4

PC Grid Computing
— Using Increasingly Common and Powerful PCs to Supply Society with Ample Computing Resources —

MAKOTO TACHIKAWA
Affiliated Fellow

1 Introduction

While the continued penetration of personal computers and the remarkable improvement of CPU processing speed, 80 to 90 percent of most PCs' processing power is untapped, according to a study[1]. This does not mean that many PCs remain turned off, but that the capacity of the CPU, the brain of the PC, is not fully utilized. In case the CPU is more extensively used when a task requiring an enormous number of operations, such as three-dimensional graphical processing, is assigned, it sits idle most of the time during word processing and Internet browsing because CPU processing speeds are much faster than the speeds of input from the keyboard or the communications line.

This fact led to the idea of virtually gathering the power of idle CPUs to use as a computer resource. In other words, this means networking numerous computers to make them work like a single high performance computer, and assigning complex processing tasks to it. The assigned task will be divided into a myriad of small tasks and allocated to individual computers on the grid-like network. Even with increasingly faster CPUs, the power of PCs are not comparable to those of supercomputers, but in a networked environment where individual PCs simply process complex task in parallel, PCs can deliver surprisingly high performance. This is the core concept of PC grid computing, and it came into a reality several years ago.

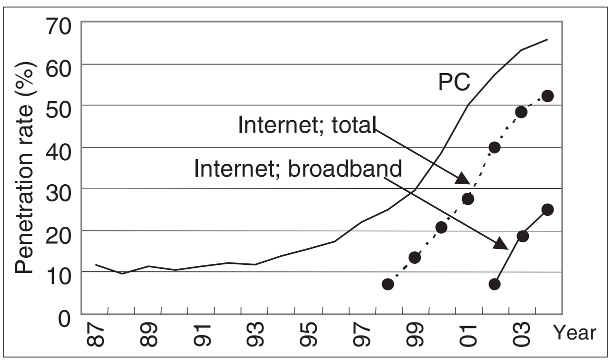
The commoditization and the increased processing speed of PCs leads the growth of idle CPU power. This will facilitate the construction of PC grid computing system along with improvement of communication environment by broadband connectivity. With these trends in mind, this article discusses PC grid computing schemes for supplying society with ample computing resources.

2 Background: Advances in PCs and networks

2-1 Penetration of PCs

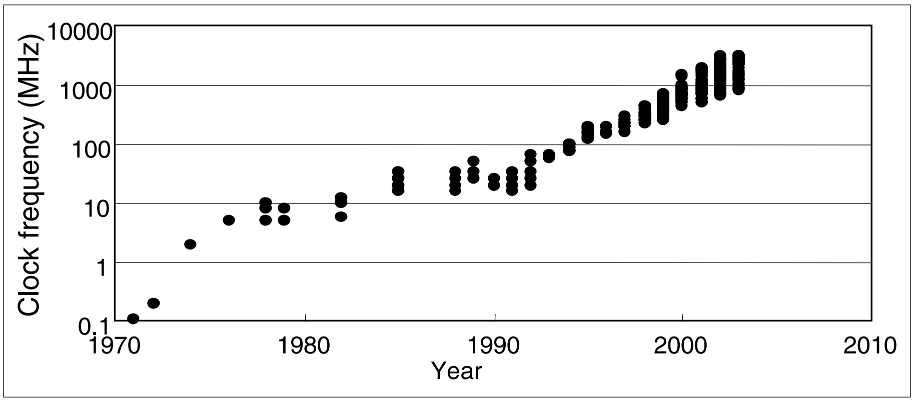
Driven by lower prices and enrich user support, PCs have rapidly come into offices, homes and schools (Figure 1). As mentioned above, growth in the number of PCs means growth in the total CPU idle time across society, because the processors in PCs are most likely under utilized while PCs are running, and many PCs are not even turned on much of the time.

Figure 1 : Penetration rate of PCs and the Internet among Japanese households



Source: Prepared by STFC based on References[2,3]

Figure 2 : Growth in CPU clock frequency (Intel)



Source: Reference[4]

2-2 CPU performance improvement

As Figure 2 shows, the CPU clock frequency has increased explosively, which is another factor of growing idle CPU processing power. Although the increasing of CPU clock frequency is said to be approaching its limit, CPU performance is expected to continue to improve, driven by dual-core CPUs, a technology that includes two logic circuits (cores) in a single processor, and by an increase in the amount of data processed by a CPU per clock cycle (from 32 to 64 bits).

