Preface

Lin Chao
Editor
Intel Technology Journal

"Women in Technology at Intel" is the theme of the Q3, 2001 Intel Technology Journal. Key women engineers and researchers at Intel have contributed papers to this issue. Technology in itself is a neutral tool and carries no particular gender bias, and in fact, perhaps the computer, and the Net as its main form of communication, could be described as the ideal gender-neutral environment. At Intel, one out of every five women is in a technical job, and for younger engineers the ratios are higher. In this issue we feature a few of the many women of science and technology at Intel Corp. These women have demonstrated ability and vision, and their papers showcase the breadth of their influence.

The first paper merges technology with anthropology. It describes ethnographic research undertaken in Western Europe to try to determine how people in Western Europe occupy their domestic spaces, what technologies are present in those spaces, and how they are used. The second paper looks at indoor wireless communications and proposes a strategy to deal with two of the main challenges in the use of unlicensed frequency bands: coexisting with other users and extracting the greatest capacity from systems that are both power- and band-limited. The third paper looks at the evolution of automation in microprocessor process development and manufacturing at Intel. And, the fourth paper discusses how the compiler for the Intel® Itanium™ processor was designed and developed in the absence of hardware. It tells how the project was started more than five years before the existence of system hardware, and it describes the role of simulators in developing the compiler.

Women at the Frontiers of Technology

By Dr. Carmen Egido
Director of the Applications Research Labs
Intel Corp.

On a worldwide scale, women are the fastest growing group of purchasers and influencers of technology purchases. It therefore makes logical business sense to involve women in all facets of technology product development and innovation. Given that innovations mirror the viewpoints, values, and desires of their inventors, the nature and impact of new technology products depend on the extent to which diverse needs are sought and understood. It is our diverse workforce that enables Intel to provide the most innovative and creative solutions to meet the needs of our rapidly changing and increasingly global markets.

In 1982, a year before I received my Ph.D., the National Science Foundation (NSF) published the first Congressionally mandated biennial report on the status of women and minorities in science and engineering. With the publication of the 10th such report in 2000, we can observe what progress has been made over the last 20 or so years. Since the first report, the relatively small percentages of women and minorities attaining science and engineering degrees, their concentration in specific fields, and the lower percentages of women in senior positions in both industry and academia still hold. However, progress has been made for women in a number of areas. For instance, the numbers and percentages of women completing high school and going on to complete bachelor's, master's, and doctoral degrees have increased. Close to half (47%) of all science and engineering bachelor's degrees, 39% of the master's degrees, and 33% of the doctoral degrees were awarded to women, with an increasing percentage of bachelor's degrees in all science and engineering fields except mathematics and computer science. Similarly, the 2000 report finds that younger women employed in science and engineering careers are as likely as men to name management as a primary or secondary work activity. However, older women who would be holding the more senior positions are less likely than men to be involved in management. Finally, similar to the first NSF report, the latest report finds that women are more likely than men to work part time and be underemployed.

The most serious concerns emerge when we look at the numbers of women in mathematics and computer science. The number and percentages of bachelor's degrees granted to women declined from 1984 to 1996 (37% to 28%), and the drop in numbers was significantly steeper than for men. The figures for master's degrees in computer science followed a similar though not as pronounced pattern. Women received a higher percentage of doctoral degrees in computer science in 1996 than in 1984 (15% as compared to 12%).

It is these very concerns that have motivated Intel to place a major emphasis on supporting educational initiatives that encourage women to enter science and engineering careers, as well as increasing its focus on growing the number of technical women in the middle to senior classifications as a means of diversifying the feeder pool of candidates for executive management positions. I am delighted to report that Intel achieved a 7% increase in the numbers of women in these classifications in the year 2000 alone.

This issue of the Intel Technology Journal celebrates and affirms the contributions of women scientists and engineers to diverse aspects of research and development at Intel. I believe that the set of papers contained in this issue, all authored by women, will give the reader insight into the remarkable breadth of issues being addressed by women engaged in technology development at Intel.

Looking Across the Atlantic: Using Ethnographic Methods to Make Sense of Europe

Genevieve Bell, People and Practices Research, Intel Architecture Labs

Index words: design ethnography, human-centered product innovation, Europe, domestic spaces, non-traditional research methods

ABSTRACT

Ethnography is a form of anthropological practice. It is both a methodology and a perspective. Through ethnography, we attempt to generate holistic accounts of cultures and peoples. At its core, ethnography relies on "participant observation," i.e., the notion that you learn by doing and by watching, and by the interplay of those two roles. This means that most anthropologists (and those from other disciplines who use ethnographic methods) do field work. They spend time in and with the cultures and peoples they are studying, engaging with the people around them, participating in every-day life, and attempting to make sense of the patterns of that culture.

In 1999, Intel researchers conducted ethnographic research in five Western European nations. We visited 45 households across small towns, larger cities, and major metropolitan centers. Our research tried to make sense of how people occupied their domestic spaces, how those spaces were embedded within the broader community, and what technologies were present in those spaces and how they were used.

INTRODUCTION[1]

American anthropology has always been preoccupied with the notion of *culture*–how to describe it, document it, theorize it, and make sense of it. For the most part, anthropologists focus their critical gaze outside western culture, taking as their object of study all manner of unexpected peoples, places, and rituals. While such work often informed policy decisions and various forms of governmental and non-governmental interventions, anthropologists sought an academic audience for their insights.

Running in parallel to these academic inquiries, and sometimes intersecting them, has also been a tradition of applied ethnographic work, with anthropologists giving guidance to the public and private sectors in the US and abroad. With a few notable exceptions, most practitioners remained within academic institutes. Recently, however, there has been a trend toward hiring anthropologists (and others who practice ethnography) in the public and private sectors.

[1] This paper follows several important ethnographic conventions, which differ significantly from traditional scientific approaches. Firstly, this paper is based on qualitative research with small non-random samples of European households. Secondly, this paper is written in the first person. For almost a century, anthropological texts have been authored in the first person, reflecting in part the discipline's tradition of reflexivity, but also acknowledging the fact that the work is not "objective." Thirdly, because most ethnographic accounts include descriptions of living cultures and real peoples, anthropologists tend not to use people's real names, or give other identifying information. In this paper, I have changed people's names to ensure their privacy. And although many of these interviews were not conducted in English, I have quoted from people in translation.

In most of their early incarnations, ethnographers working in industry focused on studying work places and work practices, providing insights into the ways in which knowledge was constituted and transmitted, and into the functioning of many different kinds of organizations (1). Over the last decade, there has been a small but growing presence of industry-located anthropologists and ethnographers turning their attention to consumer culture and non-work place practices. They join the ranks of human factors engineers, usability professionals, market researchers, and all manner of other non-traditional qualitative research workers interested in understanding peoples' behaviors and cultural practices.

For the last five years, ethnographic research techniques have been influencing technology development and innovation at Intel. The largest concentration of ethnographers can be found within People and Practices Research. This research group is comprised of a number of social scientists (trained in anthropology, psychology, communications), and designers and developers. Members of the group have conducted fieldwork in the US, Western Europe, Asia, and Latin America, spending time in people's homes and businesses, as well as in a range of other social and work spaces.

As a member of this team since 1998, I have participated in research projects in a number of locations outside of the US. In this paper, I focus on fieldwork conducted in Western Europe in the summer and fall of 1999. This work was Intel's first attempt to conduct ethnographic research outside of the US, and we developed and refined appropriate field methods for this research.

This fieldwork was prompted by an internal request from an Intel product group to determine the characteristics of home life in Western Europe relevant to the design of consumer computing products and services. Initially, the project focused on Germany, France, and the UK and was later expanded to include Spain. It was hoped that our research would focus the discussion on Europe as a critical non-US geography, and that it would supplement available market research with a sense of the household rhythms and cultural patterns of Europe. There was also a sense that getting smart about European life in this way might help frame or contextualize future product development.

As a result of this research, we identified four critical domains of cultural significance in Western Europe: (1) togetherness, (2) media experiences, (3) consumption patterns, and (4) life outside the home. In this article, I explore these domains and provide examples from our fieldwork. These domains allow us to make comparative assessments across cultural traditions and also explorations within them.

Appropriate Methodologies

This project represented a large undertaking for our group. Since it was established in 1995, People and Practices Research has been charged with the task of understanding people and their daily practices with an eye to finding new users and new uses for technology. We spend time in domestic and urban spaces, hanging out with people as they go about their daily lives. We attempt to translate insights about peoples' behavior into product concepts, technology innovations, and strategic long-range planning. Our research portfolio includes field studies with teenagers, recent retirees, empty nesters, and families with small children, as well as cross-demographic examinations of such issues as health, shopping, mobility, travel, leisure, and broadband. We have also turned our attention to different kinds of spaces: the home, small businesses, schools, the mall, etc. In all of our work, we attempt to understand peoples' experiences holistically, rather than just in interactions with technology.

Making use of ethnographic methods in non-academic settings involves both a reliance on participant observation, as well as the development of new ways of getting at "cultural contexts" (3, 4). It is important to remember, however, that the ethnographic methods are a way of getting at what people are thinking, and it is that attention to people's real desires and thoughts that makes ethnography a useful approach for the design and development of technology.

At its best, ethnographic research helps convey an experience, a sense, a feeling, a glimpse, or a window into another world. It is a way to look into people's lives that follows their own stories and interests. It is also a way of talking about deep cultural patterns that implicate everything that people do. Knowing these stories, interests, and patterns makes it possible to design and develop products and services that fit (intuitively) into people's lives.

Design ethnography is an emerging discipline that draws on many of the theories, practices, and methodologies of anthropology (as well as other social science disciplines such as psychology, sociology, and communications). Its practitioners are located within a variety of industry contexts including manufacturing, marketing, and high technology. Design ethnography shifts the focus from attempting to create holistic representations of entire cultures to generating accounts of specific social practices or social groupings. Design ethnography disrupts the assumption that people are just consumers: "they are social beings, people with desires, wishes, needs, wants–some articulated, some unrecognized." (5)

As such, it relies on a rich and nuanced understanding of what people do, what they say, and what they think.

Traditional ethnographic research takes the long view: getting to know a place, a people, a culture over months, years, or sometimes, even a lifetime. It is about acquiring a sense of subtleties and nuances, about the smallest details and about the ways all such details piece back together into a coherent or contested whole. Clearly, in an industry context, you don't have the luxury of the traditional long view, and periods of research are considerably shorter and far more focused; however, we still have a persistent interest in understanding the subtleties and complexities of social practices. This creates methodological and theoretical challenges.

Anthropologist George Marcus has suggested, that in framing research across multiple sites, you should establish or identify an appropriate vector to cross-cut the sites (2). The vectors he identifies as appropriate are deceptively simple: people, a "thing," a metaphor, a plot, story, or allegory, a particular life or biography or a conflict. In this project, we chose to pay attention to a "thing" – the home.

This project required that we pay attention to the home not only in a broad European context, but also with reference to the specificities of five very distinct national traditions and sets of cultural practices. So we developed a method of identifying and tracking important cultural values. This indexing system relied on spending time in at least three different kinds of sites within each country. We started by spending at least a week (sometimes two) in a small town, so that we could get a sense of the daily and weekly flow of life, (i.e., weekdays, weekends, school days, holidays, church days, market day, etc.) in a setting that we could easily manage. We moved from that small town to a larger population center, and then a large metropolitan center (Paris, Berlin, London, and Barcelona). We believed that if a behavior existed in smaller, and frequently more culturally conservative towns, and it persisted all the way into big cities, then it was an important behavior or activity. This indexing helped us discover important cultural patterns and spaces.

Although our European research employed a range of standard ethnographic methods: observation, participation, and interviews, as well as an attention to appropriate cultural contexts, we were aware of the need to develop research methods that redressed some of the issues that arose with such short field stints. First and foremost, we relied on the fact that the consultants who worked on this project were already embedded within particular communities, and we leveraged those connections to jump-start the process of gaining access and acceptance. This sort of "friends and family recruit" meant that we did not have a random sample of households in the countries we visited, but it did mean that we had a remarkable degree of social access. It also had the advantage of allowing us to be at multiple points within an extended social network, watching it in operation.

We tried to do everything that local people do to get a sense of what it is like to be there as a native, not as a tourist. We spent time in people's households; photo inventoried them, and tried to get a sense of daily life by asking people explicit questions and by following them through their days and their activities. We spent time in public spaces and public places watching and observing what people do there. We talked to everyone who would talk to us and asked lots of questions about what life is like and what people do there. We asked questions about shopping, catalogs, e-Commerce, travel, newspapers, TV, computers, phones, children, health, space, design – about life in general.

RESEARCHING EUROPE

In 1999, mobile phones were rapidly proliferating across Europe, and text messaging was becoming a preferred form of communication. In some countries, including several we visited, the number of mobile phones was eclipsing the number of landlines. Alternative Internet access devices were beginning to spring up and make money including Web-kiosks, interactive TV platforms, and other models of non-domestic Internet use. It was increasingly clear, from a US vantage point, that Europe was doing something different; the intersections of emerging technologies and social practices were occurring at different points, with different consequences, and with different cultural elaborations.

We theorized that these patterns of use might reflect underlying cultural practices. Getting at what some of these practices might be, and what kind of impact they could have on future technology development and consumption seemed to be an appropriate place to test our ethnographic research methods. Therefore, in the summer and fall of 1999, we spent a total of 16 weeks in Italy, the UK, Germany, France, and Spain visiting 45 households, interviewing the occupants, and taking inventory of the contents. We spent time in 15 urban centers trying to get a sense of what life was like.

For fieldwork in Germany, France, and Spain we worked with external consultants. The external consultants, who served as cultural and linguistic translators, were trained ethnographers, and we relied heavily on them to guide us in selecting field sites and in recruiting participants.

For my colleagues, the European research constituted return trips to childhood family homes, recalled halcyon high-school trips and college years spent abroad or constituted a return to previous sites of research or places of vacation. For me, this fieldwork represented my first encounters with Europe. I stood in an Italian kitchen drinking coffee and listened to a day's instructions unfold. The Europe I would find over the next five months was a patchwork of currency, dial tones, electrical outlets, newspapers, television news, bathrooms, bedding, breakfasts, the right water for the right meal, food, sounds, and color. It was the famous and mundane sights of elsewhere seen from profoundly local vantage points.

Our fieldwork in May of 1999 began with a short pilot study in northern Italy. (Fieldwork in Italy allowed us to rehearse and refine our methods and approaches.) We were concerned that research methods developed for the US might not work in Italy, and indeed, many didn't. We commenced research in a small town in northern Italy, located about 45 kilometers north of Venice. We spent ten days there, visiting a total of nine households in the surrounding area and three small businesses. We also spent time in Treviso, a nearby larger commercial center, and in Venice. This would become the format for the rest of the project. In each country, we spent time in three distinct sites: a small town, a larger population center, and a major urban metropolis.

After our initial visit to Northern Italy, we returned to the US and began planning for further visits to Western Europe later in the summer. In July, we spent a week in Scotland, in and around Glasgow, and then two weeks in and around London. We inventoried ten households, including a close knit, working class, social network in Glasgow and a household of diasporic Indians in London. We also spent time in a number of different retail and commercial spaces.

Immediately after the UK fieldwork was completed, we flew to Germany and spent two weeks in southern Germany, and then several days in Berlin. We were able to trace out an entire social network, which included an extended family, high school and university friends, and social acquaintances. We visited ten households, which included a group house in Stuttgart, several households with small children, a single-parent household, and the household of a retired couple. We also interviewed three small-business owners.

