Towards Integrated PSEs for Wireless Communications: Experiences with the S^4W and *SitePlanner*® Projects

Roger R. Skidmore[†], Alex Verstak[*], Naren Ramakrishnan[#], Theodore S. Rappaport[+], Layne T. Watson[#], Jian He[#], Srinidhi Varadarajan[#], Clifford A. Shaffer[#], Jeremy Chen[+], Kyung Kyoon Bae[$], Jing Jiang[$], and William H. Tranter[$]

[†]Wireless Valley Communications, Inc.
Austin, TX 78758

[*]Google, Inc.
Mountain View, CA 94043

[#]Department of Computer Science
Virginia Tech, Blacksburg, VA 24061

[+]Wireless Networking and Communications Group (WNCG)
Department of Electrical and Computer Engineering
University of Texas, Austin, TX 78712

[$]Bradley Department of Electrical and Computer Engineering
Virginia Tech, Blacksburg, VA 24061

October 28, 2003

Abstract

This paper describes the computational methodologies of two problem solving environments (PSEs) for wireless network design and analysis, one academic (S^4W) and one commercial (*SitePlanner*®). The PSEs address differently common computational issues such as environment specification, propagation modeling, channel performance prediction, system design optimization, and data management. The intended uses, interfaces, and capabilities of the two PSEs are compared and contrasted in a common framework. An important future direction, for these two and all future wireless system design PSEs, is resolving the fundamental impedance mismatch between physical channel modeling and upper level protocol modeling in wireless networks.

1 Introduction

In the 21st century, wireless networks are being transformed, from large-tower cellular systems, to high data-rate close range in-building networks, where coverage distances are on the order of tens of meters, not tens of kilometers as in the early days of cellular radio. To provide ubiquitous tetherless access for broadband wireless devices, the natural evolution of cellular is to replace large coverage cells with much higher densities of smaller cells.

Concomitant with these trends, wireless network design tools are also being transformed. While early CAD tools of the 1980's were primitive and used coarse mapping features and limited complexity models, today's design tools must contend with the complexities of wireless propagation, mobility, and interference characteristics (especially in indoor and combined indoor-outdoor environments); they hence cannot rely solely on the basis of large received signal strength and simple channel allocation protocols. Furthermore, such tools must support rapid design and deployment of networks where previously engineers and technicians needed to perform a significant amount of trial and error. A variety of technologies, involving databases, high-performance modeling, simulation support, and computational intelligence, have had to converge to support next generation design and deployment scenarios.

We report on our experiences in designing integrated simulation software, called 'problem solving environments (PSEs)' [12], for the modeling, analysis, and design of broadband wireless communication systems. As originally defined in [4], a PSE is a software system that provides all the computational facilities necessary to solve a target class of problems. Our focus here is on two PSEs developed over the past several years for wireless communications – S^{4}W and *SitePlanner*®. Both PSEs take as input a definition of a physical environment and provide accurate assessment of propagation characteristics and the performances of targeted wireless systems when situated in such an environment.

S^{4}W ('Site-Specific System Simulator for Wireless Communications') [6], funded by the NSF Next Generation Software Program, is an academic collaboration between computer scientists and electrical engineers and a testbed for research in novel computational technologies. SitePlanner [15], developed by Wireless Valley Communications, Inc., is the first widely used commercial software product for in-building wireless design and deployment. While these projects have independent origins, motivations, and funding sources, they have had to address similar problems, such as representing data about wireless environments, modeling propagation in a site-specific manner, characterizing the performance of wireless systems in given channels, designing and optimizing for defined performance objectives, and managing the complexities of computation involving relatively large datasets. We discuss them in a common setting here to help document important design and implementation decisions underlying integrated PSEs for wireless communications.

While the field of in-building wireless design is not new, and many corporate and academic programs have developed software tools for indoor and campus network deployment (e.g., WISE [3], Cindoor, EDX, WaveCall), S^{4}W and SitePlanner are the first software environments to systematize important facets of wireless network modeling and design. They abstract the creation and testing of wireless networks using software model libraries, help fuse measurements and simulation results, and provide facilities for automatically improving future performance by analyzing archived data from past scenarios (e.g., via data mining methodology and global optimization).

2 Background

S^{4}W adopts a model-centered perspective to the problem of designing and optimizing wireless communications networks. The PSE is organized as a library of site-specific deterministic electromagnetic propagation models as well as stochastic wireless system models for predicting the performance of wireless systems in specific environments, such as office buildings. Specific emphasis is placed on designing interchange formats that support the inclusion of new models into the PSE, composing the models to enable particular analysis and design scenarios, integration of optimization loops around the models, validation of models by comparison with field measurements, and management of the results produced by a large series of experiments. An experiment management system maintains model descriptions, simulation definitions, and simulation results in a relational database. S^4W has helped bring leading-edge computer science research in global optimization, markup languages, automated experiment management, and data mining to bear upon wireless communications system design and analysis.

Like S^4W, SitePlanner addresses site-specific wireless propagation prediction for modeling throughput, FER, BER, and other parameters in emerging wireless data networks. SitePlanner's emphasis, on the other hand, is on quick turn-around time for design and optimization; it provides a standard library of 'cookie-cutter' templates for wireless engineers, supporting popular design and management scenarios. SitePlanner relies on clever engineering assumptions and measurement integration to accurately predict network performance and equipment costs both inside and between buildings. It uses a merged indoor-outdoor site-specific representation that contains both internal and external building structural and terrain features. This enables universal prediction and visualization of wireless network performance and infrastructural layout on various scales, from macrocellular to indoor systems. In addition, the SitePlanner PSE has an editable bill of materials to track and optimize the location of various elements of network distribution systems; and provides real-time as well as non-real-time monitoring and control of network equipment. This capability highlights SitePlanner's role in network maintenance and management scenarios, in addition to design and deployment. SitePlanner is now considered an industry standard in the area of wireless network design, measurement, and optimization, and is licensed by over 250 companies worldwide.

3 Computational Components and Methodologies

To better understand the design philosophies of the two PSEs, we present their salient capabilities along eight key dimensions:

1. Physical Environment Modeling
2. Propagation Modeling
3. Characterizing Wireless Systems Performance
4. Equipment Modeling
5. Design and Optimization
6. Simulation Management
7. Simulation Output Analysis
8. Model Management

Physical Environment Modeling

An integrated PSE for wireless communications must begin by addressing the issue of how to best represent a given physical environment. Rapid advances in high-resolution satellite imagery and geographical information systems (GISs) have enabled the accurate, site-specific, capture of indoor, outdoor, and underground features, as well as terrain, vegetation, and meteorological effects. A range of data formats are available, with attendant merits and disadvantages. A wireless PSE must adopt a format that is sufficiently expressive for its purposes and, at the same time, computationally tractable.

