Workshop and CCP Proposal for Agent Based Modelling and Simulation

Prof Chris Greenough
STFC Rutherford Appleton Laboratory
email: christopher.greenough@stfc.ac.uk
tel: +44 1235 445307
Dr Paul Richmond
University of Sheffield
email: p.richmond@sheffield.ac.uk
Dr Rob Allan
STFC Daresbury Laboratory
email: robert.allan@stfc.ac.uk
tel: +44 1925 603207

1 Introduction

Most computational modeling research describes systems in equilibrium or as moving between equilibria. Agent based modeling, however, using simple rules, can result in far more complex and interesting behavior. This could be particularly relevant to studies of complex and "organic" systems, such as occur in biology which are "on the edge of chaos".

The three ideas central to agent based models are: (i) social agents as objects; (ii) emergence; and (iii) complexity.

Agent based models consist of dynamically interacting rule based agents. The systems within which they interact can create real world like complexity. Agents in a system can be recognised if they satisfy the following criteria.

- They are intelligent and purposeful;
- Situated in space and time.

Agents typically reside in networks and in lattice like neighborhoods. The location of the agents and their responsive and purposeful behaviour need to be encoded in algorithmic form in computer programs. The modelling process is best described as "inductive". The modeller makes those assumptions thought most relevant to the situation at hand and then watches phenomena emerge from the agents' interactions anc compares this with observations of "reality". Sometimes the result is an equilibrium, sometimes it is an emergent pattern (attractor), sometimes, it is purely chaotic behaviour.

Agent based models can complement more traditional computational methods. Where continuum analytic methods enable us to characterise the equilibria of a system, agent based models allow the possibility of generating those equilibria. This "generative" contribution may be the most significant of the potential benefits of agent based modelling. Agent based models can explain the emergence of higher order patterns and network structures in many systems. Such emergent behaviour is noted in studies of terrorist organisations, the Internet, power law distributions in the sizes of traffic jams, wars, stock market crashes, and social segregation that persists despite populations of tolerant people. Agent based models also can be used to identify lever points, defined as moments in time in which interventions have extreme consequences and to distinguish among types of path dependency.

Rather than focusing on stable states, the models consider a system's robustness - the ways that complex systems adapt to internal and external pressures so as to maintain their functionalities. The task of harnessing that complexity requires consideration of the agents themselves – their diversity, connectedness and level of interactions.

There are many applications of agent based technology in the mathematical, biological, physical and social science communities. Multi-agent systems are however very complex applications to program, implement and optimise for large populations of agents. The optimisation of agent interactions and the management of initial data and outputs for large scale simulations are but two areas that need research and development efforts.

There are clearly many areas within agent based methods that might benefit from collaborative activity: serial and parallel implementation of agent based applications, verification and validation of agent based models, standards for the exchange of agent based models to name a few.

During a one day awareness raising workshop in Leeds on 15/6/2010 we aim to explore the state of the art in agent based modelling and simulation from both the computational science aspects and its applications. We identify some issues with current methodologies and implementations and consider future directions appropriate to tackle research into complex dynamical systems using agent based modelling on high performance computers.

For some background information on ABMS software see `http://www.grids.ac.uk/NWGrid/ABMS/`.

2 UK Community

The currently active UK community is identified here in terms of principal application areas. We will add to this list during and after the workshop.

Cancer Research: Paul Bates and Katie Bentley, CR-UK.

Cognition: Aaron Sloman (Birmingham)

Complexity: Chris Bud (Bath)

Crowd/ Pedestrian Modelling: Paul Richmond (Sheffield), CASA group (UCL)

Ecology: MASS (below)

Economic Modelling: Chris Greenough et al. (STFC)

Evolution and Migration: Bill Sellers (Manchester)

GPUs: Paul Richmond (Sheffield)

Parallel Computing: Chris Greenough et al. (STFC)

Social Simulation: NeISS, MASS, SIMSOC (below)

Systems Biology: Rod Smallwood, Mike Holcombe et al. (Sheffield)

SIMSOC: `https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=SIMSOC`

MASS: Multi Agent Systems and Simulation Research Group (Leeds) `http://www.geog.leeds.ac.uk/groups/mass/`

CASA: Centre for Advanced Spatial Analysis (UCL) led by Mike Batty `http://www.casa.ucl.ac.uk/`

CRESS: Centre for Research on Simulation in the Social Sciences `http://alife.ccp14.ac.uk/cress/research/simsoc/cress.html` led by Nigel Gilbert (Surrey) (very old Web site)

NeISS: National e-Infrastructure for Social Simulation – a JISC funded project led by Mark Birkin (Leeds) `http://www.neiss.org.uk`

3 Awareness Raising Workshop, Leeds, 15/6/2010

3.1 Attendees

The full list of registered attendees was as follows.

