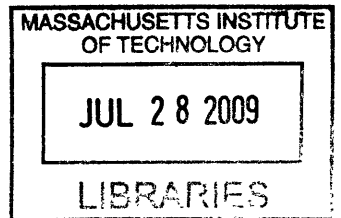# CHARACTERIZING CAPITAL AND OPERATIONAL TRADEOFFS RESULTING FROM FIBER-TO-THE-HOME OPTICAL NETWORK ARCHITECTURE CHOICE

by

Thomas Rand-Nash
B.A in Physics, University of California, Berkeley, 2005

Submitted to The Engineering Systems Division and
The Department of Materials Science & Engineering
in Partial Fulfillment of the Requirements for the Degrees of
Master of Science in Technology and Policy
and
Master of Science in Materials Science and Engineering
at the
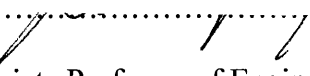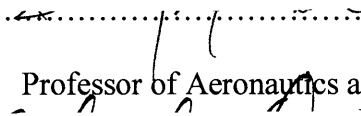Massachusetts Institute of Technology
June 2009

Signature of Author....

Technology and Policy Program, Engineering Systems Division,
Materials Science and Engineering,
Submitted 5/15/09; Defended 5/21/09

Certified by.........

Richard Roth
Director, Materials Systems Laboratory
Thesis Supervisor

Certified by.........

Randolph Kirchain
Associate Professor of Engineering Systems and Materials Science and Engineering
Thesis Supervisor

Accepted by.................

Dava Newman
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Technology and Policy Program

Accepted by ....

Christine Ortiz
Chair, Departmental Committee on Graduate Students

CHARACTERIZING CAPITAL AND OPERATIONAL TRADEOFFS RESULTING FROM FIBER-TO-THE-HOME OPTICAL NETWORK ARCHITECTURE CHOICE

By

Thomas F. Rand-Nash

Submitted to the Engineering Systems Division and the Department of Materials Science and Engineering on May 15, 2009 and defended on May 21, 2009 in Partial Fulfillment of the Requirements for the Degrees of Master of Science in Technology and Policy and Master of Science in Materials Science and Engineering

Abstract

This thesis explores the impact of relative lifecycle cost tradeoffs on technology strategy, and characterizes two factors driving these costs: population demographics, and uncertainty in component costs. The methodology developed consists of three novel components which address gaps in the current literature in the areas of large-scale network design, multi-attribute population characterization, and cost modeling. Three technologies representing near, mid, and long-term fiber-to-the-home gigabit passive optical network solutions, and seven implementation strategies are dimensioned for two significantly different population demographics, each representing large coverage regions containing millions of subscribers. The methodology is able to successfully characterize how relative network topologies changed as a function of population attributes, revealing complex cost tradeoffs between technology strategies

Thesis Supervisors:

Dr. Richard Roth

Director, Materials Systems Laboratory

Dr. Randolph Kirchain

Assistant Professor of Engineering Systems and Materials Science and Engineering

Acknowledgements

I cannot express in words my gratitude to my advisors, and my group—as they have impacted my life in ways which defy explanation. Prior to arriving at MIT, I had not enjoyed the comforts of family in many years; when I leave however, I feel that I will never be without one again. I have been privileged to learn for a living for the past four years, something very few get to enjoy for even a day. For this alone I would be grateful. However, to be surrounded by people whose curiosity is never limited by ability, and who push me to try and find that within myself, has been a humbling and extraordinary experience. You know who you are: Thank You.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

10

1 Introduction

As the demand for broadband communications continues to expand and the technologies for satisfying that demand continue to evolve, access network operators are confronted with a host of challenging questions surrounding technology choice and network deployment. Ultimately, any service provider is trying to identify a technology choice that allows them to both profitably serve their subscribers today and adapt to future market changes. Given the range of contributing issues and the rapidly evolving state of technology, it is very difficult for one firm to answer such questions alone. Currently, standards bodies are exploring a multitude of technology choices for future optical broadband network networks. Often, carriers find themselves necessarily myopic when considering which network architecture to choose. Legacy-centrism results in the construction of modeling tools which focus on a specific network design, and often, a particular architectural winner is chosen and focused on early, further obscuring the benefits a complete, general architectural comparison might provide. Understanding the cost tradeoffs resulting from technology choice free from the constraints of either existing networks, or an individual carrier's preferred architecture, requires a generalized approach characterizing the relative cost tradeoffs for a range of population/demand demographics. Next-generation gigabit passive optical network (GPON) architectures will offer not only higher bandwidths enabling more products and services, but also better quality of service, enabling more efficient and reliable networks, thereby increasing subscriber satisfaction and retention rates. These benefits will require significant upfront capital investments however, which will both "lock in" the resulting technology through standardization and component economies of scale and learning, and act as a barrier to future implementation of different technology choices. Given the decades long life cycles associated with these networks, it is important to characterize both the long-term cost implications of near-term technology choice decisions, and the long term benefits of investing in longer-term technology solutions. These issues suggest several questions for carriers and standards bodies to consider before selecting a technology:

1. How does a long-term view of network costs including operational expenditures (OpEx) impact initial technology choice decisions?

2. What are the cost implications and tradeoffs resulting from implementing GPON technologies available in the near future, versus waiting to deploy future technologies whose costs are unknown?

11

3. How should a particular technology choice be implemented to reduce total network costs?

4. How do population and data demand demographics affect technology and architecture choice?

5. How do legacy network, subscriber service penetration and discount rate impact technology choice decisions?

This work examines implementation strategies for three different GPON technology choices. These technologies were selected not only because they differ in technical specifications, but also because they present carriers with three different implementation timeframes: from technologies currently being deployed for which costs are known; to technologies which have been demonstrated in small-scale mockups but are not ready for deployment and for which costs are more uncertain; to long-term solutions with the potential to significantly reduce expensive network components, but for which cost data is either unknown or not yet available.

This thesis provides a suite of three integrated modeling tools, providing decision makers with a descriptive, rather than normative, toolset, to characterize and compare the relative lifetime network costs of each technology choice for multiple population demographics and under different technology implementation strategies. The first tool is a statistics-based population generator which parameterizes populations with different characteristics. Next, a network-modeling tool utilizes this population data and technology-specific architecture parameters to dimension large-scale network architectures. Finally, comprehensive cost models utilize cost and statistical component failure data from both manufacturers and real-world deployments to characterize how network costs change in response to technology choice decisions for a given population.

1.1 Passive Optical Network Overview

In their simplest incarnation, passive optical networks (PONs) consist of four main elements: *metro access nodes*, (MANs), which aggregate and route data for hundreds of thousands of subscribers between cities; *central offices*, (COs) which house the transmission and receiver equipment required to provide service to tens of thousands of customers; *splitters*, which enable a single transmission line from the CO to serve tens of subscribers; and the *customer premise equipment*, (CPE) which performs the optical to electronic transformation required to provide

service to individual subscribers. Linking these network components together are fiber optic cable bundles of various sizes collectively known as the fiber plant. Collectively, these elements comprise the main network sections: the *backhaul network*, consisting of all fiber bundles and equipment connecting MANs to COs and MANs to MANs, and the *local access network*, (LAN) containing the equipment and fiber links required to connect subscribers to COs and connecting COs. Figure 1 illustrates the basic "fiber-to-the-home" network topology for a architecture with two back-to-back or *cascaded* splitter stages, each enabling a single line from the CO to be split four ways. As a result, each transmission line can reach up to sixteen subscriber locations.
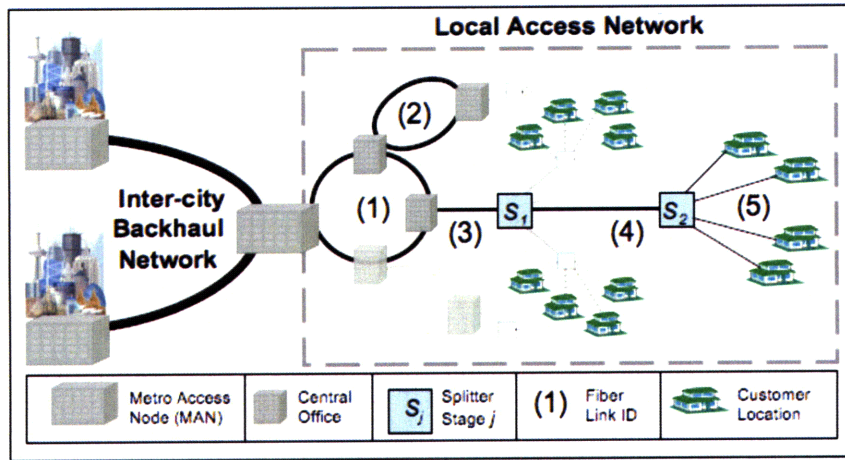
Figure 1: Basic GPON network architecture

In the literature, fiber links are classified in one of four categories depending on which network elements they connect. *Backhaul* fiber connects the MAN and COs and CO to CO; *feeder* fiber connects COs to the first splitter stage; *distribution* fiber links splitter stages; and *drop* fiber connects the final splitter stage to the subscriber, (in "fiber-to-the-home" networks) or to curbside aggregation points, which then are connected to the subscriber via copper wires, (in "fiber-to-the-curb networks"). This work focuses on the former. Table 1 classifies the relevant fiber link lengths shown in Figure 1.

Fiber Link ID	Fiber Link Description	Fiber Link
1	MAN to CO	Backhaul
2	CO to CO	Backhaul
3	CO to splitter	Feeder
4	Splitter to splitter	Distribution
5	Splitter to subscriber	Drop

Table 1: Fiber plant link descriptions

13

These networks have traditionally been called "passive" because none of the network equipment between the CO and subscriber, (splitters and fiber) requires a power source. However, recent technologies which introduce signal-boosting amplifiers in this portion of the network, (some of which will be modeled in this thesis) are also classified as passive optical networks.

In general, the maximum times a signal may be split and the total distance from the transmission equipment in the CO to a subscriber is determined by the network-specific *power budget*. This power budget is defined by subtracting the receiver sensitivity of the customer premise equipment, R_X, (a measure of the subscriber receiver's ability to detect incoming signal power) and desired safety margin, *SM*, (a built in buffer to compensate for signal power fluctuation) from the signal transmission power (T_X) (in milli-decibels[1], dBm). If the technology choice uses signal-boosting amplifiers, the signal power gain, P_{Amp}, is then added, and any related power losses due to amplifier insertion, L_{Amp}, are subtracted. Signal loss due to signal splitting, $L_{Splitter}$, scales as a loss constant, c, multiplied by the base-two logarithm of the number of splitter ports. For example, a *1:4* way splitter will result in a signal power loss given by, $L_{Splitter} = c \cdot \log_2(4) = 2 \cdot c$, while a *1:16* way splitter result in a loss of $L_{Splitter} = c \cdot \log_2(16) = 4 \cdot c$.

As a result, a network with a given power budget can *either* have a longer reach at a lower splitter port count *or* have a shorter reach but a higher splitter ports count. This tradeoff suggests that *multiple* splitter ratio/network reach *implementation strategies* are possible for a given technology choice[2].

To define network reach, (in *km*) the remaining power budget is divided by the power loss per fiber kilometer, (L_{Fiber}, an intrinsic fiber property). Symbolically, the reach determination relationship is given by:

$$\left[\frac{T_X - R_X - SM + P_{Amp} - L_{Amp} - L_{Splitters}}{L_{Fiber}} \right] \left(\frac{dBm}{dB/km} \right) = R \text{ (km)} \qquad (0.1)$$

All subscribers served by a single feeder fiber must share not only the signal transmission power of a given optical line, but also the available data rate. The maximum transmission rate for a given optical line terminal (OLT) port, (the laser and modulator combination creating and shaping the signal) and the method for combining multiple subscriber data onto a single line, (the

---

[1] A logarithmic ratio which measures signal power magnitude relative to a reference level of 1 milliwatt, defined as dBm = $10\log_{10}(P_1/P_o)$, where P_1 is the transmitter signal power and P_0 = 1 milliwatt.

[2] Several such strategies are modeled in this thesis, and are defined in greater detail in subsequent chapters

"multiplexing" strategy) are specific to technology choice. All technologies modeled in this work utilize Statistical Time Division Multiplexing, (STDM) to allocate subscriber bandwidth. STDM divides a single data stream into discrete "packets" containing both an individual subscriber's data, and the corresponding optical network terminal (OLT) address. Packets are then assigned dynamically on an as needed basis. This dynamic bandwidth allocation enables both "burst-mode" transmission, whereby an individual subscriber may be assigned large fractions of the total data stream, and "statistical gain," which effectively reduces the data burden of each subscriber under the assumption that all subscribers will not simultaneously request their maximum subscription service[3].

1.2 Overview of Technology Choices Modeled

Three fiber-to-the-home GPON technologies were selected which represent a spectrum of current and future networks. Because power budget allocation enables multiple implementation strategies for each technology (see above discussion), several such strategies were modeled.

A. "Standard" GPON. This is the least complex GPON technology choice, in that all components outside the central transmission office are completely "passive": using only chromatic splitters and without signal amplification. As a result, this choice exhibits the smallest total signal power budget, limiting both split ratio (typically *1x32* way total splitter ports) and CO to subscriber network reach (typically *20-25km*). Because this technology employs many components already in production, and is governed by an established standard, 2003 IEEE ITU-T G.984.1 GPON, it represents the nearest-term GPON solution. Therefore, it is used as a benchmark against which to value the other technologies considered.

B. Amplified GPON. This technology utilizes remote-powered semiconductor optical amplifiers (SOAs) outside the central office to increase the available power budget, enabling either longer reach (up to the theoretical *60km* distance limit of the GPON standard (Lin 2006)) or higher split (up to *1x256* total port count) architectures. The increase in reach and/or customers per transmission line due to amplification enables multiple reach/split ratio architecture combinations, some of which may reduce network

---

[3] One drawback to this method is that simultaneous maximum data demands may reduce the data available for all subscribers to levels below the agreed upon subscription rates

15

CapEx and OpEx by reducing the number of central offices required to service the coverage region. Early working models of this technology have been demonstrated (Nesset, Payne et al. 2006), and the additional network elements required have been thoroughly researched and are beginning to be manufactured, albeit not at the scale required to bring cost on par with the "standard GPON" components described above. Therefore, this technology choice represents our "mid-term" future solution.

C. <u>Long-Range GPON</u> utilizes high-gain erbium-doped fiber amplifiers (EDFAs) to significantly extend both the reach (up to *100km*) and split ratio (up to *1x1024* total ports). The extended reach enables the elimination of central offices completely, instead connecting all customers directly to the metro access node, (MAN) which traditionally serves as a gateway between local access networks (LANs) for the other technologies (connecting city to city for example). This results in the location of all transmission equipment at MAN locations, and eliminates the need for backhaul-related equipment which would otherwise be required to connect central offices in the LAN to the MANs servicing the coverage region. This technology is currently the furthest away for implementation, and is not currently covered by the existing GPON standard. Additionally, the component costs for the additional network elements required to enable this technology are currently unknown. Therefore, this choice is modeled as a "long-term" future GPON solution.

It is important to note that the technologies selected for this analysis do not constitute an exhaustive set of strategies under consideration, but are intended to be representative of the spectrum of Statistical Time Division Multiplexed (STDM) options currently being explored by carriers. Table 2 provides an overview of the implementation strategies modeled for each technology choice, for example, *"A1"* corresponds to the first implementation strategy for technology choice "A" ("standard" GPON). The assumed advantages and disadvantages of each strategy, along with more in depth technical parameters will be provided in Chapter 10.

Implementation Strategy	A1	A2	B1	B2	B3	C1	C2
Downstream Data Rate	2.5 Gbps						
Multiplexing Strategy	Statistical Time Division Multiplexing (STDM)						
Splitter Configuration	Centralized	Cascaded					
Total Splitter Port Count	1x32	1x32	1x128	1x256	1x512	1x512	1x1024
Total Network Reach (km)	25	25	60	53	46	100	88
Amplification Strategy	N/A	N/A	SOA	SOA	SOA	EDFA	EDFA

Table 2: Architecture implementation strategies by GPON technology choice

1.3 Thesis Outline

Chapter 2 presents an overview of current network modeling approaches, and discusses the gaps the thesis intends to address. Chapter 3 presents the expected research contributions, and Chapter 4 provides a brief overview of the modeling methodology used to achieve these goals. Chapter 5 introduces the population generator, which characterizes important population attributes based on input probability distributions. Chapter 6 describes the network model including inputs and the heuristic-based modeling algorithms used to dimension a virtual network, and Chapter 7 focuses on the heuristic calibration process and sensitivity analysis to the final heuristic value set. Chapter 8 explains both the CapEx and OpEx cost models and component cost data collection, while Chapter 9 performs benchmarking and validation exercises for the network and cost models, and then discusses the methodology capabilities and limitations. Chapter 10 defines the populations, architecture parameter values, methodologies, results, analyses and result sensitivities for the cases modeled. Finally, Chapter 11 provides the analysis conclusions and recommendations to firms, and Chapter 1 suggests future work.

2 Current Modeling Approaches

Modeling comparisons of network costs due to technology choices requires three things: a way to fully characterize the important aspects of a subscriber population, a network dimensioning tool which accurately reflects how fiber plant and component requirements change as a function

of technology choice and these population characteristics, and cost models which accurately characterize network *lifetime* costs, both capital and operational expenses, corresponding to each dimensioned network. This chapter examines different existing approaches to these three requirements, identifies gaps, and discusses how the MSL modeling approach addresses these gaps.

2.1 Network Modeling Approaches

Network design models must perform several functions, including siting central offices (COs), splitters, and amplifiers, and defining the fiber route topology connecting all these network components to both each other and subscribers. Additionally, many factors contribute to shape each function's output, including topological/geographic constraints of the connected region; existing/legacy facilities and/or fiber routes; and component/fiber performance, availability, cost, and capacity.

The literature reveals two classes of approaches carriers use when dimensioning optical broadband networks. The first, which we label "bottom-up," incorporates population and data demand information on every potential subscriber in the coverage region and problem-specific constraints to dimension cost-minimizing networks through optimization. The second approach, which we label "top-down," utilizes engineering rules of thumb developed over time in the telecommunications industry to dimension larger populations or specific population sub-types, (a square grid of urban city blocks for example).

2.1.1 Bottom-Up Optimization Approaches

Much work has been done to address the difficult optimization problems required to minimize network costs subject to these many constraints. Telecommunications network design problems involve multiple layers to be optimized simultaneously. For example, the physical layer concerns the geographic network topology, while the traffic layer concerns the flow of information through the physical infrastructure and the corresponding link data capacity required to meet customer demand at all times. These complex optimization problems all attempt to minimize (or potentially maximize depending on criteria) an objective function, typically total network costs, subject to constraints on link capacity, while connecting all transmission data sources to subscriber data sinks in the context of *a specific* technology choice. Problems of this type have two "flavors". The first involves siting data concentrators, (central offices and splitter sites),

18

given subscriber locations, subject to satisfying data demands, (see (Gourdin, Labbe et al. 2002). These *concentrator location* or *capacitated facility location problems* often arise in other areas such as transportation, and produce tree topologies. These problems are often modeled using mixed integer programming models and various heuristic-based constraint relaxation techniques for both tree network (concentrator to subscriber point to point connections) (Balakrishnan, Magnanti et al. 1991; Ahuja, Magnanti et al. 1993; Balakrishnan, Magnanti et al. 1995; Daskin 1995; Klincewicz 1998), and ring structures, (Klincewicz 1998; Ramaswami and Sivarajan 2002). The second problem variety assumes that concentrator and subscriber locations are known. The *Multi-capacity flow problem* takes these locations, data sources (nodes) and the rules about the allowable connections between them (links), and routes traffic from individual sinks to concentrator points and vice-versa. Such problems have also been modeled extensively using graph theoretic, and linear programming and multiple other approaches for both tree networks, (see (Boorstyn and Frank 1977; Balakrishnan, Magnanti et al. 1991; Gavish 1992; Ahuja, Magnanti et al. 1993) and ring network design, (LeBlanc, Park et al. 1996; Sanso and Soriano 1999; Sridhar, Park et al. 2000). See (Nagy and Salhi 2007) for a thorough survey of algorithmic methods and approaches throughout a variety of fields.

Often, problems of this type require decomposition into different components which are then individually solved to optimality when possible, or through heuristic approximation and constraint relaxation (Carpenter and Luss 2006). While exact solutions exist for relatively small problems, (on the order of 50-100 nodes, see (Gabral, Knippel et al. 1999; Minoux 2001), to optimize over the potentially millions of customers, multiple central offices, and hundreds of splitter sites required to serve a large coverage region becomes computationally intractable, as computational time grows rapidly as the number of nodes and edge possibilities increases[4], (Gabral, Knippel et al. 1999; Melkote and Daskin 2001; Randazzo and Luna 2001; Klose and Gortz 2006; Lee, Kim et al. 2006; Prins, Prodhin et al. 2007; Li, Chu et al. 2009).

2.1.2 Top-Down Engineering-Rule-Based Approaches

In contrast to the mathematical optimization modeling approaches described above, these methods utilize the experience of network designers and widely recognized engineering rules of thumb to examine the relative cost impacts of technology choice in networks serving small

---

[4] In general, these problems are "NP-hard," defined a class of problems for which no solution can be found using an algorithm in polynomial time. As a result, the problem scales exponentially with the number of nodes.

populations or specific population types. These models often characterize populations using information drawn from experience with previous network deployments or actual populations currently serviced. For example, a 2003 Bell Labs study (Weldon and Zane 2003) modeled several technology choices serving a small population with uniform population density occupying a square grid resembling a cluster of urban city blocks. This grid structure enables the derivation of simple fiber-routing rules, which are then used to dimension the fiber to the home network topology. The costs of different technologies are then compared for this topology at different penetration (subscriber uptake) levels. By contrast, a 2004 Corning study, (Vaughn, Kozischek et al. 2004) models the cost implications of technology choice and implementation strategies for a network serving up to *70,000* subscribers. This model breaks down the population into discrete categories by blocks of distance from three pre-located central offices. All subscribers in each category are then assigned a characteristic average fiber length, and these lengths are summed to arrive at an estimate of the total fiber required to provide data services. For example, all customers at a distance of three kilometers or less from a central office are assigned a feeder fiber length of *0.59km*. Different penetration rates are then modeled by adding or subtracting these average values and the corresponding customer premise equipment.

2.1.3 *Network Modeling Approach Comparison and Gap Analysis*

The "bottom-up" optimization and "top-down" engineering rule of thumb approaches bound the spectrum of network dimensioning tools: the former requires extensive collection of often proprietary data and computational power, but potentially provides mathematically optimal topologies for small populations, while the latter takes advantage of engineering expertise and experience and requires relatively simple computational tools, but may miss important details impacting network design by limiting analysis to population subtypes and assigning coarse fiber plant characteristics to large fractions of the population.

A method incorporating the optimization-focus of the bottom up approach with the computational tractability of the top-down approach would enable characterization of how technology choice and population demographics drive *relative* topology changes between networks. Instead of exhaustively optimizing a single network and/or technology, this method could enable large populations to be modeled for multiple technologies. Such a model could capitalize on the strengths of both optimization and engineering-rules based approaches, but

avoid the computational complexity of the former while more accurately characterizing how population and technology attributes drive network topology than the latter.

2.2 Population Characterization Approaches

Population demographics can impact network cost at least as much as technology choice (Sirbu and Banerjee 2005). Therefore, it is important to accurately characterize these relationships. For example, intuitively high subscriber population and data demand densities translates into large quantities of central offices, transmission optical line terminals (OLTs), splitter sites, and amplifiers than those required to serve sparser regions and/or those with smaller data demands. Not as immediately obvious however is the impact on fiber length of changes in population density. For example, in urban areas, the fiber length from the curb to a subscriber location is significantly less than the same link length required to span a front yard in a suburban neighborhood, or the often-significant acreage from road to urban homes. Reinforcing this effect is that a single curb to location trenched fiber can often reach multiple subscribers sharing apartment buildings in dense urban areas, while, in suburban areas there are many more single-family homes, each of which requires an individual fiber. Add to this the fact that each subscriber at a given location can require different data service tiers, potentially resulting in fewer customers served per splitter, as the maximum transmission rate per individual service line is limited. For large coverage regions with hundreds of thousands or millions of homes passed, these differences can translate into thousands of kilometers of installed fiber length—and millions of dollars in capital and operational expenses. Therefore, it is important to accurately characterize how population density, household density, and data demand density characteristics change throughout the subscriber population.

There are two population-modeling approaches in the telecommunications literature, corresponding to the two network modeling approaches outlined above: explicitly modeling every home in the coverage region, and assigning average or uniform values to population characteristics. Inherent in the "bottom-up" approach to network modeling is an assumption that every subscriber location in known. These "nodes" form the basis of graph-theoretic approaches to the *multi-capacity flow problems*, (Balakrishnan, Magnanti et al. 1991; Ahuja, Magnanti et al. 1993; Balakrishnan, Magnanti et al. 1995; Daskin 1995; Klincewicz 1998). and serve as the data sinks requiring splitter and central office siting in concentrator location problems (Boorstyn and

Frank 1977; Balakrishnan, Magnanti et al. 1991; Gavish 1992; Ahuja, Magnanti et al. 1993) (LeBlanc, Park et al. 1996; Sanso and Soriano 1999; Sridhar, Park et al. 2000) (Nagy and Salhi 2007). This makes sense, as optimization focuses on meeting the demands of a specific population given a technology choice. By contrast the "top-down" dimensioning approaches relax the constraint of explicitly characterizing every subscriber in favor of coarser metrics such as average or uniform densities, (Weldon and Zane 2003; Vaughn, Kozischek et al. 2004; Joao 2007). This also makes sense, given that these studies focus on characterizing the cost implications of different technology choices given a simple population or population subtype.

2.2.1 Population Characterization Gap Analysis

While explicitly modeling every customer location in the coverage region may ensure complete population characterization, it introduces three complications: (a) it is *data intensive*, requiring the physical location, household size, and data demands of every subscriber; (b) these large data sets require *computationally complex* optimization algorithms whose solution time scales with problem size; and (c) it is *difficult to generalize*, as the resulting resource-intensive algorithm "overfits" to satisfy the specified population. Conversely, coarse characterizations of population attributes, while enabling the relatively computationally simple dimensioning algorithms to dimension, may overlook important subtle yet important cost drivers. Between these two approaches exists an opportunity for a method which balances the detail required to characterize how these important attributes change across a population, against the abstraction required to easily model multiple technologies. Additionally, characterizing populations free from proprietary carrier data may help facilitate more detailed discussion about technology choice.

2.3 Cost Modeling Approaches

Characterizing lifetime network costs requires examining not only large capital investments, but also recurring operational costs. Traditionally, cost models have focused on the CapEx associated with technology choice, (Weldon and Zane 2003; Vaughn, Kozischek et al. 2004; Joao 2007). However, this approach obscures any potential OpEx savings associated with a particular technology choice. Often, when OpEx is considered, it is treated as a simple mark-up on network CapEx (Konrad 2007). This approach suggests that multiple technologies with similar CapEx should be considered on equal footing from an implementation standpoint; however, studies have shown that fiber to the home optical networks can provide substantial

22

OpEx savings over existing copper networks (Halpern, Garceau et al. 2004; Wagner, Igel et al. 2006). As a result, recent studies have begun to more thoroughly examine network OpEx and classify the important cost drivers (Verbrugge, Pasqualini et al. 2005). These studies either try to find ways to optimize the repair and maintenance infrastructure of existing networks, (Casier, Verbrugge et al. 2007) or focus on characterizing service-based OpEx, (Pasqualini, Kirstadter et al. 2005; Konrad 2007; Machuca, Moe et al. 2007; Prieß and Jacobs 2007; Vukovic 2007; Vusirikala and Melle 2007). Broadly, these methods comprise a set of "bottom-up" approaches, (Konrad 2007; Prieß and Jacobs 2007) which seek to characterize OpEx as the sum of network elements necessary to meet a customer demand scenario, rather than "top-down" methods which assign total network costs to network components.

2.3.1 *Cost Modeling Gap Analysis*

While many studies have done an excellent job of characterizing network CapEx, few studies have integrated CapEx with detailed network OpEx to compare lifetime network costs across technologies and/or population demand scenarios. For example, top-down methods rely on proprietary cost information for a specific, deployed network, while bottom-up approaches allocate costs to products and services for existing or mathematically optimized future networks. As a result, both methods limit the generalizability of results, the former through lack of transparency and access to data, and the latter via over-optimization or undervaluation of dynamic OpEx drivers such as maintenance (Konrad 2007; Prieß and Jacobs 2007).

3 Expected Contributions

This thesis addresses the role of OpEx in initial technology decisions for several FTTx GPON technologies, each with multiple possible implementation strategies. The methodology employed utilizes three, novel models integrated into a single methodological modeling framework: a population generator and characterization tool, a heuristic-based network dimensioning model, and a capital and statistical operational cost model. This method is then used to gain case-specific insights into parameters which drive OpEx, including population demographics and cost uncertainty of future technologies.

The first model component characterizes subscriber populations in multiple dimensions which constrain and alter network topology: location density, household density, and data demand

density. This ability to specify detailed population characteristics for multiple populations enables exploration of how these characteristics affect network costs, and the ability to model multiple populations and/or large geographic regions. The model *samples* from three, independent input population distributions, (one for each density parameter) and then uses conditional probabilities to determine how these densities change as a function of distance from the population center. Sampling reduces the data required to characterize a population while maintaining important population attributes. Additionally, modeled populations provide generic, and therefore generalizable, characteristics free from proprietary data constraints.

The second model component utilizes this data set and three heuristics to dimension a virtual network. Because fiber installation is the primary driver for network CapEx, (Wagner, Igel et al. 2006) and fiber-related outages drive OpEx labor costs, (Rand-Nash, Roth et al. 2007) the model emphasizes fiber minimization via maximum co-location. Rather than seeking a mathematically optimized solution, the model utilizes constraints based on the three population parameters defined above to determine how technology choices affect relative network topology via fiber plant and central office siting. Additionally, because sampling reduces the data intensity needed to characterize the population, the network model requires less computational power, enabling rapid modeling of multiple technology choices.

Finally, the cost models model incorporate real-world installation, failure and operations data with industry component cost and failure data to characterize how technology choice and population demographics impact lifetime network costs. By gathering cost data from multiple sources throughout the telecommunications industry, the model is able to escape proprietary point cost estimates, thereby increasing transparency and generalizeability, and assign a complete set recurring costs including scheduled maintenance and normal operations costs and probabilistic repair and replacement costs to the dimensioned network.

4 Methodological Overview

The MIT Materials Systems Laboratory (MSL) Population Generator, Network Design and Cost Models determine network lifecycle costs, and how these costs change in response to changes in assumed technological (e.g. transmitter power), demand (e.g. subscriber distribution), and operational characteristics (e.g. coverage area). Figure 1 provides an overview of the modeling approach.
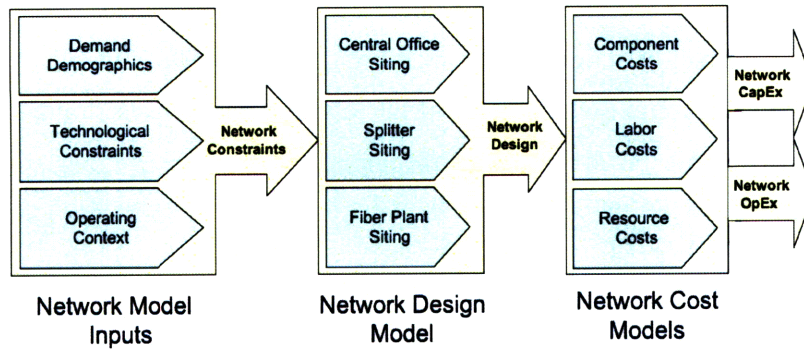
Figure 2: Modeling Overview

The demand demographic, technological characteristics, and operating context comprise the initial set of input parameters, thereby constraining the solution space of possible network designs. The population generator characterizes the subscriber demographics. The *design model* then dimensions a virtual network, defines the necessary hardware quantities, and populates the network with specific, user specified hardware/component choices. Finally, *the network cost models* map initial capital investments and recurring operational costs corresponding to each resulting network.

5 MSL Population Characterization

The MSL population model attempts to attain the balance of abstraction and detail which captures the important relationships between population characteristics and network topology while using statistics and sampling to reduce data requirements and the resulting computational burdens. The model characterizes three important population attributes, using three, independent and user defined probability distributions which together form the set of *demand demographics* for each population. The first attribute, *population density*, is determined by *sampling* from an input probability distribution. The distribution parameters then determine how this density changes as a function of distance from the population center. The reduced data intensity enabled by population sampling, while smaller than individually characterizing every subscriber, is still quite large by optimization standards, (see §2.1 for a discussion of this issue). Therefore, the model focuses on simplifying the dimensioning process by breaking the continuous coverage region into discrete grids. The second characteristic, *household size density*, is determined by first defining the types of household/business sizes available, and then determining the percentages of the population with each size, and how these household sizes are distributed as a

25

function of distance from the population center. Similarly, the final attribute, *data demand density*, is characterized by first defining the quantity and data rates of individual service tiers to be offered, and then determining the fraction of the total population subscribing to each tier, and how this tier density is distributed throughout the population.