2-3 Diffusion of Broadband Internet

Broadband Internet access has rapidly become widespread in Japan in the last few years (Figure 1). This has enabled PC users to exchange sizable amounts of data over the network, such as modeling data*1 in three-dimensional graphics. The widespread use of broadband access has also allowed running PCs to be remotely monitored in detail because many broadband users keep constant connectivity, facilitating more frequent communications.

3 PC grid computing

3-1 What is grid computing [5,6]?

The concept of the grid is derived from the electric power grid, and the term refers to "an environment in which various information processing resources (computers, storage devices, displays, experimental and observation equipment, etc.) distributed across the network are used as a virtual computer." Grid computing aims to provide the necessary amount of processing resources for its operator, on demand. Its potential benefits are as follows:

- Collection of distributed processing resources for centralized use
- Effective utilization of idle resources
- Load balancing to eliminate the need to maintain the processing capacity to meet the peak load
- Ensured fault tolerance for improved reliability

Grid computing can be divided into the following four types according to configuration.

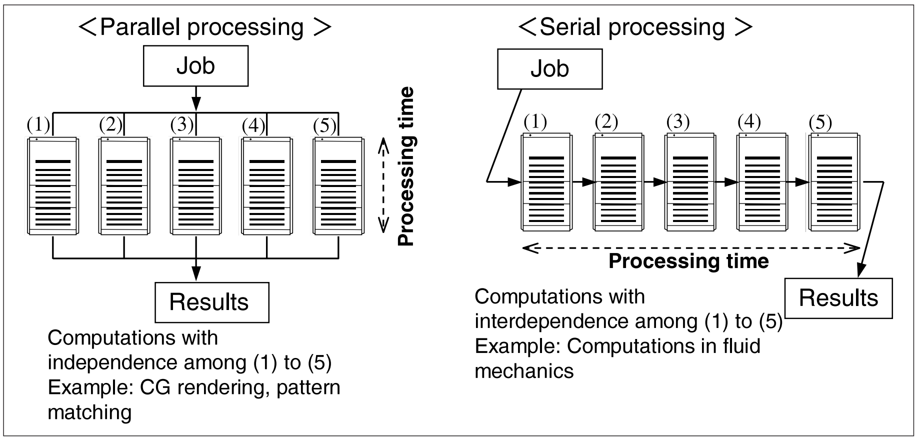
(i) Computing grid: A network of distributed high performance computers (e.g. supercomputers) working like a single huge computer.

(ii) PC grid computing: A concept similar to the computing grid. Collecting the idle CPU power of numerous PCs to perform large-scale processing.

(iii) Data grid computing: Making a grid of disk devices and file systems that is remotely accessible through the network and works like a large external storage device.

(iv) Sensor grid: A group of a myriad of distributed and networked sensors from which data can be collected for specific purposes, such as global environmental monitoring system.

Figure 3 : Parallel and serial processing



Source: Prepared by STFC based on Reference[7]

3-2 Pc grid computing's advantages and constraints

Among the above four types of grid computing schemes, PC grids are constructed PCs, which are inexpensive and widely available, and therefore have the following characteristics:

(1) Easy construction and economical operation

Computing power equivalent to a high performance computer can be achieved at very low cost using a large number of computers connected to a LAN, or the Internet. The ability to make use of the capabilities of a high performance computer with little additional investment can offer a great opportunity as PCs are becoming more commonplace in offices and households.

Since PC grids are comprised of PCs intended for other purposes unrelated to grid computing, operating them as part of a PC grid while their CPUs are idle would have little impact on their operating cost (note 1).

(2) "Automatic" improvement of the total system performance

The performance of the PC and its CPU is steadily improving, and the PCs constituting a PC grid are periodically replaced more powerful ones. This leads to automatically improve the total system performance of PC grid, even while the performance of the constituent PCs varies.

(3) Job independence

PCs are usually linked to external networked PCs via LAN, ADSL or fiber-optic cables, whose data transfer speeds are far slower than CPU processing speeds. Therefore, most PC grid applications are suitable for jobs that can easily be split into small parallel-processing jobs so that each PCs can independently execute them without much communication with other PCs.

