LinkSCEEM - Assessment of needs draft report

I. Introduction – presentation of activities

As planned in the DoW, the assessment of needs process is based principally on surveys carried out online or through direct requests for information, on the organisation of user meetings and on the setting up of user contact groups.

User surveys:

During the first stage of the project, two user surveys have been conducted; their results are described in section II:

- A simplified survey based on the circulation of a simple form requesting basic information about research interests, current use of computational resources, need for additional resources, and interest for collaboration and/or training in computational sciences.

- A detailed survey, requesting more detailed and technical information regarding HPC applications, to be provided online by participants

Requests for participation in both surveys were made to all the contacts identified within the database, and directly to the participants in the user meetings. As is described below, these two surveys have an interesting complementarity, in that the former has provided limited, basic input from a rather broad panel of potential users, while the latter provided more specific and detailed information about a more restricted panel.

User meetings:

The objectives of the meetings are:

- To provide a forum where users and providers of high-performance computing (HPC) facilities in the Eastern Mediterranean can meet and exchange views and information;
- To explore the available HPC infrastructure in the region, and the ways it currently serves the user communities;
- To provide an opportunity for computational scientists from the region (both users and providers) to learn from the experiences of leaders of HPC centers in Western Europe and the US;
- To lay the foundations of a user community around the planned HPC facility of the Cyprus Institute's Computation-based Research and Technology Research Center (CSTRC).

An initial users' meeting was organised in Cyprus on April 8th, 2008, i.e. on the day that followed the project kick-off meeting, so that all partners could easily be present. It was followed by a series of smaller scale meetings/workshops aimed at local users in various countries of the region:

- In Israel (Tel Aviv, October 26th), co-organised by CSTRC-CyI and IUCC-Tel Aviv University;
- In Jordan (Amman, Nov 6th), co-organised by CSTRC-CyI, JUNet, HIAST and SESAME, with invitations extended to potential HPC users in Jordan, the Palestine territories and Syria. The visibility and attendance of this meeting benefited from the coincidence with the 2nd EU-Med event scheduled for Nov 4th, and the EumedConnect2 Kick-Off meeting scheduled for Nov 5th;
- In Egypt (Cairo), within the framework of the 7th SESAME user meeting, a session specifically devoted to the computational aspects of synchrotron research was held on Nov 20th.

Two additional meetings are already scheduled, in Athens on February 9th, 2009, co-organised by CyI-CSTRC and GRNet, and in Istanbul on February 20, 2009, co-organised by CyI and Sabanci University. Other meetings will be organised in the later stages of the project, notably a second meeting in Egypt, aimed at all scientific fields, targeted for the spring of 2009, and a meeting in Lebanon (date TBD).

The format of these meetings is based on the following model: a first session including a few presentations by LinkSCEEM partners (eg. one on the project itself, one on the regional HPC scene and the development of CSTRC, and one or two other(s) on other relevant subject(s), such as for instance "The user community and applications portfolio of a tier-1 HPC center"), followed by a second session including a number of short presentations by users in various scientific fields, describing their research interests and projects, and the related computational needs. In addition it is useful to ask one or several leading figure(s) of the local scientific community to present an overview of the local computational science & engineering scene. Naturally other meeting formats (eg. thematic, by scientific field or with a focus on some transverse issue, such as visualisation) could also be used in the future.

User database and user contact groups:

The networking process to potential HPC users in the Eastern Mediterranean has progressed in parallel with the contacts with participants of the users' meetings, it is also of interest in its own right, as a first step of the build-up of the future CSTRC user community, and in the short term for the setting up of a user database. The current version of the user database is attached. It will be continuously updated throughout the project, and used to keep current or potential HPC users informed of the activities of the LinkSCEEM consortium, and of the development of CSTRC.

Local user contacts groups - i.e. core groups of users interested in being involved in the networking process, in co-organising meetings, channelling information to other potentially relevant contacts, etc. – are being identified in various countries of the region. The advancement of this process has been – unavoidably – quite diverse, depending on the specific context of each country. While the process has been almost immediate in certain countries (Egypt, Israel, Jordan), it turns out that more time is needed in other cases (Lebanon, Turkey).

Assessment of regional HPC resources

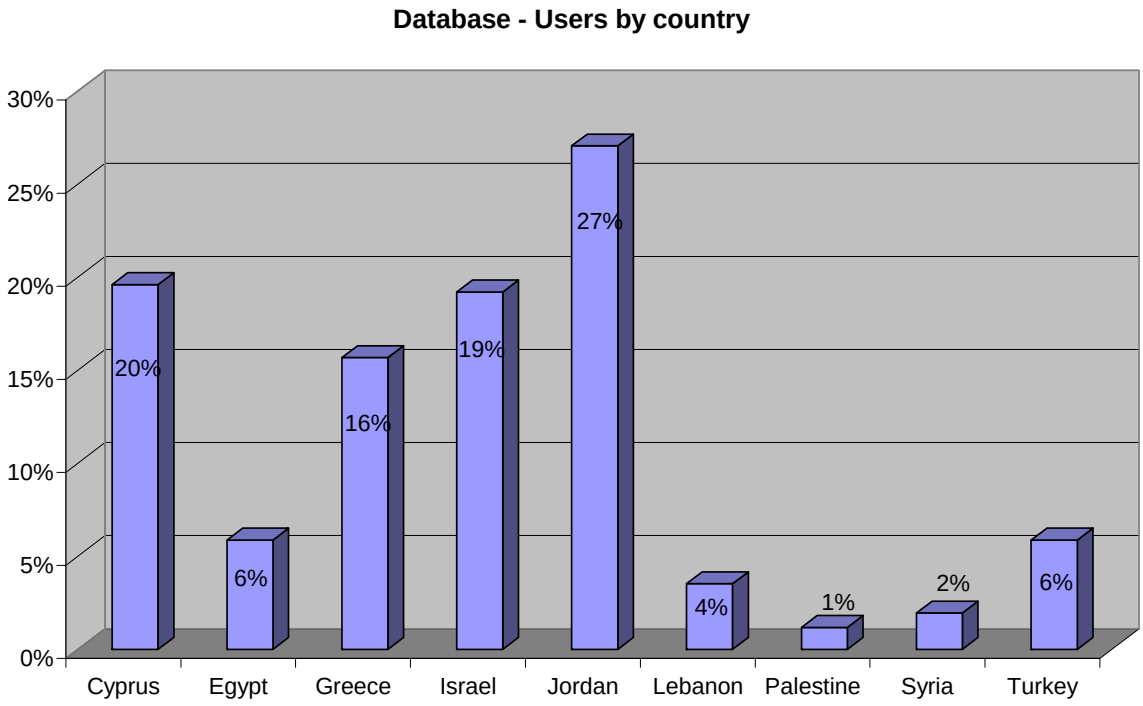
A preliminary survey of existing HPC resources in the Eastern Mediterranean has been carried out, resulting in a document describing the existing HPC infrastructure in the region; an intermediate version (see Annex 4) is available. It would be useful to rework it and provide additional insight, in particular in order to have at least partial information on the applications that are supported by the facilities that have been identified through the survey..

II. Results:

User database:

A database of current or potential HPC users in the Eastern Mediterranean has been set up through the networking process. Its current version (see Annex 1) contains 347 contacts from 9 countries of the Eastern Mediterranean (Cyprus, Egypt, Greece, Israel, Jordan, Lebanon, Palestine, Syria and Turkey). Since its composition provides a preliminary picture of the potential CSTRC users community, it is worth mentioning a few of its features:

- **Distribution by country:**

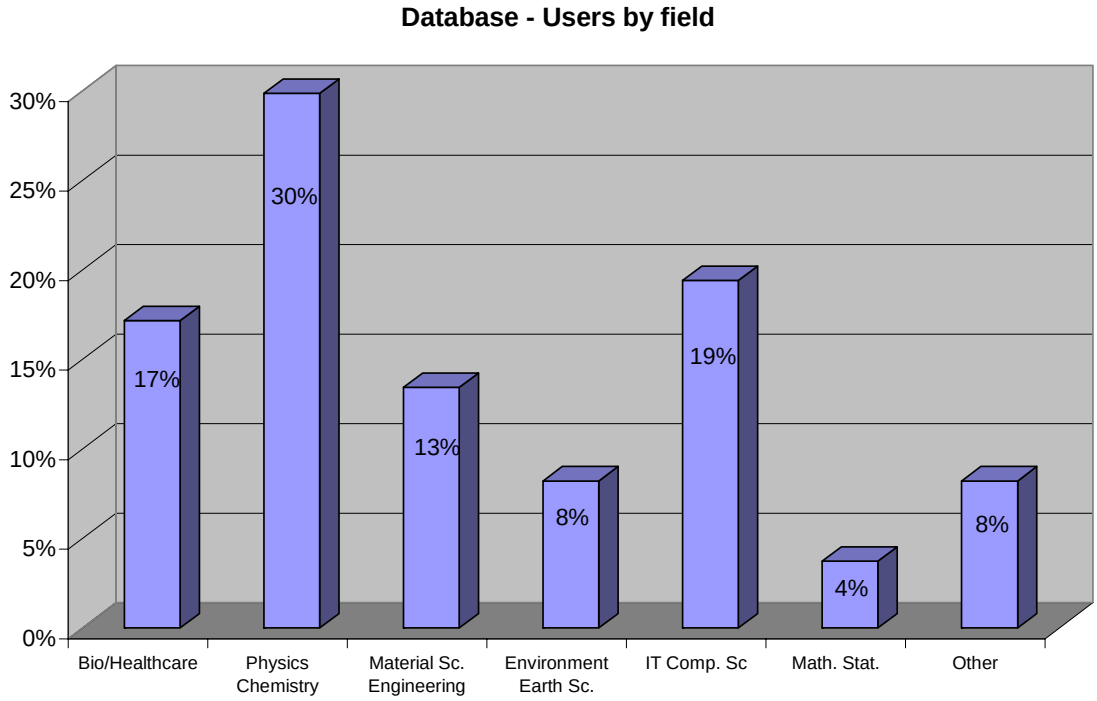


The distribution indicates that potential interest exists in all countries of the region; it seems to be influenced by the following factors:

- Firstly of course the size of each country's scientific community, which explains that the number of Greek or Israeli contacts is, not surprisingly, larger than that of Syrian or Lebanese ones.
- The influence of the users' meetings, which greatly facilitates the networking to a larger number of users. This is apparent for instance in the large number of Jordanian contacts, most of which were identified thanks to the users' meeting that was held in Amman. Conversely, the relatively small number of Egyptian contacts – a country with a very significant scientific community – is expected to increase significantly when a general users' meeting is organised.
- The case of Cyprus is naturally special, since the proximity makes it easier to identify and contact potential users.