5.1 Population Density

The first distribution utilizes the random variable, Γ_i^p, the distance from the distribution center to a subscriber location, with probability density function (pdf), $f_\Gamma^p(r)$, to site customer locations for each population, p. As an example, consider Figure 3, which depicts a coverage region with three population centers, ($p = 1,2,3$) all normally distributed but with differing standard deviations (which serve to either spread or cluster each distribution). In this specific example the standard deviations are small, on the order of 3-4 km, resulting in tight distributions (small cities). All parameter values describing each distribution, including the mean (which determines where the center of the distribution is geographically), and standard deviation are user-defined. The total coverage region shown is *100x100 km²*.
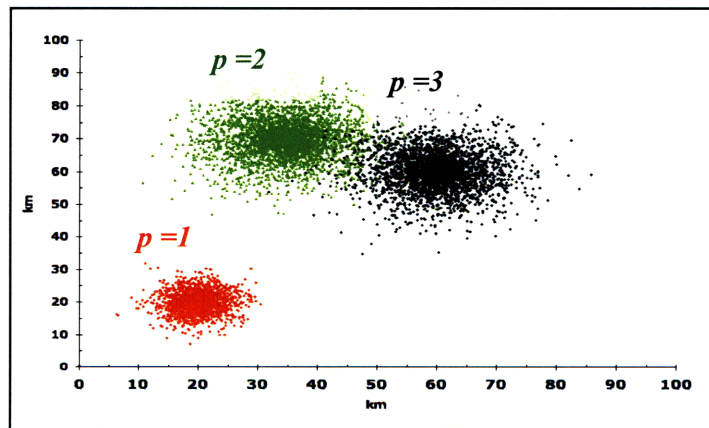
Figure 3: Sample 100km² coverage region with three population distributions

Each colored "dot" in Figure 3 does not represent a single subscriber location, but rather a statistical object representing a subscriber *neighborhood*. The neighborhoods per population, d_p,

is a user-defined input parameter[5], as is the total locations per population I_p. Together, these two parameters define the locations per neighborhood for each population, I_d^p. Table 3 presents the three $f_\Gamma^p(r)$ related parameters and values used in Figure 3.

Parameter	Definition	Value used in Figure 3
P	Total populations	3
p	Individual population ID	1; 2; 3
$T_f\left(\mu_p, \sigma_p^2\right)$	Type, mean and variance: population p	N(0,9); N(0,16); N(0,16)
I_p	Locations per population	24k; 48k; 72k
d_p	Total neighborhoods per population	2,000; 4,000; 4,000
I_d^p	Locations per neighborhood for population p	12; 12; 24

Table 3: Parameters and values of $f_\Gamma^p(r)$ in Figure 3

5.2 Population Density Modeling

Because the number of subscribers (nodes in the optimization literature, see §2.1) directly relates to the computational resources required to dimension the network, the MSL population model attempts to limit these points by overlaying a search grid onto the coverage region. The grid structure, the size each cell and the total number of cells, defines the level of detail the model can resolve. For example, Figure 4 depicts a square grid overlaid on (a) the entire population and (b) a population subset of $p=1$ in Figure 3. The grid points are uniformly spaced one kilometer apart in both the horizontal, ("x") and vertical, ("y") directions, resulting in an individual *grid point resolution* of *1 km²*.

---

[5] $\sum_{p=1}^{P} d_p \leq 10,000$

(a)	**(b)**
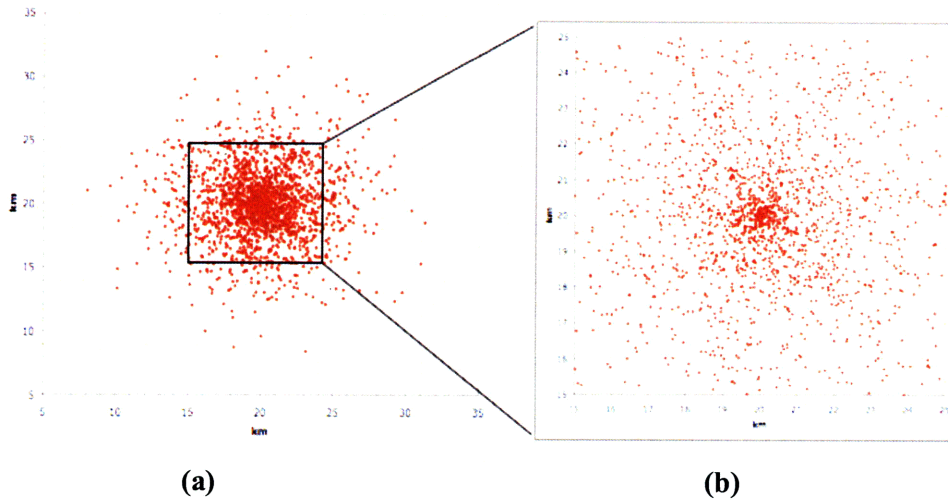
Figure 4: Square grid cells of area 1 km^2 superimposed on $p=1$ in Figure 3

As Figure 4(a)(b) illustrates, this grid structure results in a significant number of empty grids in sparse regions, and grids containing many neighborhoods in dense regions. The former results in an inefficient search, as computational resources are spent checking empty grids, while the latter results in poor resolution in much of the region, potentially obscuring important differences between individual neighborhoods. Characterizing populations requires resolution levels ranging from city blocks in dense regions, to square kilometers in rural areas. As an example, the average Manhattan city block is on the order of 0.02km^2, and contains roughly 40 locations (apartment buildings for example), while the same number of farming communities in some rural areas occupy 1km^2. Using the *1km^2* grid points shown in Figure 4, modeling the farming community would require *900, (30^2)*, total grid points, while the same resolution in Manhattan would require over *2 million ([30*50]^2)* such grid points! An effective search grid should therefore try to minimize the fraction of empty grid points while also isolating individual neighborhoods, providing the right level of abstraction to reduce computational complexity, while maintaining enough spatial resolution to fully characterize the neighborhood population.

5.2.1 Dynamic Grids

Rather than the square grid, which assigns points at *equidistant* intervals, the MSL model assigns grid points to *equiprobable* intervals of the neighborhood distribution, $f_\Gamma^p(r)$. Therefore, the grid resolution changes dynamically in response to the population density, with each grid cell *changing its size* as a function of distance from the population center. This more effective use of

resources reduces both the number of grid points required to characterize the neighborhood population and the number of empty grid cells. Characterizing the resulting grid structure requires being able to calculate three things: the grid cell size, cell spacing, and the location of each cell in the coverage region. Additionally grid cell size is constrained by location density: there are only so many locations which can be packed onto a city block, and there exists an average acreage per farm in more rural areas. Therefore, a minimum and maximum grid cell size must also be established to incorporate these constraints. The first of four network design algorithms, $\mathcal{A}_1$, performs these functions.

5.2.2 Calculating Probability Functions

Recall that within a given population, each *neighborhood* consists of a *fixed* number of locations, l_d^p, (although each location can serve multiple households and data demands). The grid cells change size such that the probability "contained" within the cell is the same for all cells. Therefore, *decreases* in neighborhood density correspond to *increases* in the neighborhood size required to service the same number of locations, and a corresponding increase in grid cell size. As a result, the algorithm *changes the grid resolution*, transitioning from many small cells in dense urban centers to fewer, larger cells in suburban neighborhoods to very few, very large cells in sparse rural areas at large distances from the city center.

5.2.3 Grid Cell Size

We model the neighborhood area, N_d^p, as a square, with side lengths equal to the average linear distance between two neighborhoods on a given ring at r (in km). To determine this distance, we begin by approximating as constant the probability in a thin annular ring of thickness dr:

$$\Pr(ring) = \left[F_\Gamma^p(r + dr) - F_\Gamma^p(r) \right] = \left[\int_r^{r+dr} f_\Gamma^p(r) \cdot dr \right] \approx \left[f_\Gamma^p(r) \cdot dr \right] \qquad (0.2)$$

The total possible neighborhoods in this ring, d_r^p is this probability multiplied by the total neighborhoods in the population, d_p:

$$d_r^p = d_p \cdot f_\Gamma^p(r) \cdot dr \quad \left({neighborhoods}/{ring} \right) \qquad (0.3)$$

29

These distributions exhibit radial symmetry; therefore, neighborhood density in each ring is *uniformly* distributed. As a result, the probability associated with a *single* neighborhood at distance r from the population center is given by:

$$\Pr(neighborhood) = \frac{1}{d_r^p} = \frac{1}{d_p \bullet f_\Gamma^p(r) \bullet dr} \qquad (0.4)$$

5.2.4 Grid Cell Spacing

We approximate the distance between two neighborhoods on a given ring as the thickness, dr, for which every neighborhood has this probability. For a ring of length $2\pi r$, this distance is given by:

$$dr \simeq \frac{2\pi r}{d_p \bullet f_\Gamma^p(r)} \quad (km) \qquad (0.5)$$

Therefore, the neighborhood size for *every* neighborhood at distance r from the population center, is given by:

$$N_d^p(r) = (dr^2) \simeq \left(\frac{2\pi r}{d_p \bullet f_\Gamma^p(r)} \right)^2 \quad (km^2) \qquad (0.6)$$

Recall that *every* $N_d^p(r)$, because it represents a single neighborhood, contains l_d^p locations. Therefore, the physical interpretation of dr and $N_d^{p=1}(r)$ for population $p=1$ in Table 5 is the total length and corresponding area required to contain $l_d^{p=1}=12$ locations[6]. This length can vary significantly depending on city size and population density.

5.2.5 Grid Cell Size, Spacing, and Density Constraints

Cell grid size is limited by the physical realities of location density, for example, there are only so many locations which can fit in on a city block, and there is an average acreage for land in rural areas. Using dense urban and sparse rural areas in the U.S., we can establish some reasonable minimum and maximum dr and $N_d^p(r)$ values reflecting these real-world constraints. To find the *minimum* neighborhood size capable of containing l_d^p locations, we model the *most populous* area in the U.S.—New York City. The average Manhattan city block size, N^{NYC}, is

---

[6] This figure was chosen to make the following example and mathematics easier to follow. This parameter is examined in greater detail including sensitivity analysis in Chapter 7

approximately 80m x 274m $\approx$ 0.02 km², and contains, on average, *forty* locations/buildings. The locations per neighborhood for a given population can be less than, equal to, or greater than forty however. Therefore, the minimum neighborhood size for a given population, N_{MIN}^p, will be the fraction, (if $l_d^p < 40$) or multiple, (if $l_d^p \geq 40$) of the minimum block size required to contain l_d^p locations.

In general, for a population p with locations per neighborhood, l_d^p, this minimum neighborhood area and length/width, dr_{MIN}^p, are given by:

$$N_{MIN}^p(l_d^p) \simeq \left(N^{NYC}\right) \cdot \left(\frac{locations\ per\ neighborhood}{locations\ per\ New\ York\ City\ block}\right) = (0.02km^2) \cdot \left(\frac{l_d^p}{40}\right)$$

$$dr_{MIN}^p = \sqrt{N_{MIN}^p(l_d^p)}$$

(0.7)

For example, for population $p=1$ in Figure 3, $l_d^{p=1} = 12 < 40$; therefore, we expect $N_{MIN}^{p=1} < N^{NYC}$:

$$N_{MIN}^{p=1}(12) = (0.02km^2) \cdot \left(\frac{12}{40}\right) = 0.006km^2$$

$$\therefore dr_{MIN}^{p=1} = \sqrt{N_{MIN}^{p=1}} = 0.075km$$

(0.8)

All neighborhoods within a distance $r \leq r_{Min}^p$ will be assigned these minimum neighborhood size dimensions. The r_{Min}^p boundary is defined as the point where $dr = dr_{MIN}^p$. Using (0.5) and (0.7), this point is given by:

$$dr = dr_{Min}^p$$

$$\left[\frac{2\pi r}{f_\Gamma^p(r) \cdot d_p}\right] = \sqrt{\frac{0.02 \cdot l_d^p}{40}}$$

$$\left[\left(\frac{d_p}{2\pi}\right)\left(\sqrt{\frac{0.02 \cdot l_d^p}{40}}\right)\right] \cdot f_\Gamma^p(r) = r$$

(0.9)

The crossover point of this transcendental equation defines r_{Min}^p.

We model the *maximum* neighborhood size capable of containing l_d^p locations by looking at the twenty *least populated* U.S states[7]. The average population density for these states is ~12 locations per km[2], resulting in an average lot size of $N^{20} = 0.083\text{km}^2$ (20.6 acres) per location. In general, the maximum resulting neighborhood size, N_{MAX}^p, and length/width, dr_{MAX}^p, again as a function of l_d^p, are given by:

$$N_{MAX}^p(l_d^p) \simeq \left(\frac{locations}{drop} \right) \cdot \left(\frac{km^2}{location} \right) = (l_d^p) \cdot (N^{20})$$

$$dr_{MAX}^p = \sqrt{N_d^{MAX}(l_d^p)}$$

(0.10)

For population *p=1* in Figure 3 these values are:

$$N_{MAX}^{p=1}(12) = (12) \cdot (0.0833km^2) = 1km^2$$

$$dr_{MAX}^{p=1} = \sqrt{N_{MAX}^{p=1}(12)} = 1km$$

(0.11)

All neighborhoods within a distance $r \geq r_{Max}^p$ will be assigned these neighborhood dimensions. The boundary, r_{Max}^p, is defined as the point where $dr = dr_{Max}^p$. Using (0.5) and (0.10), this is given by:

$$dr = dr_{Max}^p$$

$$\left[\frac{2\pi r}{f_\Gamma^p(r) \cdot d_p} \right] = \sqrt{l_d^p \cdot N^{20}}$$

(0.12)

$$\left[\left(\sqrt{l_d^p \cdot N^{20}} \right) \cdot \left(\frac{d_p}{2\pi} \right) \right] \cdot f_\Gamma^p(r) = r$$

The crossover point again defines the boundary, here r_{Max}^p. Figure 5 illustrates how $r_{Min}^{p=1}$ and $r_{Max}^{p=1}$ are determined for population *p=1* in Figure 3 with distribution parameters $f_\Gamma^{p=1}(r) \sim N(0,9)$, $d_{p=1} = 2000$ total neighborhoods, and $l_d^{p=1} = 12$ locations per neighborhood, (Figure 2).

---

[7] In decreasing order of population density: Vermont, Minnesota, Mississippi, Arizona, Arkansas, Iowa, Oklahoma, Colorado, Maine, Oregon, Kansas, Utah, Nevada, Nebraska, Idaho, New Mexico, South Dakota, North Dakota, Montana, Wyoming, and Alaska.
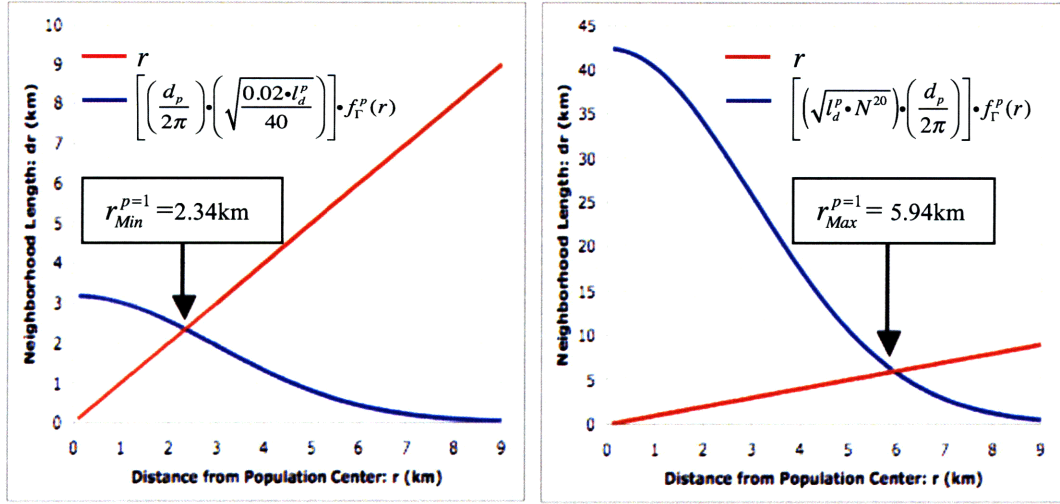
Figure 5: Graphical determination of r^p_{Min} and r^p_{Max} for $p=1$ in Figure 3

Additionally, for each population the N^p_{MIN}, N^p_{MAX}, and l^p_d parameters define the minimum and maximum *location* densities[8] in these regions (in locations per km²). The minimum density, ρ^p_{MIN}, for the region $r \geq r^p_{MAX}$ is given by, (where $N^{20} = 1km^2$):

$$\rho^p_{MAX} = \frac{l^p_d}{N^p_{MIN}} = \frac{l^p_d}{(l^p_d \cdot N^{20})^2} = \frac{1}{l^p_d} \quad \left(\forall r \leq r^p_{MIN} \right) \tag{0.13}$$

which is a function of the locations per neighborhood. The maximum density, ρ^p_{MAX}, for neighborhoods within the region $r \leq r^p_{MIN}$, is given by:

$$\rho^p_{MIN} = \frac{l^p_d}{N^p_{MAX}} = \frac{l^p_d}{\left(0.02 \cdot l^p_d \Big/ 40 \right)} = \frac{40}{0.02} = 2000 \quad \left(\forall r \geq r^p_{MAX} \right) \tag{0.14}$$

which is *independent* of the neighborhoods per population, and defines the absolute maximum location density modeled. Table 4 presents the neighborhood side length and size, and location density as a function of distance from the population center for a generalized population *p*.

---

[8] This defines the total buildings or homes per square kilometer. This is distinct from the household density, which can be significantly higher

Region (km)	dr (km)	$N_d^p(r)$ (km^2)	$\rho^p \left(locations/km^2 \right)$
$r \le r_{Min}^p$	$\sqrt{\dfrac{0.02 \cdot l_d^p}{40}}$	$\left(\dfrac{0.02 \cdot l_d^p}{40} \right)$	$\dfrac{l_d^p}{\left(0.02 \cdot l_d^p \Big/ 40 \right)} = 2000$
$r_{Min}^p < r < r_{Max}^p$	$\left(\dfrac{2\pi r}{d_p \cdot f_\Gamma^p(r)} \right)$	$\left(\dfrac{2\pi r}{d_p \cdot f_\Gamma^p(r)} \right)^2$	$\dfrac{l_d^p}{\left(\dfrac{2\pi r}{d_p \cdot f_\Gamma^p(r)} \right)^2}$
$r \ge r_{Max}^p$	$\left(l_d^p \cdot N^{20} \right)$	$\left(l_d^p \cdot N^{20} \right)^2$	$\dfrac{l_d^p}{\left(l_d^p \cdot N^{20} \right)^2} = \dfrac{1}{l_d^p}$

Table 4: Neighborhood dimensions and location density as a function of distance r

For example, if we choose three different distances from the population center for population p $=1$ in Figure 3: $r_1 =2km$, $r_2 =4km$, and $r_3 =6km$, then Table 5 provides the neighborhood dimensions and location densities at these distances.

Distance r (km)	Region	dr (km)	$N_d^{p=1}(r)$ (km^2)	$\rho^{p=1} \left(\dfrac{loc}{km^2} \right)$
2	$r \le r_{Min}^{p=1}$	0.075	0.006	2000
4	$r_{Min}^{p=1} < r < r_{Max}^{p=1}$	0.23	0.05	227
6	$r \ge r_{Max}^{p=1}$	1	1	0.083

Table 5: dr, $N_d^{p=1}(r)$, and $\rho^{p=1}$ values for population $p=1$ in Figure 3

Figure 6 illustrates how the neighborhood dimension parameter, dr, (in km), and neighborhood size, $N_d^{p=1}(r)$ (in km^2) change with distance from the distribution center over the region $0km \le r \le 3\sigma_{p=1} = 9km$.
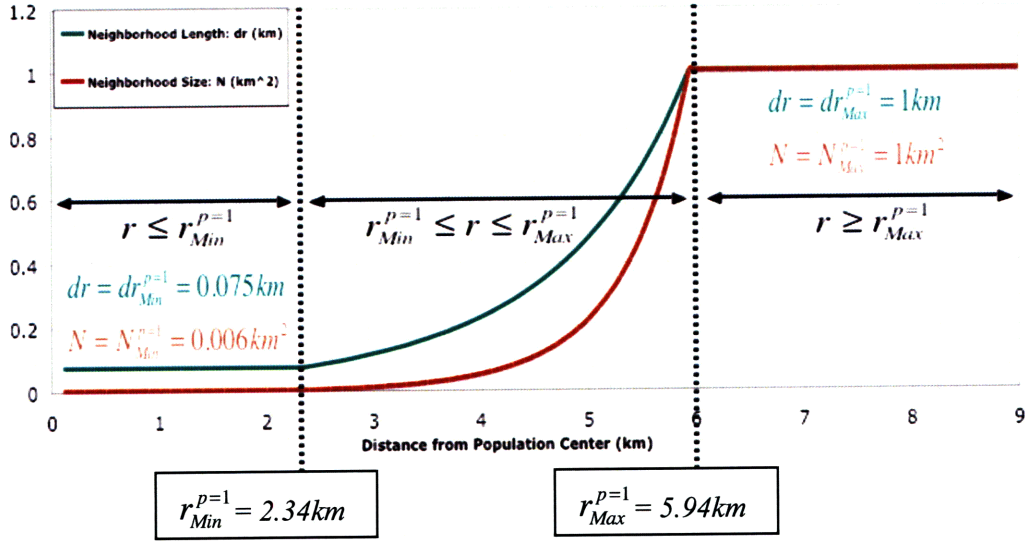
Figure 6: *dr (km)* **and** $N_d^{p=1}(r)$ *(km²)* **values as a function of *r* for *p=1* in Figure 3**

Because neighborhood size defines the maximum area corresponding to a single neighborhood, it also defines the size of a search grid cell as a function of distance from the population center. Similarly, because all neighborhoods at the same distance from the population center are the same size, *N(r)*, all search grid cells at this distance are the same size. As a result, the total grid cells associated with an annular ring at *r* for population *p* equals the total *possible* neighborhoods associated with this ring:

$$Cells\,in\,Ring\,at\,r = \frac{2\pi r}{dr} = \frac{2\pi r}{\left[\dfrac{2\pi r}{\left(d_p \bullet f_\Gamma^p(r)\right)}\right]} = d_p \bullet f_\Gamma^p(r) = d_r^p \qquad (0.15)$$

The total number of search grid cells required to model the population, is then the sum of the cells in all such rings, which is just the total number of neighborhoods in the population, d_p. This more efficient allocation of search points reduces the number of such points required to characterize the coverage region, enabling reduced computational resources.

5.3 Household Density

Each location can support multiple subscriber households, and each population density can have a different mix of household sizes. For example, dense, urban populations will have a significantly different mix of apartment buildings, businesses, and single-family homes than

sparser, rural areas. To address this issue, we define multiple household-per-location size categories. Each category, c, ($c = 1, 2, ..., C$) corresponds to a unique number of households, ϕ_c. For example, if the first category ($c = 1$) is defined as "single family home," then $\phi_{c=1} = 1$ assigns one household per location. Table 6 provides the four category definitions and values used in Figure 3 ($C = 4$).

Category ID	Name	Households per location
$c = 1$	Single-family homes	$\phi_{c=1} = 1$
$c = 2$	Apartments (small)	$\phi_{c=2} = 4$
$c = 3$	Apartments (large)	$\phi_{c=3} = 10$
$c = 4$	Businesses	$\phi_{c=4} = 15^9$

Table 6: Household size categories and values used in Figure 3

We would like to characterize how households are assigned to neighborhoods as a function of distance from a population center. First, we define the discrete random variable Φ_i: the number of households assigned to every location on neighborhood d_i^p, ($i = 1, ..., d_p$) in population p. The corresponding probability mass function, $g_{\Phi_i}^p(\phi_c)$, defines the probability that every location on a random selected neighborhood has household quantity ϕ_c. Table 7 provides the $g_{\Phi_i}^p(\phi_c)$ values used to model the populations in Figure 3.

Population	$g_\Phi^p(\phi_{c=1})$	$g_\Phi^p(\phi_{c=2})$	$g_\Phi^p(\phi_{c=3})$	$g_\Phi^p(\phi_{c=4})$
$p = 1$	20%	30%	40%	10%
$p = 2$	30%	40%	20%	10%
$p = 3$	20%	50%	10%	20%

Table 7: Specific $g_{\Phi_i}^p(\phi_c)$ values for populations in Figure 3

Individual $g_{\phi_c}^p(\phi_c)$ probabilities, when combined with the total neighborhoods, d_p, and locations per neighborhood, l_d^p, for each population p, determine the total households in a population, H_p. This provides a mechanism to increase modeled population size beyond the initial locations per

---

[9] This number represents the average workforce per business

population I_p. Table 8 illustrates this effect for the populations in Figure 3. The I_d^p values are taken from Table 3, and the ϕ_c values from Table 6.

Population	d_p	I_d^p	$g_\Phi^p(\phi_{c=1})$	$g_\Phi^p(\phi_{c=2})$	$g_\Phi^p(\phi_{c=3})$	$g_\Phi^p(\phi_{c=4})$	I_p	H_p
$p=1$	2k	12	20%	30%	40%	10%	24k	166k
$p=2$	4k	12	30%	40%	20%	10%	48k	259k
$p=3$	4k	24	20%	50%	10%	20%	72k	595k
						Totals	144k	>1Mil

Table 8: Subscriber size bins and population fractions in Figure 3

The total number of neighborhoods with households per location ϕ_c is given by $d_p \cdot h_\Phi^p(\phi_c)$. For example, in population $p=1$ in Table 8, the total neighborhoods for which every location is assigned $\phi_{c=4} = 10$ households is given by:

$$d_{p=1} \cdot g_\Phi^{p=1}(\phi_{c=3}) = 2k \cdot 40\% = 800 \tag{0.16}$$

As the final two columns in Table 8 illustrate, incorporating households per location increases the total population size modeled from one hundred forty-four thousand to slightly more than one million. To understand this how this happens, consider the first row in Table 8 representing population $p=1$ in Figure 3. Recall from Table 3 that the initial location population, $I_{p=1}=$ 24,000. We transform this into the corresponding household population, $H_{p=1}$, according to:

$$\left[d_{p=1} \cdot I_d^{p=1}\right] \cdot \left[\left(g_\Phi^{p=1}(\phi_{c=1}) \cdot \phi_{c=1}\right) + \dots + \left(g_\Phi^{p=1}(\phi_{c=4}) \cdot \phi_{c=4}\right)\right] = H_{p=1} \tag{0.17}$$

Therefore, the total *households* in population $p=1$ in Figure 3 is given by,

$$\left[2k \cdot 12\right] \cdot \left[\left(20\% \cdot 1\right) + \left(30\% \cdot 4\right) + \left(40\% \cdot 10\right) + \left(10\% \cdot 15\right)\right] = 166k \tag{0.18}$$

In general, the total households is the sum of all such population rows, P, and household size categories C, and will be greater than or equal to the total locations. Symbolically this is given by:

$$\sum_{p=1}^{P} I_p \leq \sum_{p=1}^{P} H_p \equiv \sum_{p=1}^{P} \left[\left(d_p \cdot I_d^p\right) \cdot \left(\sum_{p=1}^{P} \sum_{c=1}^{C} g_\Phi^p(\phi_c) \cdot \phi_c\right)\right] \tag{0.19}$$

Next we define the population-specific probability mass function, $h_{\phi_c}^{p}(r)$, which defines the probability that a random neighborhood d_i^p at a distance r from a population center will have household size ϕ_c. Table 9 lists the $h_{\phi_c}^{p=1}(r)$ values for population $p=1$ in Figure 3 (where σ_p is the standard deviation of the $f_{\Gamma_i}^{p}(r)$ neighborhood distribution).

r	$h_{\phi_c=1}^{p=1}(r)$	$h_{\phi_c=2}^{p=1}(r)$	$h_{\phi_c=3}^{p=1}(r)$	$h_{\phi_c=4}^{p=1}(r)$
$0 \le \lvert r \rvert \le \sigma_{p=1}$	0.10	0.20	0.50	0.80
$\sigma_{p=1} < \lvert r \rvert \le 2\sigma_{p=1}$	0.20	0.30	0.40	0.20
$\lvert r \rvert > 2\sigma_{p=1}$	0.70	0.50	0.10	0.00

Table 9: $h_{\phi_c}^{p=1}(r)$ **values for population $p=1$ in Figure 3**

Figure 7(a)(b) presents graphs of both the probability mass function $h_{\phi_c}^{p=1}(r)$ and corresponding cumulative distribution $H_{\phi_c}^{p=1}(r)$ for the values in Table 9. Table 10(a)(b) presents the $h_{\phi_c}^{p=2}(r)$ and $h_{\phi_c}^{p=3}(r)$ values for populations $p=2,3$ in Figure 3.
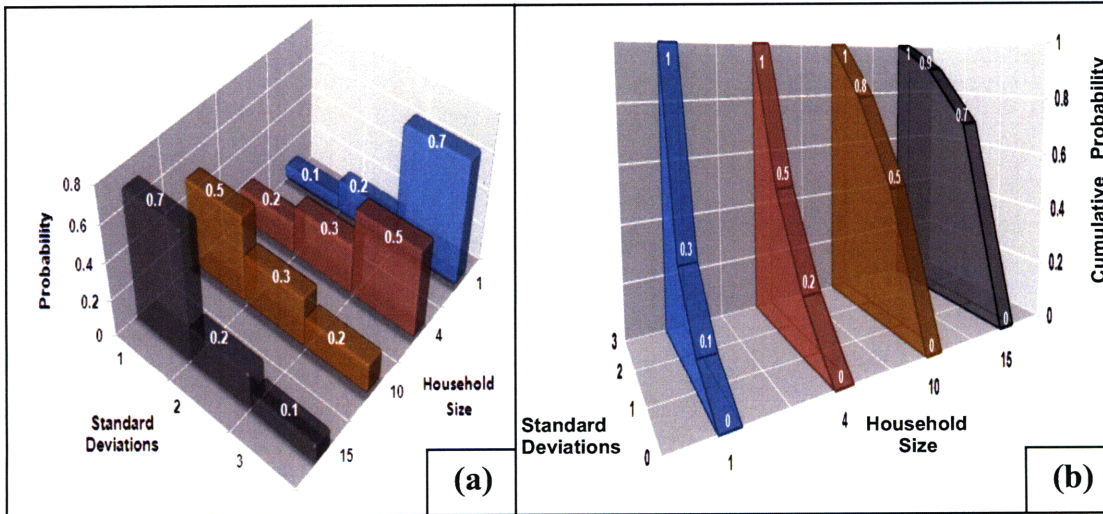
Figure 7: (a) $h_{\phi_c}^{p=1}(r)$, and (b) $H_{\phi_c}^{p=1}(r)$, for population $p=1$ in Figure 3

r	$h^{p=\{2,3\}}_{\phi_{c=1}}(r)$	$h^{p=\{2,3\}}_{\phi_{c=2}}(r)$	$h^{p=\{2,3\}}_{\phi_{c=3}}(r)$	$h^{p=\{2,3\}}_{\phi_{c=4}}(r)$
$0 \leq |r| \leq \sigma_p$	{0.25; 0.00}	{0.50; 0.25}	{0.70; 0.70}	{0.50; 0.90}
$\sigma_p < |r| \leq 2\sigma_p$	{0.50; 0.25}	{0.30; 0.25}	{0.20; 0.20}	{0.50; 0.10}
$|r| > 2\sigma_p$	{0.25; 0.75}	{0.20; 0.50}	{0.10; 0.10}	{0.00; 0.00}

Table 10: $h^{p=\{2,3\}}_{\phi_c}(r)$ **probabilities for populations** *p=2 & 3* **in Figure 3**

Given marginal distributions $f^p_{\Gamma_l}(r), g^p_{\Phi_l}(\phi_c)$, and $h^p_{\phi_c}(r)$, we can determine how household sizes are assigned to neighborhoods as a function of r using Bayes Theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(A)}$$

(0.20)

This relationship answers the question: what is the probability of event "*A*" occurring *given* that event "*B*" has already occurred? For our purposes, this question becomes: within a given population, what is the probability that every location on a random neighborhood is assigned ϕ_c households, *given* that neighborhood d^p_l is contained within a distance $\delta_p = r/\sigma_p$ from the population center (in units of standard deviations). Thus, (0.20) becomes:

$$P\left(\phi_c \middle| i \in [0, \delta_p]\right) = \frac{P\left(i \in [0, \delta_p] \middle| \phi_c\right) \cdot P(\phi_c)}{P\left(i \in [0, \delta_p]\right)}$$

(0.21)

The individual probabilities on the right hand side of (0.21) are given by:

$$P\left(i \in [0, \delta_p] \middle| \phi_c\right) = h^p_{\phi_c}(\delta_p)$$

$$P(\phi_c) = g^p_\Phi(\phi_c)$$

(0.22)

$$P\left(i \in [0, \delta_p]\right) = \int_0^{\delta_p} f^p_\Gamma(\delta_p)\, d\delta = F_\Gamma(\delta_p)$$

Therefore, the conditional probabilities are given by:

$$P\left(\phi_c \middle| i \in [0, \delta_p]\right) = \frac{h^p_{\phi_c}(\delta_p) \cdot g^p_\Phi(\phi_c)}{F_\Gamma(\delta_p)} \qquad \left(\forall\ \delta_p > 0\right)$$

(0.23)

These probabilities define the household distribution profile for any population *p* as a function of δ_p. Table 11 gives the conditional probabilities used to model population *p=1* in Figure 3 with

neighborhood distributions $f_{\Gamma_i}^{p=1}(\delta_{p=1})$, and $h_{\phi_c}^{p=1}(\delta_{p=1})$, and $g_{\Phi_i}^{p=1}(\phi_c)$ values in Table 7 and Table 9.

$P\left(\phi_c \Big| i \in \left[0, \delta_{p=1}\right]\right)$	$0 < \delta_{p=1} \leq 1$	$1 < \delta_{p=1} \leq 2$	$\delta_{p=1} > 2$
$\phi_{c=1} = 1$	0.056	0.129	0.424
$\phi_{c=2} = 4$	0.166	0.290	0.454
$\phi_{c=3} = 10$	0.556	0.516	0.122
$\phi_{c=4} = 15$	0.222	0.065	0.00
Total Probability	1.00	1.00	1.00

Table 11: Household per location profile for population *p=1* in Figure 3

For example, we expect 5.6% of all neighborhoods located within one standard deviation from the population center, (the $0 < \delta_{p=1} \leq 1$ column in Table 3) to have single-family homes at every location, 16.6% to have four-unit apartment buildings, 55.6% to have ten-unit apartments, and the remaining 22.2% to be assigned businesses.

Figure 8(a)(b) illustrates both the household size probability distribution function, (pdf) and the conditional probability profile as a function of standard deviation for population *p=1* in Figure 3. Both graphs highlight the household size distribution corresponding to $\delta_{p=1} = 1$ standard deviation.