Raster image formats such BMP, TIFF, and JPEG help overlay information such as elevation, population, and terrain type on a two-dimensional grid of pixels. They are the easiest form of information to obtain, via digital overhead imagery, photogrammetry, landuse surveys, and digital elevation models (DEMs). One of the main problems with these formats is the difficulty of distinguishing terrain regions from the outer physical structure of buildings and

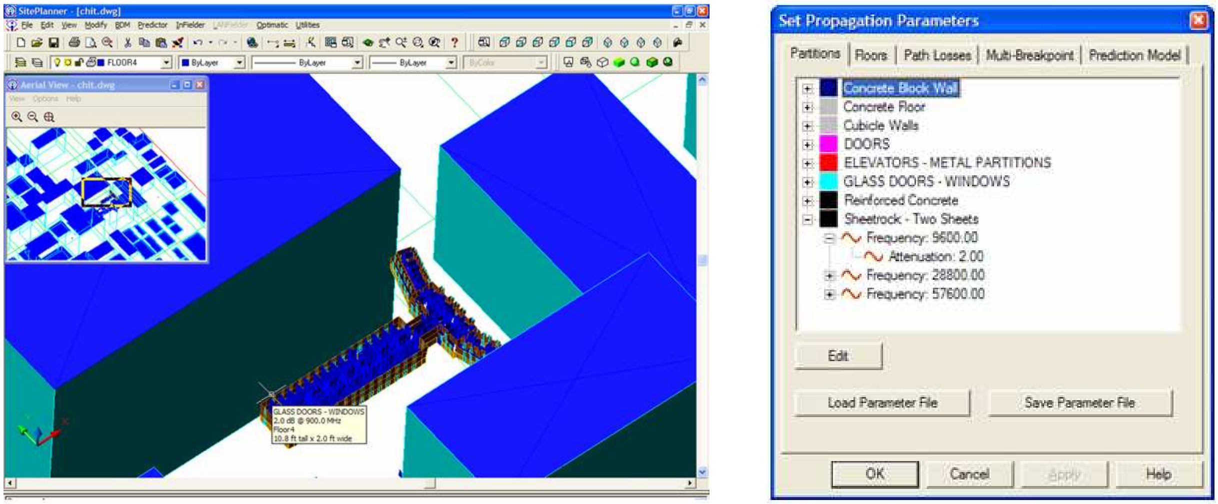

Figure 1: (left) Merged indoor-outdoor site-specific representation of a building environment created using SitePlanner. (right) Assigning electromagnetic and material properties to one or more surfaces in the site-specific model.

obstructions. Vector formats such as DXF and WMF capture a more logical organization of the physical environment, and help perform direct intersection tests between a radio wave propagating through the environment and the obstructions represented within the vector format. In contrast, intersection tests are not practical on raster datasets.

S^4W is primarily targeted at indoor environments. It uses the AutoCAD DXF format to capture a vector representation of internal building layouts. Surfaces are associated with physical obstructions and annotated with attributes such as material properties, for use in propagation modeling. SitePlanner supports both raster and vector formats of building structure and, in addition, is aimed at integrating indoor and outdoor site-specific models. The approach is to acquire external site-specific information in the form of a raster DEM, convert it into a vector format involving a triangular irregular network (TIN), apply edge detection algorithms to distinguish building surfaces from terrain, and finally overlay the indoor model using careful coordinate system transformations (see Figure 1). In addition, akin to S^4W, SitePlanner helps assign electromagnetic and material properties to each surface in the combined representation.

Propagation Modeling

With the availability of high-resolution site-specific geographical databases, deterministic radio wave propagation techniques based on ray tracing have become attractive for characterizing wireless channels [10]. Ray tracing methods are capable of estimating the complete spatio-temporal impulse response for any given receiver location, yielding high-fidelity channel models. These methods essentially approximate the solution to wave propagation equations by combining high-frequency assumptions with geometrical optics. A key issue for a wireless PSE is identifying the relevant effects that must be modeled and designing computational infrastructure for the fast computation of coverage and channel characteristics.

Both S^4W and SitePlanner support ray tracing as the primary mechanism for propagation prediction and use pyramidal beams to model electromagnetic waves. The beams are shot from geodesic domes drawn around transmitter locations. A geodesic dome is a geometric figure obtained by tesselating the faces of an icosahedron and extrapolating points to the surface of an encapsulating sphere. Each beam is a triangular pyramid formed by the point location of the transmitter and a triangle on the surface of the dome. Beams are traced through reflections and penetrations through obstructions in a particular environment. Once an intersection with a receiver location is detected, a ray is

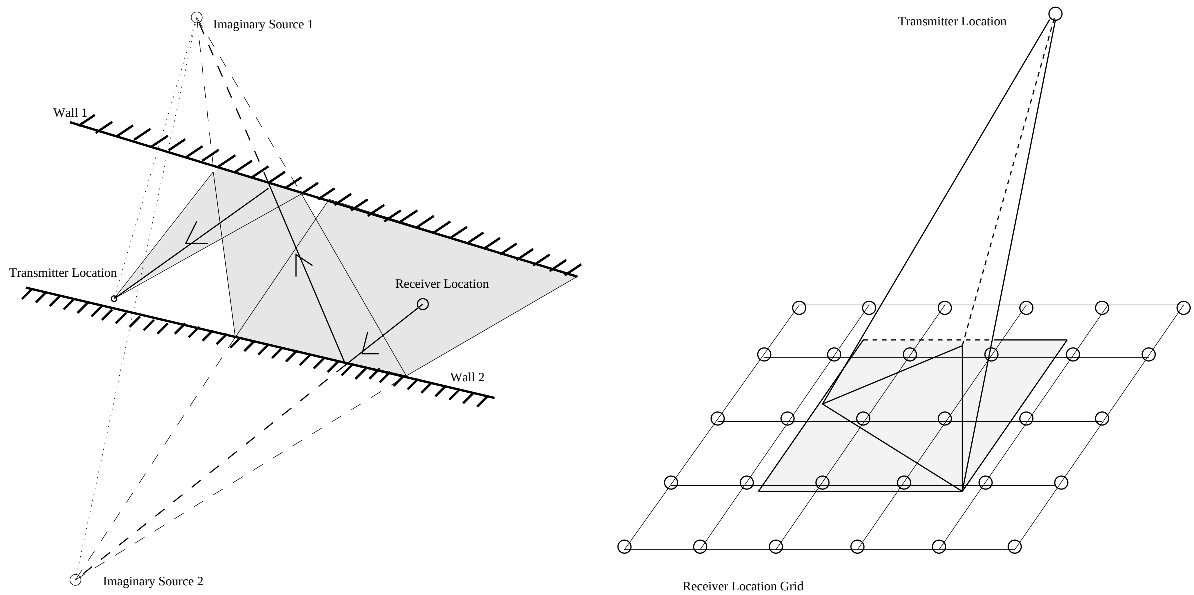

Figure 2: 2D beam tracing used by both S^4W and SitePlanner. (left) A beam (shadowed region) is traced from the transmitter location to the receiver location through two reflections. Then, a ray (bold line) is traced back. (right) Beam intersection with a receiver location grid. Only the locations inside the bounding box of the projection of the beam onto the grid (shadowed region) are tested for intersection with the beam pyramid.

traced back from the receiver location to the transmitter location through the sequence of reflections and penetrations encountered by the beam [13]. The illustration of this process in 2D is given in Figure 2 that shows an intersection test of a beam with a grid of uniformly spaced receiver locations. By default S^4W uses horizontal reception planes and SitePlanner uses reception spheres. Neither S^4W nor SitePlanner currently handle diffraction or scattering. These effects play an important role in propagation (especially around sharp edges and rooftops), but their modeling is computationally expensive.