We returned to Europe again in early September, spending time in Brittany, Nantes, and Paris. In Brittany, we spent the bulk of our time in a small coastal village, which is a major seaside destination in the summer. We spent time in five households and in one small business. We also visited a number of local markets, retail and commercial districts, and a variety of public spaces. We spent time in Nantes, interviewing two households and two small businesses, as well as spending more time in public markets and Internet cafes. We visited three households in Paris and spent time in a range of public places and public transportation sites.

In late October, we returned again to Europe, this time to Spain, where we spent the bulk of our time in the Basque region of northern Spain. We visited only three households in the Basque region, but spent time with people outside of their homes talking about daily life; we also visited several places of work. After the Basque region, we visited with families in and around Barcelona, interviewing three more households and spending time with people outside their homes.

In total, we interviewed more than 100 individuals and took more than 8,000 digital photos (and limited video), as well as collected many boxes of cultural ephemera and material artifacts (including newspapers, advertising, catalogs, free postcards). But in truth, the fieldwork was the easy part; making sense of all this material and conveying those insights back to product groups and other audiences within Intel was a much bigger challenge.

RESULTS

In thinking about this project and attempting to make sense of some of our findings, I turned to existing ethnographic accounts of Europe, hoping to take advantage of their longer time horizons and deeper understandings of particular cultures and practices. Prior to the 1980s, ethnographic accounts of Europe tended toward studies of villages and small-scale rural communities (6, 7). In the 1980s and 1990s, there have been a range of new theoretical and methodological approaches to Europe; anthropological accounts pay attention to the current complexities of identity and identity politics in Europe, with various moves (both legal and economic) that have served to deformalize national boundaries (8 9, 10). These moves complicate notions of identity, nation, and belonging, as well as community, boundaries and family. However, American anthropology does not have a strong tradition of exploring and examining the complexities of contemporary European cultures let alone of examining the intersections of social practices and emerging technologies in and around domestic spaces.

From our research, it can be said that there are a number of ways in which it is possible to generalize about European life. There is a greater awareness of environmental issues in Europe and a greater willingness to shift and moderate behaviors in order to conserve energy and preserve or renew environmental integrity. There appears to be a stronger sense of the need to balance work and home life–people were not working 60 hours a week (in fact, that is illegal in a number of EU countries) and spending time with family, actively engaging with them, was viewed as a priority. Food, cooking, and eating are also associated with strong rituals and are a source of great pleasure. Public spaces and places seem to have different use patterns and are frequently treated as an extension of the home. People's domestic spaces are, on average, considerably smaller than American homes and occupied according to distinct seasonal rounds. There is a long history of participatory and iterative design in Europe, and choices in domestic appliances reflect a strong appreciation of good design. The phone system (like everywhere else in Europe) is still metered for local calling, so Internet access is not cheap. However, for as much as all of these commonalities might exist, there is no one Europe, and to think about Europe means thinking about a range of different cultural and national traditions.

In approaching Europe, we were especially interested in what people were doing in their homes. We were interested in the "ecology" of the home, that is, the ways in which space was used, and not used, and in the ways that activities went on in and around those spaces. We were also interested in what could be found in those spaces, how people used objects within those spaces, and what people said about what they were doing. Moreover, we were curious about the relationship between the household and the broader community (11, 12).

It also became obvious that a significant portion of people's lives in Europe is conducted outside of the immediate domestic sphere. So in thinking across the domestic and the public sphere, several important cultural domains emerge. All four domains try to focus on people's lived behaviors rather than on technologies. These domains reflect one way to talk about the European material comparatively across cultures.

Domains of Cultural Significance

Here I want to talk about these domains both in abstract terms but also with reference to specific examples that arose from our fieldwork.

Togetherness. Notions of family, kinship and community are extremely important in European households. People talked about their social networks and about the kinds of relationships they had, where they took place, what they revolved around and what events were emotionally important to them. These patterns of sociality became ways of framing how people occupied their space, the kinds of activities they engaged in and with whom.

Along with her husband and three teenage children, Rina lives in a small rural village on the outskirts of a larger town. Rina says of herself, "I am not a housewife; I am an architect with three children and a house to clean." Rina's children go to schools in the neighboring towns, and Herbert, her husband, works as an accountant nearby. It is extremely important to Rina that her whole family eats together. The main meal of the day is frequently *mitagessen* (lunch), which Rina always prepares. Every day her husband comes home from the office so that he can eat lunch with her. School lets out in the early afternoon so that kids can eat at home with their parents. Rina says that she really likes it when her family can eat *alles zussamen* (all together) and talk. Meals are the time that everyone is in the same space together. This desire to eat together as a family, as a household, was expressed repeatedly in our interviews in Europe, and it became abundantly clear that meals served as social rituals.

There are other occasions when people gather together. These are often around meals, frequently in public places, sometimes at sporting events, and sometimes for leisure activities. Sometimes people spend time together simply watching television. George and Sara live in a medium-size town between Glasgow and Edinburgh. Married nearly 40 years, Sara has retired from nursing and spends her days tending her garden, walking and swimming for her health, and socializing with her friends in the local community center. George is a doctor and still continues to practice medicine every day. When George comes home in the evenings, he sometimes goes up to his study and works on the computer. But most nights, he comes downstairs to their farm-house kitchen where Sara is cooking, paying bills, washing dishes, and watching television in the background. George says, "Why would I want to sit upstairs without her? I spent all day without her; when I come home, I want to be with her." Although they have bigger televisions in the house, the little one in the corner of the kitchen is the one they watch most.

Across Europe, there is a premium on spending time with other people in person, and much energy is dedicated to make such occasions possible. And although there were marked differences in the kind of language people used to talk about their families or other important social units, and the amount of effort expended to ensure face-to-face social time, the idea of "togetherness" was present in all the households we visited.

Media experiences. There are many different kinds of interactions and experiences that people have around "media." (Here, media was understood to include the widest possible range of media types and media content.) Not only were we struck by the range of different media being consumed by individuals across Europe; we were also struck by the amount of content that was being produced. In almost every population center we visited, there were people painting, drawing, sketching, playing musical instruments, and participating in all manner of artistic production. In several of the countries we visited, this sort of activity is supported by the State, and it is actively encouraged by a range of local municipalities.

All of the homes we visited had televisions, but the ways in which television factored in those households and the lives lived within them varied across cultures. They stood in sharp contrast to the ways in which television factors into life in the US. Print media also plays an important role in daily life across Europe, and there is a genuine diversity of publications and content. Paying attention to television, and to all the other kinds of media that exist in people's homes and people's lives provides us with an interesting vantage point on European daily life.

Across Europe, a history of government subsidies means that television programming has developed along very different lines than in the US. In the UK, and elsewhere, people explicitly talk about the television as providing educational and cultural programming. Thus, in making sense of European media habits, it is important to know the history of media outlets in specific countries: are they government controlled or regulated; what is the nature of that regulation; is there compliance; what is the role of non-local content, and so on. Likewise, it is useful to have a sense of the patterns of programming and offerings: when is the watershed hour; how many channels are there; where and when are timetables published; and how is a broadcast day structured.

Mirabella is a single parent. She lives with her three teenage children on a quiet street in a small coastal town in France. She is a journalist and works long but erratic hours. She used to have cable but she unsubsribed when it became clear that her son was watching too much television and not doing enough work. Now most evenings, she gathers her children on her bed, and together they watch a movie on television. In France, at least one television station plays a movie in its entirety every night. There are no commercials and only a brief intermission at around 9:30pm. Mirabella sends her youngest children to bed at intermission and jokes with me that she is preserving the surprise for them. On Wednesdays, she sits down with her children to watch educational programming: during the summer of our visits, this consisted of a series that followed a typical day in different countries around the world.

The households we visited were strewn with all manner of print media: newspapers, broadsides, cooking magazines, catalogs, fashion and entertainment glossies, television guides, pricelists, and newsletters. Rituals of reading and use abounded, with times and locations peculiar to specific families communities, and nations. Mark is a senior partner in a well-respected global consultancy firm. He lives alone in a large penthouse flat in London where he cultivates a lush garden on his balcony. On the weekends he is in town, Mark buys three different Sunday papers: *The Financial Times*, *The Guardian* and *The Independent*. He says, "I like to know what is going on from a variety of different perspectives; reading all those newspapers gives me that." Mark is not alone in this. Many Londoners read more than one daily newspaper. There is a culture that various news and newspapers represent very different political and social positions. Picking a newspaper that tells you what you want and need to know, as well as one that reinforces or re-inscribes your identity is a common practice across Europe.

Consumption habits. Food shopping is a really important part of daily life in most European countries. There is a range of different consumption patterns and a variety of different spaces in which such consumption takes place. Unlike the US, where grocery shopping happens weekly or fortnightly, in Europe, it appears that shopping is conducted almost daily. People frequent a wider range of merchants with smaller selections than their American counterparts, and what constitutes a good deal is very different. It isn't just about saving money; it is about the perception of value (which doesn't just map directly onto money). Value, in turn, is tied up in relationships between the merchant and the customer, in levels of loyalty, trust, and familiarity.

In a number of countries there is a distinct preference for very fresh foods and thus for daily or weekly produce markets. The *marché*, or open-air produce market, is an extremely important part of French daily life. In a culture that places high value on preparing, cooking, and consuming food, the act of shopping for foodstuffs is an important one. Even when supermarkets are readily available, people will still choose to shop in the local market. Patrice lives with her husband Frederico and their two small children, in a tiny village in Brittany. In her early 30s, Patrice has retired from working in the company she and her husband started. She now runs her household and looks after her kids and husband. Like many other European women we interviewed, she expressed serious reservations about catalog shopping, saying "I like to touch and see things before I buy them." And although she lives just a five-minute drive from a town with a large supermarket, Patrice chooses to shop in the local produce markets, which cycle through southern Brittany. It is important to buy what is locally and seasonally available: "it just tastes better," Patrice says.

Shopping is not just about buying things, however, it is also about an extended network of social relations. In Spain, you shop locally, frequenting merchants in your *barrio*. Anna lives with her husband and toddler in a large Basque town near the coast. Her husband is a supervisor in a nearby tire plant and she is taking time off to raise their daughter. After her husband has gone to work, and Anna has tidied up the house and put the bedding outside to air, she goes shopping for the day's meals. She always shops at the same merchants within her *barrio*, the local neighborhood in which she lives. She goes to the butcher, the baker, the green grocer, and the fish merchant. She waits for them to serve her, taking the produce they offer her while exchanging news about her husband and their daughter and listening to stories about what is going on in the *barrio*. Anna says, "I always shop in the same stores; they know me; we trust one another; it has been like this for a really long time." This idea, that shopping is about an extended set of social relations, was repeated across many of our interviews. Shopping connected people to their communities, and it was characterized by forms of reciprocity and inter-dependence.

Life outside of the home. There is a range of significant spaces, activities, and players/forces outside of the European home. In Europe, the contemporary American distinction between home and other places is sometimes less visible. We were repeatedly struck by how much time was spent outside of the home in Europe and by the diversity of behaviors that occur in those "public" spaces.[2] People socialize in gardens, parks, museums, bars, pubs, beer gardens, cafes, markets, tavernas; they socialize along promenades and boulevards and in the plazas. People talk, read, argue, play games, flirt, admire art, create art, listen to lectures, eat, socialize, drink, watch television, politick, dance, swim, walk, and hang out.

When people's domestic spaces are small, these public, or third places function as social nodes. The frequency with which people visit these nodes, and the number of nodes that comprise their routine geography varies with age and life stage. Liz and Jim live in a duplex in a working class suburb of Glasgow. Three years into retirement, they are still settling into a non-working life together, establishing new routines of eating, sleeping, and spending time together. Some things are new: Liz now stays up late to do crossword puzzles and stays in bed til 9 a.m. Jim spends more time in his garden and can play golf twice a week. Some things remain the same: for the last thirty years, they have belonged to a bowling league, and now they go bowling with friends three nights a week. They eat at the bowling club a couple of nights a week with the friends they have made over years of working, taking holidays, and bowling. Along with a local community center, where they recently celebrated their 50th wedding anniversary, the clubhouse at the golf course, and a local pub, the bowling league is an important site of social activity. For their oldest daughter, Mary, her home in Edinburgh is augmented and extended by a tiny caravan she owns in an open field in the highlands. Mary, who works as a nurse, calls the caravan her "bolt hole" and clearly views this space as a retreat from the demands of social activity and daily life. Yet, even this caravan in its remote alpine paddock is embedded within a network of social nodes: there is a local pub and restaurant where the owners know Mary by name.

[2] Ray Oldenburg calls these sites "great good places" or third spaces–neither work nor home. He suggests that such places have a set of characteristics including neutral place without host/visitor connotations, socially inclusive membership, heterogeneous, frequented by, and visited because of, locals; stimulates good conversations, unpretentious in style and in mood, home away from home, etc. (13).

Many European homes appear to exist within a constellation of sites in which social activity transpire. In the UK, it is your "local" pub; in France, there is the *brassiere* and the café; and in Germany, the *bier garten*. In the Basque area of Spain, in which we spent time, it seemed to be less about this constellation of non-domestic spaces, and more about a redefinition of what constitutes domestic spaces. With families spread across multiple households, the *barrio* functioned as an extended domestic sphere and accorded many of the functions of home-like spaces. In the early evenings, people stroll though their barrios. When asked, people say that *el paseo* was as much about participating in the social life of their community, than it was about healthy activity. People let their children play in the *plazas* and talked about the ways in which those spaces were safe, and by some measures, private and controllable. Yet, people's homes were considered to be private, and it was rare that even close friends had visited each other's dwellings. It was far more likely that when they socialized, they did so within the *barrio*, at cafes and tavernas.

This slippage between domestic spaces and social activities is an important facet of life in Western Europe. Non-domestic ecologies provide opportunities for socializing and social activity. They also operate as sites for the consumption and production of media content, as well as significant technology use.

DISCUSSION

This research was conducted in 1999, and there have been significant shifts in technology deployment and adoption in the intervening two years. We have not had the opportunity to go back and revisit the households we interviewed to see what has changed in the last two years as new media technologies and the Internet have continued to spread. Furthermore, because the work was intended to form an ethnographic baseline, not an exhaustive analysis, there are significant gaps in our knowledge. To date, we have not conducted significant research in Scandinavia or Eastern Europe, and both of these areas have different cultural traditions and present different patterns of technology adoption.

Over the last two years, however, this research has continued to linger and influence our work. Arising out of our initial forays into Europe, we identified several themes (and critical social practices) that have continued to shape our thinking around such issues as the work/life boundary, media consumption, the role of television and mobility (14). The ethnographic research also pointed to new product opportunities such as computing in public spaces.

This research helped firmly establish at Intel that there was not "one" Europe and that to think about Europe means thinking about a range of different cultural and national traditions. This has significant implications for marketing strategies, for product development, and for design. What works in Italy might not work in Germany. What works in the UK will surely not work in France. But, clearly there is a need for further research in Europe.

The original internal partner on this project was interested in developing domestic computing appliances. Our research provided explicit feedback on the structural and social affordances of European homes. To design a computing appliance for the kitchen, for instance, involved understanding the different ways kitchens functioned within European homes. In northern Italy, for example, the kitchen was a social hub where entertainment and information content were frequently consumed. In France, by contrast, kitchens are the site of meal preparation, but not content consumption. The small size of many European kitchens also suggested the necessity for devices that did not take up valuable counter space.

