are brought together to achieve the engineering objectives. Rather than starting at the atomistic scale, a top-down approach is used that should however be able to draw upon molecular modelling and data derived from such models wherever necessary. While this may 'downgrade' the molecular model to a small cog in large wheel, it provides a well-defined route to impact.

Mechanisms and metric of impact
Since ICME connects fundamental, atomistic and molecular based models with engineering outcomes, it lends itself particularly well to looking at the economic impact, as measured in terms of metrics such as reducing the time to introduce new products, cost savings in product development and savings in manufacturing costs (materials and processes) as a result of improved understanding and design. In fact, the report [48] claims that "ICME is a technologically sound concept that has demonstrated a positive return on investment and promises to improve the efficient, timely, and robust development and production of new materials and products."

The expected return on investment (ROI) of ICME is attributed to a number of factors. Firstly, the report lists several items related to design, manufacturing and life cycle:

- Design innovation, quicker identification of materials.
- Solutions to design problems.
- Faster and less costly new product development.
- Better control of the manufacturing process.
- Improved capabilities for predicting engineering system performance or life cycle.

- Virtual engineering assessment of new materials that might be considered risky to assess with physical prototypes.
- Virtual engineering assessment in systems where the validation of materials performance by system-level testing is expensive, time consuming, or not possible.

Furthermore, there are factors related to time-to-market and satisfying market needs:

- Faster time-to-market for new products.
- Market advantage based on improved performance from incorporating materials and processes optimized for particular applications and on more precise modelling of a material's response to an application environment.

An ROI between 3:1 and 9:1 has been reported in a number of cases. Key factors contributing to successful implementations were

- Selection of appropriate engineering problems consisting of a manufacturing process, a material system and an application that steer the development of the computational tools and the infrastructure.
- Sustained investment.
- Overcoming cultural issues.

A flagship example is the Virtual Aluminium Castings software package which was developed at Ford Motor Company [53]. It has been described as a "rare technological innovation that can be used to simultaneously reduce cost, improve quality, save time, and reduce weight." The methodology was based on a holistic approach to aluminium casting component design; it modified the traditional design process to allow the variation in material properties attributable to the manufacturing process to flow into the mechanical design assessment. Fully funded by Ford to address specific power-train components, the VAC methodology was implemented by the company for cast aluminium power-train component design, manufacturing, and CAE. According to Allison et al, VAC has saved Ford millions of dollars as a result of a number of direct and indirect mechanisms:

- Selection of the most economical manufacturing process that produced components meeting the property requirements.
- Optimisation of the manufacturing process.
- Optimised design, based on the ability to predict local properties

In addition, there were impacts on the organisation and its ability to work efficiently and effectively, as the framework has provided a common tool for use by the CAE community in the company, and enabled a comprehensive knowledge capture.

At first sight, VAC may seem like a purely engineering modelling and knowledge based system. However, some of the key features that contribute to the precision and ultimately success of the tools are due to microstructural modelling at many different length scales to capture the critical features required to accurately predict properties. For example [53] one of the weaknesses of traditional casting engineering design tools has been limited accuracy and relevance due to the fact that the data needed to feed into the empirical approaches are very difficult or costly to determine. This was overcome by simulations that include ab initio and atomistic modelling. In fact, the

"method of optimizing heat treatment of alloys by predicting thermal growth" has been patented by Ford [54]. It includes first principles density functional calculations of volume changes in the precipitation-hardened alloy due to transformations.

The outputs of this processing–structure–property information are predictions of manufacturing-history-sensitive properties. For example, the design of a cylinder head casting was improved by the ability to predict the spatial variations in fatigue properties and the influence of the casting process on the location-dependent properties was key to the success. Since the variation of properties in complex components can be large (30-40%), the use of nominal figures instead of local-dependent properties can lead to either overly optimistic or conservative design, i.e. leading to durability failures and costly redesign or to unnecessarily heavy and costly products.

Hence, while ab initio and atomistic modelling may represent only a small part in ICME, it is nevertheless integral to its success, a kind of "vitamin for engineering" [9].

Another factor has been successful change management during the project. To ensure the acceptance of not only the new tools but also the new design process, extensive validation of the approach was carried out, and the necessary degree of accuracy was determined in consultation with product design engineers to build confidence in the predictions.

The move from the traditional approach of design-build-test-redesign-build-retest also helped to avoid potentially very costly delays in product launch as small manufacturing changes may lead to engine durability problems that are poorly understood. According to Allinson [48], [53], there were multiple, quantifiable benefits from introducing these improvements, including:

- 15-25% reduction in time to develop a new cylinder head or block.
- Reduction in the number of component tests required for assurance testing.
- Shorter cycle time for the casting or heat treatment process.

A cost/benefit analysis estimated a combined ROI of over 7:1 for the project. On the cost side, the VAC program involved approximately 25 people and $15 million in expenditures over 5 years, over half of which was for experimental work in either model development or validation. On the benefits side, the technology is projected to have saved Ford up to $100 million in cost avoidance by resolving product issues and between $2 million and $5 million per year in reduced testing requirements and reduced development time.