Several improvements to this basic technique are employed by S^4W and SitePlanner. S^4W uses octree space partitioning to reduce the number of intersection tests and image parallelism with dynamic scheduling to reduce simulation runtime. Image parallelism refers to the configuration where each processor has a complete copy of the physical environment and is apportioned a fraction of the total number of rays to process. This has advantages over object parallelism, where each processor works with a partition of the environment and where the communication costs can get prohibitive. S^4W's ray tracer currently operates on a 200-node cluster of workstations and is specially configured for fast parallel computing. SitePlanner's ray tracer operates in single-processor mode, uses an irregular spherical bounding volume hierarchy, and resorts to using alternating horizontal and vertical reception planes instead of reception spheres when computational cost increases. These approaches are only two possibilities to configuring a parallel ray tracer. A wealth of literature exists on other options — for instance, Kimpe et al. [10] describe a solution where the environment is partitioned according to locations with similar propagation characteristics.

To make the predicted impulse response match the output of a real instrumentation system, both S^4W and SitePlanner allow the application of antenna patterns and resampling heuristics. Both PSEs enable the calculation of gain for omnidirectional, waveguide, pyramidal horn, and biconical antennas, and convolution of the discrete impulse response with suitably designed Gaussian filters. Experimental results with channel sounders show that ray tracing predictions are consistently within 3-5dB of measurements.

In integrating ray tracing simulations in larger computational contexts (e.g., optimization with respect to received signal strength intensity – RSSI), we have come to appreciate the importance of *surrogate* models for propagation

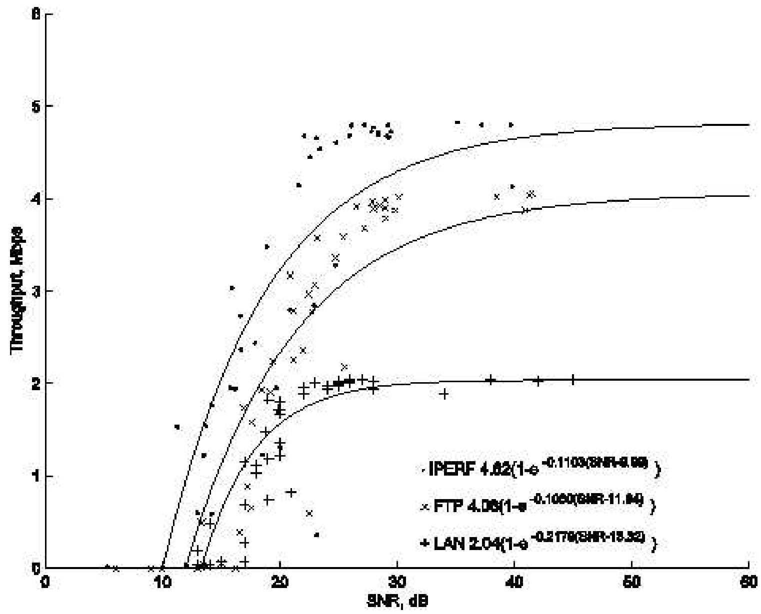

Figure 3: Throughput surrogate model, as a function of SNR, estimated from measurements made at Scholotzsky's Deli, Northcross Restaurant, Texas using a Cisco PCMCIA 350 card. IPERF is the University of Illinois's tool for estimating througput as a function of TCP-window size; LANFIelder employs a single packet-on-the-fly style; and the FTP application studies transfers of calibrated file sizes. Further details about this study are available in [1].

prediction. Surrogates are cheap mathematical approximations to the output from costly simulation codes. In S^4W, they can be derived from measured data using techniques such as curve fitting and response surface methodology. In SitePlanner, surrogates are configured using distance-dependent and partition-based path loss models.

Characterizing Wireless Network Performance

Modern wireless communications systems require more sophisticated analysis than merely ensuring a strong and reliable signal between a transmitter and a receiver. Historically, wireless engineers relied on predictions of signal strength (RSSI) and interference levels. However, it is now clear that data communication applications, wireless protocols, and air interface standards play a central role in wireless networks, and hence the characterization of metrics such as BER, PER (packet error rate), FER, and throughput becomes important. For instance, the link-layer/MAC protocols employed in a multi-user wireless network govern end-to-end capacity for a particular application, and must be taken into account while studying performance [14]. A wireless PSE can support such systems characterization using either statistical simulation or by data-driven methodologies.

In S^4W, the performance of particular wireless systems operating in a given channel is modeled via Monte Carlo simulation. The current emphasis is on WCDMA systems with optional STTD (space time transmit diversity) and convolutional codes. The simulation traces a number of frames of random information bits through the encoding filters, the channel (modeled as a linear time varying process), and the decoding filters. The inputs are hardware parameters, average SNR, channel impulse response (e.g., simulated by the ray tracer), and the number of frames to simulate. The outputs are the BER and the FER. The WCDMA simulation is computationally intensive since a satisfactory BER value ranges from 10^{-3} to 10^{-6} and depends on small-scale propagation effects that exhibit large variation with respect to receiver location. As a result, S^4W utilizes both a parallel version of the simulation and surrogates using least-squares fitting and multivariate adaptive regression spline (MARS) techniques.

In SitePlanner, network performance is modeled phenomenologically from measured data via surrogate functions.

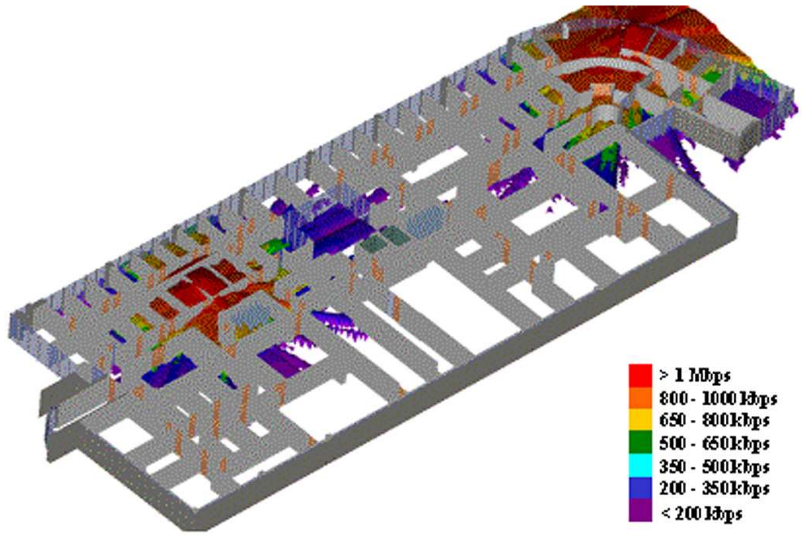

Figure 4: Throughput prediction using SitePlanner. Here the lookup tables approximate throughput given predicted SNR.