CONCLUSION

The domains spelt out above represent one cut of our data, and they help frame cross-cultural conversations, as well as provide some structure to think about and theorize about European identity, albeit one with strong local inflections.

What is perhaps missing from many accounts of the new Europe is an understanding of the role that emerging technologies are playing in shaping and creating these wider contexts. The Europe that we encountered in the summer and fall of 1999 was a place, or perhaps series of places, struggling with competing national, regional, cultural, and historical identities. Individuals that we interviewed pondered out loud about the ways in which their identities were changing or becoming more complicated. The role of information and communication technologies in these new identity politics is not well defined. But it seems to me that there is a significant shift underway, one in which we move from identifying ourselves as part of either the global or the local world to recognizing that we are embedded within a series of local contexts. This idea of translocal identity appears to be mediated through a range of technologies, including mobile phones, computers, interactive television, and a range of other household devices/appliances (including minitel, teletext, ITV). This is clearly an area for further study.

In this paper, I have reflected upon the process of doing fieldwork in non-US geographies for an industry and design audience. Throughout the year 2000, and in the first months of 2001, we have continued to conduct fieldwork in critical non-US geographies, including India, China, Ecuador, and Chile. Our methods continue to change, and our audiences get more sophisticated.

This work, however, is not without its challenges. One of the biggest challenges with this research project was finding ways to make our ethnographic work relevant, both in its specifics and in its more general interventions. It is not enough to convey results from this kind of work, you also have to teach your audience to make sense of, and make use of, the material you are generating. Training people to understand the differences between quantitative and qualitative research makes them better consumers of both kinds of research, but requires an investment of time and resources. However, it is difficult to get traction with ethnographic work when it does not always fit neatly into Intel's research and product development cycle. The material isn't "easy"; it requires time to absorb it and to think about it. The challenge here is to find ways to make the research happen, and to have it resourced and staffed appropriately and ultimately understood as just another part of good business practice.

It used to be easier to think that technology was neutral, that it had escaped the markings of culture and the politics of location. However, as technologies, from the washing machine to the mobile phone and PC, spread out across the globe, it has become harder and harder to ignore the ways in which their use is contextual. The tension here is between the globalization of technology and the localization of its use. The same object acquires different meanings, lives in different parts of our homes, conveys different messages, performs different tasks, and it even satisfies different needs.

In examining Europe from an ethnographic perspective, we were trying to pay attention to these different contexts and different meanings. This kind of knowledge can help inform and guide all manner of product development and innovation at Intel and elsewhere. Clearly, looking at Europe, we need to complicate our understanding of the domestic, identifying the principal locus of social activity, whether it be the kitchen, living room or bedroom, and design accordingly, or with enough flexibility to accommodate those other scenarios of use. (It might be the kitchen; it might be the living room.) Similarly, we need to decenter the domestic: it is no longer enough to just think about the home as the sole site for non-work-related consumption of technology. And, as we talk more and more about pervasive computing, we should do so with a sensitivity to the differing notions of public and private that might pervade that debate outside of the US, and be aware of the possibility that the "home" might not be the starting point or the center of the brave new digital world.

ACKNOWLEDGMENTS

First and foremost, I have to thank all the individuals, families, and households that so generously allowed us access to their lives, who let us into their homes and who laughed when we wanted to take photos of the insides of their fridges. Without their generosity, none of this would have been possible. I am equally grateful to the consulting anthropologists on this project: Heinrich Schwartz, Morgane Danielou, Vincent Lépinay and Steve Ybarrola who let me follow them home and helped translate my stuttering questions into intelligent remarks.

I want to thank my co-conspirators on this project: John Sherry and Tony Salvador–who were willing to be kidnapped from their families and homes for weeks at a stretch, who rose to every occasion, and who taught me a great deal. I also want to thank the other members of Peoples and Practices Research, both past and present, for their continuing enthusiasm and engagement with this project.

This paper has benefited from conversations with countless friends and colleagues, but in particular, I am grateful for the insights of Ben Anderson, Ken Anderson, Diane Bell, Elizabeth Churchill, Susan Dunn, Nicola Green, Christine Hine, Sarah Jain, Joseph "Jofish" Kaye, Scott Luning, Jim Mason, Andrew McGrath, Alan Munro, Michael O'Higgins, Julia Olson, Heather Paxson, Simon Pullman-Jones, Rick Robinson, Malene Skaerved, Don Slater, Nina Wakeford, and Steve Woolgar.

And last, but not least, I am grateful to the Intel product group that helped to fund this research and which found merit in its results.

REFERENCES

1.Eddy, Elizabeth and William Partridge, *Applied Anthropology in America*, Columbia University Press, New York.

2. Marcus, George, E., "Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography," *Annual Review of Anthropology* 24 (1995): 95-117.

3. Bernard, H Russell, *Handbook of Methods in Cultural Anthropology,*. Alta Mira Press, Walnut Creek, California.

4. Trotter, T. Robert and Jean J Schensul, "Methods in Applied Anthropology," *Handbook of Methods in Cultural Anthropology,* Alta Mira Press, Walnut Creek, California.

5. Salvador, Tony, Genevieve Bell, and Ken Anderson, "Design Ethnography," *Design Management Journal* vol. 10, #4 (1999): 9-12.

6. Wolf, Eric R., *Europe and the People Without History*, University of California, Berkeley.

7. Parnam, Susan, *Europe in the Anthropological Imagination*, Prentice Hall,New Jersey.

8. Buraway, Michael and Katherine Verdery, *Uncertain Transitions: Ethnographies of Change in the Post-Socialist World*, Rowan and Littlefield, Lanham, Maryland.

9. Darian-Smith, Eve, *Bridging Divides: The Channel Tunnel and English Legal Identity in the New Europe*, University of California Press, Berkeley.

10. Delanty, Gerard, *Inventing Europe: Idea, Identity, Reality*, St Martin's Press, New York.

11. Birdwell-Pheasant, Donna and Denise Lawrence-Zúñiga, *House Life: Space, Place and Family in Europe*, Berg, Oxford.

12. Pounds, Norman JG, *Hearth and Home: A History of Material Culture*, Indiana University Press, Bloomington.

13. Oldenburg, Ray, *The Great Good Place: Cafes, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community*, Marlowe and Co.

14. Sherry, John and Tony Salvador, Running and Grimacing: the struggle for harmony in mobile work, In Brown, B., N. Green, R. Harper (eds.) *Wireless World: Social, Cultural and Interactional Issues in Mobile Communications and Computing*. NY and London: Springer Verlag.

AUTHOR'S BIOGRAPHY

Genevieve Bell is currently a senior researcher in Intel Architecture Lab's People and Practices Research. She is a cultural anthropologist by training and has conducted fieldwork in Western Europe, Asia, the US, and Australia. She holds a Ph.D. in Anthropology from Stanford University and is working on turning her dissertation into a book. She lives in Portland, Oregon where she is learning to like the rain. Her e-mail address is Genevieve.Bell@intel.com.

Indoor Wireless Communications: Capacity and Coexistence on the Unlicensed Bands

Leslie Ann Rusch
Connected and Extended PC Lab, Technology and Research Labs,
Intel® Architecture Group, Intel Corporation

Index words: coexistence, multiuser detection, IEEE 802.11, CDMA, spread spectrum

ABSTRACT

Ubiquitous computing requires anytime, anywhere connectivity, leading to a spectrum crowded with users seeking reliable, high bit-rate communications in potentially dense geographies, especially in indoor environments. The amount of frequency spectrum available to Wireless Local Area Networks (WLAN) and Wireless Personal Area Networks (WPAN) is limited, with no relief in sight. As Wireless LANs and PANs rise in popularity, they will be required to carry increasingly larger amounts of data for multimedia and other services, while at the same time, overcrowding of these bands can lead to service degradation and undermine the goals of ubiquitous computing.

In this paper, we discuss a proposed strategy to deal with two of the challenges in the use of unlicensed frequency bands: 1) coexisting with other users, and 2) extracting the greatest capacity from systems that are both power and band limited. We first address the coexistence problem qualitatively, explaining the interference scenarios and how the proposed solution works well in this environment. We then address the capacity issue quantitatively with a throughput analysis for the proposed system and the up-and-coming IEEE 802.11a standard for WLANs. We will show how the proposed system offers similar performance to the 802.11a systems at high Signal-to-Noise (SNR), and show how it offers many added advantages, including robustness, to narrow- and wide-band interference.

INTRODUCTION

Indoor wireless communications occur on unlicensed, shared frequency bands. In the US, two such bands are the Industrial, Science, and Medicine (ISM) band at 2.4GHz and the Unlicensed National Information Infrastructure (UNII) band at 5.2GHz. These bands have similar counterparts around the world, with international regulatory bodies working to align bands and regulations (power level, modulation schemes, etc.). Since these bands are unlicensed, they can contain a wide variety of signals such as microwave ovens, RF tags, cordless telephones, Wireless Local Area Networks (WLANs), Wireless Personal Area Networks (WPANs), etc. The growing popularity of WLANs and WPANs will lead to even higher demand for these limited ISM and UNII bands. This situation leads to many challenges that frame our wireless communications research program:

1) How to optimize system throughput given constrained bandwidth and limited power (limited by regulation or by battery capacity).

2) How to coexist in a band with numerous, disparate, and uncoordinated interferers.

We are interested in several research areas holding promise for these challenges, including multiuser detection theory, smart antenna technology, and iterative methods of combining forward error correction with interference suppression and/or space-time coding of multiple antennas. In this paper, we focus on multiuser detection theory and how it contrasts with the current solutions employed in the popular IEEE 802.11 wireless standards.

After a brief introduction, we describe spread spectrum techniques used on the unlicensed ISM band at 2.4GHz to facilitate sharing of the band. The next section describes how direct sequence spread spectrum can be used for Code Division Multiple Access (CDMA). We also describe how multiuser detection can enhance performance of CDMA systems. Next, we describe Orthogonal Frequency Division Multiplexing (OFDM) on the UNII band at 5.2GHz to achieve bandwidth efficiency. In the following section we compare the use of OFDM with higher order signaling and error correction against the use of multiuser detectors to achieve parallelism in a CDMA channel. We find that the proposed system using a multiuser detector provides better coexistence and good system throughput. Finally, we mention how the new Ultra-Wide Band (UWB) technology could enjoy similar benefits, and we discuss other research areas of interest in wireless communications.

SPREAD SPECTRUM ON THE INDUSTRIAL, SCIENCE, AND MEDICINE BAND

The current IEEE 802.11 and 802.11b standards employ spread spectrum technology that allows coexistence on the Industrial, Science, and Medicine (ISM) band. Either frequency hopping or direct sequence spread spectrum is used to ensure that communications can continue in the presence of an interfering system, although throughput will suffer. The Federal Communications Commission (FCC) regulations for the ISM band give priority to the requirement for coexistence over that of capacity due to the bandwidth inefficiency of spread communications. Note that regulations for the Unlicensed National Information Infrastructure (UNII) band abandon the coexistence requirement in favor of support for higher bit-rate transmissions. Indeed, a recent Notice of Proposed Rule Making [1] from the FCC indicates they will consider relaxing the spread spectrum requirement on the ISM band for the same reason.

Direct Sequence Spread Spectrum

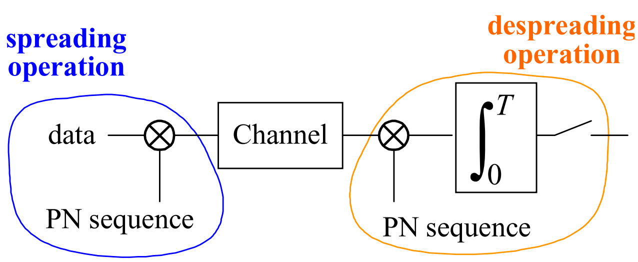


Figure 1: Transmitter and conventional receiver for direct sequence spread spectrum

Direct Sequence Spread Spectrum (DSSS) uses a secondary modulation, faster than the information rate, to spread the frequency domain content over a larger band, creating a lower power spectral density. In DSSS, each data bit is modulated by a Pseudo Noise (PN) sequence (transmitter and receiver are illustrated in Figure 1) that achieves the spectral spreading. The PN sequence is a seemingly random sequence of *chips* of plus and minus ones (see upper section of Figure 2) that achieve a second modulation of the data much faster than the data rate. PN sequences have good autocorrelation properties to allow the receiver to recover bit timing. The ratio of the time of the bit to the time of a chip is known as the processing gain of the DSSS system. The larger the processing gain, the better the autocorrelation properties, and the better the ability to reject narrowband interference. Note that the receiver multiplies the received signal by the original PN sequence as shown in Figure 1. This causes the data signal to be returned to its original form before being modulated by the PN sequence, i.e., it is despread.

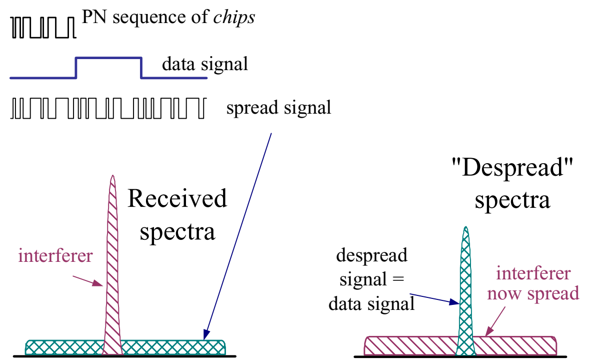


Figure 2: Direct sequence spread spectrum

In Figure 2, we see in the upper section the time domain characteristics of a DSSS signal. The frequency domain representation is given in the lower section of the figure. The over modulation of the data signal leads to a lower power spectral density covering a larger frequency band. Suppose the spread signal is transmitted in the presence of a narrowband interferer, the despreading operation at the receiver will take the wide band spread spectrum signal and collapse it back to the original data bandwidth. The receiver will also act on the narrowband interferer such that its spectrum is spread and causes much lower interference to the despread signal. This is known as jamming resistance or the natural interference immunity of spread spectrum signals.

We can see that DSSS is bandwidth inefficient in that it uses N chips to transmit a single bit of information. Without spreading the spectrum, we could have transmitted N bits in the same bandwidth. This inefficiency is the tradeoff to achieve interference rejection, or the ability to have reliable communications even in the presence of an interfering signal. It also reduces the power spectral density of the transmitted signal so that its transmission causes less interference to other systems operating at the same time on the same frequency band. Thus, you have the conflict between the two competing requirements of sharing the band and high bit-rate transmissions.

There are two other attractive features of DSSS besides interference rejection. First, DSSS can communicate well in the presence of multipath signals that are longer than a chip period, and even when delays are longer than a bit period, where other systems are crippled by Intersymbol Interference (ISI). A multipath channel is one exhibiting multiple reflections of the original signal. The *delay spread* of the channel is the average time for delays to propagate. When the delay spread exceeds the bit time of the transmission, reflections of the first bit will arrive in subsequent bits and lead to Intersymbol Interference (ISI). In order to avoid ISI, channel equalization techniques attempt to invert the channel and undo this effect. DSSS avoids this problem with the use of despreading. When despreading the first reflections of a transmission, subsequent reflections do not despread, and they have their interference reduced by the processing gain. Indeed, if multiple receivers are used, each with timing matched to a different reflection, the energy of each reflection can be combined to create a signal that is more reliable than a single receiver. This is known as a RAKE receiver. Therefore, while wasteful of bandwidth, DSSS is robust against multipath reflections.

