Intel Technology Journal

Copyright © 2010 Intel Corporation. All rights reserved.
ISBN 978-1-934053-28-7, ISSN 1535-864X

Intel Technology Journal
Volume 14, Issue 1

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Publisher, Intel Press, Intel Corporation, 2111 NE 25th Avenue, JF3-330, Hillsboro, OR 97124-5961. E mail: intelpress@intel.com.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in professional services. If professional advice or other expert assistance is required, the services of a competent professional person should be sought.

Intel Corporation may have patents or pending patent applications, trademarks, copyrights, or other intellectual property rights that relate to the presented subject matter. The furnishing of documents and other materials and information does not provide any license, express or implied, by estoppel or otherwise, to any such patents, trademarks, copyrights, or other intellectual property rights.

Intel may make changes to specifications, product descriptions, and plans at any time, without notice.

Third-party vendors, devices, and/or software are listed by Intel as a convenience to Intel's general customer base, but Intel does not make any representations or warranties whatsoever regarding quality, reliability, functionality, or compatibility of these devices. This list and/or these devices may be subject to change without notice.

Fictitious names of companies, products, people, characters, and/or data mentioned herein are not intended to represent any real individual, company, product, or event.

Intel products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications.

Intel, the Intel logo, Celeron, Intel Centrino, Intel Core Duo, Intel NetBurst, Intel Xeon, Itanium, Pentium, Pentium D, MMX, and VTune are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

†Other names and brands may be claimed as the property of others.

This book is printed on acid-free paper. ♾

Publisher: Richard Bowles
Managing Editor: Andrew Binstock

Library of Congress Cataloging in Publication Data:

Printed in United States of America

10 9 8 7 6 5 4 3 2 1

First printing: April, 2010

INTEL® TECHNOLOGY JOURNAL
ESSENTIAL COMPUTING

Articles

FOREWORD

Herman D'Hooge
Intel Labs
Intel Corporation

In the history of computing, there have been several major transitions in the use of computing, and each of these transitions has reshaped large sectors of human activity. For example, the first use of computers focused on data processing and on supporting census counts, and later, computers were used mainly for basic accounting functions. Then came the move to individual office workers who became empowered by having a personal computer on their desk to control data, analysis, and presentation. With the advent of networking, the Internet, and the Web, computing was transformed into a tool for communication and information sharing. This tool further transformed how work is done by corporations, how entire industry ecosystems operate, and the nature of the relationship that corporations have with their customers. These same transitions in computing capabilities have revolutionized consumer activity. Today, consumers have surpassed business as the driver for advances in technology.

Through each of these major transitions in computers (from mainframe to personal standalone, to networked, to connected worldwide), we have seen a radical expansion in both the number of people using computers and also in the ways in which computers are used. We may be on the verge of the largest transition of all. At this time, computers are still of limited use and capability: although computationally powerful, they are not engaged in the "soft" aspects of our lives, such as relationships, emotional well-being, and community. We are now poised on the cusp of a new explosion in computer usage, built on the capabilities of computing systems to understand our emotional, personal, and daily lives and to act meaningfully in them. We call the vision for this next wave of computing, Essential Computing.

Enjoy!

Herman D'Hooge

ESSENTIAL COMPUTING: SIMPLIFYING AND ENRICHING OUR WORK AND DAILY LIFE

Contributors

Herman D. D'Hooge
Intel Corporation

Index Words

Sensing
Sense Making
Context-aware
Perceptual Computing
Social
Personalization
Adaptation
Human-computer Partnership

Abstract

How computer systems are used is changing constantly. They have evolved from systems that are used occasionally to perform specific tasks to systems that act as our constant companions, deeply embedded in our daily lives. Intel's Future Technologies Research Lab is driving a broad-based effort to create applications, systems, and technologies that simplify and enhance many aspects of our work and daily lives. The vision that fuels these research efforts is known as "Essential Computing."

Central to the Essential Computing vision is computing that empowers us in the deep personal, emotional, and social dimensions of our lives, a decided contrast to the functional capabilities of computers (i.e., spreadsheets and word processing) that dominate computer usage today. Such Essential Computing systems will need to be keenly aware of and understand individuals, families, and communities as well as other aspects of our social environment. To achieve this, computing systems must acquire new key capabilities, which are the focus of Intel's Future Technologies Research Lab and the theme of this issue of the Intel Technology Journal.

In this article, we provide an overview of the Essential Computing vision, outline its goals and discuss the associated research challenges.

Introduction

Computing and communications technologies are steadily permeating nearly every facet of life and work. An endless stream of advances in processing power, storage capacity, bandwidth, wired and wireless communications, miniaturization, and power efficiency enable the proliferation of technology to the point that computing is always with us.

This trend towards the presence of computing in the "essence" of our lives will continue, as every "thing" in the world seems destined to become intelligent, connected, and collaborative in supporting our lives. Essential Computing is about understanding and creating the next-wave applications, services, systems, and technology capabilities that will push the use of computing and communications technology further and touch the essence of what we hold most dear: ourselves, family, community, and society. Essential Computing involves people trusting technology to mediate their human relationships and to connect to their passions and emotions on an intimate level.

Computing Becomes Essential

In this section we examine just how computing is becoming essential in our lives. We examine the current signs toward this trend, then look at how computers are beginning to sense our emotions and determine our priorities. We end with a look at the role of technology in our homes to manage our health and well-being.

"A future where services will be substantially more personally relevant and individually customized is inevitable."

Current Signs

The first time an on-line website offers you a personal recommendation for what else you should consider buying or whom you should befriend, and the first time that recommendation is "spot on" it may occur to you that technology's ability to know you has crossed a threshold. Already, our personal devices we carry with us everywhere know our location, our movements, our interests. Services are emerging that allow us to discover others in the vicinity who may share our interests (e.g., services, friends, singles interested in meeting other singles) and that allow us to discover who we may want to connect with, or perhaps avoid. While these specific examples aren't new and might even be considered mundane, a future where services will be substantially more personally relevant and individually customized is inevitable.

A second dimension of computing becoming essential is its rapidly increasing pervasiveness in our lives. Although, from a human standpoint, not an end in itself, pervasiveness is an important enabler from a technology standpoint: if technology is omnipresent; i.e., either always with us or near us, it will become useful in ways not currently possible.

Becoming Deeply Essential

At present, our computing devices have only begun to tap what they can observe and understand about us and our moment-to-moment circumstances. Such observation happens via real-time sensing of many types, and when combined with longitudinal analysis and perhaps static profile information, we have a foundation for deep sense making. From what computing devices observe, they can determine what is important to us and what our current priorities are, and they can know what it is that commands or should command our attention. In addition, they can sense our physical, cognitive, and emotional state; what we're doing, and who is around us. The computing system can then use its understanding of our context to help us accomplish what is needed in a given situation.

"When driving my car through a tricky and crowded intersection in an unfamiliar place, it is probably a bad time for my in-car computing system to do anything that would distract my attention from driving."

For example, when driving my car through a tricky and crowded intersection in an unfamiliar place, it is probably a bad time for my in-car computing system to do anything that would distract my attention from driving. An essential computing system would make that determination accurately without my having to be told explicitly. In fact, I would also expect the system to assist me proactively in spotting potential driving hazards and bring those to my attention in real-time in a way that suits the circumstances. When offering to entertain me or make suggestions about possible roadside interests while I'm driving, an essential computing application might focus on the options that are appropriate to and of interest to the current occupants of the car. Technology should be able to understand the social relationships of the people in the car and act in a way that is sensitive to those relationships.

Essential computing systems will also need to improve their ability to understand the physical world we live in. Augmented reality will find uses beyond the task-oriented uses of today, that are related to directions, tourist-guides, museum guides, and other information overlays in highly structured environments. Richer possibilities include highly personalized and social networking applications that are attached to real-word locations. Anyone will be able to decorate the physical world with their information and hyperlinks for others to discover and interact with. This will create new ways to connect with friends and strangers as well as experiences.

Technology will allow groups of people with a common passion to come together in new ways in order to reach shared goals; it will enable people to collaborate socially with those they know, but also with those they don't know but with whom they share a goal or interest.

"As our infrastructure, i.e., our electricity grids, water systems, and road systems are infused with sensors and computation and communication technology, we will be able to gather information on our personal and collective impact on our environment."

As our infrastructure, i.e., our electricity grids, water systems, and road systems are infused with sensors and computation and communication technology, we will be able to gather information on our personal and collective impact on our environment, and therefore better understand how we as individuals and groups engage with our infrastructure. When our infrastructure breaks down, as is sometimes the case in a major natural disaster, we can use the in-built technology to recover our support systems and to help coordinate relief workers.

Moreover, the whole topic of our infrastructure and our use and allocation of natural resources is replete with emotion, both good and bad. This is clearly evident in the current debate on global warming, clean and renewable energy sources, and in the controversies that surround the notion of removing hydro-electric dams to save the salmon, respecting tribal fishing rights, etc. While much of the emphasis today seems to be on the use of "smart" technology (e.g., the smart grid) to help us more efficiently manage our resources and maintain our planet, technology can also help us better understand the complex and emotional sensibilities around these topics, and in some cases resolve the inherent conflicts. How we collectively understand tradeoffs and make decisions about resource allocations that have a profound impact on our future economic development, humanity, and the environment can all be augmented with essential computing systems.

Essential Computing at Home

In the home, technology will become an essential resource for managing our health [1]. Powerful, affordable and easy-to-use diagnostic tools will be available to the individual consumer or family. These tools will have direct insight into the health or health risks of the individuals in a home: insights that would have once been thought of as complex and expensive. Technology will help us monitor, understand, and change our behaviors as they relate to physical and emotional wellness so we can live healthier and happier lives.

Finally, advances in physically or mechanically active computing systems that have the ability to actuate objects in the physical world will also contribute to making computing essential to our daily lives. Many of our household chores will be carried out by personal robots that operate alongside humans as physical partners in the cluttered and unstructured spaces we call home. They will be able to assist us with routine mechanical tasks such as loading and unloading the dishwasher, preparing meals, cleaning the house (beyond just vacuuming floors), and they will help with our mobility around the house. They will also proactively help us with cognitive tasks such as remembering where we left our keys or reminding us to take our medication. However, their use will not end there: personal robots, as embodied forms of computing, will also increasingly cater to our emotional needs such as physical security and companionship. These systems will need to understand how to live with humans, how to be curious, and will have to learn so they can better respond to human needs.

"Many of our household chores will be carried out by personal robots that operate alongside humans as physical partners."

Challenges on the Path to Essential Computing

Realizing the vision of essential computing poses far-reaching challenges across a broad range of disciplines, many of which are the focus of long-range research at the Intel Future Technologies Research laboratory:

- We need to understand, at a deep level, social and human behaviors, attitudes, aspirations, and fears, and we need to understand how these shape relevant *essential* technology solutions.
- Our devices and services will have to observe and make sense of the physical, user, and social context in which they are used.
- Systems are expected to act sensibly and adapt based on their understanding of the context in which they find themselves. Their relationships with humans are also expected to evolve eventually into a human-computer partnership, and this partnership is expected to continue to evolve over time.
- Finally, we must be able to depend on our computing system, if we are to think of it as a partner. Because the system will know virtually everything about us, we must be able to trust it to keep secrets (security, privacy), be responsive and proactive in its actions and keep them at a human pace, and be fully robust so we can allow ourselves to depend on them being there for us all the time, regardless of the place or setting we are in.

"Personal robots, as embodied forms of computing, will also increasingly cater to our emotional needs such as physical security and companionship."

Essential Implies Context-Aware

It is useful to distinguish between three facets of the context in which a system is used: these represent the progression of a system to becoming increasingly essential. The three facets are the physical environment, the person using the system, and the social setting, all represented in Figure 1.

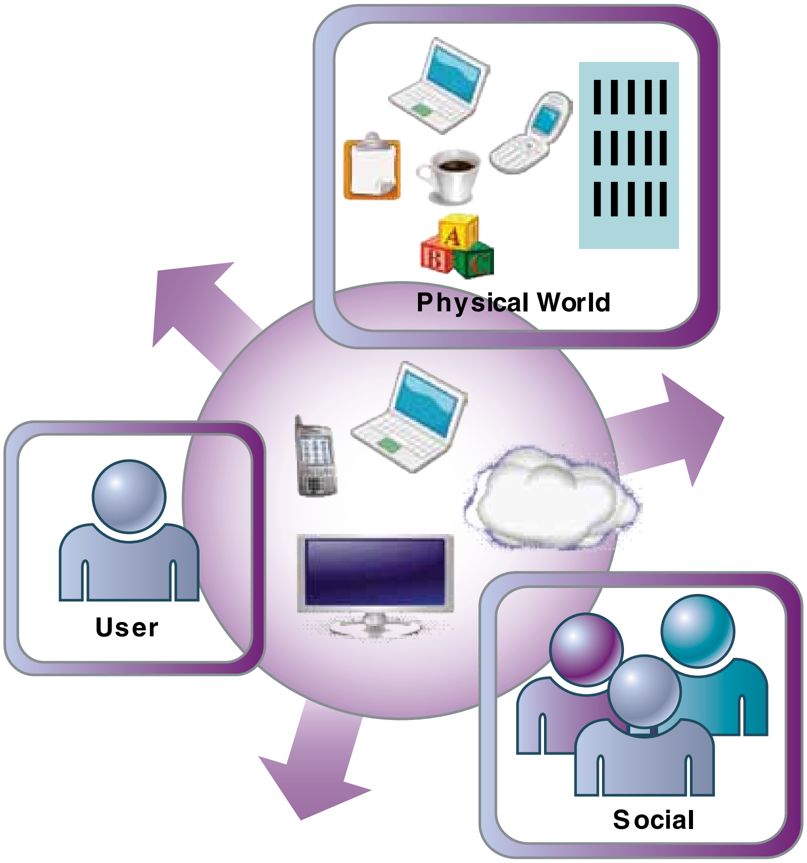


Figure 1: On the Path to Essential Computing
Source: Intel Corporation, 2010

"Combined with static map information, it is easy to infer whether someone is at home, at work, at a coffee shop, in a theatre, or moving around."

Physical-world Aware

Through the use of a global positioning system (GPS) or other dynamic location-sensing technology, the devices we carry with us today already know where they are at any moment, and therefore where we are. Combined with static map information, it is easy to infer whether someone is at home, at work, at a coffee shop, in a theatre, or moving around. It is also easy to infer what points of interest are nearby, and to augment a person's view of the physical world with information. In the augmented reality example described earlier I'm the one doing the augmentation, electronically tagging fixed-location objects in the world as a new way to connect socially.

But awareness of the physical world also must move beyond recognizing static landmarks that are locked in place; awareness of the physical world must also recognize many of the physical objects that are important in our everyday lives that are not guaranteed to stay in one place for long. To do this, object recognition, using cameras, depth sensors, and RFID sensing are necessary. Object recognition enables computing systems to understand the configuration of the world around them. In a multi-sensory system, knowledge about the fixed location environment (home, work, coffee shop) an object is in helps narrow the types of objects likely to be encountered in that environment, and it therefore increase the accuracy of the object recognition. My personal robot, for example, only needs to find its way around my house with my objects in it and to be aware of its own physical presence, so as to not inadvertently run into walls, objects, or humans.

"What one device has learned in one environment should be made available to the other devices, so they can meaningfully adapt their behavior."

Of course, the computing devices in an environment are a special class of physical objects. With more than one smart device per user in many geographies, and the number increasing every day, it is natural for users to expect that all their devices should contribute to their experience becoming more essential. What one device has learned in one environment should be made available to the other devices, so they can meaningfully adapt their behavior and coordinate their actions in ways that improve my experience as the user.

User Aware

It is clear of course that basic physical-world awareness, such as location, can already reveal quite a bit about users. Although some users will find it disturbing, if computing is to become more essential in our personal lives, our devices and services will have to know us on a very personal level.

Some of that information, our personal profile and preferences, that makes up our digital identity, is rather static and evolves relatively slowly over time. The other information, however, is a user's dynamically sensed contextual state. Motion sensing can determine in what activity we are momentarily engaged. Using cameras, facial expressions, and body posture analysis, computers can determine our cognitive and emotional state, and therefore know how we are feeling.

"If computing is to become more essential in our personal lives, our devices and services will have to know us on a very personal level."

In our larger world, as we move about, public billboards, once they become digital, smart, connected, and interactive; and the services they are tied to, will keep their content relevant to what they know about who is actually watching them. Who is watching may be as simple as the gender, age, or lifestyle of that person, but the billboard can also be aware of more personal information: for example, the billboard can have access to profile information a user has shared with the sign's service or it can have an understanding of the user's current emotional state.

Socially Aware

Social networking is clearly a major driver of technology adoption. What started out as separate communications and social-networking applications is rapidly becoming a common system capability integrated into many applications. We expect to be able to instantly and easily share every experience, based on static information about the system's understanding of our social network; that is, the system should understand our different circles of friends and from our profiles be aware of the type of information we're willing to share and with whom we're willing to share it.

"As systems get better and more effective at understanding context, we expect them to adapt their function or behavior autonomously in a way that is meaningful."

Social context is critical for social awareness, and that context is highly dynamic. It might be based on cameras, microphones and/or other sensors to determine who is physically nearby at any given time, and the system should be able to combine that awareness with an understanding of the user's social relationship with these nearby individuals. People recognition may be significantly simplified through multi-sensory techniques: if the system knows a person is at home at a given time of day, chances are that other people physically around that person are family members and not complete strangers.

Essential Implies Adaptation

Making systems more contextually aware is only useful if that contextual information is somehow being put to good use. The system may pass on certain information about context to its user and let the user decide what to do about it, if anything. As systems get better and more effective at understanding context, we expect them to adapt their function or behavior autonomously in a way that is meaningful. As past examples have illustrated, there are many choices in which a system may be able to adapt. In fact, it often feels like magic when the system really gets it right. Unfortunately, when the system gets it wrong, the value of the entire system quickly comes into question. Determining from a human-centered perspective how to best adapt and what level of direct user involvement is appropriate are major challenges in human-machine interaction design. These challenges also make the case for the importance of a machine-learning approach versus one of ready-made device behaviors.

Essential Implies a Human-Computer Partnership

"For technology to become useful and take part in many essential usages and settings, we need new and natural ways to interact with it."

For technology to become useful and take part in many essential usages and settings, we need new and natural ways to interact with it. For example, the system may whisper in our wireless in-ear headset or provide subtle tactile feedback when our personal device is in our pocket. While these new ways to interact are necessary, they are not sufficient. What we really argue is that the overall relationship we have with essential computing is moving us from Human Computer Interaction to Human Computing Partnership. The traditional user-initiated, command-and-reply mode of computer interaction will be augmented by interaction wherein the system takes the initiative and presents us with information relevant to the situation. Humans and machines become partners in decision making.

Essential Implies Truly Dependable

We must be able to fully depend on our computing systems if we are to think of them as a partner and if we are to delegate part of the decision responsibility to our computing systems. Souring a relationship, losing a friendship, damaging our personal wellness because our system wasn't there for us when we needed it are not options. We discussed earlier the awareness and sense-making side of this challenge. In this section we enumerate a few basic technical system attributes that are major challenges.

As battery-operated mobile devices are equipped with always-on, high-bandwidth sensors, their useful life can be significantly shortened. To overcome this, devices will need more sophisticated power management. For example, low-power sensors can be used to detect if any events of interest are about to happen and turn on any energy-hungry sensors only when they have something meaningful to sense.

A significant class of Essential Computing applications, including those related to personal media, robotics, and computational perception, often involve accessing and processing huge amounts of stored and streamed data, and correlating information from a diverse set of sources. For example, a social-augmented reality application may involve a data base lookup, based on GPS information, image search to find matching pictures based on the image of what the user sees, social networking profile access, and so on.

In addition, the complexity of the processing needed by these applications on the data can be significant. The user will expect the system to act instantly, to model human reaction. In our media-rich world that is ever more intensely linked and full of sensor information, our storehouses of data will continue to grow exponentially. To be viable, technologies to analyze, manipulate, and store these data must keep pace and scale with this data explosion.

"Serious concerns about privacy are central to all of the new experiences, applications, and technologies."

Finally, serious concerns about privacy are central to all of the new experiences, applications, and technologies described in this article. When computing systems reach deeply into all facets of our lives, they know more about us (i.e., who we are, who we are with, what we are doing, how we are feeling, our health records, etc.) than we might even know about ourselves; when they take autonomous actions based on that knowledge, questions about how access to that information can be used and abused, and "who controls what" are front and center. Security and privacy not only require new understanding of people, society, and our expectations, but also new technologies and architectures that must be designed and integrated throughout the whole development process?

Multi-Disciplinary Approach

The convergence of workable solutions to the Essential Computing research challenges, that humans will be eager to adopt and trust, demands a holistic approach. The social, human-computer interaction and technology viewpoints all need to be brought to bear at once. To that end, the Intel Future Technologies Research lab has adopted a multi-disciplinary approach involving social scientists, ethnographers, human-computing interaction, and researchers in a range of technical fields. This helps frame the research challenges and shape solutions with a richness of viewpoints not typically found within any single traditional academic research field.

Conclusion

We have painted a vision of computing embedding itself in and adding compelling simplification and enrichment to the essence of our personal lives—a vision that will change today's notion of how we humans relate to computing. Some of the main research challenges to realizing this vision are addressed by the other articles in this issue of the Intel Technology Journal.

We are already seeing progress toward essential computing. Just witness the explosion of new personal devices with sensors that make them aware of many facets of the context in which they are used. Complementary to that, the rise of "app stores" in the past two years that enable rapid innovation and dissemination of simple applications that exploit these sensors has unlocked a wealth of new compelling uses.

"Five to ten years in the future, there is little doubt we'll find ourselves critically dependent on essential computing services which by then will be pervasive."

In the next two to three years, we anticipate a rise in increasingly sophisticated context-aware applications and services that use more sensors, deeper long-term analysis, and combine client and cloud user perspectives. In three to five years we will see a continued proliferation of new devices in unbelievable varieties of sizes, shapes, sensors, storage/compute/networking capabilities that make computing essential to many new facets of everyday life. Five to ten years in the future, there is little doubt we'll find ourselves critically dependent on essential computing services which by then will be pervasive. Additionally, life without these services will be practically unimaginable, just as many pre-teens today can hardly imagine a world without the Internet and SMS texting. In ten years or more we are likely to recognize *essential* as the most important use of computing.

References

[1] D. F. Busch, E. Dishman, et al. “Enabling Healthcare in the Home.” *Intel Technology Journal*, Volume 13, Issue 03, 2009.

Author Biography

Herman D. D’Hooge is Director of the Intel Futures Lab, an external innovation collaboration within Intel’s Futures Technologies Research. D’Hooge joined Intel in 1981 and has held positions in technology research, development, platform architecture, industry evangelism, and management in areas ranging from multi-processor computer architectures, PC system architecture, operating systems, computer security, distributed systems, computer-telephony integration, applications research, branded consumer products (smart toys, PC cameras, and digital audio players), and the application of user-centered design and innovation to technology definition. D’Hooge received his MSc degree in Electrical Engineering and his MSc degree in Computer Science, both from the University of Ghent, Belgium.

Copyright

Intel® Technology Journal | Volume 14, Issue 1, 2010

TABLETOP ESP: EVERYDAY SENSING AND PERCEPTION IN THE CLASSROOM

Contributors

Richard Beckwith
Intel Corporation

Georgios Theocharous
Intel Corporation

Daniel Avrahami
Intel Corporation

Matthai Philipose
Intel Corporation

Index Words

Vision and Scene Understanding
Architectures for Educational Technology Systems
Interdisciplinary Projects
Pattern Recognition
Context Awareness

Abstract

Today's context-aware computers and computing devices can provide significant assistance in performing everyday tasks. Such devices "perceive" enough of your environment that they can see things the way you do—only better. Because they know who you are with, where you are, and more, context-aware computers can intelligently guide or teach you. With their greatly enhanced context awareness, these systems go far beyond "ease of use" and actually provide meaningful, real-time, relevant support to users.

Introduction

We are living in a time of mash-ups. You can listen to The Beatles's pop music mashed up with Jay-Z's rap on Dangermouse's infamous Gray Album. You can find Craigslist* apartments located on Google* Maps by surfing to *www.housingmaps.com*. Popurls'* social news aggregator at *www.popurls.com* looks across Twitter*, Delicious*, Reddit*, and others to see what is popular on the Web right now. You can blend together many things and get the benefits of each—plus synergies. In this article we discuss what happens when you mix context awareness with task assistance. Context awareness helps a machine understand where it is but not just its location. Within that location, context awareness can tell a machine who is there, what they are doing, and what resources (e.g., objects) are available. Context awareness, therefore, helps the machine by providing it with information about its context, but the trick is to make that information about context useful for users. Task assistance is all about helping users.

Within Future Technologies Research, researchers in the group project known as "Everyday Sensing and Perception," or ESP for short, focus on improving context awareness through various means, including parallelization, machine learning, and machine vision. We are also concerned with making context awareness useful. To this end, we have developed a performance support application, which we describe later in this article, that uses ESP-developed technologies to extend context awareness and task assistance to the tabletop in front of the user.

Context Awareness

Apple's iPhone* "face sensor" (or "proximity sensor") provides an example of context awareness. The face sensor is one of those things that makes an iPhone feel more advanced than a typical mobile phone. The iPhone uses its face sensor to determine whether a user's face is near the display. If a face is nearby, the phone can assume that the person is making a call and turn off the keypad so that the user's cheek doesn't make an errant keystroke, mute the call, or worse. The iPhone is not seeing a face, though. It makes an inference from a sensor event. The sensing is accomplished by use of an infrared LED and an infrared sensor, the combination of which creates a very simple but useful sensing mechanism. The LED pulses while the sensor measures reflectance. If the reflectance changes significantly, then the phone can make a series of inferences: (a) an object is very close, (b) that object is likely to be a face, and (c) the user wants the phone's display off, which will save power and save the user from random keystrokes or button presses. In this way, the iPhone uses sensors to make useful inferences.

"The iPhone uses its face sensor to determine whether a user's face is near the display."

Technology can support a wide array of inferences by using many types of sensors. Most mobile phones include a Global Positioning System (GPS), or they can use nearby cell towers or 802.11 access points for location determination [1]. The phone knows where you are and can show your location on a map or give you directions to the nearest coffee shop, for example. This is where some people think context awareness begins and ends. However, today's phones can sense location and much more. There is a large variety of sensors available for small, low-power applications. Mobile devices often include multi-axis accelerometers, gyroscopes, and magnetometers. Our group has shown that by using just a 3-axis accelerometer mounted on a TV remote control, it is possible to identify which member of a family is using that remote and personalize the TV watching experience for that person [2]. Clearly, context awareness can support many more facets of our lives than simply telling us where we are.

"As different kinds of sensors become ubiquitous in computing devices, the ability of those devices to sense context increases significantly, and the urge arises to fuse these data into a large data stream."

As different kinds of sensors become ubiquitous in computing devices, the ability of those devices to sense context increases significantly, and the urge arises to fuse these data into a large data stream. When data are fused, that is, when data are combined from multiple sources, devices can make even more powerful inferences than they can with unfused data. Luckily, the emerging ubiquity of sensors is occurring at the same time as substantial increases in the ability of today's processors to deal with large volumes of data.

"We are entering an era in which mobile computers will be able to use their video cameras for machine vision."

In fact, technology has become so powerful that we are entering an era in which mobile computers will be able to use their video cameras for machine vision. Video cameras are extraordinarily high data rate sensors, to be sure, but as devices become more powerful, machine vision becomes more tractable, and cameras become just another type of sensor that can be used for context awareness. What the camera sees can be analyzed by an application, and that analysis can be used to support a user (just like when the infrared sensor "sees" a face on an iPhone). In fact, our research group believes that computer vision, fused with data from other sources, will be key to making context awareness more useful for people across wider swaths of their lives.

Performance Support

One question that remains is what benefits these inferences might have for users. What reason would someone have for using a device loaded with sensors to make inferences about their context? One answer is that these devices will be used for context-aware task assistance (or performance support).

Electronic performance support tools have been around for decades. These tools are meant to provide what users need, when they need it. The tools are a source of "just-in-time" support. Bezanson offers a good description:

> A performance support system provides just-in-time, just enough training, information, tools, and help for users of a product or work environment, to enable optimum performance by those users when and where needed [3].

Performance support systems must collect contextual data and use those data to assist users. According to Bezanson, when a system tries to support its users but is not context sensitive, at best it can be "sympathetic." Such a system can be designed to be as easy to use as possible, but it will not be able to use contextual information to make things easier still. Alternatively, a system can be context sensitive but not support users. The best performance support is when a product is both context-aware and tries to help users, as with autopilot or some computer-based agents. For a relevant example, tutoring, unlike simple computer-aided instruction, takes contextual information and uses it to personalize a student's instruction.

"The best performance support is when a product is both context-aware and tries to help users, as with autopilot or some computer-based agents "

The kinds of performance that a device can support become more and more complex as context awareness improves. The face sensor could be considered a simple performance support tool for making phone calls. It uses contextual information, and it makes a phone system easier to use. An installation wizard for software is more complex than the face sensor, as it can sense what applications are on a machine and then ask about relevant preferences and other variables that would affect an installation. At the complex end of the continuum, LeafView* [4] is a mobile application that assists in botanical species identification. When a user aims a mobile device's camera at an example leaf, LeafView uses computer vision to select a small number of potential targets (about 5) from a large database of species (about 80,000 type specimens). A recent version of LeafView also uses the GPS on camera phones for geo-locating the sample. By using computer vision and location sensors, LeafView offers performance support that enables normal citizens to perform as scientists.

On The Tabletop

One context that ESP has been investigating is how technology can support users who are seated or standing at a desk or table. The tabletop is an important context if only because people spend so much time there. Elementary students spend much of their school day (about 1000 instructional hours a year) seated at a desk in school [5]. The 2006 American Time Use Survey [6] showed that US citizens 15 years of age and older spent an average of nearly two hours a day at surface-oriented activities. Clearly, this is a frequent context and, just as clearly, sensors and applications will need to know a lot about what people are doing in this context if our devices are to be most useful there.

Tabletop computing, as a paradigm, has garnered some interest in recent years from the research community, but not in quite the way that we are pursuing it. Tabletop computers, like Microsoft's Surface* or the SMART* Table, are really specialized computers with an integral horizontal display surface with which a user interacts. We would like to be able to support traditional activities on traditional tabletops, so we have been pursuing the use of a context-aware portable computer on a normal tabletop.

"We have been pursuing the use of a context-aware portable computer on a normal tabletop."

The applications and algorithms being developed within the ESP project use computer vision and various other sensors. All of this is done in an effort to better understand the context. Our goal is to enable devices to use this context data to improve a user's quality of life. In the case of the system described here, we are most interested in what is happening both on and around the tabletop in order to support education.

Tabletop Tutoring

Imagine if a low-cost laptop could be a world-class tutor for children, using everyday sensing and perception to understand their activities, moods, and knowledge: in other words, creating a personalized curriculum just for them. This is the vision of ClassmateAssist. Our goal is to use context-aware computing to support the current practices of teachers while, at the same time, providing them with additional capacity to instruct and evaluate their students.

ClassmateAssist is an application designed to assist students and teachers in the use of mathematics manipulatives in kindergarten through second-grade classrooms. Manipulatives are physical objects used to assist in the instruction of math concepts. They are employed every day in more than 60% of all-day kindergarten classes [7]. ClassmateAssist is designed to be a netbook-based application that (a) assigns individual students problems that have been selected to be of the right difficulty for them, (b) follows the students' progress during the process of problem solving, and (c) assists them as they work to solve these problems. The problems were developed in conjunction with kindergarten through second-grade teachers. With ClassmateAssist, teachers can continue to use physical manipulatives that are already in their classrooms and use them on desks that are already there. ClassmateAssist further supports teachers' current practices by autonomously giving tailored support to individual students so that all students in the class can work at their own pace while the teacher provides individual attention to students who need it most at that moment.

"ClassmateAssist is designed to be a netbook-based application that assigns individual students problems that have been selected to be of the right difficulty for them."

Our system has three basic components. First, a component follows the manipulatives as students use them. Second, an interaction planner uses the student's progress and history to select the problems, hints, and feedback the system will provide. Third, after selecting the content, the system uses various strategies to display the problems, hints, and feedback for the student.

Observing the Manipulatives

To follow the manipulatives, ClassmateAssist uses a computer-vision component that watches what students are doing. The application uses a standard camera (such as the one on laptop lids) to watch the manipulatives on the surface between the laptop and the student and to track these objects as the student moves them. However, the system does not need to watch everything. We worked with teachers to define pedagogically relevant, "observable" phenomena for the system to watch. For example, it watches the manipulatives to see where students are building clusters of objects selected from those on the desk and where students' movement of objects will not serve to create the appropriate clusters. It can see whether the students seem to be moving the manipulatives in such a way that they will reach a solution to the task or not. Furthermore, it can recognize common errors or non-goal directed movements.

Planning the Interaction

Taking various pieces of data, such as movement-toward-goal and time-since-assignment as input, ClassmateAssist gives the students new assignments, hints on the current task, or feedback on a recently completed task.

"Taking various pieces of data, such as movement toward goal and time since assignment as input, ClassmateAssist gives the students new assignments hints on the current task, or feedback on a recently completed task."

To determine assignments, we worked with teachers to define a range of tasks, which we verified as appropriate for kindergarten through second-grade classrooms, by deploying a version of the application that used a touch screen and virtual manipulatives. The order of assignments was heuristically determined, but a more principled approach will be used in our final version.

We want to deliver the same types of hints and feedback that teachers would deliver if they were watching the child working. To this end, we are currently analyzing videotapes of a teacher providing one-to-one instruction to students while using manipulatives. We are observing the teacher for the information she provides as well as the contextual variables that elicit communication from the teacher. That is, we are looking at what the teacher said and did as well as the aspects of the students' behavior to which the teacher was responding.

Providing Input to Participants

To reflect what a teacher might say or do, ClassmateAssist provides both auditory and visual information. It uses a speaker on the computer and a text-to-speech (TTS) engine for the auditory assignments and feedback. In practice, we have found that the teachers with whom we are working do not mind the sound in their classrooms. The application also uses the display on the laptop to show information to the student. On-screen video can be "marked up" by highlighting screen regions or superimposing information. Finally, we are pursuing a method to use an integrated pico-projector to actually project the hints or feedback directly onto the physical manipulatives.

Drilling Down on the Tabletop

How does ESP help us reach these goals just mentioned? Various research threads within our labs allow us to select among many capabilities. Among them, we use object recognition, activity detection, facial affect coding, interaction planning, camera/projection integration, and more.

> *"Because objects on a surface make up a considerable portion of the context, tabletop context awareness requires robust identification of the objects that are on the table."*

On and Around the Table

The context of the tabletop includes many contextual variables with which we must concern ourselves. There are the objects on the table, the people around the table, the actions of those people, and their internal states. Each of these things, in turn, can impact how a computer-based tutor should interact with users.

Objects

Because objects on a surface make up a considerable portion of the context, tabletop context awareness requires robust identification of the objects that are on the table. Computer-vision object recognition used to work in restricted contexts, and projects were more "proofs of concept" than useful tools. As the science is advancing and new tools are developing, the field is moving beyond proofs of concept. The systems we have developed use various strategies in concert to deliver high-probability object identities.

One strategy developed inside the ESP research team involves a figure-ground segmentation algorithm [8] that distinguishes the tabletop and other surrounding contexts from the hands and the objects that the hands are manipulating. Making this distinction between figure and ground allows us to prioritize further analyses of those objects in the context that are immediately relevant.