For instance, a general form for a surrogate function to predict throughput T is given by:

$$T = C_1[Ad + Bd^2 + C] + C_2[D(RSSI) + E(RSSI)^2 + F] + C_3 \sum_{i=1}^{M}(G_i P_i + K_i)$$

where T is the observed end-to-end application throughput or raw throughput, d is the distance in meters between transmitter and receiver (either through a straight linear path between transmitter and receiver or a traced path representing the route taken by radio waves passing through the site-specific model between the transmitter and receiver), $RSSI$ is the received signal strength intensity, M denotes either signal components from one or more transmitters or a combination of important multipath components from a collection of transmitters, G's represents processing gains for each multipath component, P's represents power levels for each multipath component, and A, B, C, C_1, C_2, C_3, D, E, F, and K's are either constants, or linear or nonlinear functions of environment types, packet sizes, manufacturer equipment types, air interface standards, modulation schemes, or other parameters [8]. The above equation provides a very general, comprehensive formulation for predicting throughput in wireless data networks, but the type and value of each parameter used are specific to the combination of equipment types employed.

For example, measurements in a public wireless LAN network (Schlotzsky's Deli) show that the above equation can be simplified into:

$$T = \alpha[1 + \beta e^{-\gamma(\text{SNR}[+\text{SIR}])}]$$

where α, β, γ are parameters to be estimated. Fig. 3 displays the fit for data collected at a Texas restaurant using three different upper-layer protocol instrumentation strategies (IPERF, FTP, and LANFielder).

Once models for desired network characteristics are available, interpolation techniques may then be used to create lookup tables, which can predict network characteristics over a region of interest (see Figure 4). It should be clear that a PSE could model/learn a wide variety of behaviors from measured data, and implement them as predictions for future network provisioning.

Equipment Modeling

SitePlanner (but not S^4W) can represent and accommodate information about individual wireless network equipment layout, interconnection, and performance. Besides base stations, elements such as mobile receivers, antennas, cables,

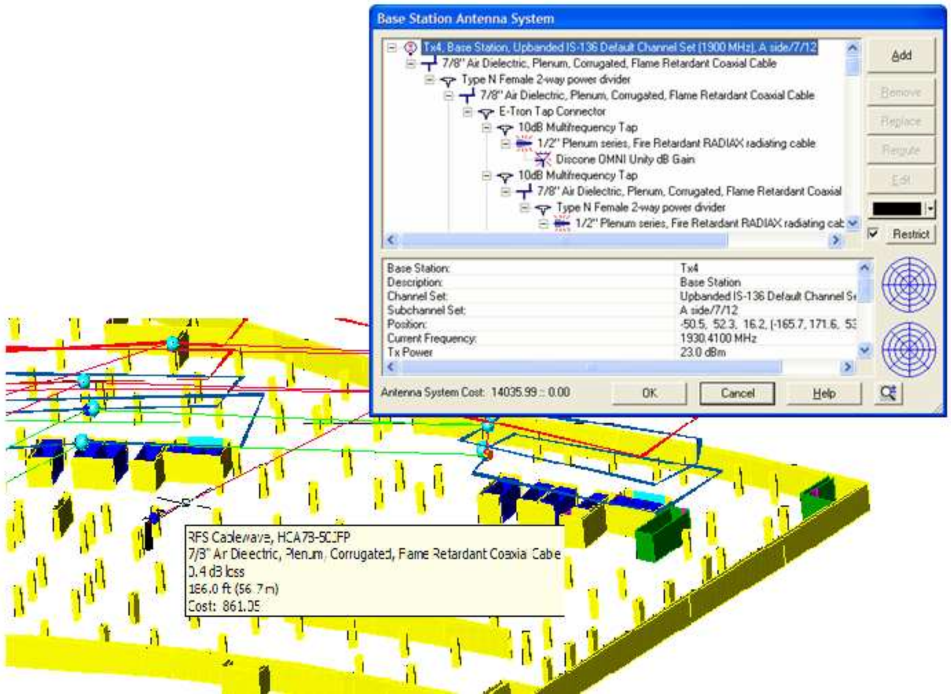

Figure 5: Using SitePlanner to plan the site-specific placement and interconnection of equipment alongside logical equipment interconnections.

connectors, splitters, amplifiers, and antennas can be modeled in SitePlanner. The type of equipment used, the manner in which the equipment is interconnected, and the combination of different transmitter and receiver types can directly affect the simulated and actual performance of a wireless communication network. For instance, it has been shown that intermixing WLAN access points from one vendor with PCMCIA WLAN cards from a different vendor will result in different performance than if both access point and PCMCIA card were from the same vendor [8]. Even more obvious, different antennas made by different vendors will provide different coverage patterns, and the placement of antennas or cable lengths impacts system performance.

SitePlanner employs an XML-based equipment library for various categories of instrumentation and describes their characteristics such as manufacturer, part number, insertion loss, coupling loss, amplification/attenuation, number and modality of input connectors, number and modality of output connectors, noise figure, frequency variable effects, radiation characteristics (for antennas), cost, and maintenance history. Since equipment behaves differently depending on the frequency of the signal driving it (e.g., the loss of a cable is different at 900 MHz than at 5.85 GHz), the equipment library fully supports and accounts for frequency-dependent effects. Positioning software is provided to help plan the layout and placement of equipment (see Figure 5). SitePlanner then translates predicted electromagnetic wave interactions between different types of transmitter and receiver equipment into an estimation of throughput or BER performance.

