Architectures for a Space-Based Information Network with Shared On-Orbit Processing

by

Serena Chan

S.B. Computer Science and Engineering
Massachusetts Institute of Technology, 1999

M.Eng. Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2000

Submitted to the Engineering Systems Division
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Engineering Systems
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February 2005

Author ....
Engineering Systems Division
January 31, 2005

Certified by ....
Vincent W.S. Chan
Professor of Electrical Engineering and Computer Science and Aeronautics and Astronautics
Director, MIT Laboratory for Information and Decision Systems
Thesis Supervisor

Certified by ....
John Chapin
Visiting Scientist
MIT Laboratory for Information and Decision Systems
Committee Member

Certified by ....
L. Jean Camp
Associate Professor of Public Policy
John F. Kennedy School of Government
Harvard University
Committee Member

Certified by ....
Daniel E. Hastings
Professor of Aeronautics and Astronautics and Engineering Systems
Director, MIT Engineering Systems Division
Committee Member

Accepted by ....
Richard de Neufville
Professor of Engineering Systems
Chair, Engineering Systems Division Education Committee

Architectures for a Space-Based Information Network with Shared On-Orbit Processing

by

Serena Chan

Submitted to the Engineering Systems Division
on January 31, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Engineering Systems

Abstract

This dissertation provides a top level assessment of technology design choices for the architecture of a space-based information network with shared on-orbit processing. Networking is an efficient method of sharing communications and lowering the cost of communications, providing better interoperability and data integration for multiple satellites. The current space communications architecture sets a critical limitation on the collection of raw data sent to the ground. By introducing powerful space-borne processing, compression of raw data can alleviate the need for expensive and expansive downlinks. Moreover, distribution of processed data directly from space sensors to the end-users may be more easily realized.

A space-based information network backbone can act as the transport network for mission satellites as well as enable the concept of decoupled, shared, and perhaps distributed space-borne processing for space-based assets. Optical crosslinks are the enabling technology for creating a cost-effective network capable of supporting high data rates. In this dissertation, the space-based network backbone is designed to meet a number of mission requirements by optimizing over constellation topologies under different traffic models. With high network capacity availability, space-borne processing can be accessible by any mission satellite attached to the network. Space-borne processing capabilities can be enhanced with commercial processors that are tolerant of radiation and replenished periodically (as frequently as every two years). Additionally, innovative ways of using a space-based information network can revolutionize satellite communications and space missions. Applications include distributed computing in space, interoperable space communications, multiplatform distributed satellite communications, coherent distributed space sensing, multisensor data fusion, and restoration of disconnected global terrestrial networks after a disaster.

Lastly, the consolidation of all the different communications assets into a horizontally integrated space-based network infrastructure calls for a space-based network backbone to be designed with a generic nature. A coherent infrastructure can satisfy the goals of interoperability, flexibility, scalability, and allows the system to be evolutionary. This transformational vision of a generic space-based information network allows for growth to accommodate civilian demands, lowers the price of entry for the commercial sector, and makes way for innovation to enhance and provide additional value to military systems.

Thesis Supervisor: Vincent W.S. Chan
Title: Professor of Electrical Engineering and Computer Science and Aeronautics and Astronautics

Director, MIT Laboratory for Information and Decision Systems

Committee Member: John Chapin
Title: Visiting Scientist
MIT Laboratory for Information and Decision Systems

Committee Member: L. Jean Camp
Title: Associate Professor of Public Policy
John F. Kennedy School of Government
Harvard University

Committee Member: Daniel E. Hastings
Title: Professor of Aeronautics and Astronautics and Engineering Systems
Director, MIT Engineering Systems Division

Acknowledgments

I wish to thank my advisor, Vincent W.S. Chan, for his patience, encouragement, inspiration, and assistance throughout my graduate studies. I also wish to thank the members of my doctoral committee, John Chapin, L. Jean Camp, and Daniel E. Hastings, for their comments, suggestions, and advice.

I am grateful to my colleagues and friends at the Laboratory for Information and Decision Systems and the Engineering Systems Division for their friendship and care.

Last, but not least, I wish to thank my husband, Abraham R. McAllister, and my family for their unwavering love, patience, and constant encouragement.

This dissertation is based upon work supported by the National Reconnaissance Office.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The concept of satellite communications was pioneered by Arthur C. Clarke in 1945 with publications in the magazine *Wireless World*. He described the use of satellites in geostationary orbit to relay radio signals. In 1946, the United States (U.S.) national security space program began with a study on space feasibility that was carried out by the RAND Corporation for the U.S. Army Air Forces [12]. The classified study revealed the possible commercial use of synchronous communications satellites, but had little effect as the report stayed secret [30]. Soon thereafter, the National Security Act of 1947 was approved by the U.S. Congress. Responsibility for all space-related pursuits was delegated to the Department of Defense (DoD) Research and Development Board's Committee on Guided Missiles, an organization overseen by the U.S. Army and U.S. Navy. Within the next couple of years, each of the American armed services (the U.S. Army, U.S. Navy, and U.S. Air Force) had initiated separate space programs [12].

The October 1957 launch of Sputnik I by the Soviet Union greatly influenced U.S. defense and security planning [12]; sparking a space race between the two countries as the technical feat of launching an artificial satellite caught the world's attention and the American public off-guard. The illusion of a technical gap provided the momentum for increased expenditure and ushered in new political, military, technological, and scientific changes to reap the benefits, profits, and prestige associated with air and space research and development (R&D). By July 1958, the U.S. Congress passed the National Aeronautics and Space Act, which created the National Aeronautics and Space Administration (NASA) to pursue aerospace activities and create technical and scientific educational programs [65].

From then on, significant Congressional and public support stimulated the growth of the U.S. space program [12].

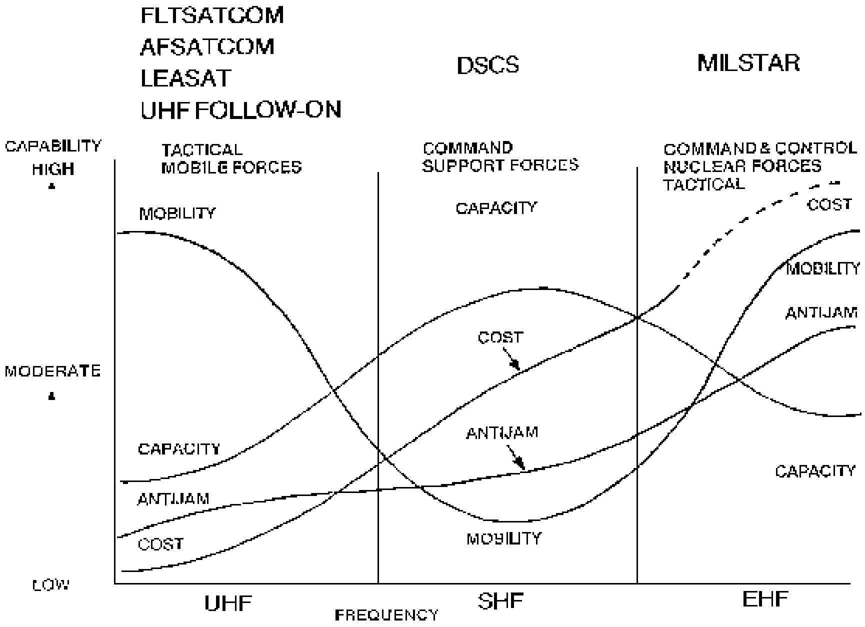
1.1 Military Satellite Communications

Military and government organizations make extensive use of satellites for a mixture of communications, remote sensing, imaging, navigation, positioning, and other services including more secret applications for intelligence work or missile guidance. Satellite communications appeal to the military because it offers a highly reliable, high capacity service over a broad coverage area. Satellite service can be available at short notice in virtually any part of the globe without any dependence on local communications infrastructures in the region. The diverse nature of military communications requires satellite systems to handle a broad range of users, traffic, and scenarios. Therefore, military satellite communications include both low data rate mobile traffic and high capacity fixed links.

Current U.S. military satellite communications systems operate in the following radio frequency ranges: (1) ultra high frequency (UHF), from 300 MHz to 3 GHz, (2) super high frequency (SHF), from 3 GHz to 30 GHz, and (3) extremely high frequency (EHF), from 30 GHz to 300 GHz [27]. A summary of the characteristics of the primary UHF, SHF, and EHF military satellite communications systems is shown in Figure 1-1. The diagram illustrates the relative features of each system type with respect to mobility, capacity, anti-jam, and cost.

Principal UHF satellites include the Navy's Fleet Satellite Communications System (FLTSATCOM), the Air Force Satellite Communications (AFSATCOM), the Navy's Leased Satellite (LEASAT) system, and the UHF Follow-On satellites that replaced FLTSATCOM and LEASAT [4]. UHF satellites mainly support mobile tactical users who use small, portable antennas and require moderately low capacity (enough to support single channel voice circuits). UHF satellites have a fairly low anti-jam capability, although they are superior to SHF and EHF satellites in terms of foliage penetration, power efficiency, and low cost user terminals [27].

The Defense Satellite Communications System (DSCS) is a SHF satellite system that supports command and control (C2) and high volume data transmissions, which include phone conversations, Internet data, and Global Command and Control System (GCCS)



Source: [4] *Army Space Reference Text*, http://www-tradoc.army.mil/dcscd/spaceweb/chap07b.htm

Figure 1-1: Comparison of U.S. military satellite communications systems.

data [4, 27]. Because users need larger satellite dishes to transmit and receive on the high bandwidth connections, SHF satellites are more expensive than UHF satellites [27]. The large size of the antennas and user terminals and the power to operate them restrict user mobility [4]. Nevertheless, SHF satellites provide much higher data rates and are more jam resistant than UHF satellites.

The Military Strategic and Tactical Relay (MILSTAR) satellite communications system is a joint service satellite communications system supporting high priority military users (e.g., nuclear forces, strategic level C2, and tactical users) with secure, jam resistant, global communications [4, 34]. MILSTAR provides both low data rate communications (75 bps to 2.4 kbps) and medium data rate communications (4.8 kbps to 1.544 Mbps) [34]. Uplink communications are provided at UHF (300 MHz) and EHF (44 GHz) while downlink communications are provided at UHF (250 MHz) and SHF (20 GHz) [13]. MILSTAR satellites have onboard processing and switching. Multiple MILSTAR satellites are crosslinked to one another to eliminate superfluous transmissions to the ground. These crosslinks operate at approximately 60 GHz [13]. EHF communications systems are a more recent development, and the cost of EHF satellites is quite high. Along with technological advancements (e.g., increased capabilities and small user terminals to enhance mobility), significant future applications may be expected, primarily for highly survivable communications needs [30].

Military reliance on space systems have considerably increased throughout the last decades. Dependence on space systems has traditionally been warranted on the basis of cost and mission effectiveness. Space systems have either been the least expensive means or the only approach to provide a vital military ability. Military space systems however are not inexpensive, ranging from hundreds of millions of dollars to billions. The relatively high unit costs, coupled with the sometimes short or limited spacecraft lifetimes, constrain the amount of military space system production. The currently fielded military satellite communications systems will require replenishment in the near future. The old budgets, old space mission designs, and justifications may no longer apply, so new methods for conducting space programs must be sought after and realized.

1.2 Issues

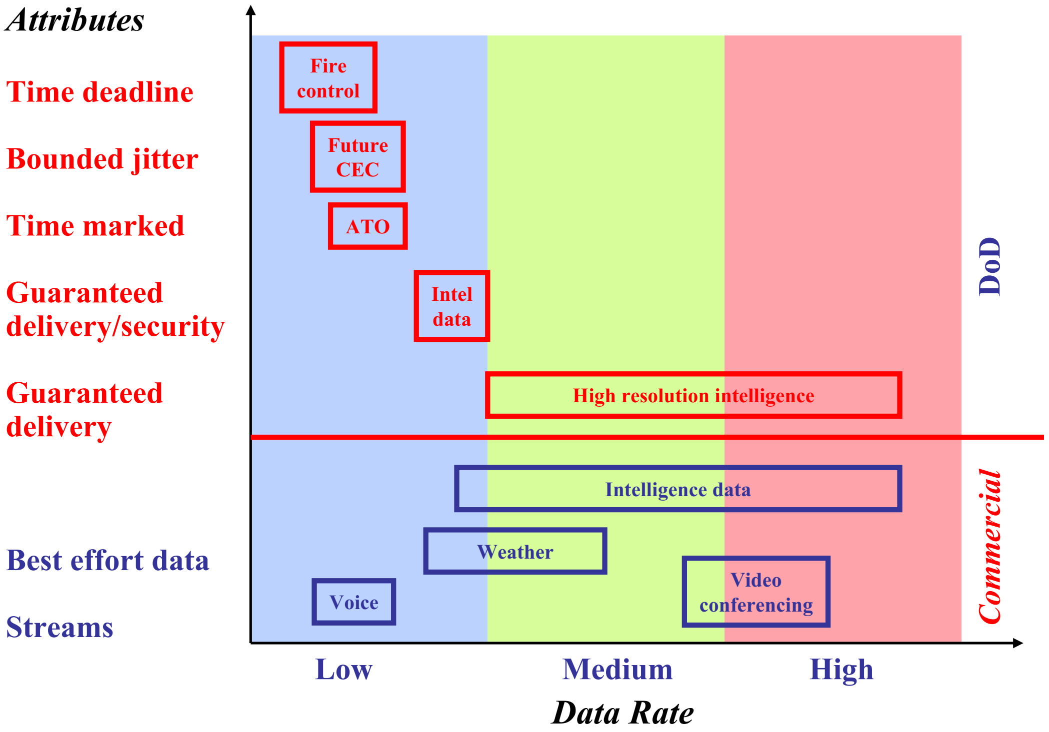
Space systems are generally characterized as being slow to design and build, very expensive, and difficult to upgrade on orbit. The only effective way to upgrade them is to implement software updates, though the new functions are fundamentally limited by the capability of the initial hardware that was deployed. Developing space systems is an expensive and risky investment as the costs can be overwhelming, with estimates ranging from $4 billion to more than $12 billion for global mobile communications space systems [10]. The desire for designing a robust space system (e.g., high reliability through radiation hardness and redundancy to mitigate effects of the space environment) with a long lifetime will further increase the cost.

Methods of cost reduction for military satellite communications include using simpler satellites with less expensive launches, using as much commercial-off-the-shelf (COTS) components, and reconsidering the degree of radiation hardening required. New choices of space modules in a spacecraft for greater performance are available due to the progress in hardware and software technologies. Satellite onboard processing and the use of infrared and optical frequencies for data transmission are growing in military and scientific interest. Currently, infrared and optical frequencies are not universally accepted as the exclusive privilege of the military, unlike the SHF and EHF bands which are commonly acknowledged in this manner within the North Atlantic Treaty Organization (NATO) [30]. Thus, an architect not only has to base his or her decisions on the options available today, but also has to keep in mind the expected changes in technology and policy during the life of the system.

Space mission analysis and design is an evolving process. Paradigm shifts are anticipated in this process as a result of increasing technological maturity, growing use of satellite onboard processing, and an ongoing emphasis on low cost missions. The demand for military satellite communications is constantly increasing as a result of greater communications needs and progressively more complex end-user demands (e.g., sensors and computers transmitting digital information as part of Command, Control, Communications, Computer, Intelligence, Surveillance, and Reconnaissance [C4ISR] networks) [30]. There are a number of military missions that can be efficiently achieved by space systems if the costs of space systems can be lowered and/or their performance capabilities increased. The total program cost of a mission can be reduced by investing in sophisticated technology that increases capacity, connectivity, and reliability of the spacecraft. There is also interest in using available commercial satellites. However, the major disadvantages of military usage of commercial satellites are issues of access and control. Commercial organizations may be unwilling to assume potentially life-threatening military communications in times of crisis. Additionally, the operation of the satellites would no longer be under national control. Because many investment opportunities have been greatly exploited and are becoming marginal, new approaches that have large potential payback margins must evolve.

1.3 Thesis Motivation

Each of the U.S. military branches has very specific and usually different applications and quality of service (QoS) requirements. Figure 1-2 illustrates the attributes of several different applications with respect to the required data rates and whether it is a military or commercial application. These differences are the consequences of varied information processing, handling, and dissemination protocols. A broad range of programs are being carried out to replace all existing satellites and launch vehicles in the near future. Figure 1-3 shows an overall roadmap of space systems from the 2001 report released by the Commission to Assess U.S. National Security Space Management and Organization. Programs span the range of communications, navigation, surveillance and threat warning, meteorology, launch, imagery intelligence (IMINT), signals intelligence (SIGINT), and relay. Systems include Advanced

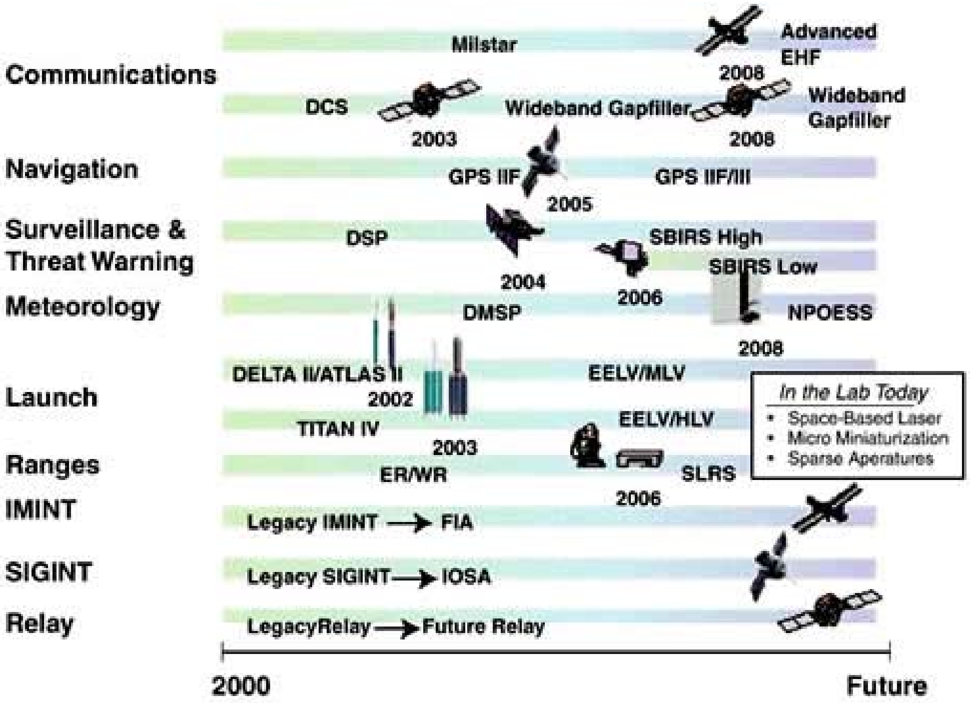


Abbreviations: CEC = Cooperative Engagement Capability; ATO = Air Tasking Order

Figure 1-2: Quality of Service (QoS) attributes for DoD and commercial communications.

Extremely High Frequency (AEHF), Wideband Gapfiller Satellite (WGS), Advanced Polar Satellites (APS), Advanced Wideband System (AWS), Mobile Users Objective System (MUOS), Transformational Communications Satellites/Architecture (TCS/TCA), Future Communications Architecture upgrades, Global Positioning System (GPS) IIF Modernization and GPS III, Integrated Overhead SIGINT Architecture (IOSA) upgrades, National Polar-orbiting Operational Environmental Satellite System (NPOESS), Space-Based Infrared System (SBIR) High and SBIR Low, Space-Based Space Surveillance System (SBSS), Space Tracking and Surveillance System (STSS), Space-Based Radar (SBR), and Future Imagery Architecture (FIA) [71].

A new space systems architecture is motivated by the desire to create a horizontally organized space-based network backbone infrastructure, as a result of adopting a networking paradigm. Networking is an efficient method of sharing communications among multiple users. A consolidation of all the different communications assets can improve spacecraft interoperability and levels of inter-spacecraft communications. As experienced with the terrestrial Internet infrastructure, the model of a horizontally integrated infrastructure can



Source: [71]. King Space Research, *Worldwide Military and Intelligence Systems in the Pipeline*, November 30, 2004.

Figure 1-3: Space systems roadmap.

improve sharing of assets and reduce costs [18]. Cost reductions from the sharing of assets can be afforded due to the elimination of duplication efforts by multiple organizations.

This dissertation provides a top level assessment of technology design choices for the architecture of a space-based information network with shared on-orbit processing with respect to their impact on the ability of the system to grow, in terms of usability, flexibility, scalability, and cost. Several architectural concepts and enabling technologies are discussed in Chapter 2. Optical intersatellite links (or crosslinks) are the enabling technology that supports high speed communications in space. Concepts of space-based data processing, shared on-orbit processing, and data transmission are presented. Leveraging COTS components is a trade-off between providing greater performance capabilities at the expense of shorter lifetimes due to the radiation exposure in the space environment. To enable component replenishment rather than complete satellite replacement, on-orbit servicing is explored. Periodic replenishment of a processing satellite provides system upgrade flexibility.

In Chapter 3, the provisioning of high speed space-to-space communications between space-based assets and networked processing resources is analyzed. The space-based information network backbone is designed to meet a number of mission requirements (e.g., high data rate, high connectivity, and low latency) at reasonable cost by analyzing several constellation topologies under different traffic models. The architectural concept of decoupled, shared, and distributed space-borne processing resources for space-based assets is a significant paradigm change in space mission design (i.e., separated processing capabilities and implementation of COTS components). This design can lead to significant performance for several space applications (e.g., SIGINT and synthetic aperture radar [SAR] processing), as discussed in Chapter 4.

Design choices and example implementations for the processing satellite are discussed in Chapter 4. The examples presented use Intel Pentium III processors and Virtex-II Pro/Virtex-II Pro X FPGAs as data are readily available. The value of space-borne processing is further enhanced by a generic nature that promotes sharing. The commonality of SIGINT and SAR applications demonstrate that the same processing architecture can be used. Novel ways of using the space-based information network with shared on-orbit processing are then briefly explored. The applications that can revolutionize satellite communications and space missions include distributed computing in space, interoperable space communications, multiplatform distributed satellite communications, coherent distributed space sensing, multisensor data fusion, and restoration of disconnected global terrestrial networks.

The role of the national government in the development of new satellite technologies and systems is assessed in Chapter 5. Case studies regarding past direct government funded defense projects (i.e., GPS, U.S. Interstate Highway System, and Advanced Research Projects Agency Network [ARPANET]) and global mobile satellite communications businesses are studied in order to recommend a technology policy regarding the development of a space-based information network infrastructure. An infrastructure built with a generic nature can satisfy the goals of interoperability, flexibility, scalability, and allows it to be evolutionary. Last but not least, conclusions are provided in Chapter 6. The design of future satellite data networks will require a networking paradigm, sharing of assets, and coordination and communications between military, economic, and political interests.

1.4 Summary

Procurement of military systems has long been recognized as very expensive. Future space systems will have to be economically designed and configured. Optical intersatellite links may be used to create an economical space network to expand the coverage area of a satellite system, remove the need for intermediate ground stations which improves overall system survivability, or connect satellites in different orbits. Advances in the computer industry and in telecommunications applications are also driving innovations in space communications and networks. On-orbit processing can increase the performance of space systems. The convergence of computer technology and space technology creates the potential for new capabilities (e.g., observation, navigation, communications, surveillance, and reconnaissance) in orbit, although there are still many engineering and business barriers to the development and deployment of space systems. Design choices have to be analyzed because they impact the ability of a system to grow, in terms of usability, flexibility, and scalability. The architecture goal of this dissertation is to provide an economical, ubiquitous, high data rate communications network with processing to enable future space systems and applications. Future military satellite communications are likely to include the exploitation of optical intersatellite links, enhanced space-borne processing capabilities, improved network management and control, cost reductions, higher bandwidths, and extended coverage.

Chapter 2

Architectural Concepts and Enabling Technologies

The envisioned infrastructure for the space-based information network is a highly connected global and heterogeneous network, integrating fiber, wireless, and space communications. Although the future space-based information network infrastructure is currently being tailored for carrying out American military missions, it has the potential of providing numerous global mobile voice and data services for the private sector. Global mobile satellite voice and data services are ideal for industrial applications such as heavy construction, defense/military, emergency services, maritime, mining, forestry, oil and gas, and aviation. Studies of the satellite multimedia business in the U.S. have predicted a convergence of broadcasting, entertainment, Internet, and telecommunications services. Worldwide current events (e.g., terrorist attacks and power grid outages) have bolstered the interest in having satellites serve as a back-up communications system for cities during times of crisis (e.g., occurrences of natural and unnatural disasters that include earthquakes, tsunamis, blackouts, and war). Satellite communications networks offer advantages over terrestrial and cellular networks as they have high availability and are difficult direct targets for man-made weapons. Satellite communications can thus arise to be a significant alternative or complementary communications technology to terrestrial and cellular communications technology worldwide.

The space-based information network architecture, connecting multiple users and space-based assets, must be designed to have a long lifetime (e.g., over 25 years), therefore it must have the following characteristics:

- *Interoperability*: Ability to operate with other networks.
- *Flexibility*: Ability to support the full range of operations and missions.
- *Scalability*: Ability to expand the number of users or increase the capabilities of the system without making major changes to the systems design or application software.
- *Evolutionary*: Ability to take advantage of multiple generations of technologies.

The design of the space-based information network builds on the idea that optical space communications at very high rates between satellites is currently feasible. The invention of such a radical technology building block can revolutionize space systems that may use the network as a critical subsystem. Examples of these space systems include those offering communications services or remote sensing. This chapter briefly discusses the enabling technologies (space laser communications and powerful space-borne processing) and satellite replenishment strategies that can be brought together to develop an integrated space network for a transformation in network performance and applications.