However, we are also concerned with the objects that are not in use but are stable on the tabletop. The work we are building on looks for similarities between the objects in view of a camera and the objects known from a database [9]. ESP applications apply color matching to known exemplars, texture matches by using Scale Invariant Feature Transform (or SIFT) [10], and 2D outline matching. As an example of the benefits provided by having these algorithms working together, consider that our work has shown that with an initial figure-ground segmentation, the SIFT algorithm gives us a 91% recognition rate, while SIFT alone gives us only a 12% recognition rate [8].

Within ESP, we have also developed parallelization methods to reduce latency. Latency here is the time between video capture and the completion of analysis, and it can be an issue with interactive machine-vision-based applications such as ours. We do not want users to perceive a delay. Such latency can make an application unusable. To ensure that users will have a good experience, we have put considerable effort into making our analyses possible in “interactive time scales” [11].

In addition to knowing which objects are on the table, we also perform analyses to determine which objects are clustered together [12].

Activities

Another important part of context is the activities of the people at the table. To determine activities, a system can use the manipulated objects to infer what people are doing [13]. Objects on a table can be used to make inferences about what kinds of tasks are possible for the user or which behaviors a user may be likely to engage in. Detection and identification are made more tractable because the objects in the context constrain the set of possible activities. If a person is seen with a pen, a system can surmise that “writing” has a high probability of occurring. To this end, researchers in ESP have compiled sets of object names associated with particular activities [14].

“Facial recognition software has improved to the point where, especially when vision systems are combined with voice recognition, it is possible to determine with near-perfect accuracy who someone is.”

People

Knowing who is in the context is valuable information for interaction planning. Knowing who the user is, which child, for example, allows an application to decide among different content. In addition, there may be more than one person and knowing whom the user is with—a parent, a teacher, a collaborating peer—can help the system decide what information to display. We use both faces and voices in determining who is in the context [15]. Facial recognition software has improved to the point where, especially when vision systems are combined with voice recognition, it is possible to determine with near-perfect accuracy who someone is.

Internal States

Faces can do more than just provide identity information. Faces can reveal something about how people feel. Internal states (ranging from confusion to interest and even distraction) are quite useful in determining how to support an individual trying to accomplish a task. A person's internal state—what activity they may find difficult or whether they are distracted—can be used to select a higher level of assistance or a simpler task.

"A person's internal state—what activity they may find difficult or whether they are distracted—can be used to select a higher level of assistance or a simpler task."

While a person's internal state is clearly something in the context, it may be less clear that we can use technology to "see" it. However, there is a set of emotions, sometimes called the basic emotions or Darwinian emotions, that not only are distinct but also are associated with specific muscle configurations. Paul Ekman's research over the past 40 years [16] has shown that contractions in specific sets of facial muscles, what Ekman calls "facial action units (FAUs)," are uniquely associated with particular internal states. For example, smiling involves a set of muscles around the mouth and the eyes. If a system thinks those muscles are in that particular configuration, it can infer two things: that the person is smiling and that they are happy.

The universality of facial configuration and internal state allows a vision system to make intelligent guesses as to a person's feelings. Machine-vision researchers have been building applications that recognize FAUs and categorize emotions [17]. We have built one of these applications into a version of our tutor.

Under the Hood

The perceptual system, as described, can be used to watch the objects, users, and actions. This is done to be context sensitive, make the interaction unobtrusive, and to help us actively move a user toward a goal, that is, for performance support. Thus far, the discussion has centered on sensing and context awareness much more than performance support. Performance support happens "under the hood." The performance support for our tutor is provided by an agent/model that receives the context data, makes inferences about both context and user status, and generates the actions of the system.

"The universality of facial configuration and internal state allows a vision system to make intelligent guesses as to a person's feelings."

As noted, if a system is to go beyond "easy to use" and actively help the user, it must be context sensitive, but it also has to do more. For any performance support application to work, it must generate content (action) that is relevant to the user and the user's current context. However, the application cannot know for sure that the inferences it is making are correct. For example, the system cannot really know what the user knows. It can know with some probability that a user is confused, but there is a complementary probability that the system is wrong. Furthermore, it may not be able to observe all of the variables that are relevant in the context. For example, relevant things can happen outside the sensors' range. Therefore, the system's actions must be undertaken with a level of uncertainty. This raises a "planning under uncertainty" problem.

The ESP group has used Partially Observable Markov Decision Processes (POMDPs) in its work on planning with faulty or incomplete information [18]. POMDPs are a special case of a Markov decision process. With normal Markov decision processes (MDPs), actions can be mapped with probabilities to future states; this means that system actions—the system's output to the user—have a certain probability of affecting the course of the user's future internal states, say. However, MDPs assume that every state variable can be observed without error [19], and we know we are dealing with probabilities, not absolutes. Fundamentally, we are saying that we can make a high probability inference as to where, for example, a student is in a problem-solving activity, and we are able to generate meaningful content to help that student, based on where she is. It may have a high probability of accuracy, but it is still an inference.

"The ESP group has used Partially Observable Markov Decision Processes (POMDPs) in its work on planning with faulty or incomplete information."

POMDPs are an alternative that allow for this uncertainty. A POMDP models the relationship between an agent and its environment with five variables: states, actions, observations, transition probabilities, and reward or cost functions. The system observables include the identities and relative positions of objects and the physical movements, including facial expressions, of users. Its actions are the teaching behaviors, including the hints and feedback that the system produces. ClassmateAssist's hidden states include the cognitive state of the user and the progress toward the goal. Its cost functions assume that the system should refrain from interrupting a user and assign a cost to any system-initiated vocalization or visualization.

ClassmateAssist uses sensor data to observe the environment and infer the user's internal states with respect to the task so that it might select actions to be used in its interactions. The system has a vocabulary of actions, and it makes assumptions about the probabilities with which those actions will affect the next state of the user.

The problem-solving activities and where a user is in a task jointly determine how the system chooses to interact. ClassmateAssist has to determine (with some degree of certainty) that a student is likely to solve a problem on her own and withhold hints. It also has to determine (with some degree of certainty) where she is in the task in order to know which hints or feedback might be appropriate, if necessary.

System Actions

We use the laptop to augment a student's interaction with objects on the tabletop. The system's actions are both auditory and visual. Sometimes it tells a user; sometimes it shows a user; and sometimes it does both. For telling a user, the system employs the by-now common method of a TTS engine. It has stored text that describes the tasks and that details the hints and feedback that provide the instruction. If the context is appropriate, for example, when the user has completed a task and needs a new assignment, the application chooses an assignment and sends the relevant text to the TTS.

Because hints often require a bit more than words alone can easily provide, the output of our tutor also involves visualizations of a type sometimes called mixed or augmented reality. That is, we combine graphic information generated by our system with the environment. Augmented reality uses context-awareness to supply the user with computer-generated imagery that visualizes further information about the context to the user.

"Augmented reality uses context-awareness to supply the user with computer-generated imagery that visualizes further information about the context to the user."

We would use this, for example, because teachers in our observations frequently used gestures over the manipulatives to help students focus on particular coins, sets of coins, or features of coins. Rather than try to use text to describe the location or some feature of a coin, we have opted to use visualization techniques. The system either displays information on the screen or projects it directly onto the objects themselves.

Screen-based Augmented Reality

One version of the system uses the screen of the computer to present the visualization. It shows the live video feed from the camera, which is facing the objects, and it then superimposes information about the objects onto the screen for the user. This information can be used to single out a particular object or set of objects. The on-screen augmentation can also highlight a region of the tabletop that a user is being asked to place objects within. In doing these things, the system can help in the performance of the task or it may help the user to learn something new.

Figure 1 shows a coin-sorting task where the user has been asked to move the dollar coin into the "box" displayed on the screen. We have also included a teacher avatar that watches what the user is doing and walks over to inspect and approve when the user does the correct thing.

Figure 1: Screen-based Augmented Reality Interface (Simulated)
Source: Intel Corporation, 2010

This screen-based augmented reality is relatively transparent, but does involve looking at the screen rather than the objects in front of the user.

Projector-based Augmented Reality

Another way to mark up the world is to project feedback and hints onto the objects themselves. Projection plus camera systems (ProCams) have been around for some time. Their utility is described by IEEE International Workshop on Projector-Camera Systems: "Systems that utilize controllable lighting systems with light-sensing devices facilitate a wide range of applications" [20]. In most cases, "controllable lighting systems" are projectors and "light sensing devices" are cameras. In the typical system, a camera is used to watch objects in an area, and a projector is used to respond to movements of those objects. The Situated Multimedia Art Learning Lab (SMALLab) [21] is an example. SMALLab uses a room-sized area and mounts projectors and cameras many feet above the floor. The camera watches the users and the objects that the users hold, while the projector changes the images projected toward the floor, based on the movements that the camera sees. Teachers design the interactions so that the projected images and actions of the users work together to teach curricular content.

"In the typical system, a camera is used to watch objects in an area, and a projector is used to respond to movements of those objects."

With the advent of low-power, short-throw, inexpensive pico-projectors, similar applications have become feasible in a portable unit. Sixth Sense [22] demonstrates the use of a pico-projector and camera in a wearable system that can watch the user's hands and nearby objects and then project relevant content onto local surfaces.

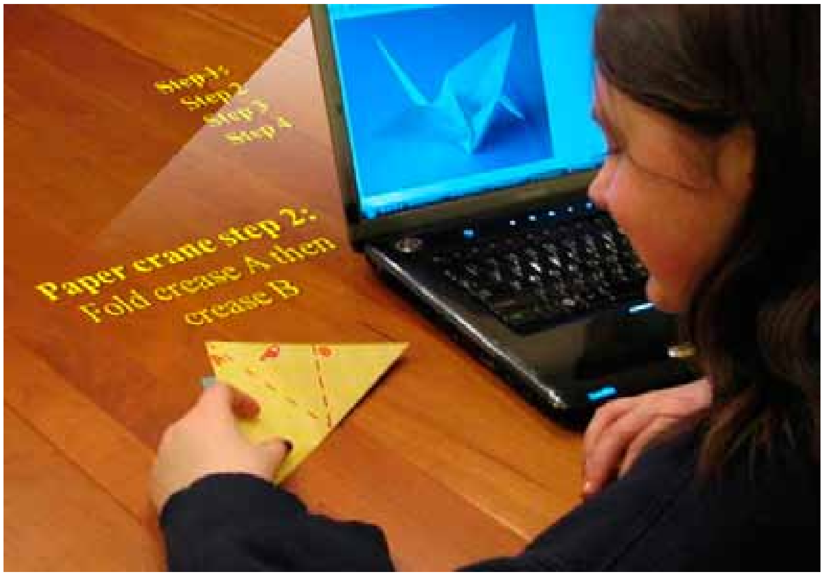


Figure 2: Projector-based Augmented Reality Interface (Simulated)
Source: Intel Corporation, 2010

Figure 2 is a simulated photo of a user who wants to do some origami. The vision system watches the paper and the user's hands and then projects instructions onto the folded paper and the table around it.

Within our group, we have developed Bonfire, a system that augments everyday laptops with projectors, cameras, and perception software [23]. This system projects onto surfaces around a laptop and watches those surfaces for gestures or objects that are then taken as commands to the system. With Bonfire, the tutor can watch as a student moves manipulatives and then it can project content onto them. Instead of focusing on the screen as they manipulate objects, students need only look at the objects themselves.

The Future of Context

The future of context is easy to see: more, faster, better. The data rate of sensor streams will no longer present problems. Computational power will make computer vision much more commonplace. In addition, sensor fusion with its multi-modal data will also be common. Technology will be able to do more analyses at interaction speed, and the quality of the models that technology can create will improve substantially.

As our technology builds more detailed models from our experience, the depth of personalization that can be delivered will increase. A system can know what tasks users have accomplished in the past and what they might want to accomplish next. It can know what level of difficulty is suitable for them as well as what kind of support they would be best with. That is, user models will not only aggregate our past experiences but they will be able to use those data to augment our current experience. Personalized technology will be able to augment our everyday experience in a way that is uniquely appropriate just for us, because the augmentation is based on the history we share with our devices—a history made available to the technology by context awareness.

"Personalized technology will be able to augment our everyday experience in a way that is uniquely appropriate just for us, because the augmentation is based on the history we share with our devices."

Summary

This article has described the current state-of-the-art in context awareness. The capabilities of today's processors allow systems to use high data rate sensors, such as video, to provide higher levels of perception. Accurate sensing of where you are, who you are with, and what you are doing opens up the capacity to deliver new kinds of services. The Everyday Sensing and Perception project will make context awareness 90% accurate over 90% of the day. Our work has focused on the identification of objects in the environment, activities of people in the environment, personal identification of those people, as well as their affective states. By providing this level of context awareness, a system can support complex tasks. Our work on tabletop tutoring shows how this emerging capability can be used to teach users new concepts or skills.

References

[1] LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G., and Schilit, B. "Place Lab: Device Positioning Using Radio Beacons in the Wild." *IEEE Pervasive*, 2005.

[2] Chang, K., Hightower, J., and Kveton, B. (2009). "Inferring Identity using Accelerometers in Television Remote Controls." *International Conference on Pervasive Computing*, 2009.

[3] Bezanson, W. *Performance Support Solutions: Achieving Goals Through Enabling User Performance.* Trafford Publishing, Victoria, B.C, Canada, 2002.

[4] White, S., Marino, D., and Feiner, S. "Designing a Mobile User Interface for Automated Species Identification." *ACM CHI*, 2007.

[5] Silva, E. "On the clock: Rethinking the way schools use time." *Education Sector*, 2007.

[6] "American Time Use Study." U.S. Bureau of Labor Statistics (2006). Available at ***http//www.bls.gov/tus/home.htm***

[7] "Student Work and Teacher Practices in Mathematics." *National Center for Educational Statistics: US Department of Education,* 1999. Available at ***http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=1999453***

[8] Ren, X. and Gu, C. "Figure-Ground Segmentation Improves Handled Object Recognition in Egocentric Video." To appear in *CVPR* 2010.

[9] Rahimi, A. and Lee, S. "Visual Object Instance Recognition." Available at ***http://people.csail.mit.edu/rahimi/projects/objrec/***

[10] Lowe, D. "Object recognition from local scale-invariant features." *International Conference on Computer Vision,* 1999.

[11] Pillai, P., Mummert, L., Schlosser, S., Sukthankar, R., and Helfrich, C. "SLIPstream: Scalable Low-latency Interactive Perception on Streaming Data." *NOSSDAV,* 2009.

[12] Shi, J. and Malik, J. "Normalized cuts and image segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22(8), 2000.

[13] Philipose, M., Fishkin, K.P., Perkowitz, M., Patterson, D.J., Fox, D., Kautz, H., and Hähnel, D. "Inferring activities from interactions with objects." *IEEE Pervasive Computing,* pages 50–57, October 2004.

[14] Wu, J., Osuntogun, A., Choudhury, T., Philipose, M., and Rehg, J. "A Scalable Approach to Activity Recognition Based on Object Use." *ICCV,* 2007.

[15] Choudhury, T., Clarkson, B., Jebara, T., and Pentland, A. "Multimodal Person Recognition using Unconstrained Audio and Video." *Audio- and Video-based Biometric Person Authentication,* March 1999.

[16] Ekman, P. and Friesen, W. V. *Unmasking the face. A guide to recognizing emotions from facial clues.* Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

[17] Bartlett, M., Littlewort, G., Lainscsek, C., Fasel, I., Frank, M., and Movellan, J. "Fully automatic facial action recognition in spontaneous behavior." *International Conference on Automatic Face and Gesture Recognition,* 2006.

[18] Theocharous, G., Beckwith, R., Butko, N., and Philipose, M. "Tractable POMDP Planning Algorithms for Optimal Teaching in 'SPAIS'." *International Joint Conference on Artificial Intelligence* (IJCAI) workshop on Plan Activity, and Intent Recognition (PAIR), July 2009.

[19] Blythe, J. "An Overview of Planning Under Uncertainty." *Lecture Notes in Computer Science,* 1600:85-110, 1999.

[20] *ProCams 2010.* Retrieved from ***http://www.cs.cmu.edu/~ILIM/ProCams2010/***

[21] Birchfield, D., Megowan-Romanowicz, C., and Johnson-Glenberg, M. "Next Gen Interfaces: Embodied Learning Using Motion, Sound, and Visuals–SMALLab." *American Educational Research Association Annual Conference* (AERA); *Applied Research in Virtual Environments for Learning* (ARVEL), 2009.

[22] Mistry P. and Maes, P. "SixthSense – A Wearable Gestural Interface." *SIGGRAPH Asia, Sketches,* 2009.

[23] Kane, S., Avrahami, D., Wobbrock, J., Harrison, B., Rea, A., Philipose, M., and LaMarca, A. "Bonfire: a nomadic system for hybrid laptop-tabletop interaction." *ACM Symposium on User Interface Software and Technology* (UIST), 2009.

Acknowledgments

We thank the entire ESP team for all of their tremendous work and supplying some of the coolest toys. We thank Nick Butko, Vimal Mathew, Nicol Schmied, and Dan Teibel for their help with the tutoring system, Tye Rattenbury for running the analyses on the American Time Use Survey, Kieran Del Pasqua for the screen-based augmented reality images, and Olivia Laing and Anthony Lamarca for helpful comments on earlier drafts.

Author Biographies

Richard Beckwith is a Research Psychologist with FTR's People and Practices Research group. He focuses primarily on education, agriculture, and privacy. He received his PhD from Teachers College, Columbia University in 1986. From 1986 to 1991 he was a research scientist at Princeton University's Cognitive Science Lab, working on WordNet. From 1991 until coming to Intel, he was a Research Faculty member at Northwestern University in the Institute for the Learning Sciences (ILS). His e-mail is richard.beckwith at intel.com.

Georgios Theocharous received his PhD degree in Computer Science in 2002 from Michigan State University. From 2002 to 2004 he was a post-doctoral associate at the Computer Science and Artificial Intelligence Lab at MIT, and in October 2004 he joined Intel as a Research Scientist. His research interests include computational models of learning and planning under uncertainty and their applications to the real world. Specific models include reinforcement learning, completely and partially observable Markov decision processes (POMDPs), semi-Markov decision processes, hierarchical POMDPs, and dynamic Bayesian nets. His e-mail is georgios.theocharous at intel.com.

Daniel Avrahami is a Research Scientist at Intel Labs Seattle. He received PhD and MSc degrees in Human-Computer Interaction from the HCI Institute at the School of Computer Science at Carnegie Mellon University and a BSc degree in Computer Science from the Hebrew University in Jerusalem, Israel. His current research looks at context-aware applications for everyday life. His other research interests include the design and implementation of new intelligent communication tools. His e-mail is daniel.avrahami at intel.com.

Matthai Philipose co-leads the Everyday Sensing and Perception project out of Intel Labs, Seattle. He builds systems that observe and reason about human activity, emphasizing detailed but tractable models of activity, very high-density sensing, and automatically acquired common sense. He is broadly interested in statistical reasoning, logic, and programming languages. He has a BS degree from Cornell University and a PhD degree from the University of Washington, both in Computer Science. His e-mail is matthai.philipose at intel.com.

Copyright

COLLECTIVE PARTICIPATION: TECHNOLOGY-ENABLED SOCIAL ACTION

Contributors

Ken Anderson
Intel Corporation

Paul Aoki
Intel Corporation

Michael Kaminsky
Intel Corporation

Renee Kuriyan
Intel Corporation

Index Words

Participation
Collectives
Social Action
Technology Adoption
Aspirations
Citizen Science

Abstract

Collective participation is the part of Essential Computing that uses technology in our social lives to turn our on-line connections into action. Participation is about enabling people to come together to achieve goals that separately they could not achieve. Technological evolution in on-line connection ultimately affects human organization.

This article provides background on what we mean by participation, and why it is emerging now. We identify current research at Intel and identify directions for future research, both technical and social.

Introduction

Have you recently updated your status on Facebook*? Received a tweet? E-mailed a group of friends about doing something together? Sent a group text from your phone? It is clear that electronic communications are today an essential part of our lives. People are now connected to former classmates, receive updates from their favorite news sources, and, all too often, receive unwanted advertising. Although we are now connected, has technology fulfilled its potential to help us accomplish joint action?

NYU researcher and pundit Clay Shirky [19] has called attention to an emerging trend of people using technology to come together in new ways. Shirky argues that information and communication technologies have begun to enable group conversations and collective actions by eliminating hurdles related to time and cost. He notes, however, that the technology tools and social research behind this growth in collective action and group conversations are still in their infancy. Still, there have been some recent prominent examples of people coming together to make something happen.

For example, the 2008 Obama for President campaign shattered previous fund-raising records by mobilizing supporters via e-mail, Facebook, Twitter*, and other Web 2.0 media to contribute money to the effort; the *Washington Post* reported that the number of small contributions was remarkable [16]. More recently, the Red Cross raised $5 million in 24 hours via text messages for earthquake disaster relief in Haiti. These examples point to the power of people acting together—to have a collective focus that can accomplish some end.

The collective participation agenda of *Essential Computing* is to enable people who desire to move beyond Web surfing and update checking to utilize technology to come together and contribute to society in a range of ways and at varying scales of action. Many of the most challenging and important problems that confront our world today, such as reducing the carbon footprint, require collective action in order to effect change. To move technology innovation to support more collective participation, we need new understanding about people and technology-enabled action and we need to apply this understanding to applications that enable social collaboration, cooperation, and action at varying scales of activity.

"Researchers view participation as an act of doing something for some end; an engagement with purpose."

The following outline presents our definition of collective participation and the parameters that we use to guide our research activities. In this article, we provide two examples of current research on collective participation from the Future Technologies Research (FTR) group at Intel, and we suggest further directions for exploration.

Participation and Collectives

Thus far, we've made repeated use of the terms "collective" and "participation." Before proceeding further, we would like to explain how we are using the terms in this article.

Enabling Participation

In its most basic sense, participation means active engagement. Henry Jenkins [13] describes the concept more fully as "participatory culture" and specifies the following conditions in which this concept can be realized: 1) relatively low barriers to expression and engagement; 2) strong support for creating and sharing with others; 3) members believe that their contributions matter; 4) members feel some degree of social connection with one another; and 5) participation has some result.

In the spirit of Jenkins' concept of "participatory culture," FTR researchers view participation as an act of doing something for some end; an engagement with purpose. Through our research we seek to empower people, to encourage engagement in (public and private) decision-making, to foster ownership of informed opinions, and to influence issues that affect everyday lives at varying scales of action. Participation can take many forms since it needs to accommodate a variety of organizational and lifestyle constraints (for example, modifying personal water consumption to conserve water, commenting on mainstream media or blog sites, providing air quality data to neighborhoods). Wikipedia* and the open source movements are both excellent examples of technology-enabled social participation that are motivated by a shared interest and have resulted in collective action.

"Wikipedia and the open source movements are both excellent examples of technology-enabled social participation that are motivated by a shared interest and have resulted in collective action."

Our goal in FTR is to create technologies that expand how people participate and to give people powerful new ways to make a difference. We want to allow people to use technology to build on their personal interests and everyday activities and participate more fully in the social world.

Creating Collectives

We use the term "collective" to denote a *group* of people taking purposeful action motivated by shared interest. Since the rise of enterprise computing, Human Computer Interaction (HCI) research has expanded its traditional focus on one "human" interacting with one "computer" to include group-oriented use of computing, a move that is consistent with the research agendas of "community informatics" and "computer-supported cooperative work (CSCW)". The HCI community has recently paid considerable attention to large-scale Web 2.0 initiatives such as Wikipedia.

"Community *implies a shared sense of values and an engagement of a long duration."*

Our own focus on collectives represents an important extension of this shift in HCI research from individual towards group computing use. We use "collective" to highlight the participation of many individuals, but not necessarily in a well-defined and persistent context such as an existing work group or community. Collectives are, after all, fundamentally different from communities (real or virtual). "Community" implies a shared sense of values and an engagement of a long duration [3]. Collectives are groups of people with some shared sense of purpose and togetherness. Unlike communities, however, collectives do not necessarily entail an extensive sense of shared values nor do they have to persist over a long period of time. Collectives do not have to be large. Collectives vary in form, sizes, temporal duration, and internally they can vary in the values held by their members. Table 1 illustrate the many forms that collectives may take :

Collective	Goal	Duration	Shared Values	Number of People
Obama Supporters	Elect Obama to the presidency	Months to Year	Some	100s of thousands
Tuangou	Buy some big ticket item	Months	Few	1000s
Portland Trail-blazers	Enable team victory	Months to lifetime	Few	10s of thousands

Table 1: Example of Range of Collectives Today
Source: Intel 2010

Obama supporters constituted a large collective that lasted a moderate length of time. Members of this collective happened to share many of the same values. They came together with the objective of electing Barack Obama for president. The *tuangou* is a collective that comes together in China to buy collectively. The members have few shared values, but their common goal is to buy the same refrigerator (or some other object) at a reduced cost. Tuangou collectives last a short temporal period and are relatively small (1000s of people). Another contrasting example is sporting fans, in this case, the Portland Trailblazers.

The collective might meet at physical events, like games, but also be in contact on-line. Although any given "season" is short, people might be part of this collective for years.

A research focus on collectives represents a radical shift in thinking about the scale of computing technologies, since the collective as a whole becomes the unit of focus, not the individual user. The collective is also the unit of analysis and design. For researchers and technologists alike, collectives encourage us to look at the social face of the challenge of "computing at scale".

"Collectives represent a radical shift in thinking about the scale of computing technologies, since the collective as a whole becomes the unit of focus, not the individual user."

In short, interest-based collectives are often organized by local, niche, and/or amateur activities that differ in some fundamental ways from standard professional organizations. Just as amateur sports leagues are predicated on a broader base of participation than professional sports, easy to use technologies lower the barriers to broaden the base for active participation with collectives for social action. In terms of collective activities like civic engagement, a recent report by the MacArthur Foundation declared that the most promising direction for involving youth was not from top-down adult mandates but through bottom-up peer based activity that develops a sense of participation by members for collective action [12]. We also believe research about enabling participatory collectives is a fruitful direction for research, though not limited to civic engagement.

Key Challenges

Research into participation builds on the copious research on mobile technologies, sensors, ubiquitous computing, ambient displays, and intelligent software. Many questions remain unanswered. Research on "participation" can focus usefully on the following:

- How can technology support collective action on topics such as environmental sustainability, assisting locally and globally when disasters strike, and other pressing social problems?
- What does scalable interaction look like as more people and devices are involved in the interactions?
- How do collectives conduct themselves?
- How do localized interactions amongst subsections of the collective happen? How do global actions become coordinated across the collective at large?
- How do people become members of collectives and how do they end their membership?
- How do conflicts occur and play out within social collectives?
- What kinds of political and social innovations would independently intercept technologically driven ones—that is, is there something different about some social movements now that can uniquely take advantage of new technology?
- What can be done to make collective action more than mere on-line tokenism?

"What can be done to make collective action more than mere on-line tokenism? "

- How do sensors, ambient displays, and mobile devices enable reliable information and coordinated action for participating in collectives?
- How can ad-hoc and temporally limited secure communications and connections be made and maintained among members of collectives?

In addressing these and other issues, research should aim to help deconstruct the nature of collectives, the nature of technologically-mediated interactions in and around participation, and the relevance of design and technology development in computing and social sciences.

Current Research

Intel's FTR group continues to work on several projects related to collective participation. In this article, we review two of these in some depth: Consumerization and Common Sense.

Consumerization

The Consumerization project is a social science research project looking at a particularly influential model of participation: i.e., consumers participating in an economy through the consumption of goods and services. This model of participation has emerged out of a long and contested historical process of consumerization that has produced, most famously, the United States as dominant national economy driven by the mass consumption of its "citizen consumers" [5]. However, the consumer-in-a-market model of participation is by no means limited to the commercial sectors of industrialized countries. This way of thinking, behaving, and feeling has been extended into domains as diverse as politics, education, healthcare, the arts, and public utilities; all of these, not without controversy, have come to be organized, conceptualized, and enacted in terms of a "product" being marketed, financed, and consumed.

Drawing upon this larger historical background, the focus of the Consumerization project is to uncover the implications of governments, businesses, and cultures around the globe adopting this consumer model of participation as the basis for large-scale economic development practices, such as efforts to foster the "emerging middle class" in developing countries and to bring "unserved and underserved" populations into the mainstream marketplace in wealthier countries. From these empirical analyses the project seeks to inform public policies around information and communication technologies (ICTs), such as PCs and mobile phones, and around economic development. These policies aim to promote both digital inclusion and economic development around the world through the use of technologies. Findings from this research are also designed to inform Intel's strategies vis-à-vis consumers, specifically to accelerate growth in emerging markets and reach first-time buyers of computers.

The question of whether technology adoption and technologies for economic development for the global poor and emerging middle classes actually lead to improved lifestyles for people in emerging societies is highly debatable. Discussions about who should be using technologies and what income groups should benefit from technology are ongoing. Consumption is a contested topic in these debates. New policies for economic development focus on individuals as "consumers" of technologies and services rather than as aid recipients. This shift is based on the logic that economic development results from a process of entrepreneurship or private-sector-led business initiatives rather than through traditional state- or nonprofit-led interventions targeting marginalized groups. Within this context, the research asks, "What are the institutional, political, and cultural processes that create consumers and what are the implications for societal participation?"

"Whether technology adoption and technologies for economic development for the global poor and emerging middle classes actually lead to improved lifestyles is highly debatable."

For the purposes of this article, we focus on a case study conducted in Kenya that reveals a set of contradictions, particularly around citizenship, as consumers are created. M-Pesa is a mobile payments service implemented by a privately-run telecom business in Kenya. Consumers of this service use their mobile phones to send and receive money and also to store or "bank" their money. Although implemented with the goal of serving the poor, or those who lacked formal bank accounts, the service became part of the telecom business' strategy to reach the broader mass market, including the wealthiest segments of the country, and the strategy was highly successful.

Method or Approach

Intel's Consumerization research has focused on the interactions among three sets of actors: governments/policy makers, businesses, and individuals. These interactions were studied primarily though first-hand, qualitative fieldwork in Kenya, India, Bangladesh, China, Russia, Mexico, and the United States. The researchers also looked across sites of interventions through analysis of development policies, business strategies, and individuals' self-perceptions through a consumptive lens and within the context of the discourses of inclusion and societal participation. Finally, this research was situated within the larger literature on development-through-entrepreneurship as well as the cultural politics of consumption.

Discourses of Inclusion and Information, and Communication Technologies

Initially, the policies around ICTs and international development revolved around the notion of the '"digital divide" and whether the Internet would serve to widen or narrow the gap between information-rich and information-poor countries [17]. The policy discussions eventually moved beyond this problematic divide between information rich vs. poor societies to whether ICTs should be used to tackle development problems [21]. However, the rhetoric and principles of inclusion and participation in the digital world continue today with programs targeting "digital inclusion," efforts by governments to reach marginalized groups with ICTs and education, and private sector efforts to reach new markets in the name of connecting and bringing broadband to remote areas.

"Intel's Consumerization research has focused on the interactions among three sets of actors: governments/policy makers, businesses, and individuals."

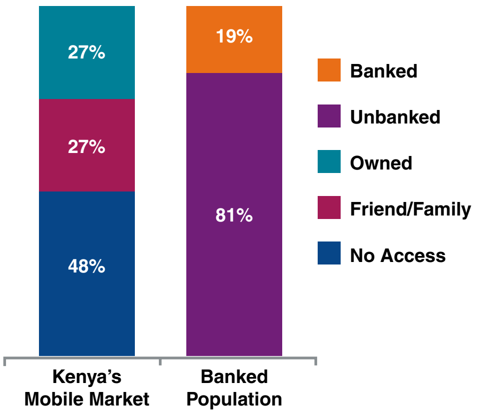


Figure 1: Gap Between Mobile Market Penetration and Banked Population
Source: Finaccess National Survey 2006 [7]

The Kenya M-Pesa project was initiated in March 2007 by Kenya's largest mobile network operator, Safaricom (which is part of the United Kingdom-based Vodafone Group). The payment service via mobile phones was introduced with the goal of "deepening financial access for the poor" [10]. The goal was to enable consumers to make payments via their mobile phones and use the network of Safaricom airtime resellers as agents to facilitate the process [10]. The service was successful in reaching a broad range of consumers in Kenya partly because the cost of the service was cheaper than traditional services (see Table 2). It also served to bridge the gap between mobile phone penetration and banking penetration (see Figure 1).

METHOD	COST (Ksh)
M-Pesa	35
PostPay (money transfer operated by the post office)	75
Akamba (bus company)	175

Table 2: Cost Comparison of M-Pesa and Other Traditional Money Transfer Services
Source: Morawczynski 2009 [15]

M-Pesa is an example of a service that was introduced in the name of inclusion to target the lower income groups in the country who lacked formal bank accounts. As the service gained penetration in the market, the image of the target consumer shifted away from the poor and from visions of them participating in a banking intervention, to an even broader understanding of inclusion that included national images of the consumer as "everyone." The service became part of Safaricom's successful strategy to reach the broader mass market, including the wealthiest segments of the country. They have now expanded the M-Pesa services beyond basic banking services to include services that provide convenience, something for which our research showed all income groups in Kenya are willing to pay.

With this shift, there is a process of abstraction that blends the challenges and politics facing the different "consumer" populations into one that is "Kenyan." Inclusion in this sense came to represent being part of a unified Kenya. The violence following the elections of 2008 resulted in hundreds of dead. The contested election result fueled tribal fears, anger, and divides. Our research found that a unified notion of what it means to be "Kenyan" and a "consumer" appealed to the mass market, to the private sector, and even to bureaucrats working within the government. Low-income Kenyans who used M-Pesa expressed a sense of belonging to a larger market. Further, people felt that they were participating in a national "action" rather than feeling like they were the "poor." With the "consumer as everyone" concept, there is an attempt (conscious or not) by the government and private sector to smooth over the divisions within Kenyan society. Such a strategy implicitly promotes equity between the poor and rich as well as between rival tribes by treating them all as equals from a consumer perspective [14].

The definition of "Kenyan" is highly contested, but the abstracted notion of "consumer as everyone" somehow transcends politics and brings people together. Further, the "consumer as everyone" promotes a sense of societal participation, and meets the goals of inclusion (at least on the surface). By no means does it guarantee equity in terms of actually reaching everyone it sets out to, but it enables the private sector and governments to at least appear to be reaching out to all. Further it enables the poor to participate in the digital world, a high stake gambit in a politically and ethnically divided environment. In this case, citizenship, identity, and nationhood are slippery concepts fraught with tensions and contradictions. However, these contradictions reveal a set of collective desires and anxieties of a nation upon which governments and businesses can strategically build. The M-Pesa banking experiment enabled participation of the poor not just in the digital world, but also as citizens. The research shows that these contradictions can be seen as opportunities for policy and market intervention.

> *"The definition of* Kenyan *is highly contested, but the abstracted notion of* consumer as everyone *somehow transcends politics and brings people together."*

Common Sense: Mobile Environmental Sensing Platforms to Support Community Action and Citizen Science

The Common Sense project's approach differs from that of the Consumerization project by (1) focusing on units of society rather than on whole societies, (2) being located in a modern-day city with concomitant infrastructure, i.e., San Francisco, and not in an emerging market, and (3) developing technology for implementation. Common Sense is developing mobile sensing platforms to support personal environmental awareness and grassroots community action. To this end, a family of hardware and software components was built that can be used with a range of applications, and new communication paradigms were developed that enable communities of non-experts to gather and produce information that is "credible enough" for experts and policy makers.