Salem Adra - s.adra@sheffield.ac.uk
Rob Allan - robert.allan@stfc.ac.uk
Joana Barros - j.barros@bbk.ac.uk
Mark Birkin - m.h.birkin@leeds.ac.uk
Mark Burkitt - m.burkitt@sheffield.ac.uk
Lee-Shawn Chin - shawn.chin@stfc.ac.uk
Maria Chli - m.chli@aston.ac.uk

Shauna Coyle - shauna@leeds.lubs.ac.uk
Andrew Crooks - acrooks2@gmu.edu
Andrew Evans - a.j.evans@leeds.ac.uk
Chris Greenough - christoper.greenough@stcf.ac.uk
Mohammed Hassouna - mohammed.hassouna@brunel.ac.uk
Mike Holcombe - m.holcombe@dcs.shef.ac.uk
Helen Hughes - h.hughes@leeds.ac.uk
Anders Johansson - a.johansson@ucl.ac.uk
Twin Kamakharm - twin@dcs.shef.ac.uk
Mariam Kiran - m.kiran@dcs.shef.ac.uk
Tomasz Kulakowsi - tomasz.mp.kulakowski@gmail.com
Jason Lander - j.lander@leeds.ac.uk
Michael Luck - michael.luck@kcl.ac.uk
Ed Manley - ucesejm@ucl.ac.uk
Katarzyna Matuszewska - kate.matuszewska@gmail.com
Peter McBurney - mcburney@liverpool.ac.uk
Phil McMinn - p.mcminn@sheffield.ac.uk
Vijayanarasimha Pakka - vpakka@dmu.ac.uk
Sang Joon Park - sjp@cs.nott.ac.uk
Ian Phillips - gy09ip@leeds.ac.uk
Jorge Ramirez - gyjr@leeds.ac.uk
Paul Richmond - p.richmond@dcs.shef.ac.uk
Joan Serras - j.serras@ucl.ac.uk
Peer-Olaf Siebers - pos@cs.nott.ac.uk
Richard Snape - jsnape@dmu.ac.uk
Richard Sibly - r.m.sibly@rdg.ac.uk
Andy Turner - a.g.d.turner@leeds.ac.uk
Cobus van Rooyen - cobus_van_rooyen@hotmail.com
Alex Voss - avoss@cs.st-andrews.ac.uk
Christopher Watts - c.watts@surrey.ac.uk
Martyn Winn - martyn.winn@stfc.ac.uk
Tao Zhang - txz@cs.nott.ac.uk

3.2 Workshop Programme

Speakers at the workshop included the following.

- Peter McBurney (Liverpool) – the challenges to computer scientists arising from ABMS
- Richard Sibly (Reading) – agent based modelling techniques and validation
- Michael Luck (King's College, London) – issues and challenges in agent based modelling

- Mike Holcombe (Sheffield) – X-Machines in FLAME and applications in systems biology
- Andrew Crooks (University College, London) – geo-spatial modelling/ crowd modelling
- Alex Voss (St. Andrews) – issues with multi-threaded parallelisation and application of Repast Simphony to modelling the population of Taiwan
- Salem Adra (Sheffield) – Automated Discovery of Emergent Mis-behavior in Agent Based Models
- Paul Richmond (Sheffield) – applications on GPUs and modelling of crowd behaviour
- Mariam Kiran (Sheffield) – EURACE, economic modelling
- Anders Johannson (University College, London) – GENESIS, pedestrian modelling and tracking
- Andy Turner (Leeds) – large scale social simulation in Java and the NeISS project

3.2.1 Chris Greenough (STFC) – introduction.

The aim of the workshop is to get the UK community together, look at issues of agent technology, parallelisation, etc. EPSRC are looking to support new communities, so we will explore the issues, challenges and hints of some applications. What can we pull together and bid to EPSRC? What will be the impacts on the community and UK plc?

3.2.2 Peter McBurney (Liverpool) – Agent Based Modelling: Challenges for Computer Science

For slides see `http://www.grids.ac.uk/Complex/Leeds.15.6.10/McBurney.pdf`.

Scientists might have a different view to other people, so I present some of the "post modern" objectives from social sciences and policy making. Firstly some definitions and challenges.

Complex adaptive systems (CAS) are systems of interacting entities which have some intelligence and autonomy. States and relationships are dynamic. This includes my own work in modelling malaria – parasites, hosts and humans evolve, build up resistance, respond to events, humans adapt and change behaviours, inetlligent? Policies may not always be followed, e.g. for financial or other reasons, e.g. too far to walk, religion, too many children (some have flu and exhibit same symptoms)...

Multi agent systems (MAS) including ABM represent systems using software entities. Not same as rational expectation models in economics. Maths is different – latter assumes entities know about the model (bizarre?). Not same as conjoint models of marketing.

Objectives of modelling. Scientists want quantitative measures, social scientists might just want relationships. Might just want some better intuition, e.g. as part of a decisino support tool.

Other objectives. In some cases behaviour depends on knowledg of the model, which contains assumptions. Referred to as "performative utterances". Models built with competitors, e.g. hedge fund model. Policy modelling helps discussions, identifies stakeholders. Epidemiology example used for forecasting impact of policies. Models of avian flue etc. at Uni. Melbourne. Research group provides forecasts but not the models, so its an independent prediction.

Formal theories of simulation lacking. Template models abstract away details to be a generic model.

Challenges. Expectations vary across disciplines. Agent Oriented Software Engineering (AOSE) immature. Upscaling and parallelisation – consider blackboards. Coupled models – e.g. climate plus economy, requires standards for interchange of information. Calibration, verification and validation. Need two independent checkers to validate design.

Always trade-offs in s/e. Need to manage expectations and agree on design. Quality vs. cost vs. speed. Good s/e practice. Application scientists not fully aware of all this.

Need to understand how models are used – coordination? Need lots of use cases "thick studies". Automated statistical or sensitivity analysis. Link parameterisation with analysis – feedback. Many of these things not yet applied. Incorporate crowd sourcing. Prediction markets – e.g. U. Iowa lots of local people making wider predictions – aggregating local specialist knowledge.