2.1 Space Laser Communication Technology

The goal of any communications system is the exchange of information from one site to another. The transmission of information is usually achieved by modulating (superimposing) the information onto a carrier (electromagnetic wave) at the source, propagating the modulated carrier to the destination, and demodulating the information that is received at the destination to recover it [37]. These systems are often designed by the location of the carrier frequency in the electromagnetic spectrum. The electromagnetic spectrum is illustrated in Figure 2-1. Frequencies in the radio spectrum and the optical spectrum are especially of interest for space communications.

Space laser communications development for very high data rate applications originated in the early 1960s. The first crosslinks for intersatellite communications were radio frequency (RF) or microwave systems. The Massachusetts Institute of Technology (MIT) Lincoln Experimental Satellites (LES) 8 and 9, carried the first 38 GHz RF crosslinks in 1976 with data rates of 100 kbps. Currently, the NASA Tracking and Data Relay Satellite (TDRS)

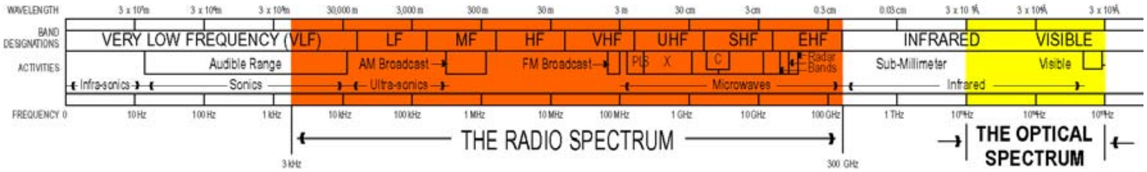


Figure 2-1: Electromagnetic spectrum: radio frequencies and optical frequencies.

system serves manned space exploration and scientific missions with high data rates of 300 Mbps and 800 Mbps. The interest in optical communications for space applications stems from the much higher operating frequency, nearly 7-8 orders of magnitude higher than RF systems [52]. At this very high range, optical frequencies can provide higher rates for data transfer. An architect must weight the advantages and disadvantages between RF and optical in deciding the operating frequency for the space-based information network.

2.1.1 Radio Frequency (RF) vs. Optical

Laser communications systems operating at optical frequencies offer many advantages over RF systems due to the very large difference in their wavelengths. Optical frequencies are thousands of times shorter in wavelength than those in the RF region, as highlighted in Figure 2-1. The beamwidth attainable with an optical communications system is narrower than that of a RF system by the same wavelength ratio in antenna diameters. Due to the very narrow beam from the transmitter, the laser beam is brighter at the receiver by the square of the ratio for a given transmitter power level. An architect can take advantage of the brighter beam or higher gain by designing an optical communications system with a much smaller antenna and transmit less power than an RF system at the same data rate. However, the narrower laser beam is a double-edged sword as it is much harder to point. Thus, acquisition of satellite terminals is much more difficult. In addition to the advantages of smaller antenna size, lower weight, and lower power, optical communications systems are capable of much higher data rates than RF (e.g., $\sim$ 100 Gbps) because the carrier frequencies of optics are very high (e.g., $\sim$ 200 THz) [20].

Needless to say, there are applications where RF communications fare better than optical communications. For broadcast applications, a much larger angular area of coverage can be provided by RF systems with broad-beam capability than with optical communications links

[53]. Naturally, there are limitations to optical technology. It can never totally replace RF systems for space-to-ground communications due to atmospheric conditions (e.g., rain, cloud coverage, scintillation, absorption, and scattering). Given that signal attenuation enters into the availability equation for optical communications links, optical communications are not as well-suited as RF for single high availability ground station applications. Downlink diversity (via multiple downlink sites and transmitters) is required for optical downlinks to obtain high availability.

2.1.2 Design Recommendations

Satellite-to-satellite communications are one area where optical communications can compete successfully with RF systems. The design issue of intersatellite links is especially important for a network of relay satellites where there will be multiple apertures. The first intersatellite links were microwave or RF systems. It can be shown that RF aperture sizes become quite large at rates above 100 Mbps [52]. RF links are generally better for data rates less than about 100 Mbps because of their lower mass and power. The Iridium global mobile satellite communications system uses RF crosslinks for the interconnection of its satellite constellation to provide voice service. At rates above 100 Mbps, optical crosslinks have a clear advantage because the carrier frequencies of optics are very high (e.g., $\sim$ 200 THz). Each optical carrier can accommodate very high data rates (e.g., $\sim$ 100 GHz). The possibility of using wavelength division multiplexing (WDM) can further increase the data rate per optical beam. For these reasons, there is no reservation that as optical crosslink technology matures, it will greatly revolutionize space systems architectures. Optical crosslinks can provide connectivity between satellites on opposite sides of the earth without expensive intermediate ground relay stations, and will be a key technology to interconnect data satellite constellations into a worldwide coverage backbone.

2.2 Data Processing

Space missions are comprised of two types of data: (1) *mission data* - the information that is produced, transmitted, or received by the mission payload and (2) *housekeeping data* - the information necessary to sustain the mission (e.g., spacecraft orbit and altitude, battery temperature and charge status, and spacecraft equipment condition and status). While mission data may be intermittent and have very high data rates, housekeeping data may be constant and have very low data rates [92].

Data delivery systems are necessary for both mission and housekeeping data. Mission data must eventually be processed before dissemination to end-users. A well-designed data delivery system is essential for transmitting large amounts of raw data from various sensors. Raw data should be efficiently transformed into valuable and useful information for the end-user in a timely fashion. The main trade-offs associated with data delivery are [92]:

- *Space vs. ground processing*: How much of the data processing is done onboard the spacecraft? How much is done on the ground at mission operations? How much is done by the end-user?

- *Central vs. distributed processing*: Is there one large central computer onboard the spacecraft? Is there one large computer on the ground that processes everything or are there several computers that communicate with each other? Distributed processing among several satellites for one application is a new space paradigm that is suggested in this dissertation.

2.2.1 Space-Based Processing vs. Ground-Based Processing

Traditionally, as satellite onboard processing was limited or non-existent, the majority of the collected data was processed at ground stations or mission operations facilities. Today, the availability of powerful satellite onboard processors provides increased capabilities in space. An architect for future space missions must understand and consider how much data to process on the spacecraft or on the ground or by the end-user. End-users may be located on the ground, in the air, or in space. The main issues associated with space-based processing and ground-based processing are [92]:

1. *Autonomy:* How much human involvement is required in order to provide intelligent analysis? Ground processing may be necessary if human interaction and interpretation is critical. Regardless, autonomous processing can be done either in space, on the ground, or by the end-users.

2. *Data latency:* How much delay can be tolerated for data delivery to the end-user? Non-critical data can handle the transmission delays for ground processing. However,

pace-borne processing is important for data that require a latency of a few fractions of a second.

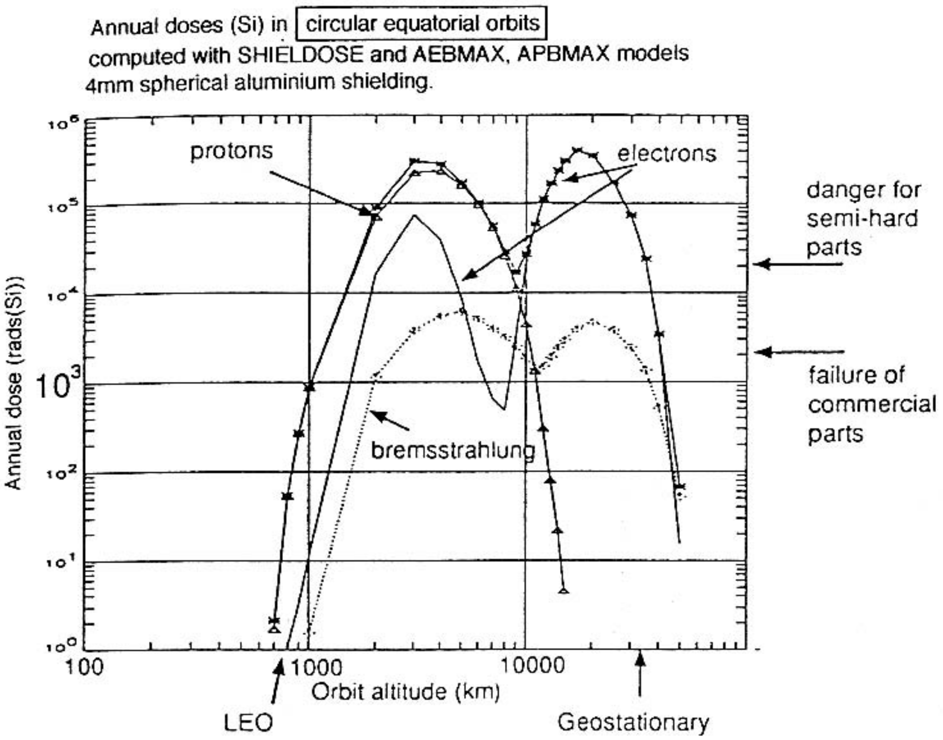
3. *Communications bandwidth:* How much data require transmission? For large amounts of raw data collected by a sensor, downlink transmission to the ground for processing will cause an enormous communications bottleneck and will drive up mission costs by requiring expensive and expansive downlinks. Therefore, data processing and compression should be conducted in space before dissemination.

4. *Location of end user:* Where is the location of the end-user? If end-users are situated in the air or in space, sending data to the ground for processing and returning the results can be costly and complex.

The space-based processing vs. ground-based processing trade is of utmost importance for future space missions. Trade-offs must be analyzed to minimize operations and end-user costs. Future space missions will employ state-of-the-art critical technologies. Long-lived space missions or time-sensitive space applications will require some level of automated processing. As remote sensing capabilities advance, they emphasize a growing need to communicate the vast amount of collected data from the field to users in other locations. Space-based processing (i.e., data compression or processing of raw data into valuable information) allows for a drastic reduction in the volume of data relayed to end-users everywhere.

2.2.2 Space Environment

The space environment can strongly influence the performance, size, weight, complexity, cost, and lifetime of operational space systems. To ensure survivability of electronic systems in spacecrafts, they are usually shielded from the naturally occurring radiation in space. There are two effects of radiation: (1) degradation due to total ionizing dose (TID) and (2) malfunctions brought on by single event upsets (SEUs). These two failure modes are induced by fundamentally different mechanisms. While both electrons and protons contribute to the TID effect, the major contribution of either trapped electrons or trapped protons depend on the orbit. As shown in Figure 2-2, protons dominate in low altitude orbits (less than approximately 800 km) while electrons dominate in high altitude orbits. *Radiation dose* is defined as the quantity of energy deposited in material and is dependent on the type of radiation and its energy as well as the material itself. The unit measure for radiation dose



Source: [35] Peter Fortescue and John Stark, ed. Spacecraft Systems Engineering, John Wiley & Sons Ltd, West Sussex, 1995, p. 31.

Figure 2-2: Radiation levels encountered in the space environment.

is *Rad*, for radiation absorbed dose. A Rad is the measure of any kind of radiation which deposits 10^{-2} J per kg of material. For electronics, the radiation dose is typically denoted in rad (Si) because silicon is the material most frequently used in their assembly [87].

2.2.2.1 Total Dose Effects

Total radiation dose is comprised of three components: proton dose, electron dose and bremsstrahlung X-ray dose. Bremsstrahlung X-ray is a product of the interaction of electrons with the shielding material. The amount of ionizing radiation that the electronic component can tolerate before failure is given by the TID, usually measured in units of krads. Device tolerance to total dose radiation provides the system designer with an estimate of lifetime survivability in the space environment. In radiation tolerance testing, devices are subjected to doses of either alpha, beta, or gamma radiation from concentrated sources of radioactive materials. Gamma-emitting sources that are often used include Co^{60} (cobolt) and Cs^{137} (cesium) [87]. To lessen the effects of total dose radiation, the general strategies include: (1) use appropriate fabrication process technology, (2) devise appropriate circuit design and layout, and (3) apply package shielding [6].

2.2.2.2 Single Event Effects

Single event effects due to radiation exposure in the space environment can be categorized into three classes [35, 51, 87, 92]:

1. *Single event upset* (SEU): A SEU arises when radiation-induced currents trigger a memory device to alter its state (i.e., a bit of data is flip from zero to one or one to zero). Although the data stored in the device is corrupted, the device is not broken and can still function correctly. Sophisticated error detection and correction codes can be implemented to guarantee the validity of data. Furthermore, SEUs are statistically guaranteed to appear on any device that proves to be vulnerable to them. Depending on the device, SEU rates can range from 10^{-10} errors/bit-day to 10^{-4} errors/bit-day in the Geosynchronous Earth Orbit (GEO) environment.

2. *Single event latchup* (SEL): A SEL can occur in many semiconductors having `npnp` or `pnpn` elements, particularly bulk complementary metal oxide semiconductor (CMOS) devices. The operation of the device can be ruined by a current loop induced by a single particle. Power cycling (turning power off and on) will reset the device and allow it to function correctly.

3. *Single event burnout* (SEB): A SEB can occur in power metal oxide silicon field-effect transistors (MOSFETs) from a large current surge (e.g., the drain-to-source voltage surpasses the breakdown voltage of the material). Unlike SEUs and SELs, a burnout causes permanent device failure.

Single event effects are simulated with the use of Cf^{252} (californium), which emits alpha particles [87].

2.2.2.3 Radiation Mitigation Techniques

A system designer has an assortment of options available to mitigate the radiation effects of the space environment. The main mitigation techniques involve parts selection, shielding, and component derating or redundancy [5, 87]:

- *Design/Parts Selection:* Selecting space-qualified components is key for designing a radiation-hardening spacecraft.

- *Shielding:* The required amount of spacecraft shielding is determined by computing the dose rate for the desired orbit as a function of shield thickness. Shielding prevents charged particles from interacting with the devices. The cost of shielding is dictated by the weight of the shielding material and the testing involved with performance validation. Spot shielding at the component or subsystem level is an alternative to shielding the entire system. Spot shielding generally influences the form factor of the device and may necessitate modifications in the device to have room for the additional layer of material.

- *Component Derating or Redundancy:* Component derating refers to designing the device to take into account its behavior under irradiation. A system designer assumes the worst-case values observed under total dose testing (with some margin) and then designs the device accordingly. Component redundancy refers to designing the system with extra circuitry or parts to serve as backups. Redundancy can improve single event tolerance. Because a SEU is caused by a single particle rather than the cumulative effect of many particles, it is much less likely that two devices will be damaged at the same time.

Systems approaches to SEU-hardening can be divided into three general categories: (1) error toleration, (2) error correction, and (3) error prevention [5, 35]. Error tolerance is expensive and entails identifying tolerance levels for various parts of the system, calculating maximum permissible error rates, and designing and building each part of the system within these constraints. Systems approaches of error correction include: error-detecting and error-correcting codes, self-checking circuits, redundant units, and serial calculation with reasonableness testing, checkpoint storage and roll-back for recovery (software solution), and repetitive execution and watchdog timers (time-related solutions). Error prevention is accomplished through the use of components that will not upset. However, there are only a limited number of SEU-hard devices.

2.2.2.4 Radiation-Hardened Processors vs. Commercial-Off-The-Shelf (COTS) Processors

Selecting suitable spacecraft electronics begins with the functional and performance requirements of the equipment and a comprehension of the radiation environment. Recall that the natural radiation environment varies with the orbital altitude. For processing in space, an architect has a choice between radiation-hardened processors and commercially available processors that may be radiation tolerant. Radiation-hardened components are generally more expensive than their commercially graded counterparts due to qualification costs and low volume manufacturing. The number of suppliers in the space electronics market traditionally has been small. Moreover, the radiation-hardened electronics marketplace does not have similar economic drivers as in the commercial marketplace. The commercial marketplace is based on "pull technologies" while the space marketplace is based on "push technologies" as it is mainly dependent on DoD and NASA funding [7]. With the DoD as the primary customer with rigorous reliability requirements and small volume needs, the technology in the space marketplace has not been pushed forward. Commercial suppliers are not motivated to enter the radiation-hardened components market to push the state-of-the-art radiation-hardened capabilities forward because of low market demand and low profits.

Processor technology development in the general-purpose computing market (e.g., personal computers, mobile phones, etc.) has been primarily influenced by market forces and economy of scale. The rapid turnaround time in technology development and marketing strategies has pushed the commercial processor industry ahead of the military electronics industry. Incorporating current commercially available advanced microelectronics and commercial practices in the development of military satellite systems can radically improve system performance. Faced with a declining budget and increased performance demands, the DoD needs to consider leveraging commercial electronics for the development of their space systems.

The use of novel and state-of-the-art commercial technologies will provide greater performance than radiation-hardened technologies. Space-qualified electronics are traditionally manufactured through the use of specialized radiation-hardened techniques. The added complexity of these specialty processes combined with a low volume market demand has led to a performance gap between radiation-hardened components and commercially available components. Space-qualified processors have an approximate 7-year performance gap (2+ generations) compared with commercially available processors, as shown in Figure 2-3. This performance gap is expected to remain roughly the same in the future. The performance metric used is the computing speed measured in million instructions per second (MIPS).

The objective of increasing processing performance of future space-based assets requires an architect to seriously consider using COTS processors instead of radiation-hardened processors. General-purpose processors (GPPs) can offer high performance, large economies of scale, and a high degree of flexibility as the same hardware can be used for a multitude of applications. Additional advantages include the ease of integration and software compatibility for multiple generations of hardware. The key to implementing GPPs in space is to determine its survivability in the harsh radiation environment.

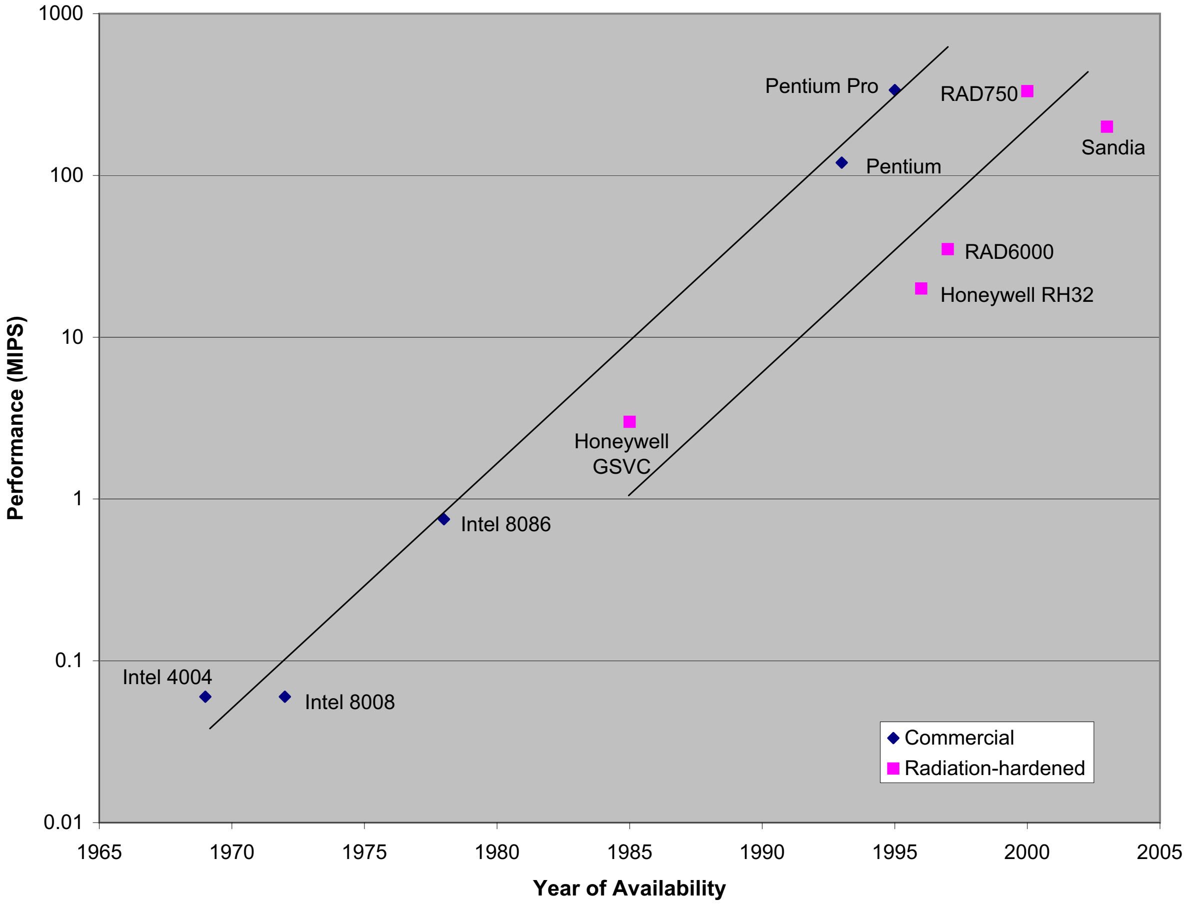


Figure 2-3: Comparison of radiation-hardened processors and commercially available processors.

Radiation testing and analysis are required to evaluate the performance of GPPs in the space radiation environment. The Intel Pentium III and AMD K7 microprocessors were tested for total ionizing dose effects in [45]. The computational power of the AMD and Intel processors used range from 550 MHz to 1 GHz. Based on minimal test data, AMD K7 processors has been shown to be TID hard to greater than 100 krad (Si) in the course of proton total dose testing. However, AMD K7 processors performed poorly to Co^{60} exposure. On the other hand, substantial data was collected for the Intel Pentium III processors undergoing proton and Co^{60} exposure. They were shown to be extremely tolerant of total dose radiation. Biased parts can survive in excess of 400 krads (Si) while unbiased parts can survive in excess of 1.6 Grads (Si). Although SEUs and functional interrupts were observed, these events can be controllable with mitigation strategies that allow the processors to function in the space environment. The most likely limiting factors to the use of GPPs in space applications are the thermal issues and power requirements. These issues must be taken into account in determining the size of the payload for the processing satellites.

Assuming that a processing satellite can be in any orbit, the annual dose rate is upper-bounded at 10^5 rads (Si), as shown in Figure 2-2. Given that the biased parts of an Intel Pentium III can survive at 400 krads (Si), the lifetime of the processor is determined to be approximately 4 years. With a very conservative safety margin of 2x, the processing satellite is considered to be designed with a lifetime of 2 years. With a 2- to 4-year lifetime, the COTS processors require little or no additional shielding if there is periodic replenishment (as frequent as every 2 years).

2.2.3 Design Recommendations

The main architectural concepts of the space-based information network with shared on-orbit processing are: (1) decoupled, shared, and distributed on-orbit processing and (2) leveraging commercial technologies. Optical intersatellite links provides the capability to connect multiple satellites that can communicate with each other at high data rates. Implementing commercially available processing technology on satellites can provide significant improvements in computational performance in space. The concept of decoupled, shared, and distributed on-orbit processing goes beyond satellite onboard processing on individual space missions. Processing capabilities can be decoupled from the mission satellites into separate processing satellites. Networking allows for the sharing of these processing resources.

2.2.3.1 Decoupled, Shared, and Distributed On-Orbit Processing

While it is possible to add processors to a mission satellite's payload, a new architectural paradigm that calls for decoupling the processing unit is suggested. A space-based backbone using optical communications enables high data rate transfers between mission satellites, backbone relay satellite nodes, and processing satellites. With this new design strategy, separate processing satellites can be built and deployed. The design of mission satellite payloads will not be affected. Mission satellites will continue to be designed to collect, digitize, and transmit data to the processing satellites. Processing satellites are responsible for activities such as processing raw data, data compression, coding, and encryption before information is transmitted to users on the ground and/or in space. The provisioning of separate processing satellites and a networked space-based backbone infrastructure enables multiple mission satellites to efficiently share processing resources. Processed information can be relayed to the ground via inexpensive RF downlinks. Data reduction in space overcomes the shortcomings of limited bandwidth on RF downlinks and reduces overall system cost as expensive high data rate RF downlinks are not necessary for mission satellites.

The high level data flow architecture for accessing a processing satellite from a mission satellite is illustrated in Figure 2-4. Data flow architectures provide some intuition as to the movement of data, starting with the collection of raw data and ending with the dissemination of processed information. Signals are collected by the mission satellite. Digitization may or may not occur on the mission satellite. If digitization occurs on the mission satellites, then the data is digitally transmitted to the nearest backbone relay satellite, otherwise analog transmission is used. The backbone relay satellite then transmits the data to a networked processing satellite. If the analog data requires digitization, then the analog-to-digital conversion occurs first before any processing computations. Additional processing functions include compression, coding, and encryption. Depending on user needs, processed information can be sent to ground stations or space users via the backbone relay satellite or by the mission satellite. In general, communications data may be relayed back to the mission satellites while sensor data may be disseminated to end-users through the backbone relay satellites.

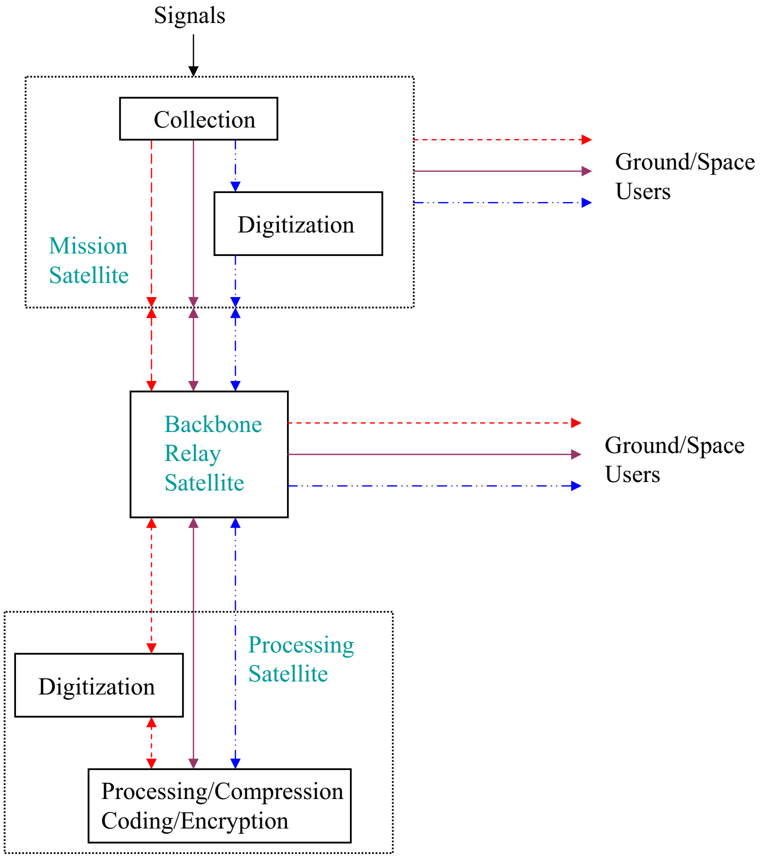


Figure 2-4: Data flow architecture for processing data.

