HP Labs Innovation Research Program

2008 Research Topics

Region 1: Americas

Universities are invited to address specific research topics as assigned to a number of regional groupings, based on HP Labs worldwide locations. Country alignments by region are summarized in the figure below. Universities located in the following regions / countries are invited to submit proposals to the research topics in this document:

- United States
- Canada
- South America
- Central America and Caribbean

Topic 1: High-speed Document Sensing and Imaging in Digital Presses 3
Topic 2: Magazine Advertisements: Measuring Effectiveness of Print versus Screen 3
Topic 3: Aqueous Dispersion Stability 4
Topic 4: Thin Film Deposition and Patterning 5
Topic 5: Liquid Crystal Colloids 6
Topic 6: Behavioral and Experimental Economics: Discovering the Principles behind Human Decision-Making 6
Topic 7: Social Computing: Methods for Harvesting the Collective Intelligence of Groups 7
Topic 8: Custom Networks on Demand 8
Topic 9: Next-Generation Multimedia Communication and Collaboration Systems 9
Topic 10: Artistic Rendering for Video 10
Topic 11: Interactive Communication and Control in an Immersive Facility 11
Topic 12: Exascale Datacenters 12
Topic 13: Nanophotonic Many-Core Architecture 13
Topic 14: Sustainable Solutions for Data Centers 14
Topic 15: Creating a Sustainable IT Ecosystem 16
Topic 16: Scalable Storage 17
Topic 17: Natural Language Processing and Computational Linguistics Tools for Taming the Information Explosion 19
Topic 18: Real-time Business Intelligence Infrastructure 20
Topic 19: Visual and High-performance Analytics 21
Topic 20: Business Optimization: Planning under Uncertainty, Resource Allocation, Procurement, Auction, and Risk Management 22
Topic 21: Personalization, Recommendation and Customer Behavior Analysis 23
Topic 22: Creating a Business Operating Environment in the Global Services Ecosystem 24
Topic 23: Dynamical nanoelectronic devices and circuits 25
Topic 24: Semiconductor Nanowire Photonics/Electronics 26
Topic 25: Signal Processing in On-Chip Optical DWDM Networks 26
Topic 26: High-Speed Low-Cost CMOS-Compatible Optical Engine 27
Topic 27: Identical Single Photons 28
Topic 28: Engineering Quantum Technologies 28
Topic 29: Integrated Si and Ge Nanowire Sensors, Electronics, and Photonics 29
Topic 30: Distributed Quantum Information Processing 30
Topic 31: Graphene Deposition 31
Topic 38: Security Analytics: Understanding the threat environment 32

Print Production Automation Lab

Topic 1: High-speed Document Sensing and Imaging in Digital Presses

The printing industry is transitioning from long-run offset print jobs, where thousands of identical pages are printed, to short-run digital print jobs, where every page can be different and customized for a user. High-speed digital presses are enabling this transition. In the past, publishers had to print thousands of books or pages and maintain an inventory, but high-speed digital presses now enable a book or single page to be printed on demand when it is requested. There is no need to maintain inventory or predict the number of books or pages that will sell, which opens up the printing market to many more content providers.

While high-speed digital presses and digital printing are offering new opportunities for print-on-demand systems, these presses have the additional overhead of confirming that each page is printed correctly. With traditional long-run offset presses, badly printed pages could be simply discarded because all the pages were identical. In digital printing, pages cannot be discarded because every page is different, but misprinted pages must be detected and re-printed. Detection of misprinted pages is complicated by the high print speed of these digital presses. Conventional imaging systems cannot integrate long-enough to get adequate signal-to-noise ratios to produce usable images. A solution to imaging high-speed document content is needed.

Research Questions

- What novel imaging and sensing architectures will be needed to maintain adequate signal-to-noise ratios for high-speed printing, where sensor integration times must be extremely short? What changes to the hardware or software will be required? How will such a system work?
- How can we measure properties of documents quickly and accurately? These properties include color, spatial content, gloss, texture, fluorescence, etc.

For more information, visit: http://hpl.hp.com/research/print_automation.html

Print Production Automation Lab

Topic 2: Magazine Advertisements: Measuring Effectiveness of Print versus Screen

The technology to print high quality magazines on high end digital presses is here! This enables customized and personalized content and advertising to be integrated into short and medium runs of magazines. It will also enable new magazine publications to be created with a variety of efficiencies for both content consumption and for targeted advertisement. This may change and shift the huge magazine market in the near future since ads are usually the major revenue and profit stream for magazines. Printed magazine ads are the 2nd biggest segment of advertising spending (second to TV and more than double the online internet ad spending in 2007).

Our goals In this research program are to analyze what readers are willing to pay for different formats of the same magazine (printed vs. displayed on screen). We seek to answer some of the following questions related to magazine consumption and ad effectiveness:

- What are the printed magazine properties that give it a sustaining edge over screen displays?
- Will our children prefer high quality images and content consumption of food, travel, home design, fashion and other such magazines in physical paper magazine format or will screen displays be the dominant distribution path in the future?
- Is the media of choice an age or cultural thing?
- What medium (printed magazines vs. screen) is most effective for remembering the content you read? The promotion? The brand?

We also seek to analyze the advertiser perspective on the future properties for effective magazine ads and answer some of the following questions:

- What are the print attributes that would maximize the printed magazine ad effectiveness?
- Can advertisers get a better return on investment (ROI) when: image enhancements are implemented? When we use special colors to improve quality or have unique colors pop out? When we customize the ad theme (for example, background color) to better engage with a predefined recipient group?
- Does HP have an opportunity to enhance its print and / or screen value proposition and presentation opportunities for the advertiser and end users?
- How should ads and content consumption be split between screen and printed (magazine) form for the different market segments HP serves in order to maximize effectiveness for both advertisers and end readers?

Research Questions

- What are the statistical methods used for cross-media measurements?
- What is the value added of using more than 4 colors for printed magazine ads?
- What is the value added for customized or personalizing ads?
- What are the advantages of printed magazine ads vs. an online version?
- What are the parameters for a differentiated pricing model of magazine advertisement?
- What is the effect of the socio-economic status (SES) of the reader on the above questions?

Commercial Print Engine Lab

Topic 3: Aqueous Dispersion Stability

Colloidal stability plays a critical role in the success of numerous applications ranging from printing and coatings to pharmaceuticals. In general, colloidal stability is perceived to be governed by the intricate balance between the attractive and repulsive physical forces among the particles in the system, and is well described by the classical DLVO theory. In recent years, water-based polymer dispersions have attracted a great deal of attention due to their increasing technological importance. However, undesirable dispersion stability problems, particularly in complex media, remain daunting challenges to overcome. Moreover, the prevention of coagulus formation within polymerization reactors still poses poorly-understood problems. It is widely acknowledged that factors affecting colloidal stability are active at the interfaces. Regarding advanced printing technology, the presence of additives and the associated complex physico-chemical interactions among these additives further complicate the scenarios. In addition, experimental tools and analysis methodology available to

researchers today are rather limited and often of low throughput. The potential power of molecular modeling has also not been fully utilized to study the problems at hand.

Hence, a research program objective is to elucidate the structural-compositions-property relationships in such a challenging matrix. Another aim of the program is to extend and advance our current understanding of the DLVO theory, and the governing kinetics and thermodynamics associated with such a system. An additional objective of the program is to look at colloidal stability issues while employing advanced molecular and mesoscopic modeling and dynamic simulation to complement the empirical understandings. Finally, a further extension of the program is to develop innovative, combinatorial approaches to perform massively parallel characterization, instrumental analysis and automation of colloidal stability prediction.

Research Questions

- What are the general rules to facilitate long-term colloidal dispersion stability for both water-based and apolar systems? What is the dominating key contributor to the stability "equation"?
- Are existing theories adequate to explain all phenomena, including physico-chemical effects?
- Could state-of-the-art molecular and mesoscopic modeling techniques and dynamic simulations be employed to satisfactorily understand the colloidal stability problem? Could they be used to predict long-term colloidal stability?
- How could we increase the throughput of various colloidal systems analyses while extracting important experimental observations? Are existing techniques adequate for fulfilling the goal?
- Could combinatorial approaches be adapted for this requirement?

Information Surfaces Lab

Topic 4: Thin Film Deposition and Patterning

HP Labs is developing technology for roll-to-roll (R2R) fabrication of electronics on flexible webs. Two of the challenges in this research area are thin film stack defects and large area 3D mastering. HP Labs is seeking proposals that address both of these issues, including TFT stack deposition on plastic films bonded to wafers, and bonding and de-bonding of these films. HP Labs is also interested in gaining access to Gen 2 TFT experimental fabrication facilities.

HP Labs currently uses thin film stacks that are R2R deposited on flexible webs. We know that these have much higher defect rates than similar thin film stacks deposited on Si wafers. We don't know which of the many differences between these two media are significant. We are looking for a partner who can bridge the gap between these differences by producing film stacks on wafers and various webs bonded to substrates under similar conditions. HP may provide various films and substrates and undertake comparison testing.

Similar capabilities are needed for large scale 3D mastering. However, in this case, a Gen 2 fab capable of experimental runs and large area lithography is also needed. We are looking for a partner whom we can work with to develop the bulk or applied film etching and deposition processes necessary to create our 3D masters on Gen 2 substrates.

