Institution: University of York

Unit of Assessment: 9 - Physics

Title of case study: Predicting the properties of materials with first-principles electronic structure software (CASTEP)

Period when the underpinning research was undertaken: 2000-2020

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	Role(s) (e.g. job title):	**Period(s) employed by submitting HEI:**
Matt Probert	Professor	Oct 2000-present
Phil Hasnip	Research Software Engineer Fellow	Apr 2005 to Jul 2006; Dec 2007-July 2012, Jan 2013 - present

Period when the claimed impact occurred: 2014-2020

Is this case study continued from a case study submitted in 2014? N

1. Summary of the impact (indicative maximum 100 words)

York researchers Probert and Hasnip comprise one third of the core development team of CASTEP, a high-performance software package using quantum mechanics to predict materials' physical, chemical and electronic properties. It supports industries worldwide to discover and optimise materials for technological applications, guide experimental design and interpret experimental data in the chemical, pharmaceutical, semiconductor, automotive and aerospace industries. During the impact period, CASTEP was sold (via BIOVIA Inc) to over 500 industrial sites, including [text removed for publication], with revenue of [text removed for publication] during the census period. For example, one business used CASTEP to design new materials for fuel cells, catalysts and batteries.

2. Underpinning research (indicative maximum 500 words)

The properties of materials and chemicals can be predicted from quantum mechanics, but wavefunction-based approaches are extremely computationally demanding. Density functional theory (DFT) is a Nobel-prize-winning reformulation of quantum mechanics using the electron density as the fundamental quantity instead. This transforms the computational complexity of the problem, allowing predictive quantum mechanical simulations of real materials.

CASTEP is a user-friendly, robust, efficient and accurate general-purpose materials modelling code built upon DFT. The seminal CASTEP paper, describing its development, was published in 2002 and Probert and Hasnip are co-authors. Hasnip was lead author on the most recent publication of CASTEP's methodology and simulation capability [3.1], with Probert a co-author. CASTEP uses DFT to predict the properties of any material [3.1], and supports experimental studies by helping to interpret experimental data. CASTEP is widely used in industry; some industrial research has been reported in the open literature and patent applications (see [5.5] for examples), but much industrial work is commercial-in-confidence. CASTEP is cited in support of 260 patents worldwide, including 36 during the census period.

The underlying research and development of CASTEP began in 1999 with the formation of the CASTEP Developers' Group (CDG), a core team of UK academics of whom Probert and Hasnip comprise one third. CASTEP was released in 2002, with new versions every year incorporating new science capabilities and calculation speed increases. Each CDG member is responsible for specific areas of functionality: Hasnip's primary responsibility is the non-linear equation solvers used (a) to solve the DFT equations themselves, which underpin all of CASTEP's other functionality; and (b) to compute the band-structures, density-of-states and experimental spectra. Probert's primary responsibility is the code for molecular dynamics and methods to predict and optimise the atomic structure of materials. This latter component is also crucial, and virtually all CASTEP materials investigations will start with a structural optimisation calculation. The core CASTEP functionality is summarised in [3.1] & [3.2], and in Fig. 1 (below). CASTEP's reliability was confirmed by a comprehensive 2016 study, comparing results to more computationally demanding methods [3.4]. The York contribution also underpins more advanced functionality, such as the calculation of the phonon dispersion and many important experimental spectra such as Nuclear Magnetic Resonance (NMR), Electron Energy Loss Spectroscopy (EELS) and Raman spectroscopy [3.3], which all use a core algorithm developed by Hasnip.