2.2.3.2 Leveraging COTS Technology

The objective of increasing processing performance of future space-based assets requires an architect to seriously consider using COTS processors instead of radiation-hardened processors. As COTS processors are years more advanced, improved efficiency and performance can be attained with their implementation. Acquisition cost savings may be obtained from the purchase of components that are produced in large volumes in the commercial marketplace. The critical drawback to flying COTS in space is their survivability in the space environment. Thus, acquisition cost savings may be offset by the cost of component validation for survivability (upscreening) in space. Upscreening includes subjecting commercial electronic equipment to electrical, mechanical and environmental stresses beyond those tested or guaranteed by the device manufacturer. Additionally, implementation of an as-is COTS component furthers the paradoxical threat of faster technology obsolescence as commercial manufacturers continue to outpace one another with newer, faster, and better components [66].

Cost continues to be a fundamental limitation to building space systems. As government space efforts may no longer be sheltered by national security and political prestige justifications, the challenge is to change the traditional ways of doing business. The price of space-qualified hardware has traditionally been very expensive as a result of the engineering effort involved. Devices have to be designed to execute specific functions and to function reliably in harsh environments. If COTS technology could be adapted, benefits can be attained by both the government and the commercial sectors. Large commercial suppliers benefit from the advanced technology and economies of scale. By using a building block approach, system designers can purchase COTS components and adapt it to meet application requirements. Although cost control is the dominant factor, minimizing development time can be achieved, along with flexibility for future upgrades and interfacing with other subsystems. Architects for future space systems can benefit from the wide variety of available components that can often be used without paying for the design of new modules as long as the modules and associated software match industry standards [73].

2.3 Data Transmission

Computer processors are not the only electronic components that can be replaced or upgraded at a much faster time scale. Other electronic components and processes (techniques) can be decoupled from the mission satellite and placed on the processing satellite. For example, consider the placement of analog-to-digital converters (ADCs). ADC placement in the mission satellites or the processing satellites impacts whether analog or digital transmission is used, as illustrated in Figure 2-5. If ADCs are built into the mission satellite, raw data can be transmitted in digital form to the processing satellites. If the quantization is lossy, then the distortion is set by the rate distortion theory in accordance to the available optical access link capacity. However, if ADCs are implemented in the processing satellites, the raw data must be transmitted via analog techniques to be digitized and processed onboard the processing satellite. Provided that the analog link has been designed with enough signal-to-noise ratio (SNR), a newer, faster, and finer analog-to-digital quantizer can be inserted into the processing satellite to reduce the distortion of the compressed signal when the technology has improved. To evaluate the choice of which data transmission

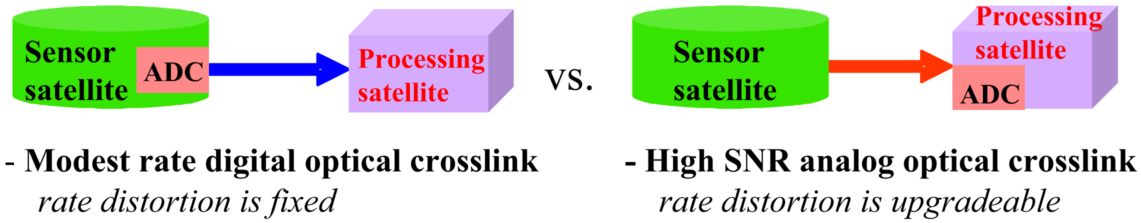


Figure 2-5: ADC placement options: mission satellite vs. processing satellite.

technique to implement, a systems designer must: (1) determine whether analog or digital signal processing techniques are more appropriate for the applications and (2) evaluate the suitability of implementing ADCs on the processing satellites which have a short life cycle.

2.3.1 Analog vs. Digital Transmission

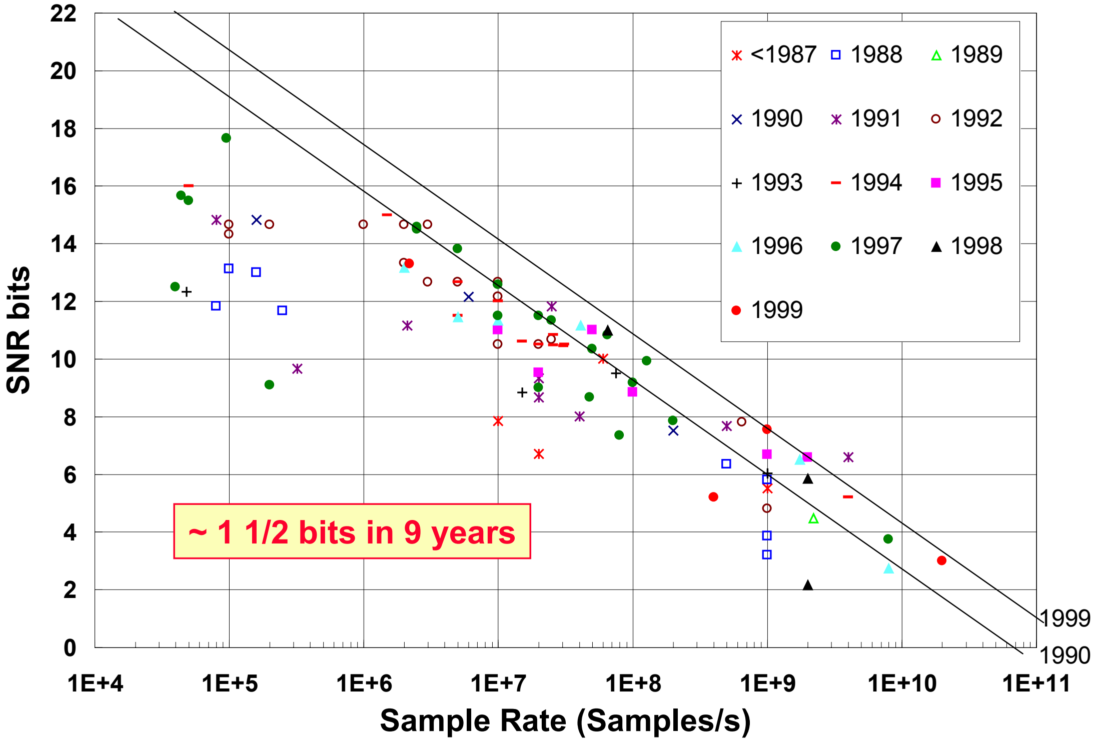
Theoretically, there is no difference between analog signal processing and digital signal processing methods. However, the number of applications using digital signal processing continues to grow as digital signal processing offers the following advantages [59]:

- There is no degradation of SNR in the stages following the ADC. However, arithmetic rounding error can cause degradation.
- Maintenance of drifts in gain, temperature and supply voltage stability are necessary for complex analog systems. These drifts can be removed by using synchronous digital systems.
    - Synchronous logic, such as a central clock for retiming after each stage, can remove differential timing and jitter problems which then allow for more complex parallel processing methods.
    - Digital multipliers can remove gain drift problems in analog systems.
- Fast, cheap, and extremely flexible digital memory is available for more real-time processing applications.
- A wide dynamic range (e.g., 12 bits or +66 dB) is possible with the use of wide data words.
    - Linear processing of signals can be done in almost any order with a wider dynamic range.

2.3.2 Analog-to-Digital Converters

A survey and analysis of experimental and commercially available ADCs has been conducted in [89]. The SNR improvement trend for ADCs is approximately 1.5 bits in 9 years, as shown in Figure 2-6. This rate of improvement is more than twice as long as the calculated lifetime of an Intel Pentium III processor in space. With ADCs improving at a much slower rate, it is not reasonable to implement them on the processing satellites if the processing satellites are replaced within 2-4 years. Consequently, for most applications, it is better to use ADCs on the mission satellites and to transmit information digitally to the processing satellites.

On the other hand, analog transmission between mission satellites and processing satellites cannot be totally ignored. Analog transmission can be considered for applications that have multiple senders of raw information (e.g., triangulation). Triangulation locates a single transmitter point by using multiple listening points to form a triangle having the unknown point and two known points at three vertices. In order to use the formula of triangulation, it is necessary for receive the three signals and then process the travel time information of the satellite emissions.



Source: [90] Robert H. Walden, Analog-to-Digital Converter Survey and Analysis, Presentation Slides, 16 July 1999, p. 22.

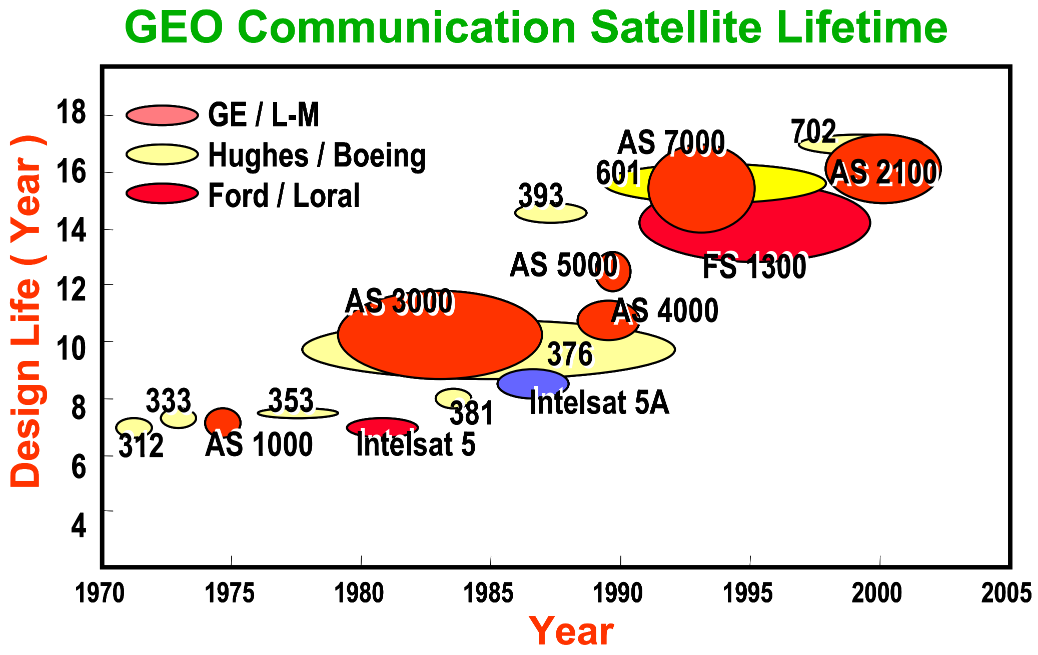
Figure 2-6: ADC performance improvement trend.

2.3.3 Design Recommendations

As digital signal processing is growing in use, digital transmission will be used for most future space-based applications. If ADC technology improvement remains much slower than improvement in processor technology, an architect should not want to implement ADCs on processing satellites which need to be replaced every 2-4 years. Yet, analog transmission can not be completely replaced by digital transmission. The performance of a sensor may be improved if the sensing function can be distributed over multiple satellites. Distributed satellite systems can significantly advance image oriented sensing and object identification from space. Geolocation applications, for example, can use two satellites to model the arm of a long baseline interferometer as long as the two sensed signals can be combined for coherent processing. This either requires phase information to be preserved through fine quantization and significant data rate transmissions or transmission transparency (coherent analog transmission at high fidelity). An analog link in an optical satellite network can provide such a service.

2.4 Satellite Replacement and Replenishment Strategies

At present, satellite systems are designed for long life and high reliability with much redundancy and no on-orbit maintenance or servicing. Examples of GEO communication satellites and their manufacturers that have been developed over the last three decades with current design lifetime averaging 15 years are shown in Figure 2-7. Space system designs are often driven by factors that include total system cost, accessibility to space and launch cost. Replacing or upgrading these unique and custom-tailed satellite systems require additional launches. Options for replacing satellites include: spare satellites (on-orbit or stored on the ground) and on-orbit servicing.



Source: [77]. James Shoemaker, Orbital Express: A Better Way of Doing Business in Space, Presentation Slides, 2001.

Figure 2-7: Design lifetime of GEO communication satellites.

2.4.1 On-Orbit Spare Satellites

On-orbit space satellites provide nearly instantaneous backup. Once a primary satellite in space is rendered non-operational, the on-orbit satellite can quickly take its place. On-orbit satellites are designed and manufactured identically to the primary satellites.

2.4.2 Full Satellite Replacement

Full satellite replacements refer to deploying spare satellites that have been stored in a ground facility. These spare satellites are typically designed and manufactured at the same time as the primary satellites.

2.4.2.1 Original Design

The advantage of deploying spare satellites that are of the same original design provides a high degree of redundancy, as the functionality is equivalent to the primary satellites. Cost of production and acquisition may be lower due to the quantity that is produced and bought.

2.4.2.2 Updated Design

An additional gain from deploying spare satellites from the ground is the opportunity to improve and upgrade the design of these spare satellites. While there may be an increase in cost in the effort to improve performance and efficiency, it allows for greater evolution of the space system as newer payloads can be implemented before a launch.

2.4.3 On-Orbit Servicing

On-orbit servicing is a relatively new and feasible technology in the early stage of evolution. Service work in space can be performed by men, machines, or a combination of both. The goal of on-orbit servicing is to enhance the operational life and capability of satellites, space platforms, and space vehicles. On-orbit servicing includes replenishment of consumables (e.g., fuel), inspection, realignment, recalibration, repair, replacement of modules/payloads. To date, maintenance of orbiting satellites required human support in space. Human interaction allows for greater flexibility as astronauts are skillful, versatile, and innovative. However, human space flight is extremely expensive and risky [91].

Past spacecraft programs typically had a non-serviceable design. Spacecrafts today still have unique system designs with specific hardware components and are vertically integrated. These systems are designed for high reliability, long life, redundancy, no on-orbit maintenance or servicing, and, if required, spacecraft replacement via another launch [91]. Autonomous on-orbit servicing is a technology that may mature in the near future. Space system designs may want to keep this in mind when designing future space systems. To leverage on-orbit servicing, future spacecraft designs must be modular. Modular designs allow a complex system to be built from independently developed components that can be plugged together.

2.4.3.1 Potential Benefits of On-Orbit Servicing

On-orbit servicing of space systems can provide many benefits for numerous satellite systems and programs. A brief summary of potential benefits for on-orbit servicing of the processing satellites [91]:

1. *Extended satellite lifetime.* The useful lifetime of the processing satellite can be extended if it was built with a modular design. On-orbit replenishment of the processing

payload will upgrade computational capabilities and increase the value of shared on-orbit processing resources.

2. *Greater mission flexibility/availability.* Changeout/repair/upgrading of payloads (e.g., optics, transponders, detectors), subsystems (e.g., power supply, communications, data handling, propulsion), and components (e.g., solar arrays, booms, antennas, sensors) can be made available through modular spacecraft designs and plans for on-orbit servicing. Mission objectives, the scope of collected data, and the quality of processed information can be adjusted or increased with the changing of spacecraft modules.

3. *Enhanced performance and reliability of critical components.* Satellite performance can be maintained at peak levels through recurrent servicing (e.g., replacing with more advanced processing technology). The reliability of other critical spacecraft components can be improved with regular on-orbit testing and servicing (e.g., calibrating payloads, checking optical, solar array, and sensor surfaces, deployment of booms, antennas, and solar arrays).

4. *Improved military mission assurance.* Mission assurance of military satellites can be supported via modular replacements. Important decisions must be made about the servicing intervals because the downtime of military satellites can affect critical mission strategic operations and availability.

5. *Reduce life cycle costs of large, long-term programs.* On-orbit servicing can reduce the amount of redundant designs required in spacecraft designs. There exists a trade-off between the servicing cost and the value added to the satellite after on-orbit servicing. Cost for servicing a satellite may be lower than full satellite replacement. Additional cost savings can be attained by requiring less space-based and ground-based replacement spare satellites.

2.4.4 Design Recommendations

In order to determine a satellite replacement and replenishment strategy that relies on either satellite replacement or on-orbit servicing for various space systems, an architect must consider the issues of the following four input parameters [91]:

1. *Satellite:* On-orbit servicing of individual satellites is influenced by the following factors: high complexity, high cost, low cost replacement units, low mean mission time, and the requirement of high reliability at a reasonable cost.

2. *Launch vehicles:* On-orbit servicing is practical if space transportation cost is low.

3. *Infrastructure:* On-orbit servicing can be leveraged if satellites are built in a modular fashion that allow for easy changeout or repair. The infrastructure for an orbital servicer is also required. Lastly, interoperability between individual satellites and the orbital servicer is necessary.

4. *Operations:* Economies of scale can be exploited for on-orbit servicing if the ability to service multiple satellite per mission exists.

Past space programs generally maintained on-orbit spare satellites or stored spare satellites on the ground. On-orbit servicing is a technology that should be considered for future space missions. Leveraging on-orbit servicing impacts the design of future missions (e.g., using modular designs). However, implementing the concept of on-orbit servicing is a chicken-and-egg type of challenge. Satellite designers are hesitant to build on-orbit serviceability into their satellites if no autonomous servicer is in existence. Entrepreneurs will find it difficult to build an autonomous servicing module if no serviceable satellites are available.

2.5 Summary

It is reasonable to believe that as more space packages are developed and extensive on-orbit operational experience increases in the next few years, the cost of high rate optical crosslinks will be substantially lower than their microwave functional equivalent. A natural next step with such a powerful enabling technology is the realization of an optical satellite network of global extent. Optical satellite communications is a transforming technology whose architectural potential has not been fully exploited. Not only will a satellite network become economically viable, but its deployment and the extraordinary services that it can offer allow for radically transforming space system architectures. A high speed optical satellite information network backbone can be used to connect different space systems and users in space rather than on the ground. This is key for transforming the stove-piped satellite communications community into a data satellite networking community serving a multitude of users.

An examination of several concepts and building blocks necessary to designing a space-based information network architecture with shared-on orbit processing has been provided. A summary of additional new functionalities that space laser communications and on-orbit processing capabilities can provide is provided in Table 2.1. The design of the high rate space-based information network backbone and shared on-orbit processing is explored in Chapter 3. The design of shared space-borne processing resources for real-time digital signal processing applications (e.g., SIGINT and SAR) are explored in Chapter 4. Space-borne processing allows for data processing and compression before transmission on the downlinks. The remaining functionalities are discussed briefly in Chapter 4 and left for future research.

New Function	Technology Enabler	Architecture Concept
1 Space information network	Laser communications	High rate backbone in space
2 Shared on-orbit processing	1 + commercial processors + cluster replenishment	Shared and rapidly upgradeable processing
3 Increased resolution, multisensor data fusion	1 + 2 + image subtraction/processing	Dissemination of processed data on limited downlink
4 Coherent distributed space sensing	1 + 2 + 3	Long baseline interferometer, fractionated optics
5 Interoperable space communications	1 + 2 + software / interconnections	Communications packages + re-programmable processing
6 Multiplatform distributed satellite communications	1 + 2	Multi-element antenna array distributed over many satellites
7 Restoration of disconnected global networks	1 + 2 + 5	Backbone reconnection
8 Rapid mission deployment	1 + 2	Reconfigurable computing in processing satellites
9 Future on-orbit upgrades	1 + 2	Shared upgradeable equipment and on-orbit servicing

Table 2.1: Summary of space architecture concepts.

Chapter 3

Network Architectures

Networking is an efficient method of sharing communications among multiple users. A space-based information network backbone is the essential building block for providing interconnections between multiple users and access to space-based processing resources. The space-based information network being considered will serve space, airborne, and terrestrial users. Space-based network users may be sensors in space, the Space Shuttle, or communication satellites themselves. The design of this network therefore may have a very different architecture than one that has been optimized to serve only terrestrial and airborne users. Major requirements that are likely to dictate differences in architecture include: (1) different data types (e.g., data streams and packets) (2) different QoS requirements (e.g., throughput and latency) and (3) location of users (i.e., users located in Low Earth Orbit [LEO], Medium Earth Orbit [MEO], GEO, and on the ground). Additionally, cost is a fundamental limitation that warrants consideration in the design of space systems.

This chapter explores the architectural design of the space-based information network backbone that acts as the transport network for mission satellites as well as enables the concept of decoupled, shared, and perhaps distributed, space-borne processing resources for space-based assets. Network architectures consist of designs for the physical network in addition to a logical topology that specifies how data move across that physical network. To provide high speed space-to-space communications between space-based assets and networked processing resources, the intersatellite backbone is built using optical communications as the enabling technology. An example of a ring backbone constellation with five satellite relay nodes with connected users and processing resources is illustrated in Figure 3-1. Laser communication systems operating at optical frequencies allow for the use of small antenna systems due to the narrow beamwidths; and thus allow for the use of low power transmitters. Moreover, a single optical communication system can transmit up to several Tbps of information using WDM technology [20]. The design goal, here, is to create a space-based information network backbone to meet mission requirements (e.g., high data rates, high connectivity, and low latency) at a reasonable cost by optimizing constellation topologies under different traffic scenarios. In this chapter, backbone constellation topologies, traffic models, and communications costs for the space-based network backbone are defined. System costs for communications and recommendations of the good architectures for various scenarios are provided. Lastly, the network connection of space-based processing satellite units to the backbone constellation are analyzed and discussed.

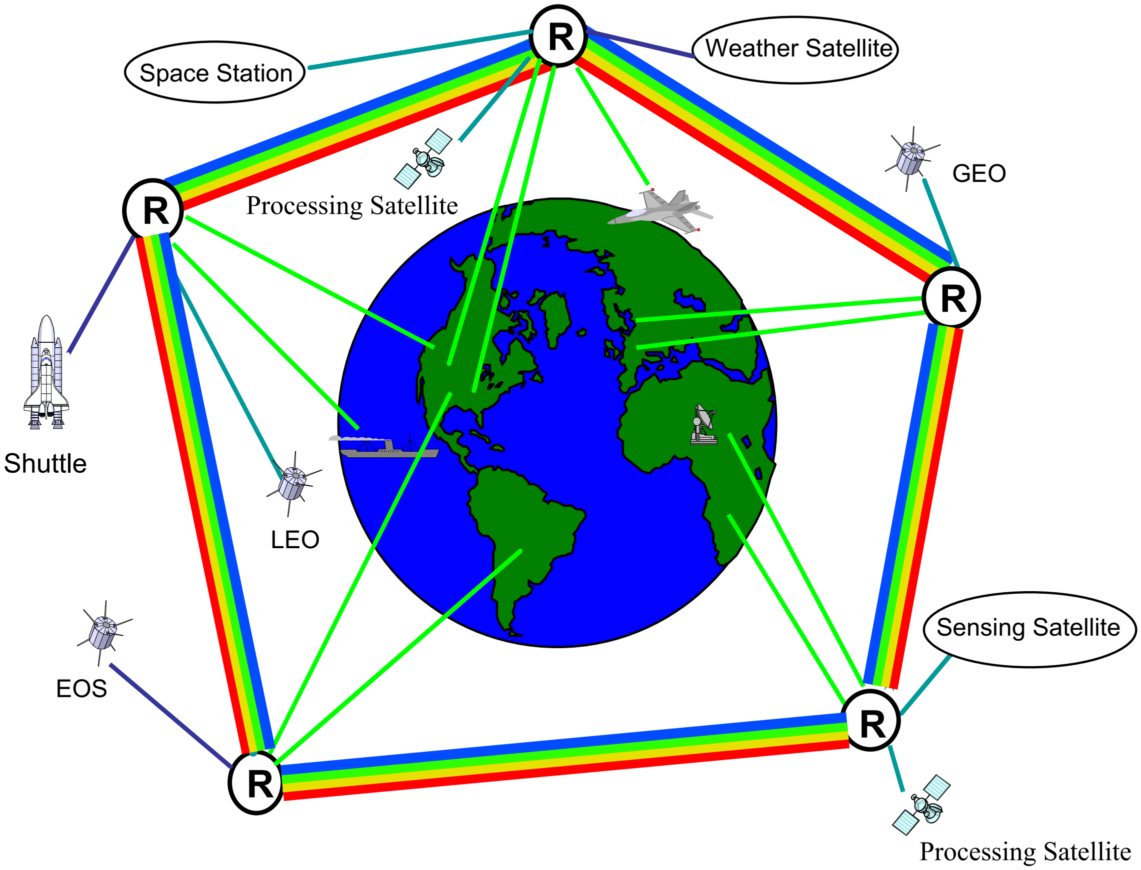


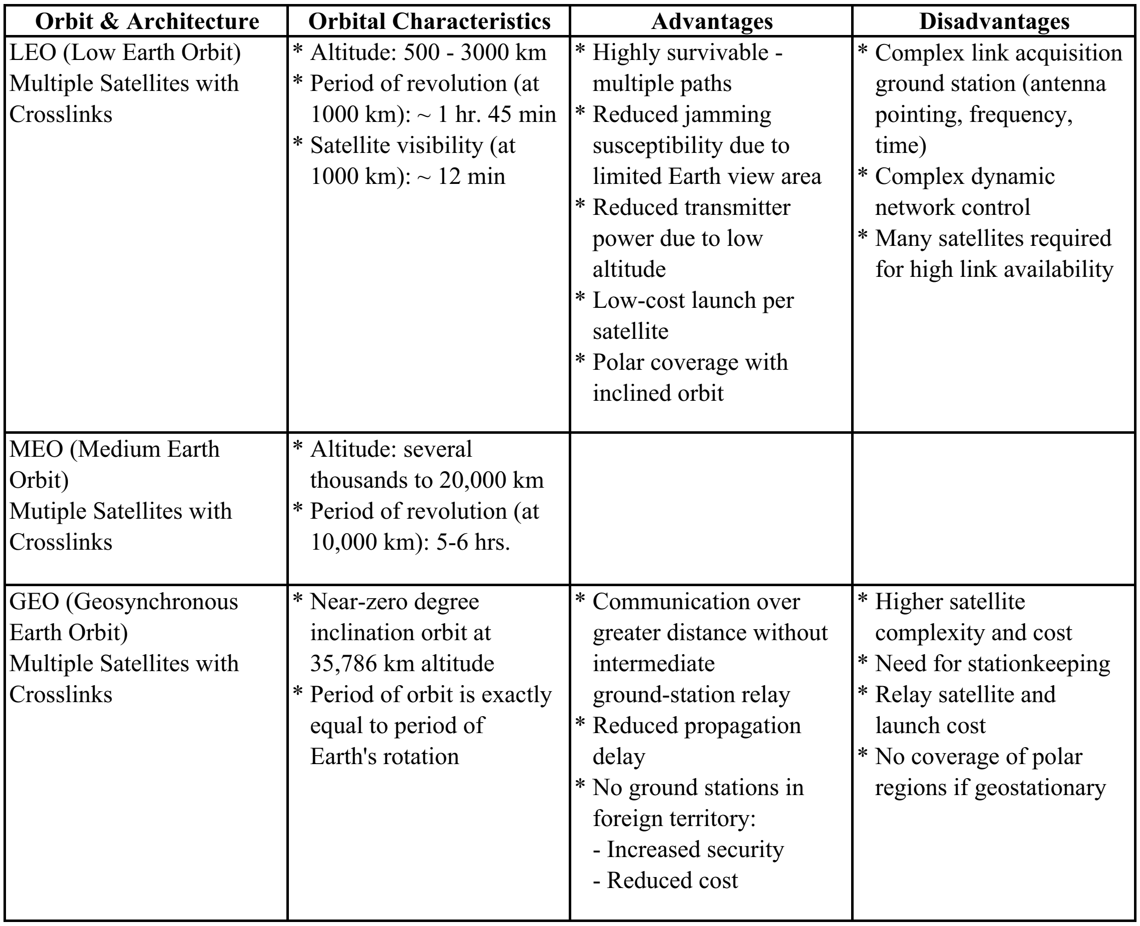
Figure 3-1: Example of a space-based information network architecture.