Research Questions

- What is the effect of substrate cleaning on defects in thin film stacks?
- Can micron deep features be etched in glass substrates with sufficient wall angles for imprint lithography?

For additional information, visit: http://www.hpl.hp.com/news/2008/apr-jun/sail_award.html

Information Surfaces Lab

Topic 5: Liquid Crystal Colloids

HP Labs is interested in colloidal systems consisting of optical materials such as pigments dispersed in liquid crystal fluids. Such systems may have interesting optical and electrooptical properties which could serve as the basis for novel display technologies with improved optical and electrical properties. We wish to find stabilization treatments that give well-behaved, long term stable dispersions in liquid crystal materials – in particular to enable non-spherical particles to align with the LC director. HP Labs is seeking proposals that identify the physical properties of such systems – in particular their optical and electro-optical properties.

Conventional colloidal systems already offer a rich variety of behaviors and properties – the use of anisotropic fluids as a medium offers the potential for extending these in novel applications. If the dispersed phase is also anisotropic – e.g. non-spherical particles – the orientations of the anisotropies could be coupled in useful ways. A key challenge in this is to find and understand 'stabilization' mechanisms: with anisotropic dispersion media, distortion energy is an additional factor in the stability and properties of a dispersion. Conventional stabilization treatments may therefore need to be modified or replaced with ones which are more compatible with the host fluid.

Research Questions

- Find novel stabilization treatments that give well-behaved, long term stable dispersions in liquid crystal materials – in particular to enable non-spherical particles to align with the LC director
- What are the physical properties of such systems – in particular their optical and electrooptical properties?

Social Computing Lab

Topic 6: Behavioral and Experimental Economics: Discovering the Principles behind Human Decision-Making

Since the 1990s, behavioral and experimental economics have become increasingly important to, and integrated into, business process research, including management science, supply chain operations and risk management. There are still many unanswered scientific questions in these contexts. For example, we have limited understanding of the behavioral effects of increasingly complex decisions.

HP Labs is interested in the fundamental science of human decision-making, and in applying to apply this knowledge to the engineering of business processes in the contexts of supply chain operations; business forecasting; behavioral risk management; and reputation mechanism design. Our research goals are to develop scientific principles and theories regarding the issues of risk behavior; decision-making in the face of complexity; sequential decisions in a dynamic environment; and strategic use of information. We also are interested in 1. experimental work that will validate or disprove candidate theories; and 2. obtaining observational evidence of behavioral effects as the basis for future theories.

Research Questions

- Standard game theory assumes agents have infinite computational power, which is obviously not true. What are the principles behind how humans deal with complexities in decisions? What is the right measure of cognitive difficulties in a decision problem?
- Standard dynamic control theory does not describe how humans make decisions in multiple-period settings, particularly in inventory control problems. What is the right way to model people's objectives if using discounted sums of expected utility as a measure is not accurate? How might we model a human's anticipation of their future actions if using backward induction is too difficult? To what degree are humans myopic in their decision-making?
- How do issues of social preference (such as trust, fairness, altruism reciprocity, etc.) affect the strategic use of information (for example, in forecast sharing)?

For additional information, visit: http://www.hpl.hp.com/personal/Kay-Yut_Chen/

Social Computing Lab

Topic 7: Social Computing: Methods for Harvesting the Collective Intelligence of Groups

The past decade has witnessed a significant transformation in the way people use computers and the Internet to interact and exchange information:

- Content is now coactively produced, shared, classified and rated on the Web by millions of people.
- Commerce, social networking, opinion formation and large collaborative efforts are increasingly taking place online.
- Mobile devices providing easy, continuous access to Web services have become common.

This collective intelligence "cloud" represents a new phenomenon which is as yet poorly understood and poorly leveraged by existing applications and services.

Harvesting knowledge from the collective intelligence represented by the web and enterprises is useful and an important problem. The objective is to understand how to harvest this information and display it, taking into account the limited attention of people, and thus to design and implement a mechanism for extracting it.

Research Questions

- How does one improve the value that users get from the collective-intelligence "cloud" in an increasingly mobile and connected world?
- How is information created, evaluated and consumed online?
- What approaches will provide incentives for and facilitate knowledge sharing?

For additional information, visit: http://www.hpl.hp.com/research/scl

Multimedia Communications and Networking Lab

Topic 8: Custom Networks on Demand

As the IT industry at large and HP move to the "Everything as a Service" (EaaS) or Cloud Services model, and as all forms of mass communication converge on the Internet platform, predictable, robust, high-performance and cost-effective networking services will be crucial. Operational networks today are straining to keep up with the volume and variety of applications, and their diverse requirements of end-to-end network connectivity with sufficient and predictable quality, security, reliability, and mobility.

Even in networks in a single domain of control, such as enterprise or datacenter networks, the currently-deployed network architectures are neither flexible nor extensible enough to provide the required performance, manageability, and adaptivity to new services. IP networking began as a best-effort bulk data transfer service. Network elements were regarded as simple devices whose only function was packet forwarding. But technological advances and the broadening of network usage have moved current networks away from this simple model. Modern applications (videoconferencing, remote visualization, cloud computing, distributed computing applications) have created a demand for quality of service (QoS), manageability, reliability and security on a networking platform that was not designed to accommodate any of these things.

The grand challenge is to rethink how networks are designed and deployed to continuously add new capabilities to, and support new economic models on network infrastructures. Recent work in overlay networks and Computing Grids have demonstrated the practicality of distributed systems as deployment vehicles for modern networking applications. Programmable network elements, in-network computing and storage are important new deployment tools for embedding distributed systems methods in the network fabric. This approach will enable us to reduce the operational complexity of networks; provide end-to-end network properties to individual sessions; support sophisticated diagnostic services; and decrease the overall power consumed by the network infrastructures.

Research Questions

- New network architectures in the home, datacenter, enterprise and public consumer networks, including wireless and wired environments:
  - Novel programmable switch/router architectures
  - In-network diagnosis of network service problems; scalable network monitoring and inference
  - Testbed design and development
  - Networks at a Terabit/second and beyond
- Energy efficient networking
- New solutions for seamless mobility, disconnected, and sporadically-connected operations, specifically including networking in developing regions
- New solutions for end-to-end security and privacy that balance flexibility, simplicity and functionality
- Delivery of end-to-end network services without necessarily the cooperation of every network component
- Novel end-user , management, and high-performance applications that highlight new requirements from the network infrastructure, including remote immersive experiences

Multimedia Communications and Networking Lab

Topic 9: Next-Generation Multimedia Communication and Collaboration Systems

HP Labs is paving the way for next-generation multimedia communication, collaboration, and entertainment systems. We welcome research proposals in the areas of creating novel algorithms, architectures, and systems for multimedia acquisition, processing, compression, streaming, rendering, and interaction for all relevant signal modalities, including acoustical, visual (both natural and synthetic), 3D, and infrared.

As transportation becomes more expensive and drains the world's resources, there is increased need for life-like and real-time communication systems, which we call remote presence. Until recently, the technology for remote presence was primitive and inadequate. The resulting systems offered so little benefit and were so difficult to use that users typically gave up on them after a few initial trials.

However, in the last five years the computational and networking resources needed to achieve usable remote presence have become available. Evidence is provided by the widespread use of low-quality zero-cost Internet-based systems, medium quality commercial video conferencing systems, and high-quality enterprise systems such as HP Halo. These systems still suffer from the fact that either the quality is low or the price is very high, and offer near-lifelike experience only in tightly controlled settings such as specially-designed and outfitted conference rooms. We would like to further improve remote presence by: improving the user experience; eliminating the need for expensive dedicated rooms; better exploiting computational and networking resources; incorporating state-of-the-art audio-visual capturing, processing and rendering; and incorporating new sensing modalities.

To achieve the ambitious goal of life-like remote presence, we must better understand key multimedia acquisition, processing, coding and networking technologies. With respect to the former two challenges, these include, for example, microphone array processing, video segmentation for background replacement and for gesture-based human-computer interaction, depth sensing, distributed coding and processing, augmented reality, and 3D rendering. With respect to the latter two challenges, there remain many questions about how to create and maintain a good user experience across a wide range of networks, media sources, and devices. We need to find the right balance between network over-provisioning (high cost, high quality, low latency, limited applicability) and algorithmic approaches (lower cost, less quality, higher latency, more flexible).

We must understand how to use these technologies to build more effective remote presence systems and how to assess the performance of those systems. For this reason, we encourage research into new algorithms, techniques, and architectures that provide a high-quality low-latency remote presence and multimedia experience at a controlled complexity.

Research Questions

- What are the underlying psychological, human factor, design and user experience principles for successful remote presence systems? How can these principles be embedded in modern computational devices and on top of modern networking infrastructure? What are the tradeoffs between computational and network resources? How can we measure the performance of a remote presence system, both subjectively and objectively?
- What new multimedia processing capabilities can dramatically improve remote presence systems, and how can they be implemented? How can multiple modalities jointly provide a performance that is larger than the sum of the components? How can we use signal processing to achieve better quality and new forms of human-computer interaction?
- How can we implement and distribute remote presence processing capabilities on adaptive configurations of participating devices as well as on non-managed heterogeneous networks? How do we guarantee fundamental algorithm and system performance in terms of latency and audio-visual quality?

