Third IEEE International Conference on e-Science and Grid Computing

Towards a Grid-Enabled Simulation Framework for Nano-CMOS Electronics

Liangxiu Han[*], Asen Asenov[‡], Dave Berry[*], Campbell Millar[‡], Gareth Roy[‡], Scott Roy[‡], Richard Sinnott[†], Gordon Stewart[†]

[*]*National e-Science Centre, University of Edinburgh, Scotland*

[†]*National e-Science Centre, University of Glasgow, Scotland*

[‡]*Department of Electronics and Electrical Engineering, University of Glasgow, Scotland*

lhan@nesc.ac.uk

Abstract

The electronics design industry is facing major challenges as transistors continue to decrease in size. The next generation of devices will be so small that the position of individual atoms will affect their behaviour. This will cause the transistors on a chip to have highly variable characteristics, which in turn will impact circuit and system design tools. The EPSRC project "Meeting the Design Challenges of Nano-CMOS Electronics" (Nano-CMOS) has been funded to explore this area. In this paper, we describe the distributed data-management and computing framework under development within Nano-CMOS. A key aspect of this framework is the need for robust and reliable security mechanisms that support distributed electronics design groups who wish to collaborate by sharing designs, simulations, workflows, datasets and computation resources. This paper presents the system design, and an early prototype of the project which has been useful in helping us to understand the benefits of such a grid infrastructure. In particular, we also present two typical use cases: user authentication, and execution of large-scale device simulations.

1. Introduction

Progressive scaling of Complementary Metal Oxide Semiconductor (CMOS) transistors, as tracked by the International Technology Roadmap for Semiconductors (ITRS) [1], has driven their phenomenal success, and especially that of the Metal–Oxide–Semiconductor Field-Effect Transistor (MOSFET), by far the most common field-effect transistor in use in both digital and analogue circuits. MOSFETs are in mass production at the current 90 nm ITRS technology node [2] and sub-10 nm transistors are expected at the 22 nm technology node, scheduled for production in 2018. 4 nm transistors have been already demonstrated experimentally [3], highlighting silicon's potential for scaling beyond the end of the current ITRS. However, it is widely recognized that the variability in device characteristics and the need to introduce novel device architectures represent major challenges to scaling and integration for present and next generation nano-CMOS transistors and circuits (Figure. 1.1) [4]. This will in turn demand revolutionary changes in the way in which future integrated circuits and systems are designed. Strong links must be established between circuit design, system design and fundamental device technology to allow circuits and systems to accommodate the behaviour of individual transistors on a chip. Design paradigms must change to accommodate this increasing variability. Adjusting for new device architectures and device variability will add significant complexity to the design process, requiring orchestration of a broad spectrum of device tools by geographically-distributed teams of device experts and circuit and system designers. This can only be achieved by incorporating e-Science technology and know-how across the whole nano-CMOS electronics design process and revolutionizing the way in which these disparate groups currently work.

The Nano-CMOS project[1] has been funded by the Engineering and Physical Sciences Research Council (EPSRC) in the UK. The main goal of the nano-CMOS project is to construct a distributed data and computing framework, with robust security mechanisms based on grid technology, to enable distributed electronics research groups to collaborate by sharing designs, simulations, workflows, datasets and computational resources. Through combining world-leading electronics design centres in the UK, including those at the University of Glasgow, the University of Edinburgh, the University of York, the University of Manchester, and the University of Southampton, we aim to revolutionize the electronics design industry. The project began in October 2006, and the first year's

[1] www.nanocmos.ac.uk

0-7695-3064-8/07 $25.00 © 2007 IEEE
DOI 10.1109/e-Science.2007.24

work, conducted by the universities of Glasgow and Edinburgh, has been targeted towards developing simulation services, including a grid-enabled atomistic device simulation service, at the process and device simulation level. In the second year of the project, these services will provide further functions to support the higher level circuit and systems simulation requirements of the project.

We have proposed and built a grid-enabled simulation framework. The implementation at this early stage of the project mainly supports the device modelling community at the University of Glasgow. This early experience provides important lessons to help us understand the benefits of using such a grid infrastructure.