3.1 Space-Based Network Backbone Constellation

In satellite constellation design, the usual trade is coverage as a measure of performance versus the number of satellites as a measure of cost [92]. A constellation of multiple satellites can support large coverage areas of the Earth with high reliability and survivability. Because a system's cost and performance are strongly affected by the constellation's size and structure, assessment of issues such as orbit characteristics, network topology, and routing schemes are required. A rational way to begin designing a constellation is to determine the satellite orbits. The choice of altitude and inclination of satellite orbits typically involves trades between parameters such as space mission lifetime, cost, space environment, viewing geometry, and payload performance.

3.1.1 Orbit Selection

Table 3.1 summarizes some orbital characteristics, advantages, and disadvantages for multiple crosslinked satellite constellations in LEO, MEO, and GEO [30, 40, 92]. A study of constellation orbits based on coverage requirements, intra-backbone link complexity, user-access link capacity, maximum slewing rate and link distance, aperture quantity, and on-board placement has favored GEO over both LEO and MEO [11]. With the selection of GEO, applying the design constraint of using all satellites in circular orbits at a common altitude and inclination allows for the period, angular velocity, and node rotation rate to be equivalent for all the satellites. Satellites in circular orbits and the same plane are always at a constant range and provide uniform coverage with no need for any slewing movement in the intersatellite links.



Orbit & Architecture	Orbital Characteristics	Advantages	Disadvantages
LEO (Low Earth Orbit) Multiple Satellites with Crosslinks	* Altitude: 500 - 3000 km * Period of revolution (at 1000 km): ~ 1 hr. 45 min * Satellite visibility (at 1000 km): ~ 12 min	* Highly survivable - multiple paths * Reduced jamming susceptibility due to limited Earth view area * Reduced transmitter power due to low altitude * Low-cost launch per satellite * Polar coverage with inclined orbit	* Complex link acquisition ground station (antenna pointing, frequency, time) * Complex dynamic network control * Many satellites required for high link availability
MEO (Medium Earth Orbit) Mutiple Satellites with Crosslinks	* Altitude: several thousands to 20,000 km * Period of revolution (at 10,000 km): 5-6 hrs.		
GEO (Geosynchronous Earth Orbit) Multiple Satellites with Crosslinks	* Near-zero degree inclination orbit at 35,786 km altitude * Period of orbit is exactly equal to period of Earth's rotation	* Communication over greater distance without intermediate ground-station relay * Reduced propagation delay * No ground stations in foreign territory: - Increased security - Reduced cost	* Higher satellite complexity and cost * Need for stationkeeping * Relay satellite and launch cost * No coverage of polar regions if geostationary

Table 3.1: Characteristics of multiple crosslinked satellite constellations in LEO, MEO, and GEO.

3.1.1.1 Geostationary Orbit

The satellites for the space-based information network backbone constellation are placed in circular geostationary orbits (GSOs) at the GEO altitude of 35,786 km. Because the footprint of a geostationary satellite covers nearly 42% of the Earth's surface, a minimum of three geostationary satellites in orbit can provide near-global coverage (between 70° South and 70° North latitude). In reality, geostationary satellites do not meet the theoretical orbital conditions of zero values for inclination and eccentricity due to disturbances by the sun, the moon, and the non-symmetric Earth. These disturbances induce the satellites to traverse a small figure of eight in the sky and can be corrected by on-board propulsion. The orbits are regarded as geostationary if the satellites sustain their inclination within less than 1 degree and their assigned longitude within a few tenths of a degree. To date, the principal disadvantages of geostationary orbits are the high launch cost, the propagation delay of approximately 0.25 seconds, the high fuel consumption to support a near-zero inclination angle to keep the orbital plane close to the equatorial plane, and the lack of coverage in the polar regions (above 70° latitude). However, satellites in a geostationary orbit appear fixed above the surface of the Earth (i.e., at a fixed longitude and latitude). This simplifies the design and operating requirements for both the satellites and the ground stations because the uplinks and downlinks are fixed (i.e., the satellite is always in view of the ground station and there is no need to track the satellite to determine where to point the antenna).

3.1.2 Constellation Topologies

In communication networks, a topology is a schematic representation of the network geometry. Network geometry can be distinguished in two ways: the physical topology and the logical (or signal) topology. The *physical topology* of a network is the actual geometric layout of the nodes and connections. The *logical topology* of a network is defined by the network protocols that direct how data are transmitted across the network without regard to the physical interconnection of the nodes. The common types of network topology, illustrated in Figure 3-2, are bus, ring, star, tree, and mesh. In a *bus topology*, all nodes are connected together by a single line. In a *ring topology*, all nodes are connected in a closed loop configuration where adjacent pairs of nodes are directly connected. In a *star topology*, all peripheral nodes are directly connected to a central node. A *tree topology* resembles the interconnection of star networks. A mesh topology can be of two arrangements: a partial mesh and a full mesh. In a *partial mesh topology*, there are at least two nodes with two or more connections between them. In a *full mesh topology*, there is a direct connection between any two nodes.

Graph theory notations and definitions are provided to assist with the terminology used in this chapter. The physical topologies considered for the space-based information network backbone constellation include connected circulant constellations (which include ring and mesh topologies) and hub constellations (i.e., star and tree topologies). Each type of constellation is subsequently defined and discussed.

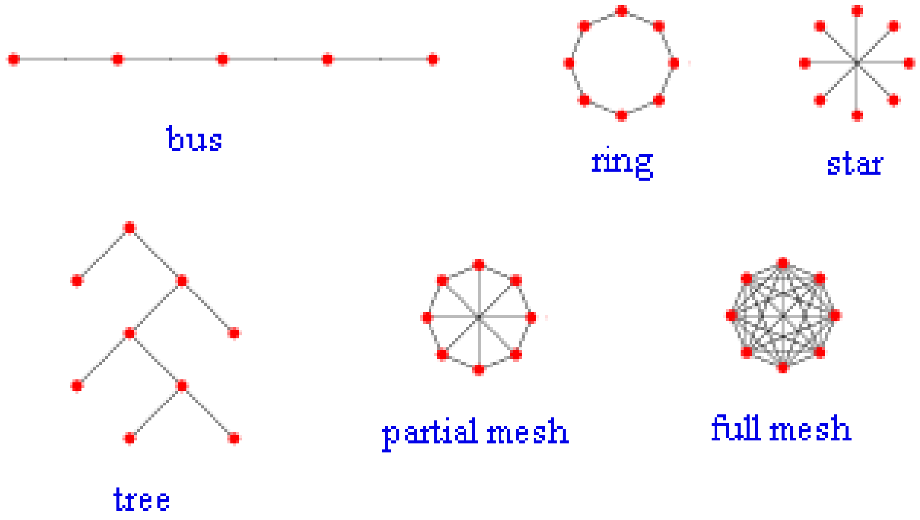


Figure 3-2: Common types of network topology.

3.1.2.1 Graph Theory Notations and Definitions

Graph theory is a branch of mathematics that can be applied to the design of a network topology for telecommunications architectures. The basic terminology of graph theory is presented here in order to understand the application of graph theory in determining and assessing the physical topology of the space-based network backbone.

A *graph* $G = (V, E)$ consists of a finite nonempty set $V = V(G)$ of p *vertices*, also known as nodes, and a set $E = E(G)$ of q unordered pairs of distinct nodes of V. The graph G has *order* p and *size* q. An *edge* of G is the pair $e = (u, v)$ of nodes in E and nodes u and v are then called *adjacent nodes*. The number of edges adjoining a node is defined as the *degree* of node v and is denoted *deg v*. Among the nodes of G, the *minimum degree* is denoted by $\delta(G)$ while the *maximum degree* is denoted by $\Delta(G)$. If the degree of all nodes equal r, then the graph is *regular of degree* r or r-regular. The *distance d(u,v)* between two nodes u and v in G is defined as the minimum length of a path joining them. If no path exists between nodes u and v, then $d(u, v) = \infty$. A graph is defined as *connected* if there is a path between every pair of nodes. A graph is defined as *acyclic* if it contains no directed cycle. The *diameter* $d(G)$ of a connected graph G is the length $\max_{u,v} d(u, v)$ of the longest shortest path (i.e., the longest graph geodesic).

3.1.2.2 Connected Circulants Constellations

A *circulant* graph is a graph on p nodes $(v_1, v_2, \ldots, v_p)$ with node v_i adjacent to each vertex $v_{i \pm n_j (\texttt{mod}\ p)}$ and denoted as $C_p(n_1, n_2, \ldots, n_k)$ where the values of n_i are called *jump sizes* [17]. The jump sizes $n_1, n_2, \ldots, n_k$ are a sequence of integers where $0 < n_1 < n_2 < \cdots < n_k < \frac{p+1}{2}$. Examples of circulant graphs are illustrated in Figure 3-3. This class of graphs has the characteristics of strong connectivity which is important for designing reliable networks and symmetry which is useful for analysis. Because circulant graphs include empty graphs and non-connected graphs, the tradespace of candidate topologies is reduced to connected circulant graphs.

Connected circulant graphs of order N and r-regular, however, are not unique, as shown in Figure 3-4(a). Thus, the tradespace of candidate topologies must be reduced again. A subset of connected circulant graphs selected for consideration is shown in Figure 3-4(b). These unique graphs share the following property: the jump sizes are an arithmetic sequence whose first term is 1. This property provides a cost-effective advantage over alternative connected circulant topologies of the same (N, r) due to the smaller graph diameter which leads to a smaller average minimum hop distance in minimum hop distance traffic routing.

Given the degree r, the number of nodes, N, in this class of connected circulant graphs can be determined by the following equation:

$$N = 2 + s(r - 1) \tag{3.1}$$

where s is the jump space. The *jump space* is a constant that denotes the difference between successive terms in the sequence of jump sizes:

$$s = |n_i - n_{i-1}| \ \forall\ i \in \mathbb{N}^+ \tag{3.2}$$

A tradespace of these special connected circulant graphs is enumerated in Table 3.2. For the space-based information network backbone, the constraints $3 \leq N \leq 20$ and $2 \leq r \leq N - 1$ for Equation 3.1 determine the finite number of candidate architectures to be examined in this chapter, as highlighted in Table 3.2. Recall that a minimum of three geostationary satellites in orbit is required to provide near-global coverage.

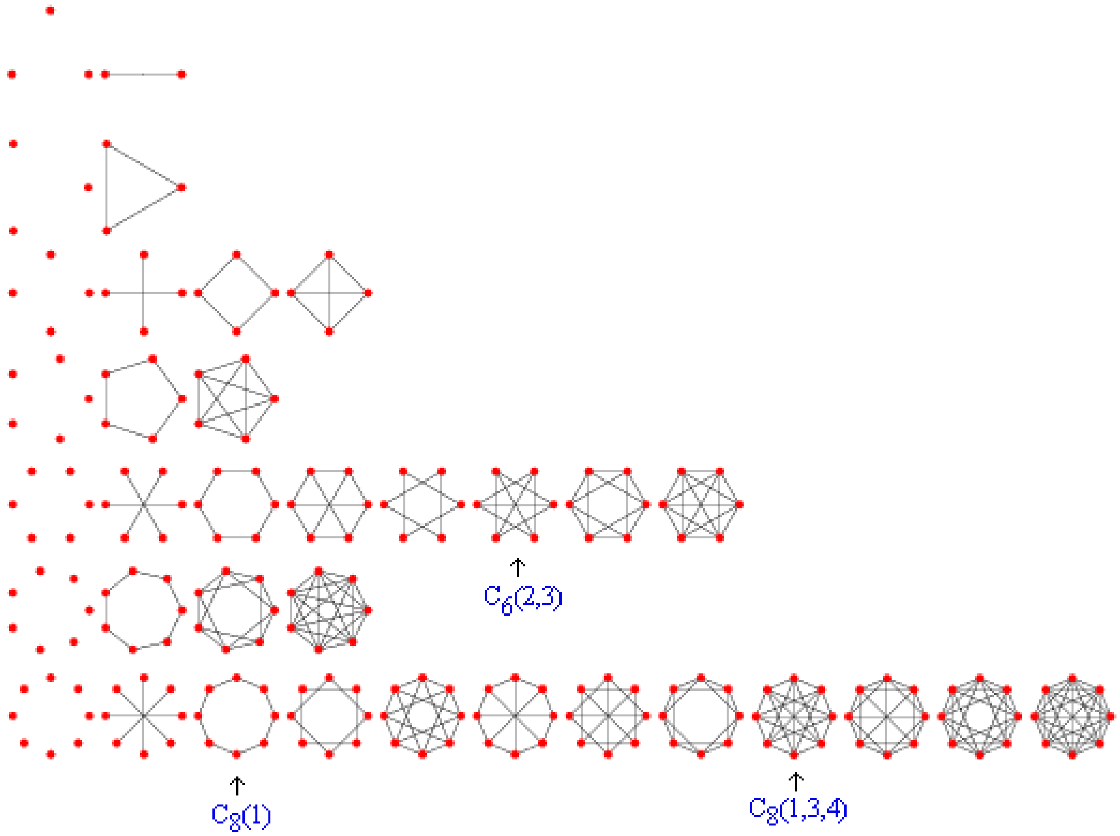
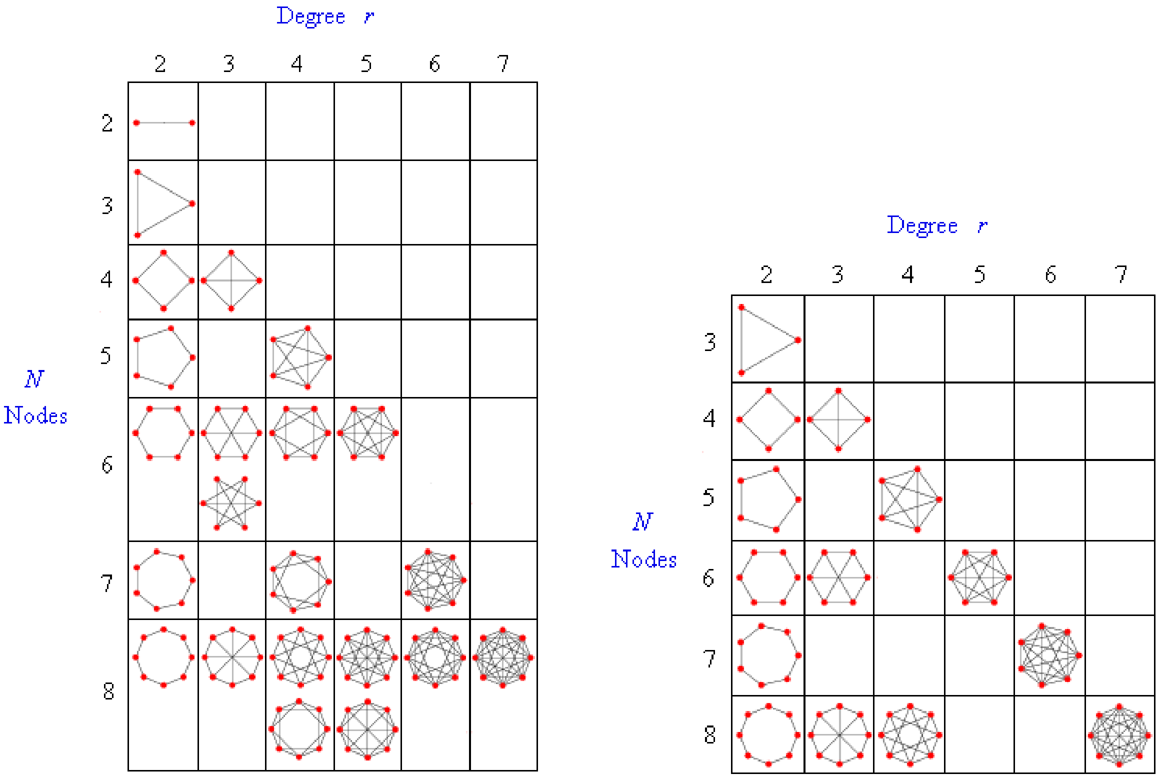


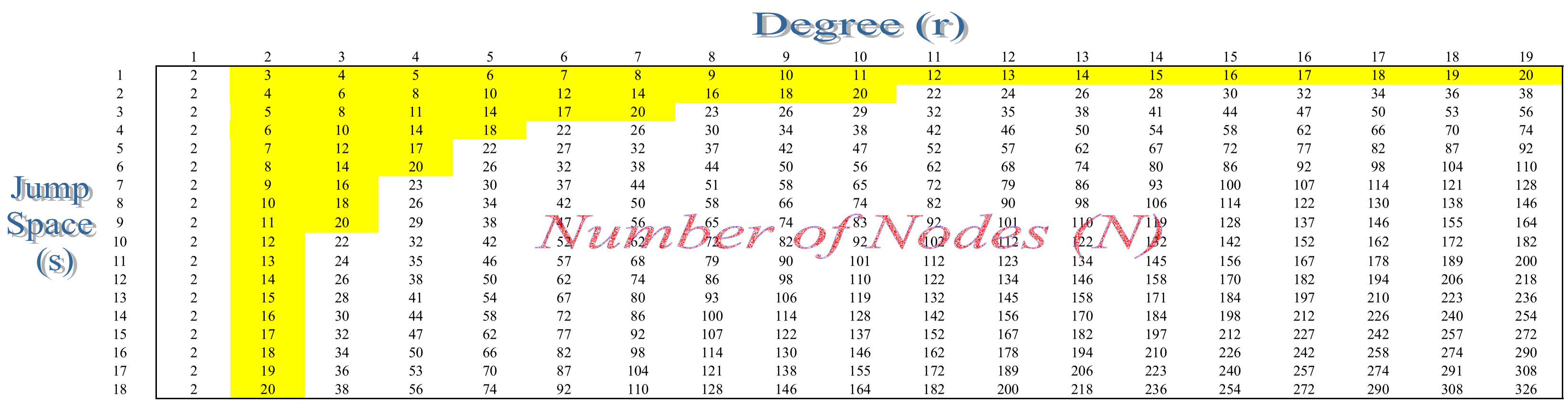
Figure 3-3: Example circulant graphs.



(a) Example connected circulant graphs.

(b) Example connected circulant graphs with uniform jump space property.

Figure 3-4: Tradespace reduction of connected circulant graphs.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38
3	2	5	8	11	14	17	20	23	26	29	32	35	38	41	44	47	50	53	56
4	2	6	10	14	18	22	26	30	34	38	42	46	50	54	58	62	66	70	74
5	2	7	12	17	22	27	32	37	42	47	52	57	62	67	72	77	82	87	92
6	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	104	110
7	2	9	16	23	30	37	44	51	58	65	72	79	86	93	100	107	114	121	128
8	2	10	18	26	34	42	50	58	66	74	82	90	98	106	114	122	130	138	146
9	2	11	20	29	38	47	56	65	74	83	92	101	110	119	128	137	146	155	164
10	2	12	22	32	42	52	62	72	82	92	102	112	122	132	142	152	162	172	182
11	2	13	24	35	46	57	68	79	90	101	112	123	134	145	156	167	178	189	200
12	2	14	26	38	50	62	74	86	98	110	122	134	146	158	170	182	194	206	218
13	2	15	28	41	54	67	80	93	106	119	132	145	158	171	184	197	210	223	236
14	2	16	30	44	58	72	86	100	114	128	142	156	170	184	198	212	226	240	254
15	2	17	32	47	62	77	92	107	122	137	152	167	182	197	212	227	242	257	272
16	2	18	34	50	66	82	98	114	130	146	162	178	194	210	226	242	258	274	290
17	2	19	36	53	70	87	104	121	138	155	172	189	206	223	240	257	274	291	308
18	2	20	38	56	74	92	110	128	146	164	182	200	218	236	254	272	290	308	326

Table 3.2: Tradespace of connected circulant constellations with uniform jump space.

3.1.2.3 Hub Constellations

Non-uniform constellations are also considered for the space-based information network backbone constellation. On these asymmetric topologies, communications between nodes are transmitted via central nodes (or hub nodes). Constellations with one hub node and with two hub nodes are taken into account.

One hub node. A 1-hub topology is essentially a star network. For a space-based network backbone of order N, one of the satellites is designated as the *central* node (or *hub* node) while the remaining nodes are known as *plain* nodes. The finite number of candidate architectures examined is again determined by the constraint $3 \leq N \leq 20$. The degree of the hub node is $r = N - 1$ while the degree of all plain nodes is $r = 1$. Example 1-hub topologies are illustrated in Figure 3-5. This type of network is susceptible to bottleneck and failure problems at the hub node.

Two hub nodes. Tree topologies can provide greater reliability than star topologies. The type of tree topology considered in this study has 2 central nodes (or hub nodes) and is 1 level deep. Example 2-hub topologies are illustrated in Figure 3-6. For a space-based network backbone of order N, two adjacent satellites are designated as hub nodes. The finite number of candidate architectures examined is determined by the constraint $3 \leq N \leq 20$ where N is even. An even number of satellites provides symmetry which allows for simpler analysis. Each hub node is connected to half of the plain nodes in the satellite network. The degree of the hub nodes is $r = \frac{N}{2}$ while the degree of all plain nodes is $r = 1$.

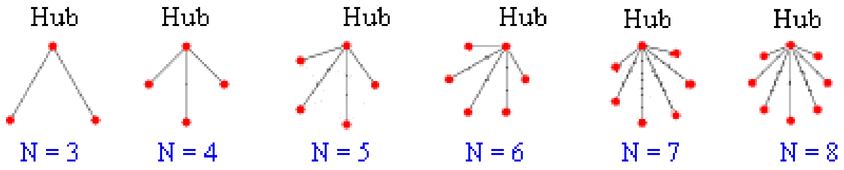


Figure 3-5: Example 1-hub topologies.

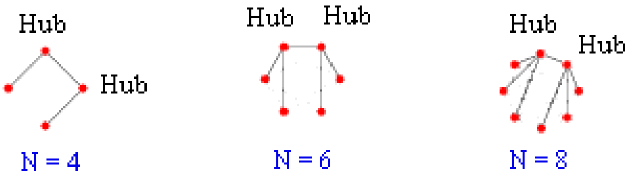


Figure 3-6: Example 2-hub topologies.

3.1.3 Traffic Models and Routing

A network cannot be designed to meet arbitrary demands with limited resources. Assumptions of network properties for the traffic demands must be made. Accordingly, a traffic model is employed to study the network, which may reflect an offline or online mode of operation, with or without blocking. The common models of a fixed traffic matrix and maximum load [70] are used in this study.

Fixed traffic matrix. In order to determine the number of wavelengths on the intersatellite links, the traffic pattern must first be defined. The network traffic is structured as a traffic matrix $T = (i, j)$, where $t(i, j)$ represents the number of wavelengths between nodes i and j. In the fixed traffic case, the entire traffic demand is determined a priori. The amount of traffic is calculated to be some nominal amount plus a marginal amount. It is a simplistic model that allows for designing the network to meet all traffic demands.