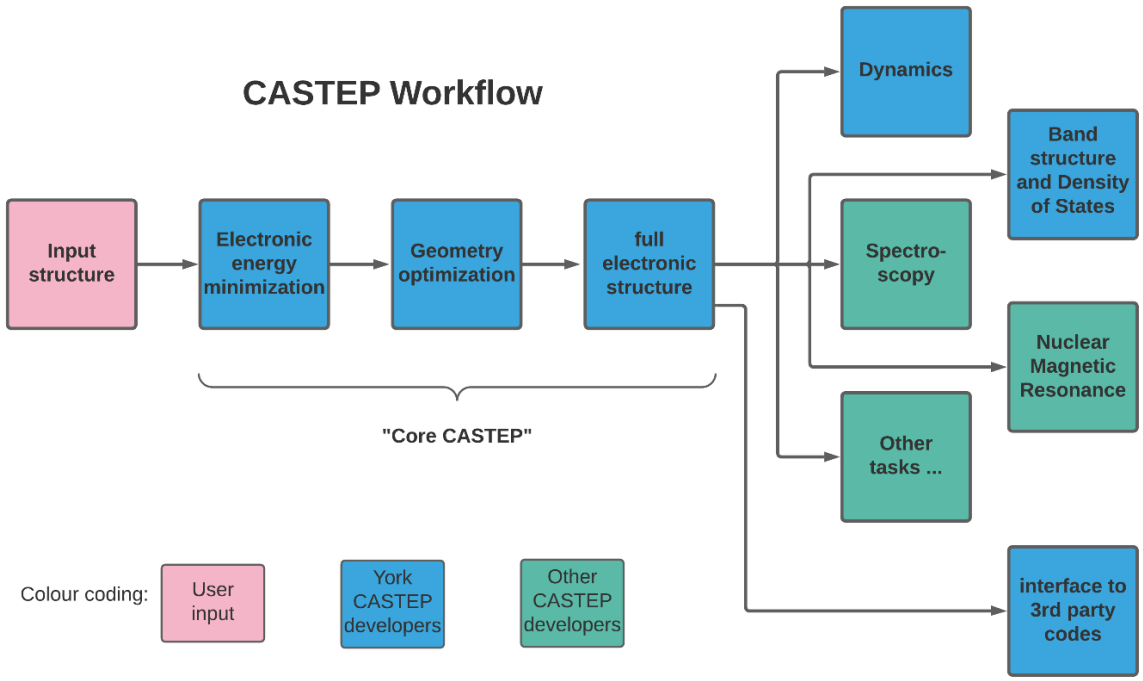

Figure 1: *CASTEP Workflow diagram, highlighting the centrality of York contributions*

The York team continues to develop new functionality in CASTEP, including a fast DFT+U (DFT with a Hubbard U term) implementation (CASTEP v18), spin-orbit coupling & non-collinear magnetism (CASTEP v16), dynamical mean-field theory [3.5], implicit solvation models [3.6] (CASTEP v18), advanced many-body van der Waals modelling (CASTEP v16), and electron-phonon coupling (CASTEP v18), as well as improved parallel performance (CASTEP v18 using hybrid OpenMP-MPI), and porting to emerging computer architectures (e.g. ARM v8 cores; NVIDIA "Tesla" GPGPU). CASTEP has also been interfaced with electronic and thermal transport programs (CASTEP v17), to calculate electrical conductivity, and electronic- and lattice thermal conductivities.

3. References to the research (maximum of six references)

3.1. "Density functional theory in the solid state", P. J. Hasnip, K. Refson, M. I. J. Probert, J. R. Yates, S. J. Clark and C. J. Pickard, *Phil. Trans. R. Soc. A* **372** (2014) 20130270; http://dx.doi.org/10.1098/rsta.2013.0270

3.2. "First principles methods using CASTEP", S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson and M. C. Payne, *Zeitschrift für Kristallographie* **220** (2005) 567 https://doi.org/10.1524/zkri.220.5.567.65075

3.3. "Electron and vibrational spectroscopies using DFT, plane waves and pseudopotentials: CASTEP implementation", V. Milman, K. Refson, S.J. Clark, C.J. Pickard, J.R. Yates, S-P. Gao, P.J. Hasnip, M.I.J. Probert, A.Perlov and M.D. Segall, *J. Mol. Struc: THEOCHEM* **954** (2010) 22; http://dx.doi.org/10.1016/j.theochem.2009.12.040

3.4. "Reproducibility in density functional theory calculations of solids", K. Lejaeghere, *et al.* including P. J. Hasnip and M. I. J. Probert, *Science* **351** (2016) 6280; http://dx.doi.org/10.1126/science.aad3000

3.5. "Many-body renormalization of forces in f-electron materials", E. Plekhanov, P. Hasnip, V. Sacksteder, M. Probert, S. J. Clark, K. Refson and C. Weber, *Phys. Rev. B* **98** (2018) 075129; http://dx.doi.org/10.1103/PhysRevB.98.075129