Design and Optimization

PSEs for wireless communications are especially important for design purposes, e.g., optimizing transmitter placement to ensure acceptable system performance within a geographical area of interest (see Figure 6). Both S^4W and SitePlanner support optimization, with respect to different criteria. Of particular importance are the choice of performing local [3] versus global [7] optimization, the nature of the objective function, and the ability to incorporate

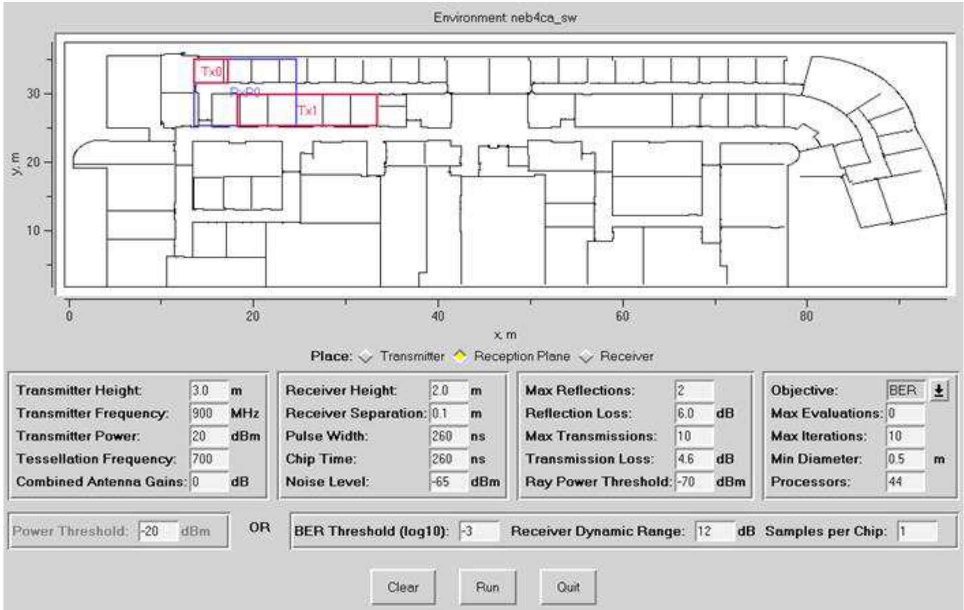

Figure 6: S^{4}W interface for global optimization of transmitter placement in an indoor environment (Durhman Hall 4th floor, Virginia Tech).

domain-specific constraints into the optimization process.

The metrics used in S^{4}W are continuous penalty functions defined in terms of power levels (e.g., power coverage) and bit error rates at given receiver locations within the covered region. Both objective functions are devised to minimize average shortfall of the estimated performance metric with respect to a threshold. In contrast to local optimization (e.g., based on gradient descent), S^{4}W performs a global optimization using DIRECT (DIviding RECTangles), an algorithm proposed by Jones et al. [9]. DIRECT is a pattern search method that is categorized as a direct search technique by Lewis et al. [11] and is named after one of its key steps—dividing rectangles. Generally speaking, “pattern search methods are characterized by a series of exploratory moves that consider the behavior of the objective function at a pattern of points” [11], which are chosen as the centers of rectangles in the DIRECT algorithm. This center-sampling strategy reduces the computational complexity, especially for higher dimensional problems. Moreover, DIRECT adopts a strategy of balancing local and global search by selecting potentially optimal rectangles to be further explored. This strategy gives rise to fast convergence with reasonably broad space coverage. The algorithm is guaranteed to converge globally if the objective function satisfies the property of Lipschitz continuity. Traditional performance criteria, such as coverage and BER, do not satisfy this condition and a careful reformulation is required in order to apply the DIRECT algorithm [6]. As mentioned earlier, surrogate models are employed to lower the computational cost.

In SitePlanner, a local form of optimization is conducted in the equipment planning module. For the case of equipment having allowable placements at many possible locations, a processing loop recursively attempts to remove already placed equipment to see if the removal undesirably affects performance. Figure 7 shows how an initial layout involving 42 WLAN access points is optimized w.r.t. RSSI.

Simulation Management

S^{4}W (but not SitePlanner) can organize and manage a large set of scientific experiments. It uses the PostgreSQL relational database system to record descriptions of simulations, particulars about how they should be run, and their

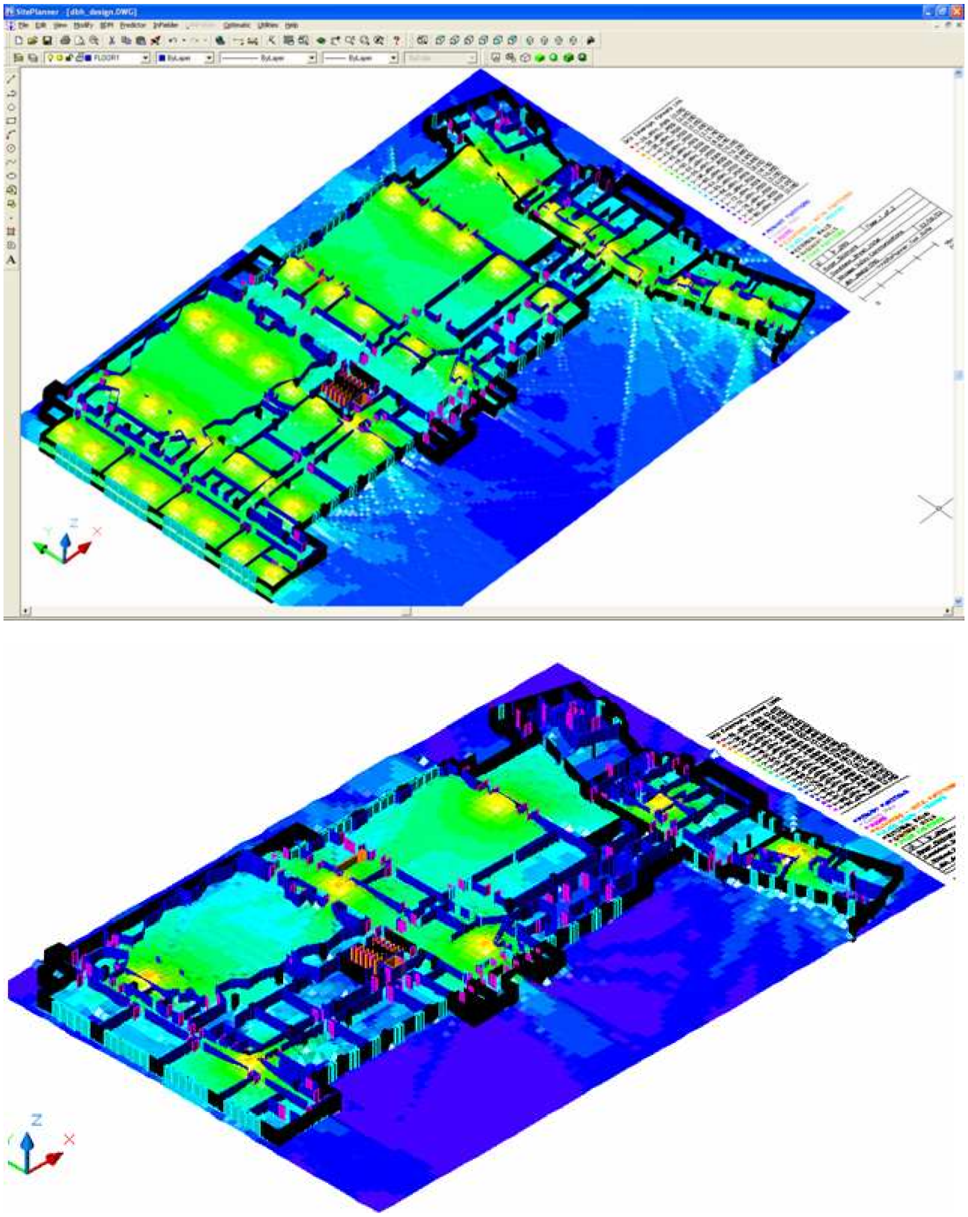

Figure 7: Using SitePlanner to plan equipment placement. (top) At the beginning of the processing loop, 42 WLAN access points have been allowed to be placed at all possible locations in the region. Notice the relatively large number of very strong hot (white) spots. (bottom) A layout with reduced number of WLAN access points that still satisfies the desired performance goals.