(a)   (b)
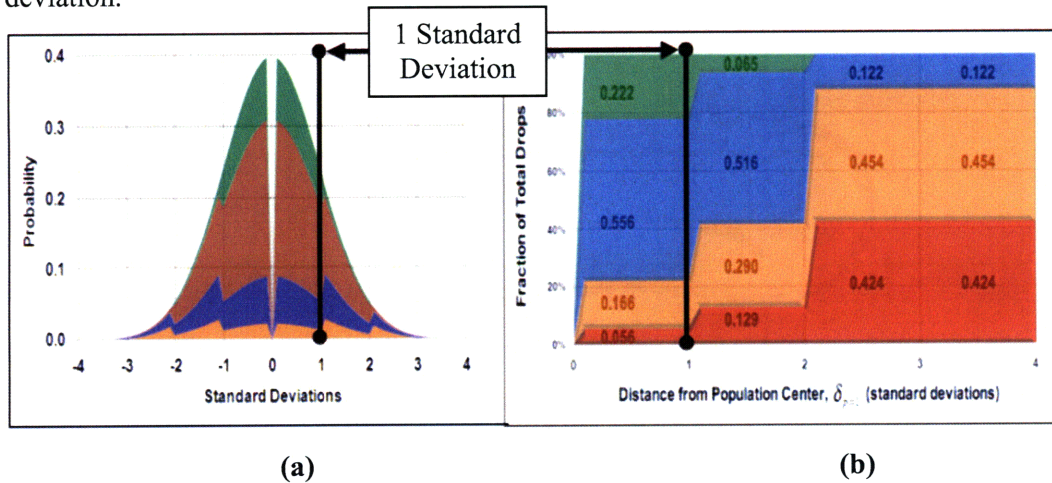
Figure 8: Household size profile: (a) pdf and (b) conditional probability profile

5.4 Data Density

Finally, we need to characterize population data demands. We first define data categories or service tiers, t, ($t = 1,...,T$) each of which corresponds to a sustained data rate, ψ_t. Each ψ_t, in units of megabits per second, (*Mb/s*) is the total data rate required to deliver a specific set of services to a single subscriber (internet, voice over IP, etc.). For example, the service tiers, and corresponding service profiles used in Figure 3 are shown below in Table 12.

Tier ID	Name	Services	Data Rate per Subscriber (Mb/s)
$t = 1$	Basic Service	Basic Internet	$\psi_{t=1} = 10$
$t = 2$	Extended Service	Internet, VOIP	$\psi_{t=2} = 40$
$t = 3$	Enterprise	High Capacity Internet	$\psi_{t=3} = 100$

Table 12: Service tiers and services for populations in Figure 3

Next, we characterize how many subscribers in each population are assigned each service tier. We define the discrete random variable, Ψ_t, the data demanded by every subscribing household assigned to neighborhood d_i^p in population p. The corresponding probability mass function, $k_\Psi^p(\psi_t)$, defines the probability that every household on a randomly selected neighborhood has data demand ψ_t. Table 13 provides the $k_\Psi^p(\psi_t)$ values used to model the populations in Figure 3.

Population	$k_\Psi^p(\psi_{t=1})$	$k_\Psi^p(\psi_{t=2})$	$k_\Psi^p(\psi_{t=3})$
$p = 1$	70%	20%	10%
$p = 2$	20%	60%	20%
$p = 3$	10%	20%	70%

Table 13: $k_\Psi^p(\psi_t)$ values for modeled populations in Figure 3

For example, 70% of all subscribers in population *p=1* are service tier *t=1*, 20% are tier *t=2*, and the remaining 10% represent business or enterprise clients.

Finally, we need to characterize how service tiers are assigned to subscribers. We define the population-specific probability mass function, $\chi_{\psi_t}^p(r)$, which defines the probability that a random neighborhood d_i at a distance r from a population center will have service tier ψ_t. Table

14 lists the $\chi_{\psi_t}^{p=\{1,2,3\}}(r)$ values for populations $p=1;2;3$ in Figure 3, (where σ_p is again the standard deviation of the $f_{\Gamma_i}^p(r)$ neighborhood distribution).

r	$\chi_{\psi_{t=1}}^{p=\{1;2;3\}}(r)$	$\chi_{\psi_{t=2}}^{p=\{1;2;3\}}(r)$	$\chi_{\psi_{t=3}}^{p=\{1;2;3\}}(r)$
$0 \le |r| \le \sigma_p$	{0.10; 0.20; 0.70}	{0.20; 0.60; 0.20}	{0.40; 0.40; 0.20}
$\sigma_p < |r| \le 2\sigma_p$	{0.20; 0.70; 0.20}	{0.60; 0.20; 0.20}	{0.40; 0.20; 0.40}
$|r| > 2\sigma_p$	{0.70; 0.10; 0.10}	{0.20; 0.20; 0.60}	{0.20; 0.40; 0.40}

Table 14: $\chi_{\psi_{t=\{1,2,3\}}}^{p=\{1,2,3\}}(r)$ values for populations $p=1;2;3$ in Figure 3

In the same way households are assigned to neighborhoods, we can determine the conditional probability distribution of how data tiers are assigned to neighborhoods as a function of r using Bayes Theorem. Table 15 presents the resulting probabilities for the three populations in Figure 3.

$P\left(\psi_t \middle| i \in \left[0, \delta_{p=1}\right]\right)$	$0 < \delta_{p=1} \le 1$			$1 < \delta_{p=1} \le 2$			$\delta_{p=1} > 2$		
Population:	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3
$\psi_{t=1} = 10\,Mbps$	0.46	0.47	0.88	0.47	0.88	0.64	0.88	0.5	0.3
$\psi_{t=2} = 40\,Mbps$	0.27	0.4	0.07	0.4	0.07	0.18	0.07	0.29	0.52
$\psi_{t=3} = 100\,Mbps$	0.27	0.13	0.05	0.13	0.05	0.18	0.05	0.21	0.17
Total Probability	1.00			1.00			1.00		

Table 15: Conditional data tier probabilities as a function of distance r

6 Materials Systems Lab (MSL) Network Model

The MSL Network Model does not focus on optimally dimensioning one *specific* network, instead characterizing *relative network component changes* as a function of technology choice and population demographics. By performing a heuristic-directed search, the network model algorithms seek to maximize co-located fiber (which implies a cost-minimizing solution, as fiber installation is the primary deployment expense). The model accounts for each subscriber location, but statistically aggregates their individual characteristics (data demand, last mile fiber distance, and size of household), thereby significantly reducing the required input data. This data reduction, combined with the directed search, substantially decreases computational complexity,

allowing for large or regional networks to be more efficiently dimensioned. Additionally, because the input data mimics actual population demographics, the model is not tied to a specific carrier's region, making it easy to explore the effects of population demographics on network design without the need to specify each subscriber's geographic location.

6.1 Network Design Model Inputs

This parameter array fully characterizes both the architectural and demographic needs of the coverage region. It contains three data sets: *demand demographics*, the outputs of the population model; *technological constraints*, the user-defined technology and implementation strategies, and *operating context,* input information about existing network components (legacy), subscriber uptake rates, and the geographic extent of the coverage region to be modeled.

6.1.1 Technological Constraints

The architectural solution space is characterized in four dimensions: network reach, data transfer rate, splitter strategy, and multiplexing strategy. The *technological constraint* parameter set identifies a unique architectural choice by specifying one value for each dimension. Each constraint is defined via a combination of several user inputs.

Table 16 defines these parameters and provides an overview of the inputs considered by each.

Parameter	Definition	Inputs Considered
Network Reach (km)	Maximum CO to subscriber fiber distance after losses	• Equipment/fiber losses • OLT transmitter power • ONT receiver sensitivity • Power margin required
Data Transfer Rate (Mb/s)	Maximum data rate per fiber	• OLT transmission rate • Statistical multiplication
Splitter Strategy	Maximum subscribers per fiber	• Splitter stage quantity • Ports per splitter
Multiplexing Strategy	How to assign multiple subscriber signals per fiber	• Method (STDM/WDM/Hybrid)

Table 16: Technological constraints

This set attempts to capture network elements driving relative changes in network topology corresponding to technology choice and/or implementation strategies. As a result, other factors, such as geography-specific routing constraints (planning around rivers etc.) are not considered,

as they are assumed to affect all choices equally. The multiplexing strategies modeled are exclusively Statistical Time Division Multiplexing[10] (TDM) for the purposes of this thesis. The statistical gain enabled by this method is modeled as a constant which is defined for each technology.

6.1.2 Operating Context

The set of *operating context* input parameters constrain network dimensioning. The first such parameter set, *legacy percentages*, allows users to independently address the impacts of legacy fiber, conduit and equipment on network deployment. The set of *fiber installation* parameters defines how much fiber will be installed underground (buried) versus overhead (aerial) in each section of the network. Typically, only a small population fraction initially subscribes to services. The *penetration rate* parameter, p_r, *(0% < p_r ≤ 100%)* defines this fraction, and reduces the demand profile accordingly. In response, carriers may wish to initially deploy a smaller or reduced capacity network. The *build percentage* parameter, *b*, *(0% < b ≤ 100%)* enables this fractional deployment, where *b =100%* is a network serving every household in the coverage region. Individual penetration and build percentage parameters are used to enable design flexibility in the face of uncertainty. For example, regulatory requirements may necessitate larger initial deployments than suggested by initial penetration forecasts. Similarly, penetration rates over time are impossible to predict, potentially leading to overbuilding based on higher than expected subscriber demand. The *geographic constraints*, {*x,y*} define the coverage region dimensions. For example, in Figure 3, {*x =100km, y=100km*}. Finally, *fiber routing type* parameters provide a way to force each fiber segment to be deployed in either a straight-line (point-to-point or star) configuration, or as loops connecting network elements. By default, the model selects the fiber-minimizing choice when dimensioning each fiber segment.

6.2 Network Design Model Algorithms

Given a set of population demographics, the corresponding grid structure, and the technological constraints, and operating context parameters, the MSL model utilizes three algorithms, $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$, to dimension a network topology. The first algorithm, $\mathcal{A}_1$, characterizes subscriber

---

[10] Two or more data streams are combined by assigning each a timeslot in a transmission packet, and the addresses of the terminal and the data are transmitted together

neighborhood size and the curb to home or "frontage" fiber required to reach all subscribers in given neighborhood as a function of distance from the population center. Next, $\mathcal{A}_2$ sites the central offices required to serve the population as a function of neighborhood density. Because fiber installation is the primary cost driver in greenfield builds, comprising up to 50% of total capital investment, (Wagner, Igel et al. 2006) the final algorithm, $\mathcal{A}_3$, utilizes three heuristics to seek fiber minimizing solutions which co-locate fiber for the furthest distances possible before splitting, and then sites all splitter stages and calculates the fiber lengths required to reach all subscriber locations.

Typically, individual fiber links are deployed in one of two ways: as "point-to-point" or "star" links, straight line distances between network elements, or as "rings," loops connecting multiple network elements using a single fiber bundle. The strategy employed determines the link length required; therefore, the model explores both strategies where appropriate, and selects the individual link topology minimizing installed fiber route length.

As an example, consider the two-stage, un-amplified cascaded splitter architecture in Figure 9, (a modified version of Figure 1 employing technology "*A2*" defined in Chapter 0). This implementation strategy results in six individual fiber links[11]: (1) *backhaul-to-CO*, (2) *CO-to-CO*, (3) *CO-to-splitter*, (4) *splitter-to-splitter*, (5) *splitter-to-curb*, and (6) *curb-to-location*. The first link, *backhaul-to-CO*, connects central offices to metro edge nodes or MANs, which connect local access networks, ("LANs", typically a city and its suburbs) to other cities and states. Not all COs contain the expensive equipment required to connect to MANs however, (40λ WMSANS for example), instead transmitting aggregate LAN-related data to other COs. Link (2) models this *CO-to-CO* transmission distance. Both links require redundancy and robustness due to the large amount of data they carry; therefore, they are modeled exclusively as loops. The remaining links (3-6), (which together with fiber link 2, CO-to-CO, comprise a LAN fiber network) are modeled as both star and ring topologies. Figure 9 illustrates relative link locations for an example network with two splitter stages, (S_j, where $j = 1,2$). Links (*1,2,6*) are modeled as loops, and (*3,4,5*) as point-to-point.

---

[11] This technology requires the most total fiber length segments, and is therefore used to illustrate the maximum fiber segments the network model will consider.
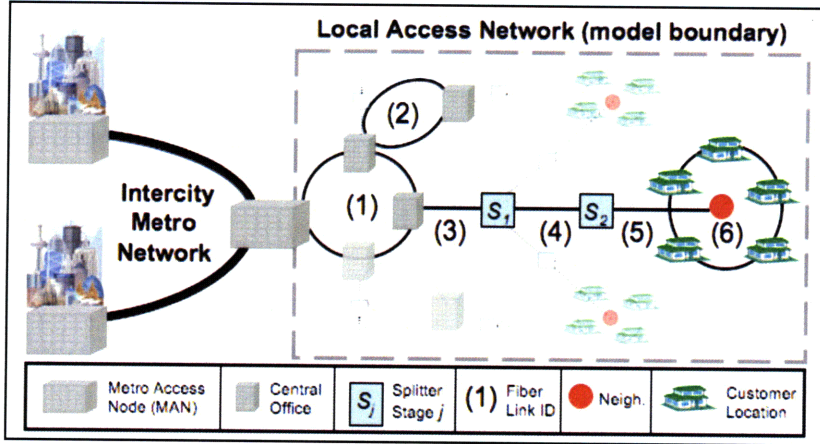
Figure 9 Network model boundaries and relative fiber link locations

Because different algorithms are responsible for dimensioning individual links, Table 17 summarizes link type, which modeling algorithm dimensions each link, and the link topology modeled by each algorithm.

Link ID	Link Description	Topologies Modeled	Algorithm
1	Backhaul to CO	Ring	$\mathcal{A}_2$
2	CO to CO	Ring	$\mathcal{A}_2$
3	CO to Splitter	Ring & Star	$\mathcal{A}_3$
4	Splitter to Splitter	Ring & Star	$\mathcal{A}_3$
5	Splitter to Neighborhood	Ring & Star	$\mathcal{A}_3$
6	Neighborhood to Location	Ring	$\mathcal{A}_1$

Table 17: Link topologies modeled

6.2.1 Neighborhood Size and Frontage Fiber Length

As neighborhoods get larger, the curb-to-home or "frontage" fiber required to reach *each* location increases (link 6 in Table 17). There can be millions of locations in the coverage region; therefore, it is important to characterize how this fiber length changes as a function of distance from the population center, r. The first algorithm, $\mathcal{A}_1$, performs this task.

Within a given neighborhood, locations are modeled as uniformly distributed and centered on the corresponding neighborhood. Figure 10 illustrates this structure for population $p=1$ in Figure 3, where each neighborhood contains twelve locations ($l^d_{p=1} = 12$).

46
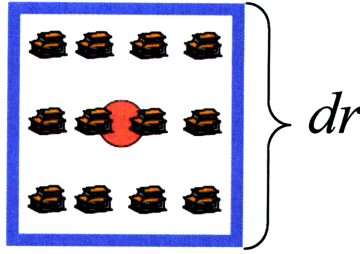
Figure 10: Uniformly distributed neighborhood for population *p=1* in Figure 3

The fiber connecting locations within a neighborhood is modeled as a fiber bundle loop, with individual frontage, or "last mile" fibers peeled away to connect the loop to customers. Uniform location distribution requires that any fraction of the total box area should, on average, contain an equal fraction of customer locations. Using this, we approximate a minimal fiber path as a loop containing half the total neighborhood area, $N_d^p(r)/2$, and therefore half the assigned customers, $l_d^p/2$. We picture a circle of radius r^* which performs this function, with an area given by:

$$\frac{1}{2}N_d^p(r) = \pi(r^*)^2 \qquad (0.24)$$

Therefore, the fiber loop will have radius:

$$r^* = \sqrt{\frac{N_d^p(r)}{2\pi}} = \frac{dr}{\sqrt{2\pi}} \qquad (0.25)$$

This radius is a function of r, with corresponding *dr* values given in (Table 4). Figure 11 below illustrates the corresponding fiber loop for the case when $l_p^d = 12$.
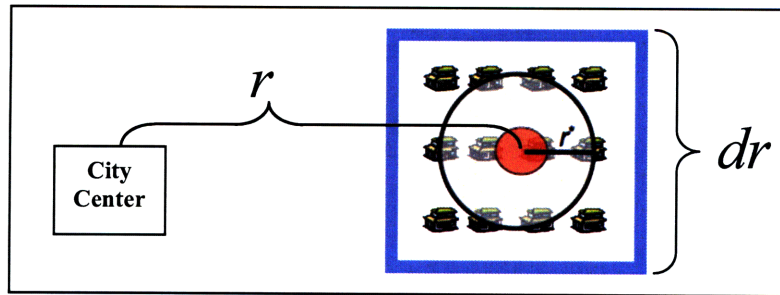
Figure 11: Fiber loop of radius *r** for a neighborhood with $l_p^d = 12$ locations

The distance from the neighborhood to the furthest subscriber location on such a fiber loop, D_{Max} is the sum of three components: the fiber required from neighborhood to loop (r^*), the loop itself ($2\pi r^*$), and loop to location (curb to home or "frontage") distances approximated as $r^*/2$. Symbolically, this is given by:

$$D_{Max} = \left[r^* + 2\pi r^* + \frac{r^*}{2} \right] = \left[r^* \left(1 + 2\pi + \frac{1}{2} \right) \right] \qquad (0.26)$$

Figure 12 illustrates these fiber components for the example neighborhood in Figure 11.
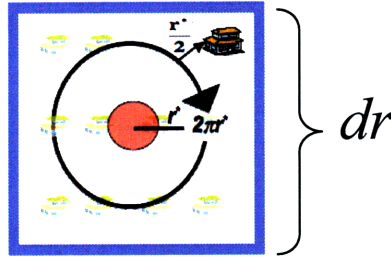
Figure 12: Neighborhood to furthest location fiber distance components

In general, there will be l_p^d locations per neighborhood, and each of these is assigned a frontage fiber length $r^*/2$. Therefore, the total installed fiber length for each neighborhood as a function of distance from the population center is given by:

$$F_d^p(r) = \left[r^* \left(1 + 2\pi + \frac{l_d^p}{2} \right) \right] = \left[\frac{dr}{\sqrt{2\pi}} \left(1 + 2\pi + \frac{l_d^p}{2} \right) \right] \qquad (0.27)$$

Where dr is defined in Table 4 as a function of r. Because dr is defined differently depending on the distance from the population center, the form of $F_d^p(r)$ will also depend on this distance. Table 19 uses (0.27) and Table 4 to define $F_d^p(r)$ as a function of r.

Region (km)	dr (km)	Installed Fiber Length: $F_d^p(r)$
$r \le r_{Min}^p$	$\sqrt{\dfrac{0.02 \cdot l_d^p}{40}}$	$\sqrt{\dfrac{1000 \cdot l_d^p}{\pi}}\left(1 + 2\pi + \dfrac{l_d^p}{2}\right)$
$r_{Min}^p < r < r_{Max}^p$	$\left(\dfrac{2\pi r}{d_p \cdot f_\Gamma^p(r)}\right)$	$\dfrac{r \cdot \sqrt{2\pi}}{d_p \cdot f_\Gamma^p(r)}\left(1 + 2\pi + \dfrac{l_d^p}{2}\right)$
$r \ge r_{Max}^p$	$\left(l_d^p \cdot N^{20}\right)$	$\dfrac{l_d^p \cdot N^{20}}{\sqrt{2\pi}}\left(1 + 2\pi + \dfrac{l_d^p}{2}\right)$

Table 18: Installed fiber length as a function of *r*

Figure 13 presents a simple, idealized example of how this fiber length grows with neighborhood size for a population *p*.
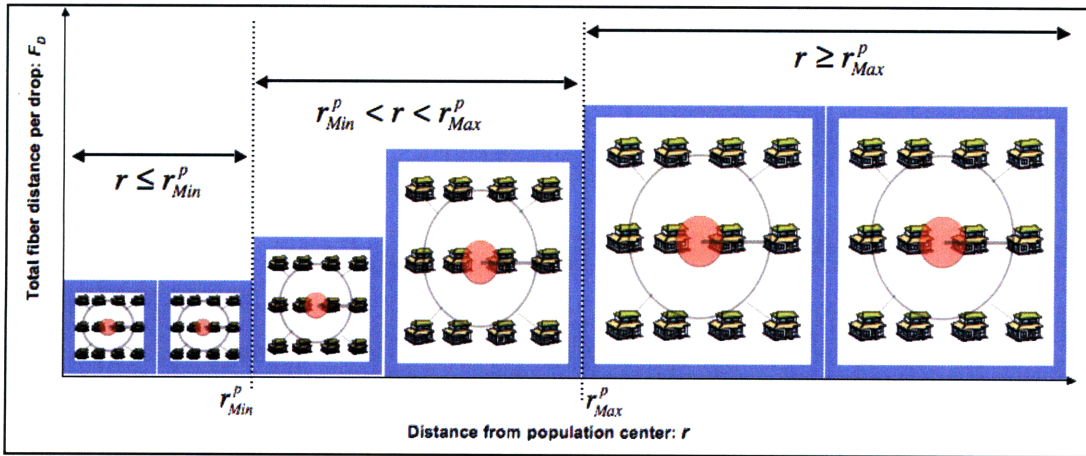
Figure 13: Neighborhood area growth as a function of *r*

6.2.2 *Central Office Siting and Neighborhood Assignment*

Once the frontage fiber length is defined for all neighborhoods in the coverage region, the second algorithm, $\mathcal{A}_2$, sites central offices, assigns neighborhoods to each office, and determines *backhaul to CO* and *CO to CO* fiber link lengths (links 1 & 2 in Table 17) linking the local access network to the intercity backbone. Central offices are assigned by population density; therefore, the first office will be placed at the neighborhood with the smallest frontage distance[12]. Once the central office is sited, neighborhoods are assigned based on available network reach. In

---

[12] Users can override this option and place central offices by hand

this context, network reach is defined as the maximum distance a coherent optical signal can travel from a transmitter, housed at the central office, to a receiver, at a customer location, less path-dependent losses, (see discussion in §1.1).

One effective way to reduce installed fiber length is co-locating many fiber bundles in individual buried trenches or aerial pole routes. When considering this deployment strategy however, one problem arises: if subscribers lie at the COs periphery (at R), then reaching them requires a direct connection from the CO, fully exhausting the power budget (and therefore reach) for this line. This connection mode is known as point-to-point, and results in inefficient fiber usage.

Recall that the power budget establishes the maximum distance, R, that any neighborhood can be from the CO (see §1.1). However, practical considerations mean that in reality the neighborhoods must be considerably closer. This is due to cost benefits that arise from co-locating feeder files coming out of the central office. As a result, fiber links between the CO and the end user do not travel in straight lines, but rather go along common pathways to minimize installation costs. These pathways have multiple steps: CO to inner splitter, inner splitter to outer splitters and out splitters to neighborhoods. To address this situation, $\mathcal{A}_2$, uses the first of three heuristics, H_1[13] ($0 \leq H_1 \leq 1$), to adjust the total possible reach, R, to a shorter distance, R_1, given by:

$$R_1 = H_1 \cdot R \tag{0.28}$$

While R remains the total possible reach for each CO, only neighborhoods within R_1 will be assigned to this CO. The areas containing customer locations, (the "inclusion area" A_I), and excluding locations (the "exclusion area" A_E) are defined as:

$$A_I = \pi R_1^2 = \pi (RH_1)^2$$
$$A_E = \pi \left(R^2 - R_1^2 \right) = \pi R^2 - A_I \tag{0.29}$$

Figure **14**(a)(b)(c) presents the $\mathcal{A}_2$ algorithmic steps for a normally distributed population: (a) central office siting at the neighborhood with shortest frontage distance; (b) total available reach, R, is calculated using the power budget as in §1.1, equation (0.1); (c) effect of heuristic H_1 on R, resulting effective radius R_1 and resulting area divisions A_I and A_E.

---

[13] This section describes each heuristic's particular function; specific heuristic values are determined via sensitivity analysis and discussed in the section *Heuristic Value Determination.*

| (a) | (b) | (c) |
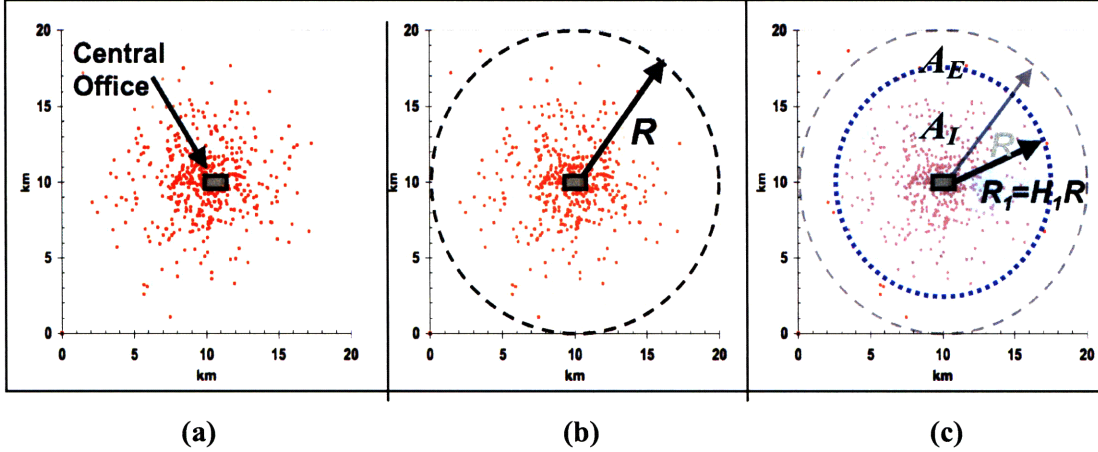
Figure 14: Central office siting and effective reach determination

Once a central office is sited, $\mathcal{A}_3$ assigns all neighborhoods within the effective radius to this office. The next CO is sited at the remaining neighborhood with the smallest frontage fiber length, or, conversely, the neighborhood with the highest subscriber location density. Unassigned neighborhoods within the effective radius of this new CO are then assigned. The procedure is repeated until all, or a specified maximum, neighborhoods are assigned to a central office.

The long "tail" of the normal distribution can result in neighborhoods which are sited further than three standard deviations away from a population center. These neighborhoods will have large frontage distances and are often isolated. As a result, connecting these customers may require a dedicated central office and point-to-point fiber links, introducing significant expense for marginal additional coverage. Therefore, $\mathcal{A}_2$ also includes a mechanism, the fractional coverage input parameter f, ($0 \leq f \leq 1$) to mitigate these disproportionate expenses by relaxing the requirement of complete coverage. For example, in a normally distributed population, on average, 1.3% of all neighborhoods will lie greater than three standard deviations from the mean, (and thus on the outer edge of the coverage region). Setting $f = 0.987$ will halt the CO siting routine once the first 98.7% of neighborhoods are assigned, (ranked by frontage distance) excluding these neighborhoods.

6.2.3 Splitter Siting & Fiber Link Length Determination

The final algorithm, $\mathcal{A}_3$, sites splitters, and determines link topology for the *CO-splitter*, and *splitter-to-neighborhood* and *splitter-to-splitter* fiber links, (links 3-5 in Figure 9). Each splitter stage, j, consists of many splitter sites s_j. A non-cascaded architecture has a single splitter stage, while cascaded architectures have at least two. For example, the local access network in Figure 9

consists of two stages, $j=1,2$, each with multiple splitter sites, $\{s_{j=1}\}$ and $\{s_{j=2}\}$. The algorithm utilizes the remaining two heuristics and three competing constraints/criteria C_1, C_2, & C_3 to site splitters, and then allocate neighborhoods and the corresponding locations, and determine the necessary fiber lengths.

6.2.3.1 Constraint Derivation

The first constraint, C_1, requires that the total distance from the CO to each neighborhood, including the distance to the furthest customer location, be less than the total network reach. Using (0.25) and (0.26), we obtain an expression for C_1:

$$C_1: \quad \begin{aligned} &\left[r + D_{\max} \right] \le R \\[4pt] &\left[r + r^* \left(2\pi + \frac{3}{2} \right) \right] \le R \\[4pt] &\left[r + \frac{dr}{\sqrt{2\pi}} \left(2\pi + \frac{3}{2} \right) \right] \le R \end{aligned} \qquad (0.30)$$

Because dr is defined differently depending on the distance from the population center, the form of C_1 will also depend on this distance. Table 19 uses (0.30) and Table 4 to define C_1 as a function of r.

Region (km)	dr (km)	Constraint Form
$r \le r_{Min}^p$	$\sqrt{\dfrac{0.02 \cdot l_d^p}{40}}$	$\left[r + \sqrt{\dfrac{0.02 \cdot l_d^p}{80\pi}} \left(2\pi + \dfrac{3}{2} \right) \right] \le R$
$r_{Min}^p < r < r_{Max}^p$	$\left(\dfrac{2\pi r}{d_p \cdot f_\Gamma^p(r)} \right)$	$\left[r \left(1 + \dfrac{(2\pi)^{3/2}}{d_p \cdot f(r)} \left(2\pi + \dfrac{3}{2} \right) \right) \right] \le R$
$r \ge r_{Max}^p$	$\left(l_d^p \cdot N^{20} \right)$	$\left[r + \dfrac{l_d^p \cdot N^{20}}{\sqrt{2\pi}} \left(2\pi + \dfrac{3}{2} \right) \right] \le R$

Table 19: Distance constraint values by distance region

The second constraint, C_2, ensures that the total customer locations on each PON is less than or equal to the number of available splitter ports. Each splitter stage, j, is assigned a user-defined per-splitter port count, N_j. An additional input parameter, E_j the splitter stage efficiency

$(0 \le E_j \le 1)$, determines the fraction of empty ports, (to allow for additional customers, repairs etc.). The total available ports for a PON after a single splitter stage, j, is given by:

$$P_j = N_j \cdot E_j \qquad (0.31)$$

Additional splitter stages will multiple each of these ports. Therefore, the total available ports on a given PON is given by:

$$\prod_j P_j \qquad (0.32)$$

Once a splitter site is established, one neighborhood, i, with its corresponding locations, l_d^p, at a time is added to the splitter site, s_j, until this maximum number of ports is reached. Therefore, C_2 has the form:

$$C_2 : \sum_{i=1}^{Max\,Drops} \left(l_d^p \right)_i \le \prod_j P_j \qquad (0.33)$$

The final constraint, C_3, ensures that the total customer data demand assigned to a PON does not exceed the total data available at this site, D_{PON}. Recall that every customer on a neighborhood is assigned the same service tier, ψ_t, by the $\chi_{\psi_t}^p(r)$ distribution, (see Table 13, Table 14 and intermediate discussion). Additionally, it is assumed that not every subscriber will utilize the entire allotted data available simultaneously, and that dynamic bandwidth allocation will reallocate unused bandwidth on a given PON. As a result, the actual data rates allocated to a given PON may be larger that the maximum OLT transmission rate. To model this behavior, we introduce the statistical multiplier, M, a user defined input which increases the available transmission rate available on a given PON. For example, if the maximum sustained bandwidth for a single PON is 2.5Gbps, a multiplier value of $M=10$ would allow customer data demands of up to 25Gbps to be allocated to this PON. The total data per neighborhood, d_i, is given by:

$$d_i = \psi_i \cdot l_d^p \qquad (0.34)$$

Therefore, C_3 can be written as:

$$C_3 : \sum_{i=1}^{Max\,Drops} d_i \le M \cdot D_{PON} \qquad (0.35)$$

6.2.3.2 Splitter Siting

One way to reduce fiber length is by co-locating fiber in a single trench as far as possible away

from the central office before splitting. Therefore, $\mathcal{A}_3$ begins by selecting the neighborhood furthest away from the central office, and tentatively placing a splitter site there. Often, this neighborhood will be located at or near the effective reach boundary, R_1, illustrated in Figure 14. Although additional reach remains, (the total reach R minus the effective reach R_1) no neighborhoods beyond this radius are assigned to the CO. As a result, all neighborhoods assigned to this splitter site will be skewed inwards toward the CO. Figure 15 illustrates this result.
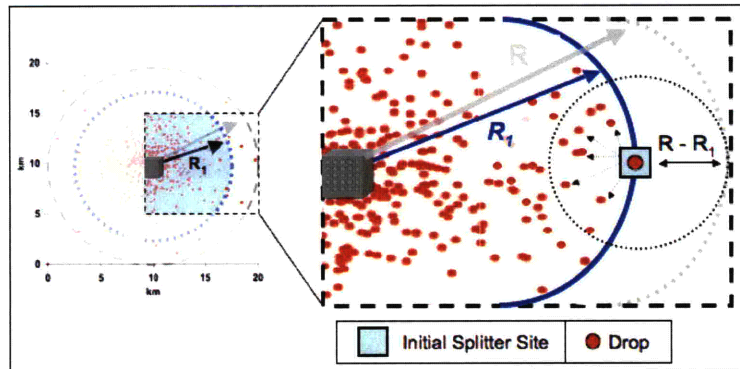
Figure 15: Initial splitter siting

One way to increase available fiber co-location is to place the splitter further inwards towards the CO, and then apply the C_1 distance constraint to every other neighborhood assigned to this CO. The splitter site would be moved again, and C_1 would be re-applied. This procedure could then repeated until a local optimum is found[14], minimizing the total distance from the CO to the splitter sites, and then to neighborhoods. This procedure would then be applied to every splitter site until the configuration minimizing the total fiber length was found. Because each splitter site can be located anywhere within A_i, (the inclusion region shown in Figure 14) and thousands of neighborhoods can be assigned to each CO, employing such an inefficient, exhaustive search algorithm quickly becomes computationally challenging, requiring enormous time and resources[15]. Additionally, this search algorithm has only considered the singular distance constraint, C_1, while full optimization requires considering the entire set competing constraints

---

[14] This procedure is similar to the "k-means" algorithm; however, in this instance the total number of splitter sites required is unknown a priori.

[15] This class of problems is defined as "NP-hard:" requiring computational time which scales exponentially with variable quantity. As of yet, only an exhaustive search ensures a global optimum, but become intractable as problem complexity and size increase

simultaneously.