3.6. "DL_MG: A parallel multigrid Poisson and Poisson-Boltzmann solver for electronic structure calculations in vacuum and solution", J. C. Womack, L. Anton, J. Dziedzic, P. J. Hasnip, M. I. J. Probert and C.-K. Skylaris, *Journal of Chemical Theory and Computation* **14 (3)**, (2018) 1412 http://dx.doi.org/10.1021/acs.jctc.7b01274

All references are peer-reviewed publications. [3.4] has been submitted to REF2021. [3.1-3.4] were supported by UKCP HEC (EPSRC funded); [3.5] funded by EPSRC; [3.6] funded by the ARCHER eCSE programme.

4. Details of the impact (indicative maximum 750 words)

CASTEP has been co-developed at the University of York since Probert arrived in October 2000. The first commercial release was in 2002, and it has steadily grown in functionality, capability and sales since then. Since Jan 2015 alone, more than 500 company sites across 46 countries have purchased CASTEP, generating commercial sales of GBPxxx during the census period [5.1] despite competition from other commercial codes, and free open-source alternatives. CASTEP is an exemplar of excellent research software engineering, and an important part of the materials modelling e-infrastructure [5.2] in academia and industry, with the York team's contributions often singled out for particular praise:

"As a frequent and avid user of CASTEP, I have admired [the York team's] work for many years. CASTEP is widely used to validate the observed structure and simulate electron energy loss (EELS) spectra for comparison with experiment" (Director, SuperSTEM Laboratory [5.3])

CASTEP's success is due to its ease of use, high-performance and the continued code innovations, with new functionality (e.g. Hasnip & Probert: many-body dispersion; electron-phonon coupling (with Clark, Durham)), new algorithms (e.g. Probert: improved geometry optimisation method) and better performance (e.g. Hasnip: optimised parallelism; fast Hubbard U potential). CASTEP is sold commercially via BIOVIA's "Materials Studio" product, and CASTEP industrial customers include [text removed for publication] [5.1].

An EU-funded 2016 study of industrial research [5.4] found CASTEP (and similar modelling codes) generated an average return-on-investment (ROI) of 8:1, and cost savings of EUR 100,000-50,000,000 per business using it. CASTEP use supports a growing number of patents (36 during the census period). Examples which use York's contributions include [5.5]:

Honda, who used CASTEP to develop "an easy-to-machine, aluminum-based alloy... to achieve high rigidity without containing hard particles" [5.5]. CASTEP simulations predicted the Young's modulus of various alloys of Al doped with Cu, Zn, Ag and Li (using the electronic energy minimisation and geometry optimisation methods of the York team); the best alloys were subsequently synthesised, and the experimentally measured Young's moduli were found to be in excellent agreement with the CASTEP predictions (mean deviation 1.1%; max. 1.7%).

Toshiba Memory Corporation (now Kioxia; a spin-off from Toshiba) used CASTEP simulations to design a new resistance change memory (ReRAM) device, by modelling the interfaces between transition-metal chalcogenide layers (the key resistance-change layers) and the insulating and conducting components of the device [5.5]. CASTEP was used to model the stress between the component materials for a range of possible interfaces and candidate materials, which was used to design several low-stress device heterostructures. (These calculations use the electronic energy minimisation and geometry optimisation methods of the York team.)

Panasonic used CASTEP to determine a new "Fluorophore and light-emitting device" with improved colour rendering and reproducibility. The key material component is a $La_3Si_6N_{11}$ phosphor doped with Ce, and Panasonic used CASTEP to determine not only the structure and properties of $La_3Si_6N_{11}$, but also which substitutional sites favour Ce incorporation [5.5].

Semiconductor Energy Laboratory Co. used CASTEP simulations to support a wide range of patent applications in the census period, including thin-film transistors for imaging and display technologies. The designs use $InGaZnO_4$ as the transistor oxide layer, and the ability to fabricate high-quality samples with low grain boundaries, all underpinned by CASTEP simulations of $InGaZnO_4$ crystals to determine the optimal crystal structure and, critically, the low-energy cleavage planes [5.5]. The lowest energy is found to be a (001) plane between two Ga—Zn—O layers, which directly informs the device design and manufacturing process. The CASTEP predictions are confirmed by X-ray diffraction of experimental samples.