- What are good end-point mechanisms for low-delay adaptive streaming over best-effort channels? What combination of source and channel coding techniques are most effective for error resilience? What are effective ways to exploit advanced compression structures such as scalable, multiple-description, and distributed coding? What are effective ways to perform channel estimation and prediction at endpoints?
- How can we achieve high-definition wireless streaming? How can we establish reliable wireless HD connectivity in a multi-hop 802.11 environment? How can we guarantee high video quality through techniques such as cross-layer optimization, traffic prioritization and management?
- How can we improve the state-of-the-art streaming architectures? Beyond basic end-point adaptation, what are the costs and benefits of coordination with infrastructure or other end-points, e.g., in a content delivery network, a peer-to-peer system, or a hybrid of these techniques? What new applications do these streaming architectures enable?
- How can we best stream compound video? What are the best ways to transmit compound video, which may contain both camera-acquired and computer generated content?

For additional information, please visit: http://www.hpl.hp.com/research/multimedia_communications.html

Multimedia Interaction and Understanding Lab

Topic 10: Artistic Rendering for Video

Home media (photos, video) is a key source of raw material for interesting entertainment. However, consumer expectations are driven by the professional media industry. How do we bridge the gap – making compelling visual experiences from home media collections? Current tools require considerable manual effort to produce pleasing results. However, many professional features are either omitted or drastically restricted in order to simplify user interfaces for consumers. Given the same source material, this leaves a large gap between what consumers can achieve with their media and what could be produced by professional graphic artists and designers. The key challenge is to reduce this gap by automating artistic and design knowledge and combining this with media understanding in ways that can be applied robustly to consumer media.

HP Labs is interested in this space as part of a future of casual, personalized entertainment provided on multiple screens around the home. Screen-based displays should present material in a compelling way by applying design knowledge and media content analysis to arrange source material both spatially and temporally in pleasing ways. In addition to arrangement issues, we believe that appropriate manipulations of the source material can add creative novelty to the display. Examples of appropriate manipulations include artistic rendering effects, automatic morphing, and cartooning. All of these can be extended in temporal ways, juxtaposing content which blends or contrasts in interesting ways.

A key requirement is a multi-disciplinary approach which combines the capabilities of computer scientists in the fields of computer vision and graphics with the aesthetic artistic and design skills from the creative disciplines.

Research Questions

We are particularly interested in the **temporal** aspects of **non-photorealistic rendering** (NPR), as applied to consumer source media. By this we mean to include the fields of **artistic rendering, morphing, cartooning** and other forms of **stylization**. We are interested in approaches for both generating temporal output media from static input images and stylizing source video clips.

In all cases techniques should target typical consumer media, much of which is people-centric. Successful approaches will involve professional designers and artists to ensure high aesthetic quality. The relevant measure of aesthetic success is appeal to typical consumers.

Specific research questions include:

- How can techniques which currently generate static outputs be extended to generate compelling video outputs?
- How can current techniques be combined to generate compelling video outputs (e.g., morphing + cartooning)?
- How can NPR techniques be successfully applied to source video clips?
- To what extent can techniques be fully automatic – parameters being selected from source media analysis?

Multimedia Interaction and Understanding Lab

Topic 11: Interactive Communication and Control in an Immersive Facility

Development of multi-view immersive environments opens new opportunities in interpersonal communication. Geometry plays a major role in this. Capture and display of multi-viewpoint video requires detailed understanding of the placement and orientation of the different devices used (in this case, cameras and projectors), or their geometric calibration. Once calibrated, an immersive environment can provide both output (display) and input (gestural communication) for participants in its space.

Key challenges in this research area include acquiring high quality understanding of the relationships among system components and users to allow metric operation in that space; developing means for facilitating understanding the intent of those users as expressed by their motions and gestures when operating in the immersive environment; and communication of the acquired information for immediate three dimensional reception across sites.

We wish to be able to represent the form and behavior of participants in such a space; to present displays of each side to the other in compelling three dimensions; and to permit users to communicate through gaze and actions. Many capabilities are implied by this, including auto-stereoscopic display; advanced calibration; geometric modeling (including kinematics); and abstraction over motions for communication and control.

Research Questions

Proposals should describe a novel and intellectually challenging project on displaying, representing, and responding to participants in an immersive multi-view communication facility. Research questions to consider include:

- The physical parameters of an immersive environment must be sufficiently accurately determined to enable remote users to interact as though in the same physical space. What developments can facilitate creation of such an environment, and how can one measure their effectiveness?
- Gesture can be subtle yet physically complex, requiring discerning judgment to discriminate among choices. Can methods be developed through visual observation and training to abstract a functional set of gestural directives that could unambiguously provide communication and control in a 3D visual interface for a specific application context?

Exascale Computing Lab

Topic 12: Exascale Datacenters

On the road to the exascale datacenter, growing server density, increasing network bandwidth and pervasive virtualization all create new challenges and open up new possibilities in system design. HP Labs is seeking good ideas that address datacenter-level issues in platform architecture, *-ability, power, and resource management, with a view to improving service quality and reducing ownership cost.

Recent technological and market trends make it important to fundamentally rethink the structure and goals of enterprise IT infrastructure. On the technical front, the number of CPU cores on a processor chip are expected to approximately double every generation. Furthermore, these processors are likely to include most of today's off-chip (board-level) functionality. At the same time, networking has been witnessing tremendous advances such as the explosive growth in bandwidth and the programmable network elements. Finally, virtualization techniques will become increasingly important to optimally utilize all these processing and networking resources. We believe that these trends will lead to unprecedented levels of performance and compaction and will thereby enable affordable exascale datacenters in the near future.

On the business front, there has been an increasing emphasis on meeting complex service level agreements (SLAs), and on minimizing the Total Cost of Ownership (TCO). This trend makes it important to pay close attention to technical and cost challenges. As the industry moves towards commodity building blocks, special-purpose changes at the hardware level are likely to continue diminishing. Thus, future innovations must be based on commodity hardware and industry standards, while delivering added value through the software and systems level.

The grand challenge is that all these trends, and the (often dramatic) changes they presage, motivate equivalent changes in emphasis of systems and datacenter research. In particular, an important "grand challenge" is the design of a radically-efficient and flexible computing infrastructure for next-generation enterprise IT solutions that provide exascale computing, through a single commodity-based high-volume compute substrate that is customizable to different markets with strong value propositions.

Research Questions

We believe that designing such a disruptive exascale datacenter requires holistic and interdisciplinary research across platform architecture, networking, virtualization, and management. Individual systems of the datacenter and their management interfaces need to be systematically redesigned, with their role in the broader data center in mind, to better leverage cross-connections across the hardware-software boundaries, and to better focus on efficiency. In addition, these new systems should be designed to leverage industry-standard and commodity components.

Key areas that HP Labs is interested in seeing proposals address include:

- Disaggregated and shared platform architectures for the datacenter
- Codesigned hardware/VM platforms and resource management
- New solutions for improved manageability, power, and availability
- End-to-end QoS management for datacenter networks
- Datacenter network convergence and management
- I/O virtualization
- VM appliances and management

- Secure, Scalable and Coordinated Cross-Layer Management
- Application of emerging technologies in the datacenter (e.g., flash)
- Architectures for unstructured data processing

Programming environment for multi-core specific areas:

- Can we fundamentally simplify threads- and locks-based parallel programming?
- Can a programming model based on, for example, transactional memory, cover the same space, but eliminate the deadlock concerns?
- Or do details related to memory models, I/O, large transactions, etc. unavoidably reintroduce similar complications?
- Alternatively, are there sufficiently simple alternate approaches that don't share memory, but still fully use the performance advantages of the existing hardware support for shared memory?
- Are there ways to improve the reliability of existing multithreaded applications without rewriting them, for example by dynamically avoiding deadlocks or other failures?
- Can we get better efficiency from a processor architecture that provides cores with heterogeneous performance characteristics? What kind of support does such a system require?

For additional information, please visit:

http://hpl.hp.com/research/exascale.html

http://www.hpl.hp.com/research/blades/index.html

http://www.hpl.hp.com/techreports/2007/HPL-2007-140.html

http://www.hpl.hp.com/personal/Hans_Boehm/c++mm/

Exascale Computing Lab

Topic 13: Nanophotonic Many-Core Architecture

CMOS nanophotonic technology may provide cost-effective ways to boost off-chip and cross-chip communication bandwidth, allowing processors to maintain reasonable system balance as core counts increase. HP Labs is encouraging proposals for new processor chip architectures, on-chip networks, and memory hierarchies to take best possible advantage of nanophotonic capabilities.

The digital data explosion continues at a furious pace, increasing at a rate of more than an order of magnitude each year. A study in 2004 predicted worldwide data volume to increase from 3.2 million exabytes (3.2 yottabytes) to 45 yottabytes in the year 2005. In one year, more data was generated than in the entire history of humankind. This trend continues, and today it would be hard to find an organization in any industry that does not collect huge quantities of data for some or all of its business purposes.