Examples of such application areas are bioinformatics for genome information analyses and searches, and CG (Computer Graphics) rendering, which involves creating images from the given numeric data describing an object or graphical figure through computation. In contrast, computations in fluid mechanics require serial processing because of the interdependence of their computational tasks. Such jobs would require supercomputers, which can process a large number of tasks much faster than ordinary PCs. This suggests that the PC grid and the supercomputer are not interchangeable, but are two different solutions to different types of problems (Figure 3).

(4) Expertise, not required to PC owners/users

While the owners or the users of supercomputers are usually computer experts, those of PCs are not. Those who intend to construct and operate a PC grid must take the scarcity of expertise among participating ordinary PC owners and users into account.

4 Mechanisms and technology

4-1 Typical configuration

A typical PC grid consists of a central server and a group of PCs on which special software has been installed, and components that assume the roles shown in Figure 4. A grid in which PCs are connected through the Internet, more project operators are choosing BOINC[8] (Berkeley Open Infrastructure for Network Computing), an open-source platform that was developed at University of California, Berkeley.

4-2 A typical mechanism of operation

A PC grid computing typically works as follows (Figure 5):

(i) Participating PC owners download special software from the Web server and install it on their PCs.

(ii) The special software requests to the central server the application programs and the data that each PC is to process as part of the grid.

(iii) The central server transmits the parallel processing programs and the data to the PCs, divided into packages of appropriate size.

(iv) The PCs run the received programs and data during their CPU idle time as their lowest priority task (note 2).

(v) When the processing is complete, the special software returns the results to the central server and requests new data.

(Steps (iii) to (v) are repeated until the entire project is finished.)

(vi) The central server collects and compiles the results returned from participating PCs into the final results.

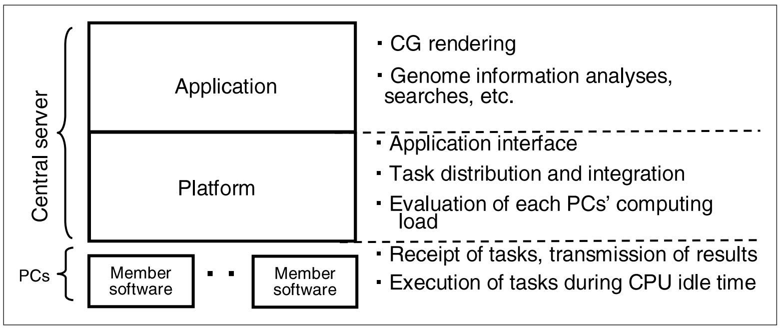
4-3 Technical requirements for implementation

(1) Security protection

Security in PC grid computing must take into account both the participants who offer their PC resources and the member of the project. Major risks for the participants are virus infection and the leakage of personal information. These risks can be significantly reduced by the ensured reliability of the connection destination (the central server in Figure 5) and complete data encryption. A more serious concern is the increased possibility of unauthorized access due to constant connection to the network. However, since this issue is not specific to grid computing, but rather common in the information society, it can be countered by introducing firewalls and frequently checking and patching security holes in software.

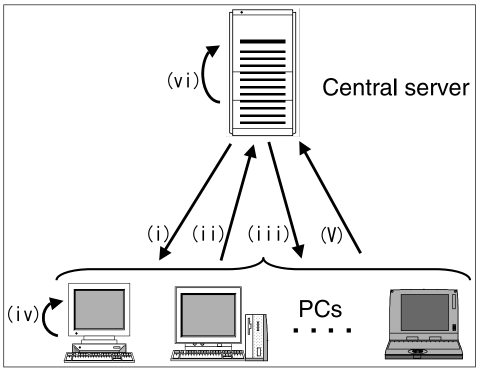
Risks for the project member include data theft and the intentional transfer of incorrect processing results by a malicious participant. Possible measures against these problems are encrypting the data to be stored on the

Figure 4 : Typical configuration of PC grid computing



Source: Prepared by STFC based on Reference[7]

Figure 5 : A typical mechanism for PC grid computing



Source: Prepared by STFC based on Reference[7] participating PCs' hard disks and recomputing unusual results.