- The case of Turkey entails a political specificity that makes it somewhat more delicate than others; it is encouraging however that promising contacts already exist, and that a user meeting is being co-organised with Sabanci University in February, which again should allow significant progress of the networking process.

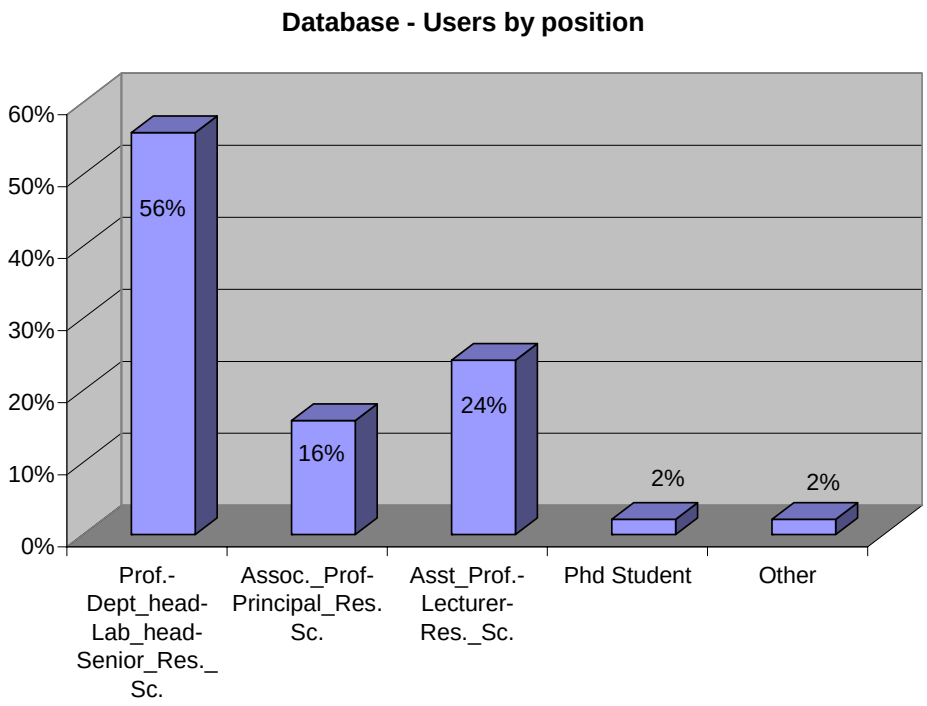
- **Distribution by field:**



This chart shows that the networking process has reached potential users in a great variety of fields. It is worth noting, however, that a few important areas are not (or only marginally) represented:

- Digital Heritage (i.e. the applications of IT to Cultural Heritage); this should be remedied quite easily in coming months, in coordination with CyI's Science and Technology in Archaeology Research Center (STARC). In particular STARC is the coordinator of the STACHEM project, an FP7 Support Action that will foster scientific networking in the fields of Digital Heritage and Archaeological Sciences; there will be obvious synergies with LinkSCEEM, and joint activities can be envisaged.
- Economy/Finance/Management; this requires a specific effort, possibly through the organisation of a thematic meeting.
- The private/industrial sector, which would probably require the carrying out of different networking activities, using specific channels. However some of the project's outreach activities (eg. within the International Conference) should help raise the awareness of decision-makers from the private/industrial sector towards HPC.

- **Distribution by position:**



This chart shows that all levels of the academic pyramid have been reached, however the predominance of relatively senior scientists seems excessive, and shows that the process is not optimal in this respect. Although it is certainly good to create and maintain strong links with scientific leaders, this chart sends a clear sign that an enhanced effort is needed towards young researchers and students. This should probably be done at grass roots level, through various actions that can facilitate direct contacts with young scientists such as user meetings and seminars (making sure that the announcements are made visible to young scientists), setting up of training programmes focusing on young researchers (including contributions to existing ones, tutorials that could be included within meetings, and online training), and targeted outreach activities.

Users surveys:

- **Simplified survey:**

The following form was circulated to all contacts in the user database, and to all participants in the user meetings:

If you are an currently or potentially an HPC user, please take a moment to fill in this form, which will provide us with valuable information on your use of and needs for computational resources, and your interest in related collaborative research and training.

Thanks in advance.

Title		**First Name**		**Last Name**	
Position		**Organization**			
Scientific field				**Email**	
Telephone		**City**		**Country**	

Research interests (personal and/or within your institution):

Current use of computational resources (personal and/or within your institution):

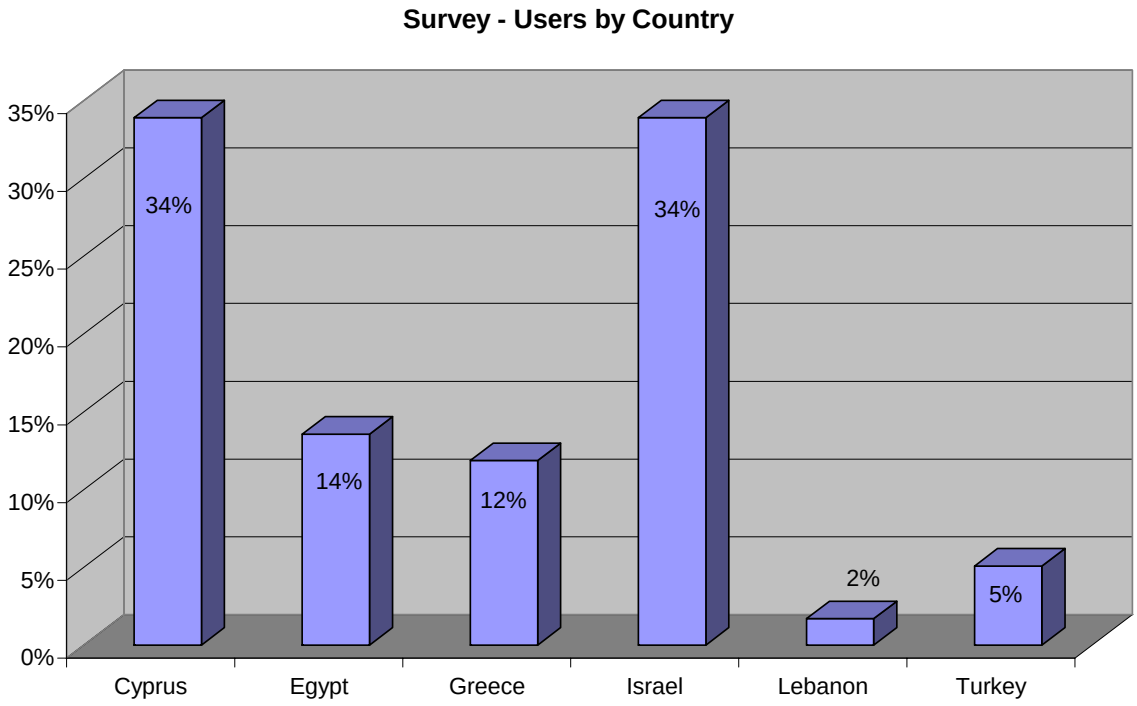
Needs for additional computational resources (personal and/or within your institution):

Interest for collaborative research in computational science (personal and/or within your institution):

Interest for training in computational science (personal and/or within your institution):

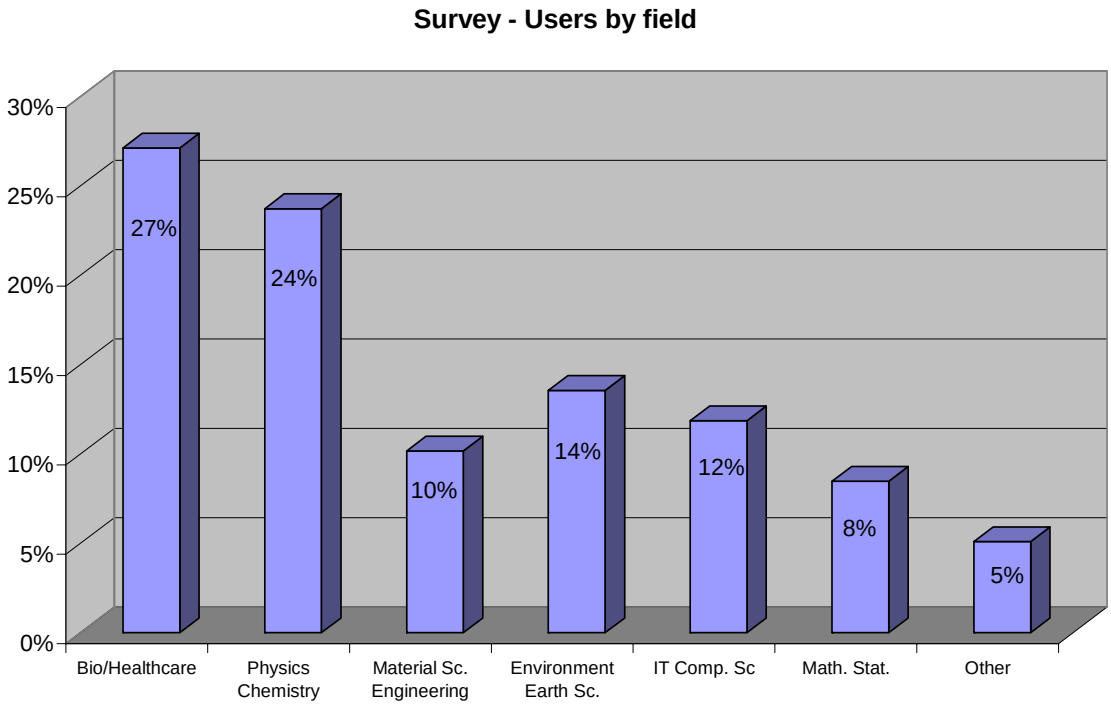
Answers from 59 users have been received. These are compiled in Annex 2; the survey participants form a subset of the user database, with the following features:

- Distribution by country:



Most of the comments that were made above about the database would apply again, with the additional remark that the influence of user meetings seems even stronger for getting feedback from potential users, as can be seen from the comparison between Cypriot and Israeli participation on the one hand, and Egyptian and Greek on the other. However that makes the absence of any participation from Palestine, Syria and especially Jordan, in spite of the user meeting that was held in Amman, even more striking. It seems that the reaction of potential users from these countries to this request for information is somewhat different, and rather cautious; this is also true of other solicitation, as we will see below.
However, combined with attendance to the user meetings, this chart confirms that interest for HPC and computational science exists in all countries of the region, with some specificities.