The rest of this paper is organized as follows. We describe the requirements and challenges of grid-enabled simulation frameworks in section 2. In section 3, we first briefly introduce the OMII-UK grid middleware framework and the related integrated grid components of OMII-UK: OGSA-DAI and GridSAM, and then we describe our open grid-enabled software system architecture based on this grid technology. As important scenarios in the project, section 4 describes two typical use cases: user authentication and running simulation services. Section 5 presents details of our specific implementation. Section 6 discusses our conclusions and outlines key future lines of the work.

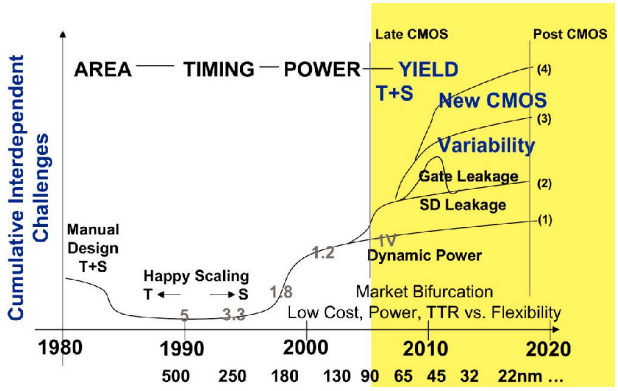

Figure 1.1. Cumulative interdependent challenges of the semi-conductor industry [5]

2. Requirements and Challenges

This section provides an overview of the data, computation and security requirements of the nano-CMOS electronics community.

2.1 Data and computation requirements

The project has adopted a hierarchical simulation methodology since this closely reflects the requirements of the nano-CMOS researchers. This simulation hierarchy determines the interaction of simulation tools and associated data sets. Figure 2.1 shows the interaction of different kinds of simulation used at different stages of the nano-CMOS electronics design process. As indicated in Figure 2.1, the process simulation is concerned with the physical steps necessary to turn a piece of silicon into a working device. Based on a given device structure, the device simulation consumes this information via targeted input files from the process simulation, and in turn generates output files. Output extracted from the device simulation can be fed into services at the circuit level, and so on.

The nature of the statistical design and simulation methodology in the project places potentially unlimited demands on access to, and usage of, computational resources. Similarly, the volume of statistical datasets and associated metadata is potentially large, and this must be accessible by each of the geographically-distributed partners.

The general data and computation requirements during the project lifecycle are shown in table 1. Specifically, table 1 shows the six geographically distributed partners and the computing resources and data they estimate will need to be stored and processed during the lifetime of the project. A key aspect of this is that the scientists themselves do not wish to immerse themselves in the intricacies of the different HPC resources where these simulations will ultimately be run.

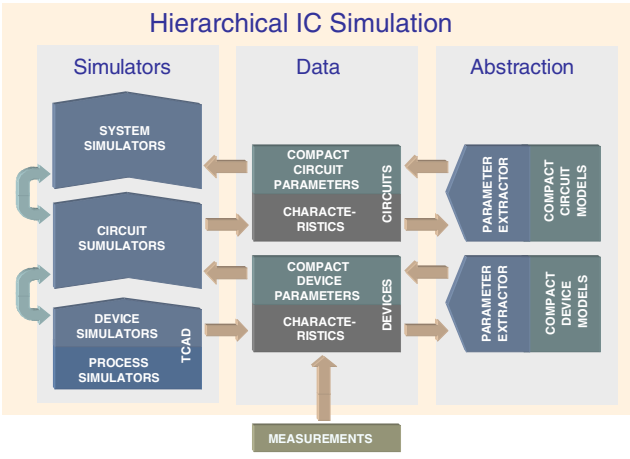

Figure 2.1. Hierarchical simulation structure

Similarly, the data sets have their own characteristics and are often targeted towards different commercial or bespoke tools. Therefore, a key focus of the project is to deliver services which allow the low level details of job submission to large-scale grid resources, such as the UK's National Grid Service (NGS)[2] and ScotGrid[3], and the associated staging of data to be hidden from the end-users (the scientists).