It is impossible to perform any CASTEP simulation without using York code, and York-led developments underpin all of CASTEP's use and impact (see Fig. 1). A typical study starts with estimated atomic coordinates & lattice vectors and uses a bespoke optimisation method

(Hasnip) to determine the electronic energy, atomic forces and cell stresses. A geometry optimisation algorithm (Probert) moves the atoms & lattice vectors to minimise the forces and stresses. All other tasks build on this foundation, e.g. simulations of the phonon, NMR, EELS or IR spectra (all using Hasnip's core algorithm). Probert also developed CASTEP's "molecular dynamics," using forces and stresses to evolve the system in time, e.g. to study dynamical processes, or temperature effects such as phase changes.

The CASTEP developers train and support scientists with many initiatives, introducing industry-leading simulation techniques and computational methods to quickly and effectively "up-skill" researchers. Of particular note are the annual user training workshops co-led by Hasnip, which have trained over 300 scientists since 2014, approximately 10% from industry (e.g. 70 participants in 2016; of the feedback respondents, 3 of 29 were commercial licence holders [5.3]). The workshops aid in expert knowledge integration and exchange between academia and industry, and have led to CASTEP being used by more research groups, including industrial groups.
"I learnt a lot about the application of DFT and how to use CASTEP. I am much more confident in how to interpret DFT results and would look to use this technique... in the future... [I am] obtaining access to CASTEP as a result" (2016 attendee [5.3])

One 2015 workshop attendee was the Principal Scientist and Research Manager of a major chemicals and sustainable technologies company. The company uses CASTEP extensively across their portfolio: *"The CASTEP code has been used widely across [the business] to guide experimental work by identifying promising materials for a wide range of applications including metal alloys for fuel cells and complex oxides for automotive catalysis and novel battery materials. Results from CASTEP calculations have contributed to generation of new IP as well as being publications in the open literature."* [5.6]

For example, the company use the molecular dynamics capability (Probert) to investigate proton mobility in Pt hydroxides [5.6]. The company chose CASTEP because "it is easy to use and provides reliable results", and examples of their work built on York's CASTEP contributions include: DFT+U approach (Hasnip) to study complex oxides, leading to new IP in Li-Ion battery materials; spin-orbit coupling (Hasnip) functionality to investigate thermoelectric materials; and the aforementioned MD study of proton mobility. They also use CASTEP NMR simulations extensively (employing Hasnip's core equation solver). NMR simulations are used to assign chemical shifts to complex experimental spectra for materials ranging from zeolites to pharmaceuticals, and they find that *"NMR crystallography has been the only approach that has enabled difficult characterisation problems to be resolved."* The company also frequently computes IR shifts (using the same solver by Hasnip) to aid the interpretation of experimental DRIFTS data.

More recently, York led the *ab initio* simulation development for the GBP20,400,000 Government-Industry-Academia project "Advanced Digital Design of Pharmaceutical Therapeutics" (ADDoPT; https://www.addopt.org/), whose industrial partners include AstraZeneca, Pfizer, Glaxo Smith Kline, Bristol-Myers Squibb, Process Systems Enterprise Limited, Britest Limited and Perceptive Engineering Ltd. The York team implemented an advanced "many-body dispersion" method to model the long-range dispersion interactions critical to understanding crystallisation processes in pharmaceutical compounds [5.7]:
"Long-range dispersion forces contribute up to 60% of the total energy in larger weight, organic molecules, therefore modelling them accurately is essential... [The many-body dispersion] code enhancement [is of] enormous benefit… the impact of calculating more accurate chemical and physical properties includes crystal surface structure; crystal cluster stability; conformational searches and analyses and disordered structures and impurity analysis, all of which are components along the drug design pipeline." (ADDoPT partner letter [5.3])

These advanced methods have kept CASTEP at the forefront of *ab initio* simulation software:
"the latest technologies are regularly updated and that the science is extremely well validated, and this keeps it at the forefront of the product portfolio" (BIOVIA [5.1])