> *"Environmental activists are continually required to produce information artifacts that are* credible enough *to engage with policy makers and the relevant bureaucracy."*

Overview

The research area of mobile participatory sensing involves the use of consumer electronics (such as mobile Internet devices and mobile phones) to capture, process, and disseminate sensor data, complementing alternative architectures (such as wireless sensor networks) by "filling in the gaps" where sensor infrastructure has not yet been installed [1,9,18]. While some sensors are already common in consumer devices (e.g., geolocation, motion, audio), there are other types of sensors that are not yet common that offer the ability to collect additional data of individual and social interest. In our research, sensors that measure gases relevant to environmental air quality were used.

To make environmental sensing useful for practical action, one must do more than just "collect" and "present" data [9]. Environmental activists are continually required to produce information artifacts that are "credible enough" to engage with policy makers and the relevant bureaucracy; appealing enough to be useful in community mobilization; and personally relevant enough to keep people interested and motivated [4]. The research seeks to enable community members to engage in collaborative "citizen science" [11] or "street science" that will be useful in interactions with government agencies and non-government organizations (NGOs).

> *"The research seeks to enable community members to engage in collaborative* citizen science *that will be useful in interactions with government agencies."*

To this end, the Common Sense project is developing the following:

- Hardware- and software-sensing platforms that allow individuals to collect environmental information. Some consist of custom data acquisition boards that contain environmental air quality sensors that are paired with commercial mobile devices; others are standalone mobile devices.
- Mobile and Internet-based software applications that allow people to analyze, share, and discuss environmental information in order to influence environmental regulations and policies.

A Family of Research Prototype Platforms

For prototyping, a set of board designs and embedded software was developed with the intention of releasing these components to the research community as they stabilize. These components can be selectively populated with off-the-shelf carbon monoxide (CO), nitrogen dioxide (NO_2), and ozone (O_3) gas sensors as well as with temperature, relative humidity, and motion (3D accelerometer or orientation) sensors. Designs so far include single sensor boards for integration with commodity mobile devices and split sensor boards for vehicular use that enable isolation of the processing electronics from exposure to the environment (Figure 2, right-side).

These boards are meant to be paired with a user's mobile device via Bluetooth. The user's mobile device uploads the sensor data (along with geolocation data) to a database server via the phone network. A standalone, handheld, environmental monitor was developed that includes its own GSM/GPRS phone module and a GPS-based geolocation module, obviating the need for a separate mobile device. All designs use microcontrollers to manage the low-power sensing; the handheld monitor is based on the Epic Core* module developed at Berkeley [6].

Figure 2: Handheld Monitor Design and Enclosure
Source: Photo by Mazzarello Media and Arts 2009

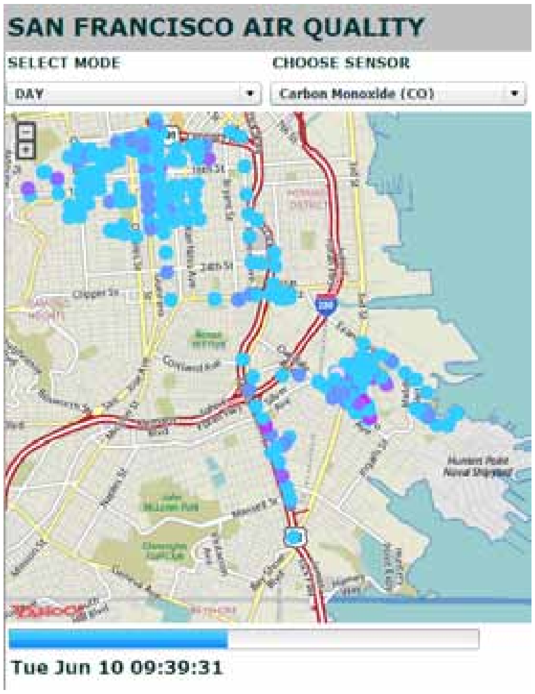
Community Sensing

To facilitate the primary research agenda of participation for "citizen sensing," a complete (standalone), handheld environmental monitor was developed. The standalone monitor is more practical in that it reduces the number of separate devices the user must charge and carry; it also eliminates the need to use mobile "user apps" in a software environment that is not really intended to support applications that run continuously for months at a time.

A trial deployment of the handheld monitors has begun in collaboration with an environmental NGO working near the Port of Oakland. A website, "Common Sense Community," has been designed to provide on-line community features to support discussion of interesting phenomena and strategies for practical action (Figure 3) as well as individual exposure information (Figure 4). The research continues by working iteratively with community members to ensure that the visualizations and features answer the type of questions frequently posed by users who are not environmental experts [20].

Just as with the San Francisco trial, a rich set of research questions emerged concerning the use of environmental data. For example, questions came up regarding different stakeholder interests and the balance between what is possible (technically) with these kinds of relatively inexpensive sensors versus what is desired (scientifically and politically). To enable informed participation by community members in environmental decision-making, it is important to consider how visualizations can appropriately represent the credibility of the data as well as facilitate citizens' understanding of the environmental conditions around them. These types of questions are what will be explored in the near future with this project [2].

Figure 3: The Common Sense Community Site Showing Data Collected by a Single User. The *My Exposure* Widget (a) and *Tracks* Visualization (b) Are Visible with the Commenting Panel (c).
Source: Intel Corporation, 2010


SAN FRANCISCO AIR QUALITY

SELECT MODE: DAY

CHOOSE SENSOR: Carbon Monoxide (CO)

Tue Jun 10 09:39:31

Figure 4: A Visualization Tool
Source: Intel Corporation, 2010

The two research projects (Consumerization and Common Sense) demonstrate the breadth of the problems associated with creating participatory collectives. The participation of Kenyans as citizens and consumers highlighted the interconnection of large social and technological infrastructures in "participation." The Common Sense project in San Francisco points to the importance of small local collectives who are enabled through technology to attempt changes in their communities. Both of these studies highlight (1) the importance of focusing on the scale for design and research, i.e., moving from the individual to the collective; and (2) the implications of our work in terms of how socio-technology interventions can support a more engaged stance toward public participation generally.

Conclusions

In 1990, Jonathan Grudin published "The Computer Reaches Out," in which he drew attention to a particular pattern in the disciplinary evolution of human-computer interaction [8]. As computer systems became more complex, as they became more widespread, and as interaction paradigms become better understood, Grudin argued, the focus of technology research attention would expand beyond the device, through interaction styles, organizational contexts, and broader social settings. Research, since the time of Grudin's publication, has shifted from what might be technologically feasible to what might be interactionally sensible, and today it focuses more on the concerns with technology systems embedded in the social practice of wider networks. Intel's research on participation is at the forefront of expanding interest in the intersection of technology, social organization, and participation. Collective formations and their desire for action pose exciting new challenges for the design and implementation of technology. Our research program on participation seeks to address the issues surrounding the integration of collective participation with the highly varied assemblies of technology that exist already in our homes and public spaces today, or that will be deployed in these areas in the near future. This research highlights the importance of understanding what is essential to people today in order to design for essential computing for tomorrow.

References

[1] Abdelzaher, T., Anokwa, Y., Boda, P., Burke, J., Estrin, D., Guibas, L., Kansal, A., Madden, S., and Reich, J. "Mobiscopes for Human Spaces." *IEEE Pervasive Computing*, Volume 6, Issue 2, pages 20-29, 2007.

[2] Aoki, P.M., Honicky, R.J., Mainwaring, A., Myers, C., Paulos, E., Subramanian, S. and Woodruff, A. "A Vehicle for Research: Using Street Sweepers to Explore the Landscape of Environmental Community Action." In *Proceedings CHI* 2009, *ACM* pages 375-384, 2009.

[3] Anderson, B. 2006. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso. ISBN-10: 1844670864.

[4] Corburn, J. 2005. *Street Science: Community Knowledge and Environmental Health Justice*. MIT Press, Cambridge, MA. ISBN-10: 0262532727.

[5] Cohen, Lizabeth. *A Consumers' Republic: The Politics of Mass Consumption in Postwar America*. Knopf. ISBN-10: 0375407502, 2003.

[6] Dutta, P., Taneja, J., Jeong, J., Jiang, X., and Culler, D. "A Building Block Approach to Sensornet Systems." In *Proceedings SenSys*, 2008, *ACM*, pages 267-280, 2008.

[7] *Finaccess*, 2006. "Understanding Kenya's financial landscape." Available at ***http://www.fsdkenya.org/finaccess/documents/FinaccessReportFINALMain.pdf***

[8] Grudin, J. 1990. "The Computer Reaches Out: The Historical Continuity of Interface Design." In *Proceedings ACM Conference Human Factors in Computing Systems* CHI'90 (Seattle, WA), 261-268.

[9] Honicky, R.J., Brewer, E., Paulos, E. and White, R. "N-SMARTS: Networked Suite of Mobile Atmospheric Real-Time Sensors." *SIGCOMM* 2008 NSDR Wksp., ACM, 2008.

[10] Hughes, N. and S. Lonie. "M-PESA: Mobile Money for the "Unbanked": Turning Cell phones into 24-Hour Tellers in Kenya." *Innovations* Winter and Spring, 2007.

[11] Irwin, A. *Citizen Science: A Study of People, Expertise and Sustainable Development.*: London, England: Routledge, ISBN-10: 0415130107, 1995.

[12] Ito, Mizuko et al. "Hanging Out, Messing Around, and Geeking Out: Kids Living and Learning with New Media." Part of the John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. *MIT Press*. ISBN-10: 0262013363, 2009.

[13] Jenkins, H. *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. MIT Press: Cambridge, MA., ISBN-10: 0262513625, 2009.

[14] Kuriyan, R. and Nafus, D. "A New Market or Development for the Poor: Depoliticizing the Consumer." *Bottom of the Pyramid Workshop*, UC Irvine, June 2009.

[15] Morawczyunski, O. and Miscione "Examining trust in mobile banking transactions : the case of m-pesa in Kenya." In *Social dimensions of information and communication technology policy : proceedings of the 8th International conference on human choice and computers, HCC8, IFIP TC 9*. Pretoria, South Africa, September 25-26, 2008. Eds. C. Avgerou, M.L. Smith, and P. van den Besselaar. Boston: Springer, ISBN-10: 0387848211, 2008.

[16] Mosk, M. "Obama Shatters Fundraising Record." *Washington Post*. October 19, 2008.

[17] Norris, P. *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge: Cambridge University Press. ISBN-10: 0521002230, 2001.

[18] Paulos, E., Honicky, R.J. and Hooker, B. "Citizen Science: Enabling Participatory Urbanism." In *Foth, M. (ed.) Handbook of Research on Urban Informatics*, IGI, pages 414-436, Global, Hershey, PA, 2008.

[19] Shirky, C. *Here Comes Everybody: The Power of Organizing Without Organizations*. Penguin, ISBN-10: 0143114948, 2009.

[20] United Nations. "Tools for Development. Using Information and Communication Technologies to Achieve the Millennium Development Goals." *UN ICT Task Force Working Paper*, 2003.

[21] Willett, W., Aoki, P.M., Kumar, N., Subramanian, S. and Woodruff, A.. "Common Sense Community: Scaffolding Mobile Sensing and Analysis for Novice Users." In *Proceedings Pervasive*, Springer 2010, to appear.

Acknowledgments

We acknowledge the assistance of our colleagues in conducting the research behind collective participation and thoughtful comments on this article: Maria Bezaitis, Prabal Dutta, R.J. Honicky, Kathi Kitner, Neil Kumar, Alan Mainwaring, Scott Mainwaring, Chris Myers, Dawn Nafus, Eric Paulos, Sushmita Subramanian, Wesley Willett, and Allison Woodruff.

Author Biographies

Ken Anderson is a Senior Research Scientist with the People and Practices Research Group. Over the last 20 years, he has explored the relationship between identity, culture, and technology in positions he has held at AT&T, MediaOne, US West, Bethel College, and Apple Computer. Ken's current research is in emerging regions, focusing on new media, mobile computing, and participatory culture. Ken is the founder of Ethnographic Praxis in Industry Conference (EPIC) and the current president of the board of directors. His e-mail is ken.anderson at intel.com.

Paul M. Aoki is a Senior Research Scientist with Intel Labs Berkeley. He holds a BS degree in Electrical Engineering and a PhD in Computer Science from the University of California, Berkeley. His research interests include mobile computing and communication systems, human-computer interaction, and computer-supported cooperative work. His current focus is on mobile sensor devices for community action and on the design of technologies for emerging regions. His e-mail is paul.m.aoki at intel.com.

Michael Kaminsky is a Senior Research Scientist at Intel Labs Pittsburgh and an adjunct faculty in the Computer Science Department at Carnegie Mellon University. Michael joined Intel in the summer of 2004 after completing his PhD in Computer Science at MIT. He received his BS degree in EECS from the University of California, Berkeley in the spring of 1998. Michael is generally interested in computer science systems research, including distributed systems, networking, operating systems and network/systems security. His e-mail is michael.e.kaminsky at intel.com.

Renee Kuriyan is a Research Scientist in the People and Practices Research group. Prior to joining Intel, she conducted research on international development topics with Microsoft Research India, the World Bank, and the United Nations Development Program. She holds a PhD degree with a focus on development studies from the University of California, Berkeley. She holds a MS degree in Public Affairs from Princeton University with a focus on international development, and on science and technology policy. Her e-mail is renee.kuriyan.wittemyer at intel.com.

Copyright

AUGMENTED REALITY ON MOBILE INTERNET DEVICES BASED ON INTEL® ATOM™ TECHNOLOGY

Contributors

Maha El Choubassi
Intel Corporation

Yi Wu
Intel Corporation

Index Words

Mobile Augmented Reality
Mobile Internet Device
Intel® Atom™ Processor
Image Match
Orientation and Location Sensor
Object Track

Abstract

Many of the latest mobile devices contain cameras, as well as location, orientation, and motion sensors. In this article, we demonstrate how mobile Internet devices, based on the Intel® Atom™ processor can access and provide information to users in a different manner than traditional web-based information sources. Our system uses a combination of camera, and location and orientation sensors, to augment live camera views on a device with available information about the objects in the view. The augmenting information is obtained by matching a camera image to a database that resides on a server, which contains half a million images. Each image has geotags identifying its location. After matching an image taken with the camera with images from the database, a combination of motion-estimation algorithms and orientation sensors is used to track objects of interest in the live camera view and to place augmented information on top of these objects. We optimized the software to take advantage of the Intel Atom processor, and we present the results of a detailed performance analysis.

Introduction

In the past few years, various methods have been suggested to present augmented content to users through mobile devices [1-5]. Many of the latest mobile Internet devices (MIDs) feature consumer-grade cameras, WAN and WLAN network connectivity, location sensors (such as global positioning systems or GPSs) and various orientation and motion sensors. Recently, several reality-augmenting applications, such as Wikitude (*www.wikitude.org*), and similar applications for the iPhone* have been announced. Although similar in nature to our proposed system, these solutions rely solely on location and orientation sensors, and, therefore, require detailed location information about points of interest to correctly align augmenting information with visible objects. Our system extends this approach by using image-matching techniques and location sensors together for object recognition and for precise placement of augmenting information. The new Google Goggles* application (*www.google.com/mobile/goggles*) also combines image matching with location information. However, Google's approach is to send the query image to remote servers for processing, which might incur long response times, due to network latency. Powered by the Intel Atom processor, our system can perform image matching on the MID itself instead of shifting all computation to remote servers.

In this article, we demonstrate a complete, end-to-end, mobile augmented reality (MAR) system that consists of MIDs, powered by the Intel Atom processor, and a Web-based MAR service hosted on a server. On the server, we store a large database of images automatically extracted from geotagged Wikipedia* pages that we update on a regular basis. Figure 1 shows a snapshot of the actual client interface running on the MID. In this example the user has taken a picture of the Golden Gate Bridge in San Francisco. The MAR system uses the location of the user along with the camera image to return the top five candidate images from the database on the server that match the image taken on the handheld device. The user has the option of selecting the image of interest, which links to a corresponding Wikipedia page with information about the Golden Gate Bridge. A transparent logo is then added to the live camera view, virtually "pinned" to the object (Figure 1). The user will see the tag whenever the object is in the camera view and can click on it to return the retrieved information.

> *"The system uses the location of the user along with the camera image to return the top five candidate images from the database on the server that match the image taken on the handheld device."*

Help build the future of Wikipedia and its sister projects! Read a letter from Jimmy Wales and Michael Snow.

Golden Gate Bridge

From Wikipedia, the free encyclopedia

The **Golden Gate Bridge** is a suspension bridge spanning the Golden Gate, the opening of the San Francisco Bay onto the Pacific Ocean. As part of both U.S. Route 101 and California State Route 1, it connects the city of San Francisco on the northern tip of the San Francisco Peninsula to Marin County. The Golden Gate Bridge was the longest suspension bridge span in the world

Figure 1: Illustration of the MAR System Showing Results when Querying the Golden Gate Bridge
Source: Intel Corporation, 2010

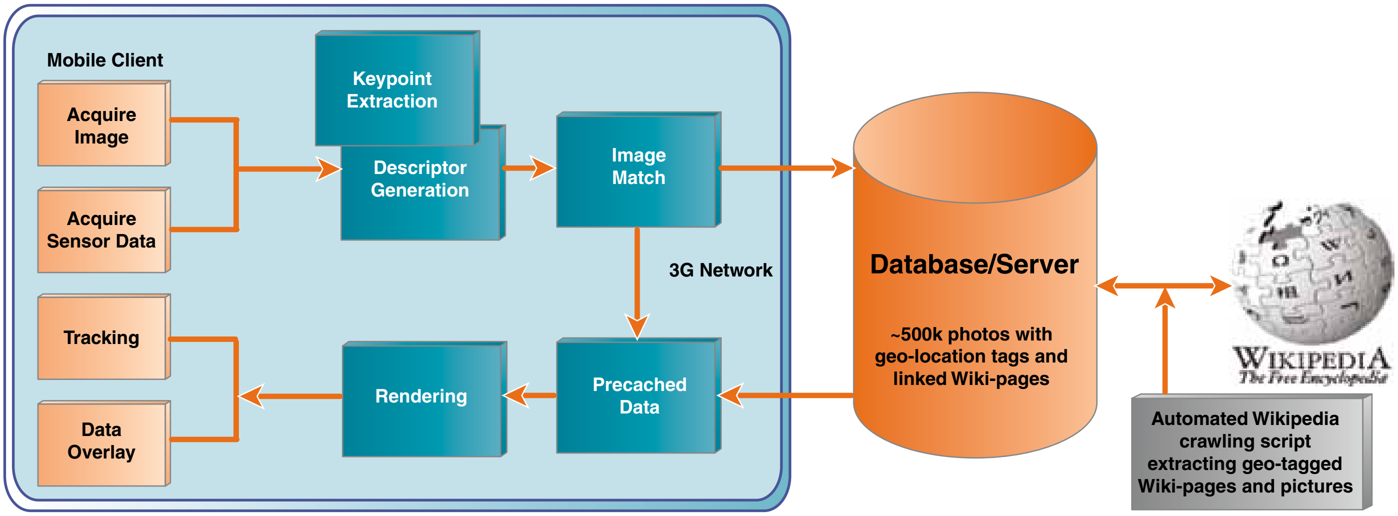


Figure 2: MAR System Diagram with Image Acquisition, Feature Extraction, Tracking, Rendering, and Image Matching Running on the Client and the Large Database Residing on the Server
Source: Intel Corporation, 2010

System Overview

The MAR system (shown in Figure 2) is partitioned into two components: a client MID and a server. The MID communicates with the server through a WAN or WLAN network.

Client

The client performs the following operations:

"As soon as a picture is taken, data from location and orientation sensors are saved to the image's field."

"Visual features are then extracted from the picture."

- *Query acquisition.* The client MID device contains a camera and an orientation and location sensor (such as a GPS). The client continuously acquires live video from the camera and waits for the user to select the picture of interest. As soon as a picture is taken, data from location and orientation sensors are saved to the image's Exchangeable image file format (EXIF) field.
- *Feature extraction.* Visual features are then extracted from the picture. We compared several well-known feature point techniques, such as Scale-invariant feature transform (SIFT) [6] and Speeded-Up Robust Features (SURF) [7]. We decided to use 64-dimensional SURF image features because their compute efficiency is higher than that of SIFT features, at comparable recognition rates. If users know their destination beforehand, they can precache a small database of image features, thumbnails, and related metadata in the corresponding neighborhood. With a precached database, image matching can run on the client MID device. Otherwise, the client sends the extracted feature vectors and recorded sensor data to the server to search for matched images and related information. The reason for working with the compact visual features instead of full resolution images is to reduce network bandwidth, hence latency, when sending the image data to the server, and also to efficiently use the memory resources on the MID, when precaching the data.

- *Rendering and overlay.* Once the matching image is found, related data, including the image thumbnail and the Wikipage link, are immediately accessed from the precached database if available, or else they are downloaded to the client via the WAN/WLAN network. The client renders the augmented information, such as the Wikipedia tag of the query object, and overlays it on the live video. In addition to constraining the image recognition based on location information, the orientation sensors (compass, accelerometer, and gyroscope) of our device are used to track the location of objects to overlay augmented information on a live camera view. As the user moves the device, the augmented information representing the query is pinned to the position of the object. This way, the user can interact separately with multiple queries by simply pointing the camera toward different locations.
- *Tracking.* When a query is made, the pointing direction of the MID is recorded by using the orientation sensors on the device. The client continuously tracks the movement of orientation sensors [8, 9]. However, tracking that uses orientation sensors is not very precise. We extend our tracking method to also include the visual information. We use image-based stabilization, which is based on aligning neighboring frames in the input image sequence by use of a low-parametric motion model. The motion estimation algorithm is based on a multi-resolution, iterative, gradient-based strategy [10], optionally robust in a statistical sense [11]. Two different motion models have been considered: pure translation (two parameters) and pure camera rotation (three parameters).

"Tracking that uses orientation sensors is not very precise. We extend our tracking method to also include the visual information."

Server

The server performs the following operations:

- *Database collection.* We automatically extracted Wikipedia pages with their concomitant GPS information [8] from the Wikipedia database. We downloaded the images from these pages, extracted visual features from images, and added images and features to our image database. Our database is constantly growing, and it now contains more than 500,000 images. This fact further emphasizes the need for and the importance of applications that can explore large numbers of images.
- *Image match.* As mentioned in the previous section, in the absence of a precached database on an MID device, the server receives the features of the query image from the client and performs the image match.

Image-matching Algorithm

Our image-matching approach combines both the GPS information and the image content. More explicitly, we restrict the database of candidate-matching images to those within a predefined search radius around the GPS value at the location of interest, i.e., the user's actual location or, in the case of precaching, the destination. Next, we use computer vision techniques to match the query image to the set of images that is near a certain GPS location. After the GPS filters the information, we downscale the database from 500,000 images to a much smaller candidate image set (10- ~100 images) to increase the likelihood of the image-based matching algorithm being successful.

"We restrict the database of candidate-matching images to those within a predefined search radius around the GPS value at the location of interest."

To recognize the top matching database images, our algorithm inspects each database image and compares it with the query image. More specifically, for each SURF keypoint in the query image, the algorithm finds the nearest and the second nearest neighbor keypoints in the database image, based on the city block (L1) distance between the keypoints' descriptors. Next, it computes the ratio of the nearest distance to the second nearest distance. If the ratio is smaller than a threshold, the algorithm decides that the nearest point is a "true" match of the query keypoint, since it is significantly closer to it, in L1 distance, than any other keypoint, as in [6]. After the algorithm detects the matching pairs between the query image and each database image, it ranks these images in descending order according to the number of such pairs. In our system, we return the top five images to the user.

Finally, if the number of nearby candidates is reasonable, i.e., in the range of a few images (~10), we perform brute-force image matching, based on the ratio of the distances between the nearest and second-nearest neighbor image features as explained previously. Otherwise, for scenarios with a high density of nearby images or with no GPS location available, the number of candidate images is large, and we use Fast Approximate Nearest Neighbor (FLANN) indexing as proposed in [12].

"If the number of nearby candidates is reasonable, i.e., in the range of a few images (~10), we perform brute-force image matching."

Building a More Realistic Database

In [8], we showed quantitative results on the standard ZuBuD test set [13], with 1005 database images and 115 query images, together with qualitative results for a small benchmark of images of iconic landmarks. Although the results on the ZuBuD set are informative, they do not precisely reflect the nature of photos taken by people under realistic conditions. To better simulate the performance of our MAR system, we selected 10 landmarks from around the world: *the Arc de Triomphe* (France), *the Leaning Tower of Pisa* (Italy), *the Petronas towers* (Malaysia), *the Colosseum* (Rome, Italy), *the National Palace Museum* (Taiwan), *the Golden Gate Bridge* (California, USA), *the Chiang Kai Shek Memorial Hall* (Taiwan), *the Capitol* (Washington D.C., USA), *the Palais de Justice* (Paris, France), and *the Brandenburg Gate* (Berlin, Germany). For each landmark, we tested the recognition performance of the MAR system. First, we reduced the 500,000-image database on the server to a smaller set that included only the images within a predefined radius. Next, we randomly picked one of the landmark images in the set as a query image, and we kept the others as database images. This provided one query image and a database for each landmark. In total, we had 10 separate databases, with a total of 777 images. The characteristics of these databases vary between sites. For instance, the size of the database is only 5 for the National Palace Museum, while it is

200 for the Palais de Justice. Also, the database corresponding to the Golden Gate Bridge has 12 relevant images of the bridge out of 19 images, while the one corresponding to the Palace of Justice has only one relevant image out of a total of 200 images. Also, images labeled by the same landmark might look different depending on the lighting conditions, the viewpoint, and the amount of clutter. To take advantage of these databases, we manually annotate them. Simulating MAR on the final databases, we obtain representative results for MAR's performance, since these databases are sampled from the Wikipedia data on our server, i.e., the actual database queried by MAR.

"One query image and a database for each landmark. In total, we had 10 separate databases, with a total of 777 images."

Matching Enhancement by Duplicates Removal

Web images have varying sizes and visual features density; therefore, many practical problems arise with realistic data that are not present in standard databases. Running MAR on the data just discussed in the "Building a More Realistic Database" section, we identified one problem in the matching algorithm that degrades MAR's performance: there are multiple keypoints in the query image matching one keypoint in the database image. Even when these matches are false, they are highly amplified by their multiplicity, which eventually affects the overall retrieval performance. We adjusted the algorithm to remove duplicate matches and to improve matching accuracy at no computational cost, as shown next.

We believe that the duplicate matches' problem particularly arises in cases where a strong imbalance exists between the number of keypoints in the database and query images, i.e., there are much fewer keypoints in the database image than in the query image. For each keypoint in the query image, the algorithm shown in our "Image Matching Algorithm" section detects the matching keypoint in the database image. Therefore, the imbalance forces many keypoints in the query image to match one single point in the database image. The standard ZuBuD data set does not have this issue, since its images have comparable numbers of keypoints. Contrarily, Wikipedia and web images in general have high variance, which emphasizes the necessity of having representative sample databases as mentioned earlier. To solve this problem, our adjusted algorithm prohibits duplicate matches. Whenever a keypoint in the database image has multiple matches in the query image, we pick only the keypoint with the closest descriptor. We recognize that duplicate matches may still be correct, e.g., repetitive structures. However, as we show in the "Experimental Results Extension" section of this article, substituting the "best" matching descriptor, i.e., the nearest, for the duplicates, improves the overall performance of our system.

"Our algorithm prohibits duplicate matches. Whenever a keypoint in the database image has multiple matches in the query image, we pick only the keypoint with the closest descriptor."

Matching Enhancement by Histograms of Distances

Originally, we match the query image to each database image one at a time, by using the ratio of keypoints descriptors distances (see the "Image Matching Algorithm" section). However, we intuitively expect the distances between the descriptors of the query and a mismatching image to be larger than those between the query and a matching image. Hence, instead of merely relying on the distances ratio, we propose further exploring other statistics, based on descriptors' distances. More explicitly, we take the top ten matching database images obtained from the original algorithm shown in the "Image Matching Algorithm" section, and for each (query image, database image) pair, we build the histogram of descriptors distances. Our purpose is to extract more information from these distances about the similarity or dissimilarity between the query and the database images. For instance, by filtering out database images with large average distance, we remove obviously mismatching images. The cost of this approach is not high, since the distances are already computed, and hence we are only leveraging their availability. We applied this approach to ten query images and their corresponding GPS-constrained databases. Results are shown in the "experimental" section of this article.

Algorithm

Our image-matching algorithm works this way: Let Q be the query image. We first apply our existing matching algorithm and we retrieve the top ten database images D_1, D_2, …, D_{10}. Next, we consider each pair (Q,D_i) at a time and build the histogram H_i of minimum distances from keypoints in the query image Q to the database image D_i. There is no additional overhead, since these distances were already calculated by the existing matching algorithm. For each histogram H_i, we obtain its empirical mean M_i and skewness S_i:

$$M_i = \frac{1}{n}\sum_{j=1}^{n} H_{i,j},$$

$$S_i = \frac{\frac{1}{n}\sum_{j=1}^{n} (H_{i,j} - M_i)^3}{\left(\frac{1}{n}\sum_{j=1}^{n} (H_{i,j} - M_i)^2\right)^{3/2}}$$

The smaller the skewness, the closer H_i is to symmetric. Our main assumptions are these:

1. If M_i is large, then many of the descriptors pairs between Q and D_i are quite distant and hence are highly likely to be mismatches. Therefore, image D_i must not be considered a match for image Q.
2. If S_i is small (close to zero), then the histogram H_i is almost symmetric. Having many descriptors in Q and D_i that are "randomly" related, i.e., not necessarily matching, would result in this symmetry. We expect this scenario when the two images Q and D_i don't match and hence there is no reason for the histogram to be biased.

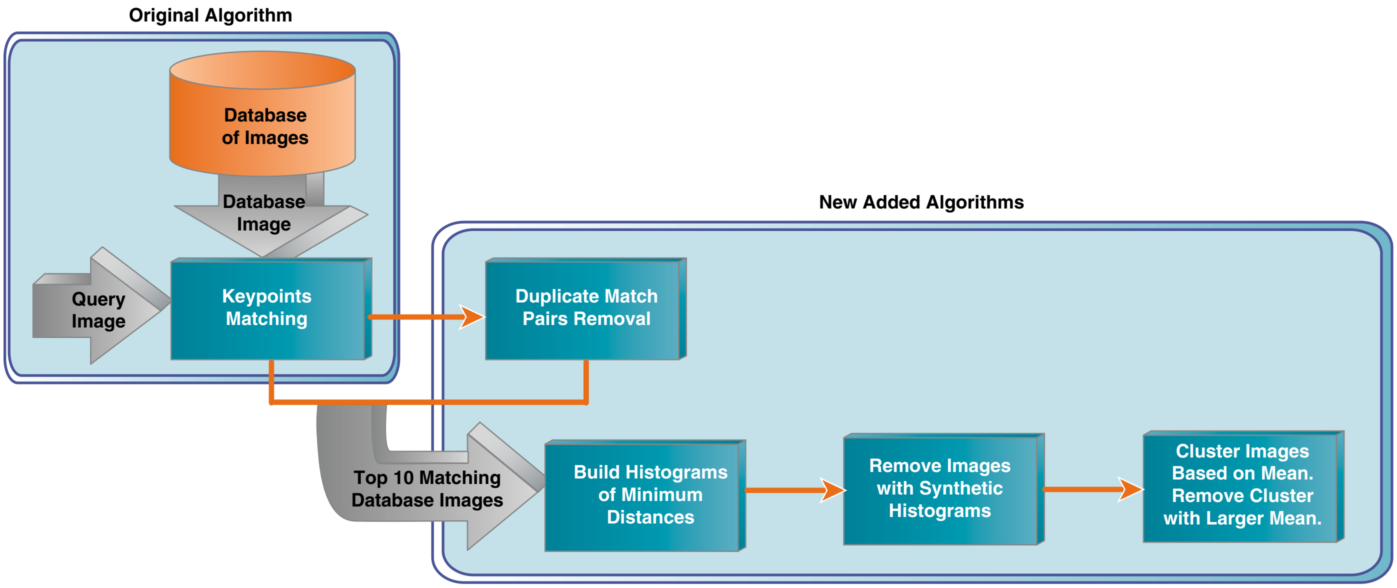


Figure 3: Block Diagram of our Algorithm for Image Matching with the Duplicates Removal and Enhancement Based on Histograms Added
Source: Intel Corporation, 2010

Based on these assumptions, we remove database images that have very low skew S_i. We also cluster the images based on the means M_1, M_2, …, M_{10} into two clusters (we used k-means). We remove the images that belong to the cluster with the higher mean. For the experimental results, please refer to the "Matching Enhancement through Histograms Distances" section of this article. Figure 3 shows a block diagram of our adjusted matching algorithm.

Matching Enhancement by Database Extension

As mentioned before, the images on our server are extracted from Wikipedia. Although Wikipedia is a good source for information in general, it is not the best source for images. For example, in the geographically restricted database of Wikipedia images in the surroundings of the Palais de Justice, only one out of 200 images in the database matches the building itself. However, not only the matching algorithm, but also the number of matching images in the database, affects the performance accuracy. Moreover, the query image might be taken from different views and under different light conditions. Having a small number of matching images, e.g., one image in the daylight for the Palais de Justice, would limit the performance of MAR. For this reason, it is preferred that the database has more images matching the query. We extend the database by combining GPS-location, geo-tagged Wikipedia pages, text in these pages, and the large resources of Web images. For each Wikipedia page, we don't only download the images on the page itself, but we also use the title as a text query on Google's image search engine. Next, we pick the first few retrieved images and associate them with the same Wikipedia page. The improvement in MAR's performance is shown in the results of the "Matching Enhancement through Database Extension" section in this article.

> *"We don't only download the images on the page itself, but we also use the title as a text query on Google's image search engine."*

Experimental Results Extension

In this section, we look at how matching can be enhanced by removing duplicate matches, by analyzing the histograms of distances, and by adding images to the landmark's database.

Duplicates Removal

To test the impact of duplicate matches' removal, we ran both the original and the new matching algorithms (see "Image Matching Algorithm" and "Matching Enhancement by Duplicates Removal") on our sample databases for various SURF thresholds. In part (a) of Figure 4, we display the average of the top 1 precision (**the precision of the top matched image**), top 5 precision (**the precision of the top 5 matched images**), and the average precision over the 10 databases. Clearly, removing the duplicate matches improves the performance.

Matching Enhancement through Histograms Distances

In part (b) of Figure 4, we display the top 10 retrieved images for the Chiang Kai Shek Memorial Hall in Taiwan and the Capitol in Washington DC from the database mentioned in the section "Building a More Realistic Database" by the updated matching algorithm with duplicates removal. We also used the histograms of distances to further refine the match, as explained previously. Based on the histograms analysis, images labeled with a red square in the figure are rejected and those tagged with a blue square are retained. From these experiments, we see that portraits and statues of people can be clearly distinguished and rejected, since most of the queries are those of buildings. Such images in general have a very symmetric histogram. They also may have a large mean. We observed similar results for the other queries mentioned previously also.

Matching Enhancement through Database Extension

For each of our 10 landmarks, we download the top 21 images returned by Google image search when queried with the title of the landmark Wikipedia page, and we add them to the corresponding landmark database. Next, we apply our matching algorithm to the improved results in part (c) of Figure 4. Obviously, the top 1 and top 5 precision values increase for almost all the 10 databases mentioned in the section entitled "Building a More Realistic Database."