Table 1. Data and computation requirements during lifecycle of the project

Partner	Simulation Task	CPU H	Storage
Glasgow Device Modelling	*TCAD* *Atomistic* *Compact models*	*1–5* *5–20* *0.05*	*1 TB* *5–10 TB* *30 GB*
Glasgow Microsystems	*Fragment* *Cell* *Extraction to STA*	*20* *50* *0.1*	*10 TB* *5 TB* *100 GB*
Edinburgh	*Mixed mode simulation*	*0.1*	*5 TB*
Manchester	*SPICE circuits sim.* *T-level sim.* *Gate-level sim.* *Behavioural sim..* *Extraction sim.*	*1–10* *10–100* *1–10* *1–10* *1–10*	*10 GB* *5–10 TB* *1 TB* *100 GB* *20 GB*
Southampton	*Circuit level sim.* *Faults Behavioural sim.*	*1–5* *1–5*	*100 GB* *10 GB*
York	*Evolutionary systems sim.*	*5–10*	*30 GB*

2.2 Security issues

To support this form of hierarchical simulation framework, we must address a key issue of this domain: security. This is critical if we are to protect the intellectual property inherent in the project's data, simulations, design processes and software licenses.

To do this in an unobtrusive manner that provides seamless security to the end-users, the project is building on results from a variety of other projects [6][7][8]. The various project resources are protected by Shibboleth, which requires that users successfully authenticate before being allowed access. With this model, when the scientists attempt to access a protected resource, e.g. the project portal, they are redirected to their home institutions and requested to authenticate. Once logged in to their home sites, authentication information, along with the various privileges they possess (provided as digitally signed attribute certificates), are returned in a Security Assertion Mark-up Language (SAML) assertion and can then be used in authorization decisions. In this model, multiple sources of attributes are expected to exist and to be used for decisions at remote protected resources. Thus we expect that given sites will want to manage their own attributes and attribute distribution policies. Site autonomy is fundamental to this model and each remote resource provider will make their own local authorization decisions.

3. The Grid-Enabled Simulation Framework

3.1 Grid technology applied

To address the challenges of distributed computing and large-scale data analysis, we have adopted the Open Middleware Infrastructure Institute (OMII-UK) middleware [9] as a platform for deploying the grid-enabled simulation framework.

The OMII-UK middleware provides an infrastructure that allows collaboration between users and providers of grid resources and applications in a trusted and secure environment, which is the main reason for adopting it in the Nano-CMOS project. OMII-UK integrates several middleware products such as OGSA-DAI, which can provide services for data access and integration, GridSAM, which can submit and manage jobs on computational resources, and Taverna, which is a workflow management and enactment tool. However, although OMII-UK provides a secured web service container and message level security, it does not support the fine-grained authorization definition and decision-handling capabilities required by the project's researchers. Therefore, in the Nano-CMOS project [13], we have adopted the Shibboleth security model for supporting fine-grained control over users' privileges on remote resources.

3.1.1 OGSA-DAI

OGSA-DAI (Open Grid Services Architecture Data Access and Integration) [9][10] is a middleware product that provides access to structured and semi-structured data resources. It allows data resources, such as relational or XML databases, to be accessed via web services that offer data integration services to clients, along with appropriate metadata. Various interfaces are provided and many popular database management systems are supported.

[2] www.ngs.ac.uk
[3] www.scotgrid.ac.uk

In the Nano-CMOS project, by using OGSA-DAI, we can access and integrate distributed data sets, and transform data sets into the correct format for the next simulation tool in the chain.

3.1.2 GridSAM

GridSAM [9][11] is an open-source job submission and monitoring web service. The aim of GridSAM is to provide a service to utilize a variety of distributed resource managers. The modular design allows third parties to provide submission and file transfer plug-ins for GridSAM. Moreover, the job management API used by the GridSAM service can be embedded into grid applications that require job submission and monitoring capabilities.

GridSAM consists of several subsystems working together to support pluggable job persistence, queuing, job launching, file staging and failure recovery. Each subsystem has a set of well-defined interfaces that are collectively known as the service provider interface. This interface corresponds to the application provider interface used by clients to demand job submission functionality, more details of which can be found at the OMII-UK website[4].

GridSAM provides the project with an easy way to use the various available distributed computing resources.

3.2 The System Architecture of the Nano-CMOS Project

The current implementation of the Nano-CMOS simulation and data framework is based on a multi-tier open architecture as shown in Figure 3.1.