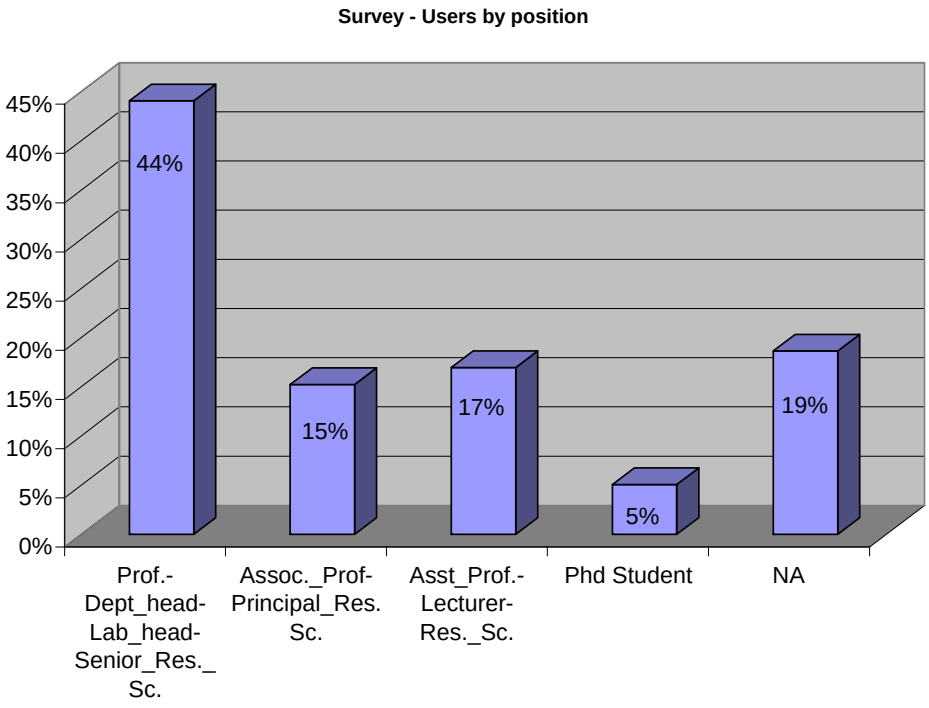
- Distribution by field:



The comparison with the corresponding chart in the description of the user database (see above) shows that the fields of Bio/Healthcare and Environmental Sciences have a stronger representation within the survey participants than they did in the database, while the opposite is true of Physics/Chemistry and IT/Computer Science. One may speculate that this is due to the fact that more scientists in the former two fields, with respect to the latter, have computational needs that cannot be met by their current resources, which constitutes an obvious motivation for participating in the survey.
Once the survey has been carried out further and more answers have been received, it would be of interest to carry out a cross analysis of the answers to the last three questions in the form within each separate scientific field and/or within each country. Another factor that may be at work here is that, as we will see below, the fields of Physics/Chemistry and IT/Comp have a strong representation within the participants in the detailed survey, which may have led them to neglect the simplified survey.; conversely, Bio/Healthcare and Environmental Sciences have a much smaller representation within the detailed survey.

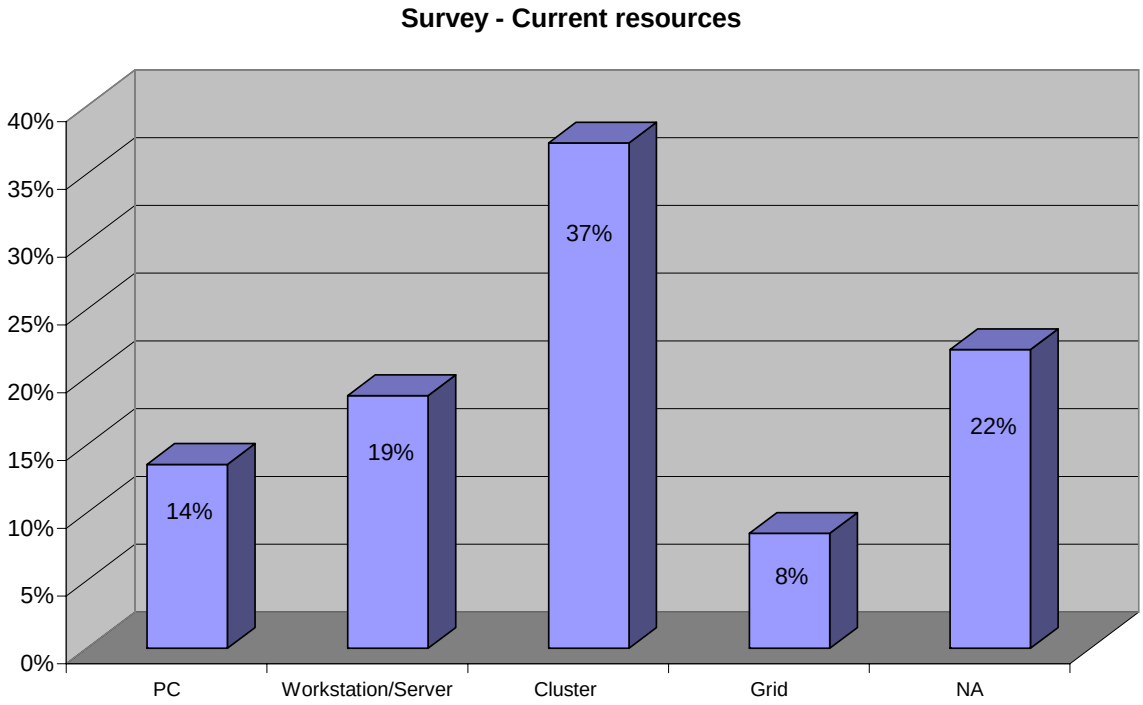
- Distribution by position:

The position chart (see the next page) is very similar to the corresponding one in the description of the database, with an enhanced representation of graduate students. This seems to indicate that all levels of the academic pyramid have more or less equivalent willingness to provide feedback to such surveys. The increased percentage participation from students indicates a willingness to participate from young scientists, which is an encouragement for a specific effort towards this target.



The distributions of the answers to the questions listed in the RFI form are as follows:

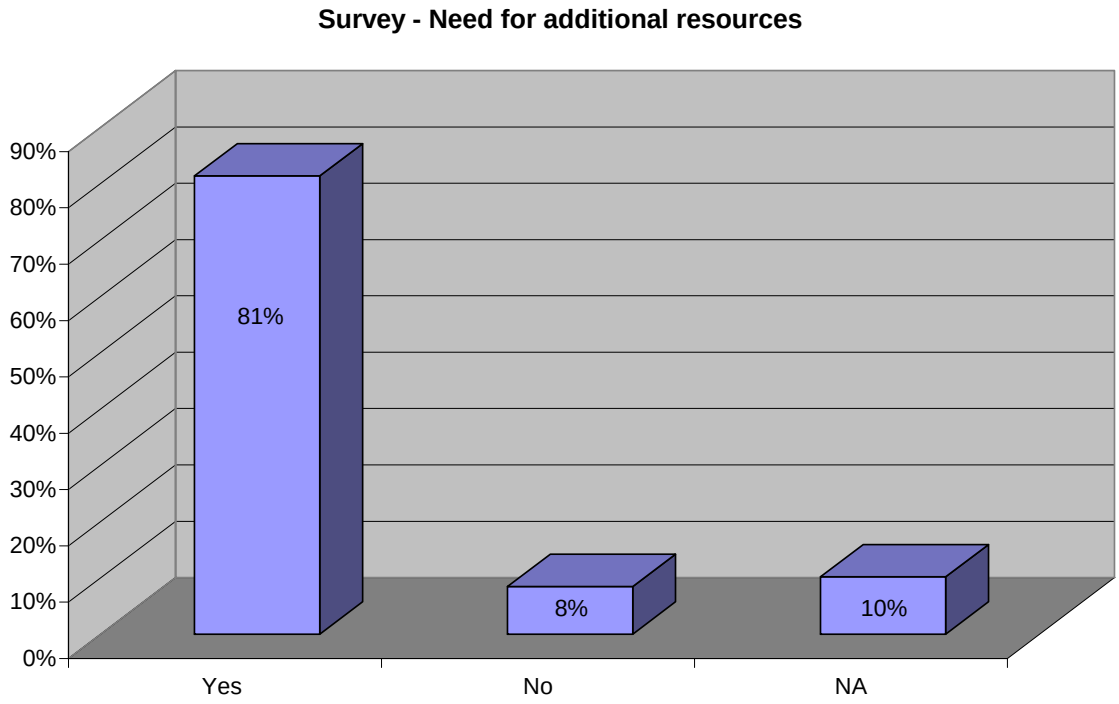
- Current resources:



The main indication of this chart is that only 37% of users have access to cluster-type facilities, implying that the majority of users only have access to very limited computational

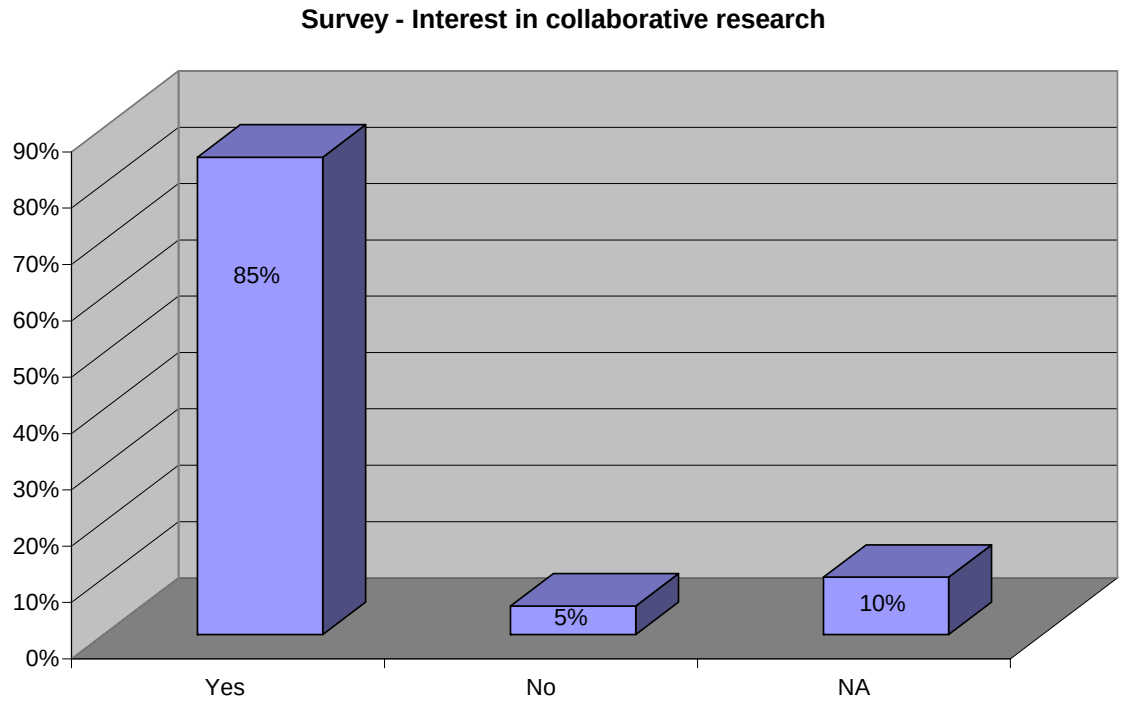
resources. This is in agreement with the answers to the next question, which shows that a vast majority of them feel the need for additional resources, although of course one may argue that any significant user of computational resources is likely to declare a need additional resources. However the detailed survey gives complementary information in the same direction (see below).