(2) Reliability of computation results

Unlike supercomputers and other special-purpose computers, which are installed in a well-controlled environment in terms of power supply, temperature and so on, PCs usually operate in a less desirable environment, and their performance is somewhat unstable. This raises the need to allow for redundancy depending on the state of participating PCs, on the assumption that the processing results returned from the PCs are not always reliable. This need can be met, for instance, by assigning the same task to multiple PCs so that the results can be checked by matching.

5 Classification by structure

5-1 Open structure

This is the most common type of PC grid computing. This type of grid is connected through the Internet and comprised of PCs owned by individuals who are willing to offer their PCs' idle processing power. Such projects sometimes involve a great many PCs from a broad range of individuals. Since participation is essentially on a voluntary basis, providing an attractive incentive is the key to success. Recognizing that sending goods or real money to individual participants is not practical because of huge delivery costs, project operators are finding other cost-effective ways of rewarding the participants, such as sending electronic money, electronic mileage points or other incentive points over the network, or lottery systems.

Instead of giving such financial incentives, other grid project operators choose to appeal to people's volunteer spirit by emphasizing the contribution to social welfare, the search for truth, and the contribution to human advancement. Such projects are sometimes called volunteer computing because participants offer their PCs' extra power for free. A famous example is SETI@home[9], a project operated by the University of California, Berkeley. It aims to search for extraterrestrial intelligence based on data collected with a radio telescope. More than five million PCs voluntarily participate in this project from around the world. The computing power is said to have reached 100 TFlops*2, which is almost comparable to the performance of IBM's Blue Gene/L (approximately 140[10] TFlops), the world's fastest supercomputer[11]. There are many other open grid projects as shown in Table 1, and many of them focus on social consensus such as searching for new anticancer drug agents.

5-2 Closed structure

PC grids in a closed structure are constructed by business enterprises and other organizations, based on their existing PCs. Organizations can have high computing power at low cost, while effective using existing resources. The benefits

Table 1 : PC grid computing projects in an open structure

Project	Research base	Goal
SETI@home	University of California, Berkeley	Find extraterrestrial intelligence
Folding@home	Stanford University	Predict how proteins fold
ClimatePrediction.net	University of Oxford	Test models of climate change
LHC@home	CERN (European Organization for Nuclear Research)	Model particle orbits in accelerator
Cancer Research Project	NCI (U.S. National Cancer Institute) University of Oxford	Search for candidate drugs against cancer
Lifemapper	University of Kansas	Map global distribution of species
cell computing	NTT Data Corporation Toagosei Co. Ltd.	Elucidation of the sequence of bases on human chromosomes: the genome project

Source: Prepared by STFC based on Reference[12] of creating this type of PC grid include the following. Once the organization decides to launch a project, there is no need to consider incentives for participants; the state of the participating PCs can be monitored and managed with relative ease; and since the each participant ID is known, security risks are better controlled than open structure. To construct a grid using PCs within a single building, an organization can purchase a software package that easily integrates PCs connected via LAN into a grid[13].

5-3 Semi-open structure

Where multiple organizations that own many PCs, these organizations can build a single network that extends beyond their boundaries to achieve high computing power. Such a grid would have a semi-open structure and allow public organizations (municipal offices, schools, etc.) and local businesses to jointly provide the local community with shared computing resources. This is a PC grid that makes computing resources available at low cost to local small-to-medium companies that hardly afford to use supercomputers. Even large companies can benefit from such a grid because owning expensive supercomputers is not always an option. Universities and research institutes in the region also enjoy this benefit. When many regions are making a variety of efforts to enhance their information infrastructure, regionally based grid computing scheme would strengthen these efforts. This is the scheme of ‘of the region, by the region, for the region’.

An example of this type of grid project is a field experiment conducted in Gifu Prefecture in February 2005. Led by Gifu National College of Technology, universities, high schools, education boards, research institutes and other organizations in the prefecture participated in the project, offering over 1,000 PCs. The experiment was designed to solve “the traveling salesman problem” for 80 municipalities in Gifu by using parallel genetic algorithms. After the experiment, the institutions involved expressed their expectations for the future if abundant computing resources were to become easily available, including research projects that would otherwise not be feasible, such as highly complex simulations. On the other hand, the experiment exposed social issues, such as whether each organization’s rules permit its PCs to be used for purposes other than the original intent, and how to compensate for the differences in security policy among the participating organizations[14].