Maximum load. A parameter that characterizes the traffic is the load, which is defined as the number of concurrent lightpaths that can exist on any link in the network. Maximum load is a key parameter because it indicates the maximum number of wavelengths that is required on the link to support the traffic.

In this architectural study, the satellite network is being provisioned before deployment. Thus, the design methodology is to calculate the number of wavelengths required on the intersatellite links and determine the size of switches in all satellite nodes needed to support the predetermined amount of traffic T. The objective is to minimize the total communications cost of the network backbone that can support the given traffic load.

3.1.3.1 Uniform All-to-All Traffic

Uniform all-to-all traffic is the type of traffic where every node sends to every other node in the network, including itself. A node can be both a transmitter and a receiver because both types of users can be located within the same satellite footprint. This type of traffic models the situation where users served by each satellite are equally likely to generate and receive traffic. Each satellite node has traffic entering and exiting the global network. The number of satellite backbone nodes in the network is denoted by N and the total amount of traffic in the network is denoted by T. Therefore, in the uniform all-to-all traffic scenario, the traffic between every node pair is $t = \frac{T}{N^2}$ units of traffic. For normalization purposes, the capacity of a wavelength is assumed to be 1 unit and T is fixed to be 400 wavelength units (max $T = 20$ nodes × 20 node pairs of 1 wavelength) for all analysis cases.

3.1.3.2 Uniform All-to-One Traffic

Uniform all-to-one traffic is the type of traffic where every node sends to or receives from a central node in the network. In this traffic scenario, the central node is either a sink or a source for all traffic generated in the network. This type of traffic models the situation where sensing satellites collect information to send to one designated node for processing or where one satellite is injecting large volumes of traffic into the network. In order to conduct a fair comparison between the two traffic models, the total amount of traffic, T, needs to remains fixed at 400 wavelength units. In the uniform all-to-one traffic scenario studied, traffic from every node is destined to the central node and is uniform (i.e., $t = \frac{T}{N}$). For consistency, the central node can be both a transmitter and a receiver (i.e., it can send traffic to itself).

3.1.3.3 Traffic Routing

The main function of the network layer is to decide which physical path the information should follow from its source to its destination. The *routing algorithm* is the software responsible for the decision. *Shortest path routing* algorithms minimize distance or path cost between nodes in the network. The shortest path metric not only refers to hops and physical distance but also labels on the edges of a graph can be computed as a function of bandwidth, average traffic, communication cost, average queue length, or estimated delay [83]. By modifying the weighting function, a shortest path routing algorithm would calculate the shortest path measured according to a set of criteria. There are several algorithms for computing the shortest path between two nodes of a graph: breadth-first search algorithm, Dijkstra's algorithm, and the Bellman-Ford algorithm [22]. In this architectural study, Dijkstra's algorithm is chosen for computing the shortest path between source-destination pairs in the network according to the criteria of minimizing the number of hops. Nodes in the acyclic network $G = (V,E)$ are first numbered in topological order such that $u < v \; \forall \; (u, v) \in E$. Dijkstra's algorithm then finds the shortest paths from the source node to all destination nodes in the network. The variant of Dijkstra's algorithm that is implemented in this chapter's analysis can be found in [1].

3.1.4 Communications Cost Model

The communications architecture of the space-based information network backbone consists of a network of satellites connected by intersatellite links. The communications cost for the satellite backbone is based on three major components: antenna costs, switch costs, and link costs. For an individual satellite, the cost equation is written as:

$$J_{\text{satellite}} = \sum_{i=1}^{r} C_{\text{antenna}_i} + C_{\text{switch}} + \sum_{i=1}^{r} C_{\text{link}_i} \tag{3.3}$$

where r is the degree of the satellite node in the network. The overall cost equation for the network is denoted as:

$$J_{\text{network}} = \sum_{i=1}^{N} J_{\text{satellite}_i} \tag{3.4}$$

where N is the number of satellite nodes in the network. If all satellites in the network are identically designed, then the total cost can be written as:

$$J_{\text{network}} = N J_{\text{satellite}} \tag{3.5}$$

3.1.4.1 Antenna Cost

The cost of an antenna (i.e., telescope cost) is mainly driven by the size of the *aperture diameter*, D, and can be written as:

$$C_{\text{antenna}} = k_0 + k_1 D^{\alpha}, \quad 2 < \alpha < 3 \tag{3.6}$$

where k_0 is the fixed cost, k_1 is the variable cost, and α depends on the antenna technology. To size the antenna, relationships among data rate, propagation path length, and transmitter power need to be defined.

In an optical space system, an optical field generated at a laser source at the transmitter from afar is directed to a receiver, as illustrated in Figure 3-7. The *propagation path length* or *intersatellite distance* is usually significant such that the laser source appears as a point source when observed from the receiver [37]. The *transmitted power*, P_t, will propagate in free space at an angle $\sim \frac{\lambda}{D}$, where λ is the operating laser wavelength. The received power, P_r, is P_t times the effective receiver antenna aperture area, A_r. Here, A_r is the ratio of the physical receiver aperture area, $\pi \left(\frac{D}{2}\right)^2$, to the receiver field area, $\pi \left(\frac{\lambda S}{2D}\right)^2$. Thus, the

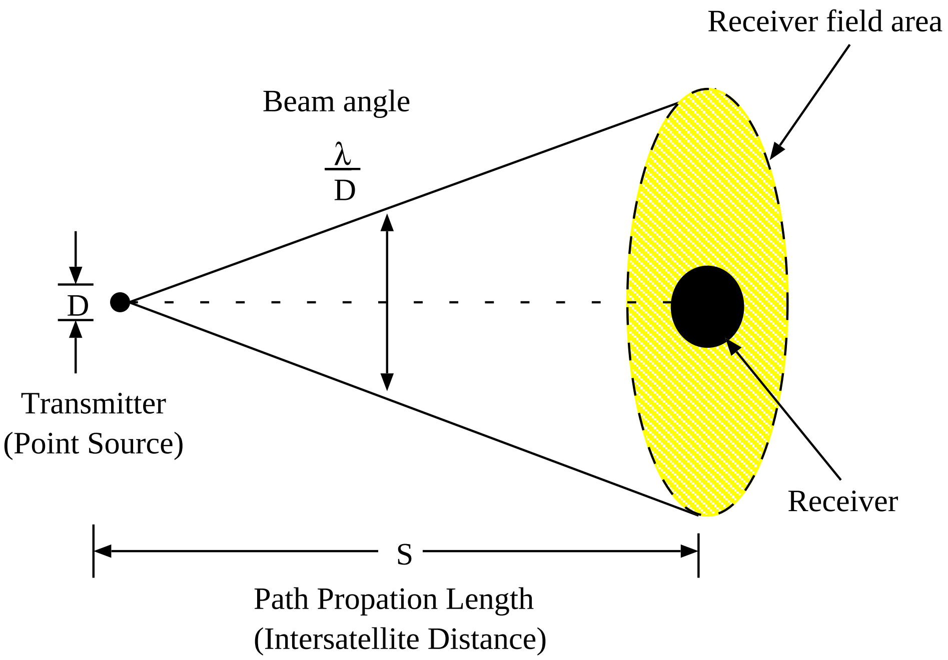


Figure 3-7: Intersatellite link between a transmitter and a receiver.

received power can be simplified in the following manner:

$$\begin{aligned} P_r &= P_t A_r \\ &= P_t \left[\frac{\pi \left(\frac{D}{2}\right)^2}{\pi \left(\frac{\lambda S}{2D}\right)^2} \right] \\ &= \frac{P_t D^4}{\lambda^2 S^2} \end{aligned} \tag{3.7}$$

The *data rate*, R, is proportional to P_t and can be written as:

$$R = \frac{P_t D^4}{\beta h \nu \lambda^2 S^2} \tag{3.8}$$

where β is a measure of receiver sensitivity and $h\nu$ is the energy of a photon. All variable definitions and their units of measure for calculating the cost of an antenna are listed in Table 3.3. From Equation (3.8), D can be rewritten as a function of R, S, and P_t:

$$D = \left(\frac{R \beta h v \lambda^2 S^2}{P_t} \right)^{\frac{1}{4}} \tag{3.9}$$

Symbol	Variable Definition	Units
P_t	Transmitter power	[W]
D	Antenna diameter	[m]
β	Receiver sensitivity parameter	$\left[\frac{\text{J}\cdot\text{s}}{\text{bits/sec}}\right]$
h	Planck's constant $= 6.6262 \times 10^{-34}$	[J · s]
ν	Frequency $= \frac{c}{\lambda}$ $c =$ Speed of light $= 2.99792458 \times 10^8$ [m/sec]	[Hz]
λ	Operating laser wavelength	[m]
S	Path propagation length (intersatellite distance)	[m]

Table 3.3: Variable definitions and units of measure for antenna calculation.

The intersatellite link distance, S, between neighboring satellites connected in a ring can be determined by using the Law of Cosines:

$$\begin{aligned} S &= \sqrt{(R_e + H)^2 + (R_e + H)^2 - 2(R_e + H)(R_e + H)\cos\theta} \\ &= \sqrt{2(R_e + H)^2(1 - \cos\theta)} \qquad (3.10) \end{aligned}$$

where θ is the geostationary arc (i.e., $\theta = \frac{2\pi}{N}$), R_e is the radius of the Earth (i.e., R_e = 6378.14 km), and H is the altitude of the satellite node (e.g., H = 35,786 km). A graphical example of the geometry involved is illustrated in Figure 3-8. A plot of the range of intersatellite link distance as a function of the geostationary arc is presented in Figure 3-9, along with the intersatellite link distances between satellite nodes in a ring for $3 \leq N \leq 20$. The longest intersatellite link distance is approximately 84,080 km and occurs when the geostationary arc is 171.2995°. An intersatellite link (also referred to as a crosslink) cannot be established beyond this point due to Earth blockage.

To determine the other crosslink distances between satellite nodes (e.g., between Satellite 2 and Satellite 4 in Figure 3-8), Equation (3.10) is slightly modified to:

$$S_{i,j} = \sqrt{2(R_e + H)^2 \left[1 - \cos\left(\frac{2\pi\epsilon}{N}\right)\right]} \qquad (3.11)$$

where ϵ is the minimum number of hops away (in a virtual ring network) between the source node i and the destination node j.

Substituting Equation (3.9) and Equation (3.11) into Equation (3.6), the cost of an antenna cost between satellite nodes i and j can be rewritten as:

$$C_{\text{antenna}_{i,j}} = k_0 + k_1 \left(\frac{2R\beta h v \lambda^2 {S_{i,j}}^2}{P_t} \right)^{\frac{\alpha}{4}} \tag{3.12}$$

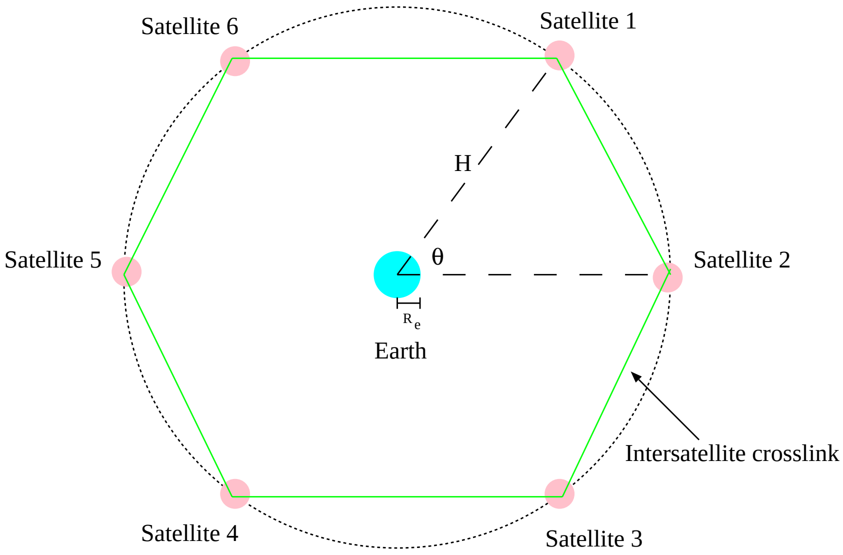


Figure 3-8: Intersatellite link distance geometry.

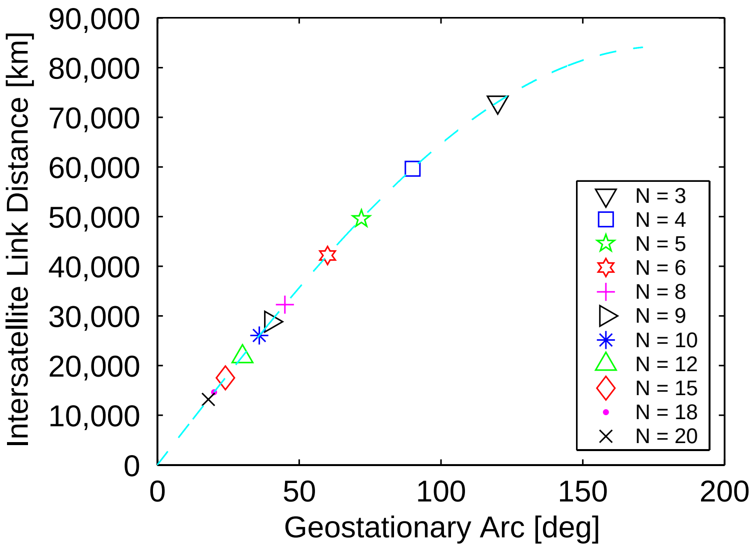


Figure 3-9: Intersatellite link distances for GEO satelllites in a ring constellation.

3.1.4.2 Switch Cost

The cost of a switch is driven by the total number of wavelengths adjoining the satellite node and can be written as:

$$C_{\text{switch}_i} = k_2 \left(\sum_{j=1}^{r+1} W_{i,j} \right)^{\gamma}, \quad \gamma \geq 1 \tag{3.13}$$

where k_2 is the cost constant, r is the node degree, γ is a constant that depends on the switching technology, and $W_{i,j}$ is the number of transmitting wavelengths between satellite nodes i and j. Notice that there are $r + 1$ connections because a satellite node can send traffic to itself.

Switch with Non-Linear Cost: Equation 3.13 is modeled for a switch with a non-linear cost structure (e.g., crossbar or 2-Dimensional [2-D] technology), i.e., values for k_2 are specific for $\gamma = 2$. The results shown in this chapter are based on the use of a switch with non-linear cost in the backbone satellite nodes.

Switch with Linear Cost: The cost effect of a switch with a linear cost structure (e.g., 3-Dimensional Micro-Electro-Mechanical Systems [3-D MEMS] technology) is also studied, where the results are shown in Appendix A.3. For a comparable switch with a linear cost function, the cross-over point between the two types of switches must first be determined:

$$k_2 W^2 \overset{?}{=} k_4 W \tag{3.14}$$

Figure 3-10 illustrates the question posed by Equation 3.14. Based on the switching industry, the cross-over point is determined to be $W^* = 64$. Substitution of W^* into Equation 3.14 yields the relation $k_4 = 64k_2$.

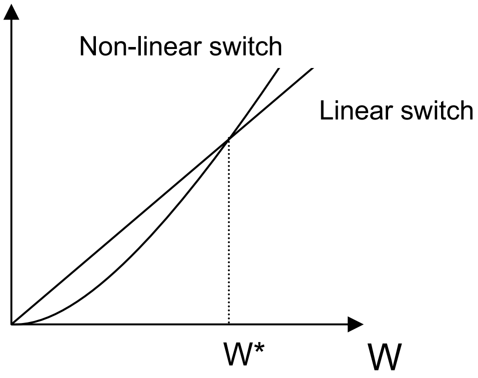


Figure 3-10: Determining the cross-over point between the linear cost switch and the non-linear cost switch.

3.1.4.3 Link Cost

The cost of an intersatellite link (i.e., transmitter/receiver cost) is driven by the total number of wavelengths on the link and can be written as:

$$C_{\text{link}_i} = k_3 \left(\sum_{j=1}^{r} W_{i,j} \right) \tag{3.15}$$

where k_3 is the cost constant, r is the node degree, and $W_{i,j}$ is the number of transmitting wavelengths between satellite nodes i and j.

3.1.5 Communications Cost Modeling Results

Results for designing the space-based information network backbone to meet demand under various traffic models and at minimal cost are provided. A matrix of the network topologies considered and the implemented traffic scenarios are shown in Table 3.4 with reference to the chapter section of their respective results. A cost comparison between designing one constellation system vs. two separate constellation systems for two user communities of varying amounts of mixed traffic is also provided.

	Uniform All-to-All Traffic	Uniform All-to-One Traffic	Mixed Traffic
Connected Circulant Constellations	Sec. 3.1.5.2	Sec. 3.1.5.5	Sec. 3.1.5.6 Part 1
Hub Constellations	Sec. 3.1.5.4	Sec. 3.1.5.3	Sec. 3.1.5.6 Part 2

Table 3.4: Matrix of network topology and traffic analyses.

3.1.5.1 Parameter Values and Cost Values

In order to evaluate the communications cost for various constellations under different traffic scenarios, parameter values must first be defined. All parameter values necessary for calculating the cost of an antenna (telescope pair) and their units of measure are listed in Table 3.5. To calculate the energy of a photon, $h\nu$, the relation $\nu = \frac{c}{\lambda}$ is used, where c is the speed of light. Values of the fixed costs and variable costs for the design variables in the antenna cost, switch cost, and link cost are listed in Table 3.6.

To determine the variable cost for an antenna, k_1, it is assumed that the variable cost is proportional to the fixed cost, k_0, with reference to a 0.15 m antenna. The relation can be written as:

$$k_{1_{i,j}} = \frac{k_0 {D_{i,j}}^{\alpha}}{0.15^{\alpha}} \tag{3.16}$$

where D is the aperture diameter that is required to connect satellite node i to satellite node j and α is a constant factor for the antenna technology, which is fixed at a value of 2.6.

Cost levels for each communication component is defined as low, medium, and high. These costs may not be attainable today but are expected in the future. Table 3.7 shows the 27 permutations of possible cost cases available for assessing the communications costs of the space-based information network backbone. This chapter will focus on the more likely future MLM cost case (medium cost antennas, low cost switches, and medium cost links). A similar set of results for the likely future HMH (high cost antennas, medium cost switches, and high cost links) cost case is also provided. Because the results for the HMH cost case show similar trends to the MLM cost case, most of the HMH cost results are presented

Symbol		**Parameter Value**	**Unit**
P_t	=	10	[W]
β	=	200	$\left[\frac{\text{J·s}}{\text{bits/sec}}\right]$
h	=	6.6262×10^{-34}	[J · s]
λ	=	1.5	[μm]
R_e	=	6378.14	[km]
H	=	35,786	[km]
R	=	40	[Gbps]
c	=	2.99792458×10^8	[m/s]

Table 3.5: Parameter values and units of measure for antenna cost calculation.

in Appendix A.1. Values of the fixed costs and variable costs for the implementation of a linear switch is provided in Table 3.8. Cost results using a linear switch is then presented in Appendix A.3.

Component	**Cost Level**	**Cost Value**
Antenna $\alpha = 2.6 \quad k_{1_{i,j}} = \frac{k_0 D_{i,j}{}^{\alpha}}{0.15^{\alpha}}$	Low Medium High	$k_0 = \$10,000$ $k_0 = \$100,000$ $k_0 = \$1,000,000$
Switch $\gamma = 2$	Low Medium High	$k_2 = \$2,000$ $k_2 = \$10,000$ $k_2 = \$100,000$
Link	Low Medium High	$k_3 = \$10,000$ $k_3 = \$100,000$ $k_3 = \$1,000,000$

Table 3.6: Design variable values for communications cost components.

LLL	LML	LHL
LLM	LMM	LHM
LLH	LMH	LHH
MLL	MML	MHL
MLM	MMM	MHM
MLH	MMH	MHH
HLL	HML	HHL
HLM	HMM	HHM
HLH	HMH	HHH

Key: $X_1X_2X_3$
X_1: Antenna Cost (Low, Medium, High)
X_2: Switch Cost (Low, Medium, High)
X_3: Link Cost (Low, Medium, High)

Table 3.7: Matrix of cost permutations for communications cost components.

Component	**Cost Level**	**Cost Value**
Switch $\gamma = 1$	Low Medium High	$k_4 = \$128,000$ $k_4 = \$640,000$ $k_4 = \$6,400,000$

Table 3.8: Design variable values for linear switch.

3.1.5.2 Connected Circulant Constellations under Uniform All-to-All Traffic

In this architectural study, traffic is routed according to the criteria of minimizing the number of hops. The *average minimum hop distance* for the space-based network backbone topology is defined as:

$$\begin{aligned} H_{\text{min}}(N, r) &= \frac{1}{N^2} \sum_i^N \sum_j^N d(i,j) \\ &= \frac{1}{N} \sum_j^N d(i,j) \end{aligned} \tag{3.17}$$

where $d(i,j)$ is the minimum length of the path between satellite nodes i and j. For a ring constellation ($r = 2$), the average minimum hop distance can be determined and written in analytical form as:

$$\begin{aligned} H_{\text{min}}(N_{even}, r = 2) &= \frac{1}{N}\left(1 + \frac{N}{2} + \sum_{v=1}^{\frac{N}{2}-1} 2i\right) \\ &= \frac{1}{N}\left(1 + \frac{N}{2} + \frac{1}{4}N^2 - \frac{1}{2}N\right) \\ &= \frac{N^2+4}{4N} \end{aligned} \tag{3.18}$$

$$\begin{aligned} H_{\text{min}}(N_{odd}, r = 2) &= \frac{1}{N}\left(1 + \sum_{v=1}^{\frac{N-1}{2}} 2i\right) \\ &= \frac{1}{N}\left(1 + \frac{1}{4}N^2 - \frac{1}{4}\right) \\ &= \frac{N^2+3}{4N} \end{aligned} \tag{3.19}$$

While there is no closed-form solution for determining the average minimum hop distance for every pair of N and r, it can be numerically calculated. The average minimum hop distance for connected circulant constellations considered is plotted in Figure 3-11(a). Ring constellations have the worst average minimum hop distance because they have high transit (pass-thru) traffic, i.e., traffic may have to go through many intermediate nodes in order to reach its destination.

For uniform all-to-all traffic, the number of wavelengths required for each intersatellite link can be calculated as:

$$W(N,r) = \left\lceil \frac{H_{\text{min}}(N,r) \times T}{Nr} \right\rceil \tag{3.20}$$

where H_{min} is the average minimum hop distance, T is the total amount of traffic in the network, N is the number of nodes, and r is the node degree. The wavelength dimensions of connected circulant constellations is plotted in Figure 3-11(b). Rings require the greatest number of wavelengths on the intersatellite links, again due to high transit traffic.

The communications costs for connected circulant constellations as a function of N nodes and node degree r for the MLM and HMH cost cases under uniform all-to-all traffic are plotted in Figure 3-12 and Figure 3-13, respectively. For each cost case, two design examples are shown. The first design example uses custom antennas on the satellites, i.e., the antenna apertures are sized to the intersatellite link distance required to make the network connection. In the second design example, every antenna on the satellites is identical or uniform (i.e., the antenna apertures are sized to the largest theoretical intersatellite link distance [$\theta = 180°$]). Subfigures (a) and (b) show the communications cost for connected circulant networks as a function of N nodes and node degree r. Subfigures (c) and (d) show which node degree r provides the lowest communications cost for every N number of nodes.

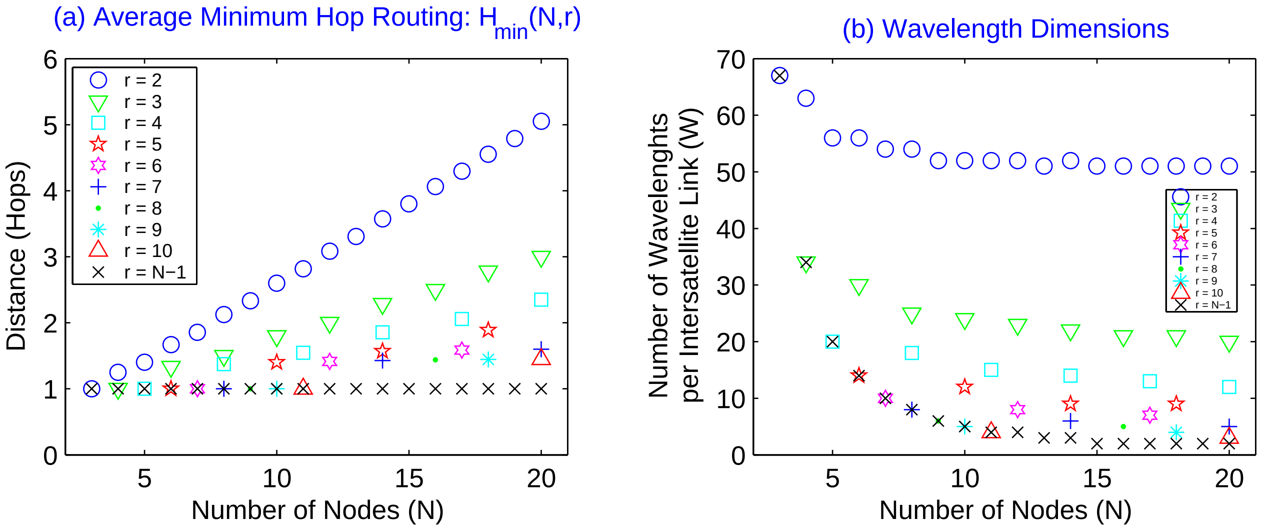


Figure 3-11: Connected circulant constellations: traffic characteristics for uniform all-to-all traffic.