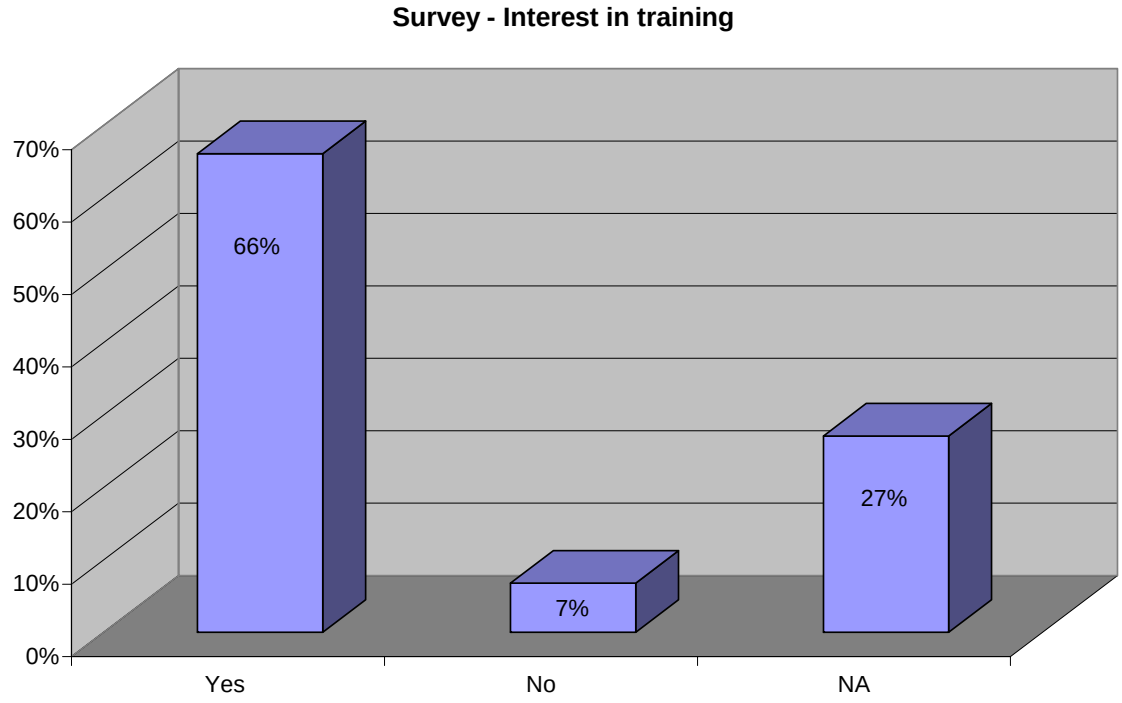
- Need for additional resources:



As mentioned above, an overwhelming majority (81%) of survey participants express a need for additional computational resources, which is of course an encouragement towards the development of the CSTRC infrastructure and the associated networking process, and an important validation of its concept as a regional center of excellence.
Strictly speaking this distribution concerns the participants of the simplified survey, but is very likely true of the user database as a whole, at least to a significant extent. Of course one may object that there is an obvious bias in that extrapolation, in that users who do not feel the need for additional resources may be less interested in participating in the survey.
Naturally this initial picture should be confirmed by obtaining feedback from a larger panel of potential users, but it is not expected that this would bring about radical modifications. It would also be of interest to monitor the evolution of this picture over time. In particular it is likely that actions directed at potential users to raise their awareness of HPC and the possibilities it offers for their research, together with the enhancement of their expertise through training programmes, will result in a significant increase of their perceived needs for computational resources.

- Interest in research collaboration and training:





An overwhelming majority of survey participants (85%) expressed interest in collaborative research in computational sciences, and a somewhat less overwhelming but still significant majority (66%), an interest in training. Again that is an important validation of the regional networking process, and of the fact that the CSTRC infrastructure will be in a position to act as a magnet and catalyst for regional scientific cooperation.

The difference between the interest in collaborative research and training is certainly influenced by the fact that the survey participants are predominantly senior and/or experienced researchers, who are naturally less interested in training. It may well diminish once the networking process succeeds in reaching more young researchers and students.

- **Detailed survey**

The detailed survey was based on the following questionnaire, implemented online by NCSA via the Zoomerang website:

1. Name of Application
2. Please provide a brief description of the application code including what types of problems are addressed with this application?
3. Who develops and maintains this code?
4. Is the code publicly available?
5. What organization supports the program?
6. Please describe the general user community(ies) of this application (who, where, number of users,etc.).
7. Give some specific scientific breakthrough(s) enabled by this code (cite publications). These can be either by you or by other members of the user community.
8. What specific scientific objectives/breakthroughs do you have planned while using this code in the next 3-5 years?
9. Do you need to have local access to data sets maintained for persistent access?
10. Approximately how many lines of code does the application currently have?
11. What methods of parallelization do you use?
12. What parallel programming models do you use?
13. What is your parallelism strategy?
14. Does your application include operations suitable for vectorization?
15. Do you have sequential parts of your code that cannot be parallelized?
16. What HPC systems have you run this code on?
17. What is your preferred HPC system and why?
18. For large application runs, what is the typical upper bound on the number of cores on which you run?
19. What is your typical total memory size requirement?
20. What is your typical per core memory requirement?
21. What is your typical run time (wall clock time) for application runs?
22. Characterize your file space requirements below for input, temporary, and output files.
23. Characterize your file sizes below for input, temporary, and output files.
24. For cluster-based applications, what is your node-level I/O bandwidth requirement?
25. What is you global I/O bandwidth requirement for most application runs?

Characteristics of execution time for PLANNED executions. Please provide information on requirements you anticipate while using this application in 3-5 years.

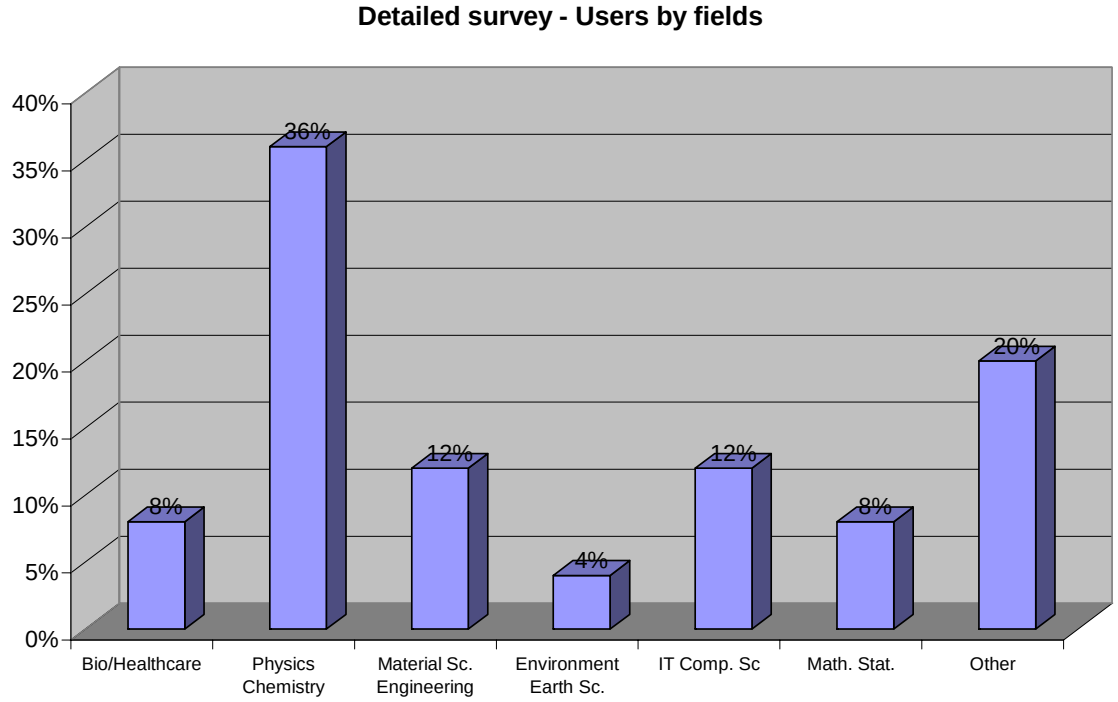
26. What HPC systems would you like to run this code on?
27. For large application runs, what is the expected typical upper bound on the number of cores on which you would like to run?
28. What is your expected typical total memory size requirement for most planned application runs?

29. What is your expected typical per core memory requirement for most planned application runs?
30. What is your expected typical run time (wall clock time) for planned application runs?
31. Characterize your expected file space requirements below for input, temporary, and output files for most planned application runs.
32. Characterize your file sizes below for input, temporary, and output files for most planned application runs.
33. For cluster-based applications, what is your expected node-level I/O bandwidth requirement for most planned application runs?
34. What is your expected global I/O bandwidth requirement for most planned application runs?

Data management: Please characterize some or your data management requirements below.