Analyzing massive amounts of data for competitive advantage will require similar growth in processing, storage, and network/interconnect capabilities. As technology progresses, both system scale and performance improve, but there are numerous show-stopping problems that must be solved in order to build systems that are capable of meeting the rampant growth in demand. The most critical challenges are: 1) improving system performance without incurring a commensurate increase in power consumption; and 2) providing increased data bandwidth to avoid starvation of the higher performance processing resources.

Copper-based on- and off-chip interconnects are the dominant culprit for both the power and bandwidth problems, since power is dependent on wire length and a wire can only carry one high-speed signal at a time. Increasing cross-sectional bandwidth implies smaller wires or higher clock rates – both of which increase power consumption. Nanophotonics is an attractive solution since power consumption is independent of path length and the use of dense wave division multiplexing improves bandwidth.

The objective in pursuing this research from a business perspective is to enable continued growth in HP's high-end datacenter business which is already being constrained by a variety of power and interconnect bandwidth problems at all system levels. Both HP and academic researchers have already developed promising CMOS nanophotonic technology. The challenge is now to develop system and device architectures that take advantage of this promising technology and to accurately quantify the benefits of these architectures.

Research Questions

- What fundamental architectural opportunities are enabled by silicon nanophotonic interconnect for on-socket, off-socket, in-rack, and inter-rack system levels?
- How can photonic interconnect be utilized to solve the datacenter cabling complexity problem?
- How can photonic interconnect be utilized to solve the main memory capacity and bandwidth problem?
- How can improved photonic communication reduce the burden on parallel program developers?

Sustainable IT Ecosystem Lab

Topic 14: Sustainable Solutions for Data Centers

The increased demand for Information Technology (IT) Services has resulted in the construction of data centers with racks of densely packed servers containing increasingly powerful microprocessors that are proliferating worldwide. HP is soliciting proposals for research projects that seek to reduce this impact by focusing on improvements in the design and management of data centers and their key resources; we are particularly interested in solutions that improve IT, cooling and power delivery resources while maintaining data center reliability and performance requirements.

Data Center Facility Grid

Data center facility infrastructure consists of power generation/conversion/delivery and cooling (chillers/CRAC units) devices. Efficient and effective generation and distribution of power and other resources among demand centers in the datacenter is crucial to data center operation. Our research is focused on designing a micro-grid for power and cooling distribution in the data center facility that meets efficiency, manageability and regulatory requirements while meeting Service Level Agreements.

The HP Labs Sustainable IT Ecosystem Lab is a multi-disciplinary team spanning computer science to mechanical engineering. Research will cover the end-to-end management of the data center and the IT services hosted by the data center, from design, synthesis and operations to end-of-life.

Data Center Design and Management

The increased demand for Information Technology (IT) Services has resulted in the construction of data centers with racks of densely packed servers containing increasingly powerful microprocessors. These data centers are proliferating worldwide as service providers struggle to keep up with demand. As a result, the worldwide carbon footprint of Information Technology is expected to increase rapidly in the coming decade, with most of this growth coming from data centers. HP is soliciting proposals for research projects that seek to reduce this impact by focusing on improvements in the design and

management of data centers and their key resources; we are particularly looking for solutions that improve IT, cooling and power delivery resources while maintaining data center reliability and performance requirements.

Research Questions

Data Center Design and Synthesis

- How can services/applications or classes of services be combined with a specified Service Level Objective (SLO) to generate an optimal computing resource solution in a generic and systematic way?
- How can requirements for performance, reliability and security be combined with installation, operation and exergy costs for computing, power and cooling in a framework that models these attributes and their interactions?
- Are there efficient, repeatable and scalable algorithms that can analyze trade-offs among the attributes in the above model to synthesize a data center design that optimizes TCO and exergy criteria?

Data Center Management

Computational Fluid Dynamics (CFD) has been accepted as the primary technique to analyze the thermal environment within data centers. While the technique is effective, it is slow and unsuitable for real-time analysis.

- How can physics-based approaches (like CFD) be improved to achieve a useful solution more rapidly?
- Can non-physics-based approaches be utilized to achieve accurate real-time results?
- Can data from real-time thermal measurements be used to improve time-to-convergence in either case?

A data center produces vast quantities of data from continuous sensor and operational measurements. These data contain vital information about the health of the data center and might be mined for opportunities to improve efficiency.

- Can trends and patterns be detected and summarized without user involvement?
- Can heterogeneous data center attributes (such as temperature and system utilization) be analyzed together to infer cross-layer trends and patterns?
- Can streaming data be analyzed with archival data to infer changes in operating conditions?
- Can efficient machine-learning models be built and deployed for effective automatic identification of root causes of anomalous behavior and predict failure conditions? And how to combine machine learning with feedback control to provide integrated diagnosis and control capabilities?
- How can one develop a distributed framework for end-to-end data center operations management where a set of local controllers, each managing a subset of the metrics using a subset of the controls, rely on only local information while collectively optimizing some global objective function for the data center, such as minimizing TCO (or exergy destruction) while satisfying application SLOs?

IT Services Management

- What performance modeling methods can be used or developed to manage multiple computing resources including CPU, memory, disk, network and storage in a data center in order to optimize TCO?
- How can traditional performance models be adapted for virtualized environments and server consolidation environments to account for virtualization overhead and interactions among co-hosted applications?

Data Center Facility Grid

- What should be the next generation local power generation /conversion /storage /delivery grid for data centers? Are the current micro-grid technologies and integration methods capable of achieving the reliability required by IT infrastructure?
- What should be the next generation cooling resource distribution grid for datacenters? Which technologies are better suited for improved datacenter reliability while reducing impact on environment.
- What other resources does a datacenter use? Can we design better distribution topologies for these resources that reduce impact on environment?
- What are the key metrics that would define (or should assess) these approaches to facility grid design?

For additional information, visit: http://www.hpl.hp.com/environment/

Sustainable IT Ecosystem Lab

Topic 15: Creating a Sustainable IT Ecosystem

HP as a company aims to play a leading role in moving towards a sustainable society, addressing the issues of climate change, resource consumption, ecosystem well-being, and quality of life. As the world's largest IT company, we seek proposals that can help us make a critical contribution to reversing the current trend by reducing the impact of the global IT ecosystem, and employing IT systems and services to measure, monitor and significantly reduce the carbon footprint of our worldwide enterprise and consumer customers.

The April 2005 Millennium Ecosystem Assessment conducted by the United Nations included 1,300 researchers from 95 countries, and concluded that 60% of the ecosystems that support life on Earth are being degraded or used unsustainably. The report also warns this is likely to significantly worsen over the next 50 years if the situation remains unchanged. As the world's largest IT company, HP can make a critical contribution to reversing the current trend by reducing the impact of the worldwide IT ecosystem, and more importantly by employing IT systems and services to significantly reduce the carbon footprint of our worldwide enterprise and consumer customers.

HP as a company aims to play a leading role in moving towards a sustainable society, addressing the issues of climate change, resource consumption, ecosystem well-being, and quality of life. To that end, HP Labs is engaged in research in three broad areas, and is looking for support to tackle specific issues related to:

- Understanding, publicly reporting and reducing the direct environmental impact of our products and services. The key research challenge here is the provision of tools and metrics which are accurate, easy to use, and incorporate a variety of sustainability factors.
- Developing methods for monitoring and reducing the overall sustainability impact of networked systems.
- ICT for sustainable business transformation. How can ICT be used to enable other business sectors to become more sustainable, particularly with respect to carbon emissions?

Research Questions

Tools and Metrics for Sustainable Solutions

- Existing approaches to life-cycle assessment (LCA) require detailed inventory and reliance on third-party databases, often with limited means of validating the vast input data. How can LCA packages be made easier to use? How can the design of products using LCA be simplified?
- Most existing LCA tools neglect issues around broader socio-economic impacts, such as the longer-term systemic impacts related to displacement of alternative methods of meeting needs or the resulting shifts in behavior. How can these be included?
- Exergy has previously been proposed in the literature as an impact metric for LCA and (separately) for broader quantification of resource consumption, waste recovery, etc. What type of advantages might be available from an extended exergy life-cycle assessment (EELCA)? Can such an approach be used to link thermodynamic process data to validate process inputs, thus allowing a more robust LCA?

Sustainable Networked Systems

- How can protocols between networked devices allow optimization of power consumption over a local area network, both in domestic and in business environments? What savings are possible? How can these be monitored and measured?

ICT for Sustainable Business Transformation

- There is much research on using ICT as an enabler of business process optimization. How can such techniques be applied to innovating, evaluating and validating the sustainability of novel business models? Appropriate areas of investigation include (but are not limited to) digital print, smart energy management, supply chains and telepresence.

For additional information, please visit: http://www.hpl.hp.com/environment/

Storage and Information Management Platforms Lab

Topic 16: Scalable Storage

The Scalable Storage program at HP Labs is researching technologies to develop and deploy systems that reliably store and retrieve very large amounts of data, support multiple interfaces (including different quality of service, availability, and consistency requirements), provide security and performance isolation between multiple tenants, and provide extensive self-management techniques at a scale that spans geographically-separated datacenters.