To reduce computational complexity, a second heuristic, H_2, is introduced which sets a limit on how close the first stage of splitter sites ($S_{j=1}$) can be to the R_1 boundary, defining the region these sites may occupy. The resulting $S_{j=1}$ boundary is given by:

$$R_2 = \left[H_2 \cdot R_1\right] = \left[H_1 \cdot H_2 \cdot R\right] \qquad (0.36)$$

where $0 \leq \left[H_1, H_2\right] \leq 1$. Just as before, the algorithm begins by selecting the neighborhood furthest from the CO but still within the R_2 boundary. The first splitter site is then tentatively placed at this location. All remaining neighborhoods within the R_1 radius are then searched a *single* time, and compared against the C_1, C_2, & C_3 constraints. Neighborhoods meeting these constraints are assigned to this site. The splitter site is then relocated to the geometric mean of these neighborhoods. The next splitter site is placed at the furthest remaining neighborhood, and $\mathcal{A}_3$ repeated. Figure 16(a)(b)(c) illustrates how the first two heuristics, H_1 & H_2, affect the relative reach available from CO to splitter site. The lone black circle in (a) represents the original total reach available to the CO, R. The additional blue circle in (b) corresponds to the effective reach, R_1, imposed by the first heuristic, H_1 (all subscribers within this reach will be served by this CO). Finally, the green circle in (c) represents the reach constraint, R_2, imposed by H_2, (all $S_{j=1}$ splitter sites will be within this region)

(a)   (b)   (c)
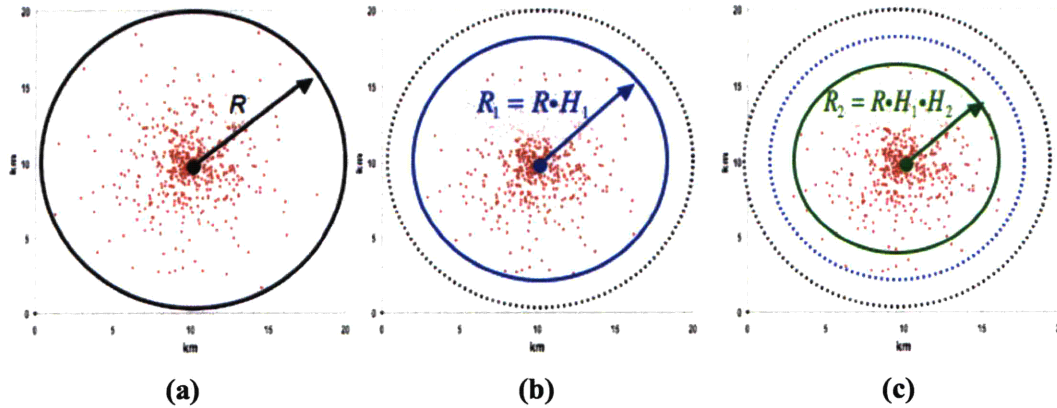
Figure 16(a)(b)(c): Relative heuristic effects

The resulting splitter site is now able to more effectively utilize the reach and more efficiently collect neighborhoods. Figure 17 (a)(b) compares illustrates this effect by comparison with the inefficient splitter siting in Figure 15, while Figure 18 illustrates an idealized representation of

the final $S_{j=1}$ splitter stage configuration.

(a)   (b)
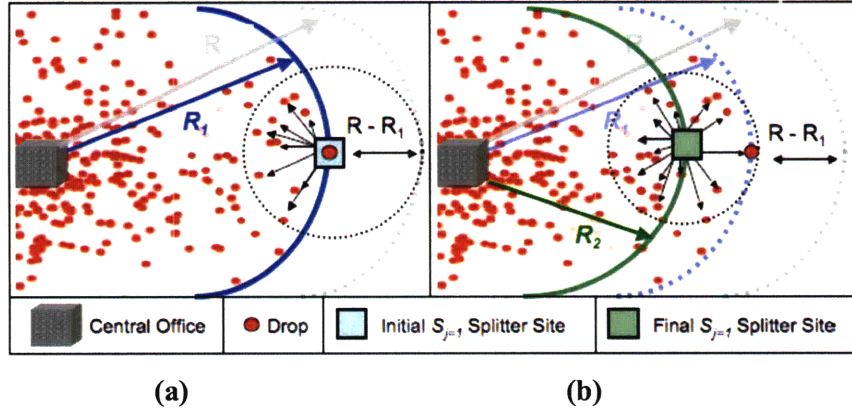
Figure 17: (a) Pre and (b) post H_2 $S_{j=1}$ splitter siting
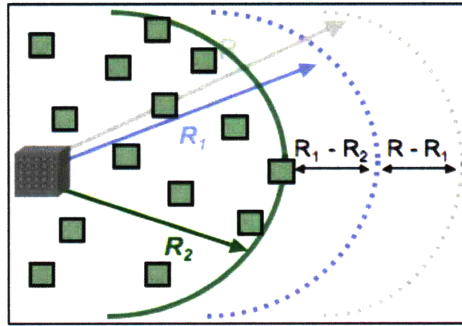
Figure 18: $S_{j=1}$ splitter stage configuration results

Once a splitter site is identified, the total installed fiber length required to reach all neighborhoods allocated to it is calculated for both star (point-to-point) and loop configurations. The architecture minimizing total installed fiber length is selected. The star configuration consists of a single fiber component: the straight lines from the splitter to the neighborhoods (S1). The loop configuration has three components: the radius length from the splitter to the loop, which is equal to the average straight-line distance from the splitter site to all included neighborhoods (L1); the loop itself, (L2); and the distance required to connect all neighborhoods to this loop, (L3).

Figure 19(a)(b) illustrates both (a) star and (b) loop configurations for a non-cascaded architecture with a single splitter stage.

56
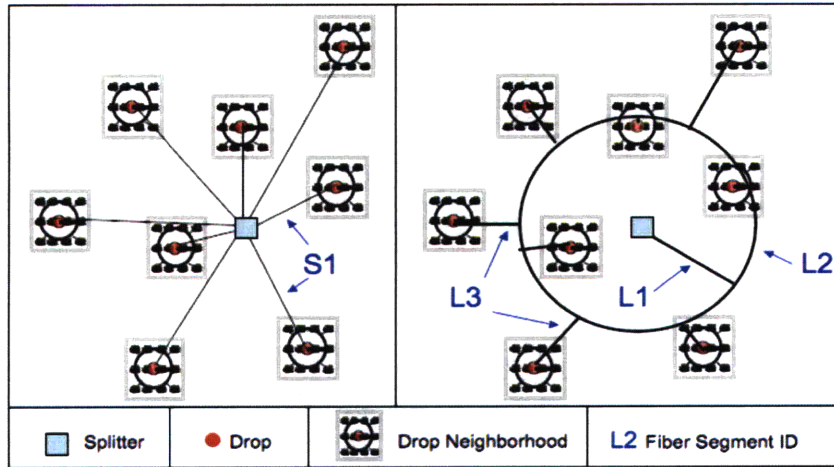
Figure 19(a)(b): Star and loop configurations for a single splitter site serving multiple neighborhoods in a non-cascaded architecture

Once the configuration with the smaller fiber length is selected, the distance from the splitter site to the CO is calculated.

Thus far, we have only considered architectures with a single splitter stage, $(S_{j=1})$; however, two-stage architectures $(S_{j=2})$ are also modeled. In these cases, the initial $j=1$ splitter sites are placed as shown in Figure 18 (as in the non-cascaded case). Once customers and data are assigned to these sites, each performs the role neighborhoods did initially: statistically characterizing the population. However, rather than a single neighborhood with l_d^p individual customer locations, each site may now represent many such neighborhoods. The result is a much smaller collection of aggregation points.

The algorithm views each $S_{j=1}$ splitter site as it previously viewed a neighborhood, and the algorithm repeats the initial $S_{j=1}$ siting process to place the $S_{j=2}$ splitter sites. As before, we are interested in co-locating PONs for the largest distance possible; therefore, the initial $S_{j=2}$ site is tentatively located at the $S_{j=1}$ site furthest from the CO. This site will typically be at or near the R_2 distance boundary imposed by the second heuristic, H_2 which ensures that no $S_{j=1}$ splitter sites can be located in the R_1-R_2 region. As a result, all $S_{j=1}$ sites assigned to an $S_{j=2}$ site near the R_2 boundary will be skewed inwards towards the CO, (just as the neighborhoods were in Figure 15). Figure 20 illustrates this situation, using the $S_{j=1}$ splitter stage configuration in Figure 18.

57
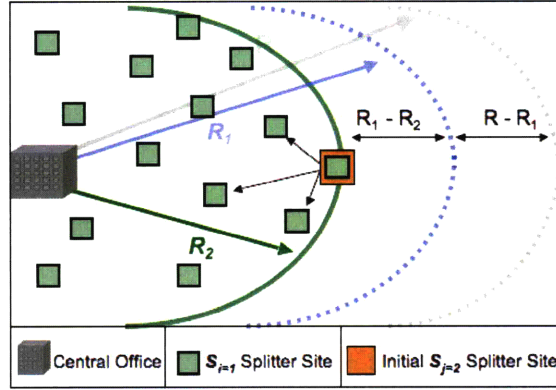
Figure 20: Initial $S_{j=2}$ siting

To address this issue, a third and final heuristic, H_3, is introduced which sets a limit on how close the second stage of splitter sites ($S_{j=2}$) can be to the R_2 boundary. The resulting $S_{j=2}$ site boundary is given by:

$$R_3 = \left[H_3 \cdot R_2\right] = \left[H_1 \cdot H_2 \cdot H_3 \cdot R\right] \tag{0.37}$$

where $0 \leq \left[H_1, H_2, H_3\right] \leq 1$. Figure 21 illustrates (a) the pre-existing R, R_1, and R_2 boundaries and neighborhoods as presented in Figure 16 and (b) the corresponding $S_{j=2}$ splitter sites and R_3 boundary.
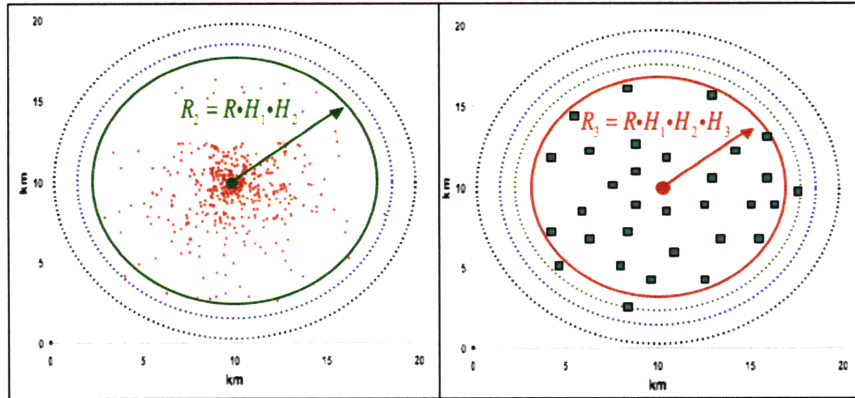
Figure 21: Final heuristic reach effects and $S_{j=1}$ configuration

The algorithm begins by selecting the $S_{j=1}$ site furthest from the CO, yet still within the R_3 boundary. The first $S_{j=2}$ splitter site is then tentatively placed at this location. All remaining neighborhoods within the R_2 radius are then searched a single time, and compared against the $C1$, the $C2$ and $C3$ constraints given in equations (0.29) (0.32) and (0.34), (with the summations taken over $S_{j=1}$ splitter sites instead of neighborhoods). First stage splitter sites meeting these

58

criteria are then assigned to the $S_{j=2}$ site. Finally, this site is relocated to the geometric mean of the assigned $S_{j=1}$ sites. Figure 22(a)(b) illustrates these two steps: (a) initial $S_{j=2}$ siting at the $S_{j=1}$ site furthest from the CO yet still within R_3 and (b) the final $S_{j=2}$ site at the geographic mean of the $S_{j=1}$ sites assigned to it.
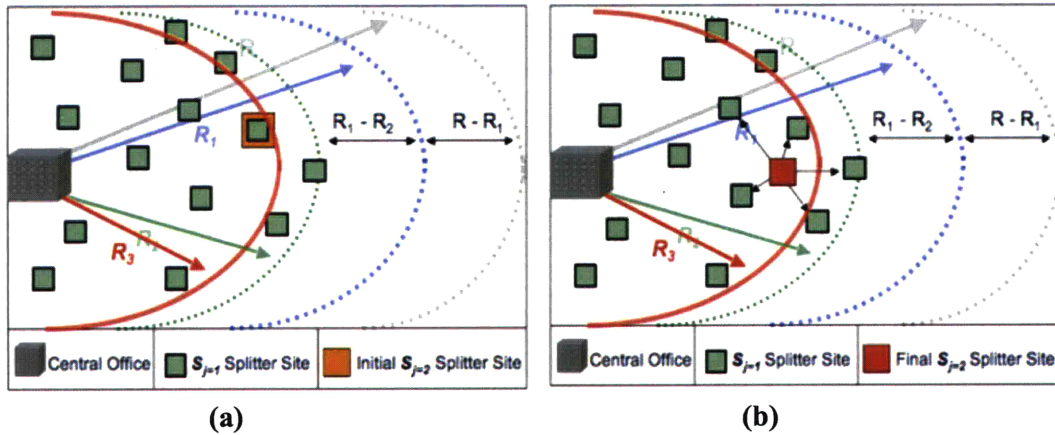
Figure 22: $S_{j=2}$ splitter-siting steps: (a) initial siting (b) final siting

The next $S_{j=2}$ site is then placed at the furthest remaining $S_{j=1}$ site, and the process repeated until all $S_{j=1}$ sites are assigned.

7 Heuristic Calibration and Sensitivity Analysis

All three model heuristics affect topology, and therefore fiber length. Recall that the first heuristic, H_1, determines the reach from the CO to the furthest customer assigned to it, and the other two heuristics, (H_2 and H_3) attempt to increase the efficiency of fiber routing by setting limits on the maximum distances from CO to splitter sites once an H_1 value has been chosen[16]. Therefore, during the heuristics calibration, we first choose an H_1 value, and then test H_2 and H_3 combinations given this value. Certain value combinations will result in inefficient and costly network topologies; therefore, calibration is required to determine which combination of values corresponds to the minimal predicted fiber length. Multiple value combinations will enable the generation of response surfaces illustrating the sensitivity of fiber length to changes in heuristic values. The calibration goal to select a heuristic combination which *balances* minimal fiber length with fiber length robustness to small changes in heuristic values (local solution stability).

---

[16] Recall that each heuristic value can range from 0 to 1.

59

7.1 Population Demographics Modeled

The population statistics were selected to represent a population characterized by urban, suburban and rural components. Therefore, a population distribution is modeled which is very dense in the center, and varies smoothly as the distance from this center increases. Additionally, a range of service tiers and household sizes are modeled using different distributions, but exhibiting the same general distance/density relationship.

Parameter	Symbol	Value
Population distribution	$f_\Gamma^p(r)$	$N(0,16)$
Households per location total bins	C	4
Households per location by bin	c_i	Table 6
Household bin distribution	$g_\Phi^p(\phi_c)$	Table 7 (row p=1)
Household size distribution	$h_{\phi_c}^p(r)$	Table 9
Household per location probability	$P\left(\phi_c \middle| i \in \left[0, \delta_p\right]\right)$	Table 11
Total service tiers	T	4
Data rate per tier	ψ_t	Table 12
Service tier distribution	$k_\Psi^p(\psi_t)$	Table 13 (row p=1)
Total neighborhoods	D_p	5,000
Locations per neighborhood	l_d^p	10
Total locations	L_p	50,000
Total households	H_p	345,000

Table 20: Heuristic calibration population parameter values

7.2 Technology Choice Modeled

The technology choice and architecture parameter values utilized for the calibration are provided in Table 17. The values reflect the baseline GPON architecture used in a widely cited 2004 Corning analysis, (Vaughn, Kozischek et al. 2004) and loss and component values gathered from industry. Some parameters are unique to this analysis, such as the statistical multiplier, (see §6.1.1) and the network reach is determined via the method outlined in §6.2.2.

Parameter	Value
Tx Power (dbm)	-28
Rx Sensitivity (dbm)	0
Loss per $\log_2$[port count] (db)	3.5
Loss per fiber km (db)	0.4
Safety margin (db)	3
Amplification (db)	0
Network reach (km)	19
Max data rate per PON (Gbps)	2.5
Multiplexing strategy	TDM
Splitter strategy	Cascaded
Splitter port count	(1:4) (1:8)
Splitter efficiency	(100%) (100%)
Statistical multiplier	10
Build type	Greenfield
Build (%)	100
Penetration (%)	100

Table 21: Heuristic calibration technology choice

7.3 Methodology

Recall that the first heuristic, H_1, modifies the total allowable reach available to a central office, (see Figure 14). The larger H_1 is, the smaller the total possible straight-line distance to customer locations. As a result, when modeling only a single population and central office, small H_1 values can result in customers which cannot be reached by the CO. Therefore, although some H_1 values may require fewer kilometers of installed fiber, thereby appearing more desirable during the calibration process, this is actually a result of reaching fewer total customer locations.

While fiber minimization is one goal of an efficient network, equally important is reaching a high percentage of customers in the coverage region. Recall that each population distribution results in a unique customer location pattern. As a result, different location fractions will lie beyond the reach limitation imposed by H_1 every time a population is generated. To correct for this variation, we required that calibration solutions produced networks reaching at least *99%* of the total customer locations. Practically this means that only H_1 values where the resulting network reached *4,950* of the *5,000* total customer locations were retained.

61

Because calibration was performed using a specific population type and technology choice, the heuristic set resulting in the fiber minimizing solution[17] for these parameters may not be the optimal set for other populations and technologies. Therefore, the final heuristic value set will produce a fiber length solution which attempts to both minimize both fiber length and be robust to small changes in the heuristic values. Additionally, although some fiber length variance is expected among different heuristic value combinations, large variation would indicate model instability. Large fluctuations are defined as individual mean average fiber length per location values varying by more than ±10% of the total average fiber length per location over all heuristic value combinations. By explicitly including fiber length variance as a decision criterion, the calibration process also enables sensitivity analysis.

The calibration process first selects an H_1 value, and then models the fiber length values resulting from different H_1, H_2 value combinations. The minimum, maximum and interval values for modeled for each heuristic are given in Table 22.

Heuristic	Minimum	Maximum	Interval	Quantity
H_1	0.55	0.75	0.05	5
H_2	0.2	0.9	0.1	8
H_3	0.2	0.9	0.1	8

Table 22: Heuristic parameter values modeled

The fast model run time enabled full characterization of the solution space, resulting in

$5 \cdot 8^2 = 320$ heuristic value combinations. The solutions were grouped by H_1 values, each of

which represented *64* total fiber length values (*8* H_2 and *8* H_3 values). The maximum, minimum, and average fiber length per neighborhood values were then calculated for each group, and the range of lengths (max - min) used as a measure of intra-group variation.

7.4 Results and Analysis

Figure 23 plots the *64* individual H_2, and H_3 fiber length results for each H_1 group and the trendline of the mean fiber length within each group.

---

[17] Minimization in the context of the different networks dimensioned by the MSL network model, not absolute mathematical minimization for all possible such networks
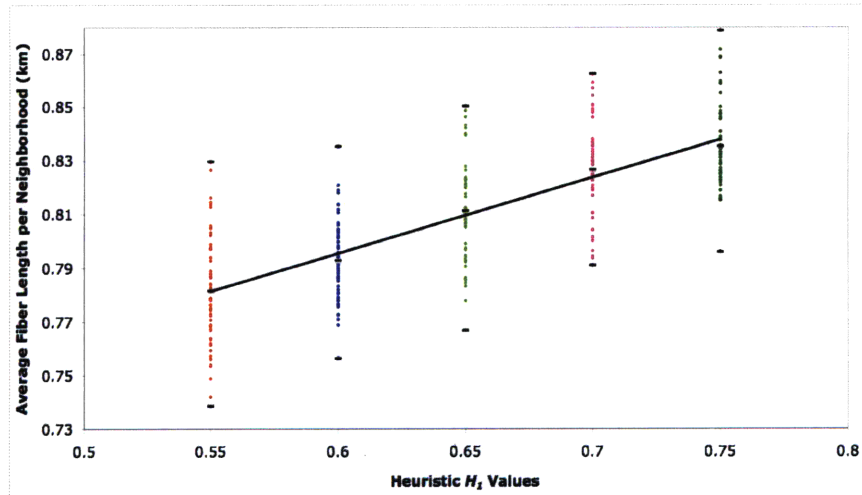
Figure 23: Fiber length values for all heuristic value combinations by H_1 group

Recall that H_1 controls the maximum available reach from CO to subscriber. Therefore, we would expect the total fiber length to increase with increasing H_1 values, resulting in a corresponding increase in the average fiber length per location. This explains the upward trending behavior we observe in Figure 23, and suggests a linear relationship between average fiber length and H_1. Within a given H_1 group, the relationship between fiber length and the remaining two heuristics is much more complex, however, all the fiber length variation for all possible H_2 and H_3 combinations falls within the ±10% threshold defined as the maximum range for heuristic robustness. Given an H_1 value form Figure 23, Figure 24 illustrates how different H_2 and H_3 values affect the average fiber length (in km) per subscriber neighborhood. For example, at $H_1 = 0.55$, (the red data points in Figure 23) the area graph outlined in red in Figure 24 illustrates how the average fiber length per neighborhood varies for different H_2 and H_3 values.

63
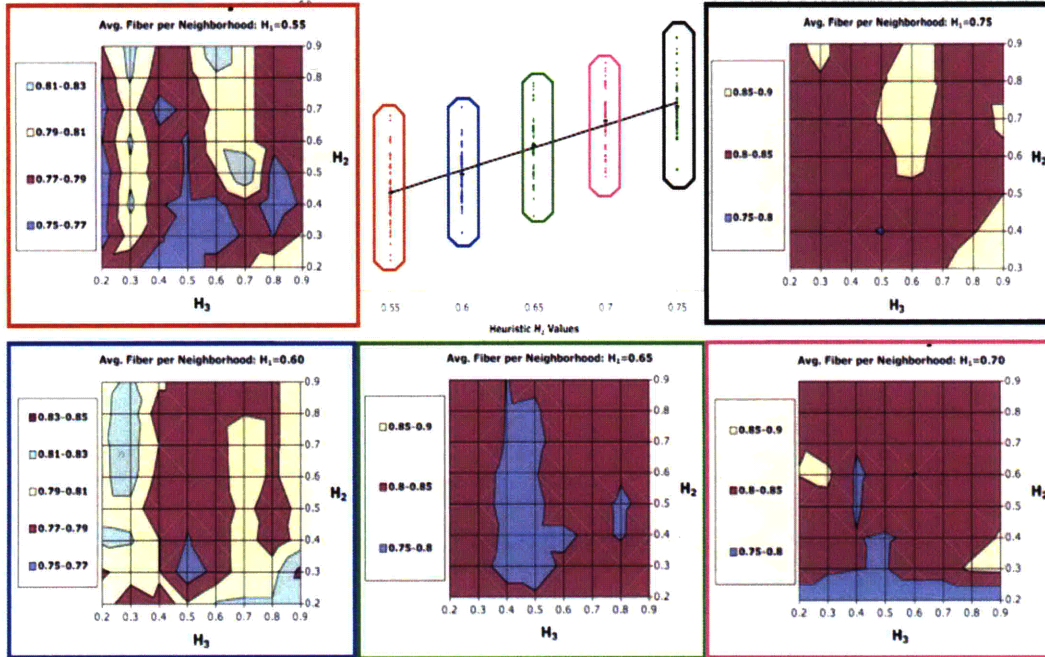
Figure 24: H_2 and H_3 average fiber per neighborhood area plots per H_1 group

The minimum fiber length in Figure 23 corresponds to the $H_1=0.55$ group; however, this group also exhibits significant intra-group fiber length variation. Because we are seeking model stability as well as fiber minimization, we instead select the $H_1=0.60$ group, (the blue data points and corresponding blue outlined area plot in Figure 24) as it exhibits tighter clustering, and therefore smaller variation, around the resulting mean fiber length. Once the $H_1=0.60$ group is selected, the H_2 and H_3 values exhibiting the smallest average fiber length per neighborhood (the minimum fiber length in the blue $H_1=0.60$ group) in the corresponding area plot are chosen. Table 23 provides the final heuristic value combination and population served, while Figure 25 illustrates their locations in the relevant H_2, H_3 area plot.

H_1	H_2	H_3	Locations Served
0.6	0.3	0.5	99.5%

Table 23: Final heuristic values
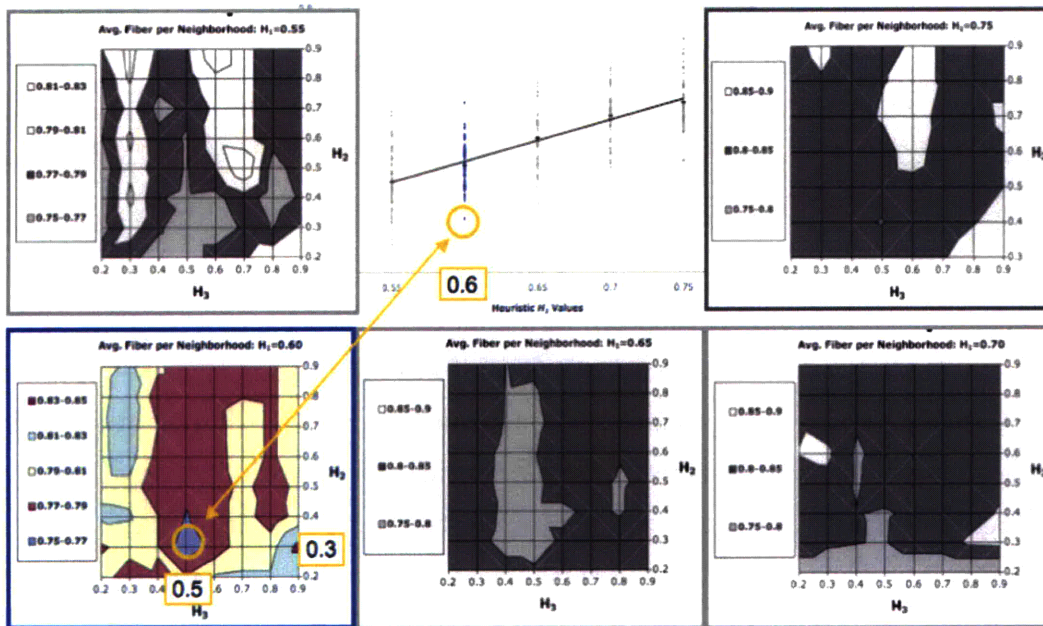
Figure 25: Final H_1, H_2, H_3 values

These heuristic values are used in all subsequent analyses.

8 Cost Models

Two cost models were created to capture the capital and operational expense (CapEx & OpEx) tradeoffs corresponding to different technology choices and population demographics. The CapEx model is comprised of an extensive database of component and installation costs which are mapped to the virtual network architectures emerging from the MSL network model. Capabilities also include build and penetration modeling and the ability to account for legacy conduit and fiber. The OpEx model database is populated with data collected both at the component level, through manufacturers, and at the operations level, through interviews and questionnaires with carriers currently operating fiber networks. This two-prong data collection approach enables detailed characterization of both *intrinsic* OpEx drivers, including manufacturing and/or materials related component failures, and *extrinsic* factors specific to the operating environment, such as fiber breakage and sag statistics or labor rates. Table 24 presents an overview of the architecture cost categories and constituent components considered by the cost models.

CO: Access	CO: Backhaul	Fiber Related	Non-CO Hardware	Customer Premises
OLT Cards/Racks	Transponder Cards	Buried Install	Splitters	ONT
Power Supplies	XFP Modules	Aerial Install	Amplifiers	Set-top box
Software	Tunable Modules	Fiber Bundles	Splicing	Install
Switch Core Cards	Mux shelves	Conduit	Enclosures	
Mgmt. Cards	Racks			
Facility	Power Supplies			
Mux shelves	Software			
	OAF Modules			

Table 24: Cost model network element categories and components

The cost models employ several parameters to characterize factors which can significantly impact network costs. The first of these, *fiber plant legacy*, adjusts each fiber length segment to ensure that re-used existing conduit and fiber is not factored into fiber plant costs. The next parameter, *fiber installation method*, determines the percentage of each fiber segment installed using existing telephone (or other) pole infrastructure, ("aerial" installation), or underground in conduit, ("buried" installation, which can be three or four times the cost of aerial installation). The third parameter, *technology up-charge*, enables characterization of the uncertainty surrounding future technologies for each network cost element, (central office transmission equipment, amplifier technology, etc.) via additional costs assigned to these elements. The *initial build percentage* parameter defines the maximum population subscriber percentage the carrier initially builds to service. For example, an initial build of *100%* means that the initial network deployment includes all network equipment and fiber plant, (less the final "drop" or frontage fiber component connecting the final splitter stage to a customer location) required to reach every household in the coverage region. The final parameter, *penetration* or *take-up percentage*, defines the percentage of the population initially built for subscribing to service. For example, if the initial build percentage is 50%, then a penetration percentage value of 100% means that all 50% are subscribers.

8.1 CapEx Model

The CapEx database utilizes data from component manufacturers and network operators to converge on realistic estimates of equipment and installation-related costs. The inputs were made possible by the close working relationships academia enjoys with these groups, and particularly as the result of input from members of the MIT Center for Integrated Systems Optical Broadband

Working Group, (OBBWG) and others including BT, JDSU, Telecom Italia, Motorola, Finisar, Neophotonics, Corning, Alphion and Deutsche Telecom. Figure 26 provides a CapEx model overview.
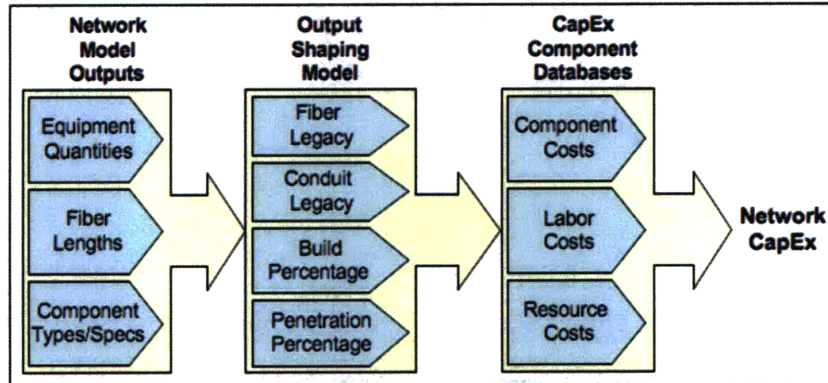
Figure 26: CapEx model overview

8.2 OpEx Model

The OpEx model combines industry-derived component failure statistics with field data from current optical network operators to construct a statistical operations model and extensive cost database. The two-tired approach provides two benefits: characterization of multiple component failure modes identified through extensive testing during manufacturing and real-world operating conditions; and characterization of failure rates for these modes. Meaningful data on many of these alternate failure modes, fiber dig up rates for example, can only come from the field, and each network component can involve multiple such OpEx parameters. OpEx accounting is divided into three main categories: energy, labor, and materials. Table 25 presents an overview of the cost categories and constituent components considered by the OpEx model.

Energy	Labor/Rent	Materials
OLT Power Supplies	Repair	Fiber Plant
Software/Computing	Replace	CO Access
CO Cooling	Maintenance	CO Backhaul
Amplification	Transportation	Non-CO Hardware
Transportation	Operations	Customer Premises

Table 25: OpEx cost categories and components

Figure 27 presents an overview of the OpEx modeling methodology.

67
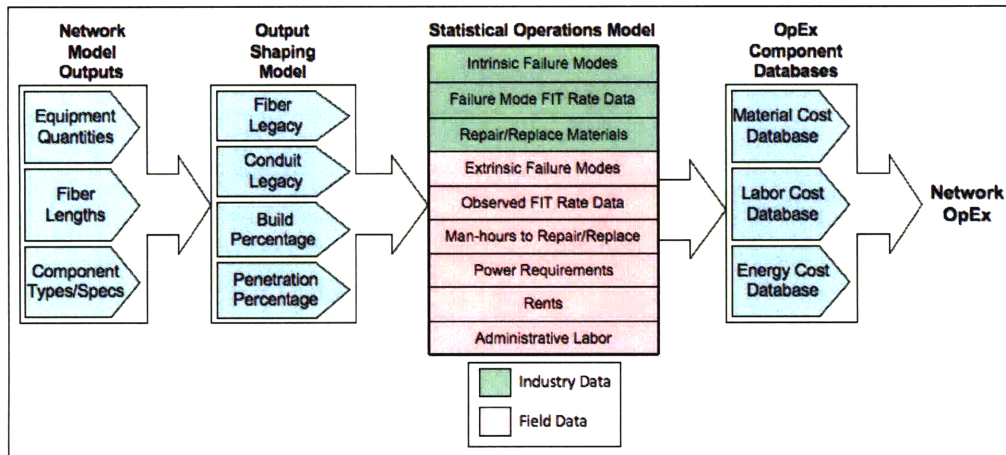
Figure 27: OpEx model overview

Interviews with carriers and municipalities already operating optical networks enabled population of these categories. These interviews provide invaluable information regarding exogenous OpEx drivers. For example, one municipal service provider contributing to this research is the Jackson Authority in Jackson, Tennessee. Table 26 provides the network architecture data, while Table 28 and Table 27 provide the population and data demographic data for this network.