35. Is the majority of the data you work with generated as output from an application run on large scale resources?
36. Do you obtain data from an instrument, detector or similar data source?
37. Do you obtain or update data from remote data stores (e.g. databases, data archives/repositories, data services)?
38. Do you retrieve from archival storage and reuse or analyze?
39. Do you use data mining techniques?
40. Do you make use of databases (even flat files)?
41. Do you need to have local access to data sets maintained for persistent access?
42. If you need to have local access to data sets, what is the total size of these data sets?

As could be expected, the number of participants in the detailed survey was smaller than in the simplified survey: 25 versus 59. The detailed answers to the questionnaire are available in Annex 3; the information they provide will be very useful for the other LinkSCEEM Work Packages, and for the design and planning process of the CSTRC infrastructure and the associated research and educational thrusts.
In particular, they give insight into the current range of applications employed and developed by prospective CSTRC facility users. These users have a wide variety of interests: industry, medicine, physics, geology, geography, engineering, mathematics and meteorology (Questions 1 and 6), using programs that have led to a number of significant scientific breakthroughs (Question 7). It is worth mentioning that when regrouped into broader scientific fields the distribution of participants (see the chart on the next page) is somewhat different from that of the simplified survey, with a larger proportion of IT/Computer Scientists, and also of physicists and chemists, which was to be expected, since these communities are known to be more versed in IT technicalities (by definition, for the former case) than others.



The survey results unfortunately do not provide explicitly distributions of participants by country or position; indirect indications show that, as could be expected, there is proportionally a larger representation of countries in which computational science and engineering is somewhat more advanced (e.g. Israel), and also a couple of representatives of large international collaborations, notably from the CERN/LHC community.

The main indications provided by the survey can be summarized as follows:

The requirements of these users vary from massive parallel computation to visualization and databasing. As a result of this diversity, the computational requirements of these fields also vary dramatically. At present, some applications are implemented on desktop systems whilst others use the most sophisticated systems available (such as the Blue Gene/P). 80% of respondents use less than 128 cores (Question 18) and 82% say they would be satisfied with 256 cores or less in the future (Question 27).

As regards the memory requirements of the applications, 96% currently use less than 256GB of total RAM memory (Question 19) and 95% would be happy with this figure in the future (Question 28). Given a 256-node system this is just half the standard per core allowance in modern systems. However, 25% of respondents say they need over 2GB of per node memory (Question 20) currently. This drops to 5% in the future (Question 29) indicating that users are either introducing or upscaling the level of parallelization in their codes.

The question of parallelization is an interesting one and is addressed both explicitly and implicitly repeatedly throughout this questionnaire. Implicitly, it is addressed in Question 3, which asks who develops and maintains the code for the applications. Few have indicated that they write all software themselves, most users (57% or more on interpretation of the responses) use commercial, community or collaboration maintained software. If it is being actively used and maintained such software is very likely to have parallel capabilities now or in the future.

However, only 44% of this software is publicly available (Question 4). The other 56% must then be made up of commercial and private software. Since only 4 respondents listed commercial software then these must be collaboration based codes. Of these codes, 52% have less than 10 000 lines (Question 10) and a further 19% with under 50 000 lines. In these cases implementation of parallelization should be reasonably quick with basic appropriate HPC skillsets. 33% of users use the Master/Worker parallel programming model (Question 12) and 68% use domain decomposition as a parallelism strategy (Question 13). 60% of users use MPI as the mode of parallelization (Question 11) but there is little indication of the use of OpenMP or hybrid methods. Given the current popularity of multi-core, tightly coupled systems, all of these factors would indicate significant room for training in this particular programming area.

Indeed 81% of users presently use loosely and tightly coupled Linux clusters (Question 16) with few using non-commodity machines such as the Blue Gene. This 81% persists when discussing future requirements (Question 26). Responses to Question 17, when discussing preferred HPC systems, also support this viewpoint. However, 84% of respondents have indicated that sections of their code are suitable for vectorization (Question 14). Such sections are the basis for the current renaissance in accelerator technologies and usage. Here, again, it would seem that there is significant scope for education. Not unexpectedly, 83% also responded that sections of their codes cannot be parallelized (Question 15), the significance of this depends on the time of execution of these sections. 65% of current utilization is under 24 hours of runtime (Question 21) and this is expected to increase to 76% in the future with a further 21% having a runtime of between 1 and 5 days (Question 30). Again such runtimes are extremely compatible with typical administration policies on Linux clusters.

Other important issues are the questions of data storage and network connectivity. Questions on current and future storage requirements again reveal a wide variation between users (Questions 21 and 31). 95% currently have less than 1TB of input files but 20% of these expect to use more input data in the future. 84% use less than 500GB of temporary storage and this is not expected to change. When it comes to archival storage, 50% currently use less than 100GB but the other 50% spreads across a very large range with 3 users using between 5TB and 25TB and in the future this picture changes little. This data is spread mostly between 10s and 100s of files from medium (100MB to 1GB) to very large files (>1TB) files (Question 23) and the only significant expected change in the future is that some users expect to increase to 1000s of files in the medium size range (Question 32).

The bandwidth requirements for these datasets is currently satisfied by 10MB/s per node in 82% of cases (Question 24) but this drops to 66% when considering future projects (Question 33). 94% of current requirements are serviced by a 10GB/s global I/O bandwidth at present (Question 25) and this remains unchanged in the future (Question 34).

Notably, 86% obtain their data from the output of work on large scale resources (Question 35), only 30% obtain data from an instrument, detector or similar device (Question 36) and 52% obtain or update data from remote stores (Question 38). 73% do retrieve and reuse/analyze data from archival storage. However 61% do not use databases (Question 40) and 59% do not use data mining techniques (Question 39). About half of users need local access to datasets for persistent use (Questions 9 and 41, which were accidentally created identical) with the size of these datasets being less than 1TB for all but one user (Question 42). It would then appear that very large scale storage resources are not immediately required but good network connectivity is necessary. Again opportunities for education in current data technologies present themselves.

Significant work is underway in this user community in a wide variety of scientific areas that have an impact on topics as varied as nozzle-clogging in the steel industry to quantum chromo-dynamics and insect conservation in Egypt to brachytherapy treatment planning (Question 8). All indications in this survey would suggest that a moderately-sized tightly-coupled Linux cluster with perhaps some accelerator technologies and a medium sized storage system could cater to most users who took this survey. Importantly however, a number of areas where there is a lack of expertise have also been exposed. The users in this group vary from the very small scale to the very large scale with the majority closer to the small scale. This, coupled with the actual number of responses received, would indicate that there is significant room for improvement with respect to scientific computing. Were a program of education in place it is highly likely that the computational requirements would increase dramatically.

User meetings:

- **Initial users' meeting, Nicosia, April 8th, 2008**

The meeting was attended by 57 participants, coming mostly from Cyprus and other countries of the Eastern Mediterranean region: Egypt, Greece, Israel, Jordan, Lebanon, Syria, as well as from the USA and western European countries. Their research interests cover computer science, algorithms, computational science, as well as applications of HPC in a large variety of fields, among which mathematics, high-energy physics (including astrophysics and lattice QCD), materials science, chemistry, electrical & mechanical engineering, climate modelling, biology, healthcare (including medical imaging).

The programme comprised a morning session devoted to presentations by LinkSCEEM partners, and an afternoon session devoted to presentations by users in the region, and to a roundtable discussion on ***"The future of scientific computing in the Eastern Mediterranean"*** **(**Co-chairs: Prof. C. Alexandrou (IGB Chair, CSTRC) and Prof. T. Dunning (NCSA)). The meeting's programme, participants list and presentations are available on the project website at http://www.linksceem.eu/joomla/index.php?option=com_content&view=article&id=6:first-users-meeting&catid=31:past-events&Itemid=34. The roundtable discussion provided interesting insight into the users community's point of view, and its summary is worth including in this report:

After a welcome to the participants from C. Alexandrou, T. Dunning made an introductory statement explaining that in the framework of the partnership that has been formed between the Cyprus Institute (CyI) and the National Center for Supercomputing Applications (NCSA), it is of great importance that the design and planning effort of the Computation-based Research and Technology Research Center (CSTRC) be based on an assessment of the needs and demands of potential users in Cyprus and the Eastern Mediterranean region. He argued that it is therefore crucial that potential CSTRC users provide input on the type of problems they address and on the applications they use, notably through meetings such as the present one and through participation in various surveys. The networking process to potential users should also involve the gathering of information on the kind of users' support and training/educational programmes that would be needed, as well as information on the environment of research activities.

The discussion then turned to the issue of the users' needs in various scientific areas. It was agreed that in general there is a clear need for more computational resources, as well as for enhanced connectivity and archiving capacities. Computational chemistry and computational biophysics were mentioned by S. Skourtis (University of Cyprus) as research areas of obvious relevance, in which the creation of research focus groups – a possibility mentioned by T. Dunning – would be of particular interest because of the interdisciplinarity that is needed, and of the variety of computational codes that need to be used. C. Alexandrou mentioned climate modelling, which was also agreed to be an obvious focus, with enormous societal importance in the region. In this field very significant input is expected from IUCC/Tel Aviv University, as a LinkSCEEM partner, and naturally also from the participants in climate modelling activities planned at CyI/EEWRC.

The possibilities of structuring the user communities around thematic research focus groups, or around local contact groups in countries of the region were both mentioned. In addition to a continuous dialog between the users communities, the team in charge of the CSTRC design and planning and the LinkSCEEM consortium, direct interaction and exchanges between users should naturally also be encouraged, as stressed by T. Dunning and A. Stupp (IUCC & Tel Aviv University). This should proceed notably through the setting-up on the CSTRC website of an online exchange system (eg. through a Wiki) where users can add information, exchange ideas, make comments, etc.
As argued by N. Nassif (American University in Beirut), it was agreed that this initial meeting needs some adequate follow-up for a sustainable activation of the users communities. As planned in the LinkSCEEM proposal, other users meeting will be organised on countries of the region, and thematic workshops and summer schools were also mentioned as interesting possibilities.
It was argued that in order to succeed the project should be based on a broad scientific community (i.e. not restricted to local users), and make plans on a timescale much larger than the lifetime of the first investment, so as to achieve sustainability.

It was agreed that, as suggested by A. Sharaf Eldin (Helwan University, Egypt), an assessment of current computational resources would also be useful, as well as a review of existing Masters and Ph.D programmes in computational science and engineering.

The issue of the access of users to the future CSTRC resources was also mentioned, including the peer review mechanism that will be set up. It was argued that the small scale needs that are encountered in the initial phase of many projects should not be neglected, since serving them adequately can be a significant factor of progress; T. Dunning explained that in many US HPC infrastructure, a specific access channel is available for very small needs (up to a certain threshold), which involves no peer review and a very simple application (one paragraph).