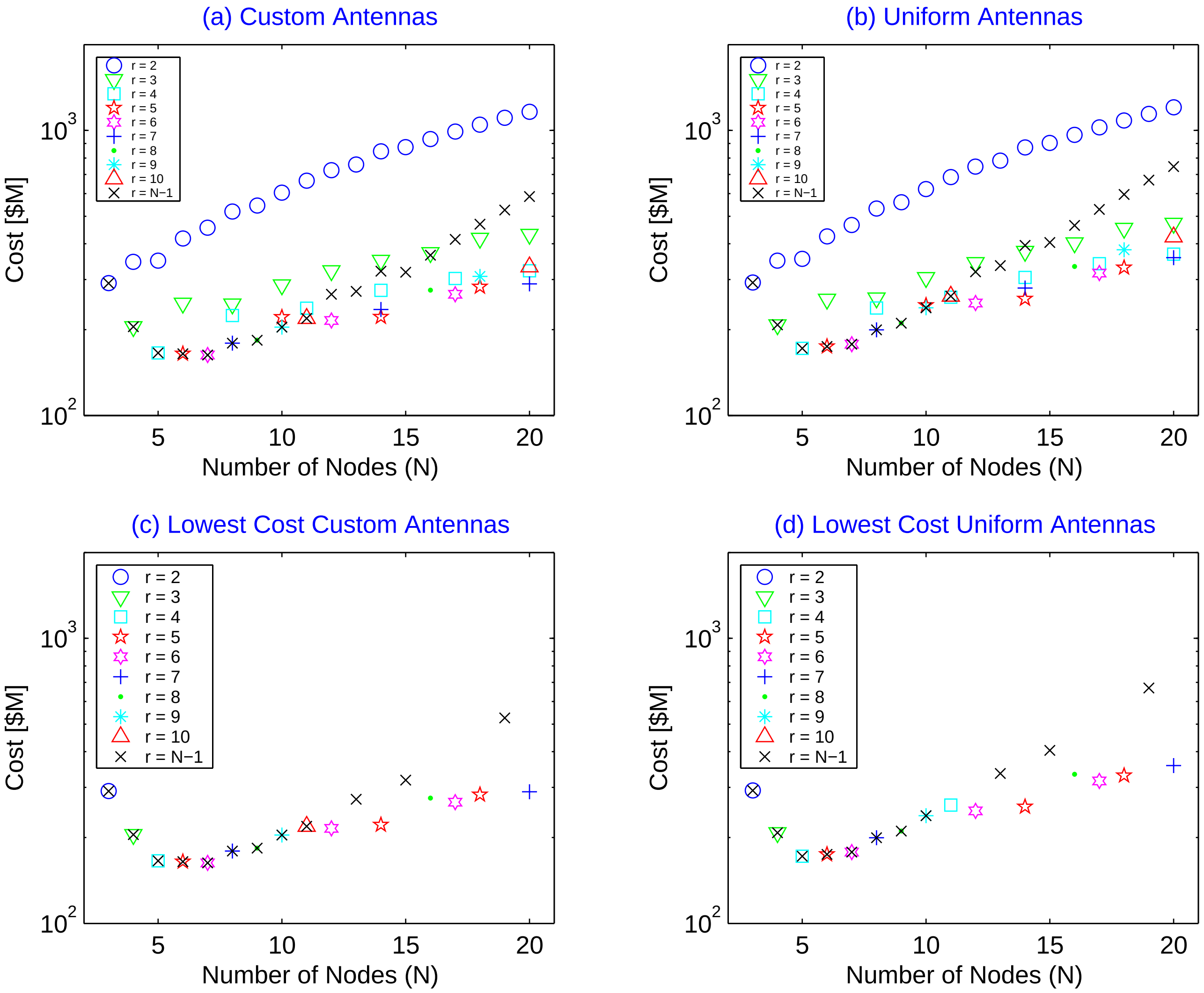


Figure 3-12: Connected circulant constellations: communications cost MLM for uniform all-to-all traffic.

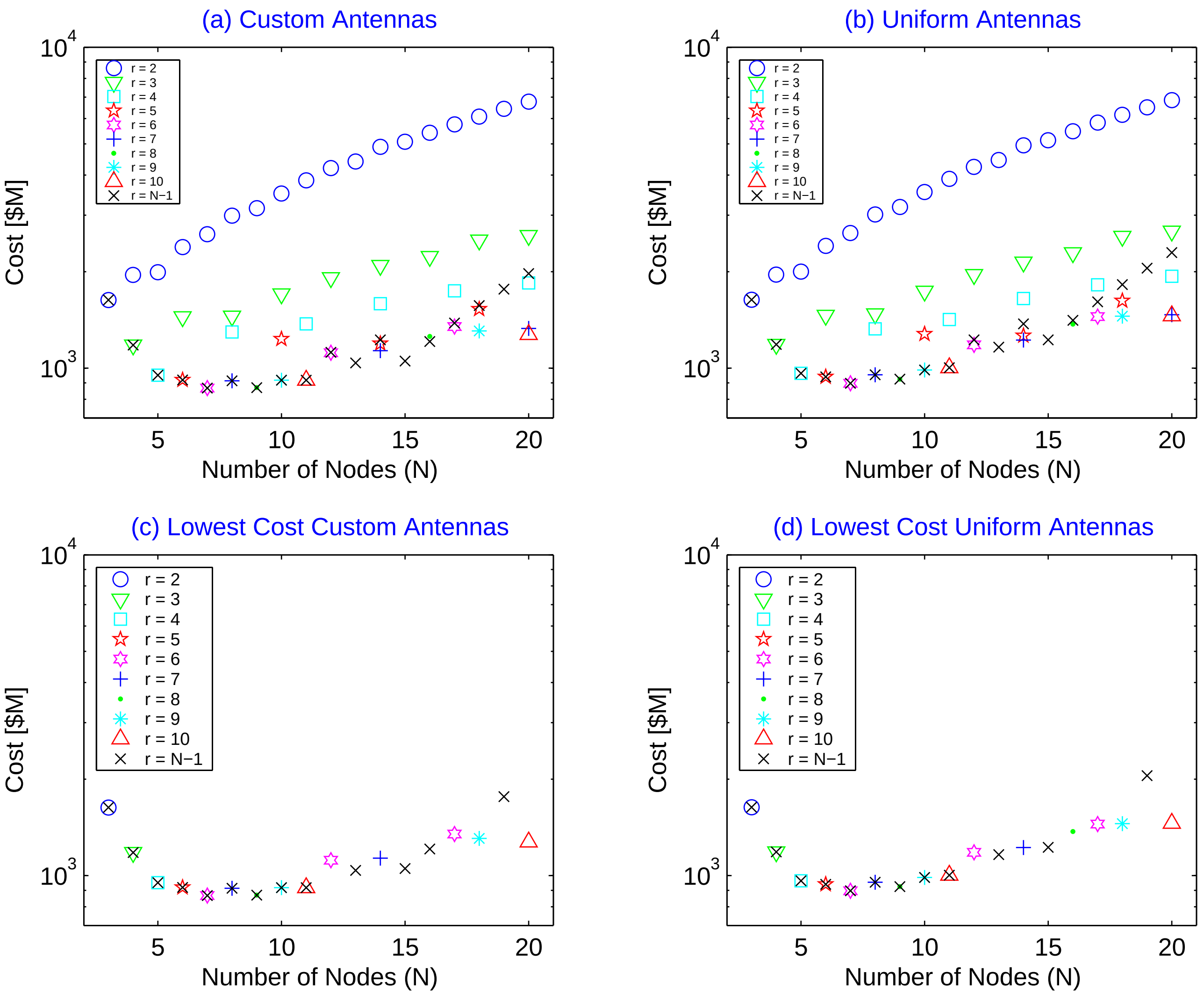


Figure 3-13: Connected circulant constellations: communications cost HMH for uniform all-to-all traffic.

The percentage increase in communications cost between using custom antennas and uniform antennas can be determined with the following formula:

$$\%\Delta J = \frac{J_{\text{uniform}} - J_{\text{custom}}}{J_{\text{custom}}} \times 100 \qquad (3.21)$$

where $J_{\text{uniform}} > J_{\text{custom}}$. The results for the MLM cost case in Figure 3-14(a) are obtained from Figure 3-12 and the results for the HMH cost case in Figure 3-14(b) are obtained from Figure 3-13. The percentage increase in communications cost for using uniform antennas is small ($< 30\%$ overall). For small constellations ($N \leq 11$), the percentage increase in communications cost for using uniform antennas is less than 10%. As uniform antennas provide the greatest flexibility and reliability for making any intersatellite connection, they will be used in the remaining analyses. Based on the parameter values in Table 3.5 and design variable values in Table 3.6, the diameter of the antenna aperture for uniform antennas is 20.29 cm.

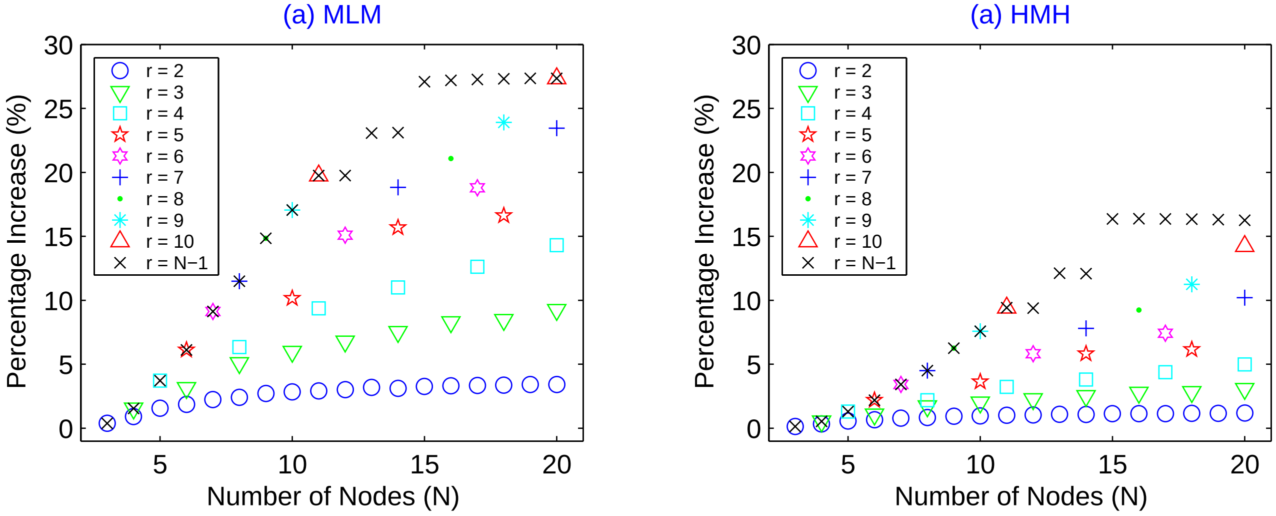


Figure 3-14: Connected circulant constellations: percentage increase in communications cost between custom antennas and uniform antennas for uniform all-to-all traffic.

The main observation in designing connected circulant constellations for uniform all-to-all traffic is that a high degree of connectivity is better for small constellations, as shown in Figures 3-11, 3-12, and 3-13. Ring constellations ($r = 2$) are the least cost-effective due to high transit (pass-thru) traffic which ties up more resources (e.g., transmitters) thereby driving up the communications cost. Full mesh constellations for $3 \leq N \leq 11$ are shown to be the most cost-effective. This trend stems from the fact that the switch is the main driver for the connectivity of the constellation architectures. The recommendation of a low cost connected circulant constellation for uniform all-to-all traffic is affected by the non-linear cost structure of the switch size. Switch size is affected by the total number of wavelengths in the system. Figure 3-15 shows the traffic characteristics of a 100 wavelength system on constellation sizes of $3 \leq N \leq 10$ while Figure 3-16 shows the resulting communications costs under the MLM cost scenario. It is observed that full meshes are no longer the most cost-effective constellation.

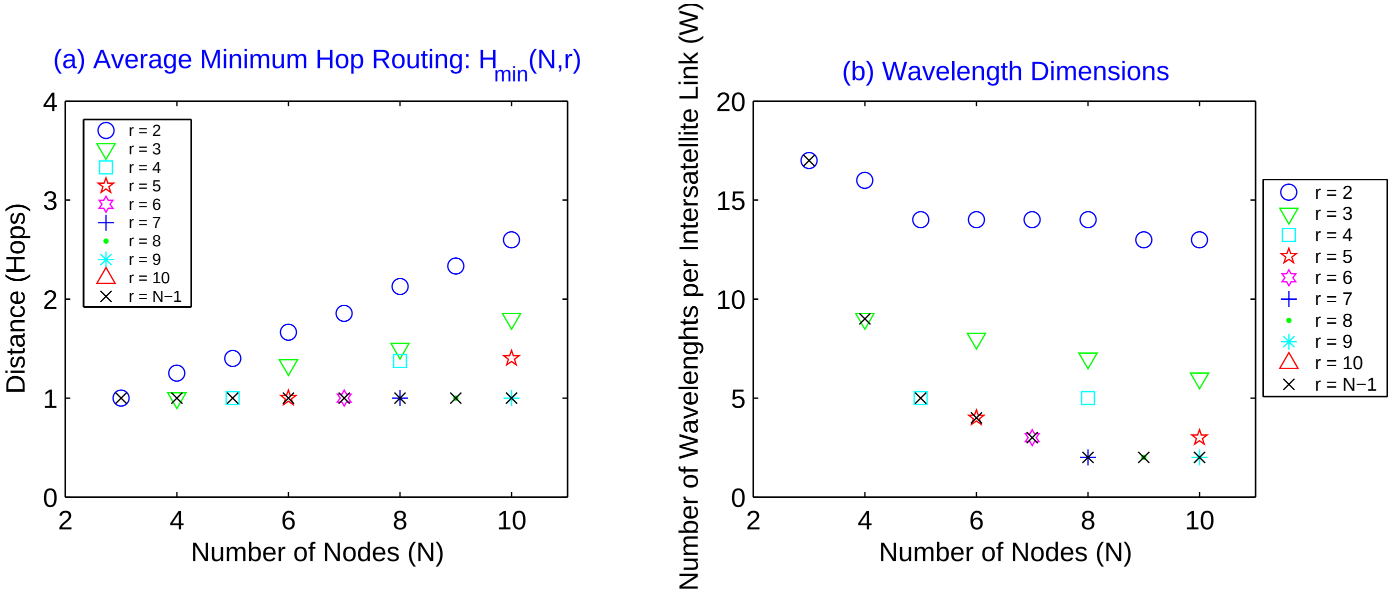


Figure 3-15: Connected circulant constellations: traffic characteristics for uniform all-to-all traffic of 100 wavelengths.

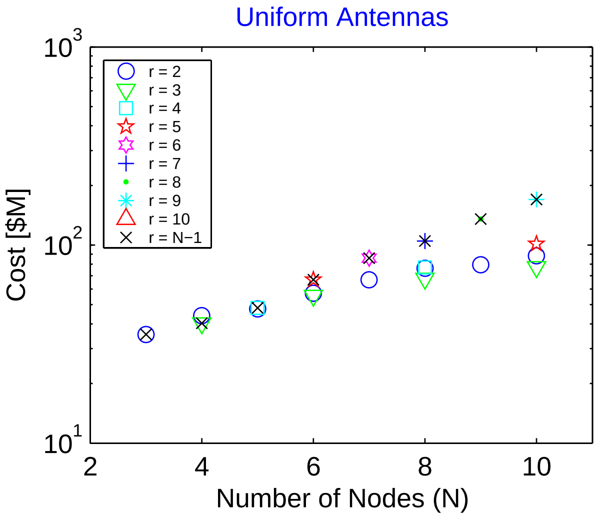


Figure 3-16: Connected circulant constellations: communications cost MLM for uniform all-to-all traffic of 100 wavelengths.

3.1.5.3 Hub Constellations under Uniform All-to-One Traffic

With uniform all-to-one traffic on 1-hub constellations, the hub node receives traffic from every node, including traffic from the hub node itself. The number of wavelengths on the intersatellite links required for such connections can be calculated as:

$$W = \left\lceil \frac{T}{N} \right\rceil \tag{3.22}$$

where T is the total amount of traffic on the network and N is the number of satellite nodes.

On a 2-hub constellation, each hub handles half the total amount of traffic in the network. The number of wavelengths required on the intersatellite links for traffic from plain nodes to hub nodes remains the same. A 2-hub constellation can provide greater reliability than a 1-hub constellation via the use of a crosslink between the two hub nodes. For example, if the downlink on hub node 1 is not operational, network traffic to hub node 1 can be transmitted to hub node 2 for transmission to the ground. To handle this possibility, the crosslink between the two hub nodes require rW wavelengths, where $r = \frac{N}{2}$ and W is derived from Equation (3.22). The wavelength dimensions for uniform all-to-one traffic on 1-hub and 2-hub constellations are shown in Figure 3-17.

For both types of hub constellations, two network communications costs can be derived depending on the types of satellites designed, as shown in Figure 3-18. In one instance, each satellite in the backbone network is custom-built to handle its respective traffic load. In the other instance, uniform or identical satellites are used to build the backbone network; therefore, every satellite must be modeled after the hub satellite. Contrast to 1-hub constellations, 2-hub constellations are more cost-effective because the two hub nodes handle half of the total amount of traffic in the network (i.e., the size of the switches on the hub nodes in a 2-hub constellation is smaller). This observation shows that the cost of the nonlinear switch is the main driver for the total communications cost of the satellite network backbone.

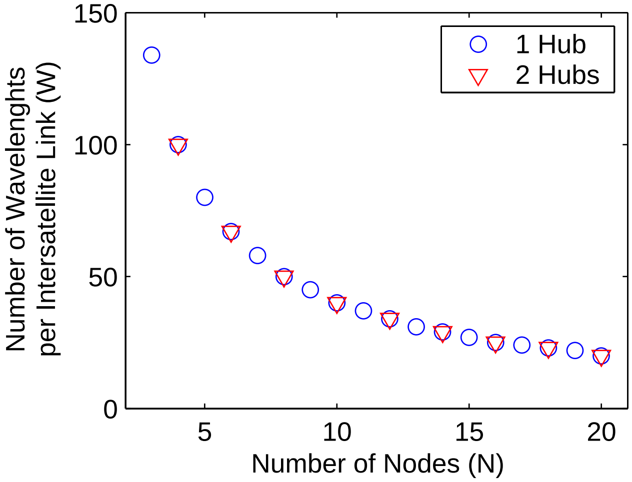


Figure 3-17: Hub constellations: wavelength dimensions for uniform all-to-one traffic.

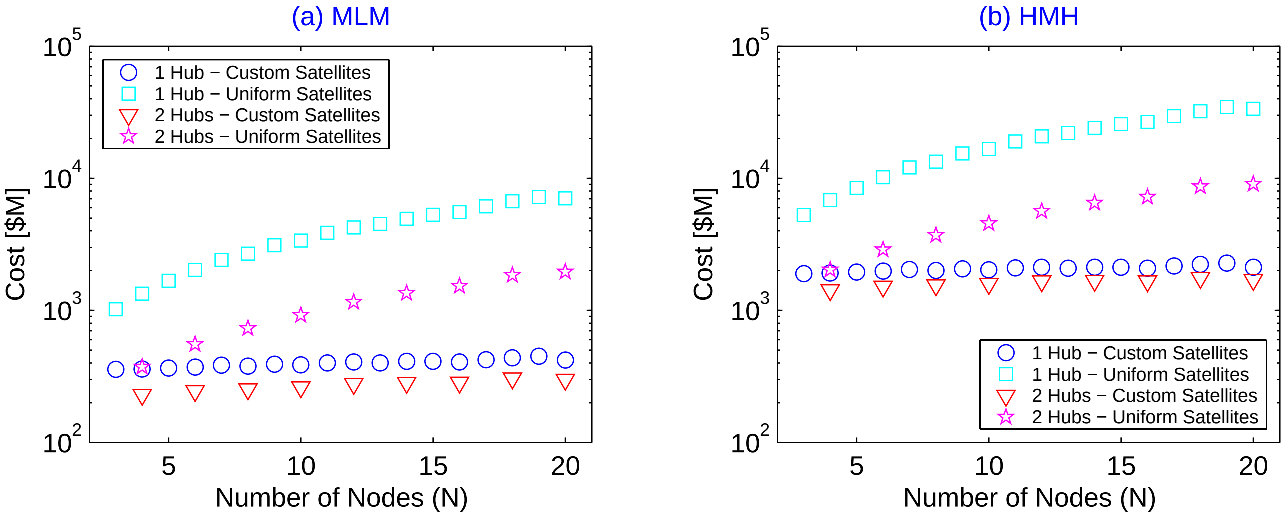


Figure 3-18: Hub constellations: communications cost for uniform all-to-one traffic.

3.1.5.4 Hub Constellations under Uniform All-to-All Traffic

In this section, the effects of uniform all-to-all traffic on hub constellations are examined. The average minimum hop distance for all source-destination pairs increases slightly, as shown in Figure 3-19(a), thereby affecting the number of wavelengths required for the intersatellite links to the hub node(s) [see Figure 3-19(b)]. Communications costs for hub constellations under uniform all-to-all traffic are shown for the cost cases of MLM and HMH in Figure 3-20 (a) and (b), respectively.

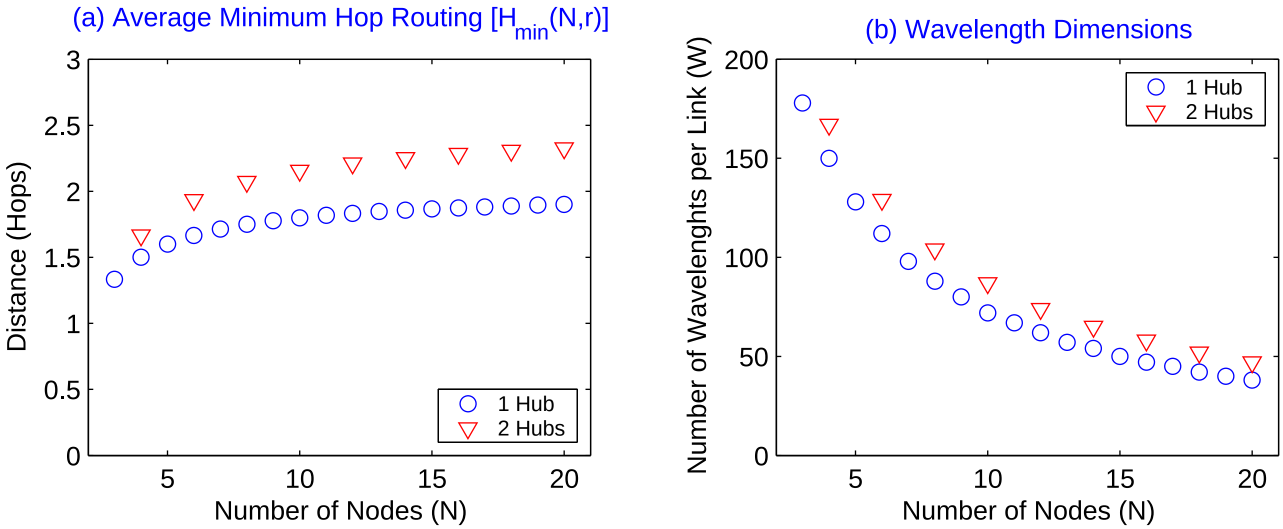


Figure 3-19: Hub constellations: traffic characteristics for uniform all-to-all traffic.

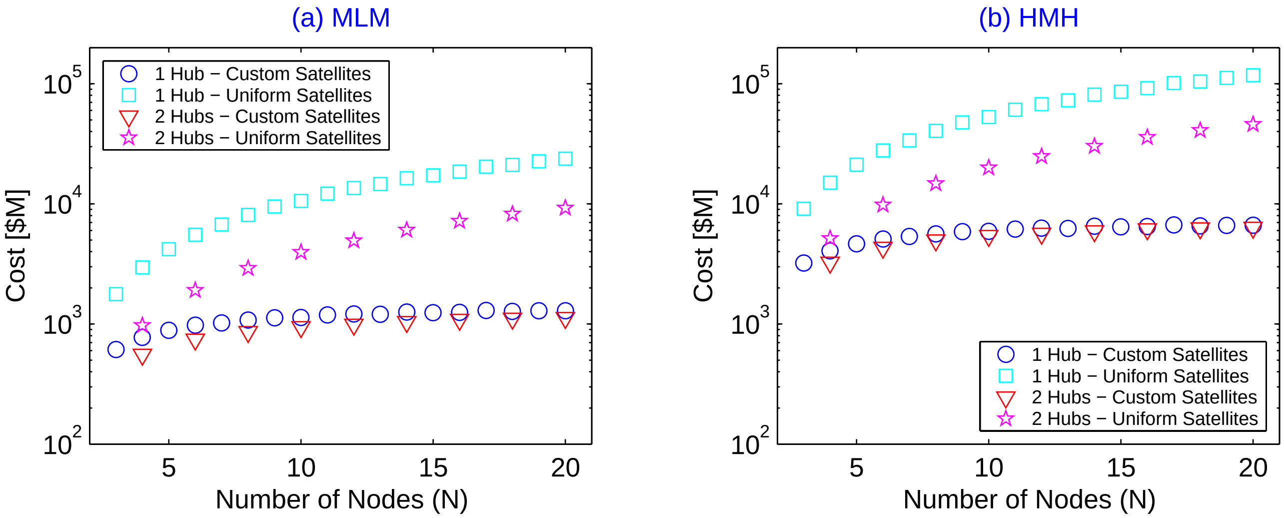


Figure 3-20: Hub constellations: communications cost for uniform all-to-all traffic.

The main observation in designing hub constellations for uniform all-to-all traffic is that the resulting cost structures are relatively flat (i.e., the cost structures under uniform all-to-all traffic have the same shape as the cost structures under uniform all-to-one traffic). No matter the traffic model, 2-hub constellations are more cost-effective than 1-hub constellations. Once again, hub constellations are less expensive when the satellites are custom-built (i.e., using two different satellite designs).