Second, DSSS is compatible with the use of Code Division Multiple Access (CDMA) and a multiuser detector that allows for the simultaneous transmission of *many* active signals at the same time and frequency. If K users could access the same bandwidth at the same time, some of the inefficiency of the spread spectrum would be corrected by the presence of K transmissions in parallel. This is discussed further in later sections.

Complementary Code Keying

The IEEE 802.11b standard introduced Complementary Code Keying (CCK) modulation [2] that abandoned spread spectrum techniques in favor of error correction. FCC regulations for the ISM band require at least 10dB of processing gain. As per [3], CCK achieves this without being a spread spectrum signal, which is usually defined as processing gain = ratio of spread bandwidth to information bandwidth. The processing gain is achieved instead via bandwidth reduction (9dB) and coding gain (2dB). Therefore up to 11 times the bit rate could be achieved in the same bandwidth with higher-level modulation and comparable interference rejection. The CCK modulation was also designed to have a spectrum very close to that of the original 802.11 systems. Unfortunately, this modulation is not compatible with Code Division Multiple Access (CDMA) and Multiuser Detection (MUD), although it is still robust to multipath and can benefit from the use of RAKE receivers. We mention this standard in passing, as it is an example of the FCC's willingness to trade coexistence (interference rejection) for capacity (bit rate).

COMPARISON OF MULTIUSER DETECTION AND OFDM

Direct Sequence Spread Spectrum (DSSS) can be combined with Code Division Multiple Access (CDMA) to achieve multiple simultaneous transmissions on the same frequency band. The interference rejection capabilities of spread spectrum are exploited here to permit the presence of more than one signal on the same band. In Figure 3, we see two separate users employing distinct signature sequences. Each of these sequences is Pseudo Noise (PN) in nature so that the autocorrelation of each is very peaked. In addition, collections of PN sequences can be chosen to have low cross-correlation and can be thought of as mutually quasi-orthogonal. In Figure 3 we see that the presence of several users on the same frequency with different PN sequences simply adds to the level of background noise. Assuming the same despreading operation in a CDMA system as that used in a single-user DSSS (known as the conventional or matched filter), the system will have performance that degrades gracefully as more active users are on the air. A longer code means greater processing gain, and therefore a larger number of simultaneous users can be supported. The performance of such systems is dependent upon the cross-correlation properties of the sequences. In a synchronous system, the sequences can be truly orthogonal and cause no interference. However, multipath returns and timing inaccuracies destroy this orthogonality, and in practice, codes with low correlation are preferred to orthogonal codes (see, for example, the IS-95 standard for mobile telephony).

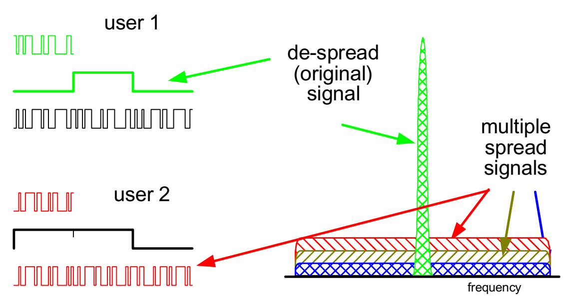


Figure 3: Code division multiple access using direct sequence spread spectrum

The term Multiuser Detection (MUD) is applied to receivers that take into account the structure of the multiple access interference in CDMA systems. By contrast, the matched filter detector, or conventional detector, ignores the presence of other CDMA users, the so-called Multiple Access Interference (MAI), or simply considers the MAI to have the same characteristics as Additive White Gaussian Noise (AWGN). In fact, the cacophony of multiple CDMA signals has a great deal of structure that can be used to sort out the received signal into components containing AWGN only, components derived from the MAI only, and finally a component containing some MAI and all of the desired signal.

The ability to support multiple, simultaneous users is a winning strategy to improve the throughput of the entire network. This is in part due to the performance of the multiuser receiver, but also in large part due to the nature of the power constraint itself. Power is constrained per user, but not per system. By having multiple users active at any given time, the total power of the network is higher, and this contributes to higher throughput.

Sergio Verdú [4] derived an expression for the optimal multiuser detector; however, its complexity is exponential in the number of active users. Several sub-optimal detectors have been proposed making various tradeoffs between complexity and performance. We consider only one such detector, the minimum mean square error detector. This detector is a linear detector, that is, the input signal passes through a linear filter that maximizes the output energy of the desired signal. While the coefficients of this filter can be calculated with complete knowledge of all system parameters (spreading signals for each active user, timing of each user, phase offset for each user, etc.), this filter can be cast in an adaptive version requiring very little side information.

Adaptive Minimum Mean Square Error Detector

The adaptive Minimum Mean Square Error (MMSE) detector is a linear filter whose coefficients are time varying. The coefficients are calculated each bit epoch via an adaptation such as the Linear Mean Square (LMS) or Recursive Least Squares (RLS) algorithms [5]. Both adaptations converge to the optimal filter; however, the RLS has the advantage of much faster convergence at the price of much greater complexity. Note that the complexity of the LMS solution is on a par with or less than that of the error correction required in the Orthogonal Frequency Division Multiplexing (OFDM) receiver for the IEEE 802.11a standard. The RLS implementation has greater complexity, but does offer increased capacity over the LMS solution for heavily loaded systems when the MMSE exact solution does not exist.

We assume the presence of a training preamble to each data transmission as this affords only minimal overhead (500 bits in an 8000-bit packet). This is consistent with the use of an LMS algorithm. RLS methods perform well even with much smaller training signals (100 bits are adequate). Blind versions of the algorithm are also possible that do not require any training sequence.

The adaptive MMSE has the advantage of not only suppressing the interference from other active users (multiple access interference or MAI), but also achieving RAKE-like combining of multipath signals that fall within the filter's range. For instance, if one-shot detection is used where a bit decision is made for an observation interval of one bit, multipaths that arrive within the bit interval and are separated by one to a few chips will be captured. Multipath signals that arrive much later (after a bit interval) require an expanded observation window for the filter to effectively capture the multipath energy.

The MMSE detector has been shown to be effective in rejecting not only multiple access interference from the CDMA system, but also other co-channel interference including both narrow and wide-band interference [6], [7]. For example, [6] reports results for a spreading code of length 63 (a maximal length sequence or m-sequence) in the presence of an interferer of equal to one-eighth-narrower bandwidth. The MMSE multiuser detector in the adaptive form was able to reject the interference even when the interferer's power was 30dB above that of the spread spectrum signal. Similar results were reported in [7] in addition to showing that both narrowband and multiple access interference could be rejected by the adaptive MMSE multiuser detector.

Therefore, the adaptive MMSE detector provides better coexistence than the natural immunity of spread spectrum systems, within the same receiver algorithm. Also, the complexity of the MMSE multiuser detector encompasses the suppression of narrowband interference as well as multiple access interference. Performance in the presence of narrowband interference will be lower than when there is none, but degradation will be graceful. Recall that OFDM systems rely on forward error correction for narrowband interference rejection, and would not be as effective as the MUD rejection in a spread spectrum system.

Hence, although the balance of this paper addresses only the performance for wide-band multiple access interference, the choice of MMSE MUD is motivated as well for its narrowband interference capacity.

OFDM ON THE UNII BAND

When moving to the new Unlicensed National Information Infrastructure (UNII) band, the focus was on achieving high-bandwidth efficiency, and the requirement for coexistence was relaxed. The Federal Communications Commission (FCC) no longer required the use of spread spectrum techniques, and the IEEE committee charged with coming up with a Wireless Local Area Network (LAN) standard (802.11a [8]) on this band opted for the use of Orthogonal Frequency Division Multiplexing (OFDM). While OFDM has no inherent immunity to in-band interference, it was chosen because of its robustness to multipath interference and because error correcting codes could be used effectively to exploit the frequency diversity of OFDM signals. Higher order Quadrature Amplitude Modulation (QAM) signaling with OFDM accomplishes the high_bandwidth efficiency at the cost of limited range. The MAC layer of these standards is the same, employing Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) that is successful when at most one user is transmitting at any time.

Comparison of Multiuser Detection and OFDM

We are interested in deriving the best strategy for pushing as many aggregate bits per second as possible through a network on the unlicensed bands. Consider these two choices: OFDM with high-order QAM modulation, and Code Division Multiple Access (CDMA) with a multiuser detector. In the case of CDMA, many parallel channels will exist, each transmitting at the same power levels as the corresponding OFDM signals. The OFDM signals increase total throughput for the wireless channel by increasing the bit rate with a single transmission at any given time (CSMA/CA protocol). The CDMA system transmits a direct sequence spread spectrum signal covering the same bandwidth as the OFDM system.

In calculating the throughput we take into account 1) the efficiency of the transmission, 2) contention, and 3) the bit error rate. By efficiency of the transmission we mean the inevitable overhead of training sequences, guard times, etc. By contention we mean the multiple access scheme and the loss in transmitted bits due to the failure of these schemes. In the case of 802.11a, the sources of contention are collisions in the contention window, which will be modeled as Poissonian with a fixed-transmission probability. (802.11a actually uses binary exponential backoff; essentially we do not take into consideration the added inefficiency due to backoff times.) The other source of collisions is the hidden terminal problem (also not considered in this analysis). We also don’t consider collisions due to uncoordinated users and/or interferers on the same frequency band.

Two broadband communications systems are compared, and we assume similar bandwidth in order to compare them fairly. The 802.11a standard uses 20MHz of bandwidth, so we use this for the CDMA system.[1] We use differential phase modulation for the CDMA signal to obviate the need for phase estimates during acquisition. However, after acquisition, we assume that phase tracking is in place, and coherent detection is used on the differentially encoded signal. The 802.11a system uses coherent detection. Spreading codes of length 31 are considered. This corresponds to a bit rate of 1.3Mb/s using Quaternary Phase Shift Keying (QPSK). This choice of spreading code is chosen for a reasonable compromise of requirements for acquisition (auto-correlation) and processing power. For now, we do not consider any error correction for the CDMA system, although this would be reasonable to add to future systems. The greatest question then becomes whether parallel channels with MUD are a better use of the peak power limited unlicensed bands than higher order signaling with error correction.

Throughput Calculation

Since these are multiple access protocols, we adopt the comparison of offered load vs. throughput [9]. We assume a Poissonian distribution on a larger base of $m \to \infty$ subscribers. For simplicity, we assume a slotted Aloha system for both OFDM and CDMA systems. Other, more stable protocols exist and could benefit each of these systems; however, for tractability we adopt slotted Aloha for this analysis.

In order to fairly compare and more easily evaluate the relative performance of these systems, we do not normalize the offered load as is customary, that is, we do not present the offered load in units of packets per slot as the slot size will be different for differing bit rates. Therefore, the offered load and throughput are presented in units of bits per second. The offered load can be thought of as the *aggregate* number of bits per second that the network is trying to communicate. We make no distinction about how this demand is structured (that is, the offered load is 50Mb/s for one subscriber demanding 50Mb/s, or five subscribers demanding 10Mb/s each, etc.). In the OFDM system, the load will be serviced by communicating with one subscriber at a time, and in the CDMA channel, there will be many subscribers active at any given moment. [2]

---

[1] In fact, there is over 100MHz of bandwidth available in the UNII bands, and the spread spectrum approach could cover greater bandwidth and yet coexist with other uncoordinated systems.

[2] In order to maintain reasonable delay, the CDMA throughput per subscriber should match the requirements of the most demanding application.

The Poissonian model gives the probability of m subscribers accessing the channel as

$$f(m) = \frac{y^m}{m!} e^{-y}$$

where y is the offered load in packets per packet slot. Note that in the plots we renormalize y to bits per second based on the bit rate being used and a fixed-size packet. The steady state throughput is defined as the expected value of the successfully transmitted packets and is given by

$$\sum_{m=1}^{\infty} m P_c(m) f(m)$$

where $P_c(m)$ is the probability of successful packet reception given there are m users accessing the channel. Once again, plots presented will renormalize the throughput to bits per second for the bit rate of the system reported and a fixed-packet length. Note that the throughput of a CDMA system will be higher than the single-user bit rate since multiple transmissions are made in parallel.

Typical throughput analysis of CSMA schemes assumes a perfect channel when there is a single transmission and bit error rate (BER) equal to one when there is more than one simultaneous transmission. In our analysis of OFDM, we assume a lost packet whenever there is more than one transmission (a collision), but we find the BER for OFDM in Additive White Gaussian Noise (AWGN) to evaluate throughput for the case of a single transmission (no collision).

For a length L packet, the correct packet rate is related to the BER by

$$P_c(m) = 1 - PER(m) = \left(1 - BER(m)\right)^L$$

where PER is the packet error rate. We assume a packet size of L=1000bytes=8000bits. For OFDM we use

$$P_c(m) = \begin{cases} 1 - PER(m) & m = 1 \\ 0 & m \neq 1 \end{cases}$$

For the CDMA system, the PER will degrade with rising values of simultaneous transmissions m in a continuous fashion, i.e., the more users accessing the channel, the larger the BER.

BER Calculation

The probability of bit error is a function of the Signal-to-Noise Ratio (SNR). For the CDMA system, the bit error rate is a function of both SNR and m. For 802.11a we assume that for m>1 the error will be one, where for m=1 the error is a function of the bit error rate, which in turn varies with the modulation, coding rate, and bit rate. The table below shows the various bit rates supported by the 802.11a standard.

For the OFDM approach, we also assume that all algorithms that might improve the link are included (for instance, soft Viterbi detection for the error correction and ideal channel estimates). The column in the table labeled Coding Gain contains values gleaned from Figure 3.11 of [10] and verified in Matlab*. This represents the effective increase in SNR due to the use of forward error correction. The bit error rate is otherwise calculated using ideal coherent detection of the modulation format in use.

Bit Rate	Bits per Symbol	Coding Rate	Coding Gain (G_c)	Effective SNR	Modulation
6 Mb/s	1	1/2	NA	SNR*G_c	BPSK
9 Mb/s	1	3/4	NA	SNR*G_c	BPSK
12 Mb/s	2	1/2	5.5 dB	SNR/2*G_c	QPSK
18 Mb/s	2	3/4	7.0 dB	SNR/2*G_c	QPSK
24 Mb/s	4	1/2	8.5 dB	SNR/4*G_c	16QAM
36 Mb/s	4	3/4	4.5 dB	SNR/4*G_c	16QAM
48 Mb/s	6	2/3	5.5 dB	SNR/6*G_c	64QAM
54 Mb/s	6	3/4	6.5 dB	SNR/6*G_c	64QAM

Parameters for 802.11a transmission rates

We further assume that these systems have a limited power budget independent of the modulation method, either due to conservation of battery power or due to regulation of the unlicensed bands. One immediate consequence of this assumption is that in going from small to large signaling sets, the energy per bit must necessarily scale down accordingly. For a given energy per bit at BPSK signaling, there would be half the energy per bit if choosing QPSK (see column "Effective SNR" in the above table).

For the CDMA system, we use a multiuser detector in the form of an adaptive MMSE algorithm with training bits every packet. The receiver can use a Linear Mean Square (LMS) algorithm for lower complexity (less than the high-rate OFDM Viterbi decoder) or a Recursive Least Squares (RLS) formulation in a systolic array for numerical stability, efficient parallel implementation, and

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

faster convergence. Steady state performance for the LMS and RLS are similar. The CDMA system uses 100 bits for training the RLS algorithm, while LMS would require on the order of 500 bits for training.