The issue of regional network connectivity is clearly of great importance; the achievements of the Eumedconnect collaboration were recalled by F. Karayannis (GRNET, Greece), and their limitations in terms of capacity and topology were recognised (Links only from Eastern-Mediterranean countries to European countries, and not within the region). It was argued that the presence of Jordanian and Syrian participants in the LinkSCEEM consortium could be a plus for finding solutions to the political difficulties that have weighted on this issue. T. Dunning mentioned that his experience in the USA is that in some cases the influence of high level political decision makers (such as State Governors) is needed to make progress on connectivity. Following a remark by M. Gharaibeh (SESAME, Jordan) the creation of the synchrotron light-source at SESAME was agreed to be a regional example of a similar nature, while the example of the conference organised by the PRACE consortium for high-level decision-makers from European industrial firms was mentioned by A. Osseyran (SARA, Netherlands) as an interesting source of inspiration, in particular for the political component of the International Conference that is projected within the LinkSCEEM project.
Following remarks by J. Towns (NCSA), it was therefore agreed that the scientific case for the need for enhanced connectivity should be prepared (climate modelling, synchrotron applications - in collaboration with SESAME - and remote training/virtual graduate schools were mentioned as obvious items). The NRENs and ISPs should be consulted about the related technical requirements, and the case for the upgrade of regional connectivity should then be taken to political decision makers in the region, through the organisation of meetings, of the International Conference and through other outreach actions.

Y. Torman (JUNET, Jordan) argued that political support could reasonably be expected in most countries of the region, but that the decision-making process could be tedious and slow in some cases.

Concerning the training/education programmes that should be set up, it was agreed that they constitute an important ingredient for the build-up of a sustainable users community, in parallel to the users support. The importance of the inclusion in the said programmes of some fundamental aspects of computational mathematics - in addition to the more obvious applications oriented computational science - was stressed by E. Kontoghiorghes (University of Cyprus). Indeed it was agreed that, as mentioned by F. Omara (University of Cairo) it is becoming more and more essential that computational science research get opportunities to interact with computer science and applied mathematics.

The importance of creating synergies with existing regional initiatives was emphasized; this naturally concerns higher-education programmes in computational science and engineering. A regional collaboration, supported by NASA and ESA, concerning astrophysics data analysis, and aimed at junior researchers, was mentioned by A. Ibrahim (Cairo Univ. & American University in Cairo).
It was agreed that engaging industrial partners in the region will also be important.

To conclude, T. Dunning stated that so many difficult problems need to be addressed for the advancement of science and the welfare of the planet, that clearly all available talents will be needed. It was agreed that precisely the Eastern Mediterranean region can contribute to this common effort its human resources, with partly unexploited reserves of talent, and large numbers of young scientists eager for opportunities to engage in, inter alia, computation based science and technology.

- Tel Aviv, October 26, 2008:

The second LinkSCEEM users meeting took place at Tel Aviv University in Israel on October 26th 2008. The meeting was attended by participants from the Cyprus Institute, from Tel Aviv University, from the Technion, and from the Israel Oceanographic and Limnological Research. The meeting consisted of three presentations regarding HPC infrastructures and 7 presentations by users. The main outcome of the meeting was that the infrastructure that is planned by the Cyprus Institute will be beneficial for all the participants, but that it is however, of major importance for the computation center to good network connectivity to the region in order for its resources to be optimally exploited.
The meeting's programme and some of the presentations are available on the project website at
http://www.linksceem.eu/joomla/index.php?option=com_content&view=article&id=53:users-meeting-in-tel-aviv-israel&catid=31:past-events&Itemid=46; the talks given during the meeting are outlined below, together with comments made by the participants.
In summary the meeting was small in scale but interesting, with high-level speakers from many fields – Physics, Engineering, Climate modeling, Socio-Economics, and Biology. The circulation of the meeting's announcement was done mostly through the leaderships of universities, which turned out to produce limited results at the level of end users; in other words the small number of attendants is a sign of the limitations of a top-down approach to regional networking, so that the future stages of the networking process should put more emphasis on the grass roots level and direct contacts with potential end users.

However, due to the small number of participants it was easier for them to get to know each other and discuss needs and possible cooperation. For instance the Dean of the Tel-Aviv University Exact Sciences Faculty, Prof. H. Wolfson, a computer scientist, had an interesting conversation with Prof. Alexandrou, which could open interesting possibilities for further collaboration between TAU and CyI.

Summaries of talks:

Prof. C.Alexandrou (CSTRC & UCy), "Working together in developing HPC in the Eastern Mediterranean Region"

Prof. Alexandrou presented the Cyprus Institute focusing on the Computation-based Science and Technology Research Center (CSTRC), the LinkSCEEM project and its objectives. She explained than in the Eastern Mediterranean region High Performance Computing (HPC) resources are poor and that the aim of LinkSCEEM is to assess the needs of the region's scientists and promote relevant solutions.
Prof. Alexandrou also discussed regional connectivity and highlighted the fact that this aspect is of major importance in order for a computational center in Cyprus to have maximal regional impact. Prof. Alexandrou concluded by stressing that the regional scientific community has the human resources needed to fully utilize such a center, which would greatly promote scientific activity in the area.

Dr. Moshe Gottlieb, Inter-University Computation Center (IUCC) "HPC in Israel"

Dr. Gottlieb presented the IUCC (Inter University Computation Center) focusing on its main areas of activity, i.e. the network, the center of digital information services, the center for learning technologies and the Isra-Grid program. Dr. Gottlieb answered a question from Prof. Weill about the reasons for the failure of the attempt to build a central HPC facility in Israel. Dr. Gottlieb answered that the main reason was that the government didn't fund the project. This raised a discussion on how Israel managed to get good connectivity to the Europe. As Dr. Gottlieb explained, this was achieved through pressure by the scientific community.

Dr. Edward Aronovich, (TAU, IsraGrid) "IsraGrid"

Dr. Aronovich presented the general ideas of Grid and Cloud computing and focused on the IsraGrid. He justified the need to share resources – both data and computation – between scientists in order for the community to have access to a larger pool. For this to be feasible, he mentioned, it is necessary for the different sites to be very well interconnected. Dr. Aronovich then talked about the EGI (European Grid Initiative) and explained that the vision of this project is for every site to have its own infrastructure, which will be connected through a high-speed network. One of the main targets of EGI is to support communities of collaborating users.
Regarding IsraGrid, Dr. Aronovich explained that one of its tasks is to enable the exchange of information, knowledge and experience between industry and academia. An inter-organization is envisioned for IsraGrid to be owned by different organizations where everyone (academia and industry) will be welcome to join. The project secured of total budget $2.6 Million for 4 years.

Dr. Aronovich then replied to a question about the security offered by the Grid computing model regarding unauthorized access to confidential data. Dr. Aronovich answered that this issue could be easily covered by using encryption techniques. Another question was about proprietary software and the handling of licenses and Dr. Aronovich answered that currently there is no global solution about this.

Prof. Benjamin Svetitsky, Tel Aviv University (TAU), "Lattice Gauge Theory for the Large Hadron Collider"

Prof. Svetitsky talked about Lattice Gauge theory for LHC, and explained that strongly coupled technicolor theories require solution of multiple sparse linear systems. For useful results he said, multiple TFlop-years are required.
Then Prof. Svetitsky reported on the resources allocated for similar studies, indicating that US-QCD dedicated 10-50 TFlop facilities at several sites with the US-TeraGrid project allocating time on 20 machines in the 50-TFlop range. He concluded by saying that neither the grid not the cloud computing paradigm are appropriate for this particular problem. In contrast, what is needed is a large scale HPC facility like the one planned by the Cyprus Institute.

Prof. Alexander Gelfgat, Tel Aviv University (TAU), "HPC and pressure-velocity linked CFD solvers"

Prof. Gelfgat explained that CFD algorithms need efficient linear system solvers in order to target the problems of solving the multiple non-linear algebraic equations, and calculating general eigenvalues. Prof. Gelfgat presented an appropriate computation algorithm for these problems, the multifrontal massively parallel solver, which is limited by memory. Prof. Gelfgat concluded by saying that CFD cannot fully exploit massively parallel systems, but rather that for this problem the most appropriate HPC infrastructure is one with as many powerful processors as possible.

Prof. Pinhas Alpert, Tel Aviv University (TAU), "HPC in Climate Modeling"

Prof. Alpert talked about High Performance Computing for climate modelling and identified a number of problems in climate prediction particularly for Israel and the Eastern Mediterranean area, highlighting that there is no HPC infrastructure in Israel that is used for climate prediction. The available resources do not allow the scientists to reach the required resolution, leading to predictions of limited accuracy, especially given the complex geography of the region. Prof. Alpert mentioned that with the available resources it is possible to get a 50KM resolution for the Israel area whereas the Earth Simulator allowed a resolution of 20KM for the whole earth. Regarding the storage requirements, Prof. Alpert explained that Climate Modelling leads to high storage requirements in the order of multiple Petabytes.

Dr. Anne Weill, Technion, "The HPC cluster for nanotechnology"

Dr. Weill talked about the HPC facility that was built in the Technion to support research in the field of Nanotechnology. In order to select the most appropriate machine as Dr. Weill explained, a study focusing on this particular code concluded that a cluster configuration

would outperform an SMP machine. Regarding the interconnection an Infiniband network.was selected. This decision was difficult to make due to the fact that the network infrastructure has a significant cost leading to a tradeoff between the computational power that will be available and the efficiency of the communication. To identify the optimal configuration the different options were evaluated, using applications from benchmark suites as well as synthetic workloads. A question asked by Mr. Stavrou regarded the decisions made for the selection of the most appropriate system that came out to be non-optimal. As Dr. Weill explained the non-optimal design issues regarded the storage.

Prof. Itzhak Benenson, Tel Aviv University (TAU), "The need for HPC in socio-economic modelling"

Prof. Benenson presented a research project which focuses on studying the ratio of parking spaces to incoming cars and its effect on the automobile market and the use of public transportation. For his experiments, Prof. Benenson explained that a large database is used, which provides the input to the simulators. For each study approximately 20000 minutes are needed to run the simulations on a single computer. The code has not been parallelized as the nature of the problem is such that parallelization will not be efficient. However Prof. Benenson indicated that an HPC infrastructure would be very helpful as it would allow multiple experiments to progress concurrently.