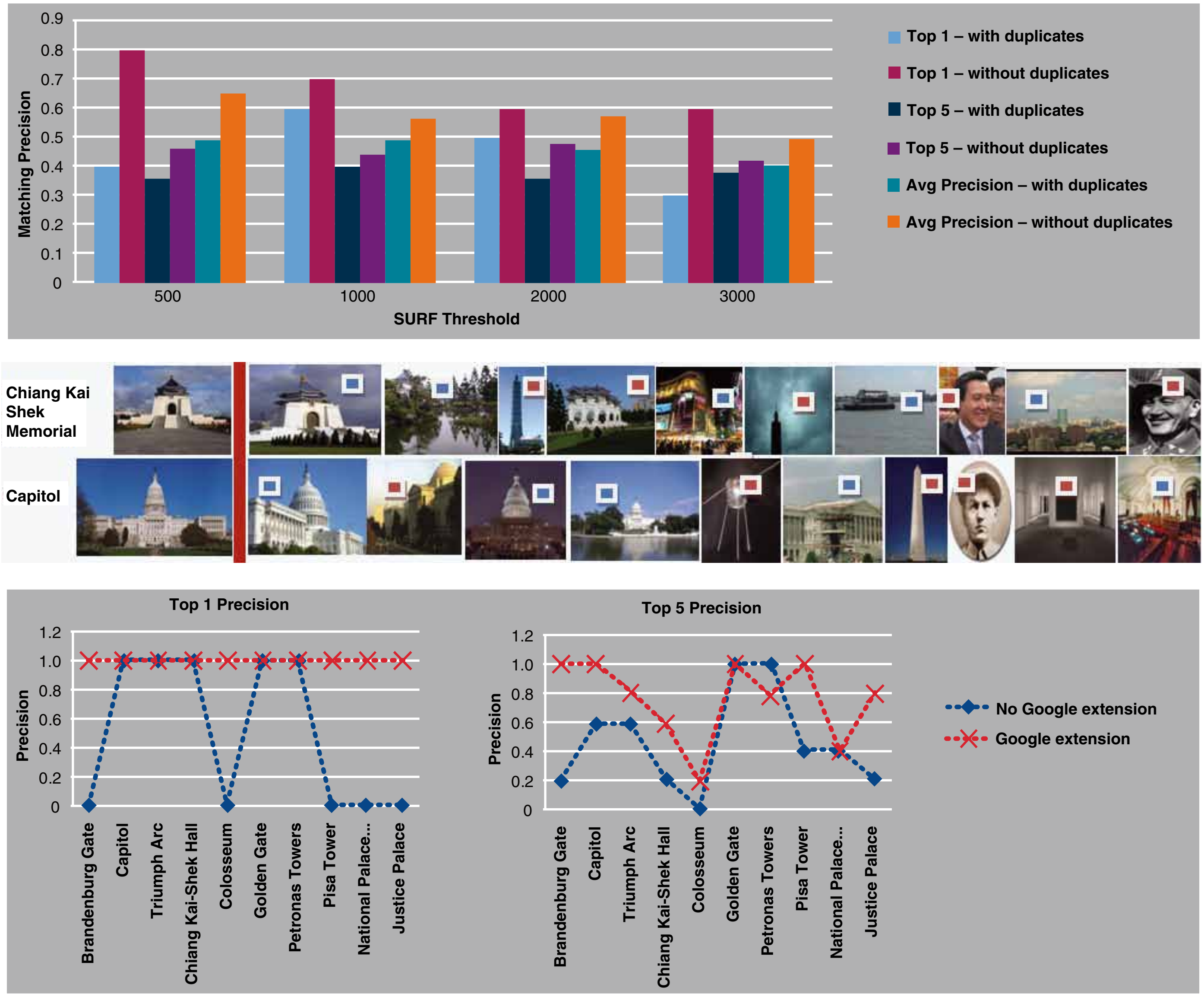


Figure 4: (a) Top 1, Top 5, and Average Precision Results for the 10 Landmarks vs. the SURF Threshold. (b) Matching Enhancement Based on Histograms of Distances. Images Before the Red Line Are the Queries. (c) Top 1 and Top 5 Precision without and with Google Database Extension for the 10 Landmarks

Source: Intel Corporation, 2010

> *"Converting keypoint detection from floating-point to integer arithmetic provided an additional 15% speedup."*

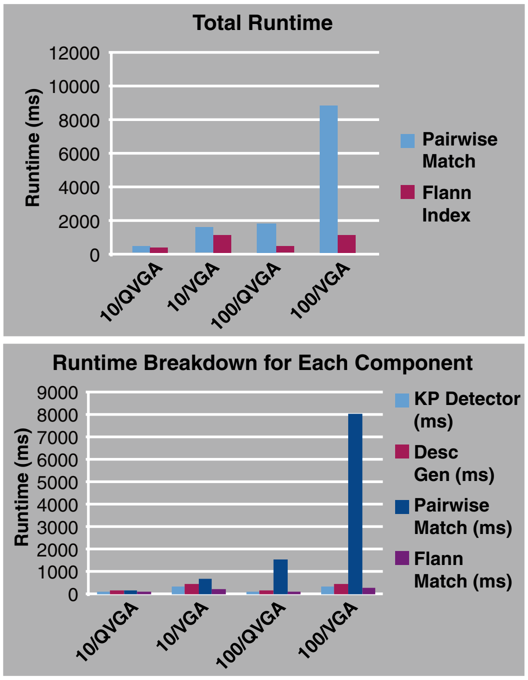


Figure 5: (a) Total Runtime Including Image Feature Extraction and Image Match; (b) Runtime Breakdown for Image Feature Extraction and Image Match
Source: Intel Corporation, 2010

Implementation and Optimization

In this section, we discuss how to optimize the algorithms to meet reasonable response time on the MID and how to take advantage of the features of the Intel Atom processor.

Feature Extraction and Image Match

As we have described in the previous sections, image feature extraction and image match (if the precached data were available on the MID) are performed on the MID client device. To perform these tasks, the MID requires significant computing power and memory storage.

Software Optimization

Software optimization includes the following components.

- *Image feature extraction.* The original SURF-based image feature extraction code is based on the OPENCV implementation (***http://sourceforge.net//projects/opencv/library***) as described earlier in this article. We identified two hotspots after using the VTune™ analyzer: keypoint detection and keypoint descriptor generation. We applied multiple optimization techniques to these hotspots to speed up the image feature extraction. We multi-threaded keypoint detection and keypoint description generation by using OPENMP, and we achieved 1.6X speedup when compared with the single-thread version on an Intel Atom processor. Converting keypoint detection from floating-point to integer arithmetic provided an additional 15% speedup. We also quantized a keypoint descriptor from float (32 bit) to char (8 bit) that resulted in a 4X reduction in the data storage requirements. Performance was improved by taking advantage of the integer operations without significantly degrading the quality of the results.
- *Image match.* We again used the VTune analyzer and identified distance calculations as the hotspot of image match. We multi-threaded keypoint detection and keypoint description generation by using OPENMP. We achieved a 1.7X speedup when compared with the single-threaded version on an Intel Atom processor. We also vectorized the distance calculation by using SSE intrinsics to take advantage of 4-way SIMD vector units in the Intel Atom processor, which provided a 2X speedup over the non-vectorized image match codes.

Performance on a Platform Based on the Intel Atom-Processor

We analyzed the software implementation on a single core, hyper-threaded Intel Atom system (800MHz, 256MB RAM 512KB L2 cache), running a Linux* operating system. Our performance analysis is conducted on four datasets with different sizes and resolutions: 10 QVGA images, 10 VGA images, 100 QVGA images, and 100 VGA images. We chose these datasets for two reasons: a) most of the current MID devices take QVGA or VGA video input; b) for a given GPS location, the visible landmarks normally range from 10 to 100. Hence, with the help of GPS localization, we need to pre-cache 10-100 landmarks for database comparison. Figure 5(a) shows the total runtime of datasets to compare performance by using pair-wise match versus FLANN indexing. Figure 5(b) lists the runtime breakdown for each component including keypoint detection, descriptor generation, and image match. For the VGA resolution, the number of keypoints per image is around 800; and for the QVGA resolution, the number of keypoints per image is around 350. From the figures, it is clear that pair-wise matching runtime increases linearly with the database size. With FLANN indexing, the runtime scales very well when the database size increases. When a query image needs to be compared against a database size larger than 10 (which is a common case), we should consider using FLANN indexing instead of pair-wise matching to get a faster response. Overall, the execution time is about one second for querying a VGA image from a 100 VGA image database.

Tracking

The tracking algorithm explained previously has been optimized by using 1) a simplified multi-resolution pyramid construction with simple 3-tap filters; 2) a reduced linear system with gradients from only 200 pixels in the image instead of from all the pixels in the images; 3) SSE instructions for the pyramid construction and the linear system solving; and 4) only the coarsest levels of the pyramid to estimate the alignment. Performance was measured on the same Intel Atom system by using VGA (640x480) input video and different options. The results are shown in Figure 6, which displays the measured frames per second (fps) for different models (displacement/camera rotation), estimation method (robust/non-robust) and resolution levels and iterations per level used. For pure displacement models, using non-robust estimation and running five iterations in Levels 3 and 4 of the multi-resolution pyramid (Level 1 being the original resolution) the performance is over 80 fps.

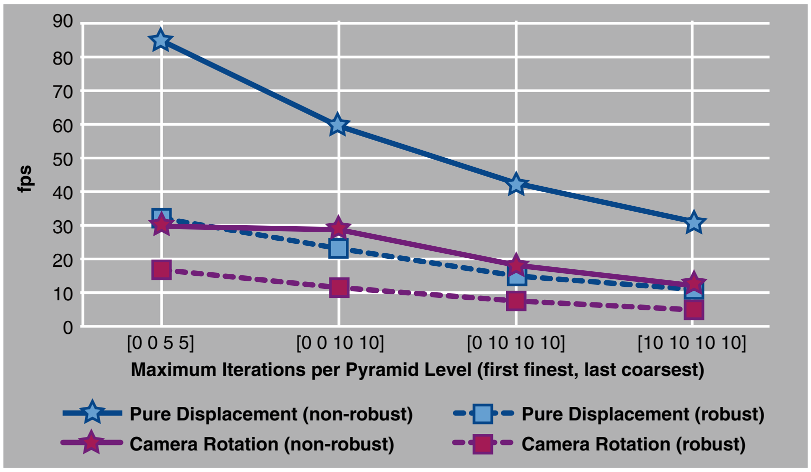


Figure 6: Performance Measured in Frames per Second (fps) of the Image-based Stabilization Method on an Intel® Atom™ System
Source: Intel Corporation, 2010

Conclusion

In [8], we presented MAR with a fully functional prototype on a MID device with an Intel Atom processor inside. In this article, we described new improvements to the matching and tracking algorithms in addition to the design of the system and its database. We also presented the code optimization benchmark results for the Intel Atom processor. With all these improvements, MAR demonstrates the powerful capabilities of future mobile devices derived from location sensors, network connectivity, and computational power.

"With all these improvements, MAR demonstrates the powerful capabilities of future mobile devices derived from location sensors, network connectivity, and computational power."

References

[1] S. Pradhan, C. Brignone, J.H. Cui, A. McReynolds, and M.T. Smith. "Websigns: hyperlinking physical locations to the web." *Computer,* Volume 34, pages 42-48, 2009.

[2] J. Lim, J. Chevallet, and S. N. Merah. "SnapToTell: Ubiquitous Information Access from Cameras." In *Mobile and Ubiquitous Information Access (MUIA04) Workshop,* 2004.

[3] Y. Zhou, X. Fan, X. Xie, Y. Gong, and W.Y. Ma. "Inquiring of the Sights from the Web via Camera Mobiles." In *Multimedia and Expo,* 2006 IEEE International Conference, pages 661-664, 2006.

[4] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W.C. Chen, T. Bismpigiannis, R. Grzeszczuk, K. Pulli, and B. Girod. "Outdoors augmented reality on mobile phone using loxel-based visual feature organization." *ACM International Multimedia Conference,* 2008.

[5] T. Quack, B. Leibe, and L. Van Gool. "World-scale mining of objects and events from community photo collections." In P*roceedings of the 2008 international conference on Content-based image and video retrieval,* pages 47-56, ACM New York, NY, USA, 2008.

[6] D.G. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision,* 60(2):91-110, 2004.

[7] H. Bay, T. Tuytelaars, and L. Van Gool. "SURF: Speeded Up Robust Features." *Lecture Notes in Computer Science,* 3951:404, 2006.

[8] D. Gray, I. Kozintsev, Y. Wu, and H. Haussecker. "WikiReality: augmenting reality with community driven websites." *International Conference on Multimedia Expo* (ICME), 2009.

[9] M. El Choubassi, O. Nestares, Y. Wu, I. Kozintsev, and H. Haussecker. "An Augmented reality tourist guide on your mobile devices." *16th International Multimedia Modeling Conference,* Chongqing, China, January 2010.

[10] BD. Lucas and T. Kanade. "An iterative image registration technique with an application to stereo vision. *International Joint Conference on Artificial Intelligence,* pages 674-679, 1981.

[11] O. Nestares and D.J. Heeger. "Robust multiresolution alignment of MRI brain volumes." Magnetic Resonance in Medicine, pages 705-715, 2000.

[12] Marius Muja and David G. Lowe. "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration." In *International Conference on Computer Vision Theory and Applications* (VISAPP'09), 2009.

[13] H. Shao, T. Svoboda, and L. Van Gool. "ZuBuD: Zurich Buildings Database for Image Based Recognition." *Technique report No. 260, Swiss Federal Institute of Technology,* 2003.

Author Biographies

Maha El Choubassi is a Senior Research Scientist at the Vision and Image Processing Research group/FTR at Intel labs. She joined Intel in February 2009 and received several recognition awards including three Intel Labs Divisional awards. She obtained her MSc and PhD degrees from the University of Illinois at Urbana-Champaign (UIUC) where she worked on signal processing and communications applications, particularly in the area of media security. Maha received multiple awards while at UIUC, including best paper award at the 2009 SPIE electronic imaging conference. She also received the James M. Henderson fellowship. Her e-mail is maha.el.choubassi at intel.com

Yi Wu is a Senior Research Scientist at the Vision and Image Processing Research group/FTR at Intel labs. She joined Intel in September 2005 and received several Intel labs awards. She received her PhD degree in Electrical and Computer Engineering from the University of California, Santa Barbara (UCSB) in 2005. Her research interests cover multimedia data processing, data mining, machine learning, and information retrieval. Her e-mail is yi.y.wu at intel.com.

Acknowledgements

The authors would like to thank Igor Kozintsev, Oscar Nestares, and Horst Haussecker for their significant contributions to this work.

Copyright

PHYSICALITY: INTERACTING WITH THE PHYSICAL WORLD, FROM ATOMS TO HUMANS

Contributors

Joshua R. Smith
Intel Corporation

Beverly L. Harrison
Intel Corporation

Xiaofeng Ren
Intel Corporation

Siddhartha Srinivasa
Intel Corporation

Index Words

Physicality
Tangible Interaction
Perception
Depth Sensing
Wireless Power
Robotics

Abstract

Physicality research at Intel is concerned with the embedding of computing systems in the physical world. The theme extends from physics-oriented topics such as energy harvesting, transmission, and storage, through robotics, to topics such as tangible interaction with digital systems and human-robot interaction.

Introduction

Computers are physical. They are subject to physical laws, such as the law of conservation of energy, and they interact with the physical world through sensors and actuators. Robots may be thought of as computers that are strongly coupled to the physical world through a particularly rich set of sensors and actuators. In fact, sufficiently capable robots, like those developed at Intel and described in this article, are so well coupled to the physical world that they can affect their own physical state in non-trivial ways, such as being able to manage their own energy supply.

The coupling of computers to the physical world is not just important for robotics. The physical user interface is ultimately what enables people to interact with and through digital systems. Mice, keyboards, and displays are physical interface devices that have become so successful that they are now completely unremarkable (though they are an enormous step forward from early physical interface devices such as punch cards and teletypes). Improvements in sensors and actuators are increasing the physicality of computer systems, bringing information out into the physical world, instead of requiring users to "enter" the computer's world. Accelerometers (in the iPhone* and Wii*) and multi-touch touch screens (in the iPhone screen) are recent examples of widely deployed physical input technologies that have enriched physical user interfaces. The trend toward physicality promises further increases in the richness of human-computer and technologically-mediated, human-human interactions.

Researchers at Intel are exploring physicality at many levels, including device-level research into energy storage and wireless power transfer, novel sensing and perception techniques, robotic systems, and richer physical input techniques. In this article we present a survey of physicality research at Intel.

Design, Interaction, HCI, and Physicality

One interpretation of "physically instantiated" user interfaces is that the physical manipulation of the devices themselves constitutes a means of interaction. In one such prototype implementation, Intel Labs, Seattle, in collaboration with Carnegie Mellon University, developed "Whack Gestures" [1]. This technique enables users to interact with their mobile device with minimal attention and effort and without even getting the device out or looking at it. Users issue simple gestural commands—in this instance a very satisfying "whack," even through pockets—and the detected gesture invokes a command (e.g., silence the device and turn off ringer). Such interactions can be subtle, non-disruptive, and undemanding in terms of user effort and visual attention. Our proposed technique is intended to be used with only a small number of common but high-value interactions; thus, there is no complex "gestural language" that a user must learn and a sensor must differentiate.

Accelerometers embedded within the device detect these distinct, abrupt gestural movements on the device and look up the gesture-command binding to determine which action to take. We tested this technique on 11 people: they wore an accelerometer on a mobile device and went about their normal routines for at least 2 hours (22 hours data collection total). We provided our recognizer model with sample gestures to be recognized (based on the accelerometer signal). We measured the accuracy of our system to determine the number of false positives (unintended) and true positives (intended gesture command). The recognizer signal model was not user specific. There was one false positive on this 22-hour "in the wild" data set (occurred during initial device attachment to the belt). We correctly classified 97% true positives. This preliminary test of the effectiveness of our interaction technique provides promising results that such simple gestures are a viable method of rapid, low effort interaction for small numbers of commands on mobile devices.

"Accelerometers embedded within the device detect these distinct, abrupt gestural movements on the device and look up the gesture-command binding to determine which action to take."

Hitting a device to make it do what you want may seem like an extremely literal example of a "physical" interface technique; we believe that this is a rich area of research, and that many additional examples, most of which will not involve whacking the computer in your pocket, will be developed in the coming years.

"Simple gestures are a viable method of rapid, low effort interaction for small numbers of commands on mobile devices."

"The problem of robotic navigation is now largely solved, if cost can be ignored; robotic manipulation, by contrast, remains unsolved at any price point."

Robotics

Robots may be the form of computing systems that are most obviously and crucially physical. Personal robots—ones that can perform useful tasks in unstructured human environments—represent a huge growth opportunity for the Personal Computer (PC) industry. Putting sophisticated compute-intensive robots in every home (not to mention offices, warehouses, hospitals, and previously un-automated factories) would also mean deploying a large number of new high-performance computer systems: the "brains" of the robots.

Personal robots are not yet feasible because of fundamental unsolved research problems in sensing, perception, machine learning, and robotics-specific topics, such as manipulation planning and control. Additionally, even when fundamental technical feasibility issues do not limit robotic applications, the cost of robots can limit the economic feasibility of these applications. Thus, robotics research will provide enormous return on investment if it can mitigate or bypass the technical and economic roadblocks.

The problem of creating robots that can function in unstructured human environments has two main components: navigation and manipulation. In recent years, excellent solutions to robotic navigation problems have been developed, enabled by the laser rangefinder sensor and probabilistic algorithms that use data efficiently, allowing the robot to manage its uncertainty about the state of the physical world [2].

Although the laser rangefinder has helped enable dramatic progress in robotic navigation, it is still an expensive electromechanical system, which limits the growth of commercial opportunities. New solid-state depth cameras promise to provide low-cost, high-quality depth images, registered with conventional Red-Green-Blue (RGB) images. Intel's research into advanced perception with depth cameras is described later in this article.

The problem of robotic navigation is now largely solved, if cost can be ignored; robotic manipulation, by contrast, remains unsolved at any price point. We believe that RGB and Depth (RGB-D) sensing will provide important benefits to manipulation, in addition to navigation. To help address the technical roadblocks to widespread deployment of Personal Robots [3], Intel Labs has created two personal robotics research platforms, HERB (the Home Exploring Robotic Butler) and Marvin. Each has a Segway mobile base, a laser range finder for navigation, a WAM arm, a Barrett Hand, onboard computing, and wireless communication. HERB was designed to support general mobile manipulation research and also has a custom spinning laser range finder for long range 3D perception. Marvin was designed to explore sensor-driven, real-time control for manipulation, and it has Electric Field Pretouch sensors, of our own design, built into its fingertips.

Electric Field Pretouch (hereafter referred to as E-Field Pretouch) is a sensing technique that is aimed specifically at improving robotic manipulation. We discuss this further in the section entitled "Electric Field Pretouch." E-Field Pretouch and RGB-D are complementary sensing techniques: RGB-D provides long-range information, and E-Field Pretouch provides information in the final stages of manipulation; i.e., in the crucial stage just before contact between the robot hand and the object. We describe several different uses of E-Field Pretouch for manipulation, and we also look at the use of E-Field sensing for detection of and alignment with electric power outlets.

Advanced Perception with RGB-D Cameras

Perceiving the physical world is crucial to physicality. For a sophisticated task, such as robotic manipulation, a robot needs to understand the environment and make decisions about how to interact with it. To build a rich model of its surroundings, a robot may utilize cameras or laser scanners: both have advantages and disadvantages. Ordinary visual cameras often suffer from robustness issues such as poor lighting conditions; typical laser scanners are expensive and return one scan line at a time, limiting their use in mobile and dynamic settings.

At Intel Labs, we have been pushing for the development of advanced perception technologies that use RGB-D cameras [4]; i.e., cameras capable of capturing both the visual appearance (RGB) and the depth (D) of a scene, in high-resolution, at full frame rates. The last years have seen dramatic improvements in the capabilities of these cameras. RGB-D cameras provide a cost-effective, solid state means to directly measure rich 3D information. They seamlessly combine the advantages of visual cameras and laser scanners, with no calibration and no moving parts. The maturation of RGB-D technologies opens up many research and application opportunities for perception.

Much of the commercial interest in RGB-D sensing so far has been for physical user interfaces. RGB-D sensors are expected to be widely deployed in commercial gaming systems to enable markerless, non-contact, full body, human-computer input [5]. We have been exploring additional applications of RGB-D sensors for robotics and physical user interfaces. (See "Tabletop ESP: Everyday Sensing and Perception in the Classroom," another article in this issue of the *Intel Technology Journal*).

> *"At Intel Labs, we have been pushing for the development of cameras capable of capturing both the visual appearance (RGB) and the depth (D) of a scene, in high-resolution, at full frame rates."*

(a)

(b)

Figure 1: An Inquisitive Robot with an RGB-D Camera Can Autonomously Capture the Shape and Appearance of New Objects
Source: Intel Corporation, 2010

In Figure 1, we show several examples of on-going research efforts at Intel Labs on RGB-D perception. One of our goals is to enable a robot to autonomously build up models of objects it encounters in its environment, so that it may continuously build up its knowledge of the world, recognizing familiar objects later on and/or manipulating them. Figure 1 (a) shows a scenario, in which an inquisitive robot picks up an object, studies it in its hand, and constructs a model with both 3D shape and appearance. Examples of the acquired object models are shown in Figure 1 (b). With the use of RGB-D cameras and a combination of matching and modeling techniques, the robot is capable of modeling a wide range of objects that are robust with regard to shape, material, and lighting conditions.

Another example of our RGB-D research is 3D modeling "at large;" i.e., building a 3D model of an entire indoor environment under a variety of lighting conditions and settings, traditionally a challenging problem for vision-based techniques and a laborious process for laser-based techniques. With RGB-D cameras and a direct 3D perception, we have developed a set of algorithms that make the modeling "easy": one can simply hand-hold a camera, walk through an indoor space freely at a normal pace, and obtain a 3D environment model on-the-fly. We have applied this technique to build a model of our lab space, shown in Figure 2 (a). The map contains a lot of visual and 3D details, as illustrated in a zoom-in in Figure 2 (b). Such a map can be useful in many ways: it can be used for robot navigation, manipulation, tele-presence, and visualization.

> *"With the use of RGB-D cameras and a combination of matching and modeling techniques, the robot is capable of modeling a wide range of objects."*

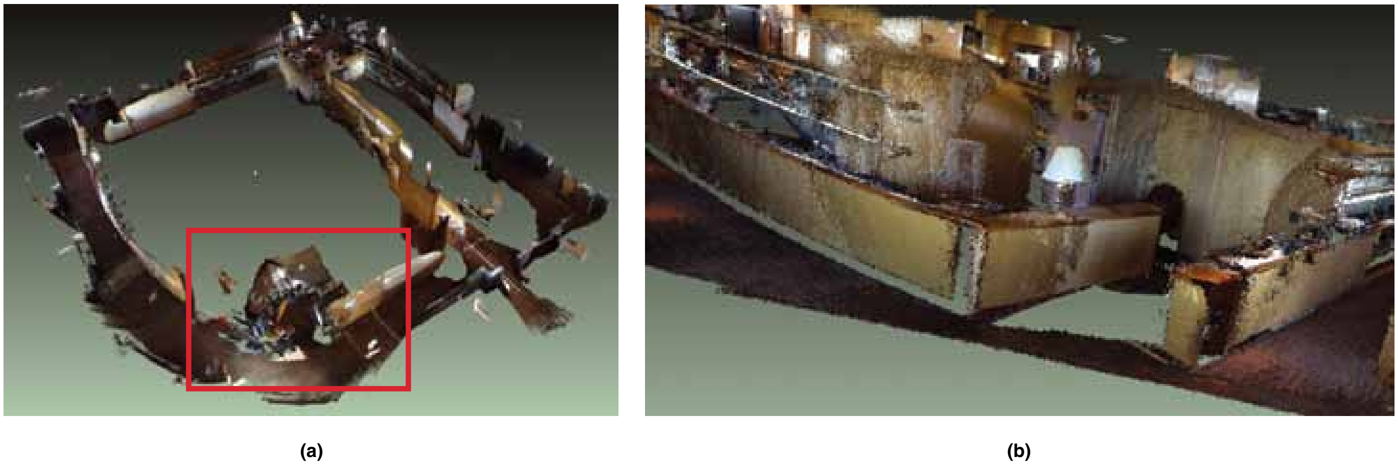


Figure 2: An On-the-fly 3D Model
Source: Intel Corporation, 2010

RGB-D cameras can be used in additional domains of physicality such as human-computer interaction or augmented reality, and they will soon be available to the general public for a reasonable price. At Intel Labs, we believe such 3D perception technologies will lead to robust solutions to many difficult perception problems, and that they will have a profound effect on the way we perceive the world and how we live in the world.

Electric Field Pretouch

RGB-D sensors, laser range finders, and cameras provide robots with rich, long-range, spatial information. This information can be used to recognize an object, estimate its pose, and then plan arm and hand motions to grasp it. Despite the richness of long-range optical sensing, it does not provide all the sensor information needed to reliably initiate a grasp with human-like precision. To understand why not, consider the following "straw man" grasping technique: capture a scene with an RGB-D camera, formulate a plan based on these data, and then execute the plan in an open-loop fashion, without collecting further sensor data. Because of image sensor noise and uncertainty in the robot's actuators, interleaving sensing and execution in the coarse-grained fashion of the "straw man" technique does not result in sufficient precision to pick up an object without noticeable manipulation errors. Even small hand-positioning errors can cause one finger to touch the object before the others, which tends to displace the object, and therefore lead to grasp failure. For a camera with 640 pixels of horizontal resolution and a 60-degree field of view, located 1.5 meters from the object, an error of just 4 pixels in the image plane corresponds to a 1-cm lateral position error in the object estimate.

"Despite the richness of long-range optical sensing, it does not provide all the sensor information needed to reliably initiate a grasp with human-like precision."

"The last centimeter before finger-object contact is the phase that is most difficult to monitor with a camera, because the fingers typically occlude the object."

Actuation errors are a

problem as well, in part because the effect of joint angle errors early in the robot arm's kinematic chain are scaled up by the arm length later in the chain. An additional shortcoming of long-range optical sensing is that the phase in which precise geometrical information is most crucial; i.e., the last centimeter before finger-object contact, is the phase that is most difficult to monitor with a camera, because the fingers typically occlude the object. Finally, depth-measuring technologies, optimized for long range sensing, tend to fail at short ranges.

Thus, the ability to close the sensing loop in the hand itself, by sensing the relative position of the fingers themselves to the object being manipulated at a very short range, could allow a robot to overcome all the problems just described.

We have been developing a new robotic sensing technique, called "Pretouch," whose purpose is to address these problems [6, 7]. Pretouch, a sense that we have built into the hand and fingers of our robot Marvin, operates at shorter range than vision, and is not subject to self-occlusion by the fingers or hand; yet, it is still non-contact, operating at longer range than touch sensing. E-Field Pretouch is a sense that humans do not possess, though some species of fish do have this sense.

To explore E-Field Pretouch, we designed electric field sensor boards and built them into the fingers and palm of a robot hand. Each electric field sensor channel uses a pair of electrodes. A 150kHz sinusoid is applied to the transmit electrode; this induces a displacement current in the receive electrode. Nearby objects affect the amount of displacement current received. The length scale of the sensor (i.e., the range at which it senses) is determined by the distance from the transmit electrode to the receive electrode: the closer the pair, the shorter the sensor range. We built electrode pairs into the robot hand to support several different sensing ranges, because different phases of the manipulation process benefit from different types of information. Early in the manipulation process, longer-range measurements are useful; as the manipulation process continues, shorter-range measurements become most useful. The longest-range sensor pairs in the robot hand measure about 10cm from palm to fingertip; the mid-range sensors measure about 3cm from the middle of the finger link to the finger tip; the short-range sensors only detect objects very close to the fingertip (less than a 1cm range). Note that these are range figures (meaning the longest range at which a non-zero signal is detected); they are not precision figures. Within their working range, the sensors can deliver mm-level precision.

Figure 3 shows a slightly different form of E-Field Pretouch. In this figure, Marvin is plugging itself into an ordinary electrical outlet to recharge itself, as described in [9]. It aligns its electrical plug with the outlet by sensing the 60Hz AC emissions from the electricity in the outlet. In this technique, there are no transmit electrodes: the building wiring serves as a transmitter of 60Hz AC electric fields. These fields function as a pervasive beacon, particularly relevant to charging applications. The fact that Marvin is able to manipulate its own state (i.e., its power state: plugged in or not) in a non-trivial way, is an exciting step. Practically, the more that robots can take care of themselves, for example by "feeding" themselves, the more useful they will be. Being able to recharge from any ordinary unmodified electrical outlet gives the robot more flexibility than requiring a custom charging dock. The robot is no longer confined to a home base; it could instead roam nomadically from outlet to outlet. From an intellectual perspective, it is exciting to have a general-purpose physical robotic platform that that can affect a truly important aspect of its own physical state, namely, whether or not it is plugged in.

Figure 3: Robot Marvin Uses Electric Field Pretouch both for Manipulation and for Sensing the Location of Electrical Outlets
Source: Intel Corporation, 2010

Mobile Manipulation

For the Mobile Manipulation component of Intel's Personal Robotics research agenda, Intel Labs, Pittsburgh, has created *HERB*, the Home Exploring Robotic Butler [10], illustrated in Figure 4. HERB is a complete mobile manipulation system that combines many advanced sensing, perception, manipulation, and planning techniques. HERB was designed as a research platform to develop "*Physical Grep*," our name for the task of autonomously fetching desired objects in an unstructured human environment. HERB can efficiently map, search, and navigate through indoor environments, recognize and localize several common household objects, and perform complex manipulation tasks such as carrying pitchers without spilling them.

Figure 4: HERB the Personal Robot Developed at Intel Labs Pittsburgh and a Snapshot of Some of the Modules
Source: Intel Corporation, 2010

HERB uses several onboard and offboard components to achieve autonomous mobile manipulation. In addition to the Segway* mobile base and Barrett WAM arm* (hardware that is common to Marvin), HERB includes cameras, a custom-built spinning laser, and a pair of low-power computers, all of which are powered by a custom-built power supply. Onboard components communicate over a wireless network with offboard, off-the-shelf PCs.

HERB implements a set of sensing, planning, and execution modules. A high-level script arbitrates their execution and error recovery. First, the robot senses its environment and sends a snapshot of the world to the planning system. Second, the planning system uses the geometry and kinematics of the robot to create a plan that avoids obstacles and meets task-specific constraints. Third, the execution system attempts to follow this global plan while compensating for dynamics, uncertainty, and error recovery.

These are some of the modules we have developed and deployed on HERB:

- GATMO: for detecting and tracking moving and movable objects to navigate in dynamic indoor environments.
- MOPED: for recognizing and localizing everyday objects from their shape and patterns, from a single image.
- Caging Planner: for reasoning quickly about opening doors by using simplified physics models.
- TSR Planner: for reaching into clutter to pick up objects and to reason about uncertainty.
- CHOMP: for producing optimized trajectories that are predictable and human-like.

HERB is a concept system for developing and deploying state-of-the art robotics algorithms. It is also an excellent system for testing the compute workloads of the future, constantly forcing us to make tradeoffs between low power consumption and high computing performance, and onboard versus offboard computing. We have learned many lessons working on HERB over the past three years, and we have been amazed by some of the behaviors that evolved autonomously.

For example, an early version of HERB puzzled us when it was grabbing coffee cups to place them in a dishwasher rack. It used a strange hand position, with one of its "thumbs" pointing down. After some scrutiny, we realized this was much like an efficient manipulation strategy used by professional bartenders who lift from underneath and pour in a single smooth motion. It is gratifying that HERB is complex and capable enough to compute solutions to physical manipulation problems that its designers had not anticipated.

Energy

Energy is becoming a crucial physical constraint on the design and operation of computing systems. Intel researchers are engaged in projects spanning a variety of energy-related topics, including energy scavenging, wireless power transfer, and energy storage.

Scavenging: Wireless Ambient Radio Power

Wireless Ambient Radio Power (WARP) turns the signals from broadcast TV into usable DC power [11]. We have powered a commercial, off-the-shelf temperature and humidity meter, with LCD display, from the signals broadcast from a TV tower 4.1km away. Figure 5a illustrates our first WARP system. This system uses a 5dBi Yagi antenna and a multi-stage voltage multiplying rectifier to scavenge around 60 microwatts from a 960kW source. TV signals are typically broadcast 24 hours per day, meaning that this energy source is available even at night, unlike solar energy. The WARP technique is better suited to populated and developed areas, which are more likely to have TV transmission; it is not appropriate for remote locations far from broadcast sources. Another limitation of the system today is that the power-scavenging antenna must be oriented toward the power source. As power consumption is driven lower, it should become possible to operate low-power devices with an omni-directional antenna.

Far Field Wireless Power Transfer

The Wireless Identification and Sensing Platform (WISP) [12-13] uses a rectifying circuit similar to that of WARP's, but rather than scavenging "wild" power (the TV signal, which was not intended to be a power source), it harvests deliberately "planted" power that is emitted from a standard commercially available UHF RFID reader (see Figure 5b). To the reader, a WISP looks like an ordinary RFID tag. RFID tags typically do nothing more than return a unique ID—conventional RFID tags are essentially electronic barcodes. Unlike conventional RFID tags, a WISP has sensors and a fully programmable microcontroller.