6 Future activities

Most existing PC grids are in an open structure, and in semi-open or closed structures have just emerged. In these structures, semi-open PC grids may be expected to have highly public effects. Building a public computing infrastructure by PC grid can lead to provide processing power as some sort of utility services. This is analogous to an environment where water and electricity are supplied as public services so that any entities in need of them can use them at reasonable cost, without having to own their own purification plant or power station. Creating local computing resources and making it accessible as needed to those who require it would bring diverse benefits to the community. This suggests that municipal governments that pursue local revitalization should pay attention to the benefits of semi-open PC grid computing as a public infrastructure. However, practical proposals on how to create semi-open PC grids and how to utilize them effectively have not been establish yet. The first step to moving beyond the status quo would be to conduct feasibility studies from many different perspectives with a view to exploring the potential of PC grid computing. Here are some examples of feasibility study in this direction.

6-1 A proposal for the construction of semi-open grids

• **Utilization of PCs in schools**

Now that computers are part of the curriculum even in elementary and secondary education, schools in every region own many PCs. A total of about 1.5 million PCs are in use in all elementary, junior and senior high schools in Japan, which works out to an average of about 30,000 PCs per prefecture[15]. If these computers were networked to form a PC grid, the resulting computing power would be huge. PCs produced in 2000 or later are usually expected to perform at 1 GFlops[11]. To consider the older machines among the participating PCs, one tenth of this figure, 100 MFlops, is used as the assumable average performance. Provided that the PCs are in an ideal state, the resulting grid would have a computing power of 3 TFlops (=100 MFlops ×30,000 machines). This result is well beyond 500 GFlops, the minimum performance for a supercomputer, indicating that a prefectural grid of school PCs would provide CPU power comparable to a supercomputer.

Despite this potential, there is a hurdle that such a grid project would have to overcome. Since most PCs in schools are intended for student education, school officials would hesitate to permit their use for purposes unrelated to education and might have trouble deciding how much a school should be involved in a regional PC grid computing project that is operated outside the boundary of ordinary school activities. One solution is that the competent authority of each school issues a policy for promoting PC grids. This should happen after some basic issues have to be solved first: finding ways of coping with increases in electricity and other operational costs; having school PCs constantly connected to the network; and establishing a system to ensure the security of connection.

6-2 Potential applications of the constructed PC grids

(1) Developing virtual learning materials for local use

A PC grid consisting primarily of PCs owned by schools and other educational institutions may be suitable for the development of virtual learning materials. For example, it would help visualize what students cannot directly observe, such as dangerous tests, motions in dynamics, and internal changes in materials. PC grids may also be useful for developing learning materials with simulating a social experiment in the local community that is hardly in the real world, and expressing the results with animation.

(2) Promoting the local content business

There is a national vision that Japan should transform itself into a country built on intellectual property. This goal is being pursued through initiatives such as the "Contents Business Promotion Policy,"s[16] which is being formulated by the government's Intellectual Property Policy Headquarters, and Nippon Keidanren's "Intellectual Property Strategic Program"[17]. One type of intellectual property that is expected to grow in significance is high quality, sophisticated digital content. A key to moving forward in this area is building a development infrastructure, as well as strengthening human resource development and facilitating content distribution. Although Japanese animation content called "Japanimation" is already valued highly across the world, the animation industry will face a growing need for computing power, as more animation produced entirely by Computer Graphics and more TV sets come to support high-definition broadcasts. The Japanese movie industry does not have a solid business base, and in particular, its animation sector is in a weak financial position, which makes supercomputers too expensive to afford to them. Therefore inexpensive and ample computing power is desirable, and PC grid computing is suitable one. Moreover, for Japan to pursue balanced local revitalization, the "Japanimation" development infrastructure should be distributed across nationwide. Regional PC grid computing would be an ideal scheme.

7 Conclusion

As society goes more digitized, it will become replete with networked devices with powerful CPUs, including home information appliances such as HDD (Hard Disk Drive) video recorders, home servers (devices that work as the central controller for all information traffic within a home) and online game consoles. These devices are all potential sources of CPU power. This suggests that future PC grids may go beyond PCs. Such changes in the social environment should be recognized as leverages of PC grid computing and taken into consideration in future discussions.