We are soliciting proposals for joint, collaborative research projects that will advance the state of the art in technologies for scalable storage systems; explore the integration of extant techniques and tools; and/or evaluate solutions by deploying, instrumenting, and stressing scalable storage systems under a range of environmental conditions.

Modern enterprises and applications are generating and finding productive ways to use petabytes of data through a variety of new access, control, and interface paradigms. The data needs to be stored reliably, made accessible as and when it is needed, and the system that does so needs to support multiple interfaces, provide security and performance isolation between multiple tenants, and provide extensive self-management techniques – all at a scale that spans multiple geographically-separated data centers. The Scalable Storage program at HP Labs is researching technologies to develop and deploy such systems.

The HP Labs Storage and Information Management Platforms Lab has a multi-decade track record of collaborating with academia, and a publication record to match. We are soliciting proposals for joint, collaborative research projects that will advance the state of the art in technologies for scalable

storage systems, or explore the integration of extant techniques and tools, or evaluate solutions by deploying, instrumenting, and stressing scalable storage systems at scale, and under a range of environmental conditions.

Research Questions

- Flexible interfaces, supporting multiple types of client interactions with a storage system; how to:
  - support different client interfaces safely, isolating them from one another; exploit virtual-machine technologies and new security schemes such as proof-carrying code;
  - allow extensible functionality ("function shipping") in both the storage system and its clients ("reverse function shipping"); and
  - determine where to execute data-intensive operations, and how to sequence, buffer, and schedule a set of such operations.

- Large-scale operations; how to assure reliable operation at:
  - petabyte scale across multiple data centers;
  - multiple different assurance, consistency, and reliability levels simultaneously; and
  - across multiple interface types and multi-gigabyte metadata structures such as directories.

- Fault handling and containment; how to:
  - automate failure diagnosis/pinpointing, mitigation, and fixes;
  - prevent data loss for datasets that span multiple data-centers;
  - provide proofs of correctness for normal-mode and failure-containment protocols;
  - cope with failures that range from element, to subsystem, rack, to data center; from hardware fail-silent to (possibly) malicious software errors, to operator mistakes.

- Storage system management; how to:
  - express desired quality of service goals; map those into SLAs and SLOs; and turn these into low-level control signals for the storage system;
  - automate provisioning, failure- and performance monitoring, data ingesting; data migration; and *all* mundane operator activities;
  - provide auditing and assurance guarantees for security, integrity, reliability, and performance, at a range of timescales (ms to decades), and in the face of both benign failures and malicious attacks.

- Information service platform; how to:
  - provide support for self-managing information-service providers;
  - support a wide range of information-centric service types concurrently on the same information sources;
  - integrate with a system like Amazon's Elastic Compute Cloud.

For more information, visit: http://www.hpl.hp.com/research/ssp/

Topic 17: Natural Language Processing and Computational Linguistics Tools for Taming the Information Explosion

(Storage and Information Management Platforms Lab and HP Labs Russia)

HP Labs is seeking research partners to work together on developing cutting-edge technologies for semantic analysis of text which will enable the next level of processing unstructured information in different languages for different industries. The outcomes of the research will be technologies for base linguistics and advanced NLP with focus on fact and event extraction, sentiment analysis, synonymous paraphrasing and word-sense disambiguation.

Unstructured information with a large amount of text is pervasive in enterprise environments due to the prevalence of emails, Wikis, websites, other portals, and, most importantly, office file formats. Viewing this text as a string of bytes (using techniques such as chunking, hashing, n-gramming, and other meaning-agnostic approaches) has outlived its usefulness. Instead, what is needed is technologies for semantic analysis of text objects, *i.e.*:

1. Base linguistics (language identification, named entity extraction, tokenization and sentence-boundary detection, word-sense disambiguation, Porter stemming, stop-word removal, …)
2. Simple natural language processing (Parts-of-speech tagging, summarization, categorization/tagging)
3. Minimal advanced natural language processing (paragraph-level pronoun or other reference resolution; fact and event extraction; sentiment analysis)
4. Specific industry verticals (Financials, Healthcare, Retail, Tourism) also need dedicated textual- and tabular data extraction from documents, e.g. tools for analyzing S.E.C. filings, Customs appeals, labor contracts, claim forms, brochures.
5. Translation tools and support of languages used by major global economies.

These are the shared needs of multiple research groups within HP Labs. Piecemeal offerings available from commercial or open source vendors are difficult to integrate. UIMA or OpenNLP compliance is therefore desirable.

While base linguistics are rapidly maturing and starting to cover almost all languages globally, NLP stacks upstream are notoriously fragmented from language to language and even capability to capability. We are calling for research partners who focus on developing, integrating and supporting state-of-the-art (primarily) open source and (as needed) commercial offerings to offer a robust platform covering 1-3 above.

In addition, we expect the respondents to partner with our research groups in exploring cutting-edge issues in computational linguistics and natural language processing, such as:

- Synonymous paraphrasing
- Expert-to-novice vocabulary mapping
- Pattern-generation and data-filling techniques for scientific, financial and medical data
- Legal contract understanding
- Yield calculators and other advanced processors for descriptions of fixed income instruments
- Slang, and the vocabulary of email, SMS, IM and 'blogs

Intelligent Information Management Lab

Topic 18: Real-time Business Intelligence Infrastructure

Business Intelligence (BI) is at a turning point today. Enterprises traditionally have used BI for offline, strategic decision making, where a small number of expert users analyze historical data and decision making cycles last weeks or months. As enterprises become more automated, real-time, and data-driven, the industry is evolving toward adaptive, operational BI systems that support online, operational decision making at all levels in the enterprise. High-quality information must be delivered in real-time to analytic applications that are integrated into the enterprise's business processes.

For example, an on-line retailer would like to analyze a user's real-time click stream data and up-to-the-minute inventory to offer dynamically-priced product bundles. A bank would like to detect and react in real-time to fraudulent transactions. A logistics provider would like to dynamically reconfigure shipping routes in response to weather conditions.

Research in intelligent information management builds on today's business intelligence and related information-management technology foundation, and dramatically expands its horizon to achieve near real-time data capture and integration; adaptive and closed-loop operational business intelligence supporting a wide user base; data-intensive parallel analytics; and fusion of business intelligence with search and information extraction over structured and unstructured information from a wide variety of internal and external sources and data feeds.

HP Labs' research draws requirements from emerging applications ranging from consumer industries delivering better customized products and services, supply chains synchronizing with real-time demand signals and scientists improving their ability to access and analyze massive amounts of data. HP Labs' goals in this research area are focused on the data pipeline, data streaming, and complex event processing; robust BI workload management and query processing; and SOA-based information services.

Research Questions

- Data pipeline, data streaming, complex event processing - real-time continuous feed of business events into business intelligence systems and stream data processing to reduce the latency between the time an event occurs and the time quality data is available for decision-making
- Robust BI workload management and query processing - techniques to enable self-managing, adaptive tuning for complex, mixed workloads, and robust query processing, storage and access methods that adapt gracefully to changing workloads
- Effective use of distributed infrastructure for information integration; fusion of search, extraction, query and analytics to enable integration of structured and unstructured information in BI solution contexts; and emerging, innovative vertical applications enabled by scalable information services

For additional information, visit: http://www.hpl.hp.com/research/information_management.html

Intelligent Information Management Lab

Topic 19: Visual and High-performance Analytics

Information is a critical enterprise asset exhibiting exponential growth; business information can be analyzed for decisions and new insights, and turned into knowledge delivered into user contexts. It powers modern businesses and scientific research and is the lifeblood of enterprise business processes.

Research in intelligent information management builds on today's business intelligence and related information-management technology foundation, and dramatically expands its horizons to achieve near real-time data capture and integration; adaptive and closed-loop operational business intelligence supporting a wide user base; data-intensive parallel analytics; and fusion of business intelligence with search and information extraction over structured and unstructured information from a wide variety of internal and external sources and data feeds. High-performance parallel analytics is critical for achieving near real-time delivery of transactional intelligence. Visualization techniques have proven to be powerful in exploring patterns, distribution, trends, and exceptions in large multi-dimensional data sets.

HP Labs' goals in this research area are focused on algorithms for scalable data-intensive analysis; use of new visual analytics techniques for concurrently managing and monitoring high-speed and high-dimensional large information space; information visualization for real-time data exploration and analytics; use of parallel architecture on multi-core clusters; and emerging memory hierarchy and storage devices in large-scale data management.

Research Questions

- Algorithms for scalable data-intensive analysis, including real-time data summarization, mining, and learning required for information extraction.
- Use of new visual analytics techniques for managing and monitoring high-speed and high-dimensional information; topics include:
  - New visual mechanisms such as multi-resolution techniques, importance-driven visualization, animation and replay mechanisms
  - Automating the selection of parameters for visualization, avoiding clutter and focusing the user's attention on the most interesting patterns through smart analytics
  - Tightly integrating automated analytics with visual methods
  - New techniques for integrating text analytics with visual methods
  - Anomaly discovery
- Use of parallel architecture on multi-core clusters and emerging memory hierarchy and storage devices in large-scale data management

For additional information, visit: http://www.hpl.hp.com/research/information_management.html

Business Optimization Lab

Topic 20: Business Optimization: Planning under Uncertainty, Resource Allocation, Procurement, Auction, and Risk Management

HP Labs is developing and exploring the next generation of analytical technologies and solutions that will bring unprecedented operational efficiencies to the enterprise. Research areas of particular interest include the management of uncertainty related to resource allocation when both supply and demand are stochastic; pricing/revenue management; and channel contract design. Possible applications are in management and optimization of large-scale service delivery. HP, as the operator of the IT industry's largest supply chain, is also interested in advanced research in mechanism design to provide analytic tools and approaches to enhance strategic, operational, and tactical decision making processes in strategic sourcing – ranging from risk-management decisions, to auction technologies related to supply contracts, to data collection and analysis for sourcing systems.