Q. Birkin. About q vs. s vs. c. Are these socially determined in some way? Are we doing problems 10x harder every 2 years? Is this an expectation. Not necessarily in mind of software engineers.

Q. Voss. CS people are involved, they might want to build the biggest (most interesting?) model, software engineers may be more fixed on practicalities. What is publishable varies with discipline.

Q. Siebers. Does each community need a theory of simulation because they have different goals? Actually we don't have it but it might be good to have. Similar problems with generic definitions. Could all be domain dependent. Then problem with coupling models.

Q. Holcombe. On work by Grimm et al. on standards. Rather loose so far. This is actually becoming a big issue as its neccessary to understand the model to know if the results are any good. This is stuff which should be available separate from the publication, i.e. archived somewhere.

3.2.3 Richard Sibly (Reading) – AMBs in Ecology

For slides see `http://www.grids.ac.uk/Complex/Leeds.15.6.10/Sibly.ppt`.

Calibration, verification and validation – greatest of these is validation. This is a view from ecology.

This is not such a small field, there are over 1,000 publications.

I am a population biologist, attending various workshops, discuss how to assess affect of chemicals on populations of animals. Big business. Pesticide safety directorate in UK, granting licenses etc. Risk assessment. EU (2000) guideline to protect populations rather than individuals, but all data comes from individuals – implies extrapolations.

Population dynamics. Why do numbers change over time? Ideal model. Euler-Lotka equation. Logistic population growth. Chemicals make this more complex. Stress affects population growth. ABMs, Roskilde (2003) and York workshop (2004). Developed 5 steps for population risk assessment. Still many unknowns. Extrapolate species with body mass. How do farmers apply chemicals? What do animals eat?

Chris Topping's spatially explicit vole model. Agent specification, emergent properties – 60,000 lines of C++. Gave similar results to classical model. Showed that ABM could be parameterised but not understood. Animals tracked on daily basis, so individual behaviour is known. Classical models are simple but can't be parameterised – can't work out how chemical will affect density dependence.

Volker Grimm is trying to gain acceptance for ABMs in ecology. Book and Science (2005) article. EU grant, CREAM, 20x Ph.D. students etc. `http://cream-itn.eu`. Can we make ABMs credible? Necessary to verify and validate so that results can be accepted by risk managers. Issues of blame. Currently using NetLogo as its easy for naive users. Another grant headed by Mark Beaumont to look at validation.

Classical evaluation of models, using R^2 methods. Bayesian approach for ABMs looks similar, calculates likelihood that the model is true. Bayesian methods are mostly in the form of MCMC, requires us to know the likelihood function, but can't write this down explicitly for ABM. Approx. Bayesian computation used instead. Approximate Bayesian Computation (ABC) allows estimation of likelihood. Bayes factor, ration of probabilities of two models – critical value of 10. Tom Kulakowsi running on 24 proc system. Start with 1,000 prior values and retain the ones which closely match the data.

Need other models to compare with to estimate likelihood.

`http://www.grids.ac.uk/Complex/Leeds.15.6.10/Beaumont_abc.pdf`.

Q. Chli. What is used for calibration? How is it done? Crucial to compare with other models.

3.2.4 Michael Luck (King's College, London) – Agent Based Modelling

For slides see `http://www.grids.ac.uk/Complex/Leeds.15.6.10/Luck.pdf`.

A broad perspective. I am from a CS multi-agent systems background. Technologies – design metaphor – simulation.

3 examples from AgentLink `http://www.agentlink.org` looked at commercial uses, e.g. EuroBios and SCA Packaging. Optimised factory got significant return, etc. Agent based model: ships and ports to provide a decision support tool – i-Scheduler. See slides for details.

In a bigger picture, agent oriented softeware, cognitive level system. Working with MoD to model effect of stimulants on pilots. Belief-design-intention (BDI) model. Jack Intelligent Agents Toolkit. These are other modelling activities.

Lost Wax went bust, now Aerogility `http://www.aerogility.com` Aerospace after market. RR sell engine plus service, goal to keep planes flying. Provide key operating metrics for contract analysis. Do reasoning about contracts. Policy, behavious, regulations. Establish, renew, terminate contracts – what are the effects?

Issues include granulatity. Cognitive approaches not typically used. But agents might be cogintive elements, individuals or teams. Need to link to real world. Process modelling – inform design of systems. What techniques? Methodological, agent oriented s/e. Higher level systems. Understanding emergence of norms. Relationship of implicit and explicit norms.

It is important when thinking about ABMS to not just focus on a limited range of topics. Its a big area with a lot of issues, contributes to other areas of science.

Q. more about cognitive elements – how are beliefs and goals included? Individual mental components. Marvin Minsky, society of mind. Can have agents of agents.

3.2.5 Mike Holcombe (Sheffield) – Biological Modelling and Validation with FLAME

For slides see `http://www.grids.ac.uk/Complex/Leeds.15.6.10/Holcombe.pptx`.

Its good to see many young people at this workshop. Work with biologists is interesting. Important to deliver information and understand core facts. CS people like to abstract from the mess, biologists however find the mess interesting! Location and geometry is important, as is physics, e.g. forces on cells. Validation is by comparison with experiments, often expensive and time consuming.

Pharoes ants, M. pharaonis are a pest. Interested in geometry of pheramone trails. Bifurcation angle is approx 57 degrees and regular, why? Rules based on physiology – view (detection via antenna), can't see other route in some cases. Blobs of pheromone and ants both agents. Simulations done on Sheffield Iceberg cluster (White Rose Grid).