While the initial development of the software required a very substantial investment in time, money and staff, updates of the system were achieved at much lower cost. For example, Allinson points out that a VAC tool set for a new alloy for aluminium cylinder heads for diesel engines was completed in less than half the time and at a cost of between 10 and 20 % of that required for development of the initial VAC tool set.

The Ford VAC example provides a well-documented case example for the impact that can be achieved if modelling tools and data across different disciplines are integrated. However, the large investment over many years as well as the cultural issues that had to be overcome are a strong reminder of the significant challenges faced in realizing this potential on a wider scale.

Impact facilitated by e-infrastructure

High Performance Computing

The huge impact of high performance computing on many areas of research, industry and society has been documented and analysed in numerous studies, e.g. [3], [25], [26]. A headline examples of the strategic importance of simulations enabled by HPC is the total replacement, at the end of the 1990's, of the French military nuclear experiments by computational simulations.

A number of studies of HPC use in industry have found that companies that have adopted such technology for virtual prototyping and data modelling have experienced significant gains in productivity. In particular, the benefits of computational models include reduced development and re-design costs, improved performance and efficiency as well as reduced waste resulting from lower emissions, noise and raw material use. IDC found [25] that almost all companies using HPC indicated that the technology was indispensable for their business. The mechanisms by which HPC is claimed to contribute to the success of research and leads to economic returns are basically the same as have already been discussed for modelling and simulation, i.e. improving research productivity, efficiency and effectiveness in coming up with solutions that produce an economic return. As in other areas such as the creative industries, retail and finance, HPC can be regarded as both a driver and multiplier of these positive attributes. With HPC, new phenomena and challenging questions can be addressed, thereby driving new discoveries and product developments, which in turn lead to growth and benefits in many other areas.

While quantification of the returns from HPC are difficult, especially in science based activities, IDC reports [3], [55] point out a number of indicators:

- A growing number of Nobel laureates have relied heavily on HPC for their achievements.
- HPC has grown from its established stronghold in the physical sciences to the social sciences and the humanities, which confirm the benefits of originally science based computing to many other areas of society.
- In an IDC study, 97% of the industrial firms that had adopted HPC said they could no longer compete or survive without it.
- In the automotive and aerospace industries, HPC has dramatically reduced the time-to-market and increased the safety and reliability of new vehicle designs.
- Some large industrial firms have cited savings of $50 billion or more from HPC usage.

The IDC strategy report [3] includes ROI scenarios for investments into hardware, software and people in the EU. The study expects the investment to make an impact not only on the HPC suppliers but also on industries that use the improved HPC infrastructure and tools to make better and more competitive products and services. Experience and data from earlier studies indicated that HPC can be a major revenue multiplier, and that the comparative level of under investment in Europe is a case in which such large returns of investment are likely. As an example, IDC quoted an estimate made by Boeing a few years ago. Apparently, HPC use had saved the company more than $60 billion, compared to a spending of well under $10 million on HPC per year, a factor of more than a 1000 even considering several years of investment at that level [25].

Along similar lines, IDC estimates that the EU can derive very large benefits from an additional total investment of €600m in HPC. Without providing any model or calculations of the economic impact,

IDC expects the incremental growth in HPC utilising industries to be in the range of 6-8 percentage points by 2020. Given a 27% proportion of these industries to the EU GDP in total, this equates to about two percentage points incremental growth of EU GDP by 2020. In 2011 terms, 2% of EU GDP is about €245bn, i.e. an ROI in the same range as that of the above Boeing example.

These are clearly very large figures, and it would be useful to see further justification and more detailed models. The figure quoted for the potential growth impact from just the HPC sector alone is questionable. The study concludes that Europe could see a 0.5%–1% growth in GDP just from the HPC sector by 2020. Assuming 0.5% for now, that would be € 61bn in 2011 terms, while the worldwide revenue of the HPC industry sector in 2011 was expected to be € 8bn. While the above impact figure presumably accounts also for government funded research etc, it looks like an inflated estimate.

With the above figures and provisos as background, the following statements from the IDC report are significant in relation to the impact of HPC in the chemistry and materials sector.

The IDC HPC [3] study lists a sample of 16 application areas targeted for multipetascale and exascale computing, of which 4 would involve atomistic modelling outside of life-sciences: Quantum chemistry, Advanced combustion modelling, Nanoscale material science, and Molecular nanotechnology. It also recommends modelling of materials/molecular dynamics as one of six target domains, along with weather and climate research, clean and sustainable energy, automotive and aerospace design, bio-life sciences, and particle physics and related fields. It emphasizes the central role of molecular modelling for a range of scientifically and economically important fields, including materials science (development of new materials, aging of materials), alternative energy (improved design for solar cells, wind turbines, etc.), drug discovery and other biomedical research, nanotechnology and product engineering.