The bit error rate of the MMSE detector in AWGN can be calculated explicitly [4], whereas Monte Carlo simulations are used to determine performance in the presence of multipath signals. Figure 4 presents the performance of the MMSE detector as a function of SNR and the number of equal power interferers accessing the channel simultaneously. These simulations assume three paths; each separated by a chip interval, with the second path one half the amplitude of the first path, and the third path one quarter the amplitude of the first.

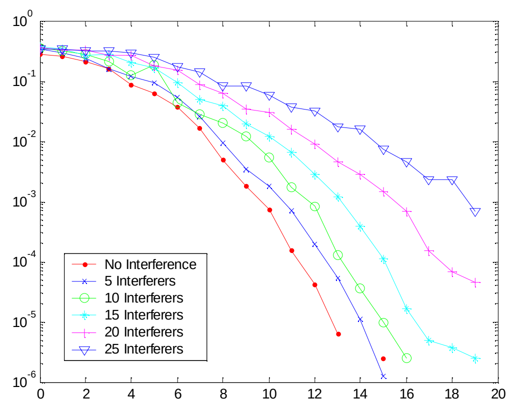
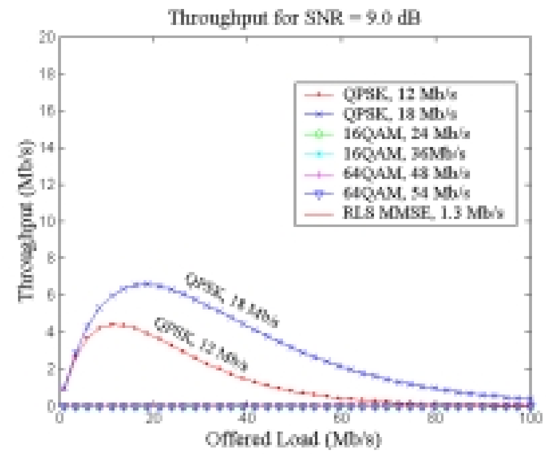


Figure 4: MMSE performance in the presence of multipath vs. SNR for various numbers of equal power interferers

Results

Using the bit error rate results for OFDM and Minimum Mean Square Error (MMSE), we are able to calculate the throughput. In Figures 4, 5, and 6 we present our results for the throughput vs. the offered load (as defined previously) for three SNR scenarios: poor, moderate, and SNR. Note that the curves for the 802.11a standard take the shape of the throughput vs. offered load of a CSMA system, with the peak throughput occurring when the offered load equals the bit rate of the transmission. The RLS MMSE curve, on the other hand, has its peak as a function of the SNR. As the SNR increases, the peak moves towards higher offered loads. This is because the CDMA system with MUD can exploit the added SNR to suppress interference and introduce more parallel channels to give higher throughput.

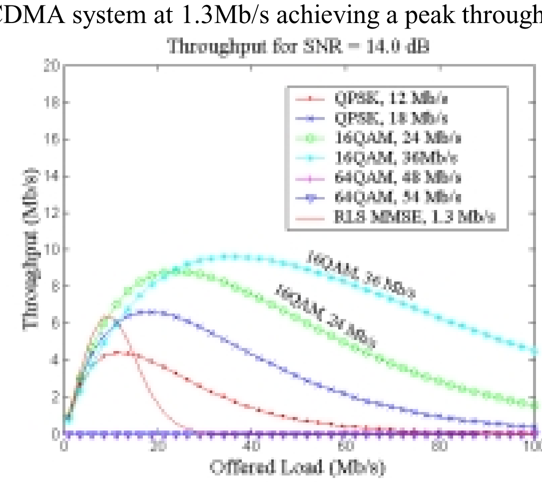
In the low SNR case (Figure 5), we see that the higher order signaling sets of 802.11a cannot achieve reliable communications; only the QPSK modulations have appreciable throughput. Similarly, the CDMA system can only support a low number of users (one or two) at reasonable bit rates, so the throughput does not compare



well with the QPSK modulations.

Figure 5: CDMA with RLS MMSE, and 802.11a in AWGN at various bit rates and poor SNR

The moderate SNR case (Figure 6) shows more modulation types giving significant throughput, with the CDMA system at 1.3Mb/s achieving a peak throughput of



6Mb/s, comparable to the 18Mb/s OFDM system.

Figure 6: CDMA with RLS MMSE, and 802.11a in AWGN at various bit rates and moderate SNR

At high SNR (Figure 7), the 54Mb/s system has lower throughput than that of the 48Mb/s system due to a higher error rate. The CDMA system achieves peak throughput slightly above the best performing 802.11a system. Recall that these plots capture the performance in AWGN with no interference presence. The coexistence properties of CDMA with an adaptive MMSE receiver would favor its performance in an unlicensed band prone to interference.

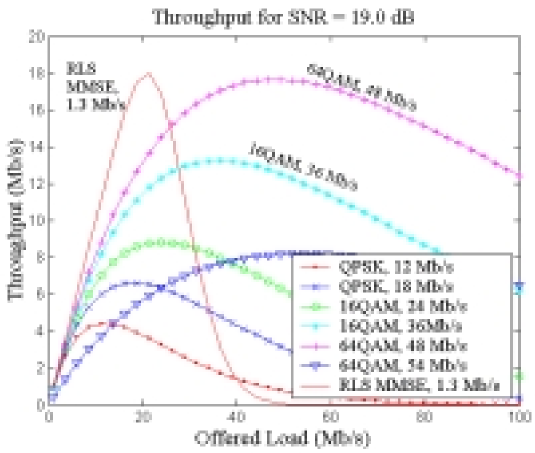


Figure 7: CDMA with RLS MMSE, and 802.11a in AWGN at various bit rates and high SNR

Future work will address the relative performance of CDMA and OFDM in a multipath environment. Van Nee [11] found that OFDM systems using error correction enjoy more robust correction as the delay spread (mean time of the reflected signals) of a multipath channel increased. This is true only until the delay spread exceeded the guard interval inserted in the OFDM signal; after this point, the intersymbol interference introduced by the channel caused significant errors. Ideal error correction properties were assumed for this study, and we need to assess the effect of multipath on the coding gain factors.

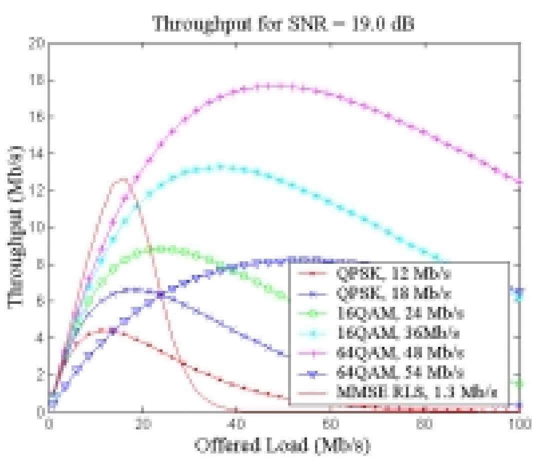


Figure 8: CDMA with RLS MMSE in the presence of multipath, and 802.11a in AWGN (no multipath)

In Figure 8, the OFDM performance is shown again for AWGN, while the CDMA system is shown for a three path channel whose BER is given in Figure 4. Now the energy of the CDMA system is assumed not to be concentrated in a single return path, but rather divided among three paths with 75% of the energy in the first path. We can see that the MMSE detector performance has deteriorated, but not significantly for these parameters. This seems to indicate that the MMSE will do well vis-à-vis the OFDM system, but more rigorous investigation is needed.

ULTRA-WIDE BAND OVERLAY

The strategy for optimizing a power limited communications system via Multiuser Detection (MUD) with parallel transmission channels can be applied equally well to Direct Sequence Spread Spectrum (DSSS) using ultra short chip pulses, such as those proposed for Ultra-Wide Band (UWB) technology. An introduction to UWB can be found at [12]. There are many technical challenges that must be overcome for UWB use of multiuser detection–most notably very high sampling rates and intense signal processing. Despite these challenges, a multiuser detector may be key to unleashing the power available in the abundant multipath return on UWB signals indoors.

CONCLUSION

We have discussed strategies to deal with challenges of coexisting with other users and extracting the greatest capacity from systems that are both power and band limited. We compared the 802.11a standard with a proposed system employing Code Division Multiple Access (CDMA) and multiuser detection. The 802.11a standard uses Orthogonal Frequency Division Multiplexing (OFDM) with higher level signaling and forward error correction to achieve spectrally efficient high bit rates in the unlicensed bands. The MAC layer uses a contention window with Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). No interference mitigation strategy is used other than allocating different Local Area Networks (LANs) to different sections of the Unlicensed National Information Infrastructure (UNII) band. We propose the use of a CDMA system employing multiuser detection, in the form of an adaptive Minimum Mean Square Error (MMSE) detector, as an active interference mitigation strategy, as well as for efficient multiple access.

The adaptive MMSE multiuser detector has good rejection properties for both narrow and wide-band interference, as documented in [6] and [7]. It requires no explicit channel estimation, and both blind and trained versions of the adaptation are available. The MMSE detector offers multipath resistance as well as a RAKE-like functionality that captures energy from the multipath signals. The complexity of the Linear Mean Square (LMS) implementation of the adaptive MMSE (depending on spreading length, sampling rate, etc.) is on a par with that of the error correction required for the IEEE 802.11a standard for OFDM.

We have shown that for an Additive White Guassian Noise (AWGN) channel the CDMA system with a Multiuser Detection (MUD) can achieve peak throughput greater than the 802.11a system for strong Signal-to-Noise Ratio (SNR). Even when the CDMA system is subject to multipath and the OFDM has no multipath, the CDMA system has comparable peak throughput. For systems with narrowband interference, the performance of the CDMA system will be markedly better than OFDM, although we only discussed this feature qualitatively. Future work should compare the two systems as a function of the delay spread of the multipath channel, and as a function of the level of narrowband interference.

The use of CDMA for multiple parallel transmission is a good strategy for power-limited systems, since at any given time the power available to the network is higher than in a system where only one network subscriber can access the channel at a given time. Finally, and perhaps most importantly, the CDMA system with a multiuser detector offers an upgrade path for higher performing systems as process technology permits the use of increasingly complex receivers and faster sampling. This approach is especially promising in a situation where greater spreading is available, such as with UWB technology.

ACKNOWLEDGMENTS

The author thanks Nihar Jindal, a doctoral candidate at Stanford University who performed MMSE simulations during a summer internship, as well as the anonymous reviewers whose suggestions much improved this paper.

REFERENCES

[1] ET 99-231; FNPRM & ORDER 05/11/01 (adopted 05/10/01); FCC 01-158 *Amendment of Part 15 of the Commission's Rules Regarding Spread Spectrum Devices, Wi-LAN, Inc. et al.*

[2] Pearson, Bob, "Complementary Code Keying Made Simple," Application Note 9850, May 2000, http://www.intersil.com/data/an/an9/an9850/an9850.pdf

[3] Andren, C. and Webster, M., "CCK Modulation Delivers 11Mbps for High Rate 802.11 Extension," Wireless Symposium/Portable By Design Conference Proceedings, Spring 1999.

[4] Verdú, Sergio, *Multiuser Detection*, New York: Cambridge University Press; 1998.

[5] Madhow, U. and Honig, M. L., "MMSE Interference Suppression for Direct-Sequence Spread-Spectrum CDMA," *IEEE Transactions on Communications*, vol. 42, pp. 3178-3188, Dec, 1994.

[6] Fathallah, H. and Rusch, L. A., "A Subspace Approach to Adaptive Narrowband Interference Suppression in DSSS," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1575-1585, Dec., 1997.

[7] Poor, H. V. and Wang, X., "Code-aided interference suppression for DS/CDMA communications," *IEEE Transactions on Communications*, vol. 45, no. 9, pp. 1101-1122, Sept. 1997.

[8] IEEE 802.11a-1999 (8802-11:1999/Amd 1:2000(E)), *High-speed Physical Layer in the 5 GHz band*, available at http://standards.ieee.org/getieee802/

[9] Raychaudhuri, D., "Performance Analysis of Random Access Packet-Switched Code Division Multiple Access Systems," *IEEE Transactions on Communications*, vol. COM-29, no. 6, pp. 895-901, June, 1981.

[10] van Nee, Richard and Prasad, Ramjee, *OFDM for Wireless Multimedia Communications*, Boston: Artech House Publishers; 2000.

[11] van Nee, R., "A new OFDM standard for high rate wireless LAN in the 5 GHz band," *Vehicular Technology Conference*, 1999, pp. 258-262, Sept. 1999.

[12] Foerster, J., Green, E., Somayazulu, S. Leeper, D., "Ultra-Wide band Technology for Short- or Medium-Range Wireless Communications," *Intel Technical Journal*, 2nd quarter, 2001, available at http://developer.intel.com/technology/itj/

APPENDIX

To help you better navigate this paper, following is a table of acronyms.

AWGN	Additive White Gaussian Noise
CA	Collision Avoidance
CCK	Correcting Code
CDMA	Code Division Multiple Access
CSMA	Carrier Sense Multiple Access
DSSS	Direct Sequence Spread Spectrum
FCC	Federal Communications Commission
ISI	Intersymbol Interference
ISM	Industrial, Science, and Medicine
LAN	Local Area Network
LMS	Linear Mean Square
LMS	Least Mean Squares

MAI	Multiple Access Interference
MMSE	Minimum Mean Square Error
MUD	Multiuser Detection
NPRM	Notice of Proposed Rule Making
OFDM	Orthogonal Frequency Division Multiplexing
PAN	Personal Area Network
PN	Pseudo Noise
QAM	Quadrature Amplitude Modulation
QPSK	Quatrenary Phase Shift Keying
RLS	Recursive Least Square
SNR	Signal-to-Noise Ratio
UNII	Unlicensed National Information Infrastructure
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network

AUTHOR'S BIOGRAPHY

Leslie A. Rusch was born in Chicago, IL and received a BSEE degree from Caltech in 1980, and an M.A. and Ph.D. degree in electrical engineering from Princeton University in 1992 and 1994, respectively. From 1994 to 2000, she was an Assistant Professor at Université Laval in Québec City. Dr. Rusch is a Senior Member of the IEEE and a member of the Optical Society of America (OSA). Her e-mail is leslie.a.rusch@intel.com

http://www.intel.com/research/mrl/people/rusch_1.htm,
leslie.a.rusch@intel.com

The Evolving Role of Automation in Intel® Microprocessor Development and Manufacturing

Kumud M. Srinivasan, Logic Technology Development, Intel Corporation

Index words: automation, 300mm, D1C, information turns, AMHS

ABSTRACT

In March 2001, Intel completed production of “first silicon” from its 0.13-micron technology, 300mm wafer development fab. Named “D1C,” this fab represents the most sophisticated level of automation in Intel today, and an inflection point in the evolution of fab automation within Intel. In this paper, we travel down this path of evolution detailing the role of automation in process development and manufacturing.

INTRODUCTION

This paper describes the pervasive role that automation is playing in 300mm process development and manufacturing. This is contrasted with the role that automation has played in Intel’s 150mm and 200mm process development and manufacturing. Additionally, we outline some of the future challenges and opportunities in this area.