results. This is advantageous for several reasons. First, analysis of simulation data stored in a database is often faster and simpler than repeated execution of the essentially unchanged simulation. Storing simulation or measurements data in the database also simplifies comparison between simulation results and measurements (validation, calibration) and comparison of several simulation runs (parameter studies, "what if" scenarios). Second, having a history of simulation descriptions eases setting up new simulations — parameters can be copied and chances of repeating past mistakes are reduced. Finally, structured storage of simulation parameters enables automated execution of simulations on a variety of computing systems, a task necessitated by today's rapidly evolving cluster computing technology.

The simulation management system in S^4W consists of three parts: database schemas and functions, execution scripts, and simulation wrappers. Database schemas define tabular data formats for both generic parameters applicable to all simulations (e.g, number of processors, execution date, execution command) and parameters specific to the given simulation (e.g., impulse response, bit error rate, number of transmitter antennas). Execution scripts, written in Tcl, interface to commonly used job schedulers such as Unix process scheduler for Unix workstations and PBS for clusters of PCs. These scripts start and stop simulations on given computing systems and update the database accordingly. Finally, simulation wrappers are scripts that convert simulation inputs and outputs between their database representation and the formats required by the simulation software. This section emphasizes systemic issues so user interface software is only mentioned briefly.

The user normally enters simulation parameters via a graphical interface similar to that shown in Figure 6. The user interface software writes simulation inputs to the database and invokes a database procedure to start simulation execution. The generic execution procedure reads the execution record from the database and invokes the appropriate execution script. The execution script tells the job scheduler to run the appropriate simulation wrapper which, in turn, reads simulation inputs from the database, communicates them to the simulation, and saves simulation status and outputs to the database as they are computed. Then, visualization software reads simulation outputs from the database and displays them in a suitable form.

Different components of this simulation management system communicate via database updates and queries. PostgreSQL provides an extremely versatile query language that is used to compute values of the objective functions for optimization, estimate interesting regions for fitting surrogate models, convolve discrete impulse responses with Gaussian filters for comparison with measurements, estimate the confidence intervals for the bit error rates computed by the WCDMA simulation, and to perform many other data transformations. This query language is also relatively simple and well suited for interactive use, and a great aid to prototyping as well as debugging.

The updates complement the queries by providing a control flow to simulation execution. For simple simulations, it suffices to insert the inputs into the appropriate table and execute the "go" query; then, the outputs will appear in the output table as they are computed. In general, S^4W supports a *controller-executor* relationship between system components where the controller iteratively writes simulation inputs to the inputs table and the executor iteratively grabs these inputs and writes the outputs to the outputs table. Executors terminate when the inputs are exhausted and all controllers have terminated. Controllers can be programmed to various input creation and termination strategies. This interplay enables the creation of problem solving abstractions such as parameter sweeps (e.g., 'simulate a given WCDMA system for a wide range of modulation schemes, rake receiver sizes, multipath channel environments') and feedback loops (e.g., 'optimize a simulation configuration for a desired performance objective', or 'find regions of the configuration space that exhibit acceptable criteria'). For more details, please see [16].

Simulation Output Analysis

Having a central facility to record and manage simulation data enables us to bring advanced data analysis techniques to bear upon the study of wireless systems. In contrast to traditional approaches for characterizing wireless system performance such as power delay profiles and BER-vs-SNR curves, S^4W pioneered the use of *data mining* [5] in wireless simulation methodology [17]. Data mining entails the 'non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data' [2]. It can be used in both predictive (e.g., quantitative

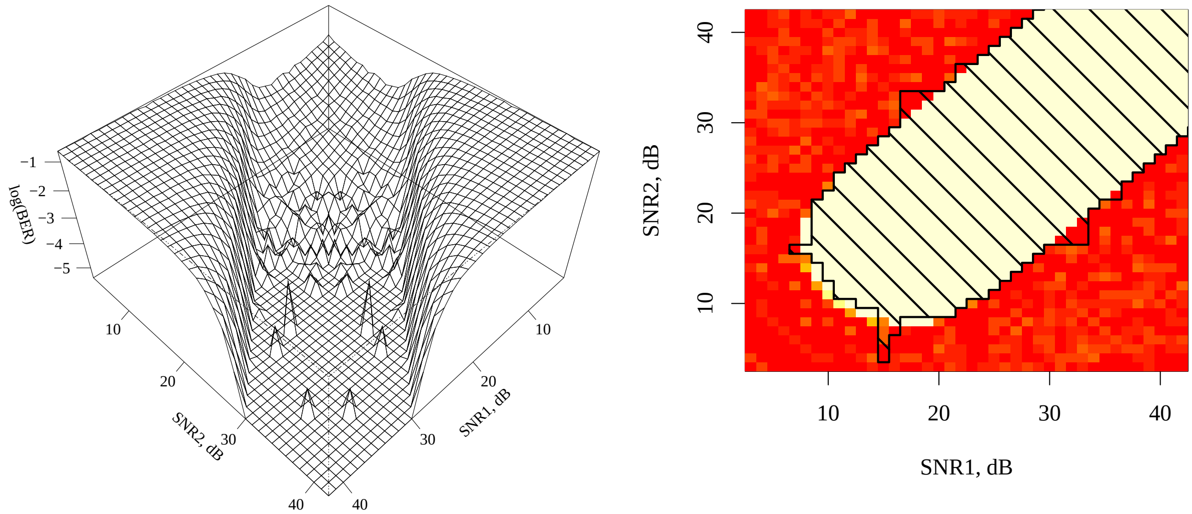

Figure 8: Using the data mining functionality of S^4W to determine configuration space regions that exhibit acceptable average performance. (left) Statistically significant estimates of the expected BERs for the configuration space. (right) Region mined by the data mining algorithm. Green (light) denotes high confidence and red (dark) denotes low confidence in acceptable average performance.

assessment of factors on some performance metric) and descriptive (e.g., summarization and system characterization) settings. In S^4W, we first organize a performance database of simulation runs that sweep over a targeted space of system configurations; this database is then mined to identify regions of configuration space that exhibit acceptable average performance.