Prof. Joan Adler, Technion, "Parallel algorithms for nanoscale atomistic simulations"

Prof. Adler belongs to the computational physics group and studies atom interaction using computational methods. Her research focuses on how the placement of atoms in a structure is affected under different conditions such as cooling and heating. In this line of research both computation and storage requirements are very high. Two different problems are being addressed, one that does not scale to more than 15 CPUs whereas the other scales very well following the throughput model.

Mr. Eyal Privman, Tel Aviv University (TAU), "Probabilistic modeling of protein evolution, and BLAST searches"

During his talk Mr. Privman indicated that his research concerns protein sequence evolution. The tool used for this study is based on a C++ application that was developed in-house. The approach of Mr. Privman is throughput based, allowing optimal scalability based on the possibility of running multiple experiments concurrently. This tool requires access to a large amount of data (multiple Gigabytes) with the output also being on the same order of magnitude. The Grid is not appropriate as it would require transfer of a very large amount of data. An HPC infrastructure however, like the one to be developed at the Cyprus Institute would be a perfect match for this problem.

- **Amman, November 6, 2008:**

This meeting took place in the Wadi Rum Auditorium at the University of Jordan and was hosted by JUNet on behalf of LinkSCEEM. The meeting was attended by over 55 participants from universities and research organizations in Jordan, Cyprus, Syria, Palestine and Lebanon. Network connectivity (both in itself and indirectly through discussions on large data transfers) was a recurring topic throughout the day. Also discussed was the necessity for sufficient stimulation of the user base to ensure full utilization of large scale facilities. It was also argued that systems, both locally and at the Cyprus Institute, should be put in place to deal with capacity as well as capability computing needs since there are institutions in the region where these needs are not being met.
The meeting's programme is available on the project website at http://www.linksceem.eu/joomla/index.php?option=com_content&view=article&id=51:users-meeting-in-jordan-amman&catid=31:past-events&Itemid=46 ; of the 9 presentations, 2 were user presentations, one from Balqa Allied University in Jordan and one from the University of Jordan. These presentations are outlined below along with relevant comments made by the participants.
It is worth noting that, although the meeting got a good attendance (circa 60 scientists), mostly from Jordanian researchers, the users' session turned out to be rather limited, with only two user presentations. This seems to confirm the impression that was mentioned above when describing the user database and surveys, namely that in certain countries potential users, even though showing interest, as the user meeting's attendance indicates, tend to show also a certain caution when it comes to active involvement, such as participating in surveys or making presentations. This is certainly due to a lack of experience both in terms of use of computational resources (so that for instance some scientists do not have a clear perception of their computational needs), and in terms of international cooperation and participation in international meetings, all of which may result in a lack of confidence. These features should certainly be taken into account in planning the networking process and the future user support mechanisms.

The panel discussion, led by a panel comprising Prof. Victor Jongeneel of the Cyprus Institute, Dr. Norbert Attig of the Julich Supercomputing Center, Dr. Assad Sakhel of Balqa Applied University and Dr. Mohamad Gharaibeh of SESAME, took place after the user presentations. During this discussion Prof. Nabil Nassif reiterated his belief, based on the experience acquired at the American University of Beirut, that significant efforts and resources should be placed in the adequate preparation and training of a suitable user base for any proposed facility.
Also discussed was the date the SESAME facility would begin operations and what its data requirements would be at that time. There was also discussion on GRID facilities in the area and the possibility of using cycle-scavenging systems to deal with some of the capability computing requirements at the university level. Prof. Jongeneel commented that multiple solutions for such a system exist but suit only a small subset of applications. Dr. Attig made the point that a smaller scale facility should exist to deal with capacity computing and also as a proving ground for users prior to applying for resources to use the larger facilities.

Summaries of presentations:

Prof. Victor Jongeneel, Cyprus Institute, "The Cyprus Institute, CSTRC and LinkSCEEM"

Prof. Jongeneel contextualized the meeting by introducing the Cyprus Institute (CyI), the Computational Science and Technology Research Center (CSTRC) and the LinkSCEEM project. He discussed how the LinkSCEEM EU funded Support Action aims to ensure that, in planning and developing the CSTRC, strong links are created both to the European ICT infrastructure and the Eastern Mediterranean scientific community. The projects thereby acts as a facilitator to build scientific and technological bridges between Europe and the Middle East and to narrow the digital gap between the Eastern Mediterranean and the Western World.
He also outlined the estimated time-line for the acquisition of facilities. In the Tier system utilized in the PRACE initiative, a Tier 2 facility (with 10s of Tflop capability) is expected to be in place by the end of 2009, with a Tier 1 facility (100s of Tflops) available in 2012. Since Cyprus is now a general partner in the PRACE initiative it can provide indirect access to Tier 0 (PetaFlop) facilities in Europe, and also in America, through its collaborative agreement with the NCSA. He also made the point that the Cypriot government has made a commitment to SESAME to analyze data created there, which will require the upgrading of network connectivity in the region to deal with the necessary data traffic.

Dr. Mohamad Gharaibeh, SESAME, "SESAME Light Source as an International Research Center"

Dr. Gharaibeh outlined the technical operation of a synchrotron facility and how it leads to a high intensity coherent light in the ultraviolet to x-ray region. Since there is a large wavelength bandwidth in the resulting light, it is possible to service multiple users simultaneously using multiple beam lines. He provided a list of the coordinators of various beam lines and their areas of expertise.
With respect to computing facilities, he detailed how crystallographers and those involved in medical imaging would have huge data storage requirements that cannot be addressed at the SESAME facility (in the medium to long term). It is intended that this data will be sent to Cyprus to be analyzed, visualized and archived there. Such a scenario would require dramatic improvements in network connectivity between the two countries.

Prof. Nabil Nassif, American University of Beirut, "Computational Sciences at the American University of Beirut"

Prof. Nassif discussed the development of computational sciences at the AUB. An undergraduate program in Computational Science was introduced in 2001 and a Masters program began in 2007. The programs focus on the mathematical basis and methods for simulation and are aimed at preparing a new generation of students to address emerging demands.
He listed the computer facilities available at the AUB, the most powerful of these being at the Center for Advance Mathematical Sciences. In this particular case, he noted that the facility (when purchased) was expensive but suffered from severe underutilization because few users with the necessary skills existed. It was for this reason that a more practical

approach was adopted, where students are instructed in the use of tools such as MatLab rather than high-level programming languages and root project development. He warned that unless significant educational support was provided, the investment in a facility in Cyprus could turn into a white elephant. Prof. Jongeneel assured him that the development of an educational program and sophisticated user support was a core function of the CSTRC.

Dr. Radha Nandkumar, National Center for Supercomputing Applications, "International Cooperation in Computational Sciences and infrastructures"

Dr. Nandkumar described the current infrastructure in place at the NCSA to assist in the growth and development of supercomputing facilities and usage at national and international levels. She outlined some of the programs that NCSA has implemented to promote collaboration and project advancement.
Particular programs included the Summer Fellowship program where researchers are invited to NCSA for a 3 month period, enabling them to become familiar with the environment and the facilities as well as develop their particular applications. This has operated on a national level for many years and has recently been expanded to the international level.
She also spoke of the partnerships that NCSA has fostered with emerging supercomputer facilities throughout the world, and stated that the cooperation with Cyprus is in a similar vein.

Dr. Alan O'Cais, Cyprus Institute, "Scientific Applications on High Performance Platforms"

Dr. O'Cais described the likely make-up of large scale computer facilities in the near future. In particular he focused on the new types of accelerators that are emerging in these platforms at present.
He discussed the basic makeup and functionality of GPUs, the Cell processor and FPGAs. Explicit examples of their successful utilization in Lattice QCD were given.
He also spoke of the necessary steps in developing an application for these platforms. Such steps include choosing a programming language, library utilization, compilers, profilers, repository use, documentation and other topics.

Dr. Norbert Attig, Jülich Supercomputing Center (JSC), "The users and applications portfolio of a tier-1 HPC center"

Dr. Attig spoke of the development of the supercomputing center in Jülich. The current installation, including a purpose-built building, was completed in 2004. The center itself has been operational for almost 20 years. The Jülich Supercomputing Center is in charge of operation of the facility. The allocation of facilities to users is implemented by the John von Neumann-Institut für Computing (NIC) which operates independently of the center. The NIC allocation is based on scientific merit of the project and the judgment process is conducted independently of the JSC. The JSC still has veto rights once the NIC has approved a project, but these are only based on technical considerations such as the actual suitability of the project for implementation on the system architecture.
Dr. Attig also outlined how the support structure for users has developed in recent times. Previously, it operated as a three tier system with a help desk providing immediate technical

support, a specialist tier for help with computational methods and optimization and an advisor to guide scientific partnerships.
The newly implemented system is that of a Simulation Laboratory (SL). An SL is a targeted research and support organization for a scientific community. It cooperates with its community and supports it in performing simulations on supercomputers. The SLs are adjusted according to the needs of the scientific communities and each SL is itself part of its community and strengthens it. An SL consists of a core group located at a supercomputer and a number of associated scientists at universities and research institutes.
Dr. Attig reinforced earlier comments on the need for a large educational effort when running a supercomputing center. He also stressed that a Tier 1 facility must be fully embedded in the educational network and that active and strong collaborations are essential to the active utilization of such a system.

Dr. Yousef Torman, Jordan Universities Network (JUNET), "Providing computational power and connectivity in the Eastern Mediterranean"

Dr. Tormans' main focus was on the network connectivity in the Eastern Mediterranean region. He listed some of the national connectivity bodies in countries in the area (such as MARWAN in Morocco and AARN in Algeria). He also discussed regional connectivity efforts and in particular the Consortium of Arab and Mediterranean Research and Education Networks (CAMREN).
He showed the extremely limited bandwidth available in the area. He stressed however that, apart from Libya, there exist dark fiber connections throughout the region, but unfortunately political and economic considerations are stunting the growth of the research connectivity.

Dr. Assad Sakhel, Balqa Applied University, "Tunneling of bosons in optical lattices: A Monte Carlo Investigation"

Dr. Sakhel described some of his recent research in the area of tunnelling in optical lattices. Optical lattices are formed by the interference of counterpropagating laser beams, which creates a periodic (in space) intensity pattern. Atoms are cooled and congregated in the potential minima. The resulting system of trapped atoms resembles a crystal in the sense that the atoms are in a periodic potential. Because of quantum tunnelling, atoms can move in the optical lattice even if the well depth of the lattice is higher than the energy of the atoms. Atoms in an optical lattice provide an ideal quantum system where all parameters can be controlled. Thus they can be used to study effects that are difficult to observe in real crystals.
With respect to his own particular research, Dr. Sakhel stressed that he had nowhere near enough computational power to simulate interesting systems (with 100000s of bosons). He has limited use of a GRID facility in Italy and his needs cannot be met locally.