Management of uncertainty is an important research field in Operations Research and practical analytical results can add significant business value. Particularly, the areas of resource allocation under uncertainty, pricing/revenue management, and channel contract design are research areas of significant interest to HP, with applications in and focus on the management and optimization of large-scale service delivery.

A related area of interest to HP is that of supply chains. Uncertainties, information asymmetries, conflicting goals are abound in the design and governance of supply chains. HP, as the operator of the IT industry's largest supply chain, is interested in advanced research in mechanism design to provide analytic tools and approaches to enhance strategic, operational, and tactical decision making processes in strategic sourcing – ranging from risk-management decisions, to auction technologies related to supply contracts, to data collection and analysis for sourcing systems.

Research Questions

- Large-scale coordination of supply and demand of people resources when both supply and demand are stochastic
  - Given sparse historical data and hidden structural changes in the data (such as project funnel data and contact center event streams), how do we forecast labor supply and demand for a services business?
  - How do we effectively model the effects of pricing policies on labor demand in large service deals (both outsourcing services and consulting services)?
  - How do we build scalable, federated optimization models for matching supply and demand, when both are uncertain?
  - How does this enable us to do medium/long-term workforce planning, as well as short-term scheduling and assignment of resources with complex service-level requirements?

- Revenue management and contract design
  - How do we model the effects of incentive contracts in value-add services on the relationships among manufacturers and distributors?
  - How do we optimally price these value-added services in a competitive environment?
  - How do we predict and optimize these effects, in both data-poor and data-rich environments?

- Modeling optimization-enabled bidding processes with expressive bidding

- What are the trade-offs in the evaluation of design alternatives for expressive bidding? How expressive should the bidding language be to optimally control the bidding incentives?
- What are effective mechanism design and cooperative game theory formulations for integer-programming duality?
- How do we evaluate (computationally and experimentally) alternative dynamic combinatorial auction processes?

- Game theoretic and microeconomic modeling of dynamics of supply chains:
  - What are the implications of supply chain design for the strategic interactions of the players?
  - How do alternative auction mechanisms for allocation of contracts with multiple parameters perform in terms of risks/rewards born by the buyer and the suppliers?

Business Optimization Lab

Topic 21: Personalization, Recommendation and Customer Behavior Analysis

HP Labs is seeking proposals focused on personalization, recommendation and customer behavior analysis which will allow for extraction of accurate models of people's interest and intent, taking into account dynamic preferences and coping with incomplete and noisy data situations. Important research outcomes include methods allowing one to objectively and cost-effectively measure improvements in user experience.

Even though the digital age has transformed the daily lives of people in ever-greater numbers, people use the PC and the web to do many of the things they have always done: get entertained, search for information, form social communities, and buy and sell products and services. However, we have seen a vast transformation in the availability and use of customer data of all kinds. This has sparked rich research in analytics related to personalization, customer behavior analysis and optimization.

Research Questions

- Large-scale recommendation analytics, optimization and customer behavior analysis
  - How do we extract accurate and predictive models of people's interests and intent?
  - How do we develop a modeling framework that encompasses the sheer diversity of customer actions across multiple market mechanisms and channels of communication?
  - How do we detect and take into account dynamic preferences and interests?
  - How do we develop techniques that work in sparse, incomplete and noisy data situations? That scale to millions of customers?
  - How do we guarantee appropriate levels of identity protection and privacy while extracting valuable information?
  - How do we develop models for policy analysis that fuse data about individual customers with data about aggregate market behavior?

- Evaluation of recommendation and personalization systems
  - How do we analyze and evaluate whether personalization and recommendation algorithms are effective?
  - What makes personalization valuable, effective?
  - How do we objectively and cost-effectively measure improvements in user experience?

Service Automation and Integration Lab

Topic 22: Creating a Business Operating Environment in the Global Services Ecosystem

In this applied research project, HP Labs is investigating what customer service lifecycles and experiences are possible in an "Everything as a Service" model and prototyping underlying intellectual property to enable them. HP Labs' goal in this research area is to address the technical challenges that must be overcome to move a business task to services over the Internet.

Shane Robison, HP's Chief Strategy and Technology Officer, has detailed a set of "Everything as a Service" predictions that he believes will shape the IT industry in years to come. One of his predictions is that "by 2012, a Fortune 50 company will research, develop, and launch a major product using only Internet-based services." This opinion is supported by information available from industry analysts, such as Gartner and IDC. In this project, we ask: "What would a corporation wishing to move to an 'Everything as a Service' model need to do?" This question poses a number of challenges, such as:

- How does the corporation find service providers that it can trust and rely upon to deliver the required business capabilities?
- How does the corporation implement and integrate business, data, and interaction flows across multiple service providers?
- What business models can the corporation implement, and are there new ways of structuring and running the corporation?

In this applied research project, we investigate what customer service lifecycles and experiences are possible in an "Everything as a Service" model and prototype underlying intellectual property to enable them. We address the technical challenges that must be overcome to move a business task to services over the Internet. We take a "business-centered" approach to research and prototyping (in contrast to an IT-centered one), where we start with some business task and address the challenges faced when the business task is moved to services available over the Internet. We propose a methodology based around a five-component model (Business Process Specification, Service Selection, Business Process Instantiation, Business Process Choreography, and Business Process Adaptation) to develop an end-to-end prototype demonstrating the entire task lifecycle.

Research Questions

- How to make requesting cloud services comprehensible to the business user?
- How to enable the business to select appropriate services over the cloud?
- How to instantiate a business process using cloud-based services?
- How to continuously demonstrate that cloud services are meeting the customer's business needs?
- How to adapt the process to changes in business needs or in the services environment?

Information and Quantum Systems Lab

Topic 23: Dynamical nanoelectronic devices and circuits

For the past 40 years, the driving force in the electronics industry has been Moore's Law. However, we are rapidly reaching the plateau of the S-curve for silicon integrated circuits, and for some metrics Moore's Law has already ended. There is an intense world-wide search for a new device that will complement the transistor to create a new technology S-curve and enable the capability of computing machinery to increase rapidly for several more decades.

Nanoscale electronic devices have recently been shown at HP Labs to posses intrinsic dynamical properties, more precisely 'memristor' behavior. A memristor is the fourth fundamental passive circuit element that completes the family of two-terminal devices: resistor, capacitor and inductor. Predicted mathematically more than 30 years ago, this device was first realized at HP Labs in 2006. The memristor differs from the standard passive resistor in that it is dynamical; in other words, its present state and functionality depend upon its history. That gives the memristor very interesting and useful non-linear current-voltage characteristics that have the potential for revolutionizing storage, memory and logic systems by attaining density and computing performance far superior to any scaled CMOS-only circuit.

Depending on how it is driven electronically, the memristor can act as either a digital or an analog switch. As a digital switch, memristor circuits may fill the growing latency gap between present DRAM memories and FLASH & magnetic disk storage systems. As an analog switch, such circuits may offer for the first time the opportunity to build intelligent adaptive logic circuits that compute using memristive electronic synapses.

Research Questions

- What nanoscale physical and chemical changes and processes are occuring in dynamical memristor nano-devices? Can the physical changes that accompany electrical switching be localized in 3D? Using innovative nanoscale testing methods, can local structural or chemical information be correlated to the electrical dynamics?
- How can circuits of dynamical nano-devices best be designed to optimize applications of (a) crosspoint memory, and (b) neuromorphic logic? What circuit and logic architectures best match the dynamical devices? Can new circuit design tools be constructed to automate large scale circuit design?
- How can neuromorphic circuits of dynamical nanoelectronic devices best be tested for intelligent adaptive computing performance? What virtual simulation environments are required to both teach and test such intelligent adaptive circuits?

References:

Strukov, Snider, Stewart & Williams, "The missing memristor found", *Nature* (1 May 2008).

Snider, "Self-organized computation using unreliable, memristive nanodevices," *Nanotechnology* 18, 365202 (2007).

Yang, Pickett, Li, Ohlberg, Stewart & Williams, "Memristive switching in metal-oxide nano-devices", *Nature Nanotech.* (2008).

Information and Quantum Systems Lab

Topic 24: Semiconductor Nanowire Photonics/Electronics

Semiconductor nanowire grown on non-latticed matched substrates has been demonstrated. It offers the advantages of heterogeneous-material integration; for example III-V nanowires on silicon substrates will allow integration of III-V photonic devices on silicon substrate with CMOS circuitry. In addition, semiconductor nanowires have also been grown on non-single crystal surfaces and this further opens the hetero-material integration to include growth on non-semiconductor surfaces. The challenges will be to further develop understanding on the junctions formed by the semiconductor nanowires to single and non-single crystal surfaces; and to study dopant to form diode and ohmic characteristics.