One study we have performed is to explore the effect of power imbalance in a two-branch STTD setup on the BER of a simulated WCDMA system across a range of average signal-to-noise (SNR) ratios. The data mining algorithm performs systematic aggregation and redescription of data into higher-level objects. Our work can be viewed as employing three such layers of aggregation: points, buckets, and regions. Points (configurations) are records in the performance database (WCDMA simulation outputs). These records contain configuration parameters as well as unbiased estimates of bit error probabilities that we use as performance metrics. Buckets represent averages of points. We use buckets to reduce data dimensionality, often to two key WCDMA configuration parameters. Finally, buckets are aggregated into 2D regions of constrained shape. We find regions of buckets where we are most confident that the configurations exhibit acceptable average performance. The shapes of these regions illustrate the nature of the joint influence of the two selected configuration parameters on the configuration performance. Specific region attributes, such as region width, provide estimates for the thresholds of sensitivity of configurations to variations in parameter values.

Figure 8 depicts the use of data mining for finding configuration space regions, after ensuring statistical significance of the data. Intuitively, the region shown in Figure 8 (right) is the largest admissible region where we can claim, with confidence at least 0.99, that configurations exhibit acceptable average performance. Red (dark) corresponds to low confidence and white (light) corresponds to high confidence that the average BER of the system meets the voice quality BER threshold 10^{-3}. The shape of this region confirms that, under a fixed effective SNR, the BER is minimal when the average SNRs of the two branches are equal. The width of this region shows the largest acceptable power imbalance. For this example, the system tolerates power imbalance of up to 12 dB. However, the width of the optimized region is not uniform. The region is narrower for small effective SNRs and wider for large effective SNRs. This confirms that configurations with low effective SNRs are more sensitive to power imbalance than configurations with high effective SNRs.

Model Management

Finally, an essential role of a PSE is to automate the tedious tasks of managing and reasoning about models, their characteristics, input and output formats, semantic properties, and in this manner codify the often unarticulated domain expertise that underlies the practice of computational science.

S^4W, in conjunction with its integrated simulation management system, uses an XML-based representation of wireless models to manage their execution and composition. A markup language called BSML ('Binding Schema Markup Language') [18] helps formalize the data interchange that happens when scientific codes are composed in the PSE. BSML helps reason if a given input is valid for a given model, performs any necessary binding to ensure that the model can be 'run,' and when inconsistencies arise, reasons about how the data can be 'massaged' in order to work with the model. Such facilities are the cornerstone of the PSE's ability to handle changes and interface mismatches. They help retain historical data and facilitate inclusion of new components.

As an example, Figure 9 depicts two slightly different schemas for antenna descriptions in S^4W. The schema at the bottom (actual schema) was our first attempt at defining a data format for antenna descriptions. This version supported only one antenna type and exhibited several inadequate representation choices. For example, polar coordinates should have been used instead of Cartesian coordinates because antenna designers prefer to work in the polar coordinate system. Antenna gain was not considered in the first version because its effect is the same as changing transmitter power. However, this seemingly unnecessary parameter should have been included because it results in a more direct correspondence of simulation input to a physical system.

The schema at the top of Figure 9 (required schema) improves upon the actual schema in several ways. It better adheres to common practices and supports more antenna types. However, this schema is different from the actual schema, while compatibility with old data needs to be retained. Figure 10 illustrates how addition of conversion and binding codes (see the `<code>` elements) to the actual schema solves the compatibility problem. A parser generated from the conversion schema in Figure 10 will recognize the actual data and provide the required binding. In this manner, In this manner, the evolution of models, data formats, and conventions can be systematically accommodated in S^4W.

Model management is emphasized to a lesser extent in SitePlanner where the primary goal is to accommodate site-specific databases and equipment models; this is because SitePlanner is targeted at rapid design, deployment, feedback, and control of network equipment, rather than to be used in a research environment supporting the evolution of computational methodologies.

4 Discussion

This survey has identified several important functionalities that an integrated PSE for wireless communications must support. Through the presentation of two state-of-the-art systems — S^4W and SitePlanner — we have seen how these functionalities are commonly provided. The importance of such simulation tools will only grow as their predictive capabilities improve, and as they continue to provide important abstractions for problem solving. For want of space, we have not discussed many other facets of PSEs, such as model validation, user interfaces, and visualization. See [19] for a comprehensive list of PSE facets (but identified in a different engineering context).

Current work is focused on three major thrusts. First, the capabilities of S^4W and SitePlanner are being extended to address newer case studies, such as MIMO communication system design, and optimization with respect to custom-defined criteria such as array and coding schemes. Second, we are beginning to investigate the knowledge-based

```
<element name='antennas'>
  <repetition>
    <element name='antenna'>
      <element name='id' type='string' min='1'/>
      <element name='phi' type='angle'/>
      <element name='theta' type='angle'/>
      <element name='gain' type='ratio' units='dB' optional='true' default='0'/>
      <code>puts stdout "%id: %phi %theta %gain"</code>
      <selection>
        <element name='waveguide'>
          <element name='width' type='distance' units='mm'/>
          <element name='height' type='distance' units='mm'/>
          <code>puts stdout "waveguide: %width %height"</code>
        </element>
        <element name='pyramidal_horn'>
          <element name='width' type='distance' units='mm'/>
          <element name='rw' type='distance' units='mm'/>
          <element name='height' type='distance' units='mm'/>
          <element name='rh' type='distance' units='mm'/>
          <code>puts stdout "pyramidal horn: %width %rw %height %rh"</code>
        </element>
      </selection>
    </element>
  </repetition>
</element>

<element name='antennas'>
  <repetition>
    <element name='antenna'>
      <element name='id' type='string' min='1'/>
      <element name='description' type='*'/>
      <element name='x' type='coordinate'/>
      <element name='y' type='coordinate'/>
      <element name='z' type='coordinate'/>
      <element name='waveguide'>
        <element name='width' type='distance' units='in'/>
        <element name='height' type='distance' units='in'/>
      </element>
    </element>
  </repetition>
</element>
```

Figure 9: Two slightly different S^4W schemas for a collection of antennas. The component requires the top schema, but the data conforms to the bottom schema. The bottom schema (a) represents antenna orientation in Cartesian coordinates, not polar coordinates, (b) lacks antenna gain, (c) requires antenna descriptions, (d) measures antenna dimensions in inches, not millimeters, and (e) covers only one antenna type.

```
<element name='antennas'>
  <repetition>
    <element name='antenna'>
      <element name='id' type='string' min='1'/>
      <element name='description' type='*'/>
      <element name='x' type='coordinate'/>
      <element name='y' type='coordinate'/>
      <element name='z' type='coordinate'/>
      <code>  <!-- convert coordinates from rectangular to polar -->
        set _r [expr sqrt(%x*%x+%y*%y+%z*%z)]
        set %theta [expr atan2(%y,%x)]
        set %phi [expr acos(%z/$_r)]
      </code>
      <code>  <!-- set default gain -->
        set %gain 0
      </code>
      <code>puts stdout "%id: %phi %theta %gain"</code>
      <element name='waveguide'>
        <element name='width' type='distance' units='mm'/>
        <code>  <!-- convert units from inches to millimeters -->
          set %width [expr 25.4*%width]
        </code>
        <element name='height' type='distance' units='mm'/>
        <code>  <!-- convert units from inches to millimeters -->
          set %height [expr 25.4*%height]
        </code>
        <code>puts stdout "waveguide: %width %height"</code>
      </element>
    </element>
  </repetition>
</element>
```

Figure 10: Actual S^4W schema from Figure 9 (bottom) after inserting conversion and binding codes. This schema describes the actual data, but provides the bindings of the required schema (top of Figure 9).

composition of wireless models to address targeted design and modeling scenarios. For instance, given a physical environment, a metric of interest, and performance constraints on prediction accuracy, what types of models must be composed in order to achieve the desired criteria? Notice that greater fidelity of models need not necessarily translate into improved accuracy, since the additional effects captured may not be pertinent in the given environment; hence simpler computational machinery might be cost-effective in certain situations. As the number of available models increases, a PSE can help recommend the choice of models for particular applications.