There are however reservations in the chemicals industry. According to one of the comments in the IDC survey [55], for companies like AkzoNobel and DSM the jury is still out if high-end computational chemistry can make a difference. Similar scepticism has been expressed in a survey of industrial requirements for thermodynamic and transport properties [5]. Key technical people in companies in the oil and gas, chemicals and pharmaceutical/biotechnology sectors did not express much interest in future developments of molecular modelling, despite its academic success. The investigation concludes that industry is not yet fully aware of the applicability of molecular modelling to address their requirements. Hence, further efforts by the academic community are required to introduce the use of advanced computational techniques, by both further refining theoretical models and showing that the results can be used to provide key data.

Given the above statements it is perhaps not surprising that the fraction of the HPC market related to chemicals and materials modelling in industry is relatively small. The share of the so-called "Chemical Engineering" sector (which includes applications such as molecular modelling, computational chemistry, process design, and chemical analysis) is about 3%. In contrast the academic sector accounts for nearly 17%, followed by biosciences (16%) and computer-aided engineering (15%). On the other hand, the Chemical Engineering sector is one of the fastest growing sectors. It kept growing during the 2005-2009 downturn when all other sectors except defence and financial shrank. It is also expected to be amongst the top performing sectors going forward, with a CAGR close to 10%.

Based on the above, an estimate of the fraction of the HPC sector which is due to molecular modelling and computational chemistry outside of life sciences can be made. The industry contribution is a fraction of the 3% that is represented by the above mentioned chemical engineering sector. Considering that chemical engineering applications includes sectors such as process design which are much larger than molecular modelling, one can assume that computational chemistry and molecular modelling application in industry account for less than 1% of the total HPC sector in industry. This relatively low figure is not surprising since many industrial application of molecular modelling are performed on small to medium size computing hardware. Also, companies collaborate with academia and computing centres wherever there is a requirement for large simulations.

The situation is quite different in the academic and government sectors however. A quick internet survey of annual reports from major computing centres suggests that 20-30% of CPU usage is due to electronic structure, computational chemistry and molecular modelling applications. For example,

- the annual report of the Swiss National Supercomputing Centre includes a breakdown into different application fields, including Chemistry, Nanosciences and Materials Science, making up 16%, 9% and 11%, respectively in 2010 [56].
- At the HLRS (Stuttgart), chemistry accounts for only about 3% of usage, but solid state physics for about 25% in 2009. While the chemistry proportion is rather small (to some extent related to the computing infrastructure in place at HLRS) the figures are still consistent with a proportion of 'materials' simulations based on atomistic and molecular structures of about 20% [57].
- At Forschungszentrum Jülich, in the context of PRACE, Chemistry and Materials Science accounted for 8% of CPU time in 2010 with a further 38% allocated to fundamental physics, likely to include a significant portion of electronic structure and similar ab initio calculations.
- At CSC in Finland, the density functional methods (VASP and GPAW) accounted for 44% of the total software usage in 2010. On the other hand, the largest programs in terms of number of users were engineering (FEM), biosciences, mathematics and linguistics programmes [58]

As a result of the above estimates in combination with the figures published by IDC on the size of the European HPC market in the academic and government sector (17% and 14%, respectively), it is estimated that atomistic and molecular modelling accounts for about between 5 and 10% of the HPC market. As a consistency check, the size of e-science funding relative to the total government funding for science and research in the UK is 4% (Total science and research resource: £4.6bn [59], e-science funding: £200m [60]). Another calculation which confirms the 5-10% share is based in software markets. According to IDC's HPC [3] report, software costs for applications and middleware (i.e. excluding operating systems) are about 75%–85% of hardware. Based on IDC figures this would suggest a software market of €61m if molecular modelling has a 5% share of HPC, which is close to general estimates of the molecular modelling software market size.

In conclusion, based on a modest 5% share of HPC, the potential impact on GDP growth resulting from further HPC investments in Europe would be in the range of €3-6bn by 2020. In general, molecular modelling applications are a major HPC sector universities and government labs. In the

chemicals industry molecular modelling based on HPC is growing at an above average rate, but from a level much below that of disciplines such as biosciences and engineering.

High-throughput computation and informatics

"High-throughput computation (HTC) involves the generation of materials structure libraries (molecules, ensembles, surfaces), followed by computation to predict key intrinsic properties including, for example, reaction energetics, surface energies, or band gaps. The resulting virtual materials database is a unique and powerful resource, allowing the identification of optimal structures or formulations. It assists experimental efforts through data mining and screening of chemical design space." [61]

High throughput computation and informatics is also a key part of the US Materials Genome Initiative [62] which supports a partnership between Harvard University and Wolfram Research, leveraging IBM's World Community Grid to accelerate the testing of millions of new, simulated organic molecules that might be used for low cost, effective and easily produced materials to conduct and store solar energy.

Other examples include the Materials Project [www.materialsproject.org] which aims to accelerate materials discovery through advanced scientific computing and innovative design tools and a UK collaborative project called iCatDesign [63], [64].