WISPs typically harvest around 60 microwatts (net) at a range of 5 feet from a reader transmitting 4W Effective Isotropic Radiated Power (EIRP). The power consumption of the WISP, when it is actively reading the sensors and communicating with the reader, is on the order of 1 milli-watt (1000 microwatts), much higher than the instantaneous power available. Nevertheless, by aggressively putting the microcontroller and sensors into sleep mode, and by retaining unused power on a capacitor, it is possible to reduce the net power consumption so that it matches the available net power. This is accomplished by having the microcontroller sleep until sufficient energy on the storage capacitor has accumulated to execute one atomic unit of workload, such as reading a sensor and communicating the result to the reader. This technique has the effect of adapting the WISP's response rate so that the average power consumed exactly matches the average power available.

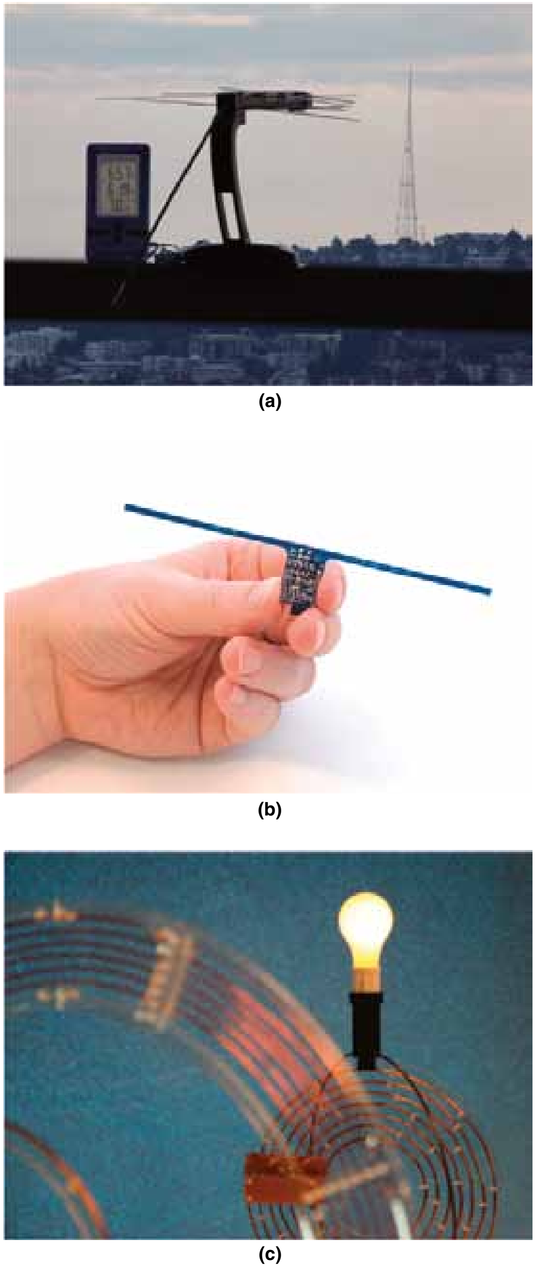


Figure 5: a: WARP; b: WISP; c: WREL
Source: Intel Corporation, 2010

We have built many generations of WISP. Our most recent WISPs communicate by using the popular EPC Class 1 Generation 2 (EPC C1G2) standard, and include a 3-axis accelerometer, a temperature sensor, a general-purpose IO pin-based capacitive sensor, and a power supply voltage sensor. Figure 5b shows an example WISP device.

WISPs are being used for a variety of sensing applications, such as neural sensing [14], sea water temperature measurement in an undersea neutrino telescope [15], and human activity recognition [16]. In addition to sensing applications, WISP has become popular as a platform for RFID security research, because it is the only available "hackable" (i.e., programmable, modifiable, extensible) UHF RFID tag. It has been used to implement the RC5 encryption algorithm [17] and to explore the design space of secure "Computational RFID" (or programmable RFID), which requires operating system functionality, such as uploading state to a (possibly un-trusted) RFID reader [18-19].

"The Wireless Resonant Energy Link (WREL) project aims to deliver tens of watts wirelessly."

Near Field Wireless Power Transfer

The Wireless Resonant Energy Link (WREL) project aims to deliver tens of watts wirelessly. The technology builds on the coupled magnetic resonance technique proposed in [20]. The image in Figure 5c shows a 60W light bulb being powered by WREL. The energy transfer mechanism of WREL is induction, just like the inductive charging mechanisms that have been around for many years. What is different about the transfer mechanism in WREL, however, is the use of coupled, high Q resonators on both the transmit side and the receive side, and the use of a control system that keeps the system at peak efficiency, even as the transmit to receive distance changes [21]. Thus WREL is able to deliver large amounts of power at greater range than inductive wireless power transder systems.

Conclusion

As computing systems become more physical, their effects become broader and deeper. The more such systems are able to "meet" us in our 3D physical world, the more useful they will become. A robot can be thought of as a computing system that, in addition to sensing, can act on the physical world by moving itself and physical objects. Compared with less physical, non-robot computers, a robot's ability to act on the world broadens the scope of what it can do and the types of effects it can have. Elder care monitoring systems, for example, can help their "clients" substantially by watching them closely and summoning assistance when necessary. However, a robot that can directly provide assistance would be able to start solving people's physical problems, not just identifying them.

"A robot that can directly provide assistance would be able to start solving people's physical problems, not just identifying them."

The research presented in this article aims to enhance the physicality of computing systems, by improving their energy storage capabilities, increasing their ability to respond to physical gestures, improving their 3D perceptual capabilities, and via robotics, giving them the ability to change the state of the world.

References

[1] S.E. Hudson, Chris Harrison, Beverly L. Harrison, Anthony LaMarca. "Whack gestures: inexact and inattentive interaction with mobile devices." In *Proceedings of the 4th International Conference on Tangible, Embedded, and Embodied Interaction*. Cambridge, Massachusetts, USA, pages 109-112.

[2] Sebastian Thrun, Dieter Fox, Wolfram Burgard, and Frank Dellaert. "Robust Monte Carlo Localization for Mobile Robots." *Artificial Intelligence*, Volume 128, Issues 1-2, Pages 99-141, May 2001. ISSN 0004-3702, DOI: 10.1016/S0004-3702(01)00069-8.

[3] Please see ***http://personalrobotics.intel-research.net*** as well as ***http://www.seattle.intel-research.net/robotics/*** for information on Personal Robotics and Mobile Manipulation research at Intel.

[4] ***http://www.seattle.intel-research.net/robotics/research.html***

[5] "Microsoft Project Natal." Available at ***http://www.youtube.com/watch?v=UkSV1rXJ0pU***

[6] Joshua R. Smith, Eric Garcia, Ryan Wistort, Ganesh Krishnamoorthy. "Electric Field Imaging Pretouch for Robotic Graspers." In *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS, 2007.

[7] Ryan Wistort, Joshua R. Smith. "Electric Field Serving for Robotic Manipulation." In *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS, 2008.

[8] Brian Mayton, Louis LeGrand, Joshua R. Smith. "An Electric Field Pretouch System for Grasping and Co-Manipulation." In *Proceedings of the 2010 IEEE International Conference on Robotics and Automation*, ICRA, 2010.

[9] Brian Mayton, Louis LeGrand, Joshua R. Smith. "Robot, Feed Thyself: Plugging In to Unmodified Electrical Outlets by Sensing Emitted AC Electric Fields." In *Proceedings of the 2010 IEEE International Conference on Robotics and Automation*, ICRA 2010.

[10] Siddhartha Srinivasa, David Ferguson, Casey Helfrich, Dmitry Berenson, Alvaro Collet Romea, Rosen Diankov, Garratt Gallagher, Geoffrey Hollinger, James Kuffner, and J Michael Vandeweghe. "HERB: a home exploring robotic butler." *Autonomous Robots*, Volume 28, No. 1, January, pages 5-20, 2010.

[11] Alanson P. Sample, Joshua R. Smith. "Experimental Results with two Wireless Power Transfer Systems." *IEEE Radio and Wireless Symposium*, RawCon, 2009.

[12] Alanson P. Sample, Daniel J. Yeager, Pauline S. Powledge, Alexander V. Mamishev, Joshua R. Smith. "Design of an RFID-Based Battery-Free Programmable Sensing Platform." *IEEE Transactions on Instrumentation and Measurement*, Volume 57, No. 11, pages 2608-2615, November 2008.

[13] Joshua R. Smith (Ed.). *Wirelessly Powered Sensor Networks and Computational RFID*. Springer, 2010. ISBN 978-1-4419-6165-5.

[14] Yeager, D.J.; Holleman, J.; Prasad, R.; Smith, J.R.; Otis, B.P. "NeuralWISP: A Wirelessly Powered Neural Interface With 1-m Range." *IEEE Transactions on Biomedical Circuits and Systems*, Volume 3, no. 6, pages 379-387, December 2009.

[15] Marco Cordelli, Roberto Habel, Agnese Martini, Luciano Trasatti. "PORFIDO: Oceanographic data for neutrino telescopes." *Instituto Nazionale di Fisica Nucleare*, Preprint LNF–10/02 (P). Available at ***http://www.lnf.infn.it/sis/preprint/pdf/getfile.php?filename=LNF-10-2(P).pdf***

[16] Michael Buettner, Richa Prasad, Matthai Philipose, David Wetherall. "Recognizing Daily Activities with RFID-Based Sensors." In *Proceedings of UbiComp*, 2009.

[17] Hee-Jin Chae, Daniel J. Yeager, Joshua R. Smith, and Kevin Fu. "Maximalist cryptography and computation on the WISP UHF RFID tag." In *Proceedings of the Conference on RFID Security*, July 2007.

[18] Benjamin Ransford, Shane Clark, Mastooreh Salajegheh, and Kevin Fu. "Getting things done on computational RFIDs with energy-aware checkpointing and voltage-aware scheduling." In *USENIX Workshop on Power Aware Computing and Systems* (HotPower), December 2008.

[19] Mastooreh Salajegheh, Shane Clark, Benjamin Ransford, Kevin Fu, and Ari Juels. "CCCP: Secure remote storage for computational RFIDs." In *Proceedings of the 18th USENIX Security Symposium*, Montreal, Canada, August 2009.

[20] André Kurs, Aristeidis Karalis, Robert Moffatt, J. D. Joannopoulos, Peter Fisher, Marin Soljacic. "Wireless Power Transfer via Strongly Coupled Magnetic Resonances." *Science*, Volume 317, no. 5834, pages 83 - 86. DOI: 10.1126/science.1143254.

[21] Alanson P. Sample, David Meyer, Joshua R. Smith. "Analysis, Experimental Results, and Range Adaptation of Magnetically Coupled Resonators for Wireless Power Transfer." *IEEE Transactions on Industrial Electronics*, to appear.

Acknowledgments

Joshua R. Smith thanks Emily Cooper, Alanson Sample, and Lenka Jelinek of the WREL project, and Louis LeGrand, Brian Mayton, and Dieter Fox of the robotics project. Xiaofeng Ren thanks Dieter Fox, Michael Krainin, and Peter Henry.

Author Biographies

Joshua R. Smith leads Intel's Physicality research effort. He is a Principal Engineer at Intel Labs, Seattle, where he leads research in wireless power and robotics. His research interests include all aspects of sensor systems, including the development of novel sensor systems; power, communication, signal processing, and security issues in sensor systems; and algorithms and applications that use data from novel sensor systems. His research has application in the fields of robotics, ubiquitous computing, and human-computer interaction. Previously he co-invented an electric field-based passenger sensing system that is used to suppress unsafe airbag deployment in all Honda cars. He holds affiliate faculty appointments in the departments of Electrical Engineering and Computer Science and Engineering at the University of Washington. He holds PhD and SM degrees from the MIT Media Lab, an MA degree in Physics from the University of Cambridge, and BA degrees in Computer Science and Philosophy from Williams College.

Beverly L. Harrison joined Intel Research, Seattle in 2005 as a Senior Scientist specializing in Human Computer Interaction and User Experience. She also holds two affiliate faculty appointments as Associate Professor at the University of Washington in the Computer Science and Engineering Department and in the Information School. She holds BA degrees in Mathematics and Computer Science, from the University of Waterloo and MASc and PhD degrees in Human Factors Engineering, from the University of Toronto, Canada. She has worked in industrial research labs, including Nortel, Alias/SGI, Xerox PARC, IBM Research, for over 15 years, and most recently at Intel Research. From 1998-2000, she worked at a successful startup company in the e-book space, SoftBook Press/Gemstar International, as the Director of User Experience. She has numerous publications, holds over 25 patents, and serves on a number of HCI-related conference committees. Her research interests include the design and evaluation of novel mobile and/or sensor-based technologies for ubiquitous computing applications. Most recently, Beverly has been focusing on applications for wearable, sensor-based systems that embed machine learning and statistical models of human behavior and context-aware user interfaces.

Xiaofeng Ren joined Intel Labs, Seattle in 2008. From 2006-2008, he worked as a Research Assistant at the Toyota Technological Institute at Chicago (TTI-C). He received his PhD from UC Berkeley, under the supervision of Jitendra Malik. His research is in the areas of computer vision and artificial intelligence. His interests are mainly in mid-level vision and its integration with both low-level image signals and high-level object knowledge. He has worked on contour grouping, image segmentation, figure/ground labeling, finding people in images as well as tracking and segmenting people in video.

Siddhartha Srinivasa is a Senior Research Scientist with Intel Labs, Pittsburgh, and holds an Adjunct Faculty position at the Robotics Institute at Carnegie Mellon University. He joined Intel in 2006 and started the Personal Robotics project, in which an anthropomorphic robotic arm and a mobile robot coordinate to accomplish useful manipulation tasks in populated indoor environments. His research focuses on enabling robots to interact faster, better, and smoother with the real world. He received his PhD degree from the Robotics Institute at Carnegie Mellon University in 2005 where he developed robust controllers for robotic manipulation. He also has a B Tech degree in Mechanical Engineering from the Indian Institute of Technology, Madras.

Copyright

ENSEMBLE COMPUTING: OPPORTUNITIES AND CHALLENGES

Contributors

Roy Want
Intel Corporation

Eve Schooler
Intel Corporation

Lenka Jelinek
Intel Corporation

Jaeyeon Jung
Intel Corporation

Dan Dahle
Intel Corporation

Uttam Sengupta
Intel Corporation

Index Words

Ensembles
Mobile Computing
Distributed Systems
Context-aware
Wireless Networking

Abstract

Since the advent of the first practical mobile computers more than 20 years ago, advances in the underlying technology have moved out of research and development into both the enterprise and consumer electronics. Today, mobile networked computers are ubiquitous. Trends that have enabled this revolution include improved processor performance, exponentially increasing storage density, and high-bandwidth standardized radios: using technologies that consume less power and are packaged in smaller form-factors than their predecessors. However, the formation of mobile collections of devices, or ensembles, can further build on the network effect to amplify an ensemble's usefulness, a property first observed with telephony, and later the Internet. Ensembles can be created by using now commonly available wireless and wired standards, providing an aggregate value to a user that is greater than the sum of the component parts. In this article, we examine the opportunities for ensemble computing; classify the main usages and application domains; review enabling component technologies, and point out the research challenges that need to be addressed to realize the full potential of the ensemble concept.

Introduction

The history of modern computing can be divided into three distinct eras, each characterized by the number of computers available to an average user. During the first era, dominated by the mainframe (1960-1980), there were typically many people using a single computer. In the Personal Computer (PC) era (1980-2000), most people had their own PC. Today, we are in the third era (2000 onwards) in which it is now common for people to own and use many computers, sometimes referred to as the era of *ubiquitous computing* [1].

In addition to the trend of increasing numbers of computing devices per person per year, we have seen improved processor performance (driven by Moore's Law) and greater network connectivity for mobile computers that have been made possible through innovations in standardized wireless technologies. These standards provide high-bandwidth data-transmission, allowing seamless coordination between applications, and they enable fully distributed and replicated programming models. By a measure called the network effect, or Metcalfe's Law [2], N devices connected in this way create a total network value proportional to the number of connections they can make multiplied by the number of devices participating; this results in the network value growing at $O(N^2)$. Orchestrated collections of computers can thus provide greater aggregate value to users than would be expected from a simple sum of the component parts.

Although distributed programming has been the subject of research for some time with its foundation in networks of traditional servers or desktop computers, mobile computing brings dynamically changing topologies, context-awareness, and more diverse collections of computing resources, adding a new dimension to this field. We call dynamically coordinated collections of computers, which include both mobile and infrastructure components, *Device Ensembles* [3], and the techniques for programming and orchestrating their applications, *Ensemble Computing*.

> *"We call dynamically coordinated collections of computers, which include both mobile and infrastructure components,* Device Ensembles.*"*

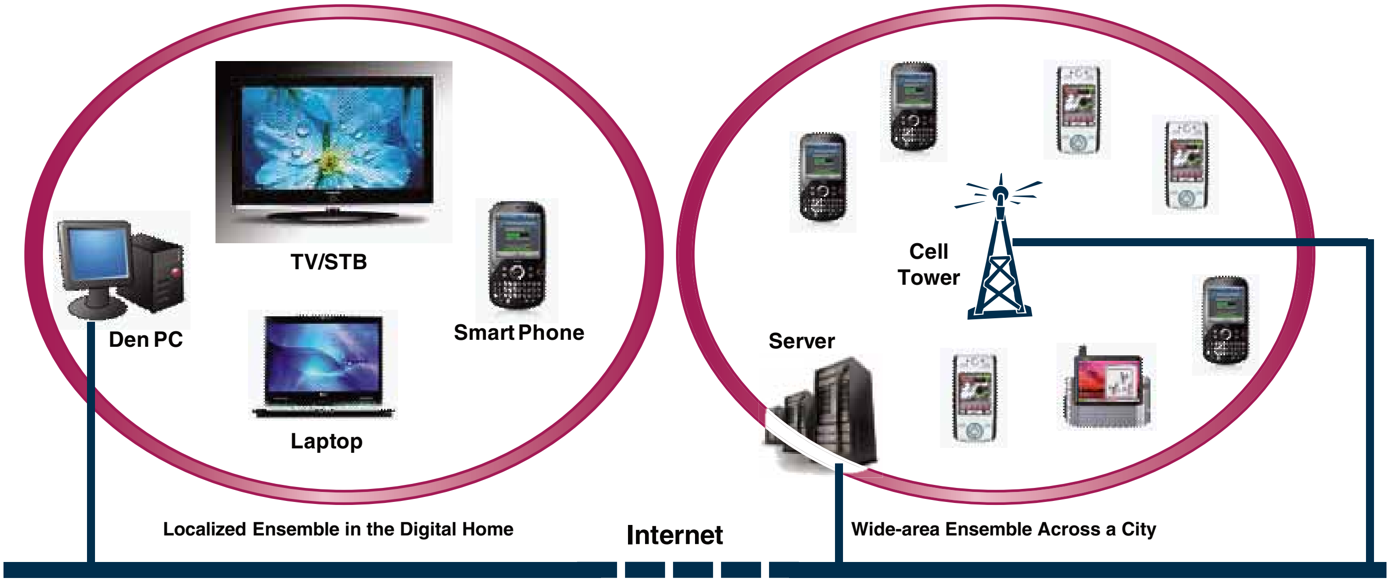


Figure 1: Ensembles of Devices a) Localized: Based on Wireless Local Area Networks (WLANs), b) Wide-area: Based on Wireless Wide-area Networks (WWANs), Properties: Dynamic Configuration/Heterogeneous/Context-aware
Source: Intel Corporation, 2010

Ensembles can be created for different scales of use, primarily determined by geographic separation. In close proximity (localized ensembles—Figure 1a), a computer can be enhanced by dynamically adding new wireless peripherals, e.g., a wireless headset. Alternatively, computers can mutually share their own system resources such as displays, networks, disks and peripherals. Processor sharing between devices is also possible through the migration of applications encapsulated in a virtual machine image, migrating them between low- and high-performance computers. Sharing resources in this way effectively builds logical computers by wirelessly piecing together new components on the fly, thus enabling a user to overcome the design limitations of their own computer by using whatever can be found nearby [4]. In localized ensembles, proximity is a prerequisite for components that support a user interface (e.g., display, camera, keyboard, mouse), as it only makes sense to share these components when they are physically accessible to a user.

> *"Sharing resources in this way effectively builds logical computers by wirelessly piecing together new components on the fly."*

In contrast, ensemble applications based on highly distributed sets of heterogeneous devices may take advantage of geographic diversity (wide-area ensembles—Figure 1b), while still dynamically adapting their membership and taking advantage of their context of use. This information can be used to create ensemble-wide inferences based on their physical separation that would have been impossible from a localized viewpoint. For example, a smart-phone vehicular traffic-monitoring application [5] can be designed so that each smart phone enrolled in the task is programmed to periodically relay information about its location to a server, which in turn builds up a real-time view of traffic flow across all the city highways, and thus can highlight areas of congestion. The traffic-service relays this information back to each of the smart phones running a traffic-monitoring client, thus enabling users to make decisions about their preferred driving routes. The resulting application has greater utility than would be possible with a set of contextually isolated devices originally designed for the primary task of making phone calls.

"In the case of WLANs, unlike wired networks, there is no connectivity between any of the peer layers in the protocol stack until a layer-2 discovery process has taken place."

Localized and wide-area ensemble applications of the types described here exist today, but there is no consistent programming strategy or architecture that can be used to capitalize on lessons learned from their design and implementation. This leaves engineers to reinvent the wheel across many of the application domains. Furthermore, these lessons are rarely captured in the tool-chains used to implement them, and therefore developers do not progressively build better tools for each generation of system design.

In the remainder of this article we examine background and related work, enabling technologies and opportunities for ensemble computing, and we discuss the challenges that need to be addressed.

Background

Distributed programming has been a subject of research and development since the 1970s with many well-understood programming models developed and in common use today.

Although wireless ensembles can make use of many proven distributed-computing abstractions such as sockets, remote procedure call (RPC), message passing, streaming, and client/server models, ensembles are unique in several important ways. In the case of WLANs, unlike wired networks, there is no connectivity between any of the peer layers in the protocol stack until a layer-2 discovery process has taken place, and until a mobile device is instructed to connect (manually or automatically). In the case of WWANs, devices are always connected, but peer-to-peer messaging and discovery protocols are restricted by the rules of a particular service provider, limiting the types of protocols in use. However, in either type of wireless network, ensembles of devices can augment the discovery process further by including metadata that can be derived from their current context (e.g., location, or physical sensor data), and device capabilities (e.g., supported communication protocols). Based on these metadata, decisions can be made about the device groupings that can form. Context information can also be derived from user input, automatically sensed, or inferred from the behavior of applications. Shared metadata of this type can define and dynamically modify connections as the context changes. We now look in more detail at some of the solutions that give rise to ensembles in both the local and wide-area environment.

"In the case of WWANs, devices are always connected but peer-to-peer messaging and discovery protocols are restricted by the rules of a particular service provider, limiting the types of protocols in use."

Localized Ensembles

Radio technologies that are designed for short-range operation implicitly define a localized ensemble of devices by the nature of the proximate connections they can establish. The first digital radio standard to achieve this was Bluetooth*, providing an effective replacement for wired peripherals (up to 10 m). Bluetooth's most popular application has been to connect a cell phone to a wireless headset, but the protocol can in fact link up to eight slave peripherals to a computer at one time [6]. Bluetooth application examples include on-body, in-room, and in-car networks supporting an ensemble of devices in those contexts.

Short-range radio systems have recently enabled new opportunities for patient care. Monitoring of chronically ill patients while out of a hospital, and hence mobile (ambulatory monitoring), has been a goal of physicians for many years. In recent years, inexpensive Body Sensor Networks (BSNs) have become viable. This is because the power consumption of many microcontrollers has been reduced to a point where it is possible to a point where it is possible to process and communicate the results within a power envelope that can be sustained by a small battery-powered device for several days [7]. BSNs can wirelessly connect a heterogeneous set of sensors through a Body Area Network (BAN) [8] to a central processor that can store, or wirelessly offload, the results to a hospital, for example, through a secondary radio, such as WiFi or a General Packet Radio Service (GPRS) [9]. However, special care needs to be taken when designing BAN antennas, because Bluetooth and WiFi operate in the 2.4 GHz band, and transmissions are readily absorbed by water, and therefore wireless signals cannot travel through the body without being severely attenuated. To support this application area, sensors, including EKGs, pulse-oximeters, and blood-pressure monitors, are now also readily available as commercial products and can be integrated into a BAN.

"Bluetooth application examples include on-body, in-room, and in-car networks supporting an ensemble of devices in those contexts."

"An ensemble, based on body-worn devices, such as a watch or a cell phone, in combination with environmental monitors in the home, can provide caregivers with the data they need."

Another use of BSNs, other than monitoring the body's vital signs, is to monitor physical activity. In recent years, there has been considerable interest in using technology to allow a larger fraction of the elderly population to remain in their homes as long as possible without needing to resort to institutional care, often referred to as aging-in-place. The problem is that aging may be accompanied by a slow mental decline, and it is often not clear to family and friends when an elderly person becomes a danger to him or herself [10]. An ensemble, based on body-worn devices, and embedded in day-to-day accessories such as a watch or a cell phone, in combination with environmental monitors in the home, can provide caregivers with the data they need. For example, when daily activity patterns change and become increasingly sedentary or erratic, these devices can provide vital clues that a person is no longer behaving in a healthful way, and that she needs further evaluation. The value proposition for the elderly is that in most cases they readily want to be able to stay in their own home as long as they are able, and a BSN provides caregivers, the physicians, and the elderly with the information needed to make this possible.

The WiFi radio standard and its various generations (b/a/g/n) also support a network of proximate wireless devices, but these standards are typically centered on an infrastructure node or access point, rather than using ad hoc communication. Although these networks can be bridged to build a wireless network of arbitrary size, a common configuration makes use of a single access point serving one location up to a radius of 100 m. For example, in-home networks typically only need one access point, and all devices that discover and connect to it can then communicate with each other. The "digital home" is another example of where a coordinated ensemble of devices can better serve a user than a set of isolated devices. By coordinating the access and playback of digital media captured and consumed on a wide variety of consumer devices, a user can see or listen to their media on a device that is most convenient, or has the best physical capabilities (e.g., size, resolution, and fidelity).

Smart spaces, or *smart rooms*, are a fertile area for experimentation that make use of the tenets of ensemble computing. The essence of a smart space is one in which local infrastructure is readily made available as a service to users and their mobile computing devices. Usually, this is in the form of support for multimedia capabilities to enhance meetings and presentations, but may extend to highly-specialized capabilities, depending on the organization that developed it. Many of the concepts that underlie smart spaces are exemplified in the EasyLiving project at Microsoft [11], and the Stanford iRoom project [12]. For example, the iRoom supports multiple, wall-sized screens that can be linked together into one contiguous display, and at the same time, is wirelessly integrated with laptops and other heterogeneous mobile devices brought into the space. The system can also be commanded through tangible objects designed to be manipulated, or gestured with, to provide intuitive user controls. This smart space demonstrated effective use of the ensemble computing concept in support of group collaboration. The EasyLiving project had similar properties, but also made use of cameras and image processing to infer activities going on in the smart space.

"The essence of a smart space is one in which local infrastructure is readily made available as a service to users and their mobile computing devices."

Wide-Area Ensembles

Ensemble computing has its origin in Grid [13] and Cloud Computing frameworks [14]. These techniques can be used to support ensemble applications, and it is likely that new programming paradigms will be discovered that further enhance the value of ensembles in the future. However, many ensemble applications can also be built by using standard distributed computing techniques that are further augmented to take advantage of mobile properties such as context.

"Ensemble computing has its origin in grid and cloud computing frameworks."

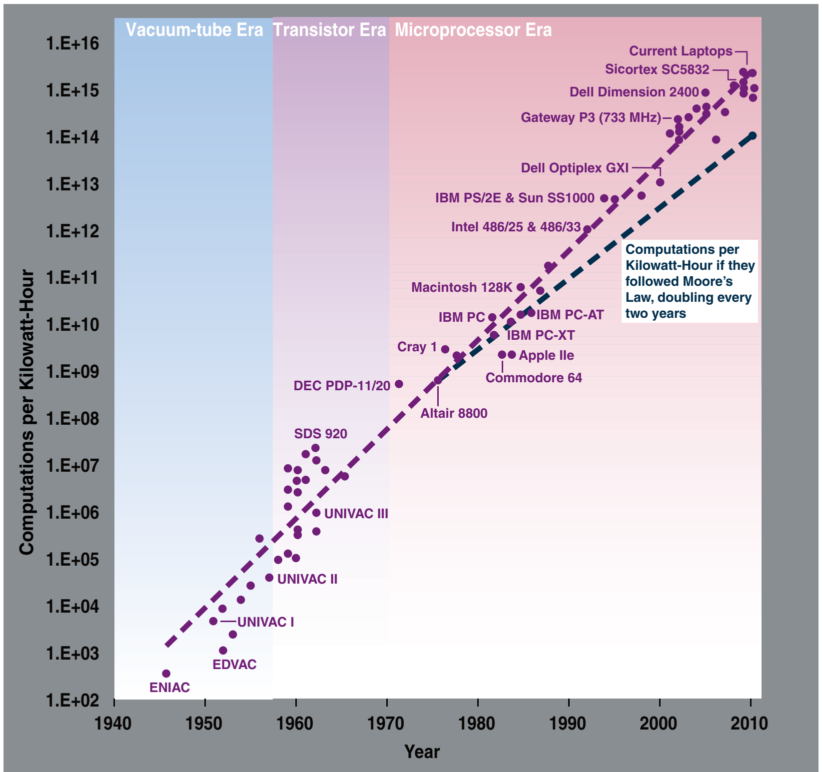
Social networking for mobile devices is an example where distributed computing merges with mobile and context-aware computing and clearly defines value for device ensembles. Consider some of the latest location-based computing applications, such as *Loopt** or *Overhere** that run on an iPhone*, and use cell-tower-based (or GPS-based) position finding. In these systems, a group of friends decide to share their location information, by allowing their devices to be tracked. Effectively, an ensemble of devices has been declared that is a proxy for group members. The resulting system lets you know if your friends are nearby so that, for example, you can serendipitously meet up over coffee or share a meal. In the case of Loopt, you can also periodically keep track of your friends' status through secondary services such as Twitter* notifications.

Sensor Networks are another example of wide-area ensemble computing in which the underlying goal is for an ensemble of devices to work together to share environmental data in order to make sense of it, and understand the larger scale trends [15]. Examples include micro-weather monitoring in order to make better weather forecasts, seismic activity monitoring to understand the effects of city-wide earthquake damage on buildings and structures, and wide-area pollution monitoring to advise on travel. Many sensor networks are composed of highly customized nodes (often referred to as motes) that sense and relay information across a wireless mesh of devices towards a collection point, or sink. From there it is routed to more capable processors that can store and analyze the data, and in turn infer the dominant trends. Through network servers the results can be redistributed to users running client programs, thereby allowing them to take appropriate action. An example of a sensor network in action is the Common Sense project [16] in Intel Labs that provides users with a device to monitor atmospheric pollution based on carbon monoxide, nitrogen oxides, and ozone. The entire system comprises both a local ensemble and a wide-area ensemble. Locally, pollution measurements are combined with location information from the monitor hardware via Bluetooth to a cell phone. Then by using GPRS, the cell phone periodically transmits these data over the wide-area network to a server where the information is aggregated; thereby, providing a service that supports all of the participating phones. In this way, users can be advised about travel options, and the airborne pollution levels currently contaminating particular routes. Such a system becomes valuable when a larger number of devices are participating and working across a wide area.

"Many sensor networks are composed of highly customized nodes (often referred to as motes) that sense and relay information across a wireless mesh of devices towards a collection point."

Enabling Technology and Trends

Ensembles have been enabled by several technology trends, each advancing the capabilities of mobile devices. The key advances are increasing processor performance (Figure 2a) at lower power, increasing portable memory density (Figure 2b), and increasing wireless networking bandwidth (Figure 2c), and all at smaller form factors for each new generation of device.



The complete study is available from the Intel Web site at http://download.intel.com/pressroom/pdf/computertrendsrelease.pdf.

Figure 2a: Trends in Processor Instructions per kWhr by Year
Source: Intel Corporation, 2010

Figure 2a shows the dramatic progress that has been achieved in increasing processor performance in terms of ever increasing numbers of instructions that can be performed with the same amount of energy. The implication is that low-power processors in a mobile device can do more each year given the amount of energy that can be stored in a small portable battery—a technology whose stored energy density is improving at a far lower rate.

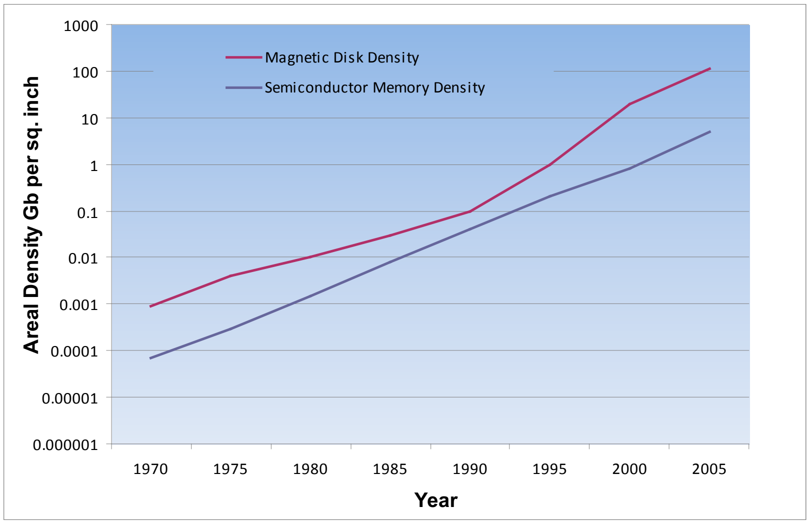


Figure 2b: Trends in Magnetic and Semiconductor Storage Density by Year
Source: Intel Corporation, 2010

Further, the yearly rate of increase in storage density for magnetic media, a rate which has been out-pacing Moore's Law, has been approximately doubling every year (Figure 2b), and this increase has also been a key enabler for ensembles. As a result, small mobile devices are able to store the gigabytes of data required by modern operating systems and their applications, and at an affordable price point. Today, many mobile computers are as capable as some of the most powerful desktop computers in use ten years ago while also running comparable operating systems.