Finally, we are working to unify the channel and physical level modeling presented here with the upper layers of the wireless network protocol hierarchy. This will enable us to use cross-layer knowledge to enable the next generation of ubiquitous wireless applications. Observe that this is a problem of not just integrating models of different types, but raises a more fundamental question of integrating diverse computational methodologies. For instance, continuous time models are typically used in physical layer modeling whereas discrete-event models are used for the network and transport layers. As a result, higher layer protocols are traditionally tested using discrete-event simulation or direct code execution whereas lower-level protocols are studied by the mechanisms described in this article. There is thus a fundamental impedance mismatch in the underlying programming models that is not easily overcome by transforming one into the other. This type of mismatch is well known in other areas of science and engineering such as multidisciplinary design optimization and multiscale materials science. Studying these issues will drive the development of new methods and techniques for the next generation of wireless networks.

Acknowledgements

This work is supported by US National Science Foundation grants EIA-9974956, EIA-9984317, and ACI-0305644 (Next Generation Software).

References

[1] J.K. Chen. *Measurements and Modeling of Throughputs and Signal Strengths of IEEE 802.11b Networks in and around Restaurants*. Technical report, Wireless Networking and Communications Group (WNCG), University of Texas, Austin, Oct 2003.

[2] U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, Vol. 39(11):pages 27–34, Nov 1996.

[3] S.J. Fortune, D.M. Gay, B.W. Kernighan, O. Landron, R.A. Valenzuela, and M.H. Wright. WISE Design of Indoor Wireless Systems: Practical Computation and Optimization. *IEEE Computational Science and Engineering*, Vol. 2(1):pages 58–68, Spring 1995.

[4] E. Gallopoulos, E.N. Houstis, and J.R. Rice. Computer as Thinker/Doer: Problem Solving Environments for Computational Science. *IEEE Computational Science and Engineering*, Vol. 1(2):pages 11–23, 1994.

[5] J. Han, R.B. Altman, V. Kumar, H. Mannila, and D. Pregibon. Emerging Scientific Applications in Data Mining. *Communications of the ACM*, Vol. 45(8):pages 54–58, Aug 2002.

[6] J. He, A. Verstak, L.T. Watson, C.A. Stinson, N. Ramakrishnan, C.A. Shaffer, T.S. Rappaport, C.R. Anderson, K.K. Bae, J. Jiang, and W.H. Tranter. S^4W: Globally Optimized Design of Indoor Wireless Communication Systems. In F. Darema, editor, *Proceedings of the Next Generation Software Systems Workshop, 16th International Parallel and Distributed Processing Symposium (IPDPS'02)*. IEEE Computer Society, Apr 2002. 8 pages.

[7] J. He, L.T. Watson, N. Ramakrishnan, C.A. Shaffer, A. Verstak, J. Jiang, K. Bae, and W.H. Tranter. Dynamic Data Structures for a Direct Search Algorithm. *Computational Optimization and Applications*, Vol. 23(1):pages 5–25, Oct 2002.

[8] B. Henty. *Throughput Measurements and Empirical Prediction Models for IEEE 802.11b Wireless LAN (WLAN) Installations*. Master's thesis, Bradley Department of Electrical and Computer Engineering, Virginia Tech, 2001.

[9] D.R. Jones, C.D. Perttunen, and B.E. Stuckman. Lipschitzian Optimization without the Lipschitz Constant. *Journal of Optimization Theory and Applications*, Vol. 79(1):pages 157–181, 1993.

[10] M. Kimpe, H. Leib, O. Maquelin, and T.H. Szymanski. Fast Computational Techniques for Indoor Radio Channel Estimation. *IEEE/AIP Computing in Science and Engineering*, Vol. 1(1):31–41, Jan–Feb 1999.

[11] R.M. Lewis, V. Torczon, and M.W. Trosset. Direct Search Methods: Then and Now. *Journal of Computational and Applied Mathematics*, Vol. 124:pages 191–207, 2000.

[12] N. Ramakrishnan, L.T. Watson, D.G. Kafura, C.J. Ribbens, and C.A. Shaffer. Programming Environments for Multidisciplinary Grid Communities. *Concurrency and Computation: Practice and Experience*, Vol. 14(13–15):pages 1241–1273, Dec 2002.

[13] S.Y. Seidel and T.S. Rappaport. Site-Specific Propagation Prediction for In-Building Personal Communication System Design. *IEEE Transactions on Vehicular Technology*, Vol. 43(4):pages 879–891, Nov 1994.

[14] S. Shakkottai, T.S. Rappaport, and P.C. Karlsson. Cross-Layer Design for Wireless Networks. *IEEE Communications Magazine*, Vol. 41(10), Oct 2003.

[15] R. R. Skidmore. *A Comprehensive Method and System for the Design and Deployment of Wireless Data Networks*. PhD thesis, Bradley Department of Electrical and Computer Engineering, Virginia Tech, April 2003.

[16] A. Verstak. *Data and Computation Modeling for Scientific Problem Solving Environments*. Master's thesis, Department of Computer Science, Virginia Tech, Aug 2002.

[17] A. Verstak, N. Ramakrishnan, K.K. Bae, W.H. Tranter, L.T. Watson, J. He, C.A. Shaffer, and T.S. Rappaport. Using Hierarchical Data Mining to Characterize Performance of Wireless System Configurations. Technical Report cs.CE/0208040, Computing Research Repository (CoRR), Aug 2002.

[18] A. Verstak, N. Ramakrishnan, L.T. Watson, J. He, C.A. Shaffer, K. Bae, J. Jiang, W.H. Tranter, and T.S. Rappaport. BSML: A Binding Schema Markup Language for Data Interchange in PSEs. *Scientific Programming*, Vol. 11(3):pp. 199–224, Aug 2003.

[19] L.T. Watson, V.K. Lohani, D.F. Kibler, R.L. Dymond, N. Ramakrishnan, and C.A. Shaffer. Integrated Computing Environments for Watershed Management. *ASCE Journal of Computing in Civil Engineering*, Vol. 16(4):pages 259–268, Oct 2002.