The impact of the approach is typically derived from (a) enhancing the productivity of the users due to automation and utilisation of grid computing resources, and (b) providing data and analysis protocols to the 'consumer' community thereby improving the transfer of knowledge.

Benefits of using high throughput computation in terms of the time saving for the modeller were quantified in a case study on battery additives by Accelrys [61] , see Figure 16. High-throughput computation was used to predict key additive properties for a combinatorial library of 7,381 compounds and to consolidate the results. The improved productivity was estimated as follows: Doing the calculations individually would require 470 hours, i.e. at least 3 weeks. The vast majority of the time is spent setting up the structures and extracting results. By contrast, the automated approach requires only 96 hours, including the time required to write setup and analysis protocols. Also, in this case the traditional approach does not lend itself to well parallelization whereas the trivial parallelisation can be used for the automated approach which can therefore be completed in 36 hours giving a total saving of more than 400 hours. See also [65] for further information on this and other cases.

Timing Comparisons for Battery Additive Study

Traditional *in silico* approach

Sketch structures:	120 sec
Setup input file:	5 sec
Setup cation file:	5 sec
Calculation:	39 sec
Extract results:	60 sec
Time for 7,381 calculations:	470 hrs

High-Throughput *in silico* approach

Sketch EC framework:	120 sec
Sketch 12 pendant groups:	1440 sec
Setup input files:	5 sec
Calculation:	39 sec
Extract results:	n/a
Time for 7,381 calculations:	80 hrs
Elapsed time with 4 cores:	20 hrs
Time to write protocol:	16 hrs

Net savings:	>400 hours

Figure 16: Time saving case example

In conclusion, high throughput computing and its related informatics infrastructure therefore perform a strong multiplier function enhancing the impact of molecular modelling.

Gaps and barriers to impact

As a wide range of studies and analyses summarized in this report have shown, there is a wealth of evidence demonstrating the economic impact of molecular modelling. In particular, it was possible to identify at least some examples of transmission mechanisms and impact from the research base to society. Nevertheless, a number of barriers to impact remain, and these have also been expressed in many publications as well as in the interviews conducted for this report.

Starting from the fundamental research level, there remain considerable scientific challenges for example [9] in the ability to calculate properties such as reaction barriers with high chemical accuracy, taking the inherent many-body nature of many problems into account. While the development of more accurate methods (e.g. Quantum Monte Carlo) is progressing it will take time before these become main-stream. Also, the ability to handle large configuration space and the bridging of scales remains challenging.

While the classical and mesoscale techniques have reached a point at which the author level achieves high accuracy and potential impact, there are still barriers to wider impact on the user community. As pointed out by Maginn [2] [66] and the Fluid Simulation Challenge community [16] the non-expert users must be capable of carrying out data-driven molecular simulations to compute properties of interest or develop and test new engineering models. In order to reach this level the following issues must be addressed [66]

- The scope of properties to be computed must be clearly defined in order to develop tools directed at these properties.
- Software and analysis tools must be integrated to compute the properties required using the best methods.
- The coverage of force fields must be improved to encompass the range of molecules and state points identified by industry as being important. Repositories must also be developed to store and distribute those force fields.
- To support validation, procedures must be developed whereby the calculation results can be archived and compared against experimental data and other benchmark calculations.
- A database of molecules and equilibrated structures similar to the Protein Data Bank - should be created to reduce the time wasted in generating initial structures.

These limitations were also echoed in a workshop on the Predictive Modeling of Nanomaterial Properties, October 9-10, 2007, Arlington, VA, which found that "progress has also been hampered by a notorious limited knowledge of the uncertainty of the most widely used computational methodologies."

It becomes clear from the survey on Industrial Requirements for Thermodynamics and Transport Properties [5] that this state of affairs severely limits the impact in downstream sectors, such as the processing industry. Despite the fact that there is an identified need to tools that can generate high-quality thermodynamic data, the survey finds that the use of molecular simulation or quantum

chemistry as an alternative source of data is currently not widely accepted by industry. According to Hendricks et al this may be due to a lack of information, in that industry is not fully aware of the applicability of these computational techniques to address their requirements. In any case, the participants of the survey "were not particularly encouraging the further developments of molecular simulation, which is a rather disappointing state of affairs."

While the ICME examples have demonstrated large economic impact in specific cases, a barrier to realizing the full potential of molecular modelling lies in the considerable investment required in implementing modelling in an industrial environment, taking the costs of software, hardware, people, training and adaptation into account. While this true for chemicals and materials research, it is even more pronounced in the engineering field. The Ford VAC software development for example [53] involved coordinating the fundamental research efforts of five universities across the United States and the United Kingdom. Substantial effort was applied to developing efficient links between the output of the casting modelling and structure and property prediction tools to feed seamlessly into the FEA codes and to facilitate reorganization of the design process.

