Draft Summary for Session II of the 7th IWFIPT
***Future Architectures for the Information Society***
September 14, 2007

In his presentation entitled "Ubiquity/Ambient Intelligence", Jean-Bernard Theeten of NXP Semiconductors defined *Ambient Intelligence* as electronic environments that are sensitive and responsive to the presence of people. The ambient intelligent system is characterized by:

- Many invisible distributed devices throughout the environment,
- The devices understand their situational state,
- The system services can be tailored toward your needs and the system can recognize you,
- The system can change in response to you and your environment, and
- It can anticipate your desires without conscious motivation.

Enabling technologies for ambient intelligent systems include; wireless, miniaturization, localization and identification/security. Ambient intelligence devices range from the autonomous such as sensors, ambient networks whose power dissipation is in the microwatt range, to portable devices such as mobile phones, wearable systems with power dissipation in the milliwatt range to stationary environments with dissipation in the watts range.

In the case of autonomous networks, power sources are needed that can provide hundreds of microwatts as needed to operate the system over time spans of a year. As an example of the power consumption for autonomous devices, state of the art transceivers operate at 1 milliwatt for 100 kbps data rates. Major advances are being made in solid-state lighting where achievable lumens per watt for incandescent lights is on the order of 15 while it is expected that Light Emitting Diodes will soon be capable of 200 lumen per watt. Localization technologies include systems like GPS, active RFID, and passive systems like paging. The e-passport was given as an example of an ambient intelligence application.

Promising domains for ambient intelligent systems in the 2005 to 2025 time frame include well-being, urban life, creative industry, elderly care and automotive. Dr. Theeten projected an ever-increasing electronic presence in the car will occur with perhaps automobiles capable of autonomous operation available in the 2025 timeframe. In the health maintenance and care area, there is likely to be continued increases in wearable electronics and remote patient monitoring.

However, there may be dark sides to the prevalence of ambient intelligent systems. This includes the possible loss of life control to 'big brother' systems on the one side to a trend to total dependence on these systems on the other hand. The approach advocated by Dr. Theeten to mitigate some of these worries is to involve the end-user during the conceptual phase and subsequent phases of system development. Finally, he concluded that ambient intelligent systems are not limited by technology but rather the lack of

available service providers. The growth of service providers for ambient intelligence ultimately depends on the creation a value chain for the customer.

Dr. Francky Catthoor of IMEC discussed challenges for future systems architectures (power, variability) and argued that continued scaling is needed to enable the ambient intelligence systems described by Dr. Theelen. An essential requirement for ambient intelligence systems is cheap, interoperable, low power, embedded software platforms. These platforms require energy efficiencies (Operations/watt) that are two orders of magnitude greater than a general purpose microprocessor and they must be available at $1/20^{th}$ the cost. They must operate with downloadable software and perform real-time stream-based processing. These platforms could contain one billion devices and execute on the order of ten million lines of code. There are two growing gaps between the dream of ambient intelligent systems and reality of nanoscale technologies. At the Electronic System Level, there is an architectural gap between system design and platform design while at the Design for Manufacturing Level, there is physical gap resulting from the unpredictability of fabricated devices.

The power efficiency for computation has been growing exponentially to about 200 Gops/Joule as transistor feature sizes has decreased, but progress is slowing. One way to continue to improve computational efficiency is to 'clean' the application software by minimizing factors such as memory transfers, data sizes, and operations. A second tactic is to exploit parallelism to reduce energy per task but this is limited currently due to the lack of software/hardware design tools. Gains on the order of factors of two or three have also been demonstrated by using multiple memories with application-specific wire sizing. With the advent of multi-core structures which do offer performance per unit energy gains, software is becoming the dominant part of the NRE costs in design.

The increased device variability with scaling is driving the need for new variation-tolerant architectures and is, in turn, driving design to statistical methodologies. Run-time adaptation can be used to address variability by operating devices in either low or high threshold modes depending on problem execution. Dr. Catthoor indicated that cooperation across many disciplines is needed if the ambient intelligence dream is to be realized.

Dr. Antonio Gonzalez of Intel addressed the challenge of developing Tera-Scale Microprocessors in the face of power budgets that remain flat at about 100 watts/$cm^2$. This power constraint stems from the very slow improvement in battery energy density and from the limits on improving cooling capability due to form factors and costs. Moreover extreme scaling is causing device performance sensitivities to processes, operating temperatures, and voltage. To complicate design challenges even further, the wear out and radiation sensitivity characteristics of extremely scaled transistors appears to degrading making it imperative to develop new schemes to design fro reliable systems that use unreliable components. Finally, the reachable portions of the chip appear to be declining due to increased communication delays thus driving a trend to more local communications on chip.

It has also been observed that current super-scalar processors can take limited advantage of instruction level parallelism (ILP). Thus, new schemes are needed to exploit parallelism. It turns out that there are several classes of applications that lend themselves to parallel computations. For example, the class of *Recognition* applications includes human-computer interfaces, multimodal detection, statistical computing, clustering and classification, and machine learning. The class of *Mining* applications includes streamlined data mining, web mining, content-based image retrieval, and summarization. *Synthesis* applications include photo-realism, real-world animation, audio/video/image synthesis, and document synthesis. It appears that each of these three classes of applications will lend themselves to parallel computing implementations, and that they are more amenable to being threaded.

The emergence of multi-core processors is underway and it can be shown that they are more power-effective than big cores. As an example, suppose that we have a single core machine with reference performance of unity and that we double the area to realize either a two core machine or a large single core microprocessor of the same area. It can be shown that a decrease of 15% in voltage and corresponding drop of frequency of 15%yields a performance increase of 1.8 for the two core machine relative to the original machine while the large core performance increase is only 1.3.

Multi-core microprocessors offer increased adaptability relative to a single core machine. For example, frequency and voltage can be adapted to maximize power performance since the multi-core machine is less sensitive to these changes. They an also be adapted to manage local variations in process, temperature, and voltage and they are resilient to faults and defects since they can be re-configured more readily. Much work remains to be done however, to take advantage of opportunities in areas such as scalability, adaptability for power efficiency, resiliency, and programmability. Intel has recently reported an eighty-tile teraflop research chip containing 100 million transistors that operates at 62 watts and 3.16 GHz. This suggests that there is significant potential for the many-core architectures for microprocessors that is accompanied by interesting technological and software challenges.

The final presentation, entitled “Human X-Sensor” was given by Dr. Kazuo Yano of Hitachi Central Research Laboratory. His thesis is that sensor information from individuals can be a change agent for human, organizational, and societal behavior that will increase productivity. Moreover, he envisions a change in the utilization of the internet from ‘download intensive’ to ‘upload intensive’ as increased utilization of sensors occurs. (This is supportive of the characteristics of ambient intelligence environments described by session speakers Theeten and Cathoor described previously in this report.) Dr. Yano described an interesting experiment in which he wore, for about eighteen months, a wrist-watch sensor that measured and recorded his motion in three coordinates. He noticed a strong correlation between his activity and his ability to generate creative ideas. In his case, he seemed to be most creative if his days were divided into times of relatively little motion followed by times of high activity. Over forty members of the Hitachi Central Laboratory were involved in this study for several

months and the motion cues to enhance productivity varied from individual to individual. Dr. Yano also contrasted the motion data from two groups, one of which had been very successful with a given project and one that had failed their assignment. He showed data that contrasted the activity levels of both groups and they were markedly different. Dr. Yano indicated that the use of individual sensor data could be an avenue to achieving significantly greater creativity for both individuals and organizations.