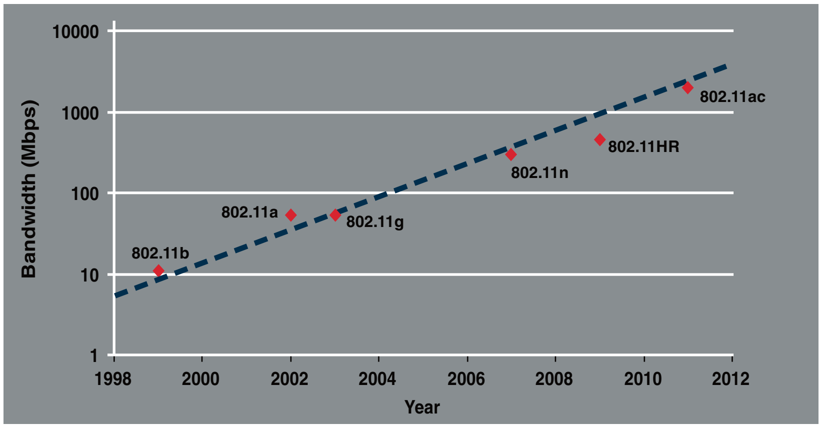


Figure 2c: Trends in Wireless Bandwidth for WLANS by Year
Source: Intel Corporation, 2010

Finally, inter-device communication, an essential component of ensemble computing, has also improved dramatically. It was only in 1999 that the first draft of the Bluetooth standard and the first implementations of the IEEE 802.11b WiFi [17] (11 Mbps) standard were established. Today, 802.11HR approaches 450 Mbps, with further extensions planned (Figure 2c). Wide-area, wireless-data transport services such as GPRS, and later 3G [18], also became available during this time. Each technology supports communication at different geographic scales and bandwidth and can be classified into various types of networks now commonly available to support ensemble computing:

- Personal Area Networks (PANs) connecting computers and peripheral components over a short range (10 m) e.g., Bluetooth 1.0 (1 Mbps).
- WLANs connecting computers to local infrastructure and other nearby computers (100 m) e.g., WiFi/802.11g (54 Mbps).
- WWANs connecting mobile computers to cell towers networked across a city, e.g., GPRS (2000 m) 20/80kbps.

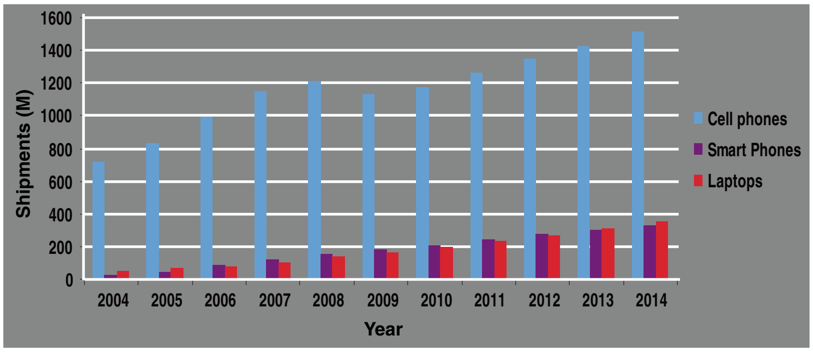


Figure 3: Shipments of Mobile Devices by Year: Actual 2004-2009 and Forecast 2010-2014
Source: Intel Corporation, 2010

Furthermore, the value proposition for mobile users has continued to increase; i.e., more capability at lower cost, resulting in the spectacular increase in market size for cell phones, smart phones, and laptop computers that we have witnessed during the last five years. Moreover, the trend is still forecasted to continue despite the recent recession in 2009 (Figure 3). The large-scale proliferation of these devices has become a key enabler for a wide variety of ensemble applications.

Opportunities

In this section we consider three major opportunities for realizing greater value for ensemble solutions in the following areas: convertible resources, super-charging performance, and added value from new emergent behaviors.

Convertible Resources

Although short-range wireless standards provide a core capability to enable localized ensembles, a higher-level construct is needed to take advantage of them. The Dynamic Composable Computing (DCC) project at Intel [4] is one model for overcoming the resource limitations of a computer, especially a small mobile device, augmenting its capabilities by utilizing the resources of more capable devices nearby. The model for DCC is to abstract away the underlying devices by using a client/server service model enabling a remote peripheral to appear as if local. There are several industry solutions to enable network access to remote displays, such as Virtual Network Computing (VNC) [19], storage [20], and USB peripherals [21], but none that coordinate all these services into a unified system. The DCC composition framework utilizes many existing solutions for a category of peripherals, advertising their presence, and coordinating operations for the user via an intuitive GUI. The opportunity for composable systems is for every computer to be able to convert itself into a much more capable device by using the best of the available resources nearby.

"The DCC composition framework utilizes many existing solutions for a category of peripherals, advertising their presence, and coordinating operations for the user via an intuitive GUI."

Super-Charging Performance

It is also possible to use high-performance computers to augment the processing capabilities of nearby small mobile devices, thus going beyond sharing system peripherals, to sharing a processor among an ensemble of devices. However, unlike peripheral sharing mechanisms such as VNC, super-charging processing has not been widely explored to date in support of mobile computing.

One of the most promising approaches to enable this capability is in the realm of virtual machine (VM) migration. It has only recently been practical to run a VM on a handheld mobile computer with the advent of low-power mobile x86 processors (i.e., the Intel® Atom™ family of processors). Mobile Internet Devices (MIDs) and Netbooks, both based on the Intel Atom processor, also run standard operating systems based on Windows* and Linux*. The next generation of Intel Atom processors will target smart phones, run standard operating systems, and use x86 applications without modification. The resulting commonality of the instruction-set allows the same code to run on mobile computers designed around the Intel Atom processor, as it does on most desktop computers, and thus standard VM hypervisors, and the associated VM migration techniques, can be employed to support the migration of computation between the two.

"The next generation of Intel Atom processors will target smart phones, run standard operating systems, and use x86 applications without modification."

Consider the benefits of VM migration in an everyday scenario (Figure 4) in which a mobile professional, Jane, wishes to maintain a consistent computing experience as her computing activities flow between home, office, and mobile working. While in her home office, Jane was sending e-mail, surfing the Web, and referring to documents about a product she wanted to research. When mobile, she would like to continue these tasks on her smart phone even while waiting for a friend in a parking lot. Later she finds the product in a store and takes some pictures with her mobile so she can study it more before buying it. As is often the case with smart phones, the camera has poor resolution, and so she employs a super-resolution [22] application in which several pictures are taken in quick succession, and then they are combined to create a higher resolution picture (a compute- and power-intensive activity).

This scenario can be implemented by running Jane's entire office PC's computational state in a VM that can be suspended, and then transferred to her smart phone where it is resumed. While mobile, her e-mail and documents can be accessed as before, and the mobile phone's camera can be used to take photos. When returning home, or at her workplace, her smart phone can suspend its VM again and transfer it back to a desktop PC where it can resume and execute at higher performance. In this example, VM migration enables effective mobility by providing session transfer and by super-charging processor performance for CPU- and power-intensive operations. By migrating code from a smart phone to a high-performance desktop computer, it is possible to reduce execution time, while also saving power on a battery-constrained device.

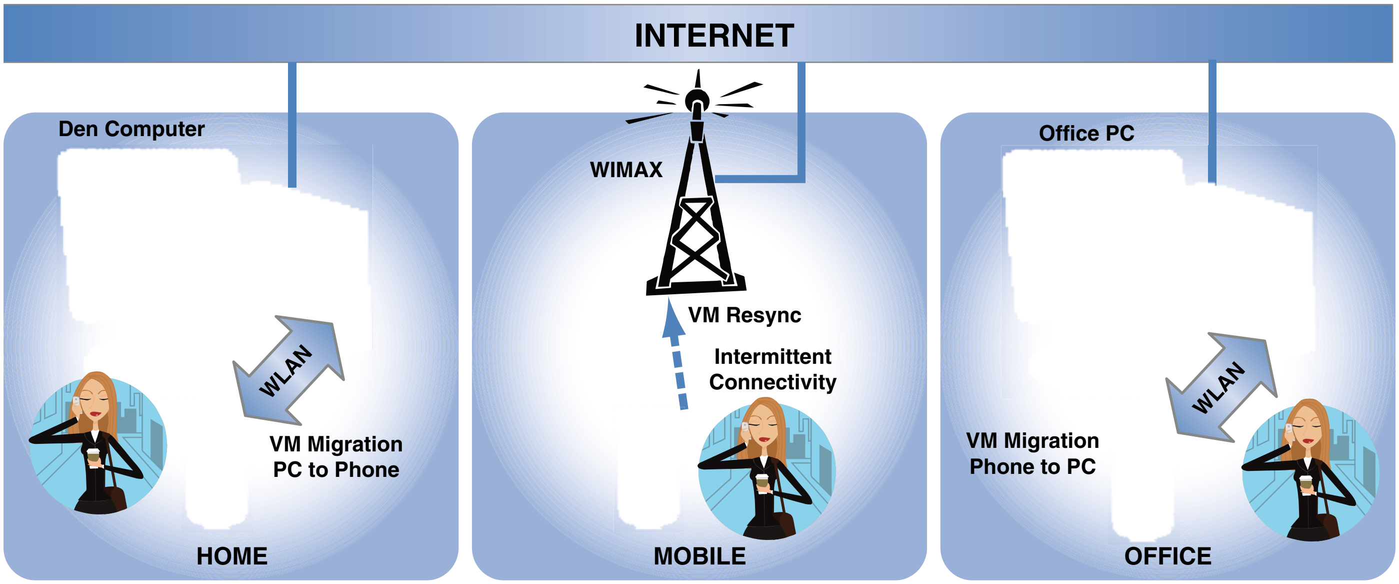


Figure 4: A Smart Phone Migrating Processing Between Home, Office PC, and Mobile Use, Encapsulated in a Virtual Machine
Source: Intel Corporation, 2010

Smart phones, based on the Intel Atom processor, make this mobile use model a reality. In a collaboration between the University of Toronto and Carnegie Mellon University (CMU), the Horatio system [23] has been developed expanding the earlier work on Internet Suspend Resume (ISR) [24] to include migration to a smart phone. This system demonstrates the opportunity for using either a network server, or a mobile device, to transfer a VM image to a secondary PC. However, the first version of the Horatio system used Advanced Risk Machines (ARM) based smart phones and therefore could not resume the VM image while mobile. In a few years, x86 smart phones will become more common, thus enabling the more complete use model.

In addition to local infrastructure, cloud computing [14] offers another opportunity to super-charge the execution performance of a mobile device. In this case computation can be moved from a mobile to a remote high-performance computing (HPC) center. For WWANs or other long-haul networks, VM migration is not a practical option as the large image size (4-20GB) results in transfer times that would likely take too long. On the other hand, ensembles that are designed for specialized tasks may use HPC services to process specific data types. These might include media, documents, and context sent to services in the cloud for processing, e.g., photos sent to the cloud for image identification, or documents translated form one language to another. In both cases these services are improved by utilizing contextual knowledge about the media's creation, which can be provided by the ensemble.

"For WWANs or other long-haul networks, VM migration is not a practical option as the large image size (4-20GB) results in transfer times that would likely take too long."

Added Value from Emergent Behaviors

Although the number of compute-capable devices in use is predicted to reach 15 billion by 2015 [25], and the number of devices connected to the Internet will more than double from 2010 to 2013 [26], the sheer availability of devices to communicate does not equate with their intelligent coordination. The opportunity to harness the collective resources of multiple devices has been maturing for some time, yet the opportunity to harness the collective *wisdom* of groups of devices awaits us.

Increasingly, devices find themselves in a sea of other devices, each of which has the ability to measure its own behaviors, to sense the environment or context in which it is operating, and to share its findings with other devices. As a consequence of gossiping data, not only are devices able to build local approximate images of global system state, but also to exhibit global properties that emerge from local properties. In these emergent collaborative systems, metrics have been shown to improve with the size of the group. In the security realm, ensembles of end-hosts collaboratively defend the enterprise network, offering improved accuracy and sensitivity of malware detection [27]. In the wireless world, ensembles have been used to great effect to mitigate wireless interference [28]. Additionally, in the arena of neighborhood-level networking, distributed trust models have demonstrated the simplification of self-configuration [29].

"Increasingly, devices find themselves in a sea of other devices, each of which has the ability to measure its own behaviors, to sense the environment or context in which it is operating, and to share its findings with other devices."

Challenges

Despite the many attractive opportunities, there are a variety of challenges when applying ensemble concepts to practical problems. Here we consider difficulties in manageability, new programming models, wireless network limitations, usability, power, and security and privacy.

Manageability Challenges

One challenge often faced by wide-area ensembles is the stringent requirement for homogeneity, i.e., that all machines agree on the same software, protocols, formats, and versioning. Thus, machine federations are often managed by one owner, who isolates the solutions from one another and can thwart efforts to support a more broadly generalized service [30]. However, proprietary solutions often reduce security and privacy concerns that arise when device federations span multiple owners and/or multiple administrative domains, even though they may not use the most effective solutions.

With the proliferation of mobile and embedded devices, there is an opportunity to revisit how we manage them. In the absence of an IT department, computers are at liberty to bootstrap services from nearby [4, 31] or like-minded devices [32, 33, 34], and/or from reputable peers [35], rather than from preconfigured servers or proxies. In other words, ensembles can take advantage of peer-to-peer mechanisms in a context where no infrastructure exists. However, ensembles also underscore the challenges inherent in establishing device trustworthiness [36], and the challenge of preserving a device's identity along with its data (see later section on Security and Privacy). For wide acceptance, these issues need to be resolved, and the level of sharing needs to be adapted to the context before enterprise IT departments will allow their mobile computers to participate in generalized ensemble applications.

New Programming Models

Platform-independent programming such as Java*, Flash*, and Silverlight* can be adapted to ensemble programming in which applications run across devices with different operating systems and heterogeneous hardware. However, ensemble computing is inherently distributed across multiple devices (unlike the traditional client-server model), and it should dynamically adapt to the presence of nearby resources. This environment, unlike the static configuration of wired computer networks, extends existing platform-independent programming models that currently may not suffice to take advantage of all the potential benefits of an ensemble.

"Ensemble computing is inherently distributed across multiple devices and it should dynamically adapt to the presence of nearby resources."

We envision that requirements for new programming models to support ensemble systems will bear similarities to those for data-parallel technologies, such as Globus* [37] or MapReduce* [38] with additional support for running algorithms efficiently on an unreliable, heterogeneous collection of mobile devices. For instance, when a failure occurs e.g., a participating device suddenly leaves the area, the departing code should be re-executed on a device that is still present.

Second, context-aware programming models [39, 40] can also be useful to write ensemble applications. These programming models allow developers to define tuples that connect a current context, derived from sensor data, to a corresponding action that allows programs to naturally adapt to their environment, and to the dynamically-changing membership of an ensemble. A friend-finder application is a good example as it uses sensed location-data to initiate an action that notifies users that are near one another in a city.

Wireless Limitations

While wireless connectivity provides great benefits in convenience and usability, it also presents significant challenges in reliability and form factor.

Reliable Connectivity

The reliability of a connection between ensemble devices is impacted by a number of factors in the ensemble's environment. Physical objects can block the wireless signal between members of an ensemble and drop the connection, much like a dropped cellular call. When ensembles include mobile nodes, reliable communication can be problematic. Also, radio frequency interference (RFI) may be caused by other wireless devices within range of the ensemble. The household microwave oven is a typical example, contributing interference in the license-free 2.4 GHz band, in which most home and consumer wireless devices around the world operate. For wide-area ensembles there are a larger number of other devices to cooperate with (and interfere with) than for local-area ensembles, and this situation can also lead to potential security risks (see Section on Security and Privacy). Furthermore, large distances between ensemble devices can affect the quality and reliability of connection. Devices that connect near the limits of their range must process weaker signals that are more difficult to decode reliably. Research on Delay Tolerant Networking [41] suggests techniques to mitigate communication link failure and boost reliability in these situations.

"Achieving greater range typically requires a more powerful transmitter that translates to larger batteries, bigger antennas, and a larger physical device."

Another approach for increasing the reliability of wireless communication is to design devices, or infrastructural communication hubs, with greater operational range. However, achieving greater range typically requires a more powerful transmitter that translates to larger batteries, bigger antennas, and a larger physical device. These factors are a challenge when designing mobile devices, resulting in a compromise of utility versus form factor.

Form Factor

By definition, given the dynamic nature of ensembles and group formation, a significant number of participating devices are likely to be highly mobile, with small user-friendly form factors. These ultra-mobile designs provide significant challenges when accommodating the antenna and power requirements of a radio subsystem that provides discovery and connectivity. High-capacity batteries that can support high-power radios, or long transmission times, incur additional device weight and size. Increasingly, mobile devices also support several communication technologies (3G, BT, WiFi) in the same package, operating at a variety of frequencies. In turn, this increases the complexity of antenna designs, also requiring shielding between each radio, and these factors contribute to larger packaging requirements for a mobile device.

Usability Issues

"Devices that transmit weak signals may frequently drop connections with other ensemble devices, thereby causing lost data and slow response times."

The broad availability and spectrum of devices that can form an ensemble, *and do so with minimal user intervention*, will dictate the usability and pervasiveness of the technology. A successful wired precedent came with the introduction of the universal serial bus (USB), which allowed users to painlessly transfer data between devices. Notably at CES 2010, manufacturers started showing concepts for wirelessly pairing mobile devices with Consumer Electronic (CE) equipment (such as TVs) to share content and control applications. We can easily imagine this notion evolving to include ensembles between (and within) automobiles, peripherals, and home appliances. Key ingredients for making an ensemble computing paradigm work well are minimal but consistent user-interfaces (UI), preference-based configuration and intelligent set-up, and context-based adaptation and connection. Further, in an optimal world, these would all require minimal user intervention to discover, connect, and share resources among participating computers.

Some of the primary advantages of wireless systems also create additional challenges in usage compared with wired networks. As described in the section on Wireless Limitations, devices that transmit weak signals may frequently drop connections with other ensemble devices, thereby causing lost data and slow response times, and these further limit connection quality, throughput, and reliability. These disruptions interrupt natural usage that negatively affects the user experience. Wireless device connections lack the visual physical indicators (i.e., a visible cable) of wired connections.

The extent of secure communication is also unclear, as eavesdropping on a wired network requires physical access to the cable, whereas eavesdropping on a wireless network only requires that the device be in range. Because wireless is invisible, these issues need to be addressed through connection status and configuration indicators to guide and support the usability model for an ensemble from a user perspective. User interface standards in these areas have not been developed at this time. Furthermore, if a device is connected to a wire, it can be designed to be powered by the same wire, simplifying the form factor design constraints, or when an internal battery is present, providing a mechanism for charging the device. Ensembles are less likely to benefit from this design approach.

Power Constraints

In the case of ensembles, where value is maximized by the seamless orchestration of multiple devices, power becomes critical.

We have all experienced the frustration of using a cell phone in a car and the phone battery dies when you're in the middle of getting directions. In this mobile situation we are further frustrated because we cannot easily use our hands to plug in a charger and redial. The reality today is that we increasingly rely on our mobile battery-powered devices to be continuously available and provide critical communication, content capture, I/O control, and storage. As the interdependency, and thus criticality, of these devices continues to grow, the expectation that they are powered and available increases. The solution is not simple; moreover, the problem is exacerbated by our expectations that ever-shrinking form factors lead to smaller batteries, and the desire for minimal dependency on plugging in. However, devices can adapt when they join an ensemble, for example, by turning off unnecessary duplicate peripheral capabilities already provided by peer devices across the wireless connection.

"Eavesdropping on a wireless network only requires that the device be in range."

Mobility of Tasks within Ensembles Based on Power Availability

One of the advantages of an ensemble is the opportunity for reliability through redundancy. In the dropped phone-call example just cited, there is a potential solution. By detecting the cell phone's low-power state and seamlessly rerouting the phone call (including headset re-pairing) to a VOIP [42] system running on a laptop in the car, an automatic handover could be achieved. Characterization and discovery of device capabilities in an ensemble, and the ability to cross device boundaries to take advantage of opportunities such as more available power, are key capabilities for enabling this approach. Techniques such as VM migration (described earlier) can also provide the infrastructure needed for process migration to devices that have sufficient power to complete a task.

"By detecting the cell phone's low-power state and seamlessly rerouting the phone call (including headset re-pairing) to a VOIP system running on a laptop in the car, an automatic handover could be achieved."

"To form an ensemble, mobile devices should be able to discover other ensemble-capable devices as they enter mutual communication range."

Mobility of Physical Power among Ensemble Members

An alternate approach uses power as a resource to be shared among devices. In this scenario, power is transferred to the most critical part of the ensemble. To continue with the dropped call on the cell phone example, the ensemble detects the low-power state of the cell phone, and it moves power from the user's laptop, or even the car itself, to the cell phone. Users can do this today under some conditions by using a USB cable; other technologies, such as wireless power, are under development. Wireless charging through inductive coupling or spatial (longer range) technologies will increasingly be adapted and improved for use in commercial applications. However, it is only likely between larger power-rich devices and nearby low-power mobile computers, due to the current inefficiency of free-space power-transfer.

Security and Privacy

Securing an ensemble system as a whole requires more effort than securing individually participating devices. In this section we discuss security and privacy issues that are uniquely associated with implementing and running ensemble systems.

Ensemble Formation

To form an ensemble, mobile devices should be able to discover other ensemble-capable devices as they enter mutual communication range. For example, an access point periodically broadcasts its SSID to allow other 802.11 clients to find the resource (and vice versa). While this type of beaconing is commonly used for automatic resource discovery, the discovery protocol could inadvertently expose the device's identity to attackers who snoop on broadcast traffic for the purpose of tracking the mobility patterns of each device. One approach for securing discovery protocols is to employ cryptographic algorithms to limit such risks [43, 44].

"A new approach to control the use of private sensitive data is via a trusted virtual domain accessible to the remote parties. Once an ensemble dissociates, any cached code or data in the trusted domain can also be deleted for added system security."

Once nearby devices are located, the next step is to establish trust between the devices before executing coordinated tasks. Given that many usage cases assume little or no institutional support for the ensemble infrastructure, both certificates and trust verification processes used in traditional networks would be impractical or too heavyweight to be applied to ensemble systems. However, existing solutions for mobile ad-hoc networks [45] could be adapted to establishing trust with ephemeral devices, but the heterogeneity of participating devices complicates distributed key management.

Ensemble Computation

If sensitive data are involved in an ensemble computation, or the code implementing applications needs to be protected from reverse engineering, it is possible to build an additional layer of security leveraging trusted hardware. For instance, a new approach to control the use of private sensitive data is via a trusted virtual domain accessible to the remote parties [46]. Once an ensemble dissociates, any cached code or data in the trusted domain can also be deleted for added system security.

Another challenge for ensemble system security is to develop dynamic security policies that can adapt to available power, platform resources, and the collection of devices present. However, the security of the ensemble is only as strong as the weakest member device; hence, striking the right balance between flexibility and security is important to secure the entire system.

Conclusion

Ensemble computing extends the general notion of distributed computing with the issues that arise from introducing collections of mobile devices into the mix. Key differentiators are the dynamic nature of the connections between devices, the use of location and other context information, and the heterogeneous nature of devices and their processing and system resources.

As a result, there are many opportunities that arise from the existence of ensembles that include converting (or augmenting) a computer by sharing remote wireless resources. Alternatively, processing can be super-charged by migrating computation from under-powered mobile devices to higher-performance infrastructure computers. Emergent behaviors in an ensemble may also solve problems that cannot be solved locally: some solutions will only be found by taking advantage of the shared wisdom derived from large numbers of devices.

"Many opportunities arise from the existence of ensembles that include converting (or augmenting) a computer by sharing remote wireless resources."

Ensemble systems also face many challenges: managing large numbers of devices, defining new contextual programming models, living with the limitations of wireless communication, designing intuitive user interfaces, mitigating power constraints, and all the while ensuring security and privacy. These problems are not intractable, and there are many promising solutions that will move us forward in this field.

Ensemble systems that exploit the available opportunities while finding new ways to effectively handle the challenges are an ongoing subject of research. At Intel Labs we continue to define new research projects that explore these issues. Time and time again, we find examples in which the aggregate value of an ensemble is greater than the simple sum of the component parts.

References

[1] Weiser, M. “The Computer of the 21st Century.” *Scientific American.* September No. 9, pages 66-75, 1991.

[2] Shapiro, C. and Varian H.R. 1999. “Information Rules.” *Harvard Business Press.* ISBN 087584863X. [Note: the original work was not archived: this reference cites a presentation in 1980]

[3] Schilit, B. and Sengupta, U. “Device Ensembles.” *IEEE Computer*, Volume 37, No. 12, pages 56-64, 2004.

[4] Want, T., Pering T., Sud S., Rosario B. “Dynamic Composable Computing.” *ACM HotMobile*, Silverado Resort, Napa Valley, California, 2008.

[5] Lunde, T. and Larsen, A. “KISS the Tram: Exploring the PDA as Support for Everyday Activities.” In *Proceedings of the Third International Conference on Ubiquitous Computing*, pages 232-239, 2001.

[6] Haartsen, J. C. “The Bluetooth Radio System.” *IEEE Personal Communications*, pages 28-36, 2000.

[7] Yang, G. “Body Sensor Networks.” *MA: Spring Science & Business Media.* ISBN 0-84628-272-1, 2006.

[8] Hanson, M. A., Powell Jr., H. C., Barth, T. T., Ringgenberg, K., Calhoun, B.H., Aylor, J.H., and John Lach. “Body Area Sensor Networks.” *University of Virginia, IEEE Computer*, pages 58-65, January 2009.

[9] Samjani, A. “General Packet Radio Service (GPRS).” *IEEE Potentials*, pages 12-15, April/May, 2002.

[10] Philipose, M., Fishkin, K., Perkowitz, M., Patterson, D., Fox, D., Kautz, H., and Hahnel, D. “Inferring activities from interactions with objects.” *Pervasive Computing*, 3(4), pages 50-56, 2004.

[11] Brumitt, B., Meyers, B., Krumm, J., Kern, A. and Shafer, S. ”EasyLiving: Technologies for Intelligent Environments.” *Handheld and Ubiquitous Computing (HUC)*, Bristol, UK, pages 97-119, September 2000.

[12] Fox, A., Johanson, B., Hanrahan, P. and Winograd, T. “Integrating Information Appliances into an Interactive Workspace.” *IEEE Computer Graphics and Applications*, Volume 20, No. 3, pages 54-65, May 2000.

[13] Foster, I., Kesselman, C. and Tuecke, S. “The Anatomy of the Grid: Enabling Scalable Virtual Organizations.” *International Journal Supercomputer Applications*, 15(3), 2001.

[14] “Amazon Elastic Compute Cloud.” (Amazon EC2). Available at ***http://aws.amazon.com/ec2/***

[15] Akyildiz, I. F., Su, W. and Sankarasubramaniam, Y. "A Survey on Sensor Networks." *IEEE Communications*, pages 102-114, August 2002.

[16] Dutta, P., Aoki, P. M., Kumar, N., Mainwaring, A., Myers, C., Willett, W. and Woodruff, A. "Common Sense: Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors." (demonstration). In Proceedings *SenSys'09*, Berkeley, CA, pages 349-350, November 2009.

[17] "WiFi." *IEEE 802.11*, February 2010. Available at ***http://grouper.ieee.org/groups/802/11/index.html***

[18] 3GPP. "3rd Generation Mobile System Release 4 Specifications." *TS 21.102*. Available at ***http://www.2gpp.org/ftp/Specs***

[19] Li, S., Spiteri, M., Bates, J. and Hopper A. "Capturing and Indexing Computer-based Activities with Virtual Network Computing." *ACM Symposium on Applied Computing*, Como, Italy, pages 601-603, March 2000.

[20] "Samba: An Introduction." 2010. Available at ***http://samba.org/samba/docs/SambaIntro.html***

[21] "Redirector: product from Incentivespro." Available at ***http://www.incentivespro.com/usb-redirector.html***

[22] Pelltier, S. "Acceleration Methods for Image Super-resolution." *Thesis*, Department of Electrical & Computer Engineering, McGill University Montreal, Canada, November 2009.

[23] Smaldone, S., Gilbert, B. Bila, N., Iftode, L., de Lara, E. and Satyanarayanan, M. "Leveraging smart phones to reduce mobility footprints." In *Proceedings ACM MobiSys'09*, Krakow, Poland, pages 109-122, June 2009.

[24] Satyanarayanan, M., Gilbert, B., Toups, M., Tolia, N., Surie, A., O'Hallaron, D. R., Wolbach, A., Harkes, J. Perrig, A., Farber, D., Kozuch, M., Helfrich, C., Nath, P., Lagar-Cavilla, H. "Pervasive Personal Computing in an Internet Suspend/Resume System." *IEEE Internet Computing*, Vol. 11, No. 2, March/April 2007.

[25] IDC. 2009. Available at ***http://community.edc.intel.com/t5/Applications-Other-Blog/15-Billion-Connected-Devices-Powered-by-the-Embedded-Internet/ba-p/613***

[26] IDC. December 2009. Available at ***http://news.softpedia.com/news/Over-One-Billion-Mobile-Devices-to-Access-the-Web-by-2013-129261.shtml.***

[27] Dash, D., Kveton, B., Agosta, JM., Schooler, E. M., Chandrashekar, J., Bachrach, A. and Newman, A. "When Gossip is Good: Distributed Probabilistic Inference for Detection of Slow Network Intrusions." *AAAI*, 2006.

[28] Huang, J., Berry, R, and Honig, M. "Distributed Interference Compensation for Wireless Networks." *IEEE Journal on Selected Areas in Communications*, Volume 24, No. 5, pages 1074-1084, May 2006.

[29] Zage, D., Livadas, C. and Schooler, E. M. "A Network-Aware Distributed Membership Protocol for Collaborative Defense." *International Conference on Computational Science and Engineering*, Volume 4, pages 1123-1130, 2009.

[30] Campbell, R., Gupta, I., Heath, M., Steven, K. Y., Kozuch M., Kunze M., Kwan, T., Lai, K., Lee Y., Lyons M., Milojicic, D., O'Hallaron D. and Soh Y. "Open CirrusTM Cloud Computing Testbed: Federated Data Centers for Open Source Systems and Services Research." HP Laboratories, *HPL-2009-134*, 2009.

[31] "Wi-Fi Direct." *Wi-Fi Alliance*, 2010. Available at ***http://www.wi-fi.org/files/20091019_WiFi_Direct_FAQ.pdf***

[32] "Wi-Di." *Wireless Display*. Available at ***http://www.intel.com/consumer/products/technology/wirelessdisplay.htm***

[33] Scheifler, R. and Gettys, J. "The X Window System." *ACM Transactions on Graphics* (TOG), 1986.

[34] Jakubczak, S., Andersen, D., Kaminsky, M., Papagiannaki, K. and Seshan, S. "Link-alike: Using Wireless to Share Network Resources in a Neighborhood." *ACM SIGMOBILE MC2R*. Volume 12, No. 4, October 2008.

[35] Devine, R. "Design and Implementation of DDH: A Distributed Dynamic Hashing Algorithm." In *Proceedings of 4th International Conference on Foundations of Data Organizations and Algorithms*, 1993.

[36] Chan, H. and Perrig A. "Security and Privacy in Sensor Networks." *Carnegie Mellon University*, IEEE Computer, pages 99-101, October 2003.

[37] Allcock, W., Bresnahan, J., Kettimuthu, R. and Link, J. "The Globus eXtensible Input/Output System (XIO): A Protocol Independent IO System for the Grid." In *Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium* (IPDPS'05) - Workshop 4, 2005.

[38] Dean J., and Ghemawat S. "MapReduce: Simplified Data Processing on Large Clusters." In *Proceedings OSDI*, 2004.

[39] Min, X., Jizhong, Z., Yong, Q., Hui, H., Ming, L. and Wei, W. "Isotope Programming Model: a Kind of Program Model for Context-Aware Applications." In *Proceedings International Conference on Multimedia and Ubiquitous Engineering* (MUE'07), 2007.

[40] Weichang, D. and Lei, W. "Context-aware application programming for mobile devices." In *Proceedings 2008 C3S2E conference*, 2008.

[41] Fall, K. "A delay-tolerant network architecture for challenged internets." In *Proceedings ACM SIGCOMM*, Karlsruhe, Germany, ISBN:1-58113-735-4, pages 27–34, 2003.

[42] Kanter, T., Olrog, C., "VoIP in Applications for Wireless Access." *10th IEEE Workshop on Local and Metropolitan Area Networks*, pages 122-125, 1999.

[43] Greenstein, B., Mccoy, D., Pang, J., Kohno, T., Seshan, S. and Wetherall, D. "Improving wireless privacy with an identifier-free link layer protocol." In *Proceedings ACM MobiSys'08*, Breacon Ridge, Colorado, 2008.

[44] Gruteser, M. and Grunwald, D. "Enhancing location privacy in wireless LAN through disposable interface identifiers: a quantitative analysis." *Mobile Networks and Applications*, Vol. 10, No. 3, pages 315-325, 2003.

[45] Eschenauer, L., Gligor, V. and Baras, J.S. "On Trust Establishment in Mobile Ad-Hoc Networks." *Security Protocols, Lecture Notes in Computer Science*, 2002.

[46] Löhr, H., Sadeghi, A., Vishik, C., and Winandy, M. "Trusted Privacy Domains – Challenges for Trusted Computing in Privacy-Protecting Information Sharing." *Lecture Notes in Computer Science*, volume 5451, pages 396-407, 2009.

Acknowledgements

We thank Dina Papagiannaki, Trevor Pering, John Gustafson, Lakshman Krishnamurthy, Anthony Joseph, and Herman D'Hooge for their technical feedback.

Author Biographies

Roy Want is a Senior Principal Engineer at Intel Labs, Santa Clara, California, and Associate Director for XTR. His research interests include mobile and ubiquitous computing, wireless protocols, embedded and distributed systems, automatic identification, and MEMS. Want received his PhD degree in Computer Science from Churchill College, Cambridge, UK in 1988. He is currently the Chair of ACM SIGMOBILE, and a Fellow of both the IEEE and ACM.

Eve Schooler joined Intel in 2005. She is a Principal Engineer in Intel Labs within its new Energy Systems Research Lab, where she is designing an intelligent control plane for Smart Homes and Smart micro-Grids. Previously she led the Distributed Detection and Inference (DDI) project, an effort that focused on collaborative anomaly detection to secure large-scale networks. Her broad interests lie at the intersection of networking, distributed systems, and scalable group algorithm design. Eve served on the Transport Directorate of the IETF, co-founded and co-chaired the IETF MMUSIC working group for many years, and is a co-author of the SIP protocol that is widely used for Internet telephony. Eve obtained an BSc degree from Yale University, an MSc degree from UCLA, and a PhD degree from Caltech, all in Computer Science. Prior to Intel, she held positions at Apollo Computers, Information Sciences Institute (ISI), AT&T Labs-Research and Pollere LLC.

Lenka Jelinek. Lenka's passion is incubating technologies, growing ideas from an inspiration into a market reality. Her most recent focus has been on making power cords obsolete through wireless power. In her prior 15 years of experience at Intel, Lenka led research on intuitive systems and low-power IA platforms. Lenka was a co-founder of the Intel Smart Toy Lab and a lead architect on the award-winning Intel® Play™ QX3™ Computer Microscope. She served as the TA to Andrew Chien, VP and Director of Intel Research. Lenka holds degrees in Electrical Engineering and Computer Science from the University of Wisconsin-Madison.

Jaeyeon Jung received her PhD degree in Computer Science from MIT in 2006, and her Bachelor's and Master's degrees from the Korea Advanced Institute of Science and Technology (KAIST). Her research focuses on many aspects of networked systems and system security. Prior to joining Intel Labs, Seattle, Jaeyeon worked at Mazu Networks and implemented her scan-detection algorithms in Mazu's Profiler product. She is an affiliate faculty member at the University of Washington and KAIST.

Dan Dahle is a Sr. Strategic Architect in Intel Labs responsible for developing and executing strategies for moving Intel research technologies into the marketplace. In the past, Dan led a team chartered with enabling technologies for anytime, anywhere computing. The team also worked in the areas of communications in mobile networks, including Wi-Fi (802.11) hotspots, WiMax, and cellular networks, addressing security, authentication, authorization and accounting (AAA) for safer, simpler, seamless access for mobile clients. Prior to Intel, Dan ran his own marketing consulting company. He started his career with AT&T in product engineering, product management, worldwide sales, and business development. A licensed professional engineer, Dahle holds a BS degree in Mechanical Engineering and Material Science and Engineering from the University of California at Berkeley, as well as an MBA from Indiana University.