There are four layers in the architecture:

- The resource layer, which includes the physical, data and computing resources used in the project, such as ScotGrid, the NGS and our own data servers.
- The grid middleware layer, which is where we build our distributed data and computing framework based on OGSA-DAI and GridSAM hosted in an OMII-UK web services container.
- The service layer, which provides different simulation services and data services to the end-user. For example, at the process and device simulation layer we have provided services to perform atomistic device simulation and compact model creation, along with the various supporting data services.
- The user interface (GUI) layer, which provides HTTP(S)-based clients that allow users to interact with the system through a standard web browser. This is based upon portal technology; specifically we are exploring the GridSphere portal framework [12].

Access to the interface layer and the underlying layers has to be secure. As noted, we are using Shibboleth and federated access control to achieve this. The Shibboleth-delivered attribute certificates are currently stored within a portal session and subsequently used by services that require such information to make authorization decisions.

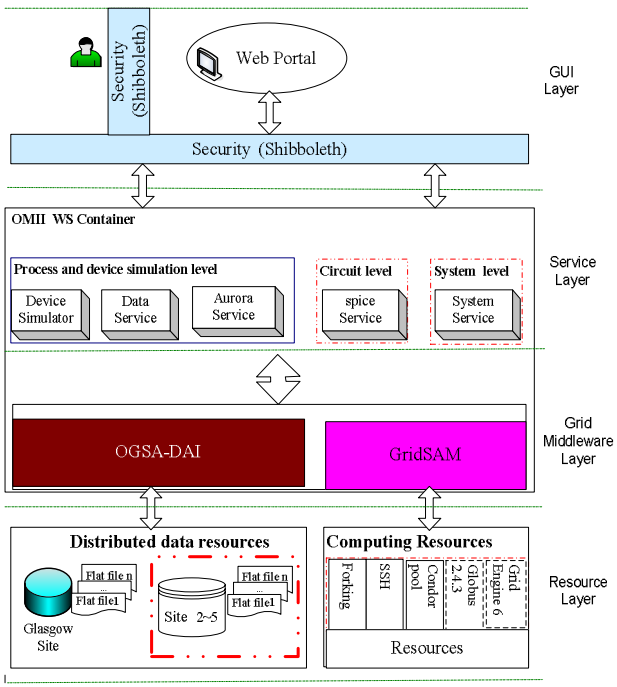

Figure 3.1. Nano-CMOS System Architecture

4. Use cases

To understand the interaction among the different kinds of services, a typical scenario based on Figure 3.1 is as follows: a scientist would like to run an atomistic device modelling simulation using a bespoke device simulator; this software generates statistically significant sets of current/voltage (I/V) curves. Once generated, these I/V curves will be fed into a commercial application (Aurora, which requires a valid software licence to execute) to generate a compact device model (which defines the parameters that model the key characteristics of that particular device). This

[4] www.omii.ac.uk

model will, in turn, be consumed by further services for circuit and system simulation.

Example use cases detailing how end-users interact include User Authentication and Simulation Execution.

4.1 User Authentication

In this project, we have adopted the Shibboleth security model for protecting the portal and simulation services [12]. In this model, local (existing) methods of authentication for remote log-in to resources are used. This model exploits several key components including federations, Identify Providers (IdP), service providers (SP) and, optionally, "Where Are You From?" (WAYF) services. By employing this technology, end-users will have single usernames and passwords from their home institutions which will provide seamless access to a range of resources at collaborating institutions and service providers. Local security policies at service provider sites can then be used to restrict (authorize) what resources authenticated users are allowed to access.

The main stages of user authentication are as follows. When a user accesses a service protected by Shibboleth, the portal will redirect to a WAYF service so that the user can select his or her home identity provider from a list of known and trusted sites. After that, the user is redirected to their site authentication server, e.g. an LDAP repository, at which point the user is prompted to log-in. After successful authentication, attributes including the user's roles are collected and sent back to the SP in a digitally-signed SAML assertion with information that they have successfully authenticated at their IdP. The SP user can make use of this information and make decisions which authorize the user to access certain services. The use case diagram for this is shown in figure 4.1.