An independent pharmaceutical modelling review recently found CASTEP had the joint-highest feature set amongst pharmaceutical modelling software, and gave significantly improved NMR predictions compared to molecular-based modelling [5.8]. CASTEP customers include all of the major multinational companies in the pharmaceutical sector [5.1] because of its advanced functionality, reliability and ease of use:
"For pharmaceuticals, CASTEP provides a simple and reliable method of allowing us to test between salt or co-crystal models using the calculated NMR spectra" [5.6]

CASTEP use is also widespread across chemical and manufacturing sectors, and its ease of use and cutting-edge functionality make it a cornerstone of BIOVIA's Materials Studio:
"CASTEP is an extremely important part of the atomistic modeling offering by BIOVIA [with many] applications in industrial R&D in: chemical, petrochemical, pharmaceutical, electronics, automotive [and] aerospace" (BIOVIA Scientific R&D Development Senior Manager [5.3])

CASTEP's success is influencing UKRI policy and supports key e-infrastructure projects. It was highlighted in the EPSRC Impact Report 2016-17 as evidence of "The UK's Research Leadership in EPS Disciplines," [5.9] showcased in the EPSRC Delivery Plan 2019 [5.9], and used as a benchmark to determine the GBP42,000,000 ARCHER2 procurement [5.10].

5. Sources to corroborate the impact (indicative maximum of 10 references)

5.1. BIOVIA letter of support; Cambridge Enterprise Ltd letter of support
5.2. I.J. Bush, *Fortran Forum*, **29** 3 (2010). https://doi.org/10.1145/1883575.1883576; T. Hey and M.C. Payne, *Nature Phys.* **11** 367 (2015). https://doi.org/10.1038/nphys3313
5.3. Letter of Support from the Director of SuperSTEM for Hasnip's EPSRC RSE Fellowship (2017); 2016 CASTEP Workshop report; Letter of Support from the Head of Formulated Products, [ADDoPT Industrial Partner] for ADDoPT eCSE project (2018); Letter of Support from BIOVIA's Scientific R&D Development Senior Manager for Hasnip's EPSRC RSE Fellowship (2017)
5.4. Goldbeck Consulting Ltd and MRIGlobal, "Economic Impact of Materials Modelling: Indicators, Metrics, and Industry Survey", 2016, https://materialsmodelling.com/2016/02/08/economic-impact-of-materials-modelling/
5.5. Activities described in patent applications underpinned by CASTEP simulations; e.g. **Honda:** "Aluminum-based alloy", H. Watanabe & S. Hirosawa, US20190024220A1 (filed 20th Jan. 2017; US patent still under consideration, Japanese patent JP2017128780 issued 28th Oct. 2020); **Toshiba Memory Corp:** "Memory Device", K. Suzuki & K. Tominaga, US9893280B2 (filed: 8th July 2015; issued 13th Feb. 2018); **Panasonic:** "Fluorophore and light-emitting device", M. Nitta, Y. Inada and N. Nagao, EP3480280A1 (filed 13th Jan. 2017; issued 8th May 2019); **Semiconductor Energy Laboratory Co Ltd:** "Imaging Device", S. Yamazaki, M. Sakakura & Y. Kurokawa, US9324747B2 (filed 9th Mar. 2015; issued 26th Apr. 2016).
5.6. Letter of support from Principal Scientist and Research Manager, [chemicals and sustainable technologies company]
5.7. "Off-the-shelf DFT-DISPersion methods: Are they now "on-trend" for organic molecular crystals?", D. Geatches *et al.* (incl. P. Hasnip & M. I.J. Probert), *Journal of Chemical Physics*, **151** (4) 044106 (2019). https://doi.org/10.1063/1.5108829
5.8. "Periodic DFT Calculations—Review of Applications in the Pharmaceutical Sciences", A. H. Mazurek, Ł. Szeleszczuk and D. M. Pisklak, *Pharmaceuticals*, **12** 415 (2020). https://doi.org/10.3390/pharmaceutics12050415
5.9. EPSRC Impact Report 2016-17 https://epsrc.ukri.org/newsevents/news/impactreport201617/; EPSRC Delivery Plan 2019 https://epsrc.ukri.org/about/plans/dp2019/
5.10. ARCHER2 procurement benchmarks https://github.com/hpc-uk/archer-benchmarks