Uttam Sengupta is a principal engineer in the Ultra Mobility Group and is responsible for defining security, manageability and virtualization capabilities of handheld platforms. Uttam's other interests include ensemble computing, location & context sensing and related technologies required for a service ready platform. Prior to Intel, Uttam was a principal engineer at Motorola, working in the area of GSM cellular technologies. He received a PhD in computer science from Arizona State University. Uttam is a member of IEEE Computer Society.

Copyright

OPPORTUNITIES AT THE INTERSECTION OF BIOLOGY AND SILICON: DECODING BIOLOGICAL COMPLEXITY

Contributors

Madoo Varma
Intel Corporation

Xing Su
Intel Corporation

Grace Credo
Intel Corporation

Oguz Elibol
Intel Corporation

Index Words

Silicon
Biosensor
Genome (or DNA)
Nanoscale
Arrays
Bioinformatics
Biomolecules
Interfaces

Abstract

Accelerating developments in biology continue to transform this formerly qualitative science into an increasingly data-driven science, fueling the need for tools to collect, analyze, and interpret molecular and system-level data. At the same time, the aggressive scaling of silicon transistors has enabled massive computational power, as well as reliable and reproducible manufacturing of components at dimensions comparable to individual biomolecules. There is tremendous potential and opportunity for both fields to affect each other: lessons from biology can be applied to improve silicon technology, for example, in creating brain-inspired cognitive computer architectures, or advances in silicon manufacturing technology can affect how biology is done by creating massively parallel, cost-effective, research and screening tools. It is conceivable that learning the precise self-assembly and proof-reading of molecular reactions may someday enable the creation of structurally programmable nanoscale devices.

In the past, silicon manufacturing technology and biology were relatively isolated by mutually exclusive domains: different tools, different approaches, different length scales. Recently, however, modern silicon manufacturing expertise has reached a stage where nanometer scale features can uniquely probe biological processes at molecular length scales. For example, the development of silicon-based sensors for electronic, label-free detection of biomolecular interactions has the potential to cost-effectively enable massively parallel genetic analysis at the single molecule level, yielding an unprecedented amount of biological information. Towards this end, we are investigating new interfaces, sensors and biomolecules, which could lead to the development of high-performance, cost-effective platforms for biomedical applications.

Introduction

Imagine a future when you would carry your baseline genome information in your wallet and have the genetic code of suspected diseased cells sequenced during a short visit to the doctor's office or a local clinic. If a master database existed where your genetic code could be screened against disease markers or drug metabolism genes on demand, cancers could be diagnosed earlier for timely treatment, the most effective medicines could be prescribed to cure an infection, or the appropriate dose of a drug could be given to lessen an acute symptom [1 - 4]. Realizing this vision requires further progress in understanding underlying molecular mechanisms in the biology of disease and health. Associations between genetic variations and various disease states need to be more thoroughly understood. In recent decades, this understanding has been a function of advanced tool development for biomolecule detection and improved data analysis, by the use of in-silico bioinformatics tools for processing the raw information [2, 6, 7]. The ability to quickly collect biological samples, analyze them at system level [8 - 11], and accurately screen them against the complex information contained in dynamic, in-silico databases would change how biology research is done today, revolutionizing targeted human disease management and enabling more informative monitoring of our natural environment and its resources [12, 13].

To make this vision a reality, current bioanalysis tools need to come down in price, and they need to show improved performance and ease of use (sample-to-answer) compared with today's technologies. As we describe in this article, we envision many opportunities to apply existing semiconductor expertise to maximizing impact where biology meets silicon. We describe an effort at Intel to develop a fully integrated, massively parallel, electronic sensing array for single biomolecule detection. A compact and easy-to-use system for sequencing and mutation detection with on-system data analysis would open up new markets in field-deployable sensing, such as detection of rapidly mutating strains of infectious agents and natural or engineered biotoxin agents, and it could help in the discovery of new sources of biofuels, among other things.

> *"We describe an effort at Intel to develop a fully integrated, massively parallel, electronic sensing array for single biomolecule detection."*

There are equally large opportunities in human and veterinary medicine and in agriculture, as well as in industrial screening, diagnostics, and monitoring. In this article, we review the complexity of biological systems in terms of molecular structures and their interactions, the current challenges in understanding the complexity, the role that silicon can play in solving these challenges, and how new approaches for decoding biological complexity can be revolutionary for biomedical, clinical, and industrial applications.

Complex Biological Systems—the Ultimate Frontier

A human individual begins in a single fertilized cell that proliferates to an estimated 100 trillion cells that self-assemble and differentiate to form functional organs and tissues [14, 15]. Within organs and tissues a variety of molecular processes occur. These subsystems within the body are highly interconnected and diverse (see Figure 1), functioning at several levels and presenting many layers of informational complexity. In the human genome, there are approximately 20,000 genes, which provide the basic instructions (expressed through molecules called mRNAs) for about one million different proteins. These proteins dictate cellular activities. Protein expression and relative activity depend on tissue types, developmental stages, and environmental stimuli. Moreover, biomolecular interaction networks or pathways regulate the functions and behaviors of cells in real time, making a biological system highly dynamic in nature. For example, the genetic makeup of cancer cells within the same tumor are not necessarily the same, and they may use different cellular signaling pathways to maintain uncontrolled growth, potentially expressing different markers at different stages of cancer within the same individual [4, 11]. Thus, accurate disease diagnosis and monitoring would benefit from the ability to detect multiple biomolecules at relatively low physiological concentrations at multiple time points [3, 16].

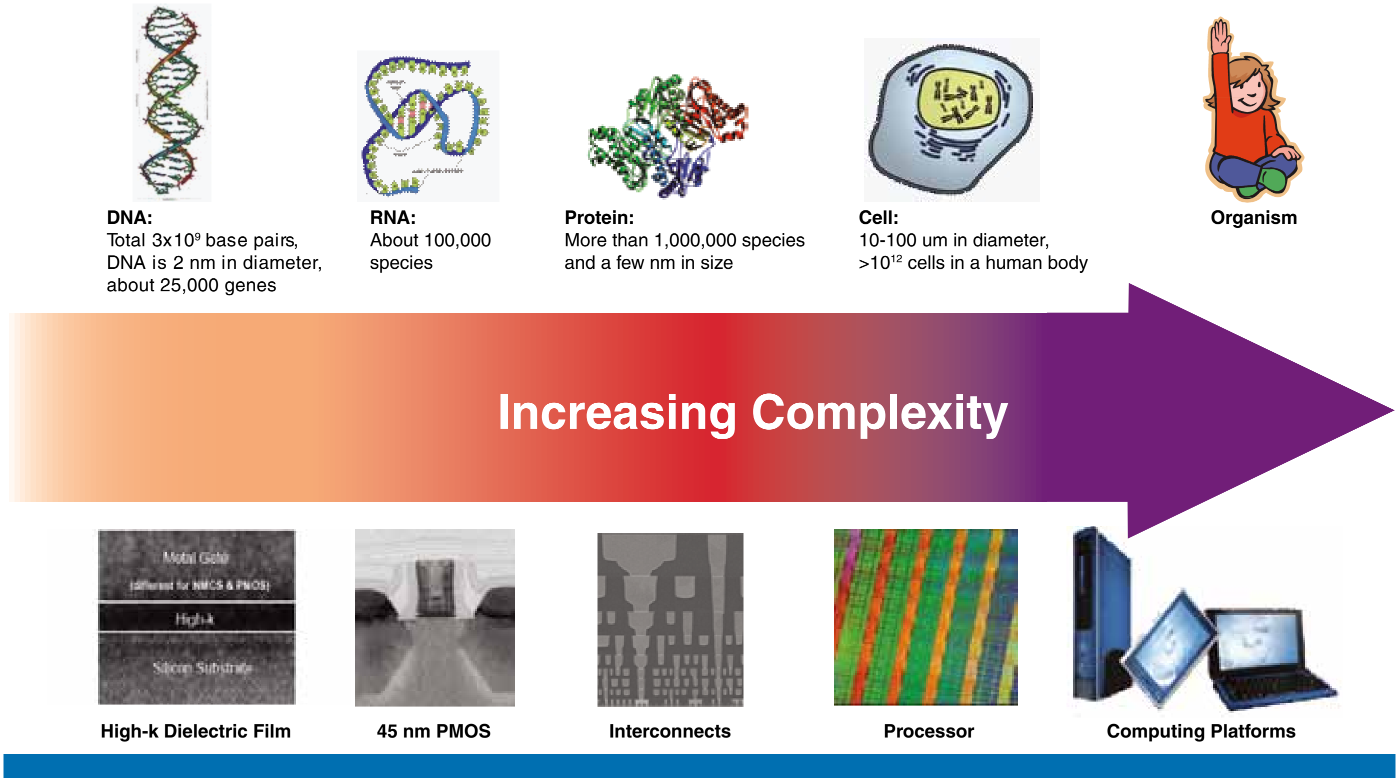


Figure 1: Biomolecules and the Silicon Components Shown in Increasing Complexity from Left to Right
Source: Intel Corporation, 2010

It is worth noting that when doctors prescribe a drug for a condition, the evaluation of its efficacy is traditionally a process of trial and error: such a process does not take into account much of the biomolecular information that could be collected from an individual or population [14, 17 - 19]. Traditional medical diagnoses and treatments focus on one subsystem or organ at a time and may use only one, preferably inexpensive, test for monitoring. A more complete and thorough evaluation of medical conditions is missing. Although the emerging paradigm of targeted medicine, based on molecular profiling, that addresses "one size does not fit all," is beginning to take shape, it is limited in scope. There is a lack of robust, cost-effective, and high-resolution tools, and many ethical questions, legal concerns, and economic business model challenges still exist.

> *"When doctors prescribe a drug for a condition, the evaluation of its efficacy is traditionally a process of trial and error: such a process does not take into account much of the biomolecular information that could be collected from an individual or population."*

The decade-long Human Genome Project, completed in 2003 at a cost of 3 billion dollars, provided a reference blueprint for human DNA, setting the stage for molecular analysis to fully describe the role that gene structure played in biomolecular function. Despite the promise of genetic sequencing, the cost was prohibitive, providing a high barrier for wide application and adoption. In recent years, efforts by a few start-up businesses and some large companies have led to the rapid progression of next-generation sequencing technologies [9] that have led to some breakthroughs regarding the function of protein-coding genes within DNA. However, these efforts are still few and far between due to the cost, among other factors. The global functional analysis of biomolecular processes at the individual and population level is currently being targeted in earnest with much of the analytical focus on a functional understanding of each person's genetic makeup.

Tools Enabling Global Profiling: the Key to Decoding Biocomplexity

The need to understand complex cellular, subcellular, and molecular interactions on a global scale has been well documented [14, 15]. In this article, we limit our scope to information gleaned at the genomic level. The information needed to synthesize proteins, enzymes, and other cellular components, ensuring a 'normal,' working body is coded by the human genome. The genome is the complete genetic make-up of an individual and is made up of double helical strands of DNA. The discovery of DNA in 1953 was followed by the development of reagents, such as modifying enzymes, and methods for DNA manipulation and genetic analysis over the next three decades. One of the most important techniques, the polymerase chain reaction (PCR), invented in the 1980s, can be used to create multiple copies of a DNA segment, allowing researchers to study specific DNA fragments in detail. At this point, researchers were limited to studying complex biomolecular phenomena one gene at a time, limiting throughput. In the 1990s, with the advent of DNA microarray chips that are based on hybridization (matching stretches of complementary DNA strands), tens of thousands of genes could be studied at one time. Disadvantages limiting the utility of microarray chips included non-specific binding during hybridization, a-priori knowledge of sequence information (in order to deposit the correct sequences on chip), and the high cost of fabrication. Sequencing has remained the method of choice against which all other methods (PCR and microarrays) of genetic analysis are measured, especially when accuracy and structural confirmation are required. In recent years, several lower-cost, new-generation, sequencing technologies have been developed, based on improved sample preparation methods (emulsion PCR and in situ PCR) and on new sequencing chemistries (luminescence or nucleotide analogs) [9, 16].

"The polymerase chain reaction (PCR), invented in the 1980's, can be used to create multiple copies of a DNA segment, allowing researchers to study specific DNA fragments in detail."

Despite rapid advances in sequencing technologies, the rate of progress is still limited by the availability of tools with the appropriate resolution and performance/cost thresholds; these are tools that can enable the necessary large-scale population studies and analytics to uncover the basis of health and disease. The latest sequencing systems on the market can sequence a human genome in eight days for $10K by using an optical detector that costs $700K [20]. Three key players in the research market sell refrigerator-sized central lab instruments that use optical detectors and consumable reagents. While the cost of sequencing has dramatically decreased over the years, the DNA analysis process and the required informatics-based assembly are still daunting tasks. Some companies are focusing on a high-throughput services model, currently targeting DNA sequencing services in the $5,000 range [21, 22]. Recently, there has been some movement toward electronic detection by two start-up companies attempting to reduce capital costs, reagent costs, and the time required for sequencing. However, the proposed sequencing methods still require cumbersome sample preparation. Sample-to-answer integration for sequencing is still beyond the capability of technologies [5].

"While the cost of sequencing has dramatically decreased over the years, the DNA analysis process and the required informatics-based assembly are still daunting tasks."

"The ultimate biomolecular analysis tool would provide full, black-box, end-to-end integration."

To reduce reagent costs and/or miniaturize instrument size, the latest DNA sequencing approaches in development focus on single molecule optical [23, 24] or electronic [25, 26] sensing. There are multiple technical challenges for various types of proposed approaches to DNA sequencing. For example, a single enzyme array-based approach scheduled to be commercialized in 2010 can successfully sequence single, short DNA molecules in nanoscale wells. However, the well array density is fundamentally limited by the yield of effective wells that have both an active enzyme and an attached DNA molecule. Moreover, the well pitch is much larger than single molecule dimensions, at the 10 μm level [23], since optical detection still requires a large microscope [24]. In the nanopore method [25 - 27], a DNA strand often moves too fast making reading accuracy a problem [27]. Finally, tip-based (AFM) methods have demonstrated the principle of distinguishing single modified DNA bases by current, but this approach has not been demonstrated on actual DNA in practical settings [28]. Realizing a "small enough probe" and being able to manipulate a full DNA strand in aqueous solution in order to read the bases rapidly and accurately still remains the holy grail of single biomolecule detection.

The Next Step: Solving Challenges in Decoding Single Biomolecules

The ultimate biomolecular analysis tool would provide full, black-box, end-to-end integration: a sample would be inserted with a particular query and a definitive answer would be delivered in real-time [29]. In the following section we focus on such a tool for DNA sequencing, but many of the same principles apply to other biomolecules. Our vision for the analysis of single DNA molecules requires single molecule sensitivity, massive parallelism, sub-molecular resolution, on-platform data analysis, and complete sample-to-answer integration.

"A single molecule can be handled or monitored by its charge properties, modified tag, and unique chemical affinity."

Single Molecule Sensitivity

The ability to probe a single DNA molecule obviates the need for time- and material-intensive DNA target amplification, and it eliminates replication errors caused by the multiplication step. These unique benefits of single molecule analysis are a key feature in recent amplification-free single molecule sequencing technologies [23, 24]. A combination of selective chemistry, high-volume nanofabrication, and integrated circuitry could enable single DNA molecule transport, positioning, and sensing by nanostructures or electronic sensor arrays. A single molecule can be handled or monitored by its charge properties, modified tag, and unique chemical affinity. A variety of approaches, when used in combination (chemical amplification, electronic signal amplification and signal confinement), can enable single molecule detection.

Submolecular Resolution (Specificity)

DNA and other biomolecules have functional units that affect their structure and function; for example, the 4 bases in DNA and the 20 amino acids in proteins. Resolving the composition and order of these units is a key objective of specific detection, but a challenging one, because the pitch between 2 adjacent DNA bases is only 3.4 angstroms, and the bases are similar in charge and mass. So far, optical detection methods have not been able to resolve these sub-molecular differences without the involvement of complicated instruments and tedious biochemistry, because sensitive detectors are physically removed from reactions on the surface (large microscopes), and biomolecules frequently need to be amplified. With electronic methods, a molecule or its reaction products can be placed directly in contact with a sensor surface and thus can be detected without diverting the signals elsewhere, thereby reducing complicated sample preparation and the need for large optical detectors. In addition, electronic sensing allows for the inclusion of signal conditioning and information-processing components on the same platform, resulting in reduced complexity, cost, and size. When combined with innovative bioassay methods that localize reactions, an electronic sensor array is a promising route for deriving sub-molecular information from complex molecular structures, such as a genome of one billion bases.

Massively Parallel Arrays with Simple Detectors

The instrument size is largely determined by the detection method, and currently the most common method is optical detection of fluorescent dye labels on DNA bases [1, 16]. Large microscopes or laser excitation, used in optical single molecule DNA sequencing, can be eliminated by using electronic sensing; with this method, the signal amplification and transduction steps are confined to a scalable semiconductor chip platform. It is estimated that to sequence a complete human genome with 1x coverage in one sequencing run, the array must have a minimum of 30 million reaction sites, assuming 100 bases can be read from each site. These values can change by 1 to 2 orders of magnitude as a function of coverage and read length.

"In addition, electronic sensing allows for the inclusion of signal conditioning and information-processing components on the same platform, resulting in reduced complexity, cost, and size."

Massively Parallel Data Analysis and Processing

In the case of DNA sequencing, the data processing required to deduce the identity of base pairs from raw data is a technological challenge [12]. Overlapping runs and multiple reads are needed to compensate for errors and to assist in DNA fragment alignment. Alignment and assembly are substantially more difficult for newer, next-generation, sequencing (NGS) data than they are for traditional Sanger sequencing (800bp); this is due to the shorter read lengths (~35 bases) [6, 30]. NGS experiments, based on optical detection platforms, generate unprecedented volumes of data (tens of terabytes of data per run), which require solutions for data management, storage, and, most importantly, processing and analysis. The tight integration of hardware with software (system-on-chip, embedded systems) will be an area in which semiconductor expertise can add value. Signal processing circuits can be

> *"Signal processing circuits can be integrated on the same die as the sensors themselves, performing the first step in transforming raw sensor data into useful information."*

integrated on the same die as the sensors themselves, performing the first step in transforming raw sensor data into useful information. Furthermore, special-purpose computation blocks for sensor data can be combined to enhance the speed and efficiency of the data analysis that inevitably follows biosensor readout.

Silicon Innovation for Decoding Biological Complexity

With the use of silicon technology, the Integrated Biosystems Research group within Intel Labs is working on creating platforms by using electronic sensor arrays for various biomedical applications such as nucleic acid sequencing (Figure 2) and protein profiling (Figure 3).

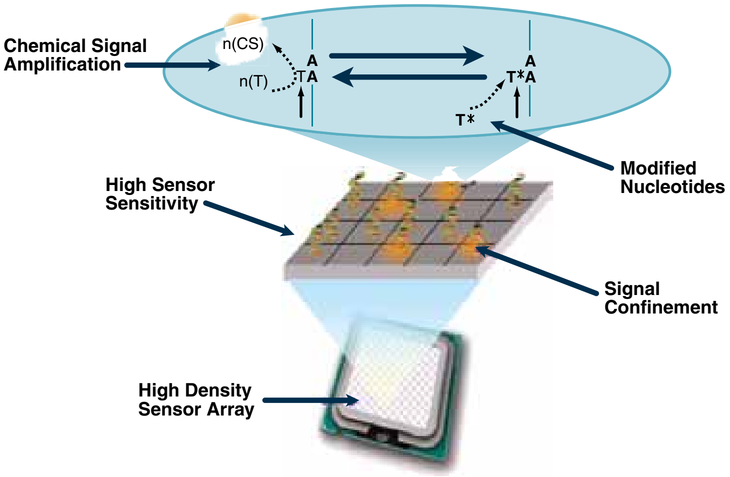


Figure 2: Overview of the General Concept of the Electronic Sensor Array Platform for DNA Sequencing
Source: Intel Corporation, 2010

Towards Nucleic Acid Sequencing with Electronic Sensor Arrays

To develop a sensitive, massively parallel, electronic sensor array, we are innovating in several inter-related areas: chemistry, biology, material interfaces, device fabrication, and circuit design. For signal generation, we are using novel sequencing-by-synthesis (SBS) chemistry designed to be sequence specific and compatible with massively parallel, single biomolecule reactions. We have demonstrated significant chemical signal amplification (>50x) off-chip. To enable single molecule detection, we are developing silicon-based nanosensor arrays that are based on devices with novel surfaces and circuit architectures.

Towards realizing a dense and sensitive array of electronic sensors, we have been exploring biomolecule detection, in collaboration with a leading academic group, by using chemically-modified field effect sensors. These sensors are promising candidates for forming fully electronic, ultra-dense, biosensor arrays due to their inherent sensitivity and scalability. The devices are sensitive to charge present on the surface [31, 32]. By immobilizing specific capture agents on the surface of the devices, we were able to detect the presence of charged nucleotide incorporation byproducts (PPi) [33]. We are investigating other promising sensor technologies and approaches, such as impedance spectroscopy [34].

Protein Profiling with Photolithographic Peptide Arrays

Towards the goal of leveraging microfabrication technology to develop biosensing platforms, we have developed technologies for lithographically patterning specific molecules on a silicon-derived surface . To realize a diverse biomolecule sensing platform, simultaneous detection of multiple species (multiplexing) is often necessary. Thus, patterning different molecules at unique sensor sites becomes a key enabling technology for realizing dense arrays at a low cost. In collaboration with a group at Stanford Medical School [35], we have demonstrated the utility of fluorescently labeled peptide (protein fragment) arrays that are lithographically patterned on silicon to profile the protein function of patient samples (diseased versus control), as depicted in Figure 3. In future studies these patterning methods will be applied to electronic sensors.

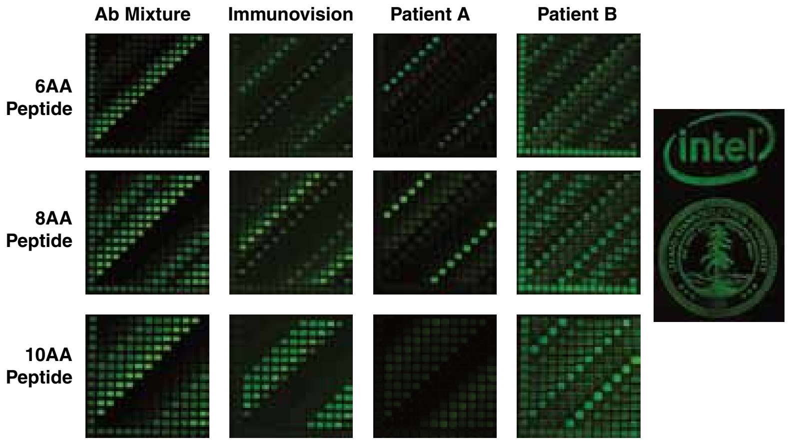


Figure 3: Photolithographic Peptide Microarrays on Silicon Wafers for Screening Patient Protein Profiles
Source: Intel Corporation, 2010

Developing Electronic Sensors for Single Molecule Sensing

We are investigating several single molecule approaches for biochemical detection, such as nanopore sequencing technology, in conjunction with our academic collaborators. This approach offers the possibility of high-throughput single molecule DNA sequencing without the use of sensors in a massively parallel array format [26]. Finally, we have also been exploring signal amplification towards single-molecule sensing by using nanogap redox cycling devices, as shown in Figure 4.

In these sensors, analytes with reversible redox properties shuttle electrons between two nanogap electrodes, when the two electrodes are biased at the corresponding reduction and oxidation potentials, amplifying the signal (electrons generated) per molecule. Results showed that 10 nM of ferrocene could be differentiated from pure buffer, which corresponds to about 50-100 molecules in the gap between electrodes, similar to reported results [36]. One of our research goals is to progressively improve the detection limit to fewer molecules in smaller gaps.

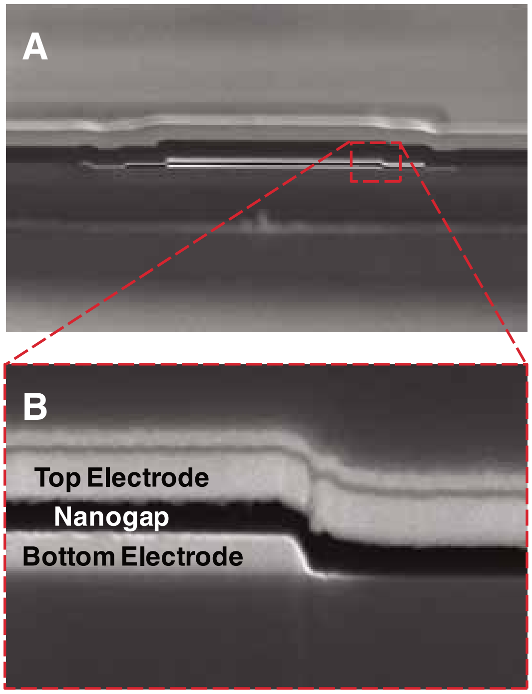


Figure 4: Cross-sectional Scanning Electrode Microscope Images of Nanogap Devices
A — Showing the Whole Cavity
B — Magnified Image Showing the Well-defined Nanogap Separating the Two Electrodes
Source: Intel Corporation, 2010

Conclusion

As we describe in this article, biological systems are complex at the organism, cellular, and molecular levels. Recent progress made by scientists and engineers from various areas of expertise has decoded the genetic blueprints of humans and many other organisms, which gives us the basic information to probe the mechanisms that underpin the activities and behaviors of living organisms. More extensive and more frequent biomolecular analysis that enables real-time answers to biology-related questions will require significant advances in several areas of technology. By addressing the challenges described in this article, we seek to bridge the gap between existing sequencing approaches and Intel's silicon expertise. Our goal is to address some of the key bottlenecks of cost, ease of use, highly parallel data acquisition and analysis, and automatic operation.

Feature resolution, high-volume manufacturing, and integrated circuitry are some of the challenges of creating consumer tools for rapid and reliable biological analysis. These challenges are regularly addressed by constant innovation in semiconductor manufacturing by the use of silicon-based technology. The strength of Intel's expertise in high-volume lithography, novel transistor materials, such as high-k dielectrics, wafer-scale deposition of thin layers, and minimization of transistor sizes, positions us strategically to take advantage of this burgeoning opportunity in decoding biological complexity.

> *"We have also been exploring signal amplification towards single-molecule sensing by using nanogap redox cycling devices."*

References

[1] Metzker, Michael L. "Sequencing technologies — the next generation." *Nature Reviews Genetics*, Volume 11, 2010.

[2] Aspinall, Mara G., Hamermesh, Richard G. "Realizing the promise of personalized medicine." *Harvard Business Review* 85, No. 10 (October 2007).

[3] Soper, S.A. "Point-of-care biosensor systems for cancer diagnostics/prognostic." *Biosensors and Bioelectronics*, Volume 21, 2006.

[4] Leary, R., et al. "Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing." *Science Translational Medicine*, Vol. 2, Issue 20, page 20ra14, February 2010.

[5] Leaman, J.H., Braverman, M., and Rothberg, J.M. "High-Throughput, Massively Parallel DNA Sequencing Technology for the Era of Personalized Medicine." *Gene Therapy and Regulation* 3(1): 15-31, 2007.

[6] Karl V. Voelkerding, Shale A. Dames, Jacob D. Durtschi. "Next-Generation Sequencing: From Basic Research to Diagnostics." *Clinical Chemistry* 55:4 641–658, 2009.

[7] Emili A, Cagney G. "Large-scale functional analysis using peptide or protein arrays." *Nature Biotechnology* 18:393-397, 2000.

[8] Shendure, Jay. "The beginning of the end for microarrays?" *Nature Methods*, Volume 5, Issue 7, 2008.

[9] Shendure, Jay and Ji, Hanlee. "Next-generation DNA sequencing." *Nature Biotechnology*, Volume 26, No. 10, 2008.

[10] Aaron Daub, Punita Sharma, and Steven Finkbeiner, "High-content screening of primary neurons: ready for prime time." *Current Opinion in Neurobiology*, 19: 537-543, 2009.

[11] Jones, S., et al. "Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses." *Science* 26, Vol. 321. no. 5897, pages 1801–1806, September 2008.

[12] Collins, F.S. et al. "A vision for the future of genomics research." *Nature*, Volume 422, 2003.

[13] B. T. Grenfell, C. S. Williams, O. N. Bjournstadi and J. R. Banavar. "Simplifying biological complexity." *Nature Physics*, Volume 2, pages 212-214, April 2006.

[14] National Human Genome Research Institute (NHGRI). Available at ***http://www.genome.gov/27532724***

[15] Hiroaki Kitano. "Systems Biology: A Brief Overview." *Science* 1, Vol. 295. no. 5560, pages 1662–1664, March 2002.

[16] Jay Shendure, Robi D. Mitra, Chris Varma, George M. Church. "Advanced Sequencing Technologies: Methods and Goals." *Nature Reviews: Genetics*, volume 5, pages 335-344, 2004.

[17] Choi, M., et al. "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing." In *Proceedings of the National Academy of Sciences*, volume 106 no. 45 19096-19101, November 10, 2009.

[18] Sergei M. Mirkin. "Expandable DNA repeats and human disease." *Nature* 447, 932-940 doi:10.1038/nature05977, June 21, 2007.

[19] Adi F. Gazdar, M.D. "Personalized Medicine and Inhibition of EGFR Signaling in Lung Cancer." *New England Journal of Medicine*, Volume 361:1018-1020, September 3, 2009.

[20] "Illumina." Available at ***http://www.illumina.com/systems/hiseq_2000.ilmn#workflow_specs***

[21] Aaron Rowe. "The First and Last Meeting of Everyone with a Fully Sequenced Genome." *Wired*. Available at ***http://www.wired.com/wiredscience/2010/02/getconference/***

[22] Drmanac, et al. "Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays." *Science*, 327:78-81, 2010.

[23] Eid, J., et al. "Real-Time DNA Sequencing from Single Polymerase Molecules." *Science*, Volume 323, Issue 5910, 2009.

[24] Harris, et al. "Single-Molecule DNA Sequencing of a Viral Genome." *Science*, 320:106–109.

[25] Astier, Y, Braha, O and Bayley, H. "Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter." *Journal of the American Chemical Society*, Volume 128, 2006.

[26] H. Bayley. "Sequencing single molecules of DNA." Current Opinion in Chemical Biology. 2006, v. 10, pages 628-637; DOI: 10.1016/j.cbpa.2006.10.040.

[27] Branton, Daniel et al. "The potential and challenges of nanopore sequencing." *Nature Biotechnology*, Volume 26, Issue 10, 2008.

[28] Chang, S., et al. "Electronic Signatures of all Four DNA Nucleosides in a Tunneling Gap." *Nano Letters*, ASAP Online, 2010.

[29] J. R. Heath. "Label-Free Nanowire and Nanotube Biomolecular Sensors for In-Vitro Diagnosis of Cancer and other Diseases." *Nanobiotechnology II*, editors, Chad Mirkin, Christof Niemeyer, Wiley-VCH, Weinheim, page 227, 2007.

[30] Mihai Pop, Steven L. Salzberg. "Bioinformatics challenges of new sequencing technology." *Trends in Genetics*, Volume 24 No.3, Center for Bioinformatics and Computational Biology, University of Maryland, MD 20742, USA, 2008.

[31] Elibol, Oguz et al. Manuscript submitted.

[32] Schöning M.J., Poghossian, A. "Recent advances in biologically sensitive field-effect transistors (BioFETs)." *Analyst*, 127(9):1137-51, 2002.

[33] Credo, Grace, et al. Manuscript in preparation.

[34] Daniels, Jonathan S. and Pourmand, Nader "Label-Free Impedance Biosensors: Opportunities and Challenges." *Electroanalysis*, Volume 19, Number 12, 2007.

[35] Yu, Jiangtao, et al. Manuscript in preparation.

[36] Zevenbergen, M.A.G. et al. "Mesoscopic Concentration Fluctuations in a Fluidic Nanocavity Detected by Redox Cycling." *Nano Letters*, Volume 7, Number 2, 2007.

Acknowledgments

We acknowledge Badari Kommandur for an insightful review of the manuscript, Jon Daniels for excellent technical feedback, and all members of the IBL team for help in generating results.

Author Biographies

Madoo Varma is the Director of the Integrated Biosystems Lab within Intel Labs and is responsible for its research and business operations. She joined Intel in 2005 as part of the Digital Health group. She earned her PhD degree in Genetics in 1981 and spent 14 years in biotech R&D and 12 years in strategic planning, corporate venture and senior business development roles at Hewlett Packard/Agilent and Applied Biosystems. In 2009, she was nominated to the National Academy of Sciences panel on Electronics and Electrical Engineering.

Xing Su is a Senior Scientist in the Integrated Biosystems Research Labs within Intel Labs. He received his PhD degree in Biology from the University of California, Santa Barbara in 1989 and completed his post-doctoral training in molecular biology at Harvard University. Xing has 20 years' experience in life sciences, and during his earlier tenure at Affymetrix was responsible for DNA microarray genotyping assay development. He holds 30 issued patents.

Grace Credo is currently a Research Scientist in the Integrated Biosystems Research Lab within Intel Labs. Grace received her PhD degree in Physical Chemistry from the University of California, Santa Barbara in 2001. She has held analytical chemistry positions at UC San Diego, North Carolina State University, and Waters Corporation. Her expertise is the use of advanced methods and tools for surface deposition and characterization.

Oguz H. Elibol received his PhD degree in Electrical and Computer Engineering from Purdue University in 2008: his dissertation work involved developing nanoscale thickness field effect sensors for bio-chemical sensing and temperature-mediated chemical reactions. Currently he is utilizing his background and expertise for the design and fabrication of electronic biosensor devices and arrays.

Copyright

CLOUD COMPUTING ON RICH DATA

Contributors

Michael Kozuch
Intel Corporation

Jason Campbell
Intel Corporation

Madeleine Glick
Intel Corporation

Padmanabhan Pillai
Intel Corporation

Index Words

Cloud Computing
Data Rich
System Software
Virtualization
Streaming
Optical

Abstract

Advances in sensing technologies are yielding vast quantities of data that must either be processed on the fly or archived for later consumption, or both. Unfortunately, new applications, eager to harvest such data, are starkly limited by current techniques that tend to be hard to scale up, difficult to adapt, and largely batch-oriented. To enable organizations and individuals to leverage their newfound wealth of information, Intel Labs is investigating key new cloud-computing technologies that will deliver higher performance for stored data at scale, lower latency for streaming data despite greater processing complexity, and a lower-cost network infrastructure. The goal is to drive exciting new breakthroughs in robotics, computational perception, personal media, mobile computing, cluster analytics, machine learning, and similar applications.

Introduction

In recent years, advances in semiconductor electronics have pushed the instrumentation of our world to unprecedented levels. Sensors are now all around us: many cell phones contain GPS receivers as well as cameras, doorways have motion detectors, stop lights sense vehicles at intersections, and satellites orbiting overhead are constantly imaging the Earth. Additionally, we have data sourced electronically: feeds from social networking sites, crawls of Web pages, repositories of medical images, results from computer simulations, etc. Many of the data objects from these sources are collected for analysis, archived, subjected to re-analysis, cross-correlated with other data objects, and processed to create additional, derived data sets.