Furthermore, for emitters, thermal issues need to be addressed and understood as the nanowires may heat up like a filament when biased at high current densities.

The final objective will be to develop nanowire electronic and photonic devices on non-single crystal surfaces that may be integrated with CMOS processes for sensors and photonic integrated circuit applications.

Research Questions

- Characterisitics of junctions between the nanowire and single and non-single crystal semiconductor
- What limits the minority lifetime of carriers in the nanowire, surface states, traps, etc
- How can heat be extracted or minimized in nanowires
- Optical absorption/emission characteristics of the nanowires as a function of temperature.

Information and Quantum Systems Lab

Topic 25: Signal Processing in On-Chip Optical DWDM Networks

Moore's Law is still a fundamental technology driver for the information technology industry: the ITRS Semiconductor roadmap shows, in the next decade and a half, a continued reduction in feature sizes from 40 nm to the sub-10-nm regime. This growth in circuit density has brought us into the multi-core CPU era, and we are on the eve of a many-core (16 or more cores per socket) era.

In the many-core era, on-chip and off-chip communication are the critical issues for sustaining performance growth for the demanding, data-intensive applications for which these many-core chips are intended. If we strive to continue delivering exponential performance improvements over a broad range of computational applications during the next decade, then we are led inevitably to an architecture with an all-to-all single-hop crossbar interconnect that allows the programmer to exploit parallelism at a high level. But this simplicity requires that we rely on an interconnect technology that is no longer limited by the physics of copper wire.

We at HP have argued that nanophotonics provides a feasible solution to this growing problem, and that on-chip optical dense wavelength-division multiplexed networks are inevitable. The workhorse of most designs of DWDM nanophotonic interconnects is the silicon microring resonator because it has small size, high quality factor Q, transparency to off-resonance light, and no intrinsic reflections.

However, there have been no comprehensive analyses of ultra-short-haul optical networks based on these components, and it is not at all clear that the usual telecom rules of thumb even apply at millimeter-to-centimeter distance scales.

Research Questions

- How should data be routed through an on-chip interconnect? Is there a role for both time-division multiplexing and WDM?
- What is the trade-off between modulation extinction ratio and pulsewidth for cm-scale networks?
- What is the trade-off between pulsewidth and chirp in such a network?
- How do imperfections in fabrication affect the on-chip signal Q?
- How does the bit-error-rate (BER) depend on the signal Q in these ultra-short-haul networks?

Information and Quantum Systems Lab

Topic 26: High-Speed Low-Cost CMOS-Compatible Optical Engine

The most economical way to create high-speed lasers for datacom may be a hybrid approach entailing wafer bonding of III-V materials to a CMOS Silicon-on-Oxide wafer. We would like to build such a few-micron-scale on-chip laser that can be modulated at speeds exceeding 10 GHz.

Optical interconnects have revolutionized communications at scales ranging from intercontinental links down to the data center. There is a strong interest in the datacom industry to push the scale of optical networks further down to the level of server boards and chip packages. This could significantly change the computer industry by fostering the era of massively parallel many-core computers with enormous memory bandwidth. This revolution will need optical components that are radically different from the ones used in the telecom industry in order to satisfy requirements for reduced cost, size, and power consumption that are the driving forces in the computer industry.

A key component for on-board and on-chip networks is an optical engine that can be integrated in a standard CMOS process. The main challenge in building such device is that while Silicon is in many ways an excellent optical material (e.g. to build low-loss waveguides on a chip) it is not an efficient light emitter. For this reason a hybrid approach entailing wafer bonding of III-V materials to a CMOS Silicon-on-Oxide wafer might prove to be the most economical way to create on-chip lasers. While this approach has already produced lab demonstrations of cw lasers [see reference, below] we believe there is a possibility to build a micron-size on-chip laser that can be modulated at speeds exceeding 10 GHz. An array of such lasers combined with other technologies that have been developed at HP labs could provide a memory link with a capacity exceeding 100 Gigabits per second while taking up less than a square millimeter of chip area and requiring no RF pins.

Research Questions

- What fabrication techniques are most suitable for a 25-100 μm^2 CMOS-compatible laser?
- How will a hybrid laser be integrated into an on-chip optical network?
- What parameters in the laser blueprint and fabrication can be optimized to obtain maximum modulation speed and wall-plug efficiency?
- What thermal management techniques will be applied to insure low-temperature laser operation to increase lifetime and reliability?

Reference: Open access publication by A.W. Fang *et al.*:
http://www.opticsexpress.org/abstract.cfm?&id=114596

Information and Quantum Systems Lab

Topic 27: Identical Single Photons

Quantum information science is an emerging discipline whose theoretical promise—for revolutionary advances in secure communications, precision measurements, and computational power—has far outstripped its experimental achievements to date. One key component (perhaps the key component) for quantum applications ranging from secure communication to more efficient computation is a device which can emit identical single photons. In schemes relying on this resource, single photons could themselves be used as quantum bits (qubits), or they could be used to convey information between spatially distributed matter qubits.

One promising approach for single-photon state preparation uses the technique of spontaneous parametric down-conversion (SPDC) in nonlinear crystals. In this case, "pump" photons are occasionally split into a pair of "signal" and "idler" photons conserving momentum and energy. If the conversion rate is low, as it must be to avoid production of multiple pairs, then detection of an idler photon heralds the creation of a signal photon with high fidelity. However, to date, the spatio-temporal modes of successive heralded single photons created using SPDC have not been sufficiently indistinguishable to allow high-fidelity quantum information processing. An experimental demonstration of spectrally-engineered SPDC producing high-quality single photons would represent a significant advance in the field.

Research Questions

We welcome proposals for experiments which will *begin* to help us answer the following long-term questions:

- How can phase matching be engineered to joint spectral amplitude of an entangled photon pair to be factored into distinct contributions from the pump beam and the phase-matching process? Is this a requirement for indistinguishability?
- How does photon temporal mode indistinguishability depend on the temporal characteristics of the pump beam?
- Is optical-fiber coupling required to produce a single spatial mode in successive photons?

Information and Quantum Systems Lab

Topic 28: Engineering Quantum Technologies

A broad class of challenges is coming into focus as we continue to develop nanotechnology for high-performance computation and communication, and as we contemplate the transition, in key industrial sectors, from classical to quantum technologies. This shift may well be gradual and might not begin in earnest for ten or twenty years, but we believe that it will be inevitable. The underlying quantum nature of the physical carriers of information in computers and communication networks—electrons and photons—cannot remain hidden forever as our relentless demand for speed drives engineers to deploy ever-smaller transistors, and ever-fainter pulses of light in fiber-optic networks.

From a classical-engineering perspective, the innate fluctuations in electronic and optical signals arising from fundamentally quantum-mechanical origins are an obstacle to further progress, but from the perspective of quantum engineering, the advent of these complexities creates radically new for functionalizing matter and energy with atomic resolution. New engineering concepts and

methodology are possible, and indeed are required at atomic scale because the physical dynamics of isolated microscopic systems are so qualitatively different from those upon which our macroscopic inventions have relied.

Forerunners of true quantum engineering can be seen in the now-ubiquitous laser and in the atomic clock, as well as more recent progress in cryptography, remote sensing and optical switching. We have also seen the first hints regarding the long-term promise of quantum engineering in the celebrated theoretical results of quantum computation, which show that fundamentally different and superior algorithms can be posited for tasks such as factoring. However, we are a long way from understanding how to engineer quantum mechanics into usable and marketable technologies, and from performing early experiments which will help us determine the best path forward.

Research Questions

We welcome proposals for experiments which will *begin* to help us answer the following long-term questions:

- How can quantum behavior be exploited for technology with performance beyond the classical limit?
- How can quantum coherence be realized (and quantum decoherence suppressed) by structural design or active control?
- Are there quantum information or quantum metrology applications that can take advantage of only a few quantum bits or states?

Information and Quantum Systems Lab

Topic 29: Integrated Si and Ge Nanowire Sensors, Electronics, and Photonics

Nanowires of Si and Ge are being proposed for many sensing, electronic and photonic devices. But, the challenge is to integrate the nanowires in an efficient and cost-effective manner to form useful and technically robust systems. The objective of the proposed work is to develop integrated systems containing nanowire sensors and related devices integrated with conventional electronics and possibly MEMS structures.

Narrow, metal-catalyzed semiconductor nanowires can be formed without fine lithography, making them competitive with increasingly expensive, conventional lithographically formed structures. They can be grown on plane substrates or on the sides of mechanically and electrically isolated posts, allowing mechanical and electrical connection, again without advanced conventional processing techniques. Field-effect and resonant sensors have been fabricated using suspended nanowires, and other nanowire sensors are possible. Because of the limited contact area between a nanowire and the substrate on which it is grown, lattice mismatch can be readily accommodated, allowing high-quality nanowires to be formed on lattice mismatched materials, and nanowires and combinations of nanowire arrays of different materials can be formed.