Similarly, the 2006 report on Simulation based engineering science [51] pointed out the need for modelling and simulation of complex, interrelated engineered systems and the acquisition of results meeting specified standards of precision and reliability. The scope of this engineering focussed approach includes much more than the modelling of physical phenomena. The differences and requirements have been outlined in a paper by Kuehmann and Olson [67]. They state that computational materials models for design purposes differ significantly from those used in traditional materials science research activities in that conventional materials modelling strives to understand general phenomena, whereas design models are used to control a materials system and optimise it for a specific outcome. This requires high accuracy, robustness and good uncertainly quantification for the specific design, as well as speed to allow a large design space to be explored. Models also need to be highly cooperative to determine the best combinations of process and composition to meet a diverse set of material objectives.

While these requirements have been addressed in some cases, the links between physical and system level simulations remain weak in general [30]. According to the report, there is little evidence of molecular models that are tightly coupled with process and device models. Barriers that are identified include the cost of experimental validation of models, and that uncertainty quantification is not being addressed adequately in many of the applications. As a result, modelling and simulation methods are mostly used to understand/explain experimental observations, but are not ideally suited for developing new products.

Regarding the codes and software infrastructure, the interoperability of software and data is seen as a major hurdle resulting in limited use of simulation software by non-simulation experts. It is argued that the utility of materials simulation codes for practical application would be enhanced dramatically by the development of standards for interoperability of codes similar to the CAPE-OPEN effort undertaken in the chemical process simulation (computer-aided process engineering, or CAPE) field (http://www.colan.org). [30]

There also still seem to be a lack in expertise and sufficiently trained people [9], [29]. While there are many brilliant scientists, there is often a gap to the communication with engineers and experimentalists. Hence there is a need for more people with a skill for translating between

industrial problems and what can be done with molecular simulation technology, supported by technology managers with the knowledge, perception and position to commit resources to simulation which are at least comparable to experimentation.

Finally, in addition to addressing gaps in the transmission due to modelling methods, software interoperability and people, there is also increasing attention on the information gap between the 'molecules', i.e. fundamental chemistry and the products. While systems are typically in place to capture product formulations, the research and development informatics, or 'e-chemistry', consists of a multitude of systems that don't connect [29], limiting the impact that models generated at the molecule level can make for example in a situation when customer preference data call for a product redesign. There are initiatives, however, to overcome this [68].

Conclusions

This review of the economic impact of molecular modelling found a range of evidence which documents a transmission of knowledge and impact from the fundamental theories via the users of software to the researchers that utilise the results and to new, engineered products. While a quantification of the impact of one methodology is very difficult, it has been possible to consider measures of impact at each stage along the above mentioned pathway.

This included publications, patents, the growth of the communities of modelling authors, users, and consumers of the technology and the ensuing software industry. Quantification of R&D process improvements due to modelling could be related to key metrics of value creation. On a macroeconomic level, it was attempted to estimate the contributions that modelling makes as a proportion of chemistry research as well as high performance computing.

One of the key pathways to realizing impact relates to the integration of molecular modelling into engineering workflows, a concept often referred to as Integrated Computational Materials Engineering (ICME). Case examples indicate a strong return on investment, but it remains the case that the benefits to industry at large remain uncertain, a fact which may induce firms to adopt a wait-and-see approach [50]. A number of fundamental economic questions still need to be addressed in more detail to assess the real costs, benefits and barriers, including:

- What are the costs of tool integration and how do they vary by industry?
- What are the expected benefits in terms of cost reduction and strategic advantage?
- What are the impacts for productivity growth?
- How do organizational and management structures affect development and adoption?

Despite these reservations, the trend towards a "reshaping of how we innovate" [68] is set to accelerate, and will include a stronger integration of molecular modelling methods along the innovation and value chain. As researchers from P&G put it [18] it is an expectation rather than a 'dream' that molecular modelling is at the core of the "next-generation laboratory." Future studies of impact should be able to corroborate this expectation.

Appendix: Metrics and ROI calculations

This appendix provides an overview of the metrics that have been used to determine impact, and includes further details of the return on investment models used by IDC [27], [28].

Metrics

- Number of publications and their growth relative to other fields
- Impact of publications
- Growth in the number of patents
- Number of users in academia and industry
- Number of consumers of modelling results
- Market size and growth of software industry
- Return on investment due to R&D process improvements
- Economic impact associated with the impact of chemistry research on the economy
- Return on investment of large integrated computational materials engineering projects
- Impact associated with high performance computing
- Benefits of high throughput computation and informatics

IDC studies for Accelrys

IDC [27], [28] developed a cost and benefit model for materials and pharmaceuticals modelling, respectively. Costs considered include software licenses, computational resource, training, IT support and labour. Three different levels of investments are considered. The low end level is based on existing staff without specialist training using standard computing equipment, whereas at the high end specialist staff is employed and equipped with powerful computing equipment.

The cost is calculated as:

$$C = SW + HW + IT + L + T,$$

Where:
C = Cost,
SW = Software,
HW = Hardware,
IT = IT support,
L = Labour,
T = Training.