A simple chemical reaction, consider Brownian motion plus reactions. Often cell assumed to be a bag of chemicals, but this is nonsense. Cells are internally highly organised. Geometrical relationships are important.

Model of NF-κB transctiption factor. Receptors on surface of cell wall. Cytokenes relased on detection of infection triggers reactions. Stimulation pathways. Adaptive immune system, basis of vaccination, also involved in cancer treatment. Very complex system and not well understood, most researched system. Model basics [see slides] led to a new discovery – Pogson et al. PLOSOne 3:6 (2008).

Epithelial tissue models all aspects including physical pressure. Used to simulate wound healing using stem cells [video in slides, Phil McMinn] Sun et al.

Salem Adra working on roles of TGF-β1 in wound healing, influences cell division and migration.

FLAME framework based on X-machine Eilenberg (1974), a mathematician. Like a Turing machine. Implemented using message boards and simple XML specification with automatic model generation. Hierarchical modelling, e.g. add external solvers etc. COPASI integration, called as a function [block diagram from slides]. Object oriented approach not used, wanted it to be parallelisable. Verification is very important. Testing agents and beginning to test overall model. X-machines important here. Biologists also seem to understand them.

Output analysis is important, there is a vast amount of data generated so how is emergent behaviour identified, extract information. DAIKON `http://pag.csail.mit.edu/daikon` uses machine learning techniques to look for invariant patterns. Can also reveal bugs (e.g. values out of range). CS can help!

ABM complements other modelling approaches. Models can be extended easily, layered, multiscale. See `http://www.flame.ac.uk`.

Q. Sibly. About verification, is it being automated? From s/e background, so raising design level as high as possible. Code generation is important. Users can design concepts. How reliable is the code? Issues with parallelism. Useful to get graph output to check model. Unit testing with CUnit. DAIKON also useful (Salem will show). All ways of addressing challenges. Biologists find visual outputs useful to compare with microscope views.

Q. Winn. Immune response, might use reaction network (Hoffman) model etc. Best model has 76 DEs tuned to fit the data. Small changes of parameters give bizarre effects – not realistic. ABM appoach did sensitivity analysis, how close do molecules have to be to react? Model seems robust. Living systems cope with variations and have evolved over long time.

Q. Was any stochastic simulation used? Yes, but mainly results from biological testing. In some cases – map-kinase – molecules are arranged as a scaffold, knowing the arrangement is crucial.

3.2.6 Andrew Crooks (George Mason Uni. formerly UCL) – GIS and Agent Based Modelling

For slides see `http://www.grids.ac.uk/Complex/Leeds.15.6.10/Crooks.pdf`.

Some stuff about non-published current work. Archeology and disaster management. Some stuff about linking GIS and ABM. GIS does not give decision making, but is for managing data. Can be useful for validating ABMs, e.g. compare with census info.

GIS represents the world as layers, e.g. road, building, car parks, agents. Agents walk between buildings.

Current research on migration of early humans to USA. Found remains in Norfolk Canyon from hunter gatherer populations. Simulated paleolithic habitat – done in OpenSim. How did people get there and when? Solutream Intercontinental Migration (Chris Rouly, Stanford) is a possibility. ABMS tests model, however 1 day of simulation takes 1 second, a big simulation. Need to run many times to test parameters. Currently populations die out, don't know why. Visualisation is difficult, very complicated model.

Other work on humanitarian relief – Haiti earthquake. Lot of Web2.0 technology gives access to instant information and geo-spatial data. Lot of ABM in USA used for this kind of apps plus military. Optimise food distribution, people compete, helpers have to implement security. Epstein + Epstein and Axtell. Example NetLogo model.

Communicating ABMs to other people is important, AgentStreet, see JASSS 12:4. Pedestrian or traffic models. Simple visual traffic simulation, look at how lights affect flow etc. Black box models.

ABMS allow us to explore how individuals interact. Can export questions about current and past events. Spatial data acts as a container for agents. Challenges include validation and linking quantitative and qualitatives methods, data, real time sim, model communication...

3.2.7 Alex Voss (St. Andrews) – Multi-threading RePast Models

For slides see `http://www.grids.ac.uk/Complex/Leeds.15.6.10/Voss.ppt`.

EUAsiaGrid. Academia Sinica, Taiwan. Running parallel models. Aim to model 22M individuals using Repast Simphony framework.

Migration is an important factor in social deveopment of Taiwan. Relationships with China, Japan and Europe have an historical effect on this. Have detailed recent census data for validation, individual level held at Academia Sinica. Restricted variables (anonymisation) used to inform rather than conflic with research goals. EU law means data must be kept in Taiwan. Will run parameter study ensembles on Taiwan Grid.

Doing tests on a simple fertility mortality model (Andy Turner). Analysed performance of single threaded version, then looked at multi threading. JProfiler used.

Serial version spends most time in scheduling code (events are scheduled to agents in RePast). Tried scheduling to a higher level context. This is important for how models are built. Not much information available about using ABMS at this scale. Parallel version now splits over contexts for worker threads, but no initial improvement. Threads blocking on monitors – sections of RePast are not thread safe, so were being synchronised, particularly RandomHelper class (random numbers using Merseene Twister). Scheduling events also a random issue, plus some agent contention. Speed up now 6x on 8x cores [for details see slides].

Lessons learned. RePast often thought of as not suitable for large models, but limitations mostly in model code, structure of model is important. Might be better to iterate over sets of agents explicitly. How to factor lessons back into RePast developers? May set up a blog or make a tutorial.