3.1.5.5 Connected Circulant Constellations under Uniform All-to-One Traffic

In this section, the effects of uniform all-to-one traffic on connected circulant constellations are examined. Alterations in the implementation of Dijkstra's algorithm can yield slight changes in the communications cost of the satellite backbone network due to the changes in wavelength dimensions, switch sizes, and link utilization. Five different algorithm implementations will be discussed to determine the routing algorithm that provides the lowest communications costs for the space-based network backbone: (1) Dijkstra's algorithm, (2) Symmetric Dijkstra's algorithm, (3) Incremental Dijkstra's algorithm, (4) Modified Incremental Dijkstra's algorithm, and (5) Symmetric Modified Incremental Dijkstra's algorithm. For all satellite constellations considered, satellite node 1 is designated as the hub node.

Case I: Dijkstra's algorithm. In the first algorithm implementation, no changes are made to Dijkstra's algorithm. The results for running Dijkstra's algorithm on the example of $N = 8$ nodes with degree $r = 4$ is shown in Figure 3-21, where subfigure (a) shows the traffic matrix, subfigure (b) shows how the matrix is derived from the loading of the source-destination paths, and subfigure (c) highlights the paths that are used for the flow of uniform all-to-one traffic. Communications cost results for the MLM cost case and HMH cost case are shown in Figure 3-22 and Figure A-1, respectively. Within each figure, three variations of satellites are used for the construction of the space-based information network backbone constellation. In the first instance, uniform satellites are used, thus all satellites are modeled after the hub node. In the second instance, two different types of satellites are built, one specific for the hub node and another for the plain nodes. In the third instance, each satellite node is custom-built to handle the traffic that it transmits and receives. In both cost cases (MLM and HMH), the communications cost for the network monotonically decreases as satellite customization increases.

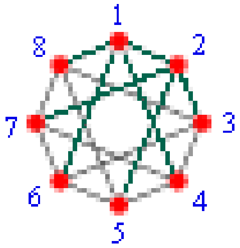
Destination j

		1	2	3	4	5	6	7	8
	1	50	1	0	1	0	1	0	1
	2	200	0	1	0	1	0	1	0
	3	0	50	0	1	0	1	0	1
Source i	4	50	0	1	0	1	0	1	0
	5	0	50	0	1	0	1	0	1
	6	50	0	1	0	1	0	1	0
	7	0	50	0	1	0	1	0	1
	8	50	0	1	0	1	0	1	0

(a)

Min. Hop	Shortest path
-	1 1
1	2 1
2	3 2 1
1	4 1
2	5 2 1
1	6 1
2	7 2 1
1	8 1

(b)



(c)

Figure 3-21: Wavelength dimensioning with Dijkstra's algorithm on the connected circulant constellation with 1 hub node for $N = 8$ and $r = 4$.

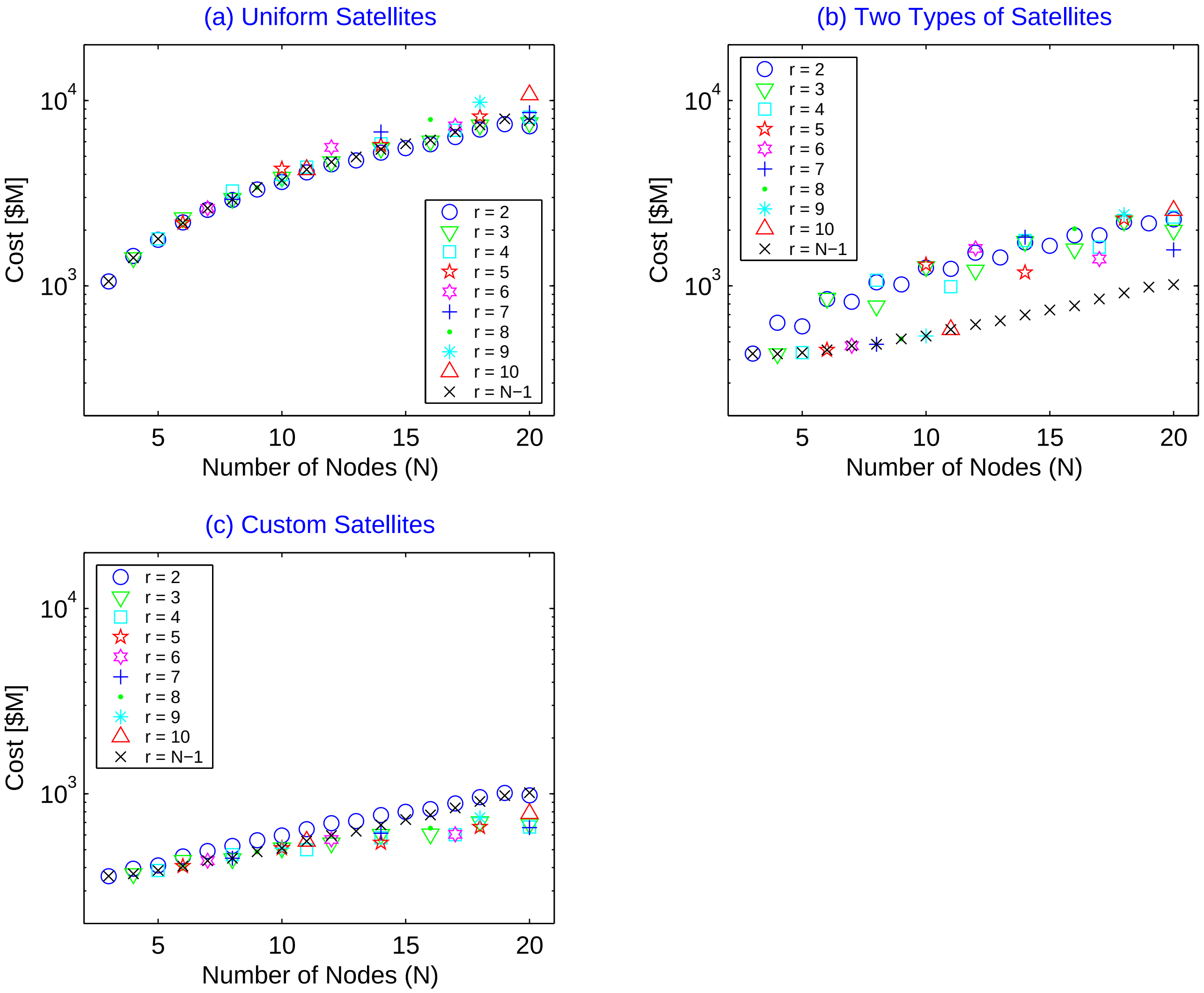


Figure 3-22: Connected circulant constellations with 1 hub node: communications cost MLM for uniform all-to-one traffic with Dijkstra's algorithm.

Evaluation of the traffic matrix in Figure 3-21(a) shows that Dijkstra's algorithm is a greedy algorithm that may not provide the optimal solution for minimizing communications cost. With Dijkstra's algorithm, because shortest paths for each source-destination pair are chosen in topological order, traffic is not symmetrically balanced.

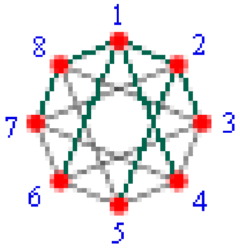
Case II: Symmetric Dijkstra's algorithm. In the second algorithm implementation, Dijkstra's algorithm is implemented in such a way that routing is symmetric on the satellite network constellation. The wavelength dimensioning results for $N = 8$ and $r = 4$ are shown in Figure 3-23 while the communications cost results for the MLM cost case are shown in Figure 3-24 and the communications cost results for the HMH cost case are shown in Figure A-2. Notice that the Symmetrical Dijkstra's algorithm yields a lower communications cost. The percentage decrease in communications costs between Dijkstra's algorithm and symmetric Dijkstra's algorithm is shown in Figure 3-25 for the MLM cost case and in Figure A-3 for the HMH cost case.

		Destination j							
		1	2	3	4	5	6	7	8
	1	50	1	0	1	0	1	0	1
	2	150	0	1	0	1	0	1	0
	3	0	50	0	1	0	1	0	1
Source i	4	50	0	1	0	1	0	1	0
	5	0	50	0	1	0	1	0	1
	6	50	0	1	0	1	0	1	0
	7	0	50	0	1	0	1	0	50
	8	50	0	1	0	1	0	1	0

(a)

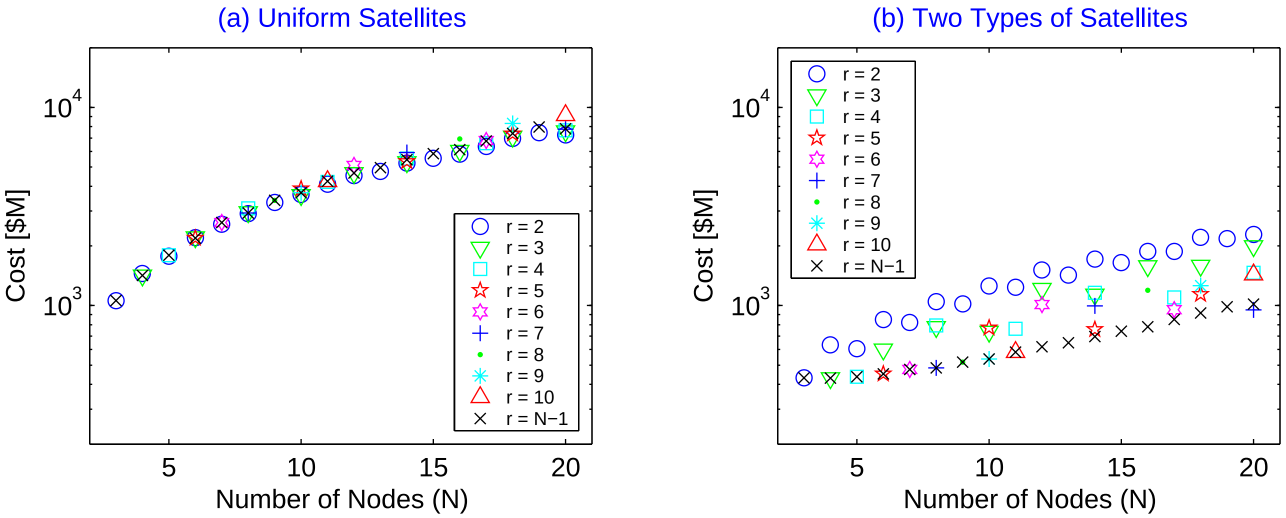
Min. Hop	Shortest path
-	1 1
1	2 1
2	3 2 1
1	4 1
2	5 2 1
1	6 1
2	7 8 1
1	8 1

(b)



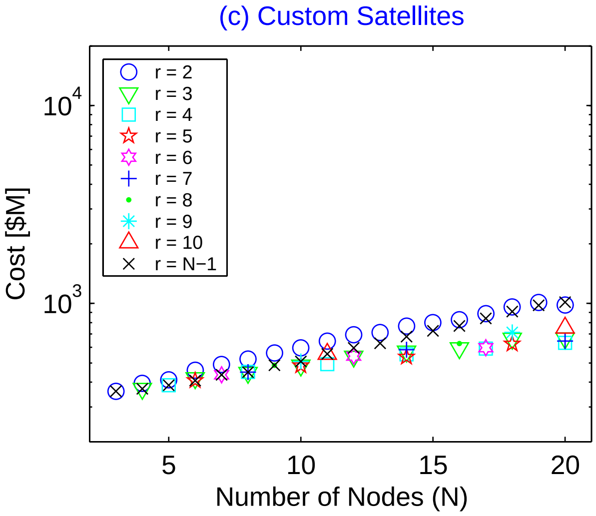
(c)

Figure 3-23: Wavelength dimensioning with Symmetric Dijkstra's algorithm on the connected circulant constellation with 1 hub node for $N = 8$ and $r = 4$.



(a) Uniform Satellites

(b) Two Types of Satellites



(c) Custom Satellites

Figure 3-24: Connected circulant constellations with 1 hub node: communications cost MLM for uniform all-to-one traffic with Symmetric Dijkstra's algorithm.

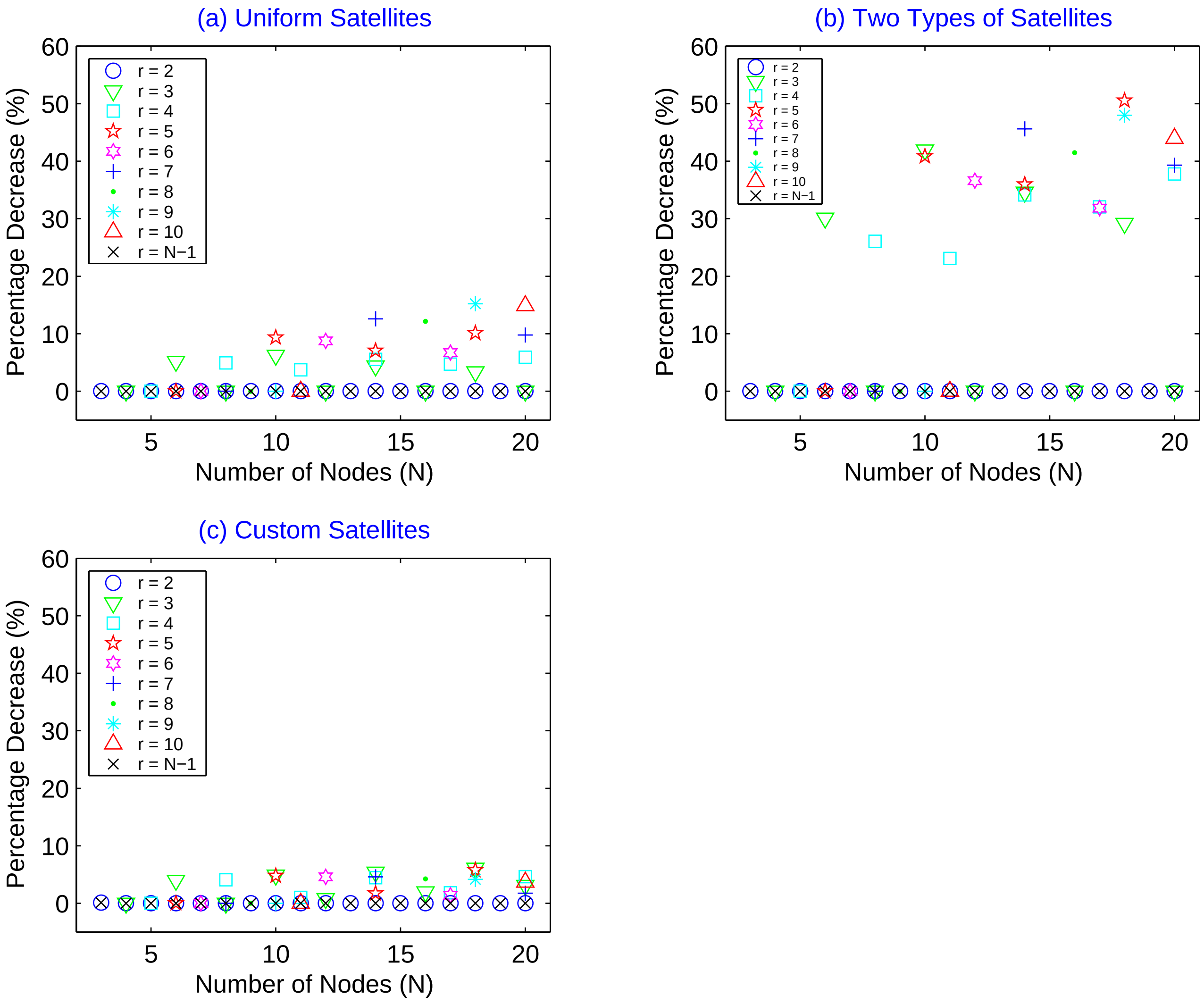


Figure 3-25: Connected circulant constellations with 1 hub node: percentage decrease in communications cost MLM between Dijkstra's algorithm and Symmetric Dijkstra's algorithm for uniform all-to-one traffic.

Both implementations of Dijkstra's algorithm and Symmetric Dijkstra's algorithm generate the shortest path for each source-destination pair independent of other traffic on the network. Consequently, these cases do not consider the objective of load balancing on the network. When load balancing is taken into account, the shortest path for each destination pair may be altered.

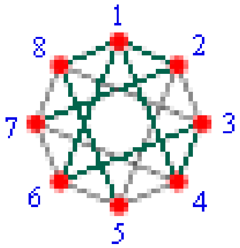
Case III: Incremental Dijkstra's algorithm. In this third algorithm implementation, traffic for each source-destination pairs are loaded onto the network in topological order (i.e., load shortest path for source node 2 to hub node 1, load shortest path for source node 3 to hub node 1, ..., and load shortest path for source N to hub node 1) while taking into account traffic that is already on the network. The wavelength dimensioning results for $N = 8$ and $r = 4$ are shown in Figure 3-26. Note that the length of the shortest path may no longer be equal to the minimum hop distance for every source-destination pair, as shown in the example in Figure 3-26(b) for source nodes 3, 4, 5, 6, and 7. With such modifications in the wavelength dimensions for the intersatellite links, the communications costs for the satellite network changes as shown in Figure 3-27 and Figure A-4 for the MLM and the HMH cost cases, respectively. Comparison for the percentage decrease in cost from the first case of Dijkstra's algorithm is provided in Figure 3-28 for the MLM cost case and in Figure A-5 for the HMH cost case. These figures show that this implementation of Dijkstra's algorithm does not provide cost savings all-around, rather some constellations are even more expensive.

	Destination j							
Source i	1	2	3	4	5	6	7	8
1	50	1	0	1	0	1	0	1
2	100	0	1	0	1	0	1	0
3	0	1	0	50	0	50	0	1
4	50	0	50	0	1	0	1	0
5	0	1	0	1	0	1	0	50
6	100	0	1	0	1	0	1	0
7	0	50	0	1	0	1	0	1
8	100	0	1	0	1	0	1	0

(a)

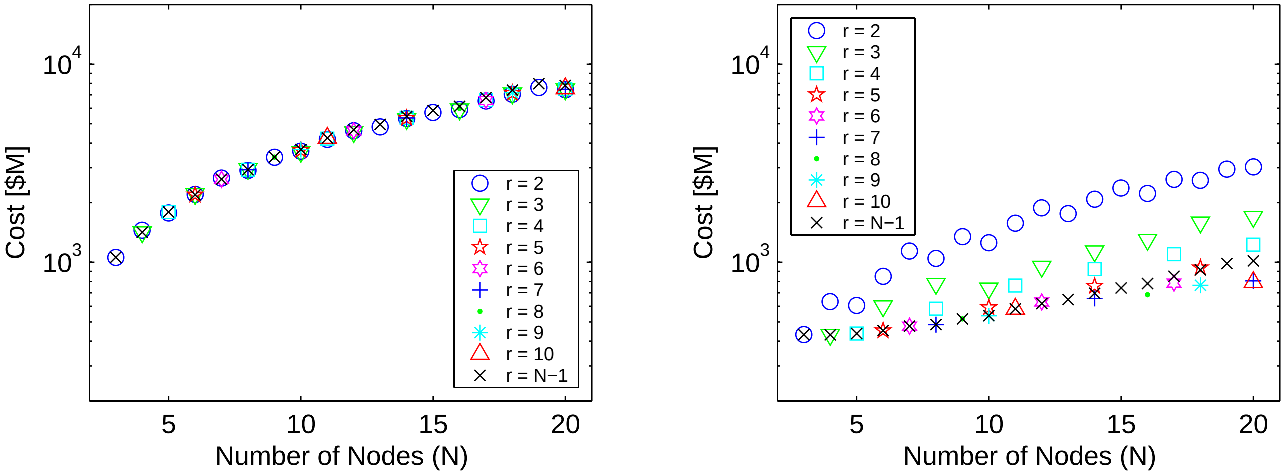
Min. Hop	Shortest path
-	1 1
1	2 1
2	3 4 1
1	4 3 6 1
2	5 8 1
1	6 1
2	7 2 1
1	8 1

(b)



(c)

Figure 3-26: Wavelength dimensioning with Incremental Dijkstra's algorithm on the connected circulant constellation with 1 hub node for $N = 8$ and $r = 4$.


(a) Uniform Satellites

(b) Two Types of Satellites

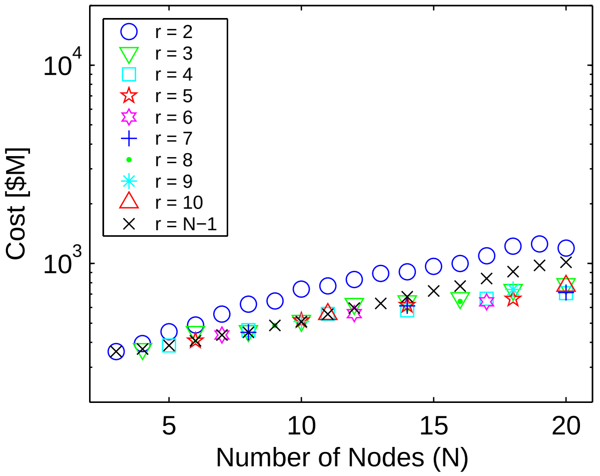

(c) Custom Satellites

Figure 3-27: Connected circulant constellations with 1 hub node: communications cost MLM for uniform all-to-one traffic with Incremental Dijkstra's algorithm.

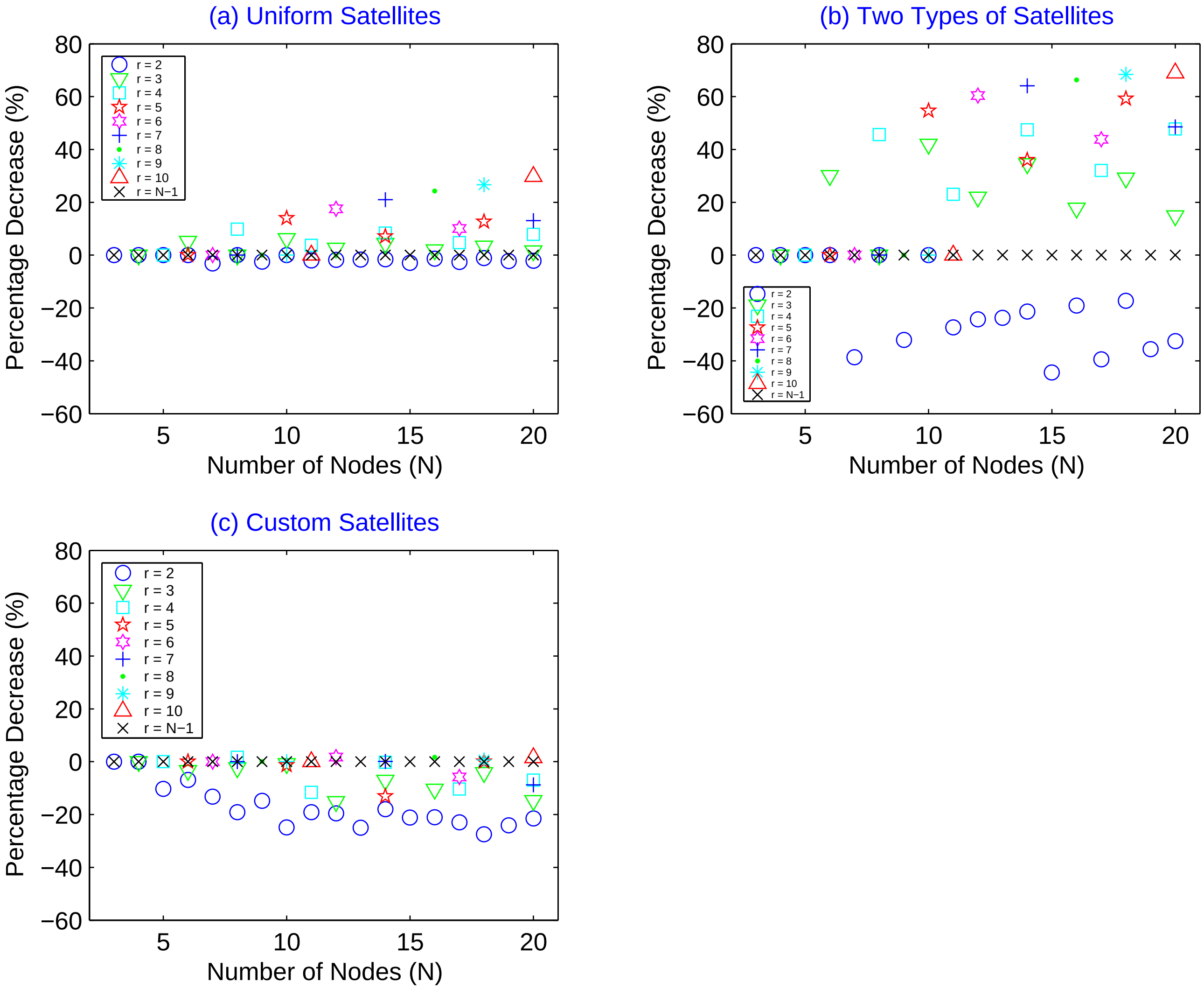


Figure 3-28: Connected circulant constellations with 1 hub node: percentage decrease in communications cost MLM between Dijkstra's algorithm and Incremental Dijkstra's algorithm for uniform all-to-one traffic.

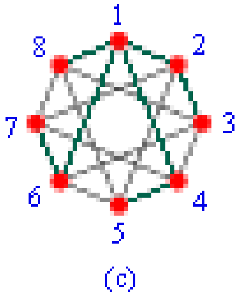
Case IV: Modified Incremental Dijkstra's algorithm. In the fourth algorithm implementation of Dijsktra's algorithm, modifications are made such that traffic loading onto the network is done based on the shortest path distance from each source to the hub. For example, all paths of 1 hop are loaded first (in topological order), then all paths of 2 hops are loaded next, and so on. Given that the traffic load on the network is taken into account during the computation of the shortest path, the number of hops in the calculated shortest path may not be equivalent to the minimum hop distance for the source-destination pair. The wavelength dimensioning results for $N = 8$ and $r = 4$ are shown in Figure 3-29. The resulting communications costs in the MLM cost case are shown in Figure 3-30 while the results in the HMH cost case are shown in Figure A-6. The percentage decrease in cost between Dijkstra's algorithm and this Modified Incremental Dijkstra's algorithm is shown in Figure 3-31 for the MLM cost case and in Figure A-7 for the HMH cost case. Of all the different implementations so far, the Modified Incremental Dijkstra's algorithm has provided the greatest percentage decrease in communications cost.