Parameter	Value
Homes Passed	27,460
Penetration	30% (6,897)
Total Data Tiers	12
Network Architecture	EPON w/powered splitters
Network Type	Point to point overlay
CO Quantity	2
Reach	10km
Installation Profile	80% Aerial, 20% Buried
Splitter Strategy	Cascaded
Split Ratio	1:12 powered, 1:8 optical
Multiplexing Strategy	TDM

Table 26: Sample municipality technology and population parameter values (Kersey 2006)

Parameter	Value
Single Family Homes	83.6%
Multi-unit Dwellings	16.4%
Average Distance to Splitter	~3km
Average Population Density	38 homes per route km[18]

Table 27: Population demographics (Kersey 2006)

Data Tier	Downstream Data Rate (Mbps)	Upstream Data Rate (Mbps)	Customer Percentage
1	0.512	0.256	41.74
2	1.5	0.256	13.15
3	2	0.384	0.69
4	2	2	0.93
5	3	0.256	4.09
6	4	0.384	23.34
7	4	4	1.21
8	6	0.384	12.48
9	6	6	0.26
10	10	1	2.09
11	12	12	0.028
12	15	15	0.013

Table 28: Data demand demographics (Kersey 2006)

Table 29 presents average outage statistics and causes, Table 30 lists the personnel and positions required to operate and maintain the network, and Table 31 provides some average OpEx figures for energy, rents, labor etc.

Outage Statistic	Quantity
Average Outages	2 per month
Average Outage Duration	45 minutes
Average Man-Hours per Outage	1.5 (2 technicians)

Table 29: Outage statistics (Kersey 2006)

---

[18] This linear population density is different than the homes/km^2 metric used to develop populations in §5.1

69

Position	Quantity
Operations staff	1
CO Technicians	1
Network Provisioning	2
Field Technicians	7
Install Contractors	8
Total Staff	20

Table 30: Network labor requirements (Kersey 2006)

OpEx Driver	OpEx
Energy Usage	$8,000/month
CO Land Rental Fees	$1,500/month
Average Repair Labor (includes truck, salary & benefits)	$112/outage
Average CO Operations Labor	$6500/month

Table 31: Average network OpEx values (Kersey 2006)

As these tables show, interviews provide detailed information about all facets of network operations, and, when coupled with industry component data, provide a high-resolution picture of practical operational costs.

An additional benefit of real operational data is that it can add extra, unforeseen cost dimensions. For example, our interviews revealed that this region frequently encounters extreme weather including tornadoes, and is located near an active fault line. One interesting OpEx consequence is that aerial portions of the fiber plant (80% of the total) are subject to extreme temperature oscillations, which induces fiber sag and deformation on a macro scale[19]. The resulting increased outages significantly impact labor and materials related OpEx.

9 Methodology Validation, Demonstration, and Limitations

Constant technological/efficiency improvements and/or economies of scale at the component level make meaningful model-result comparisons at the cost level difficult. While cost will be a primary driver to choose between technologies, the required costs for a given network are determined by the underlying network architecture, which in turn is determined by technology choice and the population to be served. Models which can replicate this structure, by specifying a

---

[19] An interesting CapEx consequence is the construction, at significant additional expense, of a fortified concrete bunker to house transmission equipment.

technology choice and accurately characterizing the population demographics of the coverage region, should produce a basis for "apples-to-apples" network cost comparisons over time.

9.1 Network Model Internal Consistency Analysis

One research goal is to characterize how *changes* in population demographics impact technology choice via network cost. Therefore, it is important to ensure the network model produces consistent results when modeling *similar* population demographics for a *fixed* technology choice. This requires that the model produce consistent results for identically parameterized population demographic distributions. This chapter describes the test used to establish model internal consistency and presents the results.

The analysis uses the technology choice provided in Table 21

To be internally consistent, for a fixed population, the network model should predict the same amount of fiber to reach a specified population. However, because we are generating populations by sampling from probability distributions, the specific geographic location of each subscriber will be different for each population, even when the distribution parameters are the same. As a result, we expect minor variations in the dimensioned fiber lengths even for identically parameterized population distributions. This section describes the method used to identify and characterize these variations, and the test used to establish if the model is internally consistent with respect to these variations.

9.1.1 Methodology

Two parameters were selected which characterize the population: the standard deviation, σ, which provides a measure of geographic population size; and the total locations per population, n[20]. Four σ and five n values were selected which characterize a large range of possible populations. The specific parameter values chosen are listed in Table 32, and Figure 28 illustrates three populations corresponding to three such combinations.

---

[20] All other population parameters, (e.g. household size density and data demands) are held constant for all populations

Parameter	Symbol	Minimum	Maximum	Step	Total Values
Standard Deviation	σ	5km	20km	5km	4
Total Population Size	n	50000	250000	50000	5

Table 32: Parameter values modeled
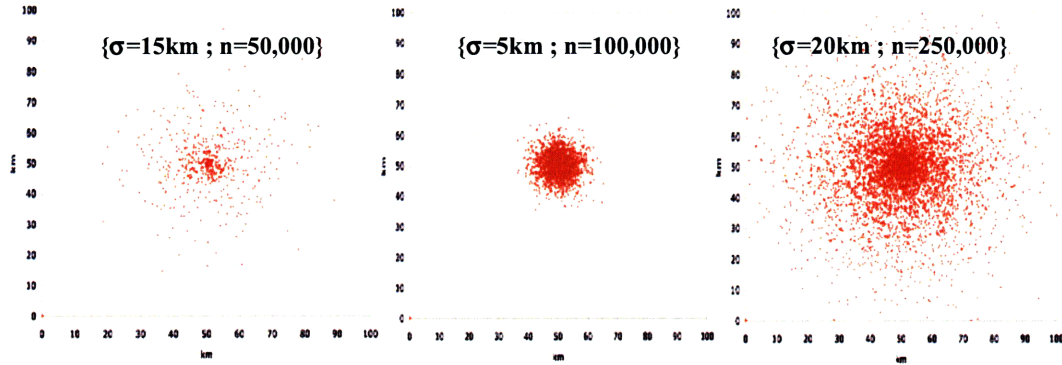
Figure 28: Example modeled populations

For each {σ, n} combination, ten populations were created by randomly sampling from a normal population probability distribution, *f(r),* (see §5.1). The network model was then run for each population, and the resulting fiber lengths recorded, and the mean fiber length determined for each of the 20 {σ, n} combinations, (4 standard deviations x 5 population sizes). The criterion for internal consistency is that all fiber length results corresponding to a particular {σ, n} combination must be within *± 10%* of the mean result for that combination.

9.1.2   Results and Analysis

Table 33 and Table 34 present the test results for all 200 model runs, reporting the maximum, minimum, and mean fiber length results for each, as well as the maximum percentage difference both above and below the mean for each case.

σ	5 km					10 km				
n	50k	100k	150k	200k	250k	50k	100k	150k	200k	250k
Max Fiber (km)	794	1100	1647	1800	2041	1100	1871	2493	3066	3168
Min Fiber (km)	761	1064	1572	1715	1958	1015	1773	2301	2979	2902
Mean Fiber (km)	778	1080	1609	1758	2005	1055	1829	2422	3011	3000
Max Diff.	1.9%	1.7%	2.3%	2.3%	1.8%	4.1%	2.2%	2.8%	1.8%	5.3%
Min Diff.	4.2%	3.3%	4.6%	4.7%	4.1%	7.7%	5.2%	7.7%	2.8%	8.4%

Table 33: Test results for σ =5km and σ =10km scenarios

σ	15 km					20 km				
n	50k	100k	150k	200k	250k	50k	100k	150k	200k	250k
Max Fiber (km)	1599	2288	3096	3991	4289	3199	4153	4692	5409	3199
Min Fiber (km)	1507	2104	2916	3801	4073	3008	3802	4611	4990	3008
Mean Fiber (km)	1560	2218	3022	3891	4137	2111	3105	3991	4656	5188
Max Diff.	2.4%	3.1%	2.4%	2.5%	3.5%	3.8%	2.9%	3.9%	0.8%	4.1%
Min Diff.	5.8%	8.0%	5.8%	4.8%	5.0%	8.2%	6.0%	8.5%	1.7%	7.7%

Table 34: Test results for σ =15km and σ =20km scenarios

The results suggest that the model is internally consistent with respect to these parameters for the criterion set: no fiber length varied by more than ± 10% of the mean value of the corresponding {σ, n} combination. Table 35 presents the summary statistics for all 200 runs.

Parameter	Value
Average Difference Above Mean	2.9%
Average Difference Below Mean	5.8%
Maximum Difference Above Mean	5.3%
Maximum Difference Below Mean	8.5%

Table 35: Summary statistics, internal consistency test

9.2 Network Model Benchmarking

Benchmarking models against real-world networks provides a way to ground results: modeling provides little insight if all models dimension the same unrealistic network architecture. To this end, two independent and unaffiliated network-dimensioning studies were selected against which to benchmark and validate the network model: (1) a study performed by British Telecom (BT) examining the costs required to convert an existing coverage region from copper to a GPON fiber-to-the-home network; (2) a 2004 Corning study which explores the value of consolidating central office equipment. The studies were selected because (a) both were performed by experts in the field with access to a wealth of real-world data, and (b) each serves a different population demographic and utilizes different network architectures to do so, enabling network model prediction performance over a range of scenarios. This section describes these studies and how we emulated the corresponding architecture and populations, and then compares the fiber lengths and equipment required to serve each population. In both cases, the network model was run utilizing the final, calibrated heuristic set.

9.2.1 BT Validation Study

The first validation exercise benchmarks the MSL network model against a real-world coverage region for which the network topology is known. Two metrics are used: the installed fiber length and splitter quantity.

9.2.1.1 Technology Modeled

Table 36 provides the technology parameters and values modeled.

74

Parameter	Value
Total Households	6,247
Total Network Reach	10km[21]
Splitter Strategy	Single Stage
Splitter Port Count	1:32
Max Data Rate per Customer	80Mbps
Maximum Data Rate per PON	2.5 Gbps (GPON)
Build type	Overlay

Table 36: BT exchange parameters

9.2.1.2 Population Modeled

The exchange population characteristics are provided in Table 37.

Distance Region	Distance Range (km)	Households
1	0 to 1	1,750 (27%)
2	1 to 2	1,250 (20%)
3	2 to 3	2,437 (39%)
4	3 to 4	810 (14%)

Table 37: BT exchange population characteristics

To prepare this data for use in the network model, the discrete populations in Table 37 were converted to a continuous population distribution function (PDF). First, a cumulative distribution function (CDF) mimicking the given household density data Table 37 was developed. From this CDF, the corresponding (PDF) was estimated[22]. Of the many distribution types examined, the beta distribution with shape and scale parameters $\alpha = 1.75$ and $\beta = 2$ provided the best fit to the data provided. Figure 29 provides (a) the cumulative distribution estimated from the values in

---

[21] Provided by British Telecom, not derived from component data

[22] Using the relationship PDF(r) = d[CDF(r)] / dr

Table 37 and the corresponding MSL fitted curve, and (b) the resulting beta PDF.
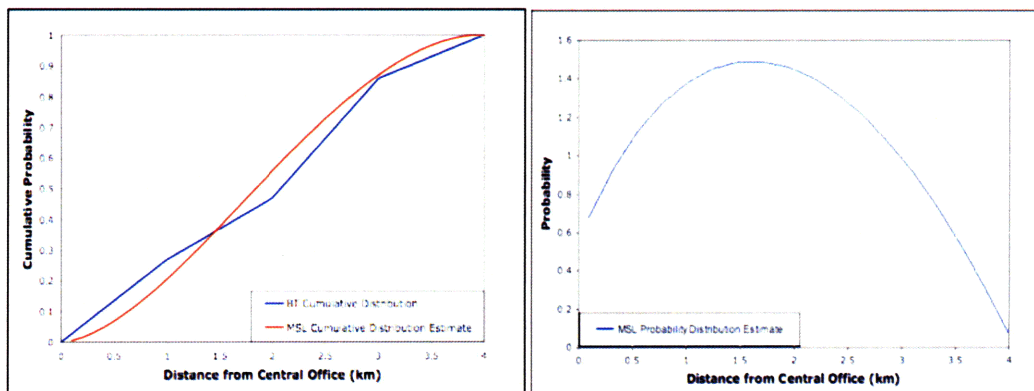
(a)   (b)

Figure 29: (a) BT and fitted MSL CDFs, (b) resulting PDF estimate

9.2.1.3   Results and Analysis

Table 38 compares the total installed fiber length and splitter quantity metrics predicted by the MSL network model against the corresponding actual values required.

Parameter	BT Results	MSL Model	Accuracy
Total Installed Fiber Length (km)	758	717	95%
Splitters Deployed	263	294	89%

Table 38: MSL and BT network metric comparison

The results suggest that the MSL model does a good job dimensioning small coverage regions, even in cases where the topology is based not on mathematical optimization but rather on knowledge of existing conduit paths and geographic constraints. The ±10% discrepancy in both metrics suggests that the MSL model tends to over-optimize slightly, when viewed from the perspective of installed fiber minimization. This is explained by the model's priority on co-locating installed fiber as far as possible from the central office prior to splitting (irrespective of existing geographic constraints) and the inclusion of as many households as possible at each splitter site (illustrated by the larger splitter quantity predicted by the MSL model).

9.2.2   Corning Study Validation

The second validation study, "Value of Reach-and-Split Ratio Increase in FTTh Access Networks" (Vaughn, Kozischek et al. 2004) assesses whether consolidating central office

76

equipment via increased splitter ratios results in CapEx savings for a given population distribution, when compared with a reference network. This two-case study covers a much larger population than the BT exchange and utilizes a two-stage, cascaded splitter architecture.

9.2.2.1 Technology Choice Modeled

The total network reach is 20km in both cases, and each utilizes a two-stage splitter cascade. The data rate, while not explicitly provided, is assumed to be ~1Mbps for both architectures. The basic FTTh PON schematic for this study is provided in Figure 1.
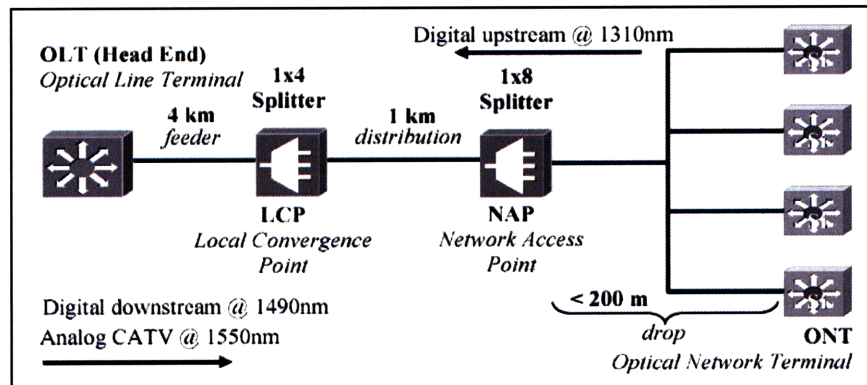
Figure 30: Corning FTTh PON (Vaughn, Kozischek et al. 2004)

As Figure 1 shows, the first splitter stage is referred to as the "local convergence point," (LCP) while the second stage is named the "network access point" (NAP). We will adopt the same terminology for this first validation exercise. Fiber length constraints and parameter values for the study are provided in Table 39.

Fiber Parameter	Value
Fiber topology	Point-to-point
Fiber installation type	Aerial
Distribution fiber max. length	0.87 km
Drop/frontage max. length	60 m
Build type	Greenfield

Table 39: Corning fiber length constraints and parameter values (Vaughn, Kozischek et al. 2004)

9.2.2.2 Population Modeled

The total homes passed in the study is *71,331* all of which are served by headend equipment located in a single central office in the consolidated case, but divided among three central offices

in the reference case. Table 40 provides the number and percentage of households served by each office in each scenario (Vaughn, Kozischek et al. 2004).

Parameter	Reference Network	Consolidated Network
Total Households	71,331	71,331
CO A: Households Assigned	30,089 (42.2%)	71,331 (100%)
CO B: Households Assigned	29,630 (41.5%)	0
CO C: Households Assigned	11,612 (16.3%)	0

Table 40: Households by central office (Vaughn, Kozischek et al. 2004)

The distribution of households in the coverage region is divided by reach into seven sub-distributions served by three central offices containing all headend equipment in the reference case, and eight sub-distributions with headend equipment located in a single central office in the consolidated network. Each sub-distribution represents a range of household distances from the closest central office, and therefore contains a fraction of the total households assigned to each office. The feeder fiber length (from central office to the first splitter stage) is constant for each distance range. Table 41 and Table 42 provide information about each sub-distribution for the reference and consolidated cases respectively, including the beginning and ending distance from the central office, the total households within this distance range from each office, and the feeder fiber lengths assigned to each sub-distribution (Vaughn, Kozischek et al. 2004).

Distribution Region: i	Distance Range (km)	Feeder Length (km)	Households (%) Assigned CO A	Households (%) Assigned CO B	Households (%) Assigned CO B
1	0 to 3	0.59	18,053 (60.0)	17,778 (59.1)	6,967 (23.2)
2	3 to 3.7	2.44	2,708 (9.0)	2,667 (8.9)	1,045 (3.5)
3	3.7 to 5	3.44	3,972 (13.2)	3,911 (13.0)	1,533 (5.1)
4	5 to 5.5	4.34	1,444 (4.8)	1,422 (4.7)	557 (1.9)
5	5.5 to 7.9	5.79	1,504 (5.0)	1,482 (4.9)	581 (1.9)
6	7.9 to 10.4	8.24	1,504 (5.0)	1,482 (4.9)	581 (1.9)
7	10.4 to 12.2	10.39	903 (3.0)	889 (3.0)	348 (1.2)

Table 41: Household distributions, reference case (Vaughn, Kozischek et al. 2004)

Distribution Region: i	Distance Range (km)	Feeder Length (km)	Households (%) Assigned CO A
1	0 to 1.5	0.59	18,053 (25.0)
2	1.5 to 3.35	2.44	2,708 (4.0)
3	3.35 to 4.45	3.44	10,016 (14.0)
4	5 to 5.25	4.34	4, 289 (6.0)
5	5.25 to 6.7	5.79	13,949 (20.0)
6	6.7 to 9.15	8.24	12,078 (17.0)
7	9.15 to 11.3	10.39	5.149 (7.0)
8	11.3 to 16.1	15.19	5,088 (7.0)

Table 42: Household distributions, consolidated case (Vaughn, Kozischek et al. 2004)

To validate the MSL network model, we again converted the discrete household distribution data into two independent PDFs, one for the reference case and the other for the consolidated network architecture. We then ran the model, and compared the installed fiber lengths and splitter quantities required to dimension the resulting networks against value ranges extracted from the reference and consolidated case studies. As in Figure 29, the reference case household distribution was emulated by curve fitting CDF to match the given household percentages by distance provided Table 41 and Table 42. Once a good fit was determined, the corresponding PDF was estimated. Figure 31(a) and Figure 32(a) provide the cumulative distribution estimated from the given values and the MSL estimated best-fit CDF, for the reference and consolidated cases respectively, while Figure 31(b) and Figure 32(b) provide the resulting estimated PDFs, used to model the population used in the validation exercise. The best-fit PDF for the reference case population was determined to be a gamma distribution with shape and scale parameters, $\kappa = 0.604$, and $\theta = 3$, while the consolidated case population was fit with a normal distribution with mean and standard deviation, $\mu = 5.2km$ and $\sigma = 5.1km$. These PDFs were used as the population distribution function inputs for the corresponding network model runs.

(a)   **(b)**
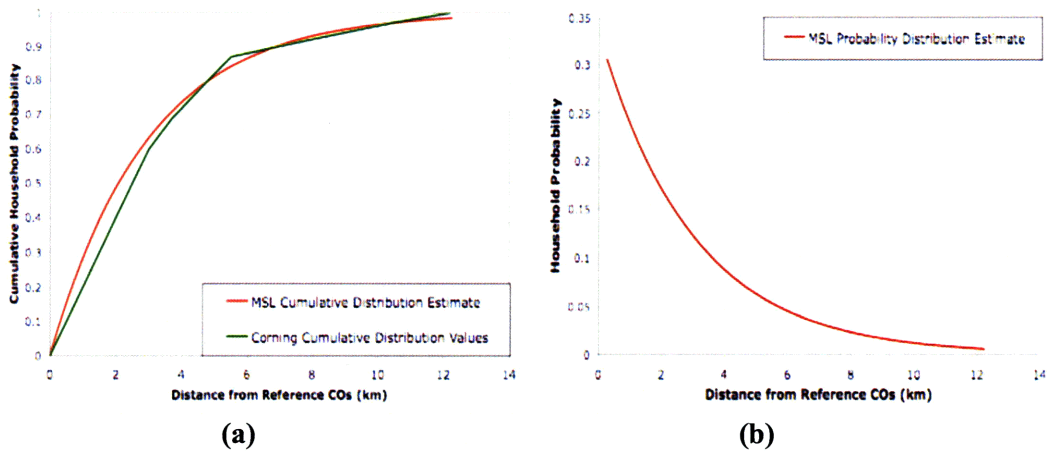
Figure 31: Reference case, (a) Corning and MSL cdf estimates (b) resulting pdf

(a)   **(b)**
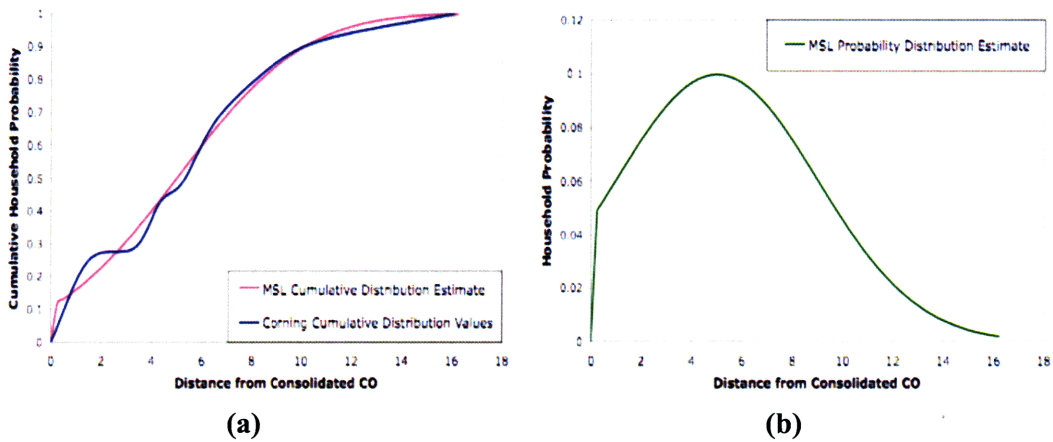
Figure 32: Consolidated case, (a) Corning and MSL cdf estimates (b) resulting pdf

9.2.2.3 Methodology

Meaningful comparison of the resulting network designs requires information about the total installed fiber length required to service the coverage region. Although these values were not explicitly stated in the Corning study, adequate information was provided to infer a range of values for each comparison metric. Therefore, we made some simplifying assumptions to estimate the minimum and maximum fiber length values for both the consolidated and reference cases to use as a basis for comparison against the network model results.

In these cases, (see Figure 30): the *feeder* fiber connects the CO to the LCP, (first stage) splitters; the *distribution* fiber connects the LCP and NCP splitter stages; and the *drop* fiber links the LCP sites to subscriber locations (see §1.1 for a discussion of the fiber length definitions). Estimating

80

the total fiber plant therefore requires characterizing how LCP and NAP sites and the population are distributed within the coverage region.

The estimation process began by estimating how many LCP, (first stage) and NCP, (second stage) splitters were utilized in the reference and consolidated cases. Table 43 provides the inferred splitter ratios for both Corning network architectures (Vaughn, Kozischek et al. 2004). The total homes per LCP site is 280, and it is important to note that each NCP splitter fiber serves *two* households, (Vaughn, Kozischek et al. 2004).

Distance from CO (km)	Reference Network			Consolidated Network		
	LCP	NCP	Homes	LCP	NCP	Homes
0-5	(1:8)	(1:4)	64	(1:16)	(1:4)	128
5-20	(1:4)	(1:4)	32	(1:8)	(1:4)	64

Table 43: Corning study splitter ratios (Vaughn, Kozischek et al. 2004)

Once the splitter ratios are established, the distribution of splitter sites as a function of distance from the CO is determined in two steps. First, because the split ratio changes at a distance of five kilometers from the CO, (Table 43) the total households for each CO are coarsely divided into those less than five kilometers from the CO and those further away. Because the total initial splitter utilization is given as 75%, the *estimated* total LCP and NAP splitters required are determined for each region by dividing the population in each of the two distance regions by 1.25 times the splitter ratio in each region[23]. Table 44 presents these results.

---

[23] The empty splitter ports are left available either to connect additional customers as the penetration rate increases over time, or to re-route customer connections during repairs/maintenance

81

Central Office	Corning Architecture	Distance from CO (km)	Total Households	Total Splitters	
				LCP	NCP
CO A	Reference	0-5	24,733	445	3,560
		5-20	5,355	100	400
CO B	Reference	0-5	24,356	435	3,480
		5-20	5,275	95	380
CO C	Reference	0-5	11,612	175	1,400
		5-20	2,067	40	160
CO A	Consolidated	0-5	30,672	550	8,800
		5-20	40,659	730	5,840

Table 44: Total *estimated* splitters required per CO by distance from CO

Once the total splitters are estimated, the second step further refines the splitter distribution, estimating the total LCP and NAP splitters per distance region, i, for each CO defined in Table 41 and Table 42, (totaling 7 for each of the three COs in the reference case, and 8 for the single CO in the consolidated case). Table 45 provides this breakdown.

Distance Region: i	Splitters CO A (reference)		Splitters CO B (reference)		Splitters CO C (reference)		Splitters CO A (consolidated)	
	LCP	NAP	LCP	NAP	LCP	NAP	LCP	NAP
1	327	2,136	318	2,088	129	840	320	2,200
2	50	321	48	314	20	126	52	352
3	71	463	69	453	28	182	180	1,232
4	28	178	27	174	11	70	77	528
5	28	178	27	174	11	70	256	1,760
6	28	178	27	174	11	70	218	1,496
7	17	107	16	105	7	42	90	616
8	--	--	--	--	--	--	90	616

Table 45: Total *estimated* splitters by population distribution region, *i*

Once the total splitters are determined for each distance region for each CO, we make some simplifying assumptions about their spatial distribution and co-location to determine estimates of fiber length.

The feeder fiber distance for each distance region, i, defined in Table 41 and Table 42, determines the distance from the CO to all LCP splitter sites in this region. However, we do not know how many splitters may be co-located at a given site in this region, a necessary element to calculate the total feeder fiber distance. Consultations with industry suggest that *50* and *200* are reasonable estimates of the minimum and maximum splitters (either LCP or NAP) co-located at a given splitter site. Maximum co-location (obtained by dividing the total LCP splitters in region i by the maximum) results in the fewest LCP splitter sites, which requires the fewest independent feeder fibers, and therefore the minimum installed feeder fiber length. Conversely, minimum co-location requires a maximum number of splitter sites, each requiring independent fiber paths, resulting in maximum total installed feeder fiber length. Table 46 provides an example of the minimum and maximum LCP splitter sites, and corresponding total feeder fiber length per distance region i, for CO A in the reference case, (using the installed feeder fiber per splitter site values in Table 41, and the total estimated splitters per distance region in Table 45).

Region (i)	Feeder per splitter site (km)	Splitters Sites CO A (reference)			
		Total LCP Sites		Total Feeder Length (km)	
		Min	Max	Min	Max
1	0.59	2	7	1.2	4.1
2	2.44	1	1	2.4	2.4
3	3.44	1	2	3.4	6.9
4	4.34	1	1	4.4	4.4
5	5.79	1	1	5.8	5.8
6	8.24	1	1	8.2	8.2
7	10.39	1	1	10.4	10.4
Totals		8 sites	14 sites	35.8 km	42.2 km

Table 46: LCP splitter sites and feeder fiber lengths per distance region i: COA reference case

To determine the distribution and drop fiber lengths, we treat each distance region, i, as a ring about the central office with inner and outer radii, r_i^{out} and r_i^{in}, given by the upper and lower bounds of the corresponding distance range value given in the second columns of Table 41 for the reference case, and Table 42 for the consolidated case. We assume that the population is uniformly distributed within each ring, enabling the calculation of a radius, r_i^*, which evenly divides the household population. Next, we assume that half the NAP (second stage) splitter sites in the ring are located in the region, $r_i^{out} - r_i^*$ and the other half in the remaining region, $r_i^* - r_i^{in}$. When assigning splitter sites to each region, a first guess might be that, on average, they are located at the midpoint of each region:

$$\frac{r_i^* - r_i^{in}}{2} \quad and \quad \frac{r_i^{out} - r_i^*}{2} \tag{0.38}$$

Figure 33 illustrates this splitter site assignment scenario, where f_i is the feeder fiber distance for ring i, as defined above.

Figure 33: NAP splitter sites at midpoints of $r_i^{out} - r_i^*$ and $r_i^* - r_i^{in}$ for single ring i of uniform population density for a generic CO

In general, the total distribution fiber for ring i with total NAP splitters sites, NAP_i, is therefore given by:

$$\left(r_i^{in} - f_i \right) \cdot \left[\frac{NAP_i}{2} \cdot \left(\frac{r_i^* - r_i^{in}}{2} \right) + \frac{NAP_i}{2} \cdot \left(\frac{r_i^{out} - r_i^*}{2} \right) \right]$$

$$= \frac{\left(r_i^{in} - f_i \right)}{4} \cdot \left[NAP_i \cdot \left(r_i^{out} - r_i^{in} \right) \right] \quad \forall \left(r_i^{in} - f_i \right) > 0 \qquad (0.39)$$

$$= NAP_i \cdot \left[\frac{\left(r_i^{out} - r_i^{in} \right)}{4} - \left| r_i^{in} - f \right| \right] \quad \forall \left(r_i^{in} - f_i \right) \leq 0$$

Dividing the splitter sites in (0.39) by the maximum, "Max," and minimum, "Min," co-located splitters (defined above as *200* and *50*) at a given site provides estimates for the minimum and maximum distribution fiber lengths.

While this appears to be a reasonable approximation to estimate the distribution fiber length, we notice that this splitter site configuration requires that all sites are able serve households up to one-fourth the ring width away $\left((r_i^{out} - r_i^{in})/4 \right)$. For example, the "widest" ring in the analysis corresponds to the final distribution ring in the consolidated case, ($i = 8$) with inner and outer radii, $r_8^{in} = 11.3 km$ *and* $r_8^{out} = 16.1 km$ (Table 42). The corresponding drop distance is therefore given by:

$$\frac{r_i^{out} - r_i^{in}}{4} = \frac{4.8km}{4} = 1.2km \qquad (0.40)$$

This value significantly exceeds the ~200-meter drop distance limit defined in the Corning paper (Figure 30). Intuitively, we would expect smaller drop distances, as in reality splitter sites will be distributed more sparsely throughout the coverage region. To characterize this distribution, we introduce a multiplicative factor, k, $(0 \leq k \leq 1)$ which determines how far, on average, NAP splitter sites are located from the $(r_i^* - r_i^{in})/2$ and $(r_i^* - r_i^{in})/2$ midpoints. For example, a value of $k=0.5$ results in one fourth of the total splitter sites located at $\pm(r_i^* - r_i^{in})/4$ and $\pm(r_i^{out} - r_i^*)/4$. Figure 34 illustrates (b) how this k value affects NAP splitter site siting relative to (a) the case where $k=0$ (Figure 33) for a single ring i of uniform population for a generic CO.
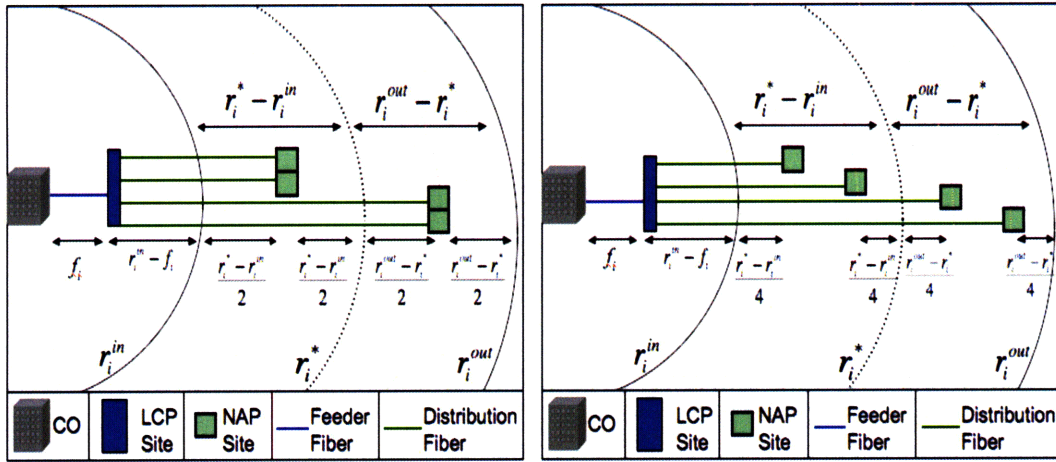
Figure 34: (a) $k=0$ and (b) $k=0.5$ NAP splitter siting and feeder and distribution fiber for single ring i of uniform population corresponding to a generic CO

While this scenario results in the same total distribution fiber length:

$$\left(r_i^{in} - f_i\right) \bullet \left[\left(\frac{NAP_i}{4}\right) \bullet \left(\frac{r_i^* - r_i^{in}}{4} + \frac{3\left(r_i^* - r_i^{in}\right)}{4} + \frac{r_i^{out} - r_i^*}{4} + \frac{3\left(r_i^{out} - r_i^*\right)}{4}\right)\right]$$

$$= \frac{\left(r_i^{in} - f_i\right)}{4} \bullet \left[NAP_i \bullet \left(r_i^{out} - r_i^{in}\right)\right] \qquad \forall \left(r_i^{in} - f_i\right) > 0 \qquad (0.41)$$

$$= \left(\frac{NAP_i}{4}\right) \bullet \left[\left(r_i^{out} - r_i^{in}\right) - \left|r_i^{in} - f_i\right|\right] \qquad \forall \left(r_i^{in} - f_i\right) \leq 0$$

Each splitter site is now required to serve households only up to a one-eighth the total ring width away. As a result, the *0.75km* drop distance found in (0.40) is now reduced to:

$$\left(\frac{r_i^{out} - r_i^{in}}{4}\right) \bullet k = \left(\frac{4.8km}{4}\right) \bullet 0.5 = 0.6km \qquad (0.42)$$

We see that, by adjusting *k* we can meet the *200*-meter Corning drop distance constraint:

$$\left(\frac{4.8km}{4}\right) \bullet k = 0.2km \quad \therefore \quad k \cong 0.33km \qquad (0.43)$$

In general, the total fiber length for a population ring *i* is given by the sum of the feeder, distribution, and drop components:

$$[LCP_i \bullet f_i] + \left[\frac{(r_i^{in} - f_i)}{4} \bullet \left[NAP_i \bullet (r_i^{out} - r_i^{in})\right]\right] + \left[\left(\frac{k \bullet N_i}{2}\right) \bullet \left(\frac{r_i^{out} - r_i^{in}}{4}\right)\right] \quad \forall (r_i^{in} - f_i) > 0$$

$$\qquad (0.44)$$

$$[LCP_i \bullet f_i] + \left[\left(\frac{NAP_i}{4}\right) \bullet \left[(r_i^{out} - r_i^{in}) - |r_i^{in} - f_i|\right]\right] + \left[\left(\frac{k \bullet N_i}{2}\right) \bullet \left(\frac{r_i^{out} - r_i^{in}}{4}\right)\right] \quad \forall (r_i^{in} - f_i) \leq 0$$

Where N_i is the households in ring *i*, (divided by two to represent two households per drop fiber, per the Corning analysis) and dividing the LCP and NAP splitter sites by the maximum and minimum co-located splitters, (defined above as *200* and *50*) at a given site provides estimates for the minimum and maximum distribution fiber lengths. The total installed fiber distance for each CO is then just the sum over all rings. Table 47 provides the minimum and maximum estimated NAP splitter sites, the feeder fiber, (calculated in Table 46) and the components required to calculate, and resulting values of, the distribution and drop installed fiber distances for the seven household density rings comprising the COA population for the reference case.

