eIUS: Microelectronics Experience Report

In the text: Some barriers/etc. in '{{…}}' to provide more contextual information.

Interviewee profiles (group interview; two interviewees)

Senior researcher and coordinator of the project, working at a UK institution: responsible for the overall architecture and development of the Grid infrastructure and ensuring that this infrastructure meets the scientific needs of the electronics research community

Researcher/PhD Student at the same institution: "I am working on the simulation of large statistical ensembles of (.) [microelectronic] devices studying the effects of intrinsic parameter fluctuations on device parameters in these ensembles. I have been working with grid and eScience people (..) to enable these simulations on large scale parallel systems."

Time spent on research

Researcher/PhD Student: 100%

Research area

Nano-technology semi-conductor device research, simulation and design; the project is funded by EPSRC for four years until September 2010 and includes academic partners at five UK institutions/universities plus six prominent industrial partners.

Research question(s)

Transistors (a semi-conductor device) become smaller and smaller which means the "individual positioning of (..) atoms on those pieces of silicone are actually influencing the behaviour of the devices, so they are becoming statistically variable", only slight differences change functionality. "Right at the bottom level [the project is] looking at the clustering of atoms and positioning of atoms and how these transistors vary", and subsequently predicting what this means. The Electronics group in Glasgow is a world leader in this kind of research. This information is then used to get to implications for the development of circuits and chips, i.e. how to build those to have them work within a specific variability, within certain tolerances.

"I guess it's just we want to look at the underlying physics that's associated with what you were talking about, so right down at the bottom, why do we see these variations in transistors and how do we then pass that up to the next people in the design flow, and I guess we try to find out how to describe that detail (..)."

In more detail/examples

"That's very much high performance computing doing large quantum level simulations of predicting thousands of these transistors and doing the averages and statistics over them, trying to understand if we did a hundred thousand of these what level of reliability might we get for example”. Systems/chips build on that basis can then be configured for “certain performance characteristic or [if] you want circuits which are low [in] power consumption”.

“The first part of the project was showing how we can build large scale simulation or a grid infrastructure shall we say from running these hundreds of thousands of simulations and one of the challenges I guess of this is it’s not just a HPC; it’s not just running big simulations on big machines, it’s actually, a lot of the stuff that we’re working with is under intellectual property [IPR] or non disclosure agreements; the (..) [large manufacturers and] whoever else might be using and designing these transistors is very wary of, if this is our data and you cannot disclose it to anyone, so the idea that we could just say, ‘put it on the National Grid Service and do some arbitrary simulation’ – they would be very scared of this (.).” This can also be a serious legal problem dealing with protected, confidential industrial data of the chip vendors (contracts to use their intellectual property have to be agreed on and signed). And “it might well be the case, the only resource we can run it on is the electronics cluster”. So even between project partners access to data and authentification are important issues that can only partially be solved through secure infrastructure. “So it’s a very, very sensitive area so even just in principal we’re trying to build infrastructures which are secure and in principal you should be able to securely share your data with the right people across a particular virtual organisation.”

“So a lot of what we’ve been doing is trying to build infrastructure which can do both, large scale simulations [and] manage the data associated with this because you’re generating (..) hundreds of thousands of files, capturing the metadata associated with this; and then the idea is that only certain people can get access to it, so if you have signed a non disclosure agreement and you agree to abide by the terms and agreements of the contracts of accessing that particular data, you can use that in simulating a given circuit.”

A range of security technologies including VOMS, PERMIS, Shibboleth, GSI and the Kerberos protocol (http://en.wikipedia.org/wiki/Kerberos_(protocol)) is used in the project.
(see: Integrating Security Solutions to Support nanoCMOS Electronics Research:

Sinnott,R.O. Asenov,A. Bayliss,C. Davenhall,C. Doherty,T. Harbulot,B. Jones,M. Martin,D. Millar,C. Roy,G. , Roy,S. , Stewart,G. , Watt,J. (2008): IEEE International Symposium on Parallel and Distributed Processing Systems with Applications, Sydney Australia, December 2008.)

Another technology we have adopted for federated data management is the Andrew File System (AFS, http://en.wikipedia.org/wiki/Andrew_File_System) which is a distributed file system running on a number of trusted servers to provide access to data. These separate technologies still have to be glued together to make it “seamless and nice and easy for these people to run the large scale simulations and manage all the data”.