Questions – what happens if model is more complex? How much can it speed up? Not much expertise in modelling and HPC in social sciences, so looking at commodity computing. Interest in making ABM frameworks usable, perhaps in collaboration with CS.

Q. Matuszewska. Why is RePast thought to be best? Identified by looking at applications (community) some had compared it with other tools. Java was also preferred as a language for encoding the models. Wanted "out of the bag" integration with development environment and user interface enabling social scientists to run the models. Alternatives are however of interest.

Q. Chli. multiple random number generators can be a problem, or other possible issues of multiple threads could change behaviour. Individual runs not repeatable but aggregate output should be same. However population dies out for some reason. Need to look at how realistic model is and test it. Not much info about how people have built models.

Q. Siebers. Repeated runs to encover rare events. Can we believe them? Could be a problem with the model or a real thing?

Q. Sibly. Why modelling 22M agents, why not start small and then scale up gradually. Should be no difference between 10k and 22M? Why model whole population? Birkin. There is literature about this. Migration studies depend on resolution across say 1,000 areas, age groups, etc. Easy to generate a large matrix of possibilities. System is rich and complex.

3.2.8 Salem Adra (Sheffield) – Automated Discovery of Emergent Mis-behavior in Agent Based Models

For slides see `http://www.grids.ac.uk/Complex/Leeds.15.6.10/Adra.ppt`.

Work with Phil McMinn. ABMS are increasingly popular. Bottom up approach with focus on individuals and emerging behaviour. Example codes include FLAME, SWARM, MASON, RePast.

Ephitheliome, part of Physiome, a worldwide public domain project. Driver for testing models. Explores roles of single parameters in complex systems which is impossible in vivo. Other examples include cardiac tissue, EU economy, Pharoe ants.

ABMs predictions can affect organisations or society. Can be complex and contain hidden errors – costly mistakes could be made. However there is no formal testing technique for ABMs. Difference between agents and objects not fully recognised. Jennings, Sycra et al. Unit invocation is different. See table [from slides] on programming approaches. Autonomy and flexibility – agents are self goverened. Modelling to expore, results often un-expected. Large and complex models. Tracing back behaviour hard.

Errors and artefacts, Galan et al (2009) JASSS 12:1 [diagram from slides].

Visual inspection showed errors in the epitheliome model. Due to abnormal lateral force, stem cell stratified and keeps dividing. Needed a rule to prevent this.

Testing strategies: 1) Application of ABM to extreme scenarios; 2) re-implementation.

ADEM – automatic detection of emerging mis-behaviour: 1) reverse engineering to determine invariants e.g. using DAIKON; 2) search based stochastic/ GA black box testing. Look for violation of invariants.

ADEM procedure and case study [see diagram from slides].

Q: Siebers. Can rev engineeering be used for very large models? Very demanding process needs a lot of resources. Maybe should focus on specific events rather than everything.

Q: Sibly. Can you detect errors just with data? Uses machine learning techniques to look for abnormal behaviour.

Q: Watts. Using GA's might be surprising. Actually its a strength. But will also look at other things. Does it beat a random walk?

Q: Kulakowski. What if bug is in comms not values? Framework will highlight errors, not locate them. Developer has to find the bug.

Q: Greenough. How do you know you've not just found emergent behaviour?

3.2.9 Paul Richmond (Sheffield) – applications on GPUs and modelling of crowd behaviour

For slides see http://www.grids.ac.uk/Complex/Leeds.15.6.10/Richmond.ppt.

Modelling pedestrian behaviour, mobile in a continuous space but with discrete time steps. GPU used to get good performance, however programming not easy. Want to abstract away GPU specifics. Extended FLAME, state based agents useful. Have made a data parallel implementation of FLAME. Models are the same, but template process is different. XSLT is used.

Tesla GPU is designed to process streams of data. Very hard to do graphic programming and debug it. New languages now available like CUDA and OpenCL. GPU is used as a co-processor. 1,000s of lightweight threads are composed into blocks (execution units) and mapped to multi-processors. SIMD style. State based implementation is nice for this. Each agent is a thread which executes a code kernel.

Implementation [from slides] 3x techniques for communication: 1) brute force, reads a message by looping over them; 2) spatially partitioned; 3) discrete agent comms, particularl used for pedestrian sim.

Model is hybrid between Reynolds flocking and Helbing social forces. Visualisation done directly from graphics card with very simple animation using levels of detail. Avoid moving data.

Environment collision avoidance. Canary Wharf example. Flow fields for long range navigation. Agent creation and death. Future work. Muli-GPU for large populations, e.g. spatial decomposition.

Q: Voss. What suggests that a model might not work on a GPU. Level of parallelism, need at least 2,000 simple agents. If too complex becomes memory limited.

3.2.10 Mariam Kiran (Sheffield) – Issues with Economic and Social Systems Modelling

For slides see http://www.grids.ac.uk/Complex/Leeds.15.6.10/Kiran.ppt.

Working with economists etc. to build models. What are the issues? How can ABM be of benefit compared with traditional models. Latter usually using DEs or games theory with maximum of 5 pairs. Too small and too many assumptions.

EURACE project had experts in labour, credit, finance, etc. Attempt to model the complete EU economy including housholds, firms, governments, banks. See diagram [from slides]. First time all markets merged together. What are the common agents and how do they interact. How to implement this? Dot file [from slide] shows processes and communications. Up to 31k agents, initially too much communication. Economists have little programming experience. Libmboard was developed (Shawn Chin), uses SPMD and sync prevents deadlocks, see picture [from slide]. Has iterators and filters to help speed up message parsing, means don't have to go through complete list for each agent.