87

Ring (i)	$r_i^{in} - f_i$ (km)	$r_i^{out} - r_i^{in}$ (km)	Total Drop Length (km)	NAP Sites		Total Feeder Length (km)		Total Distribution Length (km)		Total Fiber Length (km)	
				Min	Max	Min	Max	Min	Max	Min	Max
1	-0.59	3.0	2234.1	11	43	1.2	4.1	6.63	25.91	2285.3	2266.1
2	0.56	0.7	78.2	2	7	2.4	2.4	0.20	0.69	96.0	95.5
3	0.26	1.3	213	3	10	3.4	6.9	0.25	0.85	248.2	247.7
4	0.66	0.5	29.8	1	4	4.4	4.4	0.08	0.33	47.5	47.2
5	-0.29	2.4	148.9	1	4	5.8	5.8	0.53	2.11	174.2	172.6
6	-0.34	2.5	155.1	1	4	8.2	8.2	0.54	2.16	190.2	188.6
7	0.01	1.8	67.1	1	3	10.4	10.4	0.00	0.01	98.2	98.2
Totals			2926.1	20	75	35.8	42.2	8.23	32.05	3115.8	3139.6

Table 47: Complete NAP and installed fiber distances for COA reference case

9.2.2.4 Results

Table 48 provides the minimum and maximum estimated total installed fiber lengths for all COs in both the reference and consolidated cases.

Installed Fiber Length (km)	CO A (reference)	CO B (reference)	CO C (reference)	CO A (consolidated)
Minimum Estimate	3,116	3,071	1,212	5,927
Maximum Estimate	3,140	3,085	1,217	6,026
MSL Model	3,118	3,011	1,226	5,981

Table 48: Estimated Corning minimum and maximum total installed fiber lengths and MSL network model fiber length results for all COs

The results show that the MSL model consistently replicates the estimated Corning installed fiber values for all COs across both cases.

9.2.3 Benchmarking Exercise Conclusion

Meaningful discussions around network capital and operational network expenses rely on the ability to dimension network topologies which accurately mimic how technology choice impacts network architecture over a range of population types. The validation exercise results suggest that the MSL network model is capable of replicating the structure of both real world networks, and other modeling results containing considerable real-world data, for a range of technology choices and population demographics. Therefore, we believe this tool provides a strong foundation on which to base network cost comparisons.

9.3 Network Model Limitations

The network model suffers from several limitations. First, because the model focuses on characterizing the relative changes in network topology over a set of technologies, it does not provide an optimized network topology for any single technology. Second, because the effects of legacy fiber plant on network design are characterized in the cost modeling stage, the ways in which legacy may alter topology, and the resulting impacts on technology choice are not considered. Next, the model does not consider geographic-specific constraints affecting network topology, (lakes or mountains for example). While this may significantly alter the resulting network cost estimates for a single technology, it is assumed that these effects would affect the network design for all technologies considered in a similar way. Finally, the network model assumes that the technologies choices modeled are mutually exclusive, that is, once a particular technology is selected, no other technology may subsequently be implemented, and no migration path between technologies is possible.

9.4 CapEx Model Validation

To verify model accuracy when characterizing penetration effects on initial investment, the MSL CapEx results were compared against Corning cost results for a range of penetration values. The CapEx model was also successfully validated against cost data from the British Telecom study; however, these results will not appear in this work for proprietary reasons, (although the corresponding cost data was included both the CapEx and OpEx databases).

Although the Corning networks represent overlays onto existing copper network routes, no specific legacy values were provided for either conduit or fiber. Therefore, the MSL model

results assume no re-usable legacy fiber initially exists in any of the link segments. Additionally, the cost data utilized in the MSL model reflects cost reductions in the period 2004-2009 due to economies of scale and/or technological/manufacturing efficiency improvements. As a result, we expect per subscriber CapEx reductions when compared to the Corning results. For these reasons the resulting costs can not be directly compared; however, the point of the validation exercise is to ensure that network cost behavior for different build and penetration values is as one would expect, ("sanity check") and characterizes the Corning cost trends.

Once the MSL network model has dimensioned both the reference and consolidated networks, the resulting costs were compared as a function of service area penetration for four build values, 10%, 30%, 50%, and 80%, compared with the corresponding scenario results provided in the Corning study (Vaughn, Kozischek et al. 2004). Figure 35 compares the (a) Corning and (b) MSL CapEx per subscriber results for both the reference and consolidated cases (note that the MSL "ref." and "cons." network labels correspond to Corning's "current" and "advanced cons." terminology in Figure 35).

(a)   (b)
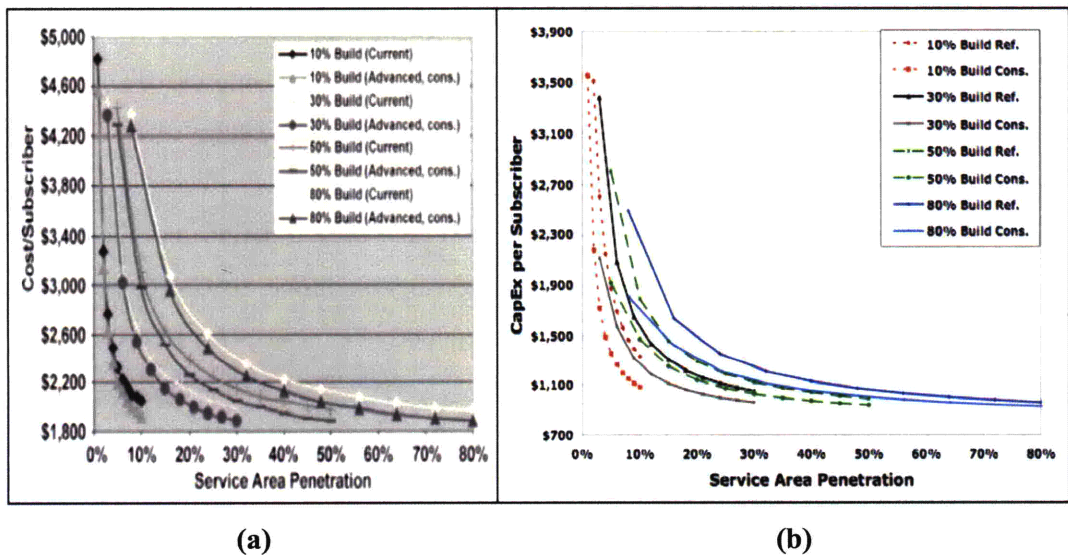
Figure 35: CapEx per subscriber for (a) Corning (Vaughn, Kozischek et al. 2004) and (b) MSL for four initial build values as a function of penetration percentage

The MSL model predicts network costs which mimic the functional behavior of their Corning counterparts but are, in general, less expensive. This is expected due to the reasons defined above. As in the Corning scenarios, the consolidated cases exhibit less CapEx per subscriber over the penetration value range at all build percentages.

90

9.5 CapEx Model Capabilities

In addition to total network and per subscriber costs, the CapEx model identifies the primary network cost drivers for combinations of build and penetration. This additional functionality, when combined with the corresponding cost database, allows for detailed network cost analysis both within and between scenarios.

As an example, Figure 36 provides two CapEx per subscriber breakdowns corresponding to the (a) consolidated, and (b) reference cases in Figure 35 (b) for 30% build and 20% penetration values, using the cost category components defined in Table 14.

(a)   (b)
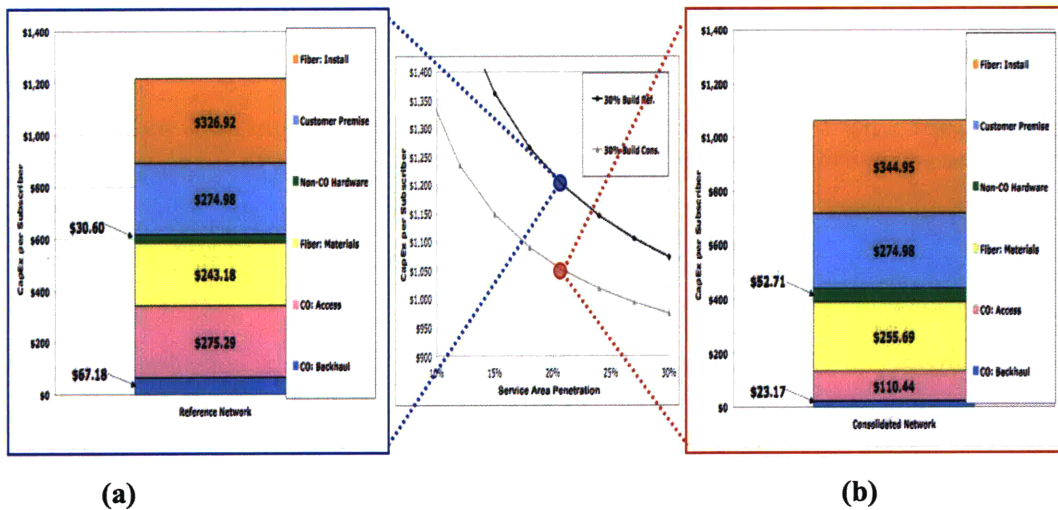
Figure 36: CapEx per subscriber for (a) reference and (b) consolidated cases with parameters build=30%, penetration=20% in Figure 35(b)

This level of resolution enables identification not only of how total network costs compare, but also of subtle cost tradeoffs. For example, although the consolidated case results in additional fiber and non-CO hardware expenses, these are more than offset by the access and backhaul related costs of two additional central offices in the reference network. Alternatively, we can characterize how network cost elements change relative to one another as a percentage of total CapEx per subscriber as a function of penetration for individual build scenarios. For example, Figure 37 examines how the CapEx per subscriber cost composite profile changes as a function of penetration for the (a) reference and (b) consolidated 30% build scenarios from Figure 36.

91

| (a) | (b) |
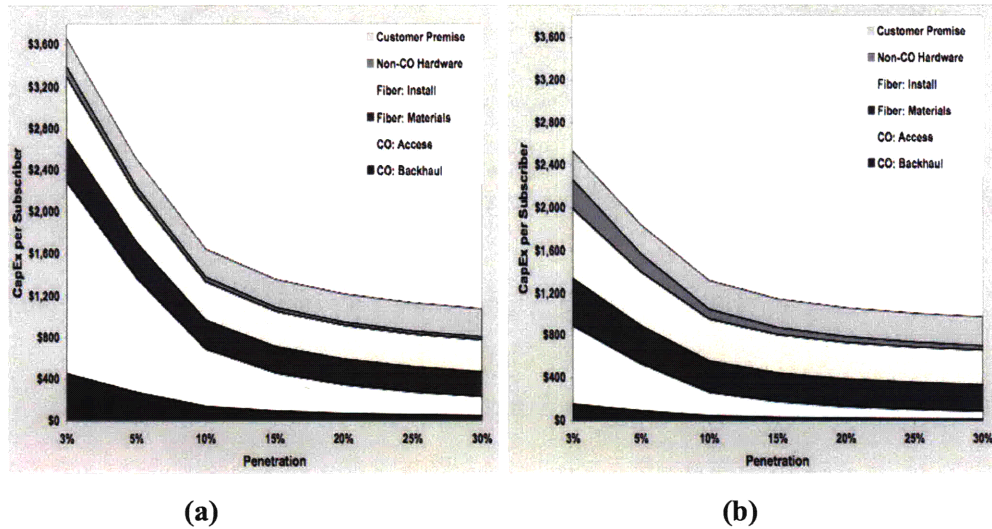
Figure 37: CapEx per subscriber costs vs. penetration by category at 30% build for (a) reference and (b) consolidated scenarios

This view highlights the dynamic tradeoffs which occur between CapEx components corresponding to technology choice, and suggest where potential opportunities lie to mitigate cost. For example, in the reference case backhual-related equipment is located in every central office, resulting in considerable additional cost when compared with the consolidated case. Once way to reduce these expenses is by restricting some central offices to only access and CO to CO functionality, thereby significantly reducing backhaul-related costs. Examining CapEx per subscriber by individual cost category can more clearly identify where consolidation or reconfiguration strategies can reduce network cost Figure 38 illustrates this cost breakdown for the (a) reference and (b) consolidated networks in Figure 37.

(a)   **(b)**
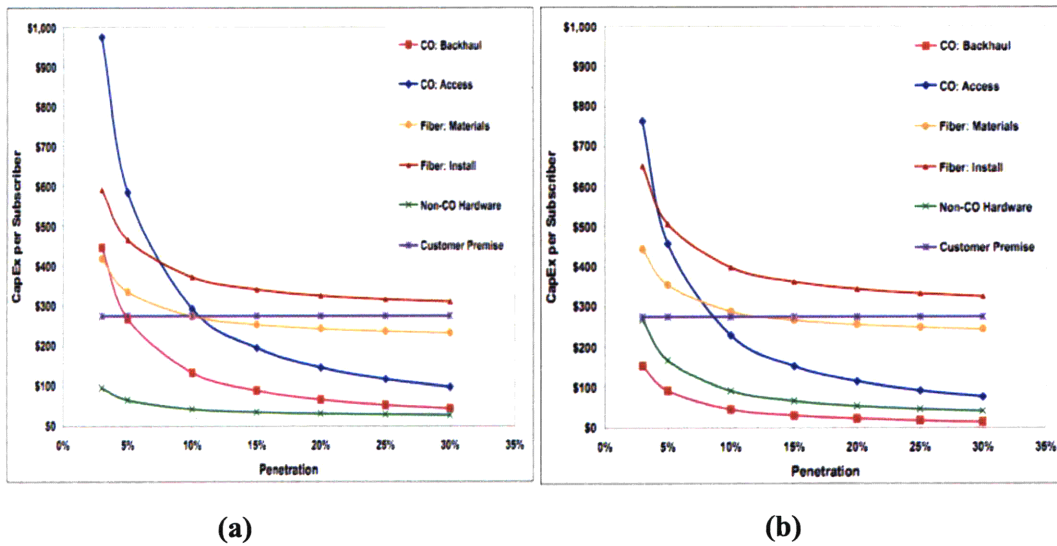
Figure 38: CapEx per subscriber by individual cost category, 30% build for (a) reference and (b) consolidated cases

These cost breakdowns confirm our earlier result that the increase in fiber length corresponding to reach extension in the consolidated case, ("Fiber: Install" and "Fiber: Materials," in Figure 38) are more than offset by the significant cost savings accompanying central office consolidation, via a reduction in both backhaul-related equipment and access network equipment, (the "CO-Backhaul" and "CO-Access" lines in Figure 38).

The cost model also considers the effects of legacy fiber and conduit on network costs, providing a way to account for existing fiber plant in each fiber link, (feeder, distribution etc.). This increases model scope, and enables identification of how legacy in specific portions of the network impact cost. For example, Figure 39 illustrates the effects of different legacy *feeder conduit* percentages on (a) the reference and (b) consolidated cases in Figure 35(b) at a fixed build rate of 80% as a function of penetration percentage.

| (a) | (b) |
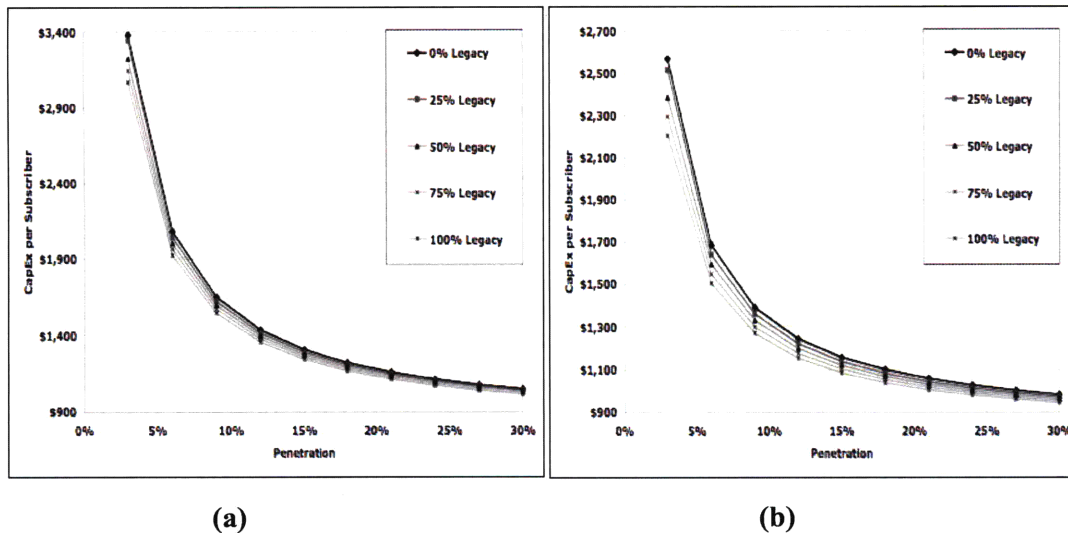
Figure 39: Effects of legacy feeder conduit on reference and consolidated cases in Figure 35(b) at 80% build vs. penetration

The results suggest that, while both networks exhibit lower costs as the quantity of legacy feeder fiber conduit increases, the possible consolidated network benefits are more pronounced at all penetration levels. This is the expected result: as the additional available reach due to consolidation results in increased CO to first splitter stage fiber distances.

9.6 CapEx Model Limitations

The MSL CapEx model, while providing informed estimates on current costs, does not consider how costs change over time. This introduces two sources of error, both of which may lead to CapEx overestimation. First, it does not capture the fact that initial network build-outs take non-trivial time to implement. As a result, the least discounted capital investment the firm sees, (today's prices) represent only a fraction of total CapEx. The MSL model considers CapEx as a one-time investment at the time the technology decision is made. The second source of error is due static component pricing in the MSL CapEx database, when in fact prices will decrease over time due to learning and economies of scale. Future CapEx model improvements will address these issues by incorporating learning curves and economies of scale parameters into the relevant cost functions.

9.7 OpEx Model Capabilities

The OpEx categories provided in Table 25 coupled with the inclusion of penetration, build, and legacy values enables detailed cost breakdowns for the major network architecture categories. For example, Figure 40 provides the OpEx cost breakdowns by (a) CapEx and (b) OpEx categories (Table 24) for the 30% build 20% penetration introduced in Figure 36.

<div align="center">(a)   (b)</div>
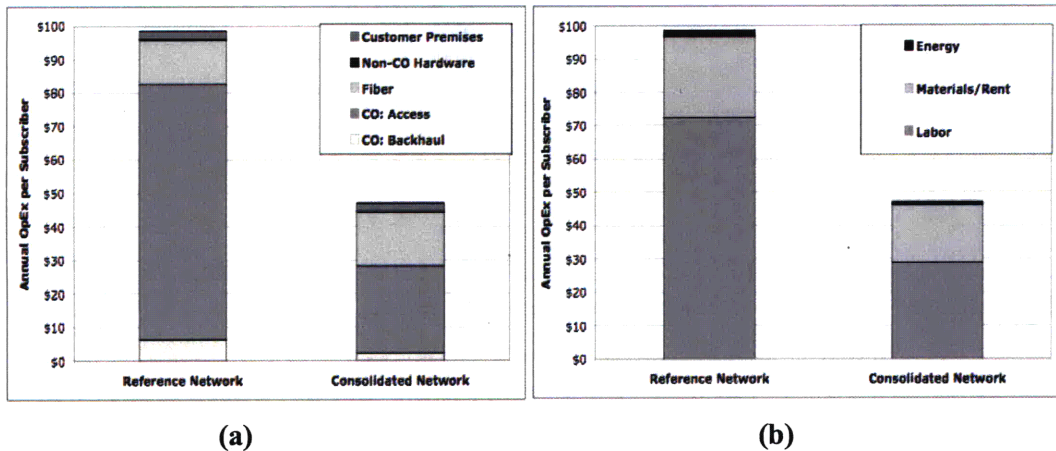
Figure 40: MSL-estimated OpEx for the 30% build, 20% penetration reference and consolidated Corning cases in Figure 36 by (a) network element and (b) cost driver

The breakdowns in Figure 40 provide two ways to look at OpEx. Figure 40(a) categorizes OpEx according to network element, (fiber, CO-access etc.) while (b) identifies the role of OpEx drivers (labor, materials and energy) across all elements. As in the CapEx results for these small networks, the "CO-Access," related costs dominate total OpEx in the reference case, as it requires three times as many COs to serve the same population.

We can also characterize how individual components contribute to total OpEx as a function of penetration. Figure 41 illustrates this relationship using MSL-generated OpEx estimates of (a) the reference and (b) consolidated Corning cases.
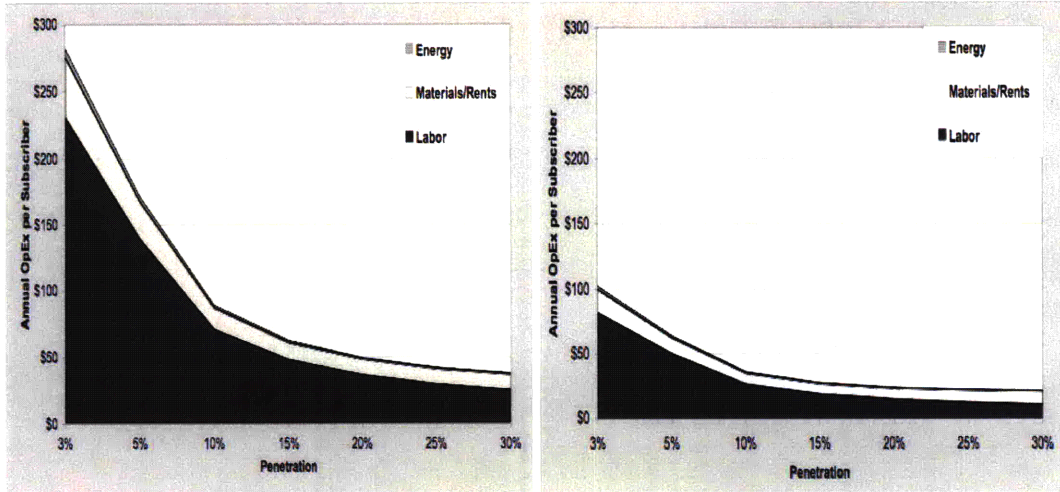
(a)   (b)

Figure 41: MSL generated OpEx components as a function of penetration for the 30% build (a) reference and (b) consolidated cases in Figure 36

Clearly labor is the dominating OpEx driver for these simple cases. The main difference between the reference and consolidated networks is CO quantity. This would suggest that CO-related costs play a significant role in OpEx, providing additional confirmation of our earlier results. It is important to keep in mind however that the extended reach enabling fewer COs results in more deployed fiber. In these geographically small networks, this fiber increase never outweighs the benefits of equipment and facility closures; however, for large coverage regions, technologies enabling extended network reaches may result in significant additional fiber related costs which limit the benefits of continued CO closure.

Finally, we can characterize the OpEx profile for the individual architecture components in Table 24. This provides additional information about where efficiency gains may make the largest OpEx improvements. Figure 42 provides these profiles for 30% build, 20% penetration (a) reference and (b) consolidated cases in Figure 36.
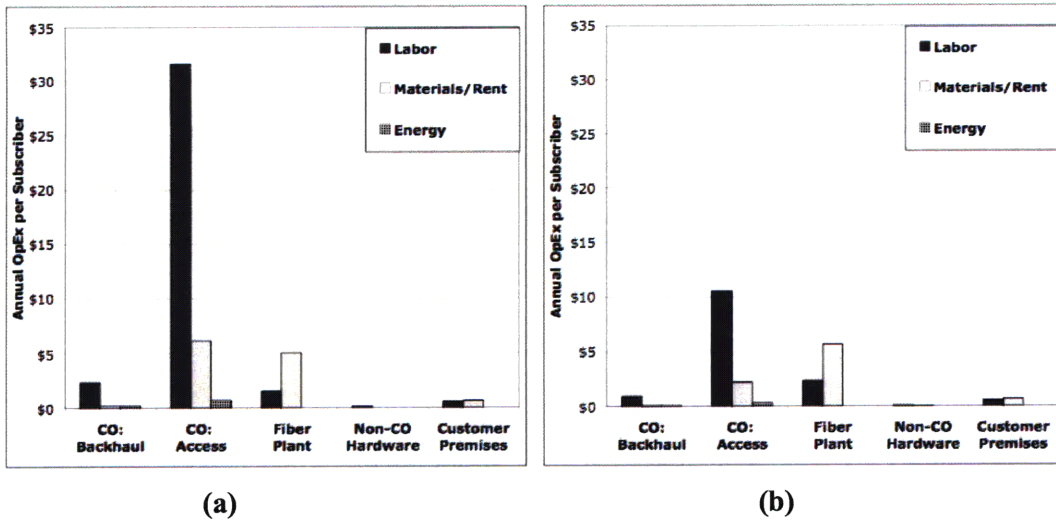
Figure 42: MSL estimated OpEx driver profile by network element for the (a) reference and (b) consolidated 30% build, 20% penetration Corning cases

9.8 OpEx Model Limitations

As with the other modeling components in this work, the OpEx model is not intended to calculate the exact operating costs of a singular technology or network, but rather to characterize relative changes in important OpEx drivers across a spectrum of technologies and populations. Therefore, the results, while realistic estimates of operating expenses for each of the technologies and implementation strategies considered, will not be exact.

10 Case Study Analyses

10.1 Methodology

Our working hypothesis is that characterizing total lifetime network costs including both OpEx and CapEx impacts initial technology choice and implementation decisions, and that population demographics play an important role in driving these costs. Additionally, we would like to characterize the impacts of uncertainty in the cost of future technologies on technology choices today. Consultation with industry and carriers elicited several technologies representing a wide range of current and future thinking regarding GPON FTTh architectures. Of these, three were selected, each with multiple possible implementation strategies, representing near, mid and long-term technology options (see §1.1). A base case network deployment scenario was developed to examine the impacts of lifetime network costs on technology and implementation strategies for

two disparate populations. The scenario assumes a fixed penetration and discount rate; therefore, we examine how technology decisions change as a function of discount rate and penetration for both population demographics. Statistical analysis is used to characterize the impacts of changes in population density and clustering on network design. Finally, the impact of uncertainty in the cost of future technologies on current technology choice and implementation strategy decisions is examined via the development of multiple pricing scenarios.

10.2 Technology Choices / Implementation Strategies Modeled

The technology parameters for the seven implementation strategies are provided in Table 49, where the "centralized" splitter strategy corresponds to strategies utilizing a single splitter stage, and the "Max subscribers per CO/MAN" category reflects the maximum number of subscribers which can be supported by a single central office or metro access node, (recall that the "long-range PON" technology strategies, *C1* and *C2*, do not use COs, but rather route all subscriber data traffic directly to MANs).

Parameter	A1	A2	B1	B2	B3	C1	C2
Transmission Power (dbm)	-28						
Receiver Sensitivity (dbm)	0						
Loss per $\log_2$[ports] (db)	3.5						
Loss per fiber km (db)	0.3						
Connector loss (db)	0.01						
Safety margin (db)	3						
Max data rate (Gbps)	2.5						
Multiplexing strategy	STDM						
Initial splitter utilization	75%						
Statistical multiplier	10						
Splitter strategy	Centralized	Cascaded					
Gain per Amplifier (db)	N/A	N/A	12			20	
Amplifier Type	N/A	N/A	SOA			EDFA	
Network reach (km)	25	25	60	43	25	100	75
Total splitter port count	1:32	1:32	1:128	1:256	1:512	1:512	1:1024
Max Subscribers per CO/MAN	50,000[24] (CO)					500,000 (MAN)	

Table 49: Technology choices and implementation strategies modeled[25]

10.3 Base Case

The base case scenario simulates a greenfield (no existing legacy fiber or conduit) initial network deployment assuming a *100%* initial build and *30%* initial penetration rate. The choice of *100%* build reflects the reality that the fiber plant installation process involves costs in addition to fiber and conduit installation, (permits, security etc.) which do not scale linearly with the amount of

[24] Max CO and MAN subscriber numbers derived from carrier input

[25] Each feeder fiber from the central office is fitted with two amplifiers

fiber plant installed. As a result, it is much less expensive to install additional dark fibers upfront and connect them later as more subscribers purchase service than to install more fiber later over potentially long distances or through crowded urban areas. The *30%* initial penetration rate reflects input from carriers on what a reasonable initial subscriber base percentage might be. Three metrics, CapEx, OpEx, and present cost per subscriber, (which includes CapEx and discounted OpEx) were used to characterize network costs, with the least-cost implementation strategy for each technology chosen as the "best."

10.3.1 Population Demographic Profiles Modeled

Table 50 provides the two dimensions of the population demographic space considered in the both the base and second cases. The two shaded population demographic profiles were used to dimension all technology choice cases.

	High Data Demand	Low Data Demand
High Population Density	I	II
Low Population Density	III	IV

Table 50: Population demographic space

These selections were chosen to bound the range of model responses and costs corresponding to realistic coverage region population diversity facing carriers, although the model enables all four to be created. For example, Region I (corresponds to multiple, high-density populations, simulating large urban and connected suburban areas such as Manhattan and the five boroughs, or Boston and the surrounding area etc. At the other end of the range, Region IV, simulates sparsely populated rural regions composed primarily of single-family homes situated on large land lots. Examples here include large swaths of farming communities. It is important to note that, although all population centers in each case have the characteristics of their profile, this need not be the case. For example, in this thesis, all cities in Region I are assumed to have high population densities with high data demands; however, because each city is independently characterized, other population profiles in this region may contain combinations of high and lower density populations where the average density is high.

In all cases, the total coverage region modeled is *100 x 100 km²* to provide enough area to explore long-range GPON solutions.

10.3.2 Region I: High Population Density / High Data Demands

This demographic profile contains three sub-populations whose central geographic coordinates, $\{x_0\ y_0\}$, are chosen at random. The resulting population map is given in Figure 3. The corresponding individual population parameter values are initially developed in chapter 5, and are summarized in Table 51.

Parameter	Symbol	Value
Total populations	P	3
Population distribution types	$f_\Gamma^p(r)$	Table 3
Households per location total bins	C	4
Households per location by bin	c_i	Table 6
Household bin distribution	$g_\Phi^p(\phi_c)$	Table 7
Household size distribution	$h_{\phi_c}^p(r)$	Table 9; Table 10
Household per location probability	$P\left(\phi_c \middle| i \in \left[0,\delta_p\right]\right)$	Table 11 (all populations)
Households per population	H_p	*552k; 864k; 992k*
Total households	H	*2.21Mil*
Total service tiers	T	*3*
Data rate per tier	ψ_t	Table 12
Service tier data distribution	$k_\Psi^p(\psi_t)$	Table 13
Service tier spatial distribution	$\chi_{\psi_t}^p(r)$	Table 14
Locations per population	l_p	*80k; 160k; 160k*
Total neighborhoods	D_p	*10,000*
Locations per neighborhood	l_d^p	*40*
Total locations	L_p	*400,000*

Table 51: Base case population parameter values

This demographic profile contains four sub-populations whose central geographic coordinates, $\{x_0\ y_0\}$, are chosen at random. The resulting population map is given in Figure 43.
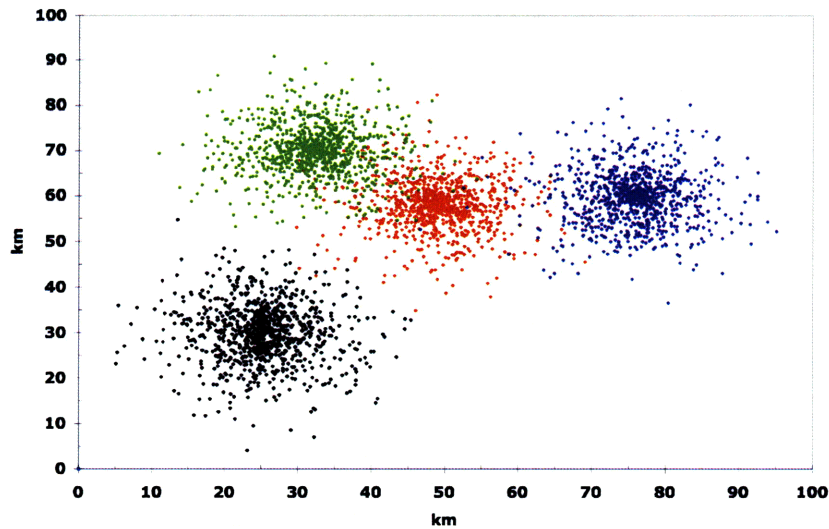
Figure 43: Region IV spatial population distribution

Four individual populations were chosen to reflect areas where multiple small towns form a large community covering a large geographic area. Each town was given slightly different populations and data demands to reflect the regional heterogeneity.