		Destination j							
		1	2	3	4	5	6	7	8
Source i	1	50	1	0	1	0	1	0	1
	2	100	0	1	0	1	0	1	0
	3	0	50	0	1	0	1	0	1
	4	100	0	1	0	1	0	1	0
	5	0	1	0	50	0	1	0	1
	6	100	0	1	0	1	0	1	0
	7	0	1	0	1	0	50	0	1
	8	50	0	1	0	1	0	1	0

(a)

Min. Hop	Shortest path
-	1 1
1	2 1
2	3 2 1
1	4 1
2	5 4 1
1	6 1
2	7 6 1
1	8 1

(b)



(c)

Figure 3-29: Wavelength dimensioning with Modified Incremental Dijkstra's algorithm on the connected circulant constellation with 1 hub node for $N = 8$ and $r = 4$.

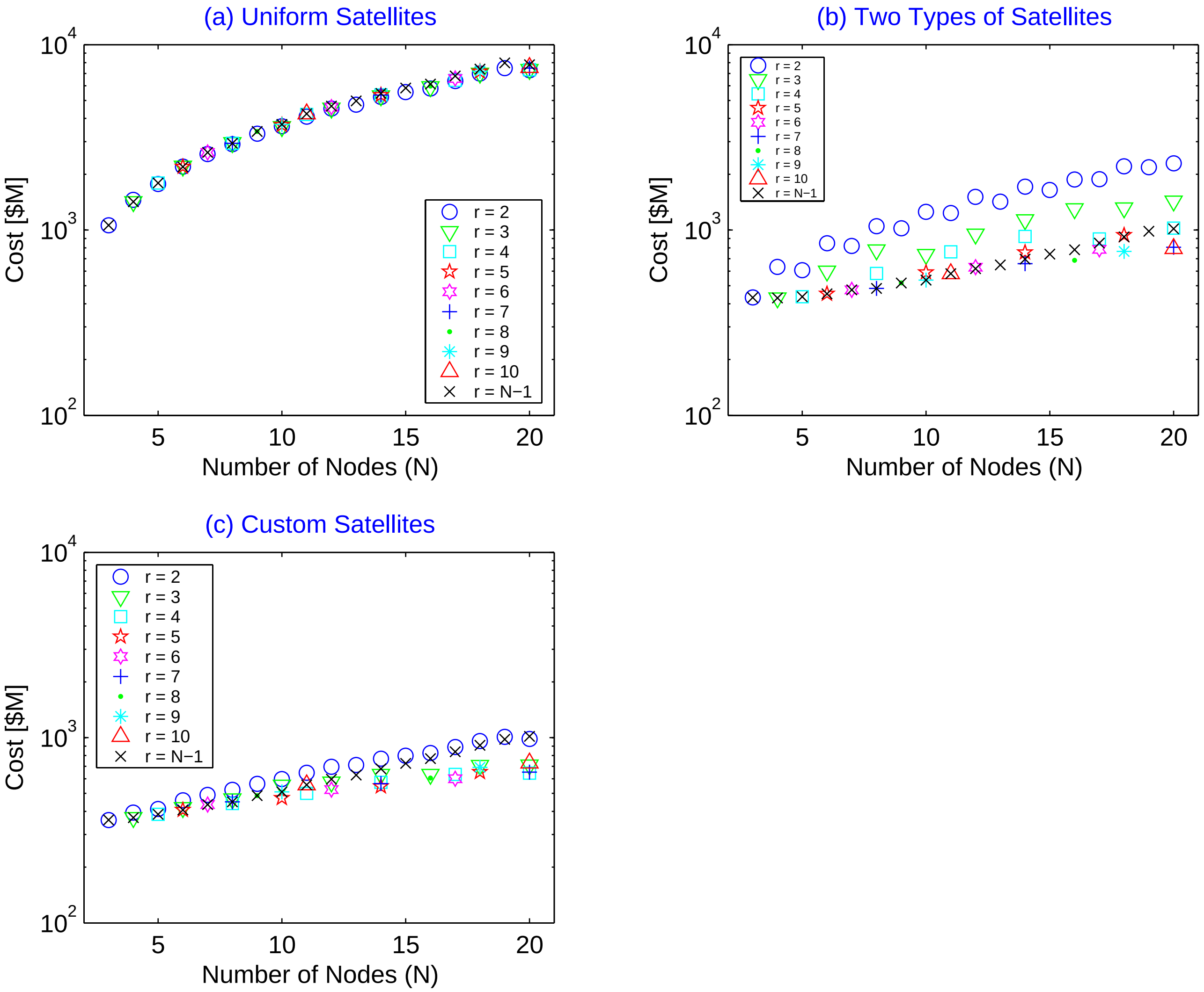


Figure 3-30: Connected circulant constellations with 1 hub node: communications cost MLM for uniform all-to-one traffic with Modified Incremental Dijkstra's algorithm.

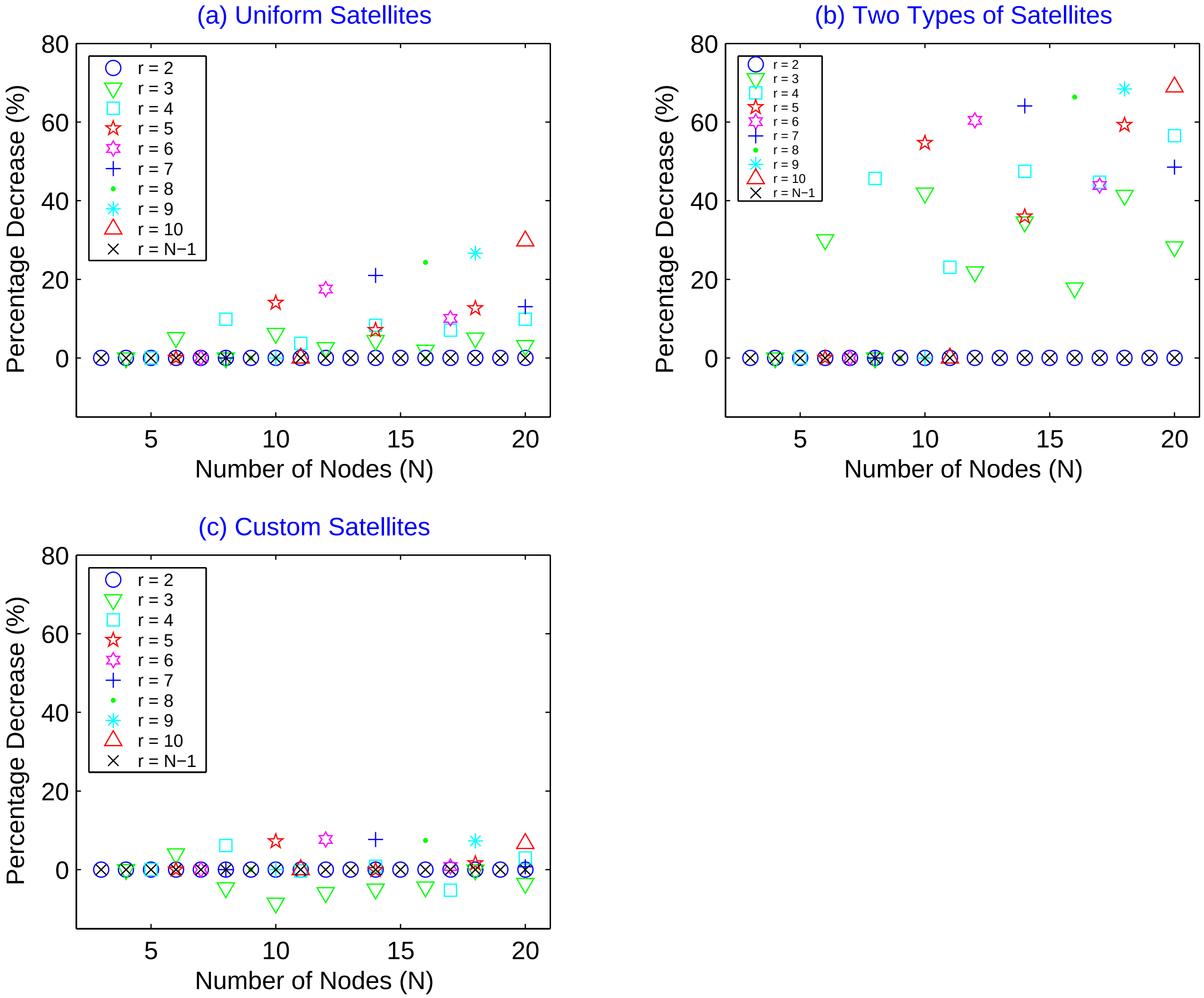


Figure 3-31: Connected circulant constellations with 1 hub node: percentage decrease in communications cost MLM between Dijkstra's algorithm and Modified Incremental Dijkstra's algorithm for uniform all-to-one traffic.

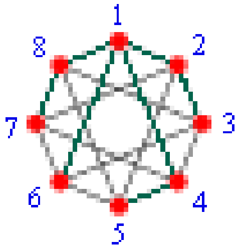
Case V: Symmetric Modified Incremental Dijkstra's algorithm. In an effort to improve both cost and performance in terms of network utilization, a symmetric modified incremental Dijsktra's algorithm is used to balance the traffic by exploiting the property of symmetry in the connected circulant constellation. First, the minimum hop distance for every node to the hub node is determined. Second, from least minimum hop distance to greatest and in topological order, for each source-destination pair, find the shortest path and then load the network with traffic along the shortest path. Symmetry is exploited to balance the traffic on the network constellation. The wavelength dimensioning results for $N = 8$ and $r = 4$ are shown in Figure 3-32. The communications costs for the MLM cost case are shown in Figure 3-33 and the communications costs for the HMH cost case are shown in Figure A-8. The percentage decrease in communications cost between using Dijkstra's algorithm and the Symmetric Modified Incremental Dijkstra's algorithm are shown in Figure 3-34 and in Figure A-9, for the MLM and HMH cost cases respectively.

		Destination j							
		1	2	3	4	5	6	7	8
	1	50	1	0	1	0	1	0	1
	2	100	0	1	0	1	0	1	0
	3	0	50	0	1	0	1	0	1
Source i	4	100	0	1	0	1	0	1	0
	5	0	1	0	50	0	1	0	1
	6	50	0	1	0	1	0	1	0
	7	0	1	0	1	0	1	0	50
	8	100	0	1	0	1	0	1	0

(a)

Min. Hop	Shortest path
-	1 1
1	2 1
2	3 2 1
1	4 1
2	5 4 1
1	6 1
2	7 8 1
1	8 1

(b)



(c)

Figure 3-32: Wavelength dimensioning with Symmetric Modified Incremental Dijkstra's algorithm on the connected circulant constellation with 1 hub node for $N = 8$ and $r = 4$.

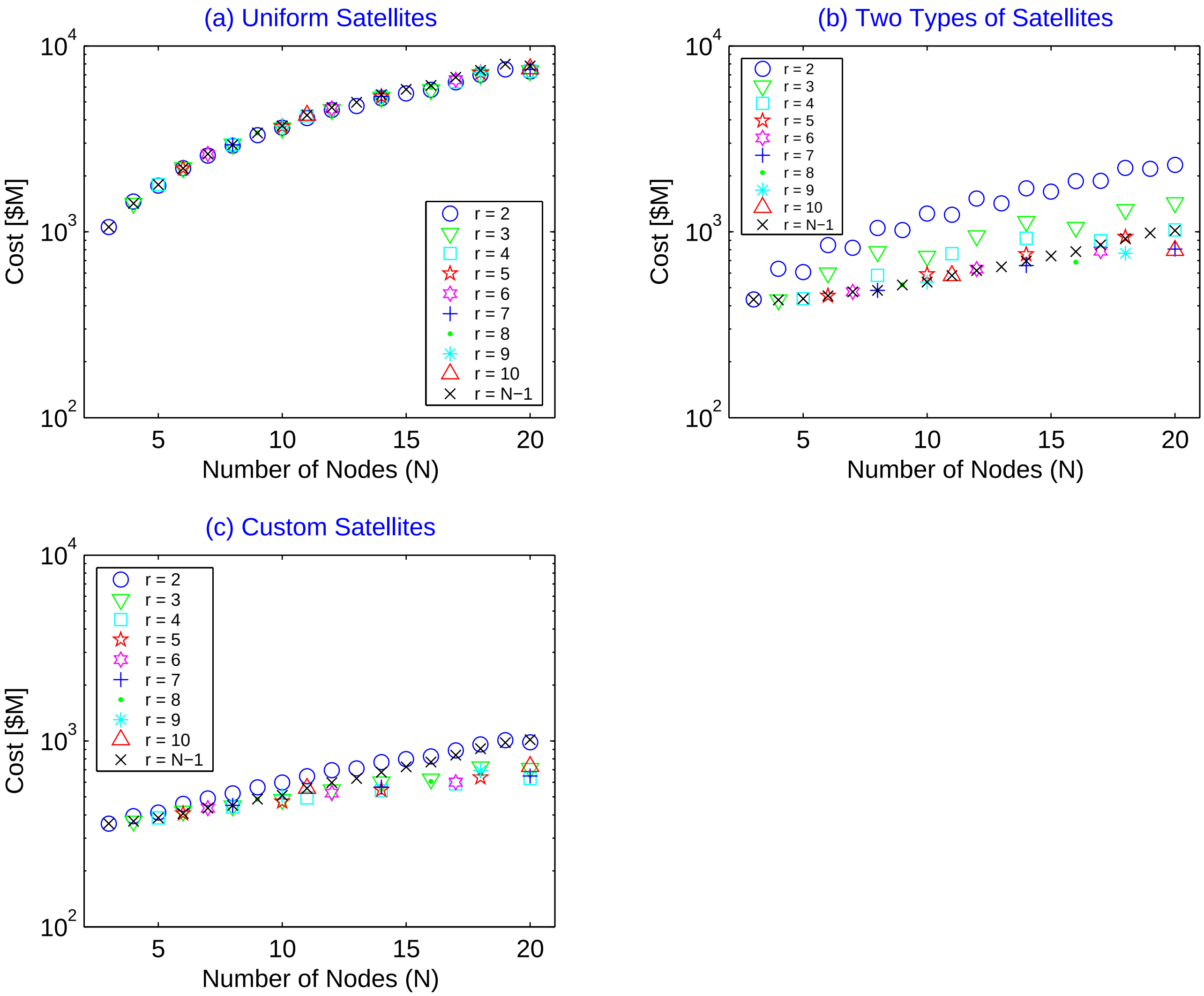


Figure 3-33: Connected circulant constellations with 1 hub node: communications cost MLM for uniform all-to-one traffic with Symmetric Modified Incremental Dijkstra's algorithm.

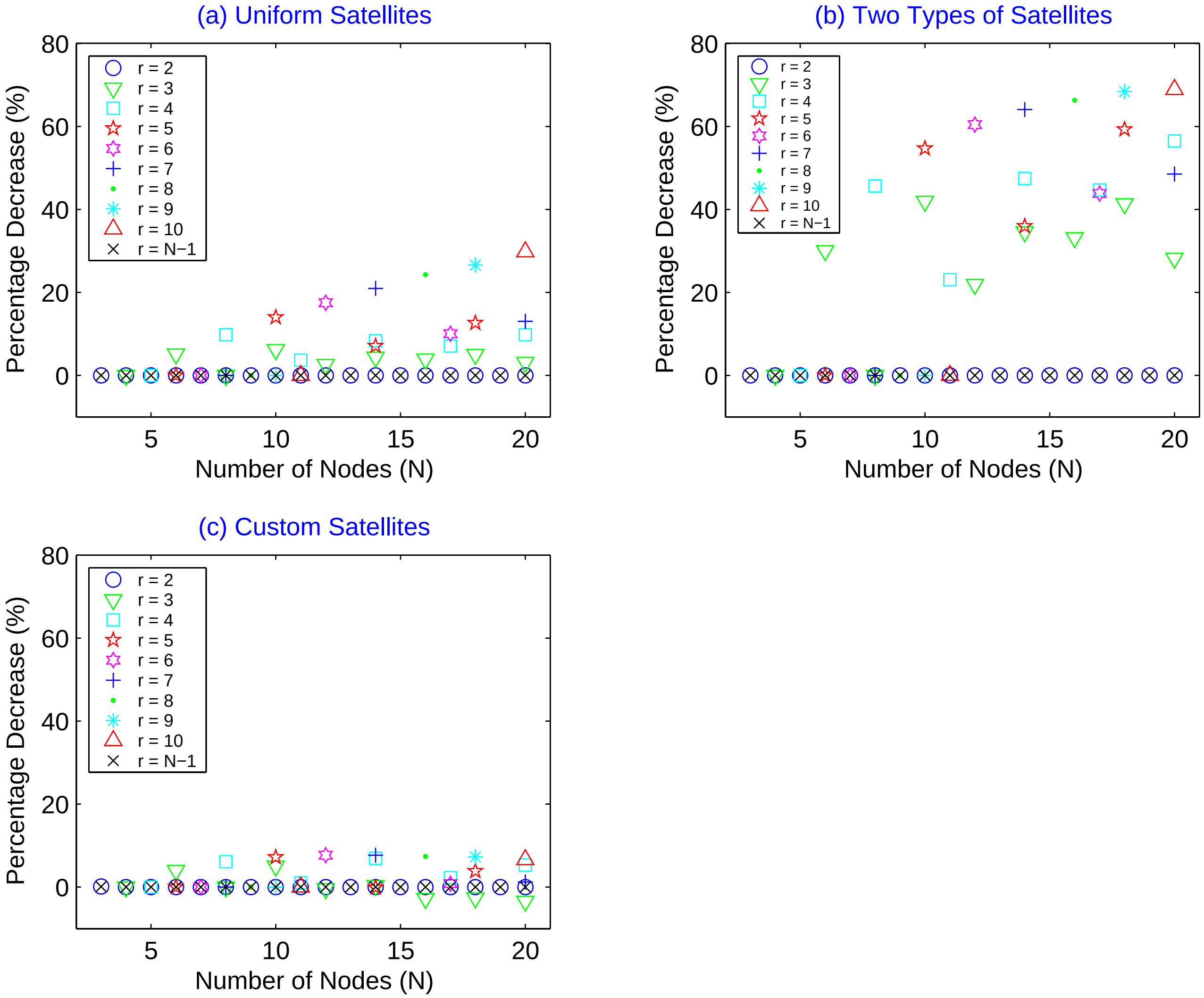


Figure 3-34: Connected circulant constellations with 1 hub node: percentage decrease in communications cost MLM between Dijkstra's algorithm and Symmetric Modified Incremental Dijkstra's algorithm for uniform all-to-one traffic.

As the costs between the Modified Incremental Dijkstra's algorithm and the Symmetric Modified Incremental Dijkstra's algorithm seem nearly identical, to determine which algorithm is actually more cost-effective, the percentage decrease in communications cost between the Modified Incremental Dijkstra's algorithm and the Symmetric Modified Incremental Dijkstra's algorithm is shown in Figure 3-35 for the MLM cost case and in Figure A-10 for the HMH cost case. Cost savings can be easily seen for constellations of degree 3 and degree 4 for custom-built satellites using the Symmetric Modified Incremental Dijkstra's algorithm.

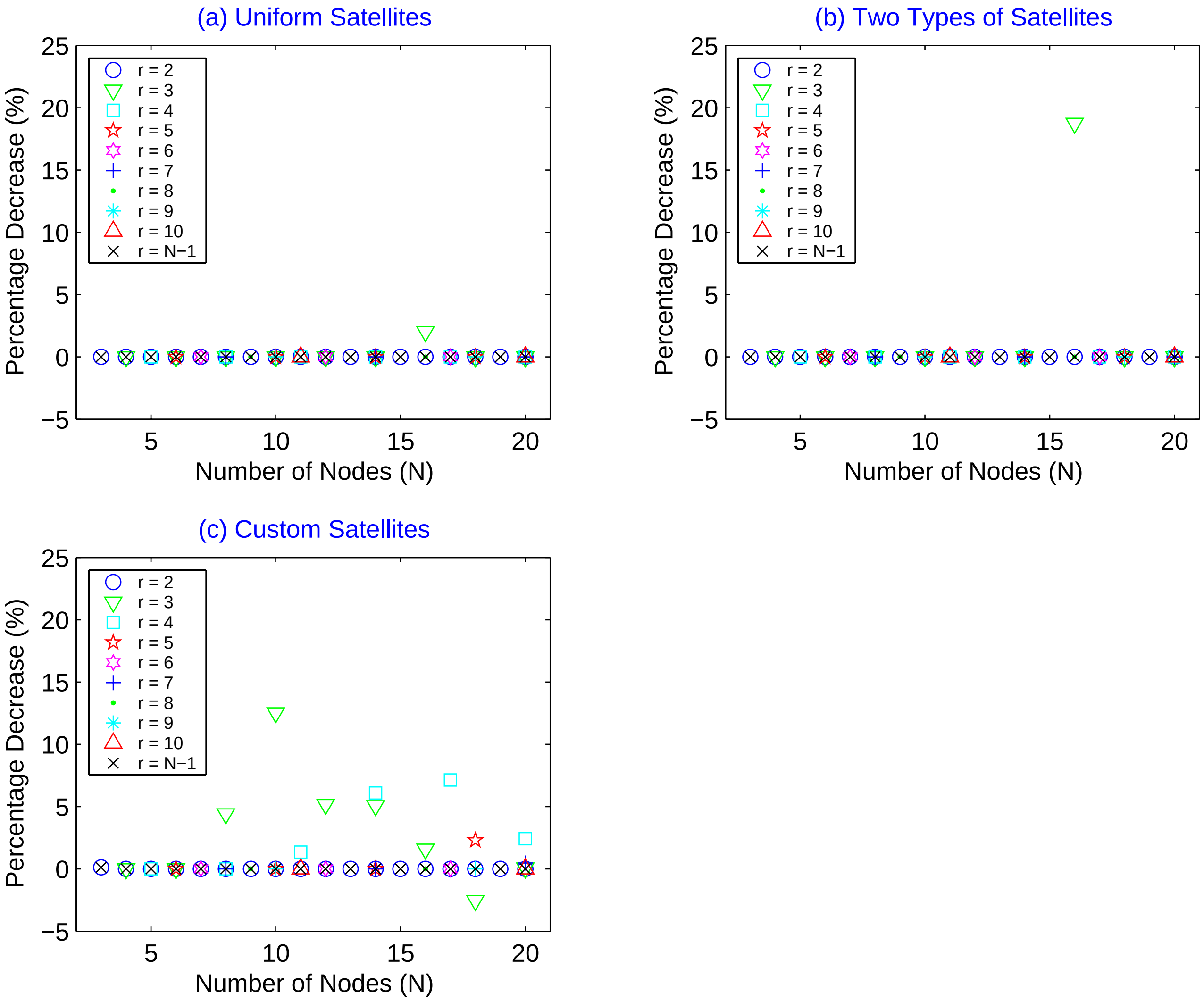


Figure 3-35: Connected circulant constellations with 1 hub node: percentage decrease in communications cost MLM between Modified Incremental Dijkstra's algorithm and Symmetric Modified Incremental Dijkstra's algorithm for uniform all-to-one traffic.

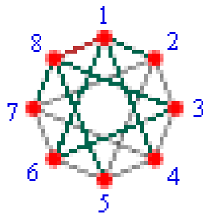
Lastly, the communication costs for uniform all-to-one traffic on connected circulant constellations with 2 hub nodes are briefly considered. The Symmetric Modified Incremental Dijkstra's algorithm is implemented for traffic routing. With 2 satellite nodes acting as the hub nodes, half of the remaining satellite nodes in the backbone constellation transmit to one of the hubs while the remaining half transmits to the other hub. The wavelength dimensioning results for $N = 8$ and $r = 4$ are shown in Figure 3-36, where nodes 1 and 8 are designated as hub nodes. Communications cost results are shown for the MLM cost case in Figure 3-37 and communications cost results for the HMH cost case are shown in Figure A-11. The percentage decrease in communications cost between using 1 hub node and and 2 hub nodes in a connected circulant constellation is shown in Figure 3-38 for the MLM cost case and in Figure A-12 for the HMH cost case. These figures show that there is a large variation in the cost difference between the 1-hub node and 2-hub node designs. Generally, connected circulant constellations with 2 hub nodes are more costly due to the built-in redundancy of having extra wavelengths on the crosslink between the two hub nodes. In the remaining architectural studies, connected circulant constellations with 2 hub nodes are not given any further consideration.

		Destination j							
		1	2	3	4	5	6	7	8
Source i	1	50	1	0	1	0	1	0	200
	2	50	0	1	0	1	0	1	0
	3	0	1	0	1	0	50	0	50
	4	50	0	1	0	1	0	1	0
	5	0	1	0	1	0	1	0	50
	6	50	0	50	0	1	0	1	0
	7	0	1	0	1	0	1	0	50
	8	200	0	1	0	1	0	1	0

(a)

Min. Hop	Shortest path
-	1 1
1	2 1
2	3 6 1
1	4 1
2	5 8
1	6 3 8
2	7 8
1	8 8

(b)



(c)

Figure 3-36: Wavelength dimensioning with Symmetric Modified Incremental Dijkstra's algorithm on the connected circulant constellation with uniform jump spaces and 2 hub nodes for $N = 8$ and $r = 4$.