Case study of Social Capital Model, replicates a social mathematical model. Involves calculation of numbers of transitive relationships, etc. bottleneck at one actor. Trying to overcome the issues. C.f. Cournot model. Geometric or round robin distribution tested. Run on Mac and Iceberg cluster.

Important to think about type of model and parallelisation of code plus distribution of agents. Is the model correct? Social or economic models, interest in how an equilibrium is reached and what happens if shocks are introducted. Large quantities of data have to be analysed locally. Run time can't be predicted which users might find annoying – needs more dialogue. `http://www.eurace.org` `http://www.eurace.groups.shef.ac.uk`.

3.2.11 Anders Johansson (University College, London) – Urban Mobility: a Data Driven Approach

For slides see `http://www.grids.ac.uk/Complex/Leeds.15.6.10/Johansson.ppt`.

Important to start with empirical data, analyse it and then build models. Sometimes done other way around however data is important.

There are 100s of different models and approaches. Microscopic to macroscopic, discrete to continuous.

Using Helbing forces model. Initially developed s/w to do video tracking. Walking experiments at Paul Sabatier Uni. doing movement tracking. For model calibration, combine model and dataset, used Brutus cluster at ETHZ. Determined interaction patterns from the data. Validation in dense crowds. Various emergent phenomena – e.g. lane formation, stop and go waves. Saudi Arabia mosque – 47k pilgrims.

Current work is on urban environment. Use data layers – map + GPS tracks, + PoI from open streetmap. Investigating different means of moving around London.

Can a model be built to re-construct known trajectories? Important to include behaviour influenced by local points of interest.

Q. Watts. Local knowledge is important (hence the use of PoIs) not all information is on the map.

3.2.12 Andy Turner (Leeds) – Large Scale Social Simulation in Java and the NeISS Project

For slides see `http://www.grids.ac.uk/Complex/Leeds.15.6.10/Turner.ppt`.

GENESIS (ESRC) and NeISS (JISC) projects, see `http://www.genesis.ucl.ac.uk` and `http://www.neiss.org.uk`, see also `http://ur1.ca/07j0y` for slides and related material.

Model characteristics: spatial 2D or 3D, temporal resolution, agent and environment attributes, constraints. Types of model include pre-historical, historical, contemporary and ones with use restricted or public data. Typology of models [from slide]. Constructing a demographic (scale years) and a traffic model (scale days). Random numbers are important. Iterators are used to go through sets of objects.

A simple test model deals with birth and death. Can we scale to 10Bn agents? Basic testing is done. Need to even out behaviour using starting conditions so not all initial births occur on day 266. Input parameters using known distributions.

GENESIS traffic simulation: There is a high spatial resolution raster network (2D square cells aligned in rows and columns) on which agent movement takes place in the 8 cardinal directions. (Yes, it might be better to be using hexagons, but anyway this resolution can be increased.) Agent movements are at fixed decimal precision between cells in this network grid. So in a time tick a moving agent moves towards and possibly through many cell centroids and its position is given to a fixed decimal precision in coordinates at a higher resolution than the high spatial resolution raster network. It is intended that Agent movement is only maybe re-considered once they reach a cell centroid in the high spatial resolution raster network. At a lower resolution there are reporting rasters which keep a record of the Agents they contain. These are for producing population density surfaces and vector networks depicting agent movements.

Discussion of computational issues code [see slides]. Not yet parallelised. Previously used MPJ Express. Issue with memory management, swapping out to disk. Agents and data structures, directory structure.

Future work: happy to help other people to try it. Look at density and flow based constraints in traffic models as well as parallelisation.

3.3 Discussion

3.3.1 Chris Greenough – Where do we go from Here?

Agent technology spans an amazing number of apps. We have been motivated by a statement from EPSRC. We seem to have a very strong agent research community in the UK, however people are focussed in their own app areas. We could nevertheless identify a number of generic problems or challenges. Validation and verification have been identified in the talks. Software is important, e.g. review from AgentLink project. EPSRC issued a call of a statement of need from communities, see below. This could be an opportunity for the community to work together. Is the need, inclination and enthusiasm to do this?

What is a CCP? See brochures and `http://www.ccp.ac.uk`. Typically brings together a community and works to its benefit. EPSRC funds most of them – core support, networks, flagship projects.

What has happened previously? AgentLink, a bit like a pan-European CCP, Multi Agent Systems Association, summer school. These are EU level projects, do we need something specific to the UK? Technology Roadmap. Meetings... EURAMAS. EUMAS, there will be a summer school and conference in Paris.

EPSRC wish to identify communities for directed funding initiatives. Posed 9 questions. Can be followed up by proposals. Deadline for statement is October. What do we need to tell them? What is the benefit? What technologies?

Next actions. We will collected our thoughts into a set of notes, will put on Web and circulate. Set up a community Wiki. Generate a draft statement of need, some people will be asked to contribute and all to comment. Ask other people to get involved.

Comments?

Q: Holcombe. Aren't the CCPs app focusssed. Yes, so far, but suggestions are that one on Agents might be of interest and they're looking for new communities, e.g. a new research programme.

Q: Turner. Anything going on already? Not that we know of.

Q: Luck. This is one particular area within a broad field of agents. Whilst its a large area its also specific. Its ABM in simulation, not complex systems *per se*.