Research Lifecycle

“I've done a bit of reading and then it’s been a lot of simulation, a lot of data analysis and a lot of post processing stuff and then trying to understand it from that. I don’t know if there’s a huge amount you could learn just from the data on its own.”

You can “run large scale simulations and generate data and look at metadata and have secure access to distributed data, the proof of the concept is there, we have versions of the systems doing it”. {{But: Proof of concept is one thing and “working with real live organisations like these electronics commercial partners or with the National Health Service” is another.}}

Literature Review

“Well I guess really most of our stuff comes from IEEE journals. There’s the big ones, ‘Transactions on Electron Devices’ which is sort of the bible for the field.” A lot of publications can be found for this particular research area, so this is the “first point of reference” – and this has to be done regularly, because “the field moves so quickly”.

Data collection process (through simulation)

“You start off with some technology provider like [large chip manufacturers] or somebody will say this is the architecture of this transistor.”

E.g.: “This is what we’re trying to design, can you simulate and tell us what happens at the quantum level with that particular transistor.” – BTW: You have “two billion transistors on a chip now” (the number on the last Intel processor in 2008).

The simulation software produces the results in one file (e.g. I-V curves, i.e. curves of current-voltage characteristic; or potential voltages), for that you also need another file with some data on how the actual physical construction of the transistor affects the results. In the end you have “two of three files for every device and that’s all just numbers and you have a hundred thousand of those” – which is a challenge for data management: “(..) you have to have some system or database or something to manage that, and I think it’s important to be able to search for specific things as well”.

There are no standards for data in this research area: “(..) these people are pretty much making it up, there are some, but when it comes down to it the actual standards for how to describe the data”, no standards exist. {{It is not possible to standardise this kind of metadata, because “it was not meaningful to try to do this because it just varies” and is different every time.}} They can only collect “some basic metadata (..) about which simulations, run by who, when, etcetera (..)”.

At “the very first project meeting we had I was talking about we built a brief prototype of what we can do at the device simulation level (..), tick this button and you’ll have line edge roughness and if you tick this button we can simulate electron mobility and everybody outside of (.) [our institute] didn’t have a clue what I was talking about (..); it’s not something they’re aware of, these people that design VLSI [Very-large-scale integration] and large scale systems, they don’t work at what’s the nuts and bolts of atomics, it’s with the system.”

But de-facto standards can “get adopted with the work”, the standard can become an API: “For the most part most of what we’re doing is industry based, (..) these tools that they’re using are the standards I guess you could say.” “Each tool in the tool chain that we’re working towards will have a completely different set of information and metadata associated with it”

{{Initially an existing technology stack (incorporating workflows like Taverna) should be adopted which seemed feasible due to the existing collaborations with the project partners and support through OMII UK. That meant running a hundred thousand simulations via services based on that technology, but in the end there were more problems than benefits.}} One of the main issues with a hundred thousand simulations is to manage this amount of data and therefore the underlying middleware has to work – but this is still a work in progress: "Right now I guess we've adopted a more pragmatic approach for what software and what technology we're applying. All of our efforts right now are trying to make this data management get solved."

In terms of infrastructure there are two themes in the project: Three RA are building grid middleware for the project. Two of them are based at two different institutions, "looking at data management and metadata management". The third at another institution is "building the job submission and the whole grid infrastructure" for the huge number of simulations.

The ScotGrid (http://www.scotgrid.ac.uk/) is used as an infrastructure to do much of those large-scale simulations. The interface software to submit the jobs on the Grid is called 'ganga' (http://www.scotgrid.ac.uk/wiki/index.php/Glasgow_Ganga_Quickstart_Guide). Overall there are different ways of how to do these simulations and where to submit them, e.g. the NGS, electronic clusters with various Grid engines or Condor pools (http://www.cs.wisc.edu/condor/). All this was tested for small batches of Jobs (especially by one RA), but on a larger scale and including proper security and authentification is more challenging. There are now systems in place that support this.

Data analysis (simulation data)

The data analysis is a combination of "data mining type thing" plus statistical operations.

"Last year I used a lot of the stats program R to do my analysis. In the last six months I've been using more of Python which is a scripting line in which there's a bunch of scientific tools that go with that which I've found quite useful but yes, very much you have to write your own code to do the analysis, there's nothing that really exists."