Many aspects of this automation evolution are explored. We describe automation’s changing role from a support function, to a technology module, to an enabler of 300mm operations. We characterize the role of automation in “information turns” and discuss its ability to speed up the process development cycle time. A view of the increasing richness of capabilities, the re-use profile, and the improvement in reliability and availability over technology generations is also given.

AUTOMATION IN PROCESS DEVELOPMENT AND MANUFACTURING

Automation and Information Turns

Why is automation important? During process development, quick information turnaround is required to hit process targets and to sustain the ever-increasing speed of process technology conversions. From Figure 1 below, we can see that quick information turnaround means increasing information and WIP turns. (WIP turns are defined as the number of wafers processed through an activity in twenty-four hours, divided by the total wafer inventory.) The goal during the experiment-execution phase, or the manufacturing phase, is to increase WIP turns. An increase in WIP turns automatically increases information turns. Additionally, information turns are impacted by the cycle time of information during the experiment-design phase as well as by the time it takes to analyze data after a lot leaves the fab.

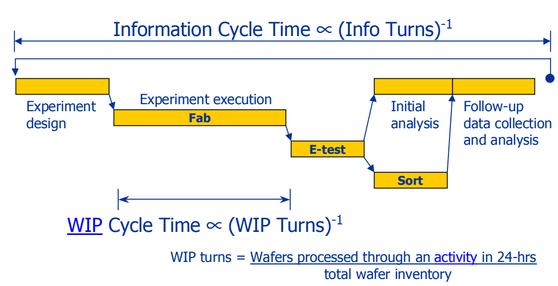


Figure 1: Information turns and WIP turns

Automation systems can speed up information turns in many ways. As Figure 2 shows, experiment definition systems help design and specify experiments. Manufacturing execution systems enable experiment definition and execution and overall work flow control in the fab. Equipment control systems automate tool control, and wafer handling is automated via material-handling systems. Decision support systems help monitor fab performance. Data analysis systems enable analysis of E-test, yield, and Sort data. The challenge of automation systems, then, is to enable, via these systems, ever-increasing information and WIP turns.

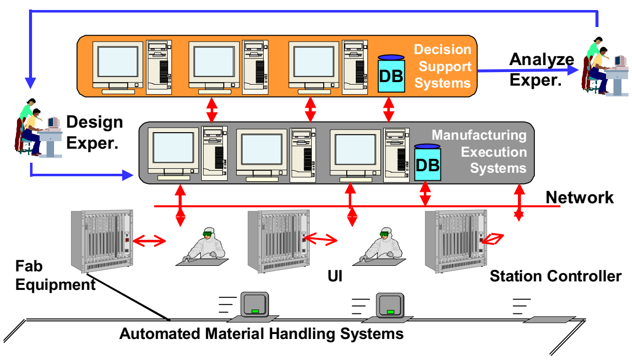


Figure 2: Automation in process development and manufacturing

Automation as an Enabler of 300mm Operations

With 300mm technology, the role of automation has taken on even more significance. We now have much larger wafers, about 12 inches in diameter, which is more than twice the surface area of 200mm wafers, which are about eight inches in diameter. These wafers provide 240 percent more area for printed die per wafer than standard 200mm wafers, thereby lowering production cost per chip. This increase in the size and weight of wafers has created new challenges for automation systems.

Ergonomic challenges caused by the increased size and weight of wafers have resulted in the need for automation of all wafer handling, and in the emergence of automation systems to transport wafers directly to and from each tool in the factory. Today, Intel's 300mm fab, D1C, features the world's first fully automated system for handling wafers. Tools with mini-environments and slot selection capabilities enable wafer handling to be completely automated: stockers and carousels store wafers; a monorail, overhead vehicles, and automated guided vehicles transport wafers to and from tools and stockers. Contrast this with a typical 200mm fab where WIP racks are used in conjunction with stockers to store wafers, WIP carts are used in conjunction with interbay systems to transport wafers, and wands are used for wafer handling. The high level of mechanical automation in the 300mm fab is accompanied by a tight integration of data automation systems–equipment control, scheduling and manufacturing systems–resulting in the ability of 300mm fabs to progress from basic point-to-point delivery of wafers to continuous, uninterrupted processing, and eventually to lights-out operations.

The pervasive automation of wafer handling has enabled the movement of wafer processing away from the bay to a central location in the chase, or to a command center, which makes for a safer environment as it reduces human contact with material-handling systems. In 200mm fabs, technicians run tools from the front in bays. The automation in the 300mm fabs enhances productivity as it enables one technician to run several tools from a single location. It enables dense packing of tools with fewer tool interfaces thus supporting changes in the design of user interfaces and workstations. The user interface is evolving from one that is designed to be highly interactive with the technician to one that facilitates the remote management of clusters of tools. Likewise, workstations are being designed to support mobility via wireless devices and tool operations via command centers.

The use of test wafers represents yet another operational change in 300mm fabs. The higher cost of wafers has resulted in the need for their extensive re-use through regeneration and reclaim. The inability to manually handle wafers has resulted in the need for extensive tracking of wafers at the wafer level. The pervasive use of automated intrabay systems has resulted in the need for systems that comprehend complex workflows where a tool is held "down" while test wafers are pre- and post-measured. Equipment control and manufacturing execution systems are being extended to address these challenges.

Figure 3 shows the presence and impact of automation systems in a typical 300mm bay-chase.

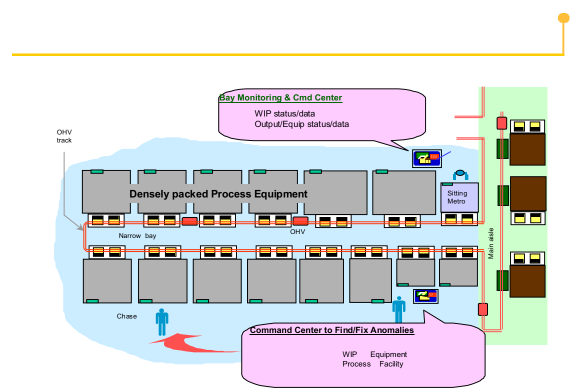


Figure 3: Automation in the 300mm clean room

KEY COMPONENTS AND ARCHITECTURE

So, what is meant by *automation*? As shown in Figure 4, automation components typically belong to one of five primary systems: process equipment interface, manufacturing execution, automated material handling, engineering analysis, and infrastructure (a pervasive layer of hardware, networks, and user interfaces).

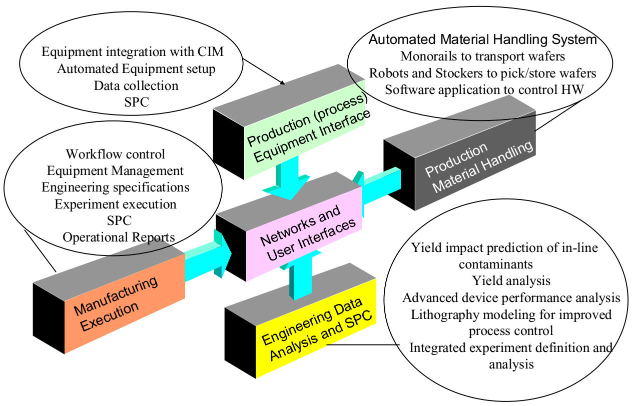


Figure 4: Key components of automation

These systems integrate and interact with each other via the use of industry standards, wherever possible, and through custom adapters, where no standards exist. Figures 5 and 6 show two views of the integration architecture as it is implemented in D1C today.

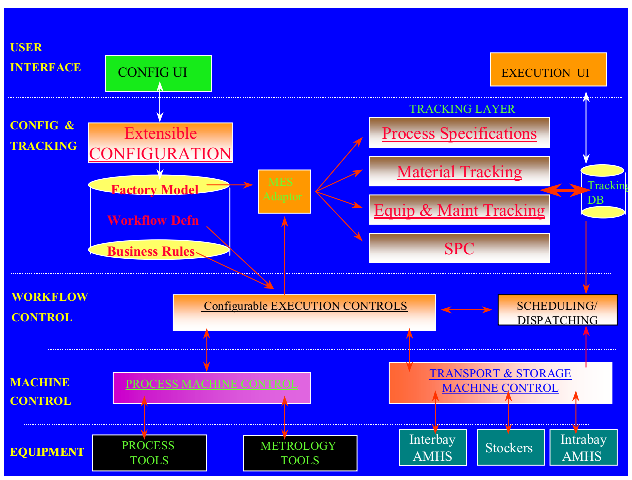


Figure 5: Automation systems integration architecture

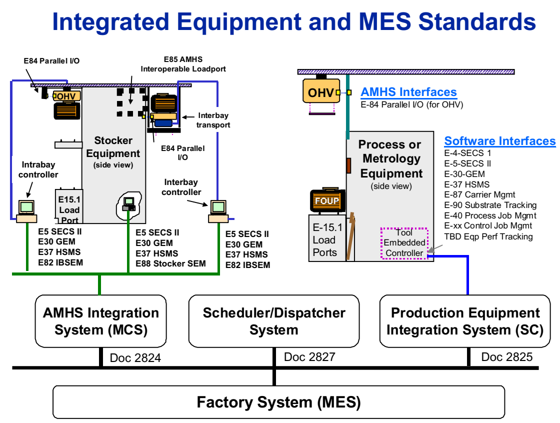


Figure 6: Use of standards in systems integration

INTEGRATION WITH PROCESS TECHNOLOGY

Automation Technology Cycle

As automated systems have become more complex, so have the inherent methodologies and processes. Process development has, over several technology generations, developed a rigorous methodology and lifecycle as described in Figure 7. The key philosophy behind this methodology is as follows: as a process generation moves from research through development into manufacturing, the willingness to take risks diminishes rapidly because the costs of making mistakes increases dramatically. Because of this, the technical focus changes from evaluation during research to integration followed by replication during the manufacturing ramp. Likewise, management's focus changes from planning to synchronization to control.

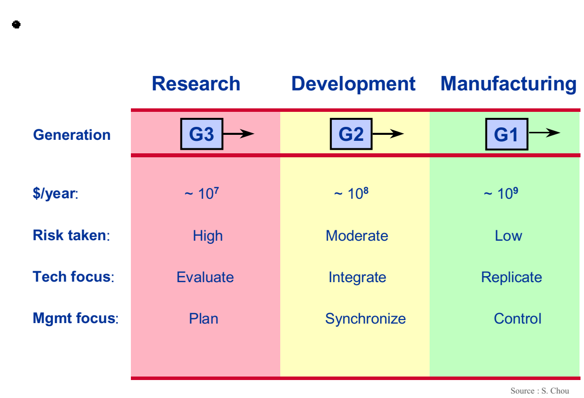


Figure 7: Technology life cycle

Automation today has adopted a methodology and lifecycle similar to that of process development. In

150mm and early 200mm technology generations, automation systems were introduced when they were ready, not in a manner that was necessarily synchronous with the process development cycle. The 0.18-micron technology saw the emergence of automation as a technology module, and the adoption of a methodology for managing innovation and risk that mirrored the one that had been successfully used for process development. This resulted in the establishment of an Automation Module Target Spec for a process technology generation. The spec captured automation capabilities to be transferred with a process to high-volume manufacturing sites. The spec established performance and reliability targets for automation systems. Also adopted was the re-use methodology that is used to quantify the extent of change in process equipment from one generation to another. Development of systems was managed by a program as is done for process. A milestone was defined for development completion and integration known as an Automation Baseline. This was pegged at a year prior to ramp start to enable sufficient time for systems to harden for ramp. Transfer and training of personnel at receiving sites were facilitated by adoption of the "seeds" program. This enabled receiving sites to send personnel to the technology development site for training assignments of 6-12 months duration. This technology cycle is captured in Figure 8.

As mentioned previously, 0.18um was the first technology generation in which automation adopted a cycle that mirrored the one used in process. Since then, this methodology has been used successfully in two subsequent generations, 0.13um in 200mm and in 300mm. This has now become the *de facto* standard.

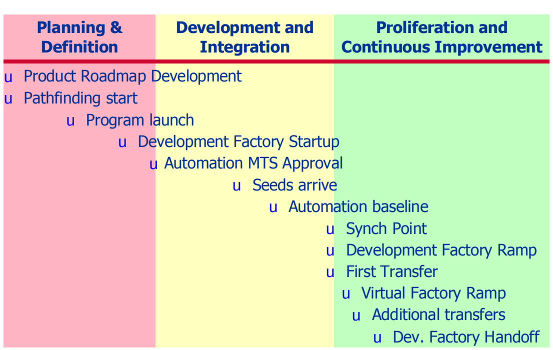


Figure 8: Automation technology cycle

Management Structure and Development Methodology

As expected, the automation management structure has also evolved to accommodate the technology cycle described in the previous section. Figure 9 describes this structure. Basically, the structure and ownership change as the cycle moves through the phases of research, selection, development, and proliferation, enabling the technology during each phase to get an appropriate level of management focus.

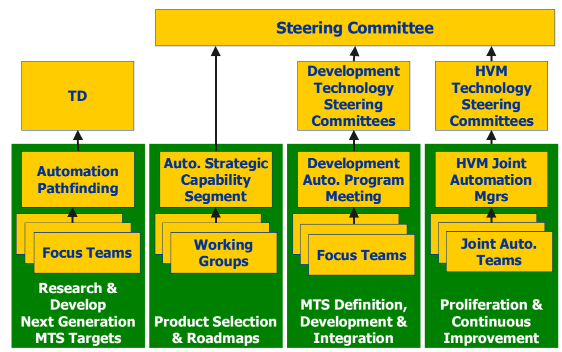


Figure 9: Management structure

We now move on to the development and implementation methodology used today, as shown in Figures 10 and 11. The pervasiveness of automation systems in the fabs today has necessitated a high level of rigor in development and implementation methodology. This is because, typically, when an automation system is down, fab processing is halted, and potential revenue is lost.

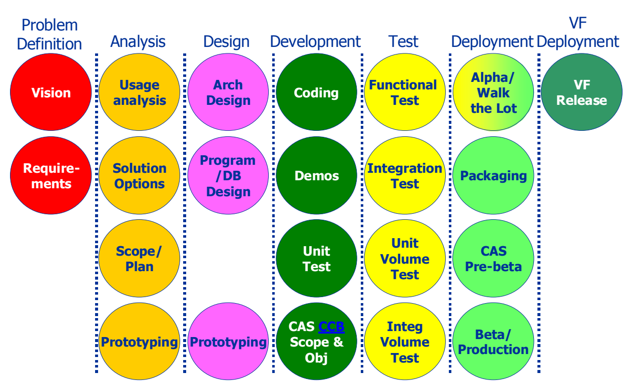


Figure 10: Development methodology

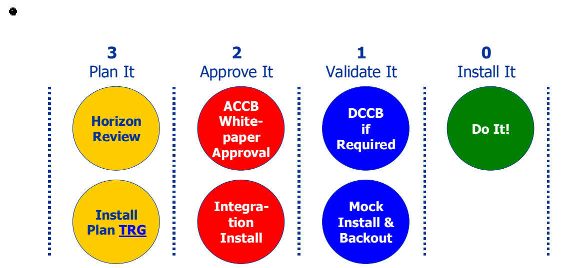


Figure 11: Automation change methodology

The impact of the technology cycle, management structure, and rigorous methodologies is evident today in the reliability of automation systems. Reliability has improved dramatically from a mere five days between unscheduled full fab downs in the 0.35um generation to a record 253 days in the 0.13 generation!