Q: Voss. CCP12 is also a broad area.

Q: Barros. Would we consider working on a platform, or guidelines e.g. for validation? A platform might be more identifiable with technology. Current CCPs develop community based s/w or will develop something specific. E.g. CCPP will also educate their community to develop better s/w as well as maintaining some codes.

Q: Siebers. How to link this to UK specifically? Make UK world leading in this area. There is also other work in EU and US which we could bring in and make relevant to our community. EPSRC are involved in large EU projects, e.g. PRACE. Impact is the big thing.

Q: Sibly. Don't see any issue making the case that ABMs are important. Calibration, validation and verification are important topics. Translating between languages is also important, e.g. enable use of different tools and visual techniques. Test that s/w is doing what we expect. There are currently different methods to check that models are trustworthy.

Q: Birkin. There are lots of oingoing activities, so EPSRC might ask what is new. Idea of exascale is interesting and links into PRACE. Masses of data across all disciplines could pull things together. Validation and verification is then a natural focus because its not straightforward to use large resources, serial bottlnecks become a problem and data management is already an issue. Using HECToR might be tempting for , however the number of people using large models is currently quite small.

Q: Siebers. We are mostly academic. We should include industrial prople – can EPSRC help? Progress requires building trust in the modelling.

4 Future Research Challenges

4.1 Modeling and Simulation of the Exascale

An NSF report was published in 2007 entitled: *Modeling and Simulation at the Exascale for Energy and the Environment. Report on the Advanced Scientific Computing Research Town Hall Meetings on Simulation and Modeling at the Exascale for Energy, Ecological Sustainability and Global Security (E3).* `http://www.sc.doe.gov/ascr/ProgramDocuments/ProgDocs.html`. This notes as one of its goals *to identify emerging domains of computation and computational science that could have dramatic impacts on economic development, such as agent based simulation, self assembly, and self organization* and suggests this can be addressed by *Math and Algorithms. Advancing mathematical and algorithmic foundations to support scientific computing in emerging disciplines such as molecular self assembly, systems biology, behavior of complex systems, agent based modeling, and evolutionary and adaptive computing.*

More specifically it lists a number of challenges which have to be addressed. Agent based modeling [3, 2] must be advanced along several directions before it can present a viable approach for addressing exascale application needs as follows.

- Distributed query resolution to allow agents to flexibly and repeatedly find other agents and recognize affordances for interaction in a dynamic environment with a continually and endogenously evolving structure (e.g., non-reified networks);
- Situational activation of agents based on contextual factors and associated real location to workings sets of processors with appropriate interprocessor locality;
- Efficient implementation of periodic fine grained interactions between agents where the payoffs from the interplay are endogenously defined as a function of the ongoing interactions

themselves, such that players are free to enter and leave the interactions at idiosyncratic times;

- Distributed time scheduling at a level of parallelism beyond the current approaches;
- Extremely high volume data warehousing to allow efficient exploration of huge numbers of large model runs;
- Efficient directed sweeps across huge model parameter spaces with appropriate adaptation as results are discovered;
- Domain decomposition techniques for parallel agent based simulations where the computational load per agent is variable, in time for the same agent, as well as from agent to agent, and where the geographical locality has no relation to the nature and volume of communication between agents.

4.2 The FuturIcT Knowledge Accelerator

There is growing momentum in Europe for an EU ICT Flagship proposal to expore social life on Earth, see `http://www.futurict.ethz.ch/FuturIcT`. This includes a white paper [4] explaining the background and potential impact of such a project which would run for at least 10 years and be on the billion EUR scale. A number of members of the UK community have already expressed interest.

Quite a lot of the discussion of this paper focusses on the use of agent based modelling. In particular Section A2.3 is as follows.

Besides massive data mining capabilities, it is required to build up suitable supercomputing capacities for the simulation, optimization, and management of sustainable techno-social and economic systems. Gigantic computer power is, for example, needed for large scale computational analyses in the following areas.

- Massive data mining, e.g. real time financial data analysis;
- Network research, community detection;
- Monte Carlo simulations of probabilistic system behavior;
- Multi-agent simulations of large systems (e.g. "whole earth simulation", which may involve up to 10 billion agents and complementary environmental simulations);
- Multi-agent simulations considering human cognitive and psychological processes (e.g. personality, memory, strategic decision-making, emotions, creativity etc.);

- Realistic computer simulations with parameter-rich models (coupling simulations of climate and environmental change with simulations of large techno-social-economic-environmental systems);
- "Possibilistic" multiple world view modelling (to determine the degree of reliability of model assumptions and to improve the overall prediction capability);
- Calibration of parameter rich models with massive datasets;
- Scanning of multi-dimensional parameter spaces;
- Sensitivity analyses (e.g. k-failures);
- Parallel worlds scenario analyses (to test alternative policies etc.);
- Visualization of multi-dimensional data and models of complex systems
- Optimal real time management of complex systems ("guided self organization", "self optimization").

It should be underlined that most challenges addressed by the FuturIcT Flagship concern a combination of the above points.

4.3 From the Leeds Workshop

Other directions in research were considered during the Leeds workshop of 15/6/2010. These include the following.

- Issues of applications to large and complex systems;
- Performance of ABMS – parallelisation technology;
- Verification and validation of models;
- Statistical inference for agent based models;
- Representation and standards for model exchange;
- Identification of future application areas.

5 A Collaborative Computational Project on Technology for Agent Based Simulations

5.1 CCP Review and Call for Statements

Dr Emma Jones, EPSRC, on 12/5/2010 noted the following.