Recall that the Region I population statistics were developed in Chapter 5; however, this second set of demographics requires a new set of distribution characteristics (see Chapter 5 for a discussion of derivations and nomenclature). Table 52 provides the relevant customer location and neighborhood data.

Parameter	Symbol	Value
Total populations	P	*4*
Population distribution types	$f_{\Gamma}^{p}(r)$	*N(0,49)* (all four)
Locations per population	l_p	*50k* (all four)
Neighborhoods per population	d_p	*2,500* (all four)
Total neighborhoods	D_p	*10,000*
Locations per neighborhood	l_d^{p}	*20*
Total locations	L_p	*200,000*

Table 52: Region IV customer location data

The distribution of households in this profile is dominated by single-family homes, with a small percentage of business and multi-unit dwellings in the smaller city centers. The resulting household distribution profile is given in Table 53.

102

Category ID	Name	Subscribers per Location
c = 1	Single-family homes	1
c = 2	Multi-unit buildings	4
c = 3	Businesses	10

Table 53: Region IV household size categories and values

The household size probability mass function values for all four towns are given in Table 54.

Population	$g_\Phi^p(\phi_{c=1})$	$g_\Phi^p(\phi_{c=2})$	$g_\Phi^p(\phi_{c=3})$
1	95%	2.5%	2.5%
2	90%	5%	5%
3	85%	7.5%	7.5%
4	80%	10%	10%

Table 54: Region IV household size distribution values

The resulting total households per population are provided in Table 55, while the spatial distribution of household sizes is given in Table 56.

Population	Households
1	16,250
2	20,000
3	23,750
4	27,500
Total	87,500

Table 55: Region IV total households per population

r	$h_{c=1}(r)$	$h_{c=2}(r)$	$h_{c=3}(r)$
$0 \le |r| \le \sigma$	0.10	0.70	0.90
$\sigma < |r| \le 2\sigma$	0.20	0.30	0.10
$|r| > 2\sigma$	0.70	0.00	0.00

Table 56: Region IV spatial household distribution

The resulting households per location conditional probability profiles are given for populations one and two in Table 57, and three and four are Table 58, (see Table 11 and surrounding discussion).

103

$P\left(\phi_c \middle\| i \in \left[0, \delta_{p=1}\right]\right)$	$0 < \delta_{p=1} \leq 1$		$1 < \delta_{p=1} \leq 2$		$\delta_{p=1} > 2$	
Population:	P_1	P_2	P_1	P_2	P_1	P_2
$\phi_{c=1} = 1$	0.70	0.53	0.95	0.9	1	1
$\phi_{c=2} = 4$	0.13	0.21	0.0375	0.075	0	0
$\phi_{c=3} = 10$	0.17	0.26	0.0125	0.025	0	0
Total Probability	1	1	1	1	1	1

Table 57: Region IV households per location conditional probability profile by distance region, populations 1 & 2

$P\left(\phi_c \middle\| i \in \left[0, \delta_{p=1}\right]\right)$	$0 < \delta_{p=1} \leq 1$		$1 < \delta_{p=1} \leq 2$		$\delta_{p=1} > 2$	
Population:	P_3	P_4	P_3	P_4	P_3	P_4
$\phi_{c=1} = 1$	0.41	0.33	0.85	0.8	1	1
$\phi_{c=2} = 4$	0.26	0.29	0.11	0.15	0	0
$\phi_{c=3} = 10$	0.33	0.38	0.04	0.05	0	0
Total Probability	1	1	1	1	1	1

Table 58: Region IV households per location conditional probability profile by distance region, populations 3 & 4

The profile consists of two data service tiers, given in Table 59.

Tier ID	Name	Services	Data Rate per Subscriber (Mb/s)
1	Basic Service	Basic Internet	$\psi_{t=1} = 5$
2	Extended Service	Internet, VOIP	$\psi_{t=2} = 20$

Table 59: Region IV data service tiers

The rate probability mass function values (the percentages of each population receiving each service tier) are given in Table 60.

Population	$k_\Psi^p(\psi_{t=1})$	$k_\Psi^p(\psi_{t=2})$
1	80%	20%
2	70%	30%
3	60%	40%
4	50%	50%

Table 60: Region IV service tier population percentages

The spatial distribution of data tiers is the same for all four populations, and is given in Table 61.

r	$\chi_{\psi=1}(r)$	$\chi_{\psi=2}(r)$
$0 \le |r| \le \sigma$	0.3	0.8
$\sigma < |r| \le 2\sigma$	0.4	0.2
$|r| > 2\sigma$	0.3	0

Table 61: Region IV data tier spatial distribution

The resulting data tier per location conditional probability profiles for populations 1 & 2 and 3 & 4 are given in Table 62 and Table 63 respectively.

$P\left(\psi_t \Big| i \in \left[0, \delta_{p=1}\right]\right)$	$0 < \delta_{p=1} \le 1$		$1 < \delta_{p=1} \le 2$		$\delta_{p=1} > 2$	
Population:	P_1	P_2	P_1	P_2	P_1	P_2
$\psi_{t=1} = 5\,Mbps$	0.6	0.47	0.89	0.82	1	1
$\psi_{t=2} = 20\,Mbps$	0.4	0.53	0.11	0.18	0	0
Total Probability	1	1	1	1	1	1

Table 62: Region IV spatial distribution of data tiers, populations 1 & 2

$P\left(\psi_t \Big| i \in \left[0, \delta_{p=1}\right]\right)$	$0 < \delta_{p=1} \le 1$		$1 < \delta_{p=1} \le 2$		$\delta_{p=1} > 2$	
Population:	P_3	P_4	P_3	P_4	P_3	P_4
$\psi_{t=1} = 5\,Mbps$	0.36	0.27	0.75	0.67	1	1
$\psi_{t=2} = 20\,Mbps$	0.64	0.73	0.25	0.33	0	0
Total Probability	1	1	1	1	1	1

Table 63: Region IV spatial distribution of data tiers, populations 3 & 4

10.3.3 Base Case Results and Analysis

The networks cost results were examined using three metrics: CapEx, OpEx, and lifetime discounted network costs per subscriber. The base case assumes no uncertainty in technology cost; therefore, the cost models use current cost estimates for all technologies. Figure 44 presents the CapEx per subscriber results for (a) Region I and (b) Region IV population densities, each broken down by network element, (see Table 24 for individual included costs) for all seven implementation strategies defined in Table 49.
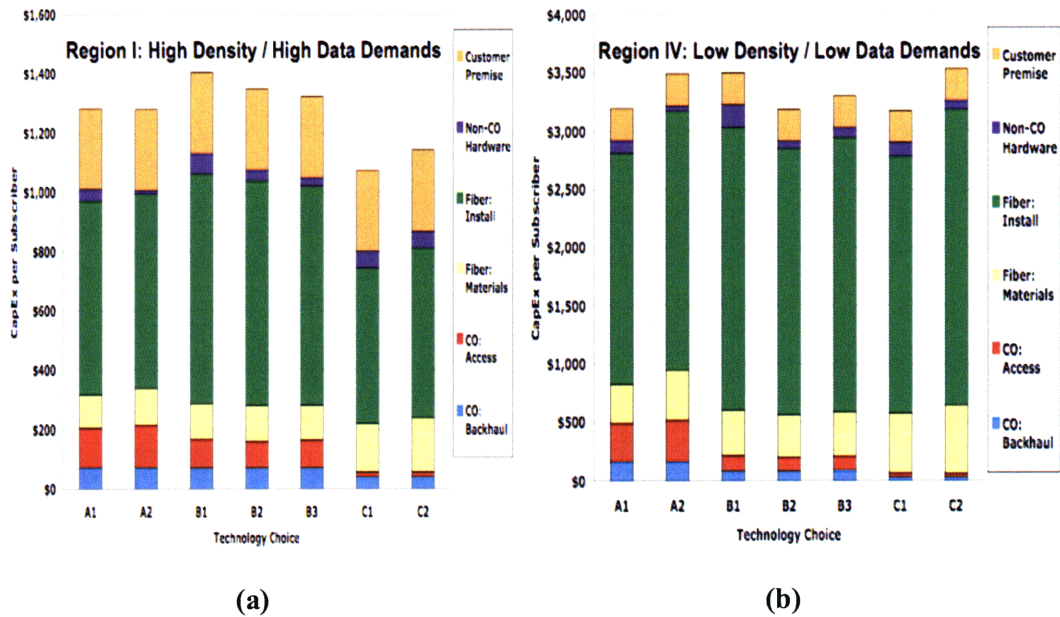
(a)   (b)

Figure 44: Base case CapEx per subscriber for (a) high density, and (b) low density population demographics

The results suggest some interesting CapEx-specific tradeoffs between technologies and implementation strategies both within a given population, and between demographics. For example, for the sparse population in region IV, the increased reach enabled by signal amplification in technology "B," results in a reduction in transmission and backhaul-related equipment: more customers serviced per line, either via reach extension or higher split ratios requires fewer lines, which reduces the total required central offices. As a result, the "CO-Access," and "CO-Backhaul" costs for all three technology "B" implementation strategies in Figure 44 (b) are significantly less than the un-amplified (and therefore shorter reach and lower splitter port count) technology "A" solutions. One would expect similar behavior in the high-density region I CapEx; however, the *50,000*-subscriber maximum per CO forces most of the COs to remain, muting this amplification-related cost advantage. Another interesting, and unexpected result involves the CapEx reduction accompanying the advanced long-range technology choices, (*C1* and *C2*). In the high-density region, one might expect that the extremely long reach enabled by the EDFA amplifiers would result in extra fiber-related expense; however, our analysis suggests that this reach will enable more fiber to be co-located for greater distances, reducing total installation costs. By contrast, the fact that subscribers are located further apart in region IV means that the increased reach will be used to add customers further away to each line,

resulting in additional fiber installation expense. In both cases however, fiber *material* costs do increase however, as aggregating all data at metro access nodes, (which are often not located in a city center) instead of COs results in larger fiber bundles over longer distances.

As described in , the tradeoff between using the power budget for reach extension versus increasing splitter port count can result in significantly different CapEx results depending on how a single technology choice in implemented. For example, in region I, using the extra power budget enabled by amplification for technology "B" to increase splitter port count (by moving from *B1* to *B3*) reduces the associated CapEx, while in region IV, implementation strategy *B2*, with a *1:256* way splitter port count, exhibits the least associated CapEx. These results suggest that *how* a technology is implemented may be just as important as which technology is selected, and that within a given technology optimal implementation strategies may both exist, and vary according to the population demographics to be served.

Finally, although both technology "C" strategies in region I exhibit the smallest associated CapEx per subscriber, this is not the case in sparsely populated region IV, as the associated increase in fiber installation costs offset the corresponding reduction in central office related costs. Additionally, only considering CapEx in this region suggests that implementing strategies *A1*, *B2* and *C1* results in similar network costs. However, as we shall see, the OpEx associated with these three choices is not the same. As a result, choosing a strategy based only on CapEx may result in significant additional costs over the network lifetime.
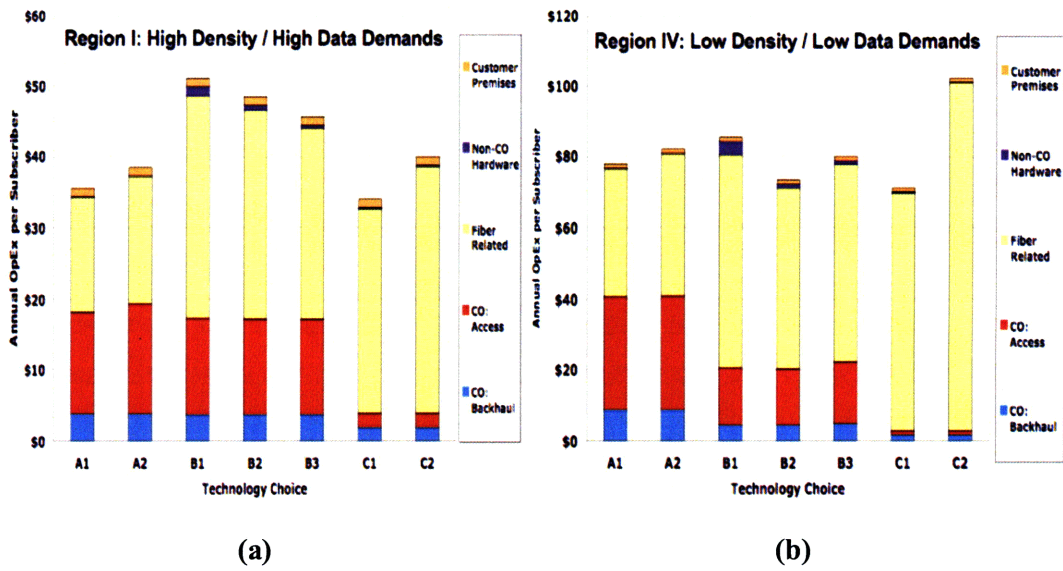
Figure 45: Base case OpEx per subscriber for (a) high density and (b) low density population demographics

The results indicate that fiber and central office related costs dominate OpEx, but in different ratios depending on technology and implementation strategy. In the *high-density* population case, the short-range, low-split technology A strategies require many central offices, (with not all reaching the *50,000* subscriber maximum) resulting in significant CO-related OpEx. However, the fact that all subscribers are located within twenty-five kilometers from the CO results in less overall installed fiber than the amplified technologies, tempering fiber-related OpEx. By contrast, the inability of technology B to utilize its amplified power budget to close central offices, (due to the maximum CO subscriber limits discussed above) results in the "worst of both worlds" from an OpEx perspective, requiring almost the same CO-related OpEx as the un-amplified strategies, while also deploying more installed fiber to either connect more subscribers to a single line, or reach subscribers further away from each CO. The decreasing fiber-related OpEx corresponding to increased splitter port counts as we move from *B1* to B3 (from 1:128 to 1:512 total split ratio per line) reflects the fact that densely populated areas benefit less from using the available power budget to increase network reach, as subscribers are clumped together, and that this power budget should be used instead to increase the total splitter port count per line. Interestingly, this trend of decreasing network OpEx with increasing splitter port count appears to end at a total splitter port count of 1:512. For example, strategy *C1* is characterized by a total splitter port count of 1:512 per line, and a total reach of *100km*, while *C2* enables up to 1:1024

108

subscriber per transmission line, but at the price of reducing total network reach to *85km*. Even considering that the primary driver for OpEx reduction in the long-range strategies is the closing of central offices, we would expect that the increased splitter port count of strategy *C2* (1:1024) would result in lower OpEx; however, as Figure 45 (a) illustrates, pursuing strategy *C1*, with the smaller 1:512 split ratio results in significantly less OpEx per subscriber. This suggests that the additional fiber length expense associated with adding additional subscribers to fill the large splitters in strategy *C2* eventually outweighs the benefits of additional subscriber per line in high-density regions. These results indicate that an optimal tradeoff exists between network reach and total splitter port count may exhibit an optimal value. Finally, because the long-range solutions require large fiber bundles for much longer distances than the other strategies, the long repair times and expensive replacement materials associated with these larger bundles result in strategies which are dominated by fiber-related OpEx.

The *low-density* population case, while exhibiting similar behavior, results in a different strategy ranking based on total annual OpEx, with amplified, 50km, 1:256 total split strategy *B2* exhibiting the smallest OpEx per subscriber. Several tradeoffs are responsible for this shift. First, amplification enables more subscribers per line at further distances from the CO, which results in an increase in fiber when compared with the non-amplified strategies; however, the reduced number of subscribers in the coverage region means that more customers per line will enable fewer COs to serve the same number of customers without running into the *50,000* subscribers per CO maximum. Second, the tradeoff between splitter port count and total network reach in the high-density case suggests OpEx savings accompany reach extension and increased total splitter port count up to maximum values of 100km and 1:512. In the low-density case, both *B2* and *C1* exhibit minimal relative OpEx per subscriber. This suggests that strategy *B2* is able to preserve the benefits gained via closing central offices, while also striking the right balance between network reach and splitter port count, thereby minimizing total fiber distance. Finally, as in the high-density case, the long-range solutions are dominated by fiber-related OpEx as a result of larger fiber bundles extending over larger distances. Although subscribers are spaced further apart, strategy *C1* still exhibits the optimal splitter strategy/ reach combination, while the splitter ratio in *C2* appears to be to large, resulting in long fiber lengths from the last splitter stage to subscriber locations.

Although the CapEx per subscriber suggests equivalence between multiple technologies and/or implementation strategies, the OpEx per subscriber suggests that this is due to incomplete information. For example, both technology "A" implementation strategies in Figure 44 (a) exhibit similar CapEx; however, the fiber reduction enabled by the single-stage consolidated splitter architecture in *A1* results in non-trivial OpEx savings. Similarly, the results in Figure 44 (b) show that, although the CapEx for strategies *A1*, *B2*, and *C1* are almost identical, (see Figure 43 (b)) the associated OpEx results are quite different. Therefore, using CapEx as the single metric may result in technology choices which result in significantly higher total costs over the network lifetime.

Because OpEx is a recurring cost, its value must be characterized over time. Incorporating the discount rate provides a way to capture this effect, by enabling characterization of total lifetime costs for each strategy. Figure 46 provides a sensitivity analysis illustrating these lifetime costs[26] for the base case deployment strategies as a function of the discount rate.

(a)   **(b)**
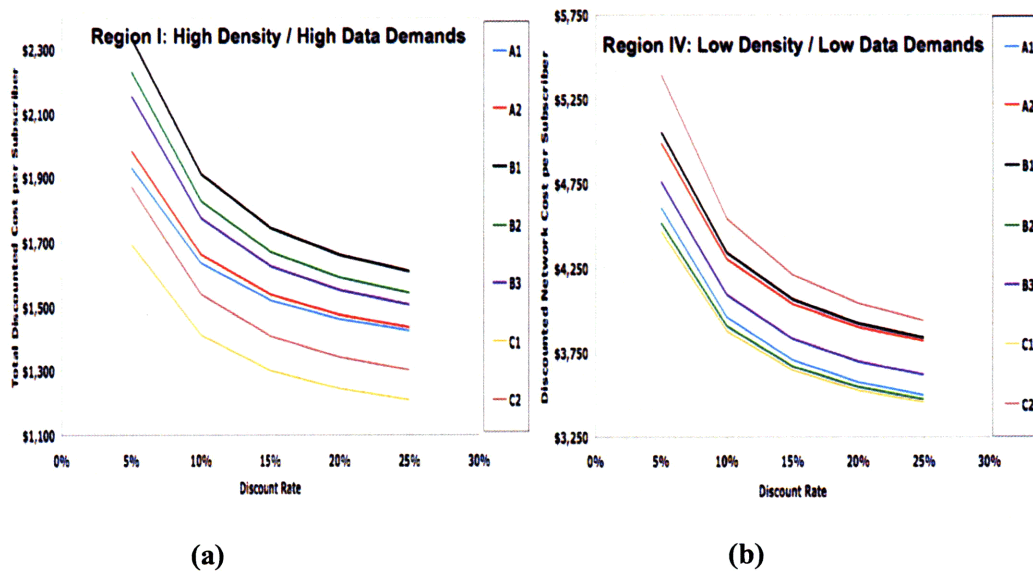
Figure 46: Discounted lifetime costs per subscriber as a function of discount rate for the (a) high density and (b) low density regions

These results also provide a way to value the impacts of OpEx on technology decisions, as higher discount rates reduce the associated OpEx burden over time. For example, a firm discount rate of 5% in the low-density case suggests that strategy *B2* provides the cost-minimizing network solution due to the relative OpEx advantage of this strategy shown in Figure 45 (b).

---

[26] Network costs discounted over a 50 year time horizon

110

However, increasing the discount rate to 25% erases this OpEx advantage through discounting. As a result, CapEx again becomes the dominant criterion for technology choice, resulting in the three-way tie between *A1*, *B2*, and *C1* observed in Figure 44 (b).

When taken together, these results indicate that, for the base case, not only can characterizing OpEx add value to technology choice and implementation strategy decisions, but also that population demographics play a significant role in both technology choice and the particular way in which the technology should be implemented.

10.4 Effects of Penetration and Discount Rate on Technology Strategy

The results in Figure 46 provide a "ranking" of technology and implementation strategies for fixed build, penetration and legacy fiber plant values, (recall that these are 100%, 30% and 0% respectively for the base case). However, because penetration affects both CapEx and OpEx, we would like to generalize our earlier results to characterize these effects. Additionally, because we have seen that discount rate plays an important role in the impact of OpEx on lifetime costs, we examine the effects of penetration and discount rate on technology choice and implementation strategy. Figure 47 presents a map of these effects for both the (a) high density region I, and (b) low density region IV population demographics.
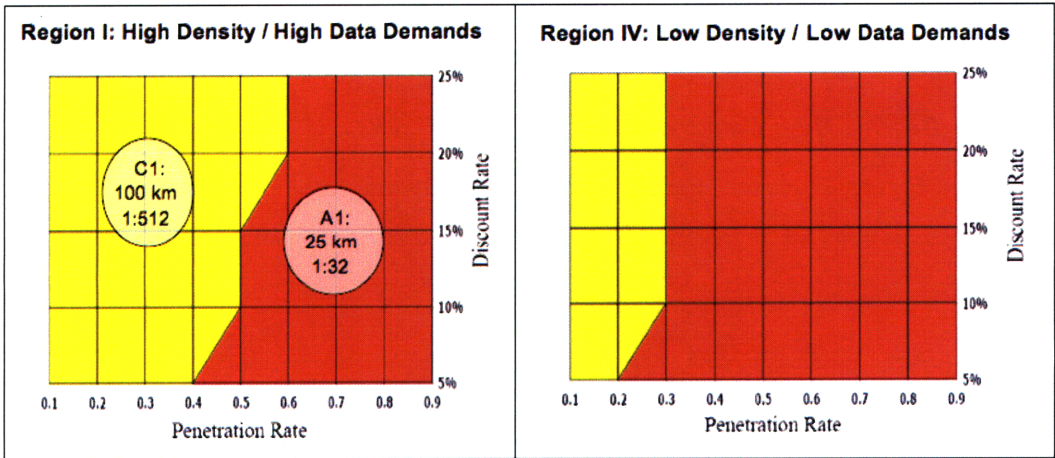

(a)   (b)

Figure 47: Technology strategy as a function of penetration and discount rate for (a) high density, and (b) low density regions

The results suggest that two deployment strategies dominate in both population cases: the short-range, low-split *A1*; and the long-range, high-split *C1*. These results do not seem to match the behavior observed in the base case. For example, Figure 46 (a) suggests that, at a penetration of 30%, (the base case deployment), *C1* and *C2* are the two least-cost technology strategies to serve the high-density population, while *C1* and *B2* are the least-cost options in the low density case, (Figure 46 (b)). Because OpEx is the only cost which is discounted over time, as we increase the discount rate, the effect of OpEx on total network costs diminishes. Therefore, for a fixed penetration rate, any change in technology strategy as we vary the discount rate is directly attributable to the difference in lifetime OpEx per subscriber between the technologies on either side of the boundary. However, technology strategy shifts corresponding to changes in penetration at a fixed discount rate are due to tradeoffs between CapEx and OpEx, as each technology strategy responds to additional subscribers, (which will depend on the tradeoff between network reach and splitter strategy). The least-cost strategy in both the high and low-density cases shown in Figure 47 exhibit dependence on both OpEx, as evidenced by shifts in strategy as we change the discount rate holding penetration constant, (vertical lines within given scenario) and how each strategy adapts to service additional subscribers, (horizontal lines across a given scenario).

To characterize how these effects impact technology strategy, we examine the total discounted cost per subscriber as a function of penetration at fixed discount rates for both the high and low population densities modeled in the base case.

10.4.1 High-Density, High Data Demand Population Results

Figure 48 and Figure 49 present the total discounted network costs as a function of penetration for all seven technology strategies modeled, with enlarged regions where

Figure 48: Total discounted network cost per subscriber as a function of penetration for the high-density case at discount rate 10%, and 100% build with penetration region 0.35 to 0.7 shown expanded. Base case penetration shown for reference

Figure 49: Total discounted network cost per subscriber as a function of penetration for the high-density case at discount rate 20%, and 100% build with penetration region 0.45 to 0.85 shown expanded. Base case penetration shown for reference

These figures illustrate that cost crossovers exist where the per-subscriber total discounted costs for the long-range strategies stop decreasing. At a 10% discount rate, this crossover begins at a penetration value of 0.4 for *C1*, and 0.5 *C2*, and ends at penetration values of 0.6 and 0.7

respectively, while when the discount rate is 20%, both crossover regions shift towards higher penetration values, beginning at 0.5 and 0.6 for *C1* and *C2*, and ending at 0.725 and 0.825 respectively. This behavior provides three insights which, when combined, explain the strategy map shown in Figure 47 (a). The only fiber component which changes with penetration in our model is the link connecting the final splitter to the neighborhood. Therefore, how each strategy handles this distance determines how network costs change with penetration. All strategies address this distance by deploying additional fiber from splitter sites which are not fully utilized, (recall that an initial build of 100% installs splitters with dark fibers at all splitter sites to connect new subscribers).

The increased reach and splitter port count of the long-range strategies both eliminates COs and results in splitter sites which are located further away from subscribers than the other two technologies, whose limited reach and splitter ports result in central offices and splitters which are closer to subscribers. As a result, the additional installed fiber length required to serve new subscribers as penetration increases will be greatest for the long-range strategies. As penetration continues to increase, this effect gets larger until it begins results in an increase in the average discounted per subscriber network costs. This is why the long-range curves begin to trend upwards rather than asymptoting to a minimum value. The reason why the technology strategy behavior observed in Figure 47 (a) is different than that suggested by the base case is that at the base case penetration of 30% no crossovers have been observed. The expanded graphic in Figure 48 provides the evidence for the observed strategy behavior: below a penetration value of 0.5, strategy *C1* results in the least total discounted cost per subscriber; however, due to the *C1* crossover, *A1* becomes the dominant strategy for penetrations greater than 0.5.

10.4.2 Low Density, Low Data Demand Population Results

In the low-density case, this crossover effect is much more subtle, as there are significantly fewer total possible subscribers. Additionally, only strategy *C1* exhibits a cost crossover in the case. Because the population is far more geographically spread out than in the high density case, the extremely high split ratio enabled by *C1*, coupled with extended reach, results in many large fiber bundles running over large distances. This in turn costs requires additional upfront investment, and translates into additional OpEx over time. As a result, this strategy loses the CapEx and OpEx advantages exhibited in the high-density case. Figure 50 and Figure 51

illustrate the *C1* to *A1* cost crossover point as a function of penetration for discount rates 10% and 20%.
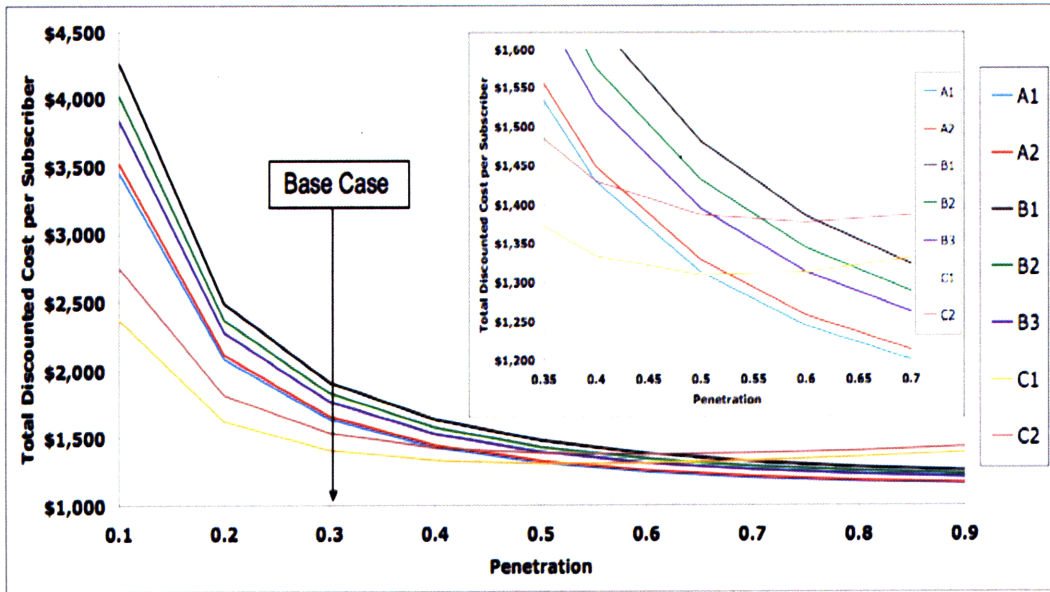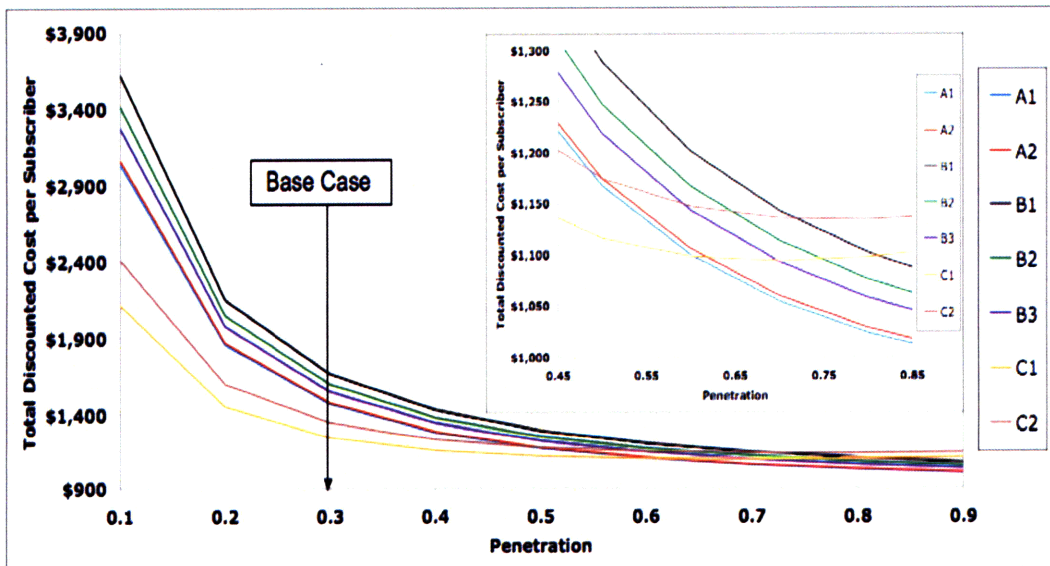

Figure 50: Total discounted network cost per subscriber as a function of penetration for the low-density case at discount rate 10%, and 100% build with penetration region 0.29 to 0.34 shown expanded. Base case penetration shown for reference

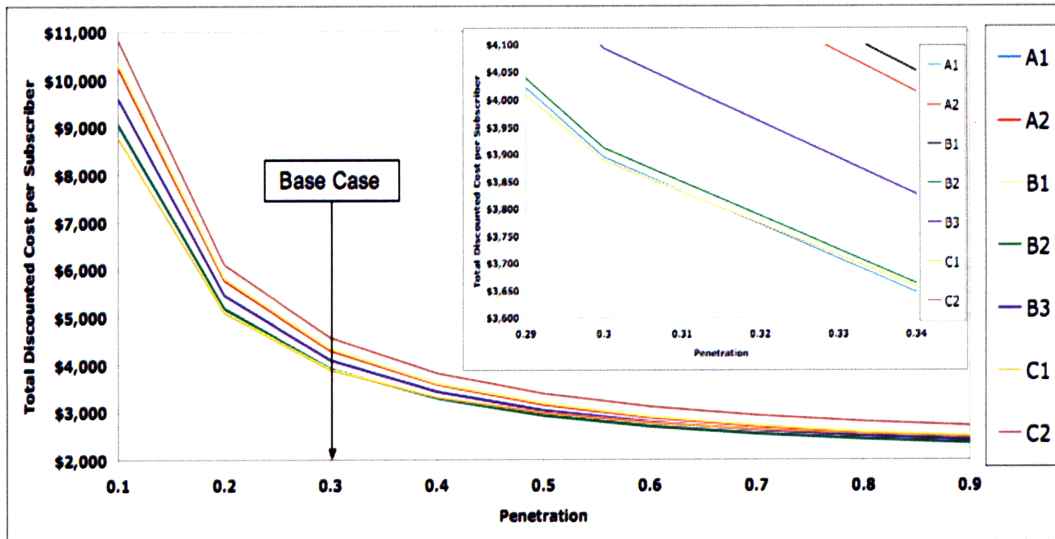**Figure 51: Total discounted network cost per subscriber as a function of penetration for the low-density case at discount rate 20%, and 100% build with penetration region 0.3 to 0.4 shown expanded. Base case penetration shown for reference**
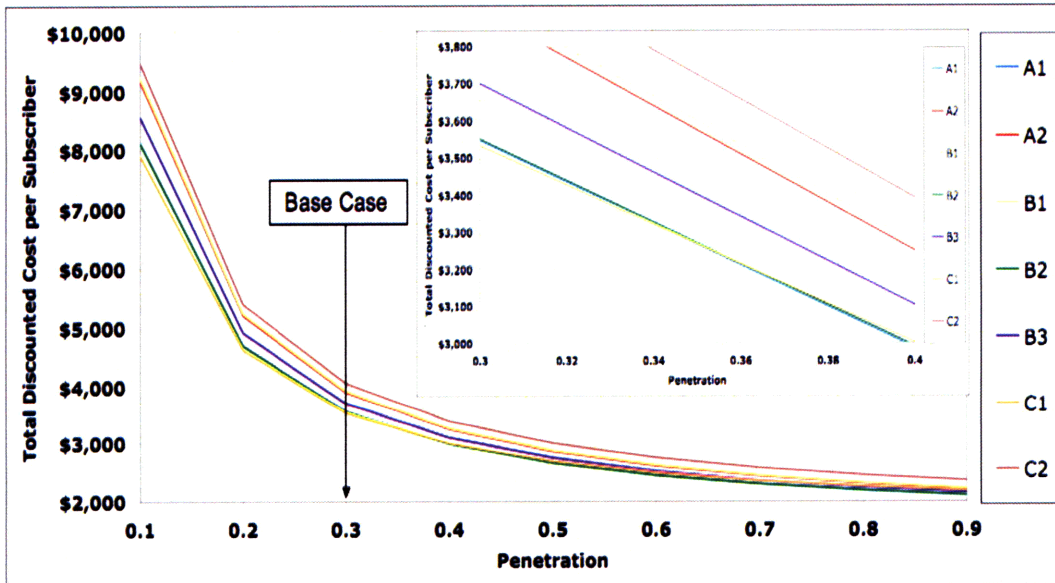
The behavior in the low-density case exhibits both similarities and differences when compared with the high-density scenario, which is driven by two, additive effects As in the high-density

case, the additional installed fiber length required to serve new subscribers will be greatest for the long-range strategies. However, unlike the high-density case, the costs associated with this increase in fiber length are not enough to cause the discounted cost per subscriber to increase at higher penetration levels, (as shown by the positive curvature in both the *C1* and *C2* lines after penetration values of 0.5 and 0.4 in Figure 48), but instead to asymptote at a minimum value. Recall that *A1* and *C1* exhibit almost identical CapEx in the low-density base case, (Figure 44 (b)) yet have significantly different OpEx structures, with *A1* exhibiting less fiber-related OpEx and less OpEx overall, (Figure 45 (b)). As a result, as the fiber length differential between *A1* and *C1* increases at higher penetrations, so will the associated fiber-related OpEx. These two effects combine to reduce the total discounted cost per subscriber for *A1* at a faster rate than *C1*, leading to the observed cost crossover.