4.2 Simulation Service Execution

When a user needs to run a certain simulation, he or she first needs to be authenticated through the user authentication mechanism outlined above.

After the end-user is authenticated, the returned attributes are used to personalize the portal, i.e. to show the client portlets that the end user is authorized to access and use. When the user prepares to run a simulation service, he or she will first interact with the associated data service. The user can do this directly, simply by invoking the client side portlet. However, it is often the case that data preparation is needed, e.g. to parameterize the particular simulation execution with appropriate values for the simulation. This might include the number of simulation runs they wish to perform, or whether they wish to have quantum mechanical simulation turned on. To do this, the user will select the associated data service and generate an appropriate input file. When submitting the job, the input file will be staged to where the simulation code is running via the job execution module of the associated simulation service.

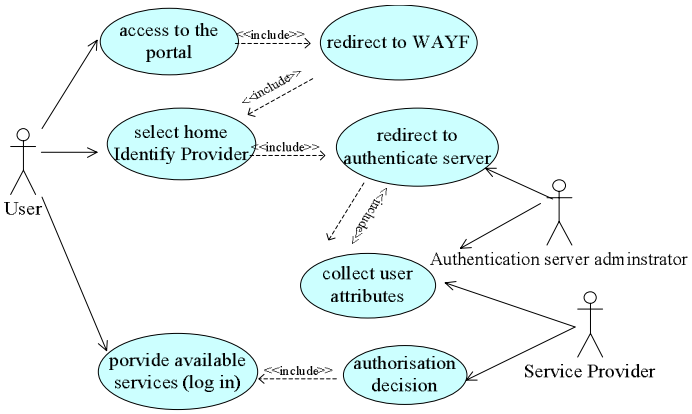

Figure 4.1. Use Case of User Authentication

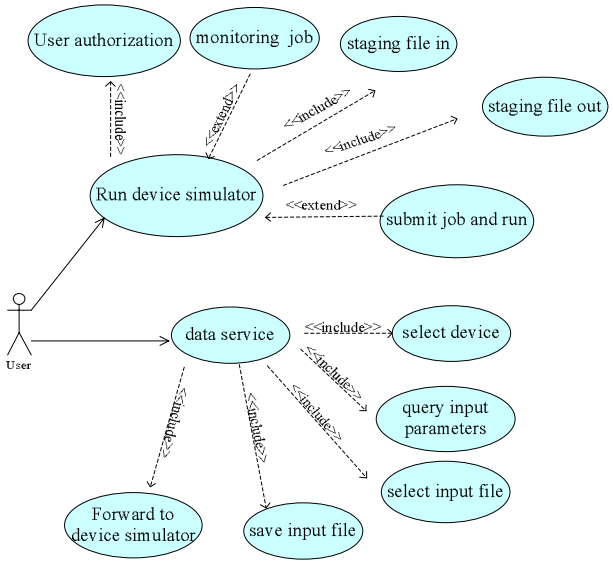

Figure 4.2. Use Case of Device Simulator Service Run

The job execution module will then pass the job to the selected resource's job manager. The job monitoring module of the simulation service will track the status of the job, allowing users to see what stage of execution has been reached. Once all of the jobs are complete, the job execution module will stage the output files from the remote grid resources to the appropriate local locations, from where they can be fed

into the next simulation service in the tool chain. Figure 4.2 shows the use case diagram of this scenario.

5. Implementation Status

The prototypes of the first nano-CMOS services have been built and made available within a project web portal which is protected by Shibboleth.

The current GUI client has been developed using the GridSphere portal framework which provides an open-source portlet-based web portal.

As mentioned in the previous section, a user must log-in and be authorized before he or she can access any available services. Figure 5.1 shows the user log-in stage with the various attributes used for authorized access to the services available. Figures 5.1, 5.2, 5.3 and 5.4 show our initial services, including the device simulation service, the data service and the compact model generation service.

All of these services have been developed and deployed using the integrated components of OMII-UK software, i.e. OGSA-DAI and GridSAM.