Benefits were evaluated based on a number of mechanisms which describe R&D efficiency and effectiveness in general, and which have been documented to be impacted by computational modelling. These so-called scenarios include

- More efficient experimentation.
- Broader Exploration and Deeper Understanding.
- Saving a Product Development Project and/or Accelerated Product Development.
- Risk Management through Safety Testing.

For each scenario the benefit was estimated by a formula of the form:

$$B = VT \times Np \ (C) \times R$$

Where:

B=Benefit attributed to modelling.

VT = total commercial value of the respective mechanism.

Np (C) = Number of projects that involve modelling, which is a function of the resource available, hence cost.

R (C) = Percentage of modelling projects that make an impact on the mechanism, which is also found to depend on resource, with highly skilled/trained staff with good equipment more likely to make an impact.

Cost data for the study were taken from typical Accelrys software license cost, and typical market rates of hardware at the time the studies were carried out. Obviously there have been considerable changes in costs since then, with much reduced hardware costs for more powerful computing capabilities, as well as reduced software costs. Benefit scenarios and values were estimated on the basis of a number of interviews carried out, as well as industry typical data.

The spread sheet below includes the data for low end, occasional user and the high end specialist cases as in the IDC study. In addition, a third column provides an updated scenario for a typical modelling user today with a lower cost base due to the reduced price of hardware as well as software since the original study. On the other hand, it was also assumed that the user will not be as highly skilled as some of the experts that were interviewed by IDC, hence the number of projects impacted was reduced from 18 to 10. Nevertheless, the return on investment is substantial.

1	**Efficient Experimentation**	Low	High	Updated
	Cost per experiment	13000	13000	13000
	Experiments per project	10	10	10
	Number of projects impacted	4	18	10
	Reduction on experimentation	15%	35%	50%
	Benefit	**78,000**	**819,000**	**650,000**
	ROI	**1.56**	**2.34**	**2.81**

2	**Innovation due to broader exploration**	Low	High	Updated
	Total market size for product category	100,000,000	100,000,000	100,000,000
	Market share increase resulting from project	1%	1%	1%
	Percentage of Projects Generating a Product Improvement	7%	20%	20%
	Number of projects impacted	4	18	10
	Contribution from modelling	15%	15%	15%
	Benefit	**42,000**	**540,000**	**300,000**
	ROI	**0.84**	**1.54**	**1.30**

3	**Saving stalled projects**	Low	High	Updated
	Percentage of projects saved	0.20%	1.25%	1%
	Value of save, Development cost per project	6,500,000	6,500,000	6,500,000
	Number of projects impacted	4	18	10
	Benefit	**52,000**	**1,462,500**	**812,500**
	ROI	**1.04**	**4.18**	**3.52**

4	**Risk Management Through Safety Testing**	Low	High	Updated
	Percentage of Projects with a Hazard or Safety Element	1%	3%	3%
	Value of Hazard or Liability Avoidance	2,000,000	2,000,000	2,000,000
	Number of projects impacted	4	18	10
	Benefit	**80,000**	**1,080,000**	**600,000**
	ROI	**1.60**	**3.09**	**2.60**

DIRECT BENEFITS	Low	High	Updated
More efficient experimentation	78,000	819,000	650,000
Broader exploration	42,000	540,000	300,000
Saving stalled projects	52,000	1,462,500	812,500
Risk management	80,000	1,080,000	600,000
Potential TOTAL DIRECT BENEFITS	**252,000**	**3,901,500**	**2,362,500**

DIRECT COSTS	Low	High	Updated
Software licenses	35,000	90,000	40000
Hardware	6,000	100,000	30000
Training	7,000	2,000	3000
IT support	2,000	8,000	8000
Labour	0	150,000	150000
TOTAL DIRECT COSTS	**50,000**	**350,000**	**231,000**
ROI estimate	**3**	**9**	**7**

References

[1]   P. Warry, "Increasing the economic impact of Research Councils," Jul. 2006.

[2]   E. J. Maginn, "From discovery to data: What must happen for molecular simulation to become a mainstream chemical engineering tool," *AIChE Journal*, vol. 55, no. 6, pp. 1304-1310, Jun. 2009.

[3]   E. C. Joseph, C. Ingle, C. Meunier, S. Conway, G. Cattaneo, and N. Martinez, "A Strategic Agenda for European Leadership in Supercomputing: HPC 2020 — IDC Final Report of the HPC Study for the DG Information Society of the European Commission," Sep. 2010.

[4]   European Commission, "ICT Infrastructures for e-science," Brussels, Mar. 2009.

[5]   E. Hendriks et al., "Industrial Requirements for Thermodynamics and Transport Properties," *Industrial & Engineering Chemistry Research*, vol. 49, no. 22, pp. 11131-11141, Nov. 2010.

[6]   K. E. Gubbins and J. D. Moore, "Molecular Modeling of Matter: Impact and Prospects in Engineering," *Industrial & Engineering Chemistry Research*, vol. 49, no. 7, pp. 3026-3046, Apr. 2010.