THE EVOLUTION OF INTEL'S FIRST 300MM FAB

Let us now take a look at the evolution of automation capabilities over process generations and see how they have come a long way from the days of the 150mm generation. Figure 11 illustrates this evolution. Incremental capabilities can be seen in all components: equipment control, manufacturing execution systems, material-handling systems, data analysis systems, and system infrastructure and architecture.

	Equipment Interface	Manufacturing Execution	AMHS	Engineering Analysis
0.8um (6")	DOS-based Eqp Control	Experiment tracking	Interbay	Centralized file management system
0.6um (8")	Unix-based eqp control	Skip lot sampling	Diffusion Intrabay	
0.4um	Statistical process control			
0.25um	Excursion Protection		Litho Intrabay	
0.18um	Stand-alone User Interface, Advance Process Control	Rich Specs	Scheduler	Yield Analysis System
0.13um		Experiment definition, Cu segregation		
0.13um (12")	NT-based eqp control, user interface for tool clusters	Test wafer tracking, FOUP management	Pervasive Intrabay	MES-insulated DSS

Figure 12: Evolution of automation capabilities

Equipment interfaces have evolved from DOS-based controllers to NT*-based ones in today's D1C. Along the way, the capability set has expanded from recipe control and data collection to excursion protection and advanced process control. The architecture has evolved from one that had the user interface hard-wired into it to one that is able to handle the user interface as a separate layer. Manufacturing execution systems have given fabs the ability to define and track complex experiments, FOUPs, and test wafers, as well as the ability to view rich specs and the controls to support copper segregation. Material-handling systems have expanded from intrabay in Diffusion and Lithography only to one that is pervasive, and integrated with a fab-wide scheduler. Decision support systems have made significant headway in yield analysis and in architectural insulation from manufacturing execution systems. Through these generations, the architecture has evolved from a single monolithic system to a multi-layered system with a powerful middleware layer: it has gone from dumb terminals to wireless PCs in the clean rooms. Additionally, the technology applied has evolved from DOS to NT, from COBOL to Java* and C++, and from hierarchical to OO databases.

D1C, Intel's first 300mm fab, represents the most sophisticated level of automation in Intel today. It has seen a step-level increase in automation capability from the previous technology generation. This is best illustrated by the re-use profile since 0.25um. A technology generation typically re-uses 75-85% of the previous generation's automation capabilities, and it sees 10-15% new capabilities. Intel's 300mm, 0.13um generation has re-used only ~55% of the capabilities of the equivalent 200mm generation, while it has seen ~50% new capabilities. A look at capital investment shows a similar story: re-use of capital investment has been a mere 25%, with ~100% new investment.

The D1C changes are pervasive, and they cover many components from state-of-the-art decision support system with data warehouse, data mart and OLAP technologies, to sophisticated and tightly integrated scheduling and control systems. But, the highlight is the automated material-handling system. Figures 13 and 14 show examples of some of the material-handling systems in use in D1C today.

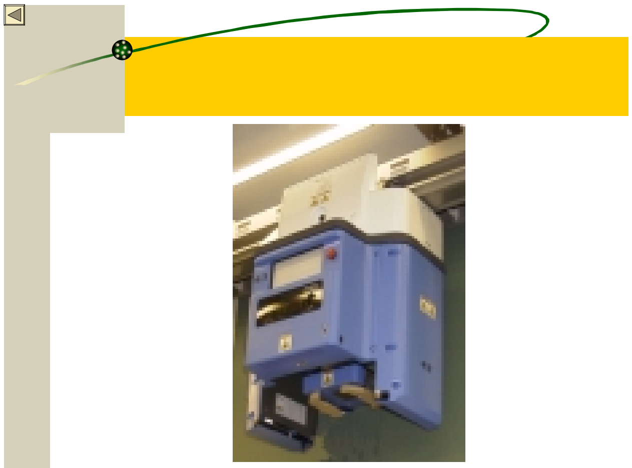


Figure 13: Overhead vehicle

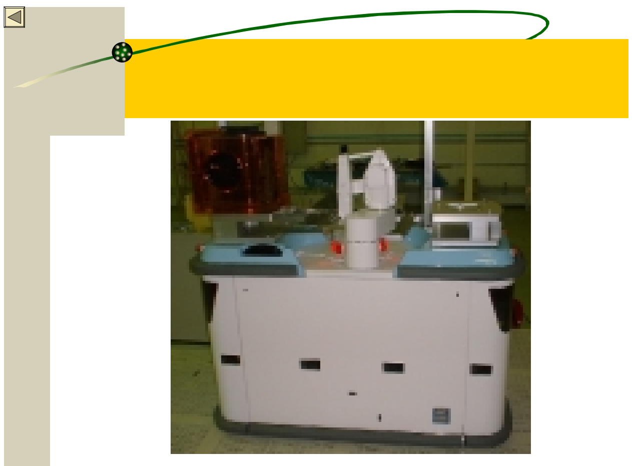


Figure 14: Automated guided vehicle

CHALLENGES AHEAD

As we look ahead, we see some primary trends. As process margins shrink, automation capabilities required to enable new process technologies are increasing. Wafer size increases have significantly increased the demand for full automation across the fab. Moreover, computing technology and available power continue to grow at a rapid pace.

Given these trends, one of our primary challenges is to obtain a fundamental understanding of technology development and manufacturing operations across Fab, Sort, Assembly, and Test to ensure we bring the right solutions at the right time. We need to increase our technical expertise in, and stay current with, computing technologies. We need to become effective system architects and integrators. We need to demonstrate a consistent use of effective development methodologies. And last but not least, we need to maintain our focus on operational excellence.

There are many examples of real opportunities ahead. One is the replacement of the 0.13um manufacturing execution system with a next-generation system that is capable of tracking workflow at a wafer level. Other examples are improvement in storage management (prioritized storage), improved integration with lot sorter tools, and automation of remaining manual areas. More sophisticated equipment control, pervasive advanced process control, remote factory operations, and increased automation reliability requirements are yet other examples of future opportunities.

CONCLUSION

Factory automation at Intel is at an inflection point. The performance of the 300mm fab of the future will be intricately linked with the performance of its automation systems. This makes it an appropriate time to pause to reflect on the path that automation has traveled and the role that it has played in Intel's microprocessor development and manufacturing. It helps prepare us for the many exciting challenges ahead. And, last but not least, it has made me appreciate the opportunity that I have enjoyed in my years at Intel, the opportunity to influence the direction that automation has taken and to contribute to its rich history.

ACKNOWLEDGMENTS

I want to thank Sunit Rikhi, my partner and co-manager of LTD Automation, who helped me select the topic for this paper. In addition, many thanks go to the contributors to the "Introduction to LTD Automation" class: Wayne Carriker, Ross Giddings, Hazem Hajj, Kris Hegg, Chandra Mouli, Sunit Rikhi, and Sara Runnings. Much of the information in this paper has been taken from the class materials. Special thanks to Wayne Carriker for the invaluable shots of D1C's AMHS. Last, but not least, thanks to Arnie Sparnins for his input on the role of industry standards.

REFERENCES

[1] "Introduction to LTD Automation" class. Contributors include: W. Carriker, R. Giddings, H. Hajj, K. Hegg, C. Mouli, S. Rikhi, S. Runnings, and K. Srinivasan.

[2] Carriker, Wayne, "D1C AMHS Status," *LTD Automation Communications Meeting*, August 18, 2000.

[3] Mouli, Chandra and Byrd, Tom, "MES Model," Px60 Ops. Review, WW10.

AUTHOR'S BIOGRAPHY

Kumud Srinivasan is the manager of Logic Technology Development automation in the Technology and Manufacturing Group. She shares responsibility with Sunit Rikhi for managing the development and integration of automation for technology development and manufacturing. Srinivasan joined Intel in 1987 in the Technology and Manufacturing Group's second high-volume manufacturing fab in Albuquerque, where she developed, and later managed, several key automation initiatives and teams. In 1991, she joined the Intel Mask Operation group, and later, Intel Test Tooling Operation group, as manager of software and business development. In 1995, she joined the Portland Technology Development Group. Currently, her scope includes wafer processing and wafer sort and test automation. Srinivasan graduated in 1984 with an M.S. degree from Syracuse University, School of Information Studies, and did a post-graduate stint at UC, Berkeley. In 1981, she received a B.S. degree in Economics from Calcutta University, India. Her e-mail is kumud.srinivasan@intel.com

The Making of a Compiler for the Intel® Itanium™ Processor

Carole Dulong, Intel Compiler Lab, Intel Corporation
Priti Shrivastav, Intel Compiler Lab, Intel Corporation
Azita Refah, Intel Compiler Lab, Intel Corporation

Index words: compiler, Intel® Itanium™ processor, SDK, SoftSDV, OSV, ISV

ABSTRACT

Intel has developed an Itanium compiler that compiles code written in C/C++ or Fortran languages and generates the assembly instructions for the Itanium™ architecture on Windows NT* and UNIX* platforms. This paper describes how the Intel Itanium compiler was designed and developed in the absence of hardware. The project started more than five years before we had any hardware, and the work on simulators presented many challenges. We discuss three different aspects of the Itanium compiler: the compiler performance analysis, which focused on demonstrating the Intel® Itanium™ processor (ITP) performance on the Spec benchmarks; enabling Operating System Vendors (OSVs) such as IBM to port their Operating System (OS) to the Itanium architecture; and Independent Software Vendor (ISV) enabling work, which focused on enabling ISVs to port real applications on the Itanium architecture.

For each area we describe the methodology used in the absence of hardware, the results achieved, and the lessons learned.

GOALS OF THE INTEL ITANIUM COMPILER

The project goals for the Intel Itanium compiler have evolved over time. Initially the goal was to create a reference compiler, which would demonstrate the Itanium processor performance on a few key benchmarks. A few years into the project, the strategy was changed, and we decided to develop a robust product compiler that would not only demonstrate performance, but also enable Independent Software Vendors (ISVs) and Operating System Vendors (OSVs) to port their applications and Operating Systems (OS) to the new Itanium architecture. Making the Intel compiler a product would give it credibility among our other compiler vendors, and it would create an incentive in the compiler industry to achieve as high performance as possible on this new architecture.

We knew from the beginning that many ISVs would not want to switch vendors, so it was important to incite the other compiler vendors to expend their resources on the performance aspect of the Itanium compilers.

Time Line

It was well understood from the beginning of the Itanium project that compiler technology was a key ingredient of the project. Therefore, the compiler design was started at the same time as the architecture definition and the chip design: i.e., in late 1994.

Key Phases and Milestones

The key phases and milestones for the project were as follows:

- 1994 – 1997: Design
- 1996 – 1999: Development
- Q4, 1996: Windows NT boots to command prompt with Intel's compiler in two months.
- Q3, 1997: First Non-NT OSV Engagement

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

[2*] Other names and brands may be claimed as the property of others.

- Q2, 1999: Monterey64 boots with Intel's Compiler on the simulator
- 1999 – 2001: Performance tuning on simulator, then on hardware
- Q1, 2000: First access to hardware

Software Development Kits (SDKs) were released to firmware (FW) developers, ISVs, and OSVs very early in the project:

- Q1, 1996: SDK 0.1 first FW release
- Q4, 1997: SDK 0.3 64 bit analyzer (for ISVs to make their application 64-bit clean)
- Q2 - Q4 1998: SDK0.4 and SDK 0.5, enabling of OS and FW development
- Q1, 1999 SDK0.6 release criteria included Intel assembler and building of NT OS 4.0
- Q3, 1999: SDK1.7 first release with strong emphasis on applications. Release criteria applications included 3D Studio Max (C/C+), Games and Nag F90 (Fortran)
- Q1, 2000 SDK2.0 last simulator-based release
- Q3, 2000 SDK5.0 first HW-based release, enabling large application vendors such as Ansys, Nastran, Oracle, SQL, and Mentor Graphics

Tools

When the Itanium compiler and OS development started, a simulation environment was defined to run programs on a functional simulator called Gambit. The environment, called application mode, was used to test the very first Itanium compiler by providing basic runtime capabilities including file I/O and memory management. The compiler was used to get the Windows* and System V UNIX OSs up and running on the functional simulator. The OS support was limited to loader, OS kernel, and basic OS functionality.

A Software Development Vehicle (SDV) was designed to include OS drivers and a loader to start an OS on the functional simulator under a debugger. The tool was named SoftSDV [2]. It was widely used to debug and test the OS, compilers, and 64-bit applications. A complete Windows OS and System V UNIX OS along with runtime support was developed and run on SoftSDV (see Figure 1).

[*] Other names and brands may be claimed as the property of others.

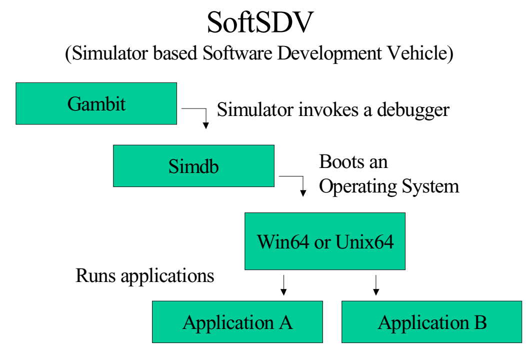


Figure 1: SoftSDV

A performance simulator called Emerald was also developed. It was run in application mode, on top of Gambit. For performance analysis of the compiler on Emerald, we mostly used the SPEC CPU95* benchmarks. To run these benchmarks on the simulator in less than 24 hours per benchmark, we had to use the lite inputs, a subset of the reference inputs that was borrowed from HP. Results from Emerald include cycle count and micro-architecture statistics such as cache hit rates, branch mispredict, and instruction distribution. However, it is important for a compiler to know where events such as cache misses and branch mispredicts happen in order to prevent them. A special version of Vtune™, called Vtune++, gathered event information, including the Instruction Pointer (IP) from the Emerald simulator. It used the core GUI functions of Vtune, but could provide information at the bundle level, such as cache misses or time-based IP sampling

PERFORMANCE ANALYSIS

The compiler performance analysis has several goals:

- *Performance projection*.
- *Guiding and validating the optimization development*. Analysis of compiler-generated code shows where and what optimizations are needed. Performance analysis is also important to check if optimizations are doing what they are supposed to do, after they have been developed.
- *Performance regression testing*. Interactions between different optimizations can be complex, and

™ Vtune is a trademark or registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.

regression testing is needed to ensure that optimizations keep working as the compiler evolves.

The SPEC CPU95 benchmarks were used for performance analysis and projections. By the end of 1997, the compiler reached the long awaited level of 30 SPECint95*. At that time, the chip design went through two die diets in a row, and the compiler performance stayed at the 30 SPECint95 level for more than six months. New compiler optimizations were making up for lost silicon performance. Floating-point performance (shown with squares in the graph) was not affected as much by the die diets as was integer performance. However, the compiler team developed a new High-Level Optimizer (HLO) to take advantage of the Itanium™ architecture, but the HLO did not come on line before the first half of 1999 [1]. Figure 2 shows the performance improvements over time on SPEC CPU95 with the Intel Itanium compiler.