With these capabilities, the challenge is to integrate the nanowires in an efficient and cost-effective manner to form useful and technically robust systems. For example, preliminary calculations [see reference, below] show that a sensor array can be fabricated in a small area, with local CMOS electronics adjacent to each sensor. Forming the integrated structure using compatible processing will be a challenge. In other cases, integrating an individual sensor with adjacent actuation and detection circuitry will require novel and innovative techniques. Extension to photonic and photovoltaic devices to interact with the environment and to power a small system is promising, but challenging.

The objective of the proposed work is to develop integrated systems containing nanowire sensors and related devices integrated with conventional electronics and possibly MEMS structures.

Research Questions

- How can efficient, robust, and cost-effective, single-function sensors (including associated actuation, detection, and electronics) be fabricated using metal-catalyzed nanowires and possibly integrated with conventional electronics?
- How can nanowires be integrated with other components of a sensor system in a multi-variable sensor system (including peripheral functions, such as energy harvesting, energy storage, and data communications)?
- What are the most valuable sensing functions that can be performed using metal-catalyzed nanowires?
- How can the physical, electrical, and photonic properties of nanowires be controlled to form repeatable sensors?

Reference: T. I. Kamins, et al, "Metal-catalyzed, bridging nanowires as vapour sensors and concept for their use in a sensor system,' *Nanotechnology* **17**, S291 (2006).

Information and Quantum Systems Lab

Topic 30: Distributed Quantum Information Processing

Fundamental effects in quantum physics can enable new information gathering, processing and communication technologies. These have the ability to outperform their conventional counterparts. The irreversible disturbance caused by measurement of quantum systems enables physically secure communications, guaranteed by Nature. The possibility of quantum systems being in many states at once enables massively parallel computing. The correlations between entangled systems enable transmission of quantum information, and thus distributed quantum processing. Such correlations also enable detection and measurement with sensitivities beyond conventional bounds. Research is on-going in many groups worldwide to turn these ideas into reality; this is an exciting, expanding and competitive field, with recognized disruptive technology potential.

The grand, long-term vision of Quantum Information Processing (QIP) Research at HP Labs is a whole new class of quantum information-based technologies. Imagine a world where quantum-enhanced sensors, consumer devices, businesses, service providers - in the end, everything - is connected on a hybrid network, communicating quantum information and quantum-secured conventional data. This quantum network is enabled with entangled photons and photonic buses, extending over distance using quantum repeaters. Powerful processors, built from quantum toolbox components in an integrated and defect tolerant manner, run novel quantum algorithms. This complex, distributed quantum network delivers new analysis, results and services, not possible with conventional technologies alone.

There are some crucial "next questions" to address to progress towards this long term goal: scaling quantum information processors beyond the present small-scale devices is very challenging, as communication between parts of the processor, single site addressability and scaling are difficult to reconcile. So how can this be achieved and what are the critical steps along the way? What can small scale distributed quantum processors be used for?

Research Questions

- What are the fundamental building blocks needed for a distributed, hybrid, information processing system (using a hybrid "analogue/digital" architecture of quantum buses and bits) and how should these be achieved? What forms of buses can be utilized?
- How does one design an efficient high bandwidth quantum repeater network capable of transmitting quantum information over large distances? What are the core components and what routes exist for their implementation? What is likely to be the optimal number of qubits per node and how does the throughput of the device scale with the resource per node?

For additional information, visit: http://www.hpl.hp.com/research/qip/

References:

T. P. Spiller, Kae Nemoto, Samuel L. Braunstein, W. J. Munro, P. van Loock and G. J. Milburn, "Quantum Computation by Communication", *New J. Phys*. 8, 30 (2006)

T. D. Ladd, P. van Loock, K. Nemoto, W. J. Munro and Y. Yamamoto, "Hybrid Quantum Repeater Based on Dispersive CQED Interactions between Matter Qubits and Bright Coherent Light", *New J. Phys*. 8, 184 (2006)

Information and Quantum Systems Lab

Topic 31: Graphene Deposition

Graphene has shown great promise for interconnects and active devices, but current formation techniques cannot produce the large areas of graphene needed for practical applications. The objective of the proposed work is to find and develop a method of forming graphene over large areas using deposition techniques, such as chemical vapor deposition or molecular beam deposition.

In the past three years graphene (the two-dimensional, unrolled version of the carbon nanotube) has been shown to possess outstanding properties. Unlike carbon nanotubes, the two-dimensional structure of graphene makes integration into planar devices possible. It is being considered as a low-resistance, integrate-able electronic interconnection. Narrow ribbons of graphene behave as a semiconductor, and are being explored for electronic devices. Graphene has recently been shown to have a unique combination of optical transparency and low resistance. Graphene was included in the 2007 edition of the International Technology Roadmap for Semiconductors (ITRS), and is currently being considered for expanded coverage in the upcoming edition of the ITRS.

To date, however, graphene has been formed by techniques that produced only small areas of material. These techniques formed material that showed the outstanding properties of graphene, but are not compatible with practical, large-scale production. To make the use of graphene practical, it must be formed by deposition techniques, probably from the vapor phase. Its highly anisotropic structure is likely to enable forming graphene's two-dimensional structure by these deposition techniques.

The objective of the proposed work is to find and develop a method of forming graphene over large areas using deposition techniques, such as chemical vapor deposition or molecular beam deposition.

Research Questions

- How can graphene be formed over large areas using deposition techniques?
- Can graphene be deposited by chemical vapor deposition or molecular beam deposition?
- What deposition method is preferred?
- On what insulating substrates can graphene be deposited?

Systems Security Lab

Topic 38: Security Analytics: Understanding the threat environment

Chief Information Security Officers (CISOs) face increasing challenges to justify how they spend their budgets. Two drivers for this are the evolving threat environment and increasing complexity in the technology and systems used to provide information services. The problem is not restricted to securing the infrastructure. As employees become more mobile, with services accessed from the "cloud" and information shared across increasingly blurred boundaries, it will become more challenging to understand the attack surface, and to manage the gap between security policy and operations.

On one side we are interested in contributions that explain and characterize the evolving threat environment. More specifically, we want to learn how to use available (or potentially available) information to "measure" aggregate and individual threats, both in general and for enterprises in particular. For example, we can count and categorize software vulnerabilities and associated patches, exploits and malware -- but how do these things relate to the amount of website defacement, network disruption, fraud and data loss that occurs? A particular challenge is whether and how to compare threats from different levels; or, if the concern is sophisticated attacks, whether to be concerned about their combined effect.

This is challenging both because so much of the relevant information is not shared (e.g. breaches, criminal activity) and because it is difficult to characterize (e.g., subtly changing commercial and employment patterns and cultures). Moreover, there is little understanding of the economic framework within which threats to systems and services are encountered, leading to poor economic understanding of the investments required to defend systems and services at appropriate and identifiable levels of risk.

On the defense side we are interested in meaningful quantitative metrics that capture the security exposure of an enterprise. Unlike QoS metrics in networking, quantitative measures are generally missing in security discussions. Security metrics in the literature retain a qualitative flavor – most things are classified as high, medium or low.

Best-practice standards recommend monitoring and recording low-level detailed information at the level of network traffic and database transactions. Not only is this data expensive to collect, store and manage, but it is not clear what is to be done with all this low level data. Intrusion-detection systems, and monitoring alerts are notorious for the high rates of false positives; the result is that their output is mostly ignored. Moreover, even when they work correctly, the results of these systems can only be used in a reactive and operational manner.

Contributions in these areas will primarily support HP Labs' work in security analytics, which aims to produce tools and methods to improve all enterprise IT security decision-making, including where and how to invest in security policy and technology.

Research Questions

We are interested in innovative combinations of disciplines and stakeholders. For example, whilst any analysis is likely to require significant knowledge of IT security in the context of modern enterprise IT environments, we also suspect there are approaches from other disciplines (e.g. cognitive science, philosophy, economics) and problem domains (e.g. physical security, criminology, forensics) which will apply. Similarly whilst much of the necessary empirical data can be gleaned from experts and public information, we expect that certain stakeholders (e.g. customers, vendors, service providers) are likely to hold relevant information.

We are interested in proposals that attempt to explain and characterize the evolving threat environment and/or, on the defense side, create meaningful quantitative metrics that capture the security exposure of an enterprise.

On the threat side, example questions we'd expect such approaches to be able to provide insights on include:

- How to move from expert, intuitive viewpoints, based on unshared data, to a science having an appropriate framework for understanding of cause, effect and likelihood of threats in given IT situations?
- How to measure and compare the change in threats as we move to "cloud" computing and Web2.0 ways of using IT assets?
- How to develop the microeconomics of threats and investments in defenses (e.g., incentives, information asymmetries, games, mechanism design)?
- How to develop the macroeconomics of threats and investment in defenses (e.g., aggregative models of threats to CIA and matching defensive investments)?
- How to integrate models of systems and services with models of their economic and threat environments?

And on the defense side, we seek metrics that are

- measurable (i.e. computable from data that is easily collected); and
- meaningful (e.g. system-wide properties aggregated into measures that abstract away operational details).

To get there, we are interested in proposals that are potentially multi-disciplinary, but that certainly detail innovative ways of deriving metrics. Similarly, studies that investigate the possibility of using metrics as significant shortcuts in managing security will be of interest.

May 2008