[7]   "The economic benefits of chemistry research to the UK," Sep. 2010.

[8]   E. Wimmer et al., "Ab initio calculations for industrial materials engineering: successes and challenges," *Journal of Physics: Condensed Matter*, vol. 22, no. 38, p. 384215, Sep. 2010.

[9]   E. Wimmer, "Impact of modelling." Mar-2012.

[10]   G. Martyna, "Impact of modelling." Mar-2012.

[11]   "Nanoscale Device Modelling Research Positions IBM India - Computational Modelling Group," 2012. [Online]. Available: http://cmg.soton.ac.uk/vacancies/23/. [Accessed: 13-Mar-2012].

[12]   J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, "Towards the computational design of solid catalysts," *Nature Chemistry*, vol. 1, no. 1, pp. 37-46, Apr. 2009.

[13]   P. D. Haynes, A. A. Mostof, C.-K. Skylaris, and M. C. Payne, "ONETEP: linear-scaling density-functional theory with plane-waves," *Journal of Physics: Conference Series*, vol. 26, pp. 143-148, Feb. 2006.

[14]   R. Meier, "Multi-scale modelling in a chemistry and polymer environment how far have we got, what do we hope for?" Brussels, Sep-2011.

[15]   "Industrial Fluid Properties Simulation Collective." [Online]. Available: http://fluidproperties.org/.

[16]   F. H. Case et al., "The sixth industrial fluid properties simulation challenge," *Fluid Phase Equilibria*, vol. 310, no. 1–2, pp. 1-3, Nov. 2011.

[17] S. Gupta and J. D. Olson, "Industrial Needs in Physical Properties," *Industrial & Engineering Chemistry Research*, vol. 42, no. 25, pp. 6359-6374, Dec. 2003.

[18] "Procter & Gamble's Story of Suds, Soaps, Simulations and Supercomputers," Nov. 2009.

[19] A. Browning, "Utilization of Molecular Simulations in Aerospace Materials: Simulation of Thermoset Resin/Graphite Interactions," 2009.

[20] Chemical Industry Vision2020 Technology Partnership, "Chemical Industry R&D Roadmap for Nanomaterials By Design: From Fundamentals to Function," Dec. 2003.

[21] Y. Samson, "Report on the workshop "Towards a multi-scale, multi-phenomena modelling-simulation-design-engineering environment & tools," Brussels, ERC building, Place Madou, Sep. 2011.

[22] B. Hall, J. Mairesse, and P. Mohnen, "Measuring the Returns to R&D," vol. 2, Elsevier, 2009.

[23] D. Hanson, "Valuing R&D," *Chem. Eng. News*, pp. 22-23, Jan-2010.

[24] "Physics and the UK Economy," London, Sep. 2007.

[25] E. C. Joseph, J. Wu, S. Conway, and S. Tichenor, "Benchmarking Industrial Use of High Performance Computing for Innovation," May 2008.

[26] E. C. Joseph, S. Conway, and J. Wu, "Massive HPC Cloud Redefine Scientific Research and Shift the Balance of Power Among Nations," Framingham, MA, 2009.

[27] M. Swenson, M. Languell, and J. Golden, "Modeling and Simulation: The Return on Investment in Materials Science," Jun. 2004.

[28] A. Louie, M. Brown, and A. Kim, "Measuring the Return on Modeling and Simulation Tools in Pharmaceutical Development," Jan. 2007.

[29] M. Doyle, "Impact of modelling." Apr-2012.

[30] S. Glotzer et al., "International assessment of research and development in simulation-based engineering and science," Baltimore, 2009.

[31] G. Lewison, "Beyond SCI citation - new ways to evaluate research," *Current Science*, vol. 89, no. 9, pp. 1524-1531, Nov. 2005.

[32] G. Lewison, "The percentage of reviews in research output: a simple measure of research esteem," *Research Evaluation*, vol. 18, no. 1, pp. 25-37, Mar. 2009.

[33] B. M. Althouse, J. D. West, C. T. Bergstrom, and T. Bergstrom, "Differences in impact factor across fields and over time," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 1, pp. 27-34, Jan. 2009.

[34] P. Dederichs, P. Mavropoulos, and D. Tunger, "The size of our field of ab initio simulations," *Psi-k Monthly Update*, vol. 3, Feb-2010.

[35]   P. Dederichs, "Growth and Regional Strength of the Field of Ab-initio Calculations," *Psi-k Chairmen's Update*, vol. 8, Jul-2010.

[36]   P. Dederichs, "Ab-initio Publications of European Countries No. 11 (October 2010)," *Psi-k Chairmen's Update*, vol. 11, Oct-2010.

[37]   "High-density Energy Storage: Better Batteries through Simulation," May-2012. [Online]. Available: http://www.scientificcomputing.com/news-HPC-High-density-Energy-Storage-050812.aspx. [Accessed: 09-May-2012].