Acknowledgements

The author would like to thank the following people for their very valuable suggestions and advice, and for the provision of valuable reference materials for this article: Ryoichi Shibata, an associate professor of the Department of Architecture, Gifu National College of Technology; Tetsuhiko Yoshida, Chief Scientist of the Corporate Research and Development Department, Toagosei Co., Ltd.; and Shoji Yarimizu, deputy executive manager of the Cell Computing Business Promotion Office, Business Incubation Sector, NTT Data Corporation.

Glossary

*1 Modeling data
Data that define the shape of a model (object), such as the coordinates of vertices and the parameters of equations expressing contours and faces.

*2 Flops (Floating point number Operations Per Second)
A measure of computer performance, equal to the number of floating-point calculations (real calculations) a machine can perform in a second.

Notes

1: If PCs are to operate only for the PC grid, an electricity and other operating costs will be incurred. But these costs would still be required if the target processing were performed by any other means.

2: PCs usually process multiple tasks in parallel, but they actually divide the CPU processing time into very short cycles and allocate them to the given tasks one by one, according to their priority. By assigning the lowest priority to the PC grid tasks, PCs can only offer their CPU idle time for grid computing with no impact on their other tasks.

References

[1] Miller,Michael. Discovering P2P. SYBEX Inc., 2002.
[2] Supervised by Internet Association Japan. Internet White Paper 2004. Impress Corporation, Jul. 2004. (Japanese)
[3] Economic and Social Research Institute, Cabinet Office. Consumer Confidence Survey., the diffusion rate for 2005 (surveyed in March): http://www.esri.cao.go.jp/jp/stat/menu.html#shohi-z (as of July 26, 2005).
[4] Yamada, Hajime. Management of Technology. NTT Publishing Co. Ltd., 2005. (Japanese)
[5] Watari, Masao. Trends in Grid Technology. Science & Technology Trends — Quarterly Review. No. 6, Feb. 2003.
[6] Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology (edition). Grid. Maruzen, 2004. (Japanese)
[7] Yarimizu, Shoji. The Present and Future of PC Grid. Joho Shori. Vol. 44, No. 6, 2003.
[8] je2bwm. Overview of BOINC: http://boinc.oocp.org/intro.php (as of July 14, 2005).
[9] 2001 SETI@home: http://setiathome2.ssl.berkeley.edu/ (as of July 26, 2005).
[10] TOP500 SUPERCOMPUTER SITES: http://www.top500.org/lists/plists.php?Y=2005&M=06 (as of Aug. 12, 2005).
[11] Contributing to science through @home. Nikkei Science. August 2005. (Japanese)
[12] Grassroots Supercomputing. Science. Vol. 308, May 6, 2005.
[13] DNP trading Co., Ltd., “AD-POWERs Gaiyou” (AD-POWERs Overview), http://www.ad-powers.jp/index_ie.html (as of July 19, 2005).
[14] Shibata, Ryoichi and Koike, Hirotaka. Grid System Development Using Local Computing Resources-A report on the Gifu Grid Project Field Experiment. Transcript of the Symposium on Advanced Computing Systems and Infrastructures (SACSIS) 2005, IPSJ Symposium Series Vol. 2005, No. 5, May 18, 2005. (Japanese)
[15] Ministry of Education, Culture, Sports, Science and Technology, “School Computer Ownership and Internet Connectivity: http://www.mext.go.jp/a_menu/shotou/zyouhou/04120301.htm (as of July 27, 2005).
[16] Intellectual Property Policy Headquarters.

Contents Business Promotion Policy (Draft)-National Strategy in the Soft Power Age:
http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/contents/dai5/5siryou5-1.pdf (as of July 27, 2005).

[17] Nippon Keidanren. Toward the Formulation of 'Intellectual Property Strategic Program 2005' :
http://www.keidanren.or.jp/japanese/policy/2005/013/honbun.html (as of July 27, 2005).

Makoto TACHIKAWA
Affiliated Fellow of NISTEP
Research Institute for System Science, Research and Development Headquarters, NTT Data Corporation
http://www.nttdata.co.jp/rd/riss/index.html

He researches information systems from a broad perspective transcending computers and communications and examines their correlation with society, focusing on topics such as individual-oriented health promotion support systems.

(Original Japanese version: published in September 2005)