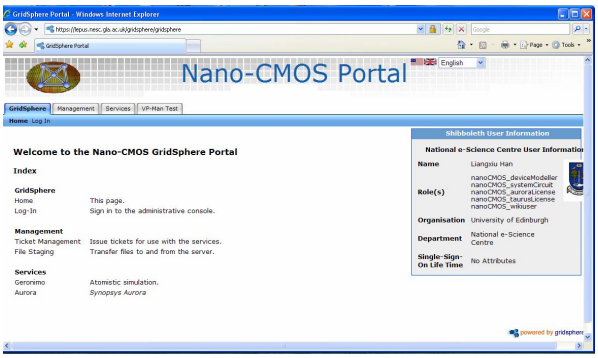

Figure 5.1. User Log-in and Security Attributes

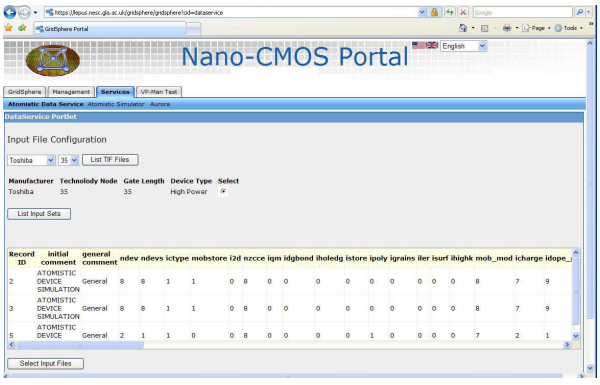

Figure 5.2. Device Modelling Data Service

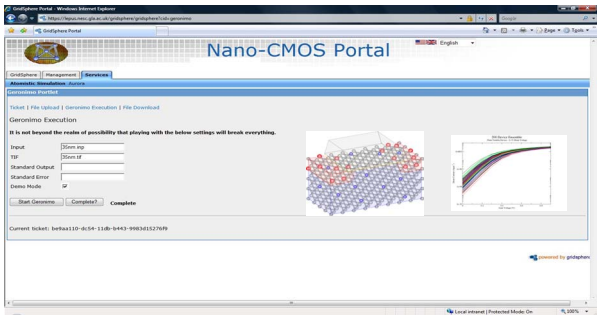

Figure 5.3. Device Simulation Portlet

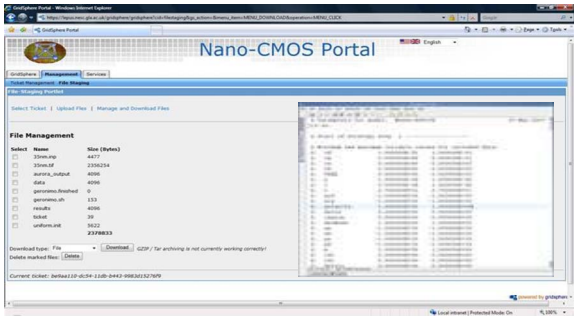

Figure 5.4. Aurora Compact Model Portlet

6. Conclusion and Future Work

This paper has proposed a grid-enabled simulation framework to meet the challenges of decreasing transistor scale and the increase in device design flexibility in the electronics community in the UK. The early prototypes based on off-the-shelf OMII-UK grid middleware offer grid-enabled atomistic device modelling and compact model generation services to support researchers at the University of Glasgow.

By using GridSAM, we can now submit simulation jobs to a variety of resources, including the NGS, the ScotGrid cluster at Glasgow, Condor pools and clusters running Sun Grid Engine. The details of the job submission process for each of these various clusters are completely hidden from the scientists who simply select the number of jobs or simulations they wish to run through the client portlet.

By using OGSA-DAI, we can query data and transform it to the format required for input to the various scientific simulation and analytical tools. Furthermore, we have deployed the Shibboleth security model for supporting multiple attribute authorizations and restricting access to services and data to privileged users. The implementation at the early stage of the project provides an important function by helping us to understand the benefits of using such a grid

infrastructure, and thus has helped to move the project forward.

The Nano-CMOS project is still at an early stage of its development, with over three years still to run. As such, much work remains to be done, including enhancing the existing services to support simulation of larger ensembles of devices. In particular, we will pay further attention to metadata management at the process and device simulation level, the circuit level and the system level, and also to workflow management to orchestrate different simulation services.