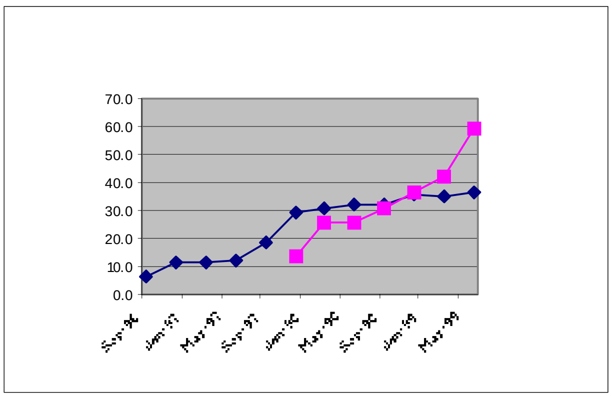


Figure 2: SPEC CPU95* performance over time

The chart in Figure 3 shows the difference between simulator-predicted performance and actual hardware performance. A negative number means that the simulator was optimistic and that the hardware did not perform as fast as the simulator. SPECint95 performance was predicted to be 13% faster than actual performance. The difference was partially due to the use of lite inputs on the simulator, which did not always correlate well with the reference inputs used on hardware. The other source of difference was inaccuracies in the simulator.

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

™ Itanium is a trademark of Intel Corporation or its subsidiaries in the United States and other countries.

® Intel is a registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.

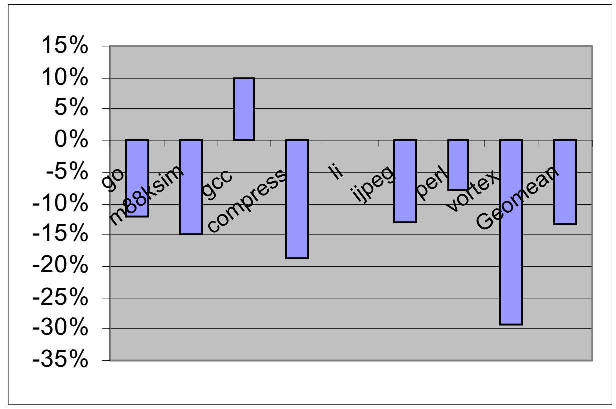


Figure 3: Simulator accuracy on SPECint95*

Overall, the floating-point performance projection was more accurate than the integer projection, see Figure 4; since it came within 3% of the hardware performance. However, the simulator was very inaccurate on benchmarks that were memory intensive. The memory model of the simulator was fairly simplistic and did not model overlapping requests to memory. We were fortunate to have two benchmarks grossly underestimated, and two benchmarks grossly overestimated. The errors compensated each other and gave an overall good prediction.

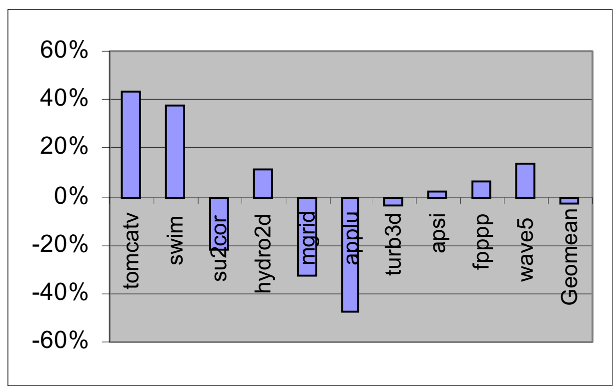


Figure 4: Simulator accuracy on SPECfp95*

ENABLING AN OPERATING SYSTEM VENDOR

The first enabling Software Development Kit (SDK) shipped in November 1997 to multiple Operating System Vendors (OSVs). The goal of the SDK was to provide OSVs with the tools necessary to port their Operating Systems (OSs) to the Itanium architecture. The initial

™ Itanium is a trademark of Intel Corporation or its subsidiaries in the United States and other countries.

contents of the Enabling SDK were the compiler, the assembler, and a functional simulator called Gambit. These tools were all UNIX-based. But in September 1998, the simulator no longer supported the UNIX platforms; it only supported Windows platforms. Every one working on a port to the Itanium architecture had to have a Windows-based system to be able to continue. This was a hard sell to non-NT OSVs.

During the life of the Enabling SDK project (10/97 to 12/99), there were seven major releases, occurring approximately once per quarter.

The Monterey Story

Although the Enabling SDK was provided to several OSVs [3], Gemini64 (the Operating System provided by SCO) was our reference platform. We held weekly meetings to discuss status, progress, plans, and of course, compiler bugs or needed features.

In October 1998, the compiler teams engagement with IBM accelerated. IBM had already decided to port their AIX* Operating System using the Intel compiler. At the same time the announcement was made, SCO, IBM, and Sequent decided to collaborate on their OS effort, and they started the Monterey Operating System project. This OS was designed to take advantage of strengths in each code base.

Intel compilers had never before been used to port operating systems, so the testing done in the compiler group did not always catch problems seen while porting the Monterey OS. We were working closely with the OS team and providing them pre-release compilers to test the kernel. After a few releases, the compiler team decided to make the OS build a part of the compiler release criteria.

The main challenge on the Enabling SDK Project was that there were no constant factors:

- Aggressive performance-driven optimizations were under development, making the compiler unstable at times. A compiler simply meant for OSV-enabling would have been more conservative in its optimization work.
- The OS was also under constant change and development.
- Development was underway on the simulator, where the changes for new External Architecture Specification (EAS) revisions were incorporated.

Because of all the changes, debugging was a challenge.

Testing the UNIX SDK was another challenge:

- There was no runtime environment available on UNIX. We had written a minimal set of runtime functions to interface with SoftSDV for our testing purposes. Each OS vendor had a different runtime environment.
- Our UNIX tool chain did not include a full-feature linker. Again OSVs were using their linker, which often behaved differently from what we were using.
- Not having a UNIX-based SoftSDV forced us to change our testing tools to work across platforms.

In Q3 1999, the bring-up plan gave all OS vendors a two-week period to bring up their OS on the Itanium Hardware in Intel Dupont. Monterey was scheduled for the week starting September 14. It took Monterey less than 24 hours to bring up the kernel on SDVs. According to IBM, this was the fastest they ever booted an OS.

The Monterey binary that booted on Itanium hardware was built with the Electron Compiler with optimizations enabled. No workarounds were required! This was quite a success story for the compiler and SoftSDV. [http://aix5l.ihost.com/innovations/index.shtml]

The presence of compilers for the porting effort along with simulators and other tools (enabling SDK) are what made the difference when hardware was available.

PRE-SILICON COMPILER TESTING AND VALIDATION METHODOLOGY

The Intel Itanium compilers were tested in two different environments to cover the C/C++ and Fortran compiler language functionality and to test optimizations and code generation, based on Itanium.

Testing Environments

Compiler testing was divided into two parts: Application-Mode testing on the functional simulator Gambit, and System Mode testing using SoftSDV on the Win64* Operating System (OS).

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

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

® Intel is a registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.

™ Itanium is a trademark of Intel Corporation or its subsidiaries in the United States and other countries.

Application-Mode Testing: Compiler language conformance tests, coverage for optimizations using C/C++, and Fortran compiler test suites were mostly done in application mode. This resulted in fast turnaround on test results and overnight test runs. However, application mode testing was limited to basic runtime file I/O, and memory management. It could not be used to exercise the C/C++ language features, which required a complete Operating System (OS) runtime environment. It also could not test recovery code, since Gambit did not simulate page faults.

SoftSDV-Testing and Debugging Environment: The C/C++ functionality in the areas of structured exception handling, C++ exception handling, runtime type information, and C++ I/O required a complete OS runtime support. SoftSDV was the only way to exercise these C++ functionalities and the compiler and operating system stack unwinding support. Unalign and NAT exceptions specific to Itanium are handled by the OS, and so the compiler-generated code for speculation can only be tested with the OS. We relied on SoftSDV to have comprehensive test coverage with the Itanium compilers.

Compiler tests running on a simulated Win64 operating system had a slow turnaround time. This environment enabled a SoftSDV cross-testing environment for the compilers. Tests were compiled on the host system; SoftSDV then booted the operating system on the simulator and executed the compiler-generated executable.

Debugging programs on SoftSDV was very challenging. The debugger and OS were still in the development phase. Moreover, the OS/debugger interface was not very stable. Several times we ran into OS problems as we debugged compiler issues. Compiler debug engineers had to understand the interaction between the debugger and the OS. They also had to learn the OS interface for trapping the exceptions. The compiler functionalities, which were the hardest to debug, were structured exception handling and C++ exception handling.

A good example of this would be that we would set a breakpoint to trap an exception raised by the an exception-handling test case, and we would end up in the OS exception dispatcher, having trouble in unwinding. We would then enable OS breakpoints to debug the unwinding problems, to uncover the compiler or OS stack unwinding issues.

Netbatch Tools

To improve the test coverage for the compiler, a networked batch-testing software was integrated with the compiler test tool, widely known as Netbatch. With the Netbatch compiler, testing was completed in 16 to 18 hours. Netbatch allowed regression testing to be run nightly. This allowed fast turnaround on the bugs introduced daily.

Configuration: The test tools were modified to use the Netbatch APIs. The client systems known as Garcons were set up with SoftSDV tools, a Win64 OS, and drivers. The compiler-testing tool called TC, is written in Perl and uses the Netbatch capability to distribute compiler tests across all the machines in the Netbatch pool. TC works in the SoftSDV environment and in application mode.

Challenges with changing OS and SoftSDV: OS development was still in progress, and SoftSDV was constantly being modified for additional support for new drivers. Every upgrade for a new version of the OS and SoftSDV required changes to be made to the testing environment and testing tools, such as TC. We had to re-install the new versions of SoftSDV, the OS, and related tools on all Netbatch client systems before testing the compiler in the new environment.

SDK Development and Test Cycle

Intel compiler and OS teams worked closely with Microsoft, HP, and SCO to define Software Conventions and Runtime Architecture for the Itanium architecture. An NT Application Binary Interface (ABI) was defined to support the Windows NT operating system features for Itanium. We also defined a System V UNIX Processor-specific ABI for the Itanium architecture. The conventions, and OS ABI, formed a norm for the OS and for tool development on NT and UNIX platforms. On the NT platform, a complete Software Development Kit (SDK) was being developed, see Figure 5. It included the OS, SoftSDV, and the development tools: compilers, the assembler, the linker, and OS runtime libraries

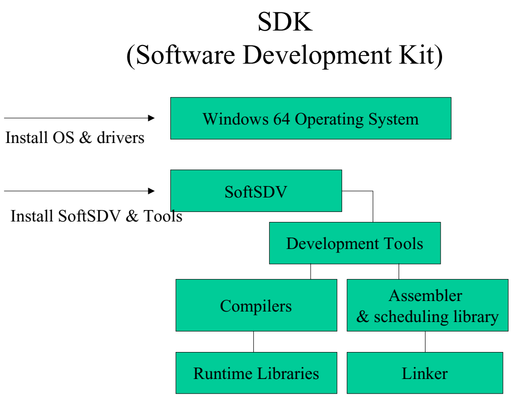


Figure 5: Software Development Kit

Microsoft worked with the Intel linker team to implement the ABI support and extended their executable file format to include 64-bit virtual addresses. The new file format was defined as an extension to the existing PE32 and was called PE32+. The compiler and assembler implemented the support for the PE32+ and the NT ABI. For example, support for function pointers known as "plabel," unwind tables, exception handling tables, and various 64-bit relocations were added to the tools and to the OS runtime. The SoftSDV included firmware and drivers, which were compatible with the Windows NT ABI. Figure 6 describes the dependencies within the SDK components. These dependencies made it hard to stabilize the SDK across engineering groups. All the OS and compiler ABI compatibility issues were tested on the simulator to be silicon ready. The pre-silicon releases enabled porting applications to 64-bit, so they would be ready to run on the Itanium hardware.

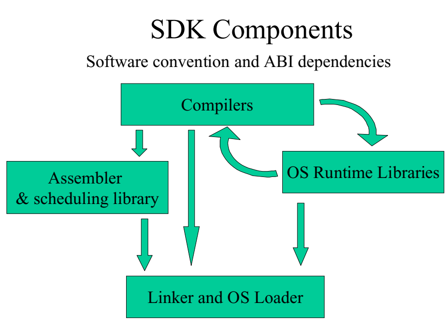


Figure 6: Dependencies within SDK components

Compiler Language Extensions for the Itanium Architecture

We defined extensions to the Windows NT C compiler to support the Win64 ABI. Among them were the following:

- pragma data_seg to define data as short or long was introduced in the Microsoft and Intel C compilers
- __ptr64 and __ptr32 to support 64-bit and 32-bit pointer types
- several compiler intrinsics were implemented to support the NT operating system functionality requirements in place of inline assembly

SoftSDV was used to test the mixed 32-bit and 64-bit pointer support by forcing programs to be loaded at higher than 2**32 bit address.

CONCLUSION

Developing a compiler in a virtual environment presented many challenges, but we were able to do everything that we would have done on hardware: performance projections, nightly regression testing, and Operating System Vendor (OSV) and Independent Software Vendor (ISV) enabling. Since the Intel Itanium processor was the first of a new architecture, no other methodology was possible, but it had a high cost in people to develop and maintain all these complex tools. We also had to limit the scope of performance analysis and stability testing to accommodate the simulator speed. We have learned that hardware is needed at least six months before the first release of a compiler product in order to analyze the performance and stability of real applications, and not only benchmarks and test suites.

ACKNOWLEDGMENTS

We acknowledge the work of all members of the Intel Compiler Lab. Special thanks to Tom Joyce, Bill Savage, Suresh Rao, and Steve Skedzielewski for their thorough reviews.

REFERENCES

[1] Carole Dulong, Rakesh Krishnaiyer, Dattatraya Kulkarni, Daniel Lavery ,Wei Li, John Ng, and David Sehr, "**An Overview of the Intel IA-64 Compiler**," *Intel Technology Journal,* Q4, 1999.

® Intel is a registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.

™ Itanium is a trademark of Intel Corporation or its subsidiaries in the United States and other countries.

[2] Richard Uhlig, Roman Fishtein, Oren Gershon, Israel Hirsh, and Hong Wang, "**SoftSDV: A Pre-silicon Software Development Environment for the IA-64 Architecture**," *Intel Technology Journal*, Q4, 1999.

[3] Kathy Carver, Chuck Fleckenstein, Joshua LeVasseur, and Stephan Zeisset "**Porting Operating System Kernels to the IA-64 Architecture for Pre-silicon Validation Purposes**," *Intel Technology Journal*, Q4, 1999.

AUTHORS' BIOGRAPHIES

Carole Dulong has been with Intel for over eleven years. She is a principal engineer and staff member of the Intel Compiler Lab. Prior to joining Intel's Microcomputer Software Laboratory, she was with the Itanium architecture group, where she headed the Itanium multimedia architecture definition and the Itanium experimental compiler development. Her e-mail is carole.dulong@intel.com

Priti Shrivastav joined Intel's Itanium compiler code generator group in September 1995. She moved on to the Enabling SDK and Development Group as the technical lead in 1996. For three years she was the Itanium Compiler Evaluation team manager and now co-manages the IA-32 and IPF Product Compiler team within the Intel Compiler Lab. Priti received her Masters degree in Computer Science from Ohio University in 1983. Her e-mail is priti.shrivastav@intel.com.

Azita Refah has been with Intel for over eleven years. She leads the Linux development team in the Intel Compiler Lab. She graduated from Northeastern University (in Boston) with a B.S. degree in Computer Science. Her e-mail is azita.refah@intel.com