As agreed at the last CCP Steering Panel meeting, a call for statements of need for CCPs in the EPSRC remit has been published on the EPSRC web site. See `http://www.epsrc.ac.uk/funding/calls/open/Pages/computationalprojects.aspx`

The closing date for statements of need is 4pm on 1/10/2010.

The main discussion of the CCP review at the CCPSP was as follows.

- The purpose of this review is to assess which scientific areas are of highest priority to include within the EPSRC supported portfolio of CCPs – it is not to review the CCP mechanism itself.
- the review will also assess levels of support (core support, networking activities and software development projects) required by each CCP. In addition to the Daresbury core support provided through the SLA between EPSRC and Daresbury, EPSRC will also make funding available for networking and software development projects identified as high priority through this call for statements of need. N.B. Proposals for these projects will be invited following the outcome of the prioritisation of the statements of need.
- A statement of need is required for all CCPs that the community would like EPSRC to support, whether they have current funding from EPSRC or not.

5.2 Background

The purpose of the Collaborative Computational Projects (CCPs, see `http://www.ccp.ac.uk`) is to bring together leading UK expertise in key fields of computational research to tackle large scale scientific software development, maintenance and distribution tasks. Each project represents many years of intellectual and financial investment. The aim is to capitalise on this investment by encouraging widespread and long term use of the software, and by fostering new initiatives such as High End Computing consortia. The Collaborative Computational Projects cover a range of domain areas spanning several Research Councils.

The CCPs are also a convenient instrument for the UK to participate in international activities, for instance as a lever to funding networks of excellence. An example is Psi-k, `http://www.psi-k.org`, which has strong support from members of CCP9.

Agent based models have been used since the mid-1990s to solve a variety of business and technology problems. Agent based simulation already plays a vital role in many areas of the UK academic and industrial research community. Applications range from the economic and financial sector through logistics to behavourial science, from the modelling of biological systems to the simulation of complete market economies. There are a vast number of applications being addressed in the UK and worldwide through agent based modelling. Most of these applications use similar technology and encounter similar issues. Some of the major generic ones are: model verification and validation, simulation performance, the use of high performance computing...

There have already been activities to bring together the agent community to share knowledge and experiences and there are often sessions at the major computer science and complex systems conferences. One recent notable pan-European initiative was AgentLink, for a time funded by EU but unfortunately no longer active.

The recent call by EPSRC for a "Statement of Need for a Collaborative Computational Projects" referenced above provides an excellent opportunity for the agent community to collaborate and inform EPSRC of its needs. A good case to EPSRC on the requirements of the agent based modelling community could open the doors to funding for network activities and the collaborative development of software and technology address generic issues such as verification and validation.

5.3 Remit

The main activities of a CCP are typically as follows.

- Carry out flagship generic code development projects
- Maintain and distribute code libraries
- Organise training in the use of codes (sometimes in the form of annual schools)
- Hold meetings and workshops
- Invite overseas researchers for lecture tours and collaborative visits
- Issue regular newsletters

Flagship projects represent innovative software developments at the leading edge of a CCP's area of science or engineering. They normally last for three years and may support a PDRA associated with the project. At the end of a flagship project, the resulting software usually becomes part of the code library. The focus of this CCP could initially be on tools for validation and verification of ABMs.

CCPs maintain, distribute and develop the new code according to demand from member and user research programmes. This flagship model suits most CCPs. It provides a mechanism for responding to advances in the appropriate subject area and maintains the interest of participating staff in cutting edge research. The collaborative approach makes the community almost uniquely able to adapt and respond to developments in computer science, information technology and hardware. One of the strengths of the scheme is that the focus of each CCP has evolved to maintain international scientific topicality and leadership within its community. The CCPs are increasingly represented in science and engineering as advances in computational techniques and hardware make it feasible to tackle problems of real practical significance.

Each CCP has a Chair and a Working Group, which sets the scientific agenda, decides the work programme and monitors progress. Currently, more than 300 groups participate in the CCPs, including several outside the UK and in industrial research. There are probably more than 1,000 individual researchers and research students in the UK supported by CCPs.

The whole CCP programme is overseen by the CCP Steering Panel, under the current chairmanship of Professor P. Coveney. The Steering Panel comprises the Chairs of all CCPs and two international members (Professors G. Ciccotti and Professor Himeno), together with Dr. R.J. Blake, as Director of STFC's Computational Science and Engineering Department, and the Director of CECAM. The Steering Panel looks actively at areas for co-working between CCPs; joint meetings, workshops and training events are common.

5.4 What we will require

Volunteers to form a Steering Committee that will solicit input from the community and draft a statement of need for the community. A separate Wiki has been set up to take the discussion forward and develop documents for submission to EPSRC. See `http://www.softeng.rl.ac.uk/wiki/abm-ccp`.

References

[1] M.J. North and C.M. Macal *Agent Based Modelling and Simulation for Exascale Computing* SciDAC Review (Feb'2008) `http://www.scidacreview.org/0802/html/abms.html`

[2] M.J. North and C.M. Macal *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation* (Oxford University Press, New York, 2007)

[3] M. Woolridge *An Introduction to Multi Agent Systems* (Wiley and Sons, 2002)

[4] D. Helbing et al. *The FuturIcT Knowledge Accelerator: Unleashing the Power of Information for a Sustainable Future.* (ETH Zürich, 2010) `http://www.futurict.ethz.ch/data/futurict.pdf`