Other challenges that remain to be addressed include providing more intelligent use of the computational and data resources. In particular, we intend to focus upon supporting computational steering of applications running across resources. In supporting this, we intend to develop services allowing users to visualize the results of particular simulations in real time. This will exploit the results of the OMII-RAVE project[5] in which the National e-Science Centre at Glasgow is involved; this is a continuation of the original RAVE [14] project led by the University of Cardiff.

Acknowledgements

We gratefully acknowledge funding from EPSRC for the work described in this paper. We also wish to thank all project partners (both academic and commercial) for their input.

References

1. "International Technology Roadmap for Semiconductors", Sematech.

2. R.Khumakear et al. "An enhanced 90nm High Performance technology with Strong Performance Improvement from Stress and Mobility Increase Through Simple Process Changes", *2004 Symposium on VLSI Technology, Digest of Technical Papers*, 2004, pp. 162–163.

3. H.Wakabayashi. "Sub 10-nm Planar-Bulk-CMOS Device using Lateral Junction Control", *IEDM Tech. Digest*, 2003, pp. 980–991.

4. A. Asenov et al. "Meeting the Design Challenges of Nano-CMOS Electronics", EPSRC UK-eScience Pilot Project, Octobor, 2006.

5. G. Declerck. "A Look Into the Future of Nanoelectronics", *Digest of Technical Papers for the 2005 Symposium on VLSI Technology*, June 2005, pp. 6–10.

6. "Shibboleth. Web Single Sign-On and Federating Software", *Internet2 Middleware Initiative*, 2007.

7. J. Watt, R.O. Sinnott, J. Koetsier, A.J. Stell. "DyVOSE Project: Experiences in Applying Privilege Management Infrastructure", *Proceedings of the UK e-Science All Hands Meeting 2006*, Nottingham UK, September 2006, pp. 669–676.

8. R.O. Sinnott, J. Watt, D.W. Chadwick, J. Koetsier, O. Otenko, T.A. Nguyen. "Supporting Decentralized, Security focused Dynamic Virtual Organizations across the Grid", *2nd IEEE international Conference on e-Science and Grid Computing*, Amsterdam, December 2006.

9. M. Atkinson et al. "Web Service Grids: An Evolutionary Approach", *OMII-UK*, July 2004.

10. K. Karasavvas et al. "Introduction to OGSA-DAI Services", *Lecture Notes in Computer Science*, Volume 3458, May 2005, pp. 1–12.

11. J. Bradley et al. "The OMII Software Distribution", *Proceedings of the UK e-Science All Hands Meeting 2006*, Nottingham UK, 2006, pp. 748–753.

12. J. Novotny et al. "GridSphere: An Advanced Portal Framework", *Proceedings of the 30th Euromicro Conference*, September 2004, pp. 412–419.

13. R. Sinnott, A. Asenov, et al. "Grid Infrastructures for the Electronics Domain: Requirements and Early Prototypes from an EPSRC Pilot Project", In P*roceedings of the UK e-Science All Hands Meeting 2007*, Nottingham UK, September 2007.

14. I.J. Grimstead et al. "RAVE: Resource-Aware Visualization Environment", *Proceedings of the UK e-Science All Hands Meeting 2004*, Nottingham UK, August – September 2004.

[5] http://www.wesc.ac.uk/projectsite/rave/index.html

Author/s:
Han, L;Asenov, A;Berry, D;Millar, C;Roy, G;Roy, S;Sinnott, R;Stewart, G

Title:
Towards a grid-enabled simulation framework for Nano-CMOS electronics

Date:
2007

Citation:
Han, L., Asenov, A., Berry, D., Millar, C., Roy, G., Roy, S., Sinnott, R. & Stewart, G. (2007). Towards a grid-enabled simulation framework for Nano-CMOS electronics. Fox, G (Ed.) Chiu, K (Ed.) Buyya, R (Ed.) E-SCIENCE 2007: THIRD IEEE INTERNATIONAL CONFERENCE ON E-SCIENCE AND GRID COMPUTING, PROCEEDINGS, pp.305-+. IEEE COMPUTER SOC. https://doi.org/10.1109/E-SCIENCE.2007.78.

Publication Status:
Published

Persistent Link:
http://hdl.handle.net/11343/28838