"A lot of it's just done on the workstation, but there are a couple of ones; because there's so much data it's taking so long to go through it that I've just split it up and just said run on the cluster and just take a bunch of nodes." ScotGrid or NGS are used for this as well: "I'll do it the way it's easiest; if it's easiest on my work station then on a cluster and then if I have to I'll put it on ScotGrid.".

For 2D visualisation the plot programme Grace is used as well as OpenDX for 3D, both run on the interviewee's desktop machine. {{But at the same time "looking at the chips with two billion transistors on them, how do you visualise?"}}

{{"We did have a project working with people down in Cardiff, they have this resource aware visualisation environment, RAVE, and a project funded by OMII looking at having the visualisation service for the OMII stack, and we looked into this but it wasn't really solving [our visualisation problems]."}}

{{Initially we "wanted to be able to run simulations, do analysis and get an image, and then you want to be able to tweak that and it's going to change those simulations, change that analysis, give you a different image" But right know if a researcher can run “big simulations and he's got just some in-house tool for analysing his data and understanding why you’re getting the results you're getting, then right now that’s pretty much all we’re really aspiring [to as it represents the way they want to work]”.}}

Example on simulation (data collection) and first analysis steps, also standards

“So the idea is he runs a hundred thousand simulations or several hundred thousand and it will generate collections of I-V curves for that particular transistor that you're trying to simulate, and then they want to be able to understand which parameters associate with that particular device are causing that to happen and this is where these kinds of standards are coming in, there are agreed parameters which you can use to be able to say these seven parameters influence that, the shape of those I-V curves that have been generated and it’s those kinds of things, those parameters which you then want to be able to feed up to the next level, so the guys who are designing circuits don’t really care about a hundred thousand I-V curves, current voltage, etcetera, they really care about what the actual physical properties of those devices that you are simulating are resulting in, that variability, because I need to know those physical parameters, and so we’re really trying to extract that information up and there’s a (..) [way] to do this, which (.) [our institute] opened.”

Collaboration

In the interviewees institution the electronics people “have weekly meetings where everyone says ‘oh, I've done this and this this week, this is the results I forgot, here’s some pictures, blah, blah, bah, oh yes, this is good, we can write this up, send it to me, we can work out what to put in the paper’, things like that.” The project coordinator and his RA also usually attend these. Additionally these are followed by weekly (usually Wednesdays) management audio conferences by phone with the project partners to coordinate efforts. Not all work is directly related, but some results of simulations by the interviewees institution are used by other partners, especially one partner university – there are additional bilateral meetings scheduled for this.

Access Grid is used occasionally and mostly for technical meetings. Sometimes project meetings are also f2f. Skype is used as well.

On f2f and AG meetings: “It’s really when it needs to be done it’ll be done and there are people who do come up from another site (..).”

“Data that’s being shared is quite a small subset of the data that we have”, usually just zipped and emailed (about models to one partner university). Also a secure webpage and an encrypted file system are used to exchange data securely and more seamlessly for direct input into simulation software (e.g. with the same partner university again). Overall the data of one researcher does not make sense instantly to another researcher “it has to be very much abstracted up to the level that they work at”, i.e. “the circuit guys don’t really understand the devices”. So the direct exchange of data often is not that useful. {{Overall the project is quite diverse and “there isn’t a single person who’s got the complete knowledge of how it all fits together perfectly”. E.g. the interviewees institution focuses pretty much on their end of things and for a more complete understanding “I guess what we’re doing right now is we’re starting to slowly work our way up”.}}

Dissemination

An article was recently submitted to the ‘Transactions on Electron Devices’ IEEE journal, which is the most important one. “The device modelling group itself has got a lot of papers on the subject.” The IEEE Electron Device Letters is also quite popular and a journal article has been accepted for the Philosophical Transactions of the Royal Society journal, coming from the UK All Hands Meeting 2008 publication.

“So a lot of it is really just graphs, it’s more at that level, it’s about the results as opposed to the background. (..) If you haven’t got anything novel at the end which no-one’s done before it’s not a publication.”

“It seems that it’s getting more and more to be the type of journal that also has a repository, so normally you publish the paper but you would also publish part of the results; maybe it’s not the raw data, maybe it’s just summary tables with important information, it’s not the whole thing and maybe it’s not enough in order to replicate the experiment but it gives you quite a lot.” – “If you wanted to exchange raw data in a meaningful way you would have to have agreements between all the people involved.”