10.5 Effects of Uncertainty in Future Technology Costs on Technology Strategy

Thus far, our analyses have treated all technology strategies as though they could be deployed today, and that the estimates we are using for component pricing represent actual costs. In reality however, there is significant uncertainty about how much future technologies will cost to implement on a large scale. This analysis explores the effects of this uncertainty on technology strategy.

10.5.1 Methodology

Technology-related cost uncertainty can either increase actual network costs if underestimated, or reduce them if overestimated. We model this uncertainty by introducing the technology-dependent multiplicative factor, $M_i = (1 + m_i)$ (where i = A, B, or C depending on the technology choice, and $-1 \leq m \leq 1$ represents the variance percentage) which modifies the relevant *CapEx* related cost categories (Table 24) for the each technology[27]. While we expect technology-related cost uncertainty to affect almost all areas of the network, we do not expect significant variance in the price of fiber materials or installation over time. Additionally, because all the technologies modeled in this thesis are assumed to utilize fiber with the same optical and loss characteristics, fiber-related costs will remain the same as in earlier analyses. For example,

---

[27] OpEx cost structures are not modified

116

if the actual cost to implement technology B is 20% higher than the cost estimates used in the base case analysis, then this analysis will multiply all CapEx categories except fiber for all technology B strategies by $M_B = (1 + m_A) = (1 + 0.2) = 1.2$.

All seven technology strategies are modeled for the two population scenarios defined for the base case. We assume that technology A is the closest to implementation, followed by technology B, and finally technology C. We also assume that cost uncertainty increases with time; therefore, technology A cost estimates are modeled as invariant, $(M_A = (1 + m_A) = (1 + 0) = 1)$[28] and technology C is assigned the largest multiplier, (such that M_B < M_C is always true).

Three scenarios are compared for both the high and low-density populations developed in the base case: Scenario 1 assumes that the base case cost estimates are too high for technologies B and C, resulting in actual deployment costs which are smaller than expected; Scenario 2 uses the base case results, which assume that all cost estimates equal actual deployment costs; and Scenario 3 assumes that the base case cost estimates for technologies B and C are too low, such that actual costs are higher than expected. By modeling different variance values, we can characterize technology strategy sensitivity to uncertainty in technology cost for different discount rates and penetration values in both the high and low-density population cases. Table 64 provides a scenario overview and the corresponding cost variance and multiplier relationships.

Scenario ID	Technology A		Technology B		Technology C	
	m_A	M_A	m_B	M_B	m_C	M_C
1	0	1	$m_C < m_B < 0$	$M_C < M_B < 1$	$m_C < m_B < 0$	$M_C < M_B < 1$
2	0	1	0	1	0	1
3	0	1	$m_C > m_B > 0$	$M_C > M_B > 1$	$m_C > m_B > 0$	$M_C > M_B > 1$

Table 64: CapEx uncertainty scenarios considered for each population by technology

10.5.2 Results and Analysis

Recall that OpEx is the only cost which is discounted over time. Therefore, as we increase the discount rate, the effect of OpEx on total network costs diminishes. Therefore, for a fixed penetration rate, any change in technology strategy as we vary the discount rate is directly attributable to the difference in lifetime OpEx per subscriber between the technologies on either side of the boundary. However, technology strategy shifts corresponding to changes in

---

[28] Therefore, the cost results for strategies A1 and A2 are the same as in the base case

penetration at a fixed discount rate are due to tradeoffs between CapEx and OpEx, as each technology strategy responds to additional subscribers. Adjusting the CapEx component cost estimates for technologies B and C, will impact this tradeoff

To characterize how these effects impact technology strategy, we again examine the total discounted cost per subscriber as a function of penetration at fixed discount rates for both the high and low population densities modeled in the base case.

10.5.3 Low-Density, Low Data Demand Population Results

Figure 55 illustrates the minimum total discounted cost technology as a function of both penetration and discount rates for the low-density, low-data demand population case. From left to right, Figure 55 represents (a) Scenario 1, in which current component costs *overestimate* actual deployment costs by 5% for technology *B* and 10% for technology *C* (actual costs are *less* than expected); (b) Scenario 2, in which current costs equal actual deployment costs for all technologies, (Figure 47 (a)); and (c) Scenario 3, in which current component costs *underestimate* actual deployment costs by 5% for technology *B* and 10% for technology *C*, (actual costs *greater* than expected). Table 65 provides the corresponding variance, m_i, and multiplier, M_i, values for all three scenarios.
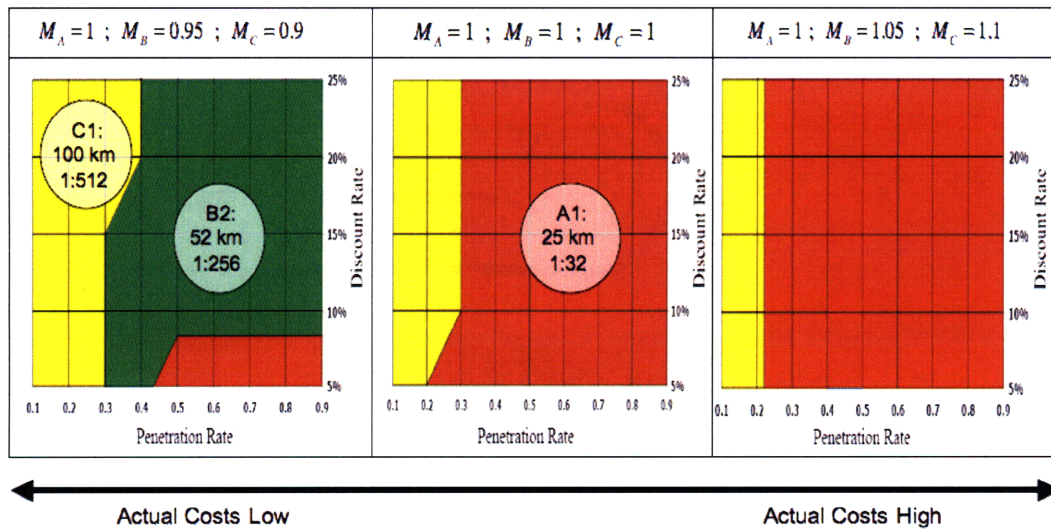

Figure 52: Low-density technology strategies exhibiting least total discounted network costs as a function of penetration and discount rate for (a) Scenario 1: M_B=0.95, M_C=0.9; (b) Scenario 2: M_B= M_C=1; and (c) Scenario 3: M_B=1.05, M_C=1.1

118

Scenario ID	Technology A		Technology B		Technology C	
	m_A	M_A	m_B	M_B	m_C	M_C
1	0	1	-5%	0.95	-10%	0.9
2	0	1	0	1	0	1
3	0	1	+5%	1.05	+10%	1.1

Table 65: Variance, m_i, and multiplier, M_i, values modeled for the low-density, low-data demand case

We immediately notice that even these relatively small changes in component cost estimates result in significant changes in the lowest-cost strategy, particularly when costs are overestimated, (Scenario 3). The least-cost strategy also shows dependence on both OpEx, as evidenced by shifts in strategy as we change the discount rate holding penetration constant, (vertical lines within given scenario) and how each strategy adapts to service additional subscribers, (horizontal lines across a given scenario). However, both effects are scenario dependent, and indicate that, as CapEx is undervalued by underestimating component costs, the importance of OpEx on technology strategy diminishes.

Because the cost multiplier, M_i, only operates on CapEx-related component costs, a map of OpEx as a function of penetration and discount rates fully characterizes OpEx for all penetration and discount rates for a given population. Additionally, because discount rate only impacts OpEx, a graph of CapEx as a function penetration rate provides a complete picture of CapEx for a given population. As a result, separating the total discounted network costs into (a) CapEx as a function of penetration and (b) discounted *total* OpEx costs per subscriber as a function of both discount rate and penetration, and then comparing these against the total discounted cost per subscriber, provides insight into (1) why three different technology strategies all produce least-cost solutions at a single discount rate, (2) why this shrinks to two strategies at higher discount rates, and (3) why underestimating component costs in Scenario 3 results in two strategies irrespective of the discount rate, but dependent on penetration.

Figure 53 illustrates these breakdowns for Scenario 1 in Figure 52, while Figure 54 provides the total discounted cost per subscriber as a function of penetration for a discount rate of 7%, (the discount rate in Figure 52 (a) at which three different strategies provide least-cost solutions depending on penetration).

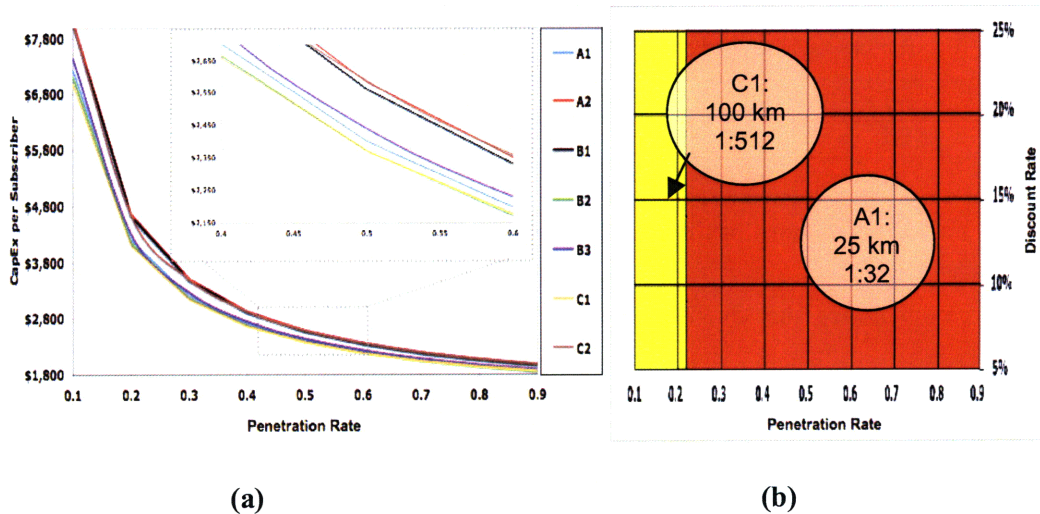(a)   (b)

Figure 53: Low-density, low data-demand (a) CapEx per subscriber vs. penetration with crossover emphasized, and (b) discounted OpEx per subscriber vs. both penetration and discount rate for Scenario 1
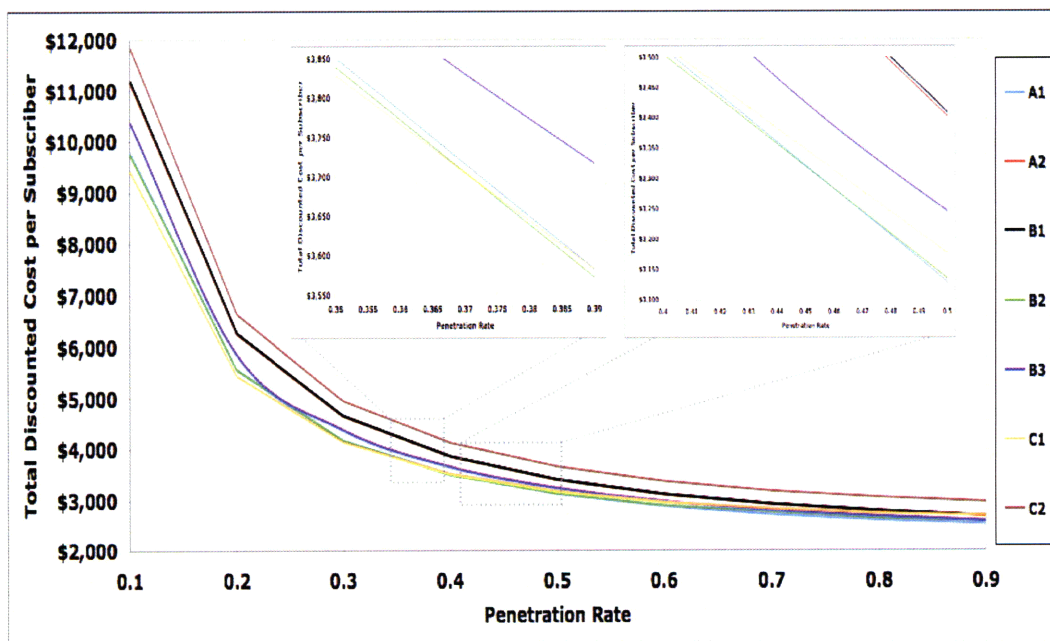
Figure 54: Low-density total discounted cost per subscriber as a function of penetration rate for discount rate 7% with crossovers emphasized, Scenario 1: M_B=0.95, M_C=0.9

We explore the three issues above in order, beginning with, (1) why three different technology strategies, *C1*, *B2*, and *A1*, all produce least-cost solutions for discount rates less than ~7%. When *M=1* for all three technologies, (estimated costs equal actual costs), recall that strategies *A1*, *B2*, and *C1* exhibit almost identical CapEx at low penetrations in the low-density case,

(Figure 44 (b)). Additionally, our assumption is that, if technologies B and C are overvalued, then C is *more* overvalued than B, (because costs of implementing C are more uncertain). As a result, when component CapEx for *C1* is overestimated, ($M_C < 1$ as in Scenario 1) this strategy becomes the least expensive at low penetration rates from a CapEx per subscriber perspective, (as shown in Figure 53 (a)). Additionally, as shown in Figure 53 (b), strategy *C1* also exhibits smaller OpEx per subscriber at all discount rates for penetration rates less than ~20%. These effects combine to make *C1* the dominant strategy at low penetration rates. As penetration increases however, the 100km network reach, coupled with the 1:512 split ratio, results many, long-distance fibers. The resulting fiber installation costs erode *C1*'s CapEx advantage, while fiber maintenance and repair erodes the OpEx advantage. Because strategy *B2* is also overvalued, at low penetration rates it is also less expensive than *A1* from a CapEx perspective, (although still more expensive than *C1*). Additionally, the total reach and splitter ratio associated with *B2*, 60km and 1:256, are significantly smaller than *C1*. As a result, as penetration increases the fiber required to connect additional customers in strategy *B2* is less than *C1*, until the resulting CapEx *and OpEx* per subscriber for B1 drops below that of *C1*[29]. The CapEx per subscriber effect results in the crossover from *C1* to *B2* at a penetration of 0.5 observed in Figure 53 (a); however, the additional *OpEx savings* causes this crossover to happen at the lower penetration value of ~0.37 observed in Figure 54. The fact that only one CapEx per subscriber-related crossover exists in Figure 53 (a), yet two crossovers are observed when we look at total costs per subscriber in Figure 54 means that the second crossover is due *exclusively to OpEx savings*. In this case, the OpEx savings associated with strategy *A1*, (shown in Figure 53 (b)) drives down the total costs per subscriber as penetration increases until *A1* becomes the least-cost strategy at a penetration of ~0.47, (the second crossover point observed in Figure 54).

Next, we address issue (2): why, when component costs are overvalued, (as they are in Scenario 1) is only a single strategy crossover observed as the discount rate increases? As discussed above, the second crossover point, indicating a shift from strategy *B2* to *A1* occurs as a result of the OpEx savings associated with strategy *A1*. Because OpEx is discounted over time, these savings diminish as the discount rate increases, resulting in only a single strategy crossover from *C1* to *B2*.

---

[29] Although strategy B1 results in less OpEx per subscriber at penetrations greater than 0.2, strategy A1 results in the *least* OpEx per subscriber of all the strategies; therefore, this OpEx result is not shown in Figure 53 (b), as this map only shows the least-cost OpEx strategy at each penetration and discount rate

Finally, issue (3) concerns why moving from overestimating component costs to undervaluing them, (from Scenario 1 to 3) results in different technology strategies which are independent of discount rate, but dependent on penetration. As discussed above, when $M=1$ for all three technologies, (estimated costs equal actual costs), strategies *A1*, *B2*, and *C1* exhibit almost identical CapEx at low penetrations in the low-density case, (Figure 44 (b)). As we move from overestimating actual deployment costs to underestimating them, our assumption that component costs for technology C are more uncertain than those for B translates to technology C strategies being *more undervalued*. However, the $m_c=10\%$ undervaluing in this Scenario 3 example, (corresponding to a multiplier $M_C = 1.1$ in column three of Table 65) is not large enough to change the strategy ordering at low penetrations. However, the cost premium of 5% for strategy *B2*, ($M_B=1.05$ in column two of Table 65), coupled with the OpEx advantage enjoyed by *A1 is* enough to eliminate the small penetration window in which *B2* was the least-cost solution, (between the two crossover points in Figure 54), leaving *A1* and *C1* as the two remaining strategies. This result is independent of discount rate because the reduced CapEx difference between *A1* and *B2*, (resulting from the extra component costs assigned to *B2*) is smaller than the discounted OpEx savings of *A1* over *B2*.

10.5.4 High-Density, High Data Demand Population Results

Figure 55 presents the three valuation scenarios provided in Table 65 for the high-density, high-data demand population case.
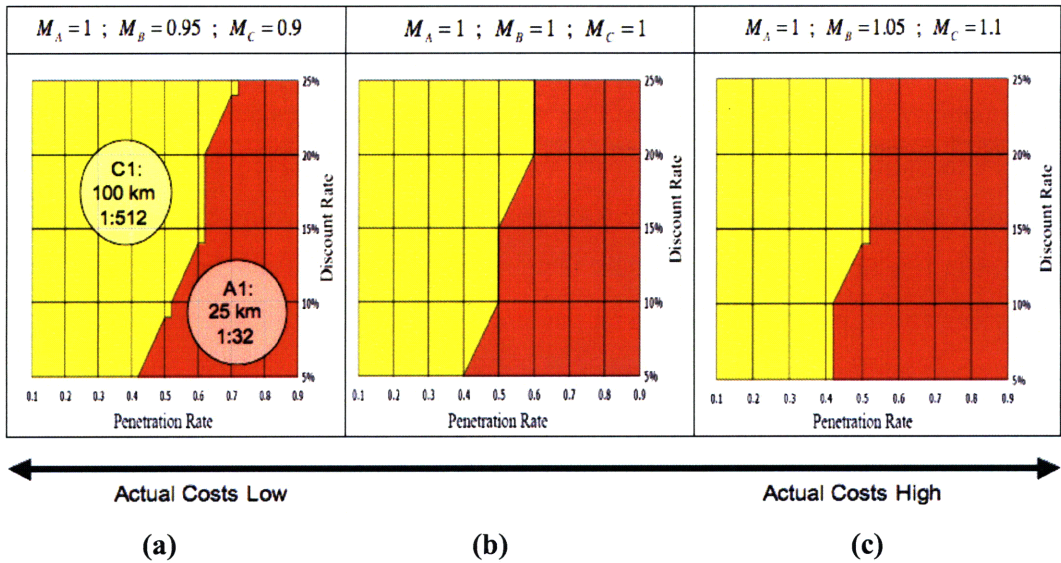
Figure 55: High-density technology strategies exhibiting least total discounted network costs as a function of penetration and discount rate for (a) Scenario 1: M_B=0.95, M_C=0.9; (b) Scenario 2: M_B= M_C=1; and (c) Scenario 3: M_B=1.05, M_C=1

The high-density results do not display the result complexity observed in the low-density case, instead exhibiting only minor variance from the Scenario 2 results developed in §10.4.1. The corresponding (a) CapEx per subscriber as a function of penetration and (b) OpEx as a function of penetration and discount rates are provided in Figure 56.
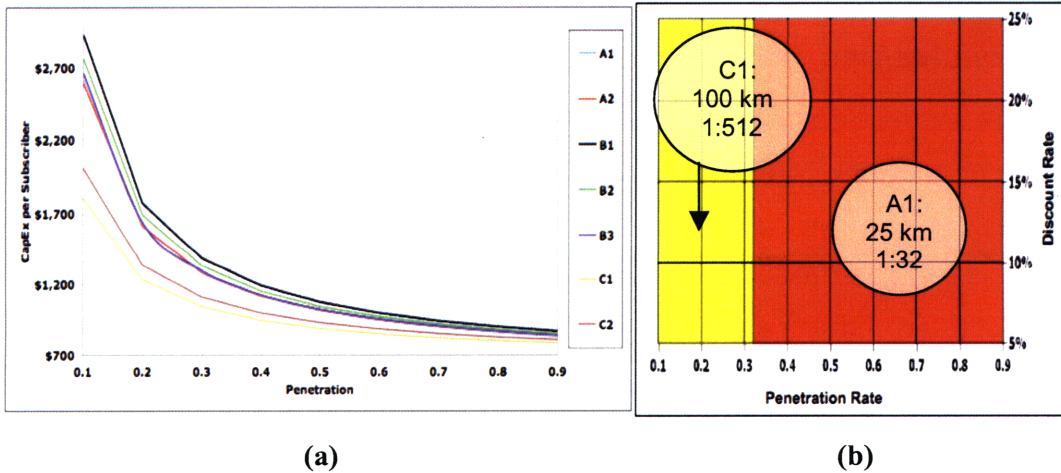

Figure 56: High-density, high-data demand (a) CapEx per subscriber vs. penetration with crossover emphasized, and (b) discounted OpEx per subscriber vs. both penetration and discount rate for Scenario 1

The results indicate that, in high-density regions, only two strategies provide least-cost alternatives, *A1* and *C1*. Whereas in the low-density case the high splitter port count of strategy *C1* was a liability, as subscribers are more spread out, resulting in longer fiber lengths to attach subscribers to splitters, in the high-density case, this high split ratio is an asset at higher penetrations, as subscribers are densely packed and therefore assigning them to splitter sites requires much less fiber. However, because the populations modeled also have regions of lower subscriber density, as penetration increases, the long-range strategy begins to require much more fiber to connect additional subscribers to splitter sites than *A1*, with its low splitter port count, (1:32) and corresponding short reach, (25km). While installing the additional fiber never drives the CapEx per subscriber for *C1* up enough to make *A1* the least-cost strategy in terms of CapEx, (no CapEx-related crossover in Figure 56 (a)) maintaining and repairing this additional fiber results in significant additional OpEx per subscriber for strategy *C1* at penetration rates greater than ~30%, (the OpEx-related strategy transition shown in   (b)). The effects of over or undervaluing technology B has no effect in this case, as the resulting CapEx and OpEx for all technology B strategies are significantly higher than either *A1* or *C1*. The effects of over or undervaluing technology C, as observed in Figure 55 are simply to shift the point at which the OpEx savings gained by implementing *A1* equals the initial CapEx savings *C1* provides.

11 Conclusions and Contributions

11.1 Conclusions

As the demand for broadband communications continues to expand and the technologies for satisfying that demand continue to evolve, access network operators are confronted with a host of challenging questions surrounding technology choice and network deployment. Often, Understanding the cost tradeoffs resulting from technology choice free from the constraints of either existing networks, or an individual carrier's preferred architecture, requires a generalized approach characterizing the relative cost tradeoffs for a range of population/demand demographics. Next-generation gigabit passive optical network (GPON) architectures will offer not only higher bandwidths enabling more products and services, but also better quality of service, enabling more efficient and reliable networks, thereby increasing subscriber satisfaction and retention rates. These benefits will require significant upfront capital investments however, which will both "lock in" the resulting technology through standardization and component

economies of scale and learning, and act as a barrier to future implementation of different technology choices. Given the decades long life cycles associated with these networks, it is important to characterize both the long-term cost implications of near-term technology choice decisions, and the long term benefits of investing in longer-term technology solutions. This thesis presented the hypothesis that characterizing lifecycle cost tradeoffs between CapEx and OpEx, and the role of population demographics and technology cost uncertainty as cost drivers, would impact not only technology choice, but also the way in which a particular technology is implemented.

The thesis results support the hypothesis that these often-complex tradeoffs impact technology strategy decisions. Additionally, we have demonstrated that these impacts exhibit a strong dependence on the subscriber demographics of the coverage region, including population and data demand densities. Therefore, we conclude that gaining insight into the value of different technologies requires characterizing these important population characteristics. Finally, our results indicate that even small changes in network component costs can dramatically alter technology strategy outcomes. Therefore, it is important to thoroughly characterize the cost uncertainty surrounding future components.

11.2 Contributions

This thesis explored the impact of relative lifecycle cost tradeoffs on technology strategy, and characterized two factors driving these costs. The methodology developed consists of three novel components which address gaps in the current literature in the areas of large-scale network design, multi-attribute population characterization, and cost modeling. Three technologies representing near, mid, and long-term FTTx GPON solutions, and seven implementation strategies were successfully dimensioned for two significantly different population demographics, each representing large coverage regions containing millions of subscribers. The methodology was able to successfully characterize how relative network topologies changed as a function of population attributes, revealing complex cost tradeoffs between technology strategies.

12 Carrier Recommendations

The results of our analysis suggest several recommendations to help carriers better characterize the cost tradeoffs accompanying technology choice and implementation strategies for GPON FTTx network deployments.

1. It is important to characterize lifetime network costs, including OpEx, for each technology under consideration: Multiple networks may exhibit similar CapEx but significantly different OpEx structures, potentially leading to non-trivial additional costs over time

2. When modeling the relative costs of multiple technologies, it is important to thoroughly characterize the population demographics of the coverage region: our work has demonstrated that minimum-cost technology strategies vary considerably depending on multiple population attributes, including population, household, and data demand densities and how these densities change throughout the coverage region

3. It is important to characterize component cost uncertainty: our analysis has shown that even small over or underestimates of component costs can significantly impact technology strategy decisions

13 Future Work

The next steps in this research will focus on expanding model capabilities to incorporate multiple sources of uncertainty, and identify sources of flexibility enabling technology strategy migration over time. The network model assumes once a particular technology is selected, no other technology may subsequently be implemented, and no migration exists between technologies. In reality however, sources of flexibility may exist enabling migration over time, or staged deployments. We hope to characterize these sources of flexibility and characterize their impacts on technology strategy over time. This analysis only modeled, in a coarse way, cost uncertainty surrounding components enabling future technologies. Future research will both better characterize this uncertainty, leading to a more accurate picture of how it affects technology choices, and incorporate and characterize additional sources of uncertainty affecting real-world networks, such as the demand for service. Finally, incorporating time-related production effects including efficiency improvements due learning, and cost reductions due to economies of scale, will provide a more realistic estimate of how costs impact technology choice.

126

References

Ahuja, R. K., T. L. Magnanti, et al. (1993). Network Flows: Theory, Algorithms and Applications. Englewood Cliffs, Prentice-Hall.

Balakrishnan, A., T. L. Magnanti, „ et al. (1991). "Models for planning capacity expansion in local access telecommunications networks." Annals of Operations Research 33: 239-284.

Balakrishnan, A., T. L. Magnanti, et al. (1995). "A decomposition algorithm for local access telecommunications network expansion planning." Operations Research 43: 58-76.

Boorstyn, R. R. and H. Frank (1977). "Large scale network topological optimization." IEEE Transactions on Communication 25(1): 29-46.

Carpenter, T. and H. Luss (2006). Telecommunications access network design. Handbook of optimization in telecommunications. M. Resende and P. Pardalos. New York, Springer.

Casier, K., S. Verbrugge, et al. (2007). Extending operational models to perform micro optimizations. 33rd European Conference and Exhibition on Optical Communications. Berlin, Germany.

Daskin, M. S. (1995). Network and discrete location: Models, algorithms, and applications. New York, Wiley.

Gabral, V., A. Knippel, et al. (1999). "Exact solution of multicommodity network optimization problems with general step cost functions." Operations Research Letters 25: 15-23.

Gavish, B. (1992). "Topological design of telecommunications networks." Annals of Operations Research 33: 17-71.

Gourdin, E., M. Labbe, et al. (2002). Telecommunication and Location. Facility Location: Application and Theory. Z. Drezner and H. W. Hamacher. New York, Springer.

Halpern, J., G. Garceau, et al. (2004). Fiber-to-the premises: revolutionizing the Bell's telecom networks, Bernstein Research and Telecordia research study.

Joao, P. R. P. (2007). A cost model for broadband access networks: FTTx versus WiMAX. Broadband Access Communication Technologies II, The International Society for Optical Engineering.

Kersey, K. (2006). Senior Vice President, Telecommunications, Jackson Energy Authority. T. Rand-Nash. Cambridge, MA.

Klincewicz, J. G. (1998). "Hub location in backbone/tributary network design: A review." Location Science 6: 307-335.

Klose, A. and S. Gortz (2006). "A branch-and-price algorithm for the capacitated facility location problem." European Journal of Operations Research 179(3): 1109-1125.

Konrad, N. (2007). Increasing transparency using process oriented OpEx modeling methods and standardized process frameworks. 35th European Conference and Exhibition on Optical Communications. Berlin, Germany.

LeBlanc, L. J., J. S. Park, et al. (1996). "Topology design and bridge-capacity assignment for interconnecting token ring LANS: A simulated annealing approach." Telecommunications Systems 6(1): 21-43.

Lee, Y., Y. Kim, et al. (2006). FTTH-PON splitter location-allocation problem. Seoul, Korea University, Kangwon National University.

Li, J., F. Chu, et al. (2009). "Lower and upper bounds for a capacitated plant location problem with multicommodity flow." Computers and Operations Research 36(11): 3019-3030.

Lin, C., Ed. (2006). Broadband: Optical Access Networks and Fiber-to-the-Home, Wiley.

Machuca, C. M., Ø. Moe, et al. (2007). Modelling of OpEx in network and service life cycles. 35th European Conference and Exhibition on Optical Communications. Berlin. Germany.

Melkote, S. and M. S. Daskin (2001). "Capacitated facility location/network design problems." European Journal of Operations Research **129**: 481-495.

Minoux, M. (2001). "Discrete cost multicommodity network optimization problems and exact solution methods." Annals of Operations Research **106**: 19-46.

Nagy, G. and S. Salhi (2007). "Location-routing: issues, models and methods." European Journal of Operations Research **177**: 649-672.

Nesset, N., D. Payne, et al. (2006). Demonstration of Enhanced Reach and Split of a GPON System Using Semiconductor Optical Amplifiers. Ipswich, UK, British Telecom.

Pasqualini, S., A. Kirstadter, et al. (2005). "Influence of GMPLS on network providers' operational expenditures: a quantitative study." IEEE Communications **43**(7): 28-38.

Prieß, S. and B. Jacobs (2007). OpEx modelling: a bottom-up approach for network operators. 35th European Conference and Exhibition on Optical Communications. Berlin, Germany.

Prins, C., C. Prodhin, et al. (2007). "Solving the capacitated location-routing problem by a cooperative Lagrangean relaxation-granular tabu search heuristic." Transportation Science **41**(4): 470-483.

Ramaswami, R. and K. N. Sivarajan (2002). Optical networks: A practical perspective, Morgan Kaufmann Publishers.

Rand-Nash, T., R. Roth, et al. (2007). Characterizing the cost tradeoffs in next generation fiber-to-the-home networks. 35th European Conference and Exhibition on Optical Communications. Berlin, Germany.

Randazzo, C. D. and H. P. L. Luna (2001). "A comparison of optimal methods for local access uncapacitated network design." Annals of Operations Research **106**: 263-286.

Sanso, B. and P. Soriano, Eds. (1999). Telecommunications Network Planning, Springer.

Sirbu, M. and A. Banerjee (2005). Towards technologically and competitively neutral fiber to the home (FTTH) infrastructure. Broadband Services: Business Models and Technologies for Community Networks. I. IChlamtac, A. Gumaste and C. Szabo. New York, Wiley.

Sridhar, V., J. S. Park, et al. (2000). "LP-Based heuristic algorithms for interconnecting token rings via source routing bridges." Journal of Heuristics **6**: 149-166.

Vaughn, M. D., D. Kozischek, et al. (2004). "Value of Reach-and-Split Ratio Increase in FTTH Access Networks." Journal of Lightwave Technology **22**(11): 5.

Verbrugge, S., S. Pasqualini, et al. (2005). Modeling operational expenditures for telecom operators. 2005 Conference on Optical Network Design and Modeling.

Vukovic, A. (2007). Enablers of cost-efficient network operation. 35th European Conference and Exhibition on Optical Communications. Berlin, Germany.

Vusirikala, V. and S. Melle (2007). OpEx benefits of digital optical networks. 35th European Conference and Exhibition on Optical Communications. Berlin, Germany.

Wagner, R. E., J. R. Igel, et al. (2006). "Fiber-based broadband-access deployment in the united states." Journal of Lightwave Technology **24**(12): 4526-4540.

Weldon, M. K. and F. Z. Zane (2003). "The Economics of Fiber to the Home Revisited." Bell Labs Technical Journal **8**(1): 181-206.