The result is that we live in a world that is *data rich*. In this article, we consider two types of data sources: stored and streaming. A stored data object is just that, information that has been archived in some way. A corpus of digital images stored on a collection of magnetic disks would be an example of stored data. Streaming data objects have a real-time component; a live video feed is the canonical example of streaming data. The two types of data present different processing challenges in that applications operating on stored data are often throughput-sensitive, while those operating on streaming data are often latency-sensitive. While the two types of data present subtly different performance constraints, both require significant, scalable computing resources. For example, an image search application operating on stored images may need to scale out depending on the number of images or complexity of the search. Similarly, an application executing a face-detection algorithm on live video may need to scale out if faces are detected and more compute-intensive face recognition algorithms are invoked.

Cloud computing technologies enable many users to share modern computing clusters while providing mechanisms for scaling applications as needed. As a result, researchers in Intel Labs are investigating what challenges arise when leveraging cloud computing technologies in the context of rich data applications operating on either stored or streaming data, and what solutions may address those challenges. This research program includes support of the Open Cirrus* research test bed, development of an open source software stack for operating on stored data, development of a runtime system for operating on streaming data, and exploration of the benefits resulting from integration of optical networks in compute clusters.

Supporting Open Research in Cloud Computing: Open Cirrus*

When considering the topic of cloud computing on large data sets, many questions suggest themselves at an early stage. How should the data be organized and stored? What are the important system software components that enable access to the data? How should those components be organized? What are the appropriate user interfaces? How can the data be processed in the most efficient manner possible?

Answering such a broad array of questions is difficult for a single research organization. Tackling a broad research agenda naturally requires a vibrant research community. To help provide cloud computing resources to this community, Intel, HP, and Yahoo!, in collaboration with the National Science Foundation, sponsored the Open Cirrus* cloud computing testbed [1]. (Open Cirrus is a trademark of Yahoo!, Inc.)

A Unique Collaboration Between Diverse Institutions

The goals of the Open Cirrus project are these:

- Foster systems-level research in cloud computing.
- Encourage new cloud computing applications and applications-level research.
- Collect and share experimental datasets.
- Develop open-source stacks and APIs for the cloud.

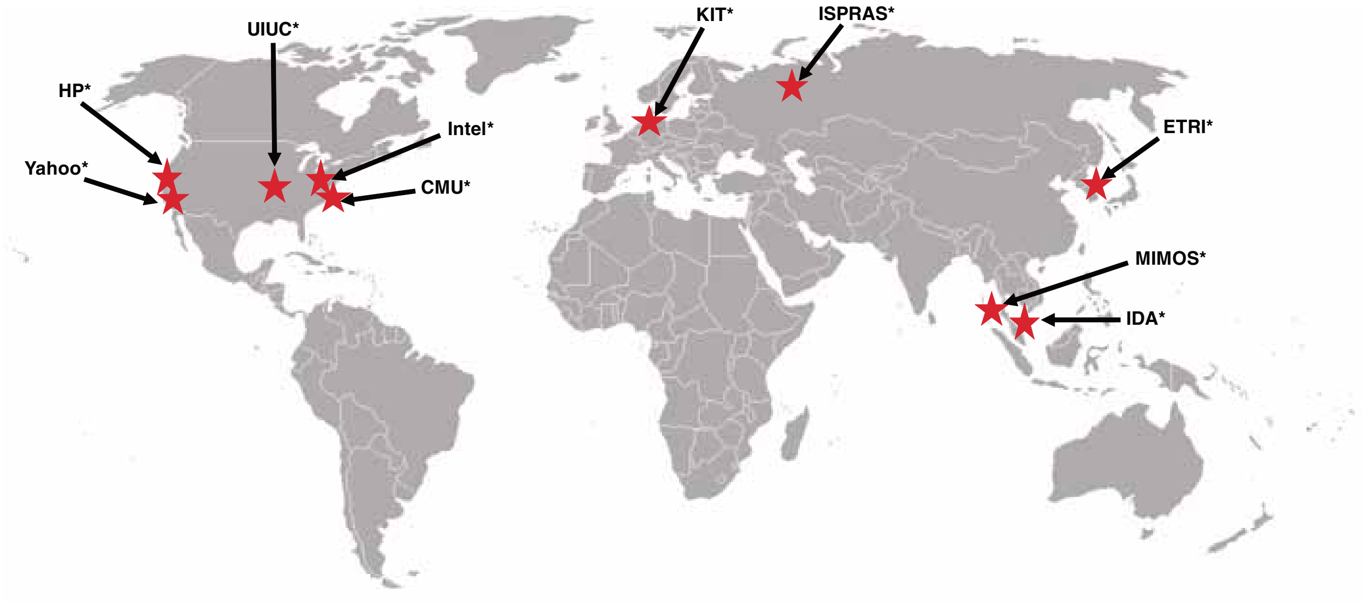


Figure 1: Open Cirrus* Consists of Ten Sites World-wide
Source: Intel Corporation, 2010

To achieve those goals, Open Cirrus provides a world-wide, federated collection of cloud computing sites and a software architecture designed to unify those sites into a coherent platform. The sites (shown in Figure 1), each of which includes a cluster of at least 1,000 cores, are provided by the Open Cirrus collaborating institutions: HP, Intel, Yahoo!, the University of Illinois (UIUC), Karlsruhe Institute of Technology (KIT), the Infocomm Development Authority (IDA) of Singapore, the Russian Academy of Sciences (RAS), the Electronics and Telecommunications Research Institute (ETRI) of South Korea, the Malaysian Institute Of Microelectronic Systems (MIMOS), and Carnegie Mellon University (CMU).

The Open Cirrus* Service Architecture

The Open Cirrus testbed is intended to support research at various levels of the cloud-computing stack from the lowest layers that interact directly with hardware to the highest application layers. However, research in the upper layers requires that lower-level software be available and stable. Consequently, the Open Cirrus community has adopted a common software service architecture; the core services of this architecture are shown in Figure 2. All core software components are open-source projects managed by the Apache Software Foundation. Intel Labs has made significant contributions to the development of Zoni and Tashi, and Yahoo! has made significant contributions to the development of Hadoop and HDFS.

The primary responsibility of the lowest software layer, Zoni, is to partition the cluster into *domains*. A domain is a set of compute servers that are network-isolated from the rest of the servers in the cluster (by the use of VLANs). When users experiment with system software that interacts with key networking components, such as DHCP services, they, or the cluster system administrators, will first use Zoni to create an isolated domain for the experiment; in this way, if the experiment goes awry, it cannot affect the normal operation of the cluster (and other activity in the cluster cannot interfere with the experiment).

> *"The data sets in the cluster are potentially valuable to many users, and consequently, are ideally not stored in virtual machine images."*

Most of the research that does not interact with core networking services, however, will take place in the *primary domain* of the cluster. This domain is considered to be for production use. Experiments in the primary domain are isolated from other activity by operating in virtual machine environments, and the virtual machines are managed by a cluster management layer such as Tashi. Tashi enables users to rapidly deploy virtual machine instances in the cluster by specifying attributes of the virtual machine (such as number of processors and the amount of memory) as well as the software that should run within that virtual machine.

However, the data sets in the cluster are potentially valuable to many users, and consequently, are ideally not stored in virtual machine images. Instead, the Open Cirrus core services include a cluster file system that resides beneath the virtual machine layer. In this way, data stored in the cluster file system are accessible from any of the virtual machines operating in the primary domain. After evaluating many cluster storage options, the Hadoop File System (HDFS) best fit the needs of Open Cirrus.

By leveraging the virtual machine layer, the cluster administrators can provide any number of application-level services. The Open Cirrus software service architecture explicitly suggests one such application runtime: the Hadoop map/reduce framework. This framework is particularly suitable for enabling cluster users to process data stored in the cluster file system.

Naturally, the utility of these clusters would be quite limited if they only hosted the development of these core services. Fortunately, many of the cluster users are not involved directly in research on cloud computing; instead, they simply use the Open Cirrus clusters as computing resources in the course of conducting research in some other field. This use of the cluster is welcome and encouraged, because these users provide a realistic context for evaluating the system software by providing authentic data-rich workloads.

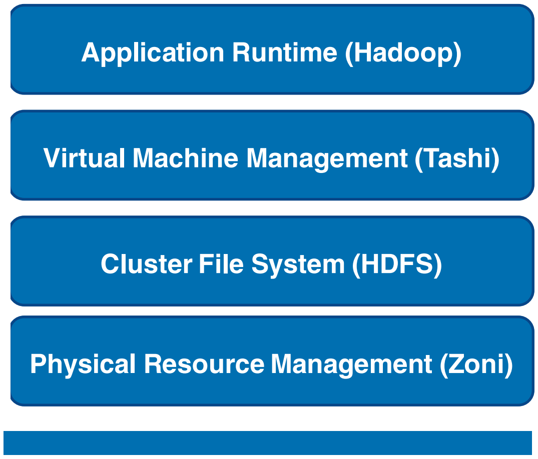


Figure 2: Open Cirrus* Software Architecture Core Services
Source: Intel Corporation, 2010

Infrastructure for Operating on Big Data in Compute Clusters

Many of the most interesting data-rich applications are those that compute on large data sets. These "Big Data" applications, beyond simply operating on data that is big, are also typically *data hungry* in that the quality of their results improves with the quantity of data available.

As a result, one of the challenges associated with these applications is constructing a data storage system that scales in both capacity and delivered throughput. That is, as more data become available, or more throughput is required, administrators need the ability to expand easily the infrastructure that stores the data as well as the computing resources that will operate on it. Fortunately, these applications tend to be easy to parallelize. Further, they tend to operate on large data objects, and as a result, they are typically limited by the delivered bandwidth—not the seek performance—of a storage system.

"Without careful engineering, however, a commodity cluster network can become the bottleneck for big data applications."

Therefore, commodity disk-based, cluster hardware, deployed at scale, is a good fit for supporting big data applications. For example, the Open Cirrus cluster at Intel Labs Pittsburgh consists of a modest 200 servers; yet, it provides more than 1300 computing cores and over 600 TB of disk storage—enough to accommodate many big data problems. Here, each server acts as both a storage and compute server. Without careful engineering, however, a commodity cluster network can become the bottleneck for big data applications.

"Can we deliver application performance that matches the throughput provided by the storage devices without building an expensive cluster network?"

For example, the core networking infrastructure of the Intel Open Cirrus cluster consists of commodity 1-Gbps Ethernet components. There are approximately 10 racks with 15–40 servers per rack. Each rack has a top-of-rack (TOR) switch to which the servers in that rack are connected via 1-Gbps connections; the TOR switches in turn are connected to a central switch via (trunked) 4-Gbps connections. If a parallel big data application were forced to process its stored data by transmitting them all between racks, this central switch would present a bottleneck, and the maximum throughput would be 10 * 4 Gbps = 40 Gbps (5 GB/s). However, the data objects are actually stored on over 700 magnetic disk drives, each of which can deliver close to 50 MB/s, for a theoretical throughput of approximately 35 GB/s. If the storage devices were solid state devices, such as the Intel® X25-M which can sustain 250 MB/s, rather than magnetic media, the theoretical throughput could be even higher (175 GB/s).

Of course, upgrading the network (to 10-Gbps Ethernet components, for example) could reduce the disparity between the throughput delivered by the storage devices and that provided by the network, but a reasonable question to ask is this: can we deliver application performance that matches the throughput provided by the storage devices without building an expensive cluster network?

"If big data applications can be written such that each block of their data is consumed on the server where it resides, the amount of data sent over the network can be drastically reduced."

Location-aware Storage Systems

If big data applications can be written such that each block of their data is consumed on the server where it resides, the amount of data sent over the network can be drastically reduced. In this case, significant throughput can be delivered to the application without requiring a high-throughput network. In fact, the application performance should approach the throughput delivered by the storage system, which is the limiting factor in many big data applications, regardless of network performance.

This observation was one of the motivating factors for runtimes, such as Hadoop, that are based on the map/reduce paradigm. The assumption here is that the application's (big data) input can be organized into large blocks, say 64 MB, and that these blocks need not be processed in any particular order. With this assumption, the data can be processed by sending the computation to each of the servers containing data blocks that need to be processed and by operating on the blocks on that server in whatever order is most efficient.

The key here is that the cluster file system, such as HDFS, exposes location information describing where the data objects are stored; thereby enabling applications (or application runtimes) to be location-aware. In other words, the application task running on some server, N, can query the file system to determine which data blocks are stored on server N.

Virtualization

By enabling applications and runtimes to query the file system for data location information we enable the application to consume data in place, avoiding network transmission—provided the application and file system both understand the same representation of location information. This common understanding is not necessarily found in a system such as the Open Cirrus software architecture, shown in Figure 2, because the application executes in a virtual environment while the file system does not. In such a scenario, an application instance running inside a virtual machine with IP address, *A*, generally will be unable to determine whether or not it is running on the physical host with IP address, *B*, which contains a desired data block.

"The key here is that the cluster file system, such as HDFS, exposes location information describing where the data objects are stored."

To solve this problem, researchers from Intel Labs collaborated with a team from Carnegie Mellon University to develop two additional cluster services [2]: the Data Location Service (DLS) and the Resource Telemetry Service (RTS). The DLS provides a standard interface for applications to query a cluster regarding the location of a data block regardless of the file/storage system storing that data block.[1] Typically, the location response will be a hostname or IP address (or set of such locations if the block is replicated). The application may still be unable to interpret the location information if, for example, it is executing in a virtual environment. The RTS was designed to provide the missing information by being a single clearinghouse for cluster-wide location information. Applications access the RTS through a general interface that provides the relative distance between two location identifiers. Applications, such as Hadoop, can use the RTS information to assign the processing of data blocks to particular compute tasks in such a way that data transfers over the network are minimized. Further, the application can make this assignment in the presence of virtualization and without having detailed information regarding the operation of the underlying file system.

---

[1] So far, we have only considered HDFS, but other cluster file systems, such as Lustre, PVFS, or pNFS, are also possible.

Energy Efficiency

Operators of cluster computing systems are concerned not only with the performance of these systems; they are also concerned with their operational cost, particularly in the area of power consumption. A common goal is *power proportionality*; i.e., the energy consumed by the cluster should be proportional to the work completed. For example, if the processing of 100 TB of data consumes 40 kilowatt-hours, the processing 1 TB of data would ideally consume 400 watt-hours.

"Operators of cluster computing systems are concerned not only with the performance of these systems; they are also concerned with their operational cost, particularly in the area of power consumption. A common goal is power proportionality*; i.e., the energy consumed by the cluster should be proportional to the work completed."*

An important implication of power proportionality on today's systems is that, if there is not enough work available to keep the entire cluster busy, some servers should be powered down (or placed in a low-power state). However, because cluster file systems such as HDFS distribute their stored data across many servers in the cluster (often randomly), powering down servers when their processing capability is not needed may be impractical; such actions may render the data stored on those servers unavailable. This is a limitation even when the data are replicated; if every block is replicated *K* times in the cluster, and the blocks are placed randomly, it may be impossible to power down more than *(K-1)* servers without some block becoming unavailable.

One solution to this problem is to modify the random placement algorithm for data blocks. Researchers at Intel Labs, in collaboration with colleagues from Georgia Institute of Technology and Carnegie Mellon University, have designed such a placement strategy [3]. The central idea is to organize the storage servers according to the *equal*-work layout policy wherein the servers are divided into *K* pools, where *K* is the replication factor. Within a pool, blocks are placed on servers such that, for whatever number of servers is active, the servers can process equal amounts of work.

Processing Streaming Data with SLIPstream

Different technologies are needed for streaming data than for the stored data discussed so far. Video understanding and computer-vision techniques are computationally intensive tasks that are becoming increasingly important for a variety of applications. These include both throughput-oriented applications, such as multimedia search and retrieval, and latency-sensitive interactive perception applications such as surveillance, gaming, videoconferencing, and vision-based user interfaces. Enabling these types of interactive perception applications requires not only new algorithms and techniques, but new runtime systems, that optimize latency as well as throughput, and make effective use of parallel computing machinery. The Scalable Low-latency Interactive Perception on streaming data (SLIPstream) research project seeks to develop the APIs, runtimes, and tools that are needed to create and effectively execute latency-sensitive streaming applications on parallel and distributed computing resources [9].

"Because cluster file systems such as HDFS distribute their stored data across many servers in the cluster (often randomly), powering down servers when their processing capability is not needed may be impractical."

A Runtime for Streaming Perception Applications

A major effort in the SLIPstream project is to develop a set of APIs and a runtime environment to rapidly develop and deploy streaming perception applications on parallel hardware. Such applications, which employ computer vision techniques to understand the environment (such as detecting when a user is visible in a video stream) or interact with a user, pose some interesting challenges. In particular, they are very computationally demanding, require very low latency in interactive settings, and often have time varying and data-dependent execution characteristics. On the other hand, they generally have characteristics that make them amenable to parallel execution: they are often structured as a series of processing steps or transformations on a stream of data items (e.g., image frames), the items are often independent or nearly so, and even processing within a frame may be decomposable into smaller units. A developer with a good understanding of the algorithms involved and with expertise in writing parallel applications should be able to make effective use of parallel hardware for perception applications. The real challenge is enabling the computer vision and video analytics domain experts, who may not have expertise in parallel application development, to write and run their code on clusters of multicore machines.

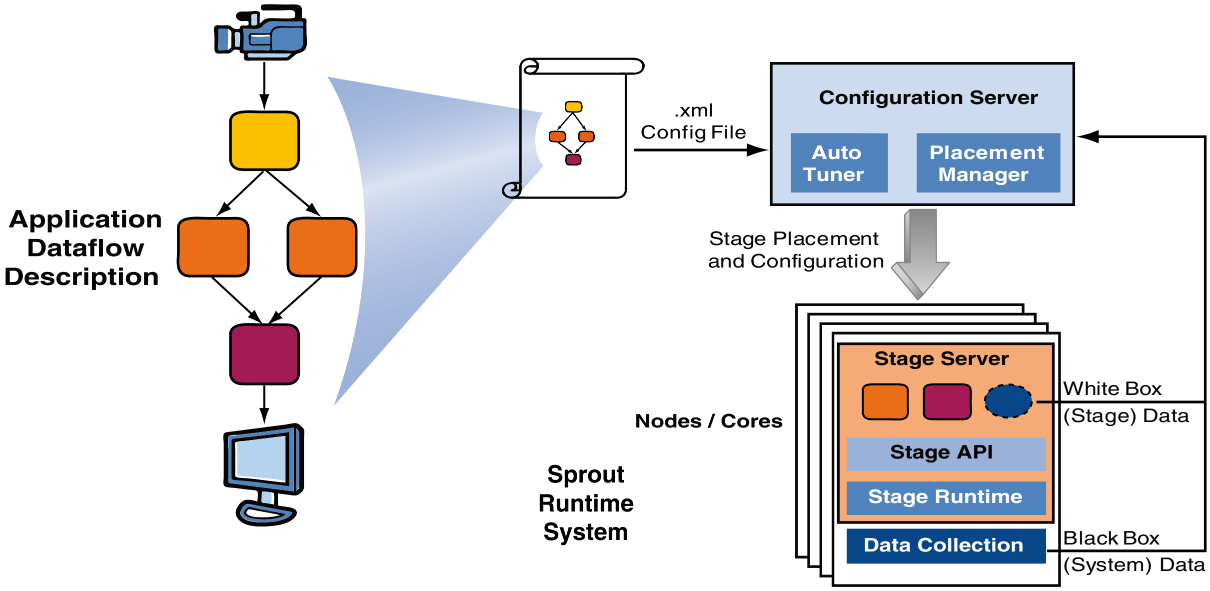


Figure 3: Sprout Architecture
Source: Intel Corporation, 2010

To this end, the SLIPstream team at Intel Labs Pittsburgh has developed the Sprout runtime and APIs, which seek to preserve the ease of writing sequential code, but which expose and exploit parallelism in streaming perception applications. The high-level architecture of Sprout is shown in Figure 3. This system is designed to exploit coarse-grained parallelism in the application, at the level of algorithms and processing steps, rather than fine-grained parallelism at the instruction or loop level. The Sprout programming model requires the application be expressed as a data flow graph, with vertices corresponding to processing steps, or stages, in an application, and the edges corresponding to explicit data dependencies between these stages. This model is suited to video analytics and computer vision on video streams, as these applications inherently have such structure. Sprout provides an easy-to-use API for defining the stages, which essentially wraps the sequential implementations of each algorithm used in the application in an object class that provides strongly typed inputs, outputs, and methods to adjust tunable parameters at runtime. Sprout also provides a library of common stages and stage templates, e.g., round-robin splitters, image tilers, camera capture stages, etc., and it incorporates support for useful data types, e.g., OpenCV image types; user-defined types are easily incorporated into the system. The Sprout stage interfaces hide much of the complexity of writing and running parallel and distributed code by automating the coordination and transfer of data between stages.

"Coarse-grained parallelism is sufficient to effectively exploit parallel computing resources in perception applications."

A human readable configuration file specifies how the application is constructed from the stages, essentially describing the data flow graph of the application. This file is used by the Sprout runtime to orchestrate the deployment and execution of the application on multiple cores and machines. The runtime also monitors execution times and latencies, by using both white box (stage data) as well as black box system information. The runtime supports both manual and automatic runtime adaptation and tuning of the application at runtime.

A key aspect of the Sprout runtime, and SLIPstream research in general, is the focus on coarse-grain parallelism at the level of processing stages. One point this research seeks to demonstrate is that coarse-grained parallelism is sufficient to effectively exploit parallel computing resources in perception applications. Breaking up applications into task and data parallel pieces at this level of granularity is easily accomplished by non-experts in parallelization, unlike low-level vectorization and loop parallelization techniques. Furthermore, this alleviates much of the need for explicit concurrency management and the complexity, correctness, and performance issues that it entails. Finally, the mechanisms employed are complementary to low-level parallelization techniques, and experienced developers are free to implement threaded and vectorized stages, as well as employ parallel libraries such as Intel IPP. Effective use of both low- and high- level parallelism for a complex vision task, using SLIPstream, has recently been demonstrated [10].

"Algorithms in these applications typically have many adjustable parameters that can have a significant effect on both the fidelity and computational costs. In addition, the degree to which computations are parallelized may be flexible."

Automatic Tuning of Distributed Streaming Applications

As SLIPstream is primarily concerned with interactive perception applications, ensuring low end-to-end latency in the applications is paramount. Algorithms in these applications typically have many adjustable parameters that can have a significant effect on both the fidelity and computational costs. In addition, the degree to which computations are parallelized (that is, the number of data parallel instances of a stage) may be flexible. If dynamically adjustable, such parameters provide an opportunity to control latency, but how one configures a parallel application to run optimally on a particular set of cluster resources remains an open problem.

To manage latency in streaming applications, an automatic tuning system has been developed that can dynamically adjust application parameters and degrees of parallelization of stages to bound application latency while maximizing fidelity. As applications can have multiple parallel stages and dozens of tunable parameters, the system first uses performance monitoring to identify the critical stages that contribute most to latency. It further reduces complexity by analyzing, in a grey box manner, which set of parameters and degree of parallelization controls affect the critical stages. Online machine learning is employed to learn and dynamically update performance models for the critical stages as a function of parameter settings. These models, in conjunction with the structure of the application provided in the data flow graph, are used to estimate end-to-end application latency. A solver is then used to either minimize latency or, given a fidelity model, select parameters to maximize fidelity subject to a latency requirement. Ongoing research is investigating how to learn such models quickly, and how to trade off time spent refining the models versus making use of them.

"Most computer vision code today is written in either a purely mathematical form, with little regard to implementation efficiency, or it is written with very specific hardware in mind."

SLIPstream Algorithmic Research Efforts

In addition to the systems research, SLIPstream also involves research efforts on perception algorithms and how they are implemented. In particular, most computer vision code today is written in either a purely mathematical form, with little regard to implementation efficiency (other than in the asymptotic complexity sense), or it is written with very specific hardware in mind. The former has little chance of running efficiently on any system, while the latter often takes algorithmic shortcuts and sacrifices accuracy, robustness, or generality to run fast on a target platform. With the availability of a scalable infrastructure and programming model provided by Sprout, we believe there is great scope for developing perception applications between these extremes—applications that do not sacrifice algorithmic properties, yet can be readily parallelized and scaled up with additional hardware to achieve performance goals.

"More data used correctly can mean greater accuracy or robustness in various applications. Processing more data can also allow simpler algorithms to be effective, potentially increasing further the processing capacity."

Parallelized applications on computing clusters provide the opportunity to process much more data than it is possible to process on single machines. This ability to use more data can have significant ramifications on the algorithms and applications. More data used correctly can mean greater accuracy or robustness. Processing more data can also allow simpler algorithms to be effective, potentially increasing further the processing capacity. Finally, by using more data with computation, we can enable entirely new capabilities, such as the synthesis of accurate novel arbitrary views of a scene by merging information from dozens of different camera views.

Accelerating Rich Data Transport by Using Optical Interconnects

"Optical switches have the property of bit rate transparency, *meaning that the data rate of the information being transmitted does not affect the switch performance."*

As mentioned previously, data-center network bottlenecks present a significant impediment to efficiently processing data rich sources, regardless of whether they are stored or streaming. While the optimized software infrastructures we described are designed to extract as much performance as possible from traditional data-center networks, innovations in the underlying hardware can also be used to augment existing electrical designs and improve performance. As data centers become increasingly large, with tens of thousands of servers, the communications infrastructure and interconnect become more critical for optimum performance. We seek to leverage the well known high bandwidth capabilities of optical components and transmission over optical fibers to build optical systems that improve data-center performance. In particular, we are investigating optically switched systems and advanced modulation formats [4,5,6,7,8].

While optical network transmission components are already in widespread use, switching still occurs primarily in the electrical domain. As the bandwidth of the links between switching fabrics increases, and more optical interconnects are used, it becomes logical to investigate the possibility of using optical switches, particularly to avoid the costly and power-consuming electrical-to-optical and then optical-to-electrical conversions necessary at the switch interface.

Optical switches have the property of "bit rate transparency," meaning that the data rate of the information being transmitted does not affect the switch performance. This property is very advantageous for scalability; the advantages of optical switching in terms of power/bit and cost/bit increase with increasing data rate.

"In this architecture, a small number of rack-to-rack optical links provide high bandwidth for long-lived network flows, while the traditional electrically-switched network handles low-bandwidth, latency-intolerant communication."

A limitation of optics, however, is that a commercially available equivalent to random access memory does not exist, although there is considerable research into the field of optical buffering technologies. Further, no optical switches are currently available that reconfigure on packet timescales and are practical (having sufficiently high integration and port counts) for data-center applications. However, optical MEMS switches are currently commercially available with relatively high port counts (currently 100s with development underway for scaling)—although they require relatively long reconfiguration times. To compensate for these long reconfiguration times, we have been investigating a hybrid network architecture that augments a traditional packet-switched electrical network with a circuit-switched optical network [4, 5]. In this architecture, depicted in Figure 4, a small number of rack-to-rack optical links provide high bandwidth for long-lived network flows, while the traditional electrically-switched network handles low-bandwidth, latency-intolerant communication.

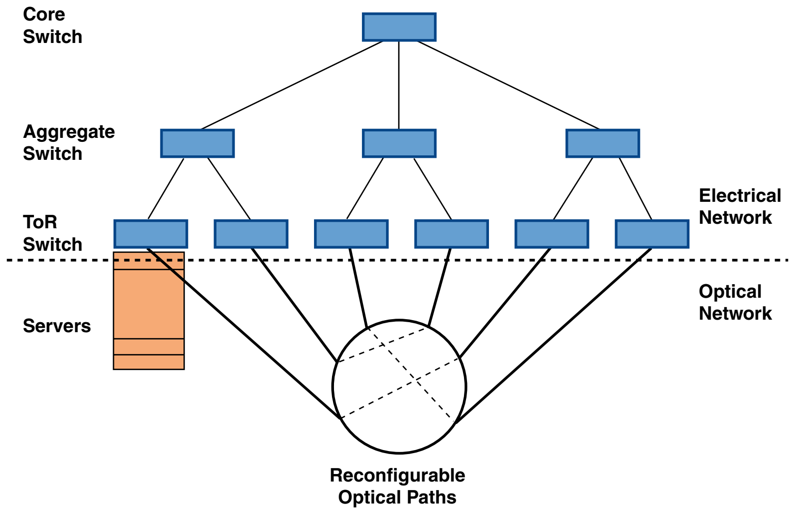


Figure 4: Hybrid Network Architecture: an Optical Network Connects the Top-of-rack (ToR) Switches and Augments the Electrical Network
Source: Reference [5], 2009

This architecture can be effective because the characteristics of many data-center workloads do not require packet time-scale switching speeds, and the hybrid network may relieve bandwidth bottlenecks and improve performance without increasing cost and power consumption inordinately. In the longer term, highly integrated semiconductor-based optical switches with packet scale reconfiguration times should relieve time constraints and offer more flexibility [8].

In parallel, we are developing a technology to increase the bandwidth of the optical interconnect. Interconnects based on vertical cavity surface emitting lasers (VCSELs) and parallel optical fibers are the most commercially-advanced technology and have shown considerable progress in data rate, power consumption, and packaging technologies. Multi-wavelength solutions offer high bandwidth on a single fiber. This research explores the increasingly interesting option of whether digital signal processing and higher-order modulation formats can increase the transmission data rates from a single laser [6, 7]. This technique can be applied to the different kinds of lasers, and it may be particularly useful when using lower-cost, multimode fiber systems that suffer from intermodal dispersion; thereby, limiting the bandwidth-distance product.

Conclusions

As our world becomes increasingly data rich, new technologies are required to support the applications that process data sources—whether they are stored or streaming. In both cases, cloud-computing technologies provide an infrastructure that enables a large number of users to process shared data sets. However, the bandwidth and/or latency requirements of these applications dictate that special care must be taken when designing systems for these applications. Research undertaken within Intel Labs is beginning to discover the technologies needed to provide high-performance computing on data-rich sources.

References

[1] Roy Campbell, Indranil Gupta, Mike Heath, Steve Ko, Michael Kozuch, Marcel Kunze, Thomas Kwan, Kevin Lai, Hing Yan Lee, Martha Lyons, Dejan Milojicic, David O'Hallaron, and Yeng Chai Soh. "Open Cirrus Cloud Computing Testbed: Federated Data Centers for Open Source Systems and Services Research." *HotCloud*, June, 2009.

[2] Michael A. Kozuch, Michael P. Ryan, Richard Gass, Steven W. Schlosser, David O'Hallaron, James Cipar, Elie Krevat, Julio López, Michael Stroucken, and Gregory R. Ganger. "Tashi: Location-aware Cluster Management." *First Workshop on Automated Control for Datacenters and Clouds*, *(ACDC)*, 2009.

[3] Hrishikesh Amur, James Cipar, Varun Gupta, Gregory R. Ganger, Michael A. Kozuch, and Karsten Schwan. "Robust and flexible power-proportional storage." Accepted to *Symposium on Cloud Computing (SOCC 2010)*, June 2010.

[4] Madeleine Glick, David G. Andersen, Michael Kaminsky, and Lily Mummert. "Dynamically reconfigurable optical links for high-bandwidth data center networks." *Conference on Optical Fiber Communication 2009 (OFC 2009)*, 22-26 March, 2009.

[5] Guohui Wang, David G. Andersen, Michael Kaminsky, Michael Kozuch, T. S. Eugene Ng, Konstantina Papagiannaki, Madeleine Glick, and Lily Mummert. "Your Data Center Is a Router: The Case for Reconfigurable Optical Circuit Switched Paths." *Hot Topics in Networks* (HotNets-VIII) October 22-23, 2009, New York City, NY.

[6] Yannis Benlachtar, Philip M. Watts, Rachid Bouziane, Peter Milder, Robert Koutsoyannis, James C. Hoe, Markus Püschel, Madeleine Glick, and Robert I. Killey. "Real-Time Digital Signal Processing for the Generation of Optical Orthogonal Frequency Division Multiplexed Signals." Accepted to the *IEEE Journal of Selected Topics in Quantum Electronics*.

[7] Yannis Benlachtar, Philip M. Watts, Rachid Bouziane, Peter Milder, Deepak Rangaraj, Anthony Cartolano, Robert Koutsoyannis, James C. Hoe, Markus Püschel, Madeleine Glick, and Robert I. Killey. "Generation of Optical OFDM Signals using 21.4 GS/s Real Time Digital Signal Processing." *Optics Express*, Volume 17, Issue 20.

[8] Howard Wang, Ajay S. Garg, Keren Bergman, and Madeleine Glick. "Design and Demonstration of an All-Optical Hybrid Packet and Circuit Switched Network Platform for Next Generation Data Centers." *Optical Fiber Communication Conference*, 2010.

[9] Padmanabhan Pillai, Lily Mummert, Steven Schlosser, Casey Helfrich, and Rahul Sukthankar. "SLIPstream: Scalable, Low-latency Interactive Perception on Streaming Data." In the *Workshop on Network and Operating Systems Support for Digital Audio and Video* (NOSSDAV), June, 2009.

[10] Ming-yu Chen, Lily Mummert, Padmanabhan Pillai, Alex Hauptmann, and Rahul Sukthankar. "Exploiting Multi-level Parallelism for Low-Latency Activity Recognition in Streaming Video." In *Proceedings of the ACM Multimedia Systems Conference*, February, 2010.

Author Biographies

Michael Kozuch is a Principal Engineer and Manager of the Systems Research and Engineering team for Intel Labs Pittsburgh. His interests lie predominantly in the areas of computer architecture and system software. He received a PhD degree from Princeton University in 1997 and a BS degree from the Pennsylvania State University in 1992, both in Electrical Engineering. His e-mail is michael.a.kozuch at intel.com.

Jason Campbell is a Senior Researcher at Intel Labs Pittsburgh. His research interests include sensor networks, autonomous robotics, distributed systems, embedded computing, and computer vision. Previously he served as a senior executive in several startup companies, including as Chief Technical Officer of Zack Systems and as Vice-President of Technology for Strategic Economic Decisions, Inc. Jason has worked as a manager, programmer, and consultant with a wide variety of firms in the fields of high-technology, finance, economic forecasting, commercial printing, and manufacturing. He holds a BS degree in Electrical Engineering from Stanford University. His e-mail is jason.d.campbell at intel.com.

Padmanabhan (Babu) Pillai joined Intel Labs Pittsburgh in September, 2003, after finishing his PhD degree in Computer Science and Engineering at the University of Michigan, where he worked with Professor Kang Shin at the Real-Time Computing Lab (RTCL). He received his MS degree from the University of Michigan in 1999 and a BS degree in Electrical and Computer Engineering in 1996 from Carnegie Mellon University. His e-mail is padmanabhan.pillai at intel.com.

Madeleine Glick is a Principal Engineer at Intel Labs Pittsburgh and an Adjunct Professor in the Department of Electrical and Computer Engineering, Carnegie Mellon University. Her research interests focus on optical interconnection networks, including signal processing and coding for data centers and high-performance computers. She is an Associate Editor of the *IEEE Photonics Technology Letters* and the *Journal of Optical Communications and Networking*. Madeleine is a Fellow of the Institute of Physics and on the Board of Governors of the IEEE Photonics Society. She received her PhD degree in Physics from Columbia University, New York, in 1988. Her e-mail is madeleine.glick at intel.com.

Copyright

THIS PAGE INTENTIONALLY LEFT BLANK