[38]   T. RÖLLE, F.-K. Bruder, T. FÄCKE, M.-S. Weiser, D. HÖNEL, and C. Diedrich, "SELECTION METHOD FOR ADDITIVES IN PHOTOPOLYMERS," .

[39]   M. Makowski, "Modeling at PPG Industries - An Interview with Michael Makowski, Scientific Computing Group," 2003. [Online]. Available: http://accelrys.com/resource-center/case-studies/archive/studies/ppg_int.html. [Accessed: 08-May-2012].

[40]   T. Yamauchi et al., "Method of manufacturing silicide layer for semiconductor device," .

[41]   A. Strachan, G. Klimeck, and M. Lundstrom, "Cyber-Enabled Simulations in Nanoscale Science and Engineering," *Computing in Science & Engineering*, vol. 12, no. 2, pp. 12-17, Mar. 2010.

[42]   G. Klimeck, "Impact of modelling." May-2012.

[43]   JRC. European Commission, "The 2011 EU Industrial R&D Investment Scoreboard." [Online]. Available: http://iri.jrc.ec.europa.eu/research/scoreboard_2011.htm.

[44]   A. B. Richon, "Mergers and Alliances within Computational Chemistry - Part 2," Mar-2001. [Online]. Available: http://www.netsci.org/Science/Compchem/feature17c.html. [Accessed: 30-Apr-2012].

[45]   A. B. Richon, "A History of Computational Chemistry: Part 2," Mar-2001. [Online]. Available: http://www.netsci.org/Science/Compchem/feature17a.html. [Accessed: 30-Apr-2012].

[46]   T. Parish, "The Technology Value Pyramid," Washington: Academies Press, 1998.

[47]   R. Gilbert, "Nanotechnology ROI at e2v Technologies - an Interview with Dr Richard Gilbert, Principal Scientist (Biosensors)." 2004.

[48]   Committee on Integrated Computational Materials Engineering, "Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security," Washington, 2008.

[49]   C. Magee, "Towards quantification of the Role of Materials Innovation in overall Technological Development," 2009.

[50]   "Retooling Manufacturing: Bridging Design, Materials, and Production," Washington, 2004.

[51]   T. Oden et al., "Simulation-Based Engineering Science: Revolutionizing Engineering Science through Simulation," May 2006.

[52]   J. Fish, "Bridging the scales in nano engineering and science," *Journal of Nanoparticle Research*, vol. 8, no. 5, pp. 577-594, Sep. 2006.

[53]   J. Allison, M. Li, C. Wolverton, and X. Su, "Virtual aluminum castings: An industrial application of ICME," *JOM*, vol. 58, no. 11, pp. 28-35, Nov. 2006.

[54]   C. Wolverton and J. Allison, "Method of optimizing heat treatment of alloys by predicting thermal growth," .

[55]   E. Joseph, S. Conway, C. Ingle, G. Cattaneo, N. Martinez, and C. Meunier, "D 2 Interim Report: Development of a Supercomputing Strategy in Europe," Jul. 2010.

[56]   CSCS Swiss National Supercomputing Centre, "Annual Report 2010," 2011.

[57]   High Performance Computing Centre, "HLRS BI-ANNUAL REPORT 2008/ 2009," Stuttgart, 2010.

[58]   CSC – IT CENTER FOR SCIENCE LTD, "CSC Annual Report 2010," 2011.

[59]   Department for Business, Innovation and Skills, "THE ALLOCATION OF SCIENCE AND RESEARCH FUNDING 2011/12 TO 2014/15," Dec. 2010.

[60]   D. Tildesley, "A Strategic Vision for UK e-Infrastructure," 2011.

[61]   G. Fitzgerald, "High-Throughput Computation for Materials Discovery," San Diego, 2010.

[62]   C. Wadia, "New Commitments Support Administration's Materials Genome Initiative," May-2012. [Online]. Available: http://www.whitehouse.gov/blog/2012/05/14/new-commitments-support-administration-s-materials-genome-initiative. [Accessed: 15-May-2012].

[63]   M. Sarwar, "Designing New Fuel Cell Catalysts: A theoretical and experimental approach." London, May-2010.

[64]   J. Gavartin et al., "Exploring Fuel Cell Cathode Materials: A High Throughput Calculation Approach," 2009, pp. 1335-1344.

[65]   G. Fitzgerald, "Accelrys Community: Accelrys Blog: High-Throughput, Quantum Mechanics, and Lithium Ion Batteries." Apr-2010.

[66]   E. J. Maginn, "Transforming Molecular Simulation into a Mainstream Chemical Engineering Tool," *Chemical Engineering Progress*, vol. 105, p. 12, Jun-2009.

[67]   C. J. Kuehmann and G. B. Olson, "Computational materials design and engineering," *Materials Science and Technology*, vol. 25, no. 4, pp. 472-478, Apr. 2009.

[68]   R. McDonald, "Inside P&G's digital revolution," *McKinsey Quarterly*, Nov-2011.