Dr. Iyad Al-Qasir, University of Jordan, "Thermal and Neutronic Properties of Crystalline Materials"

Dr. Al-Qasir described his work which is based on first principle calculations in nuclear engineering. He uses the density functional theory formalism to address these problems and utilizes the VASP software package to perform simulations. This package performs ab initio quantum mechanical molecular dynamics (MD) using Vanderbilt pseudopotentials.
He commented that the University had contacted IBM to assist in setting up a center for computational physics but that a solution for funding it had not been found.

- **Cairo, November 20, 2008:**

This meeting took place at the Faculty of Science at Cairo University and was hosted by SESAME as a joint LinkSCEEM-SESAME specific session within the 7th SESAME User Meeting. The main objective of the session was to assess the computational needs related to processing SESAME synchrotron data, and how CSTRC could contribute to serving them. There were a total of 5 presentations. Two of these were by invited speakers: an overview of computing in the Eastern Mediterranean, and of the LinkSCEEM project by Prof. C.Alexandrou and a keynote presentation reviewing the computing requirements and facilities at the European Synchrotron Radiation Facility (ESRF) by Prof. Rudolf Dimper. There were then 3 user presentations from the Egyptian academic community by Drs. Tariq Mahmoud, Ahmed Shawky Moussa and Mohamed Abouelhoda.

Summaries of presentations:

Prof. Constantia Alexandrou, Cyprus Institute, "Working together in developing Computational Science and Technology in the Eastern Mediterranean Region"

This presentation looked at the current stage of development of computing in the Eastern Mediterranean region in relation to the rest of the world. It highlighted the current initiatives in Europe and the United States to develop petascale computing systems open to a large academic community.
It is envisioned that the Computation-based Science and Technology Research Center (CSTRC) would act as a national *and* regional computational facility that could provide a gateway to even larger computational infrastructure in Europe and the USA (through PRACE for the former and through its cooperation with NCSA on the latter). Prof. Alexandrou also briefly described the research centers of the Cyprus Institute and how the CSTRC functions in a supportive role with respect to the other Centers. She outlined the goals of the LinkSCEEM project and how it can be used to make contacts with scientists in the area and promote collaboration.

Rudolf Dimper, European Synchrotron Radiation Facility, "The IT infrastructure at the ESRF – today and tomorrow"

This presentation gave an overview of the current make-up of the ESRF, in particular in relation to the amount of data the facility generates and how that data is handled. Particular examples were given where raw data of ~400GB was processed for 1000 CPU hours and generated >4TB of processed data. Mr. Dimper outlined the current data storage policy of the ESRF and noted that due to the sheer volume of data produced that data is only stored for a maximum of 6 months with archiving left to the users which has led to a number of recent proposals to the problems this policy creates.
The presentation also outlined the actual infrastructure of the facility in terms of network, compute and storage facilities. With current expansion rates it is expected that in 10 years the facility will have to deal with up to 1 PB per day! He outlined the software, computational and storage advancements that this would require.

Dr. Tariq Mahmoud, University of Munich, "GRID computing for LHC"

Dr. Mahmoud discussed current GRID computing initiatives at the LHC that are being developed to deal with the enormous datasets that will be developed at the LHC. It is expected that 15PB of data per year will be generated at the site.

Dr. Ahmed Shawky Moussa, Cairo University & Ministry of Communications and Information Technology, "Computationally intelligent HPC for computational science"

Dr. Moussa gave a presentation on the interesting idea of fuzzy (or soft) computing, which refers to a set of computational techniques in computer science which study, model, and analyze very complex phenomena: those for which more conventional methods have not yielded low cost, analytic, and complete solutions. Hard computing is bound by a computer science concept called NP-Complete, which means, in layman's terms, that there is a direct connection between the size of a problem and the amount of resources needed to solve the problem (there are problems so large that it would take the lifetime of the Universe to solve them, even at super-computing speeds). Soft computing aims to surmount NP-complete problems by using inexact methods to give useful but inexact answers to intractable problems.

Dr. Mohamed Abouelhoda, Nile University and Cairo University, "HPC for bioinformatics"

Dr. Abouelhoda discussed the use of cluster computing to service the needs of bioinformatics. He spoke of particular projects including the PCAP program for assembling the mouse genome, protein structure prediction and protein folding based on genetic algorithms. Packages that are used in particular in bioinformatics include BLAST and COCONUT. At the Nile University these are implemented on a 14 node cluster, each node has 2 quad-core 1.6Ghz processors

III. Comments, conclusions:

A. Assessment of needs:

The main conclusion of this report is of course the existence within the Eastern Mediterranean ofsignificant interest and needs for computational resources, and for related research and training programmes; this has been clearly demonstrated during the networking process, notably through the simplified survey and during the presentations and discussions that occurred in the user meetings. This constitutes an important validation of the concept of CSTRC as a regional HPC facility and Center of Excellence for computational science.
The results described in section II will provide useful input for other Work Packages of the project, notably for the preparation and incubation of research and educational collaborations (WP4) and the Network Study (WP5). Indeed many opportunities for collaborative research can be derived from the content of the user database, from the simplified survey and from the user presentations in the various meetings. In addition obvious needs for training and education are apparent from the users surveys, since whilst there are certainly a number of users proficient with parallel computing, the majority are lacking basic skill-sets in this area. Education must therefore be a core component of the CSTRC's development; it should take a number of forms, and even a tiered structure, to adapt to the different skill levels already existing in the community:

- Education in community-maintained and commercial software packages and tools
  For a large number of users, package software is already available to address their research needs and does not require major programming contributions on their part. An educational program that informs Users of such software, and how it is used, could greatly benefit users in the short term and is a highly appropriate first step for those unfamiliar with HPC environments.
- Education in parallel programming models and tools
  For those who already have programming experience, but not necessarily in a parallel environment, it would be useful to have introductory courses on parallel programming models, concepts and tools. These might include, for example, an introductory course in MPI and in the use of optimisation tools.
- Education in emerging HPC technologies and tools
  For those familiar with a parallel environment there is still scope for further education. Following the example of PRACE, one can learn of emerging parallel-specific languages such as Chapel, Fortress or X10 as well as programming techniques for accelerator technologies like GPUs and FPGAs.

A related question is how the users' perception of their own needs may evolve in time, in parallel to the increase of their resources and expertise; it will be of interest to monitor that evolution, and perhaps to get insight by evaluating the said needs before and after taking part in a training programme.
It is also worth mentioning that the Computer Science community, interested in large systems, large databases, etc. from a theoretical or practical computer science research perspective, could make a very valuable contribution to future training and advanced user support activities, that could gain significantly from an active interface with basic IT/Computer

Science research. This component should be kept in mind in relation to the incubation of the research and education collaboration.
It is also clear that CSTRC must actively introduce the regional scientific community to HPC. It should not only provide an educational structure that addresses all levels of experience, as mentioned above, but also create a user policy and process that encourages usage, particularly for new users. In this respect the setting up of a specific access channel for very small needs, with a simplified application process and enhanced user support where needed (e.g. for new users) could be very favourable.
It is worth noting that the considerations in the two previous paragraphs illustrate the interesting complementarity of the two surveys: potential new HPC users certainly exist within the participants of the simplified survey, while insight into the interface between advanced user support and research on emerging HPC tools can be provided by the detailed survey.

Concerning the Network Study, there is no doubt that the assessment of needs process has confirmed its relevance, since network connectivity has emerged as an obvious concern for the scientific communities in the whole region (only Greece, Israel and possibly Turkey have a different status in that respect, although scientific communities in these countries are also interested in regional connectivity). This topic was stressed in the Nicosia, Amman and Cairo meetings, and was also raised in the Israel meeting. It seems that the consensus among users is that this problem is more political than technical. The solution requires both large investments by governments and cooperation among them. Even if financing could be found to lease more bandwidth, connecting through Israel will be problematic for several countries. The situation concerning Turkey is not clear, and a user meeting there will certainly help to clarify it.

In addition the outcome of the assessment of needs, and notably the analysis of the detailed survey (see section II above), will also provide interesting input for the design of the future CSTRC facility. At the current preliminary stage, it seems that the user surveys indicate that a tightly coupled Linux-based cluster is the system of choice for the majority. This would also be a very prudent choice given that such a system is the most prolific worldwide and a great deal of support and software is publicly available, making the transition to such a system easier for novice users. Accelerator technologies can also be built into such a system meaning that there will still be plenty of raw power to satisfy more advanced users. According to the development plan of CSTRC, a tier-2 machine will be implemented in 2009-10, followed by a much larger tier-1 machine within the next few years; having this type of tier-2 facility will provide a highly appropriate breeding ground for advanced users that can then utilise the necessarily more complicated structure of future tier-1 facility. It can also, given its cluster nature, help address some of the capacity computing requirements of the regional user community.

B. The User Community:

The current version of the database (see Annex 1) provides a preliminary picture of the user community; it seems clear that the setting up of the database and the carrying out of the surveys have been assisted dramatically by the holding of User Meetings. Participation in both of these activities increased significantly once direct contact had been made through the User Meetings. The Amman user meeting is an interesting case, however, where there were a large number of additions to the database without any contributions to the user surveys. This highlights a certain reluctance for exposure in some countries but also a definite willingness

for participation; as mentioned above the specificities of each community should be taken into account in the planning of future activities. .

The database will also serve for the follow-up to potential users and as a basis for future actions. As mentioned above, it appears that, at least in certain countries, such as Israel, an enhanced networking effort, carried out at grass root level and favouring direct contact with end users would be in order. This should furthermore be adapted to have significant impact on Graduate students and young researchers., For certain scientific fields, thematic meetings may turn out to be an appropriate scheme.

Follow-up activities will be carried out in order to inform the contacts that are identified within the database about future LinkSCEEM events and activities, about the development of CSTRC and about other relevant issues. Contact and coordination will be ensured with the members of the user contact groups, in order to involve them in relevant activities, for instance in the preparation of the project's international conference.
Direct contacts between users should also be facilitated, possibly through the setting up of an interactive online platform within the project's website.

Annexes:

- Annex 1: Current version of the users database.
- Annex 2: Compiled answers to simplified survey
- Annex 3: Answers to detailed survey
- Annex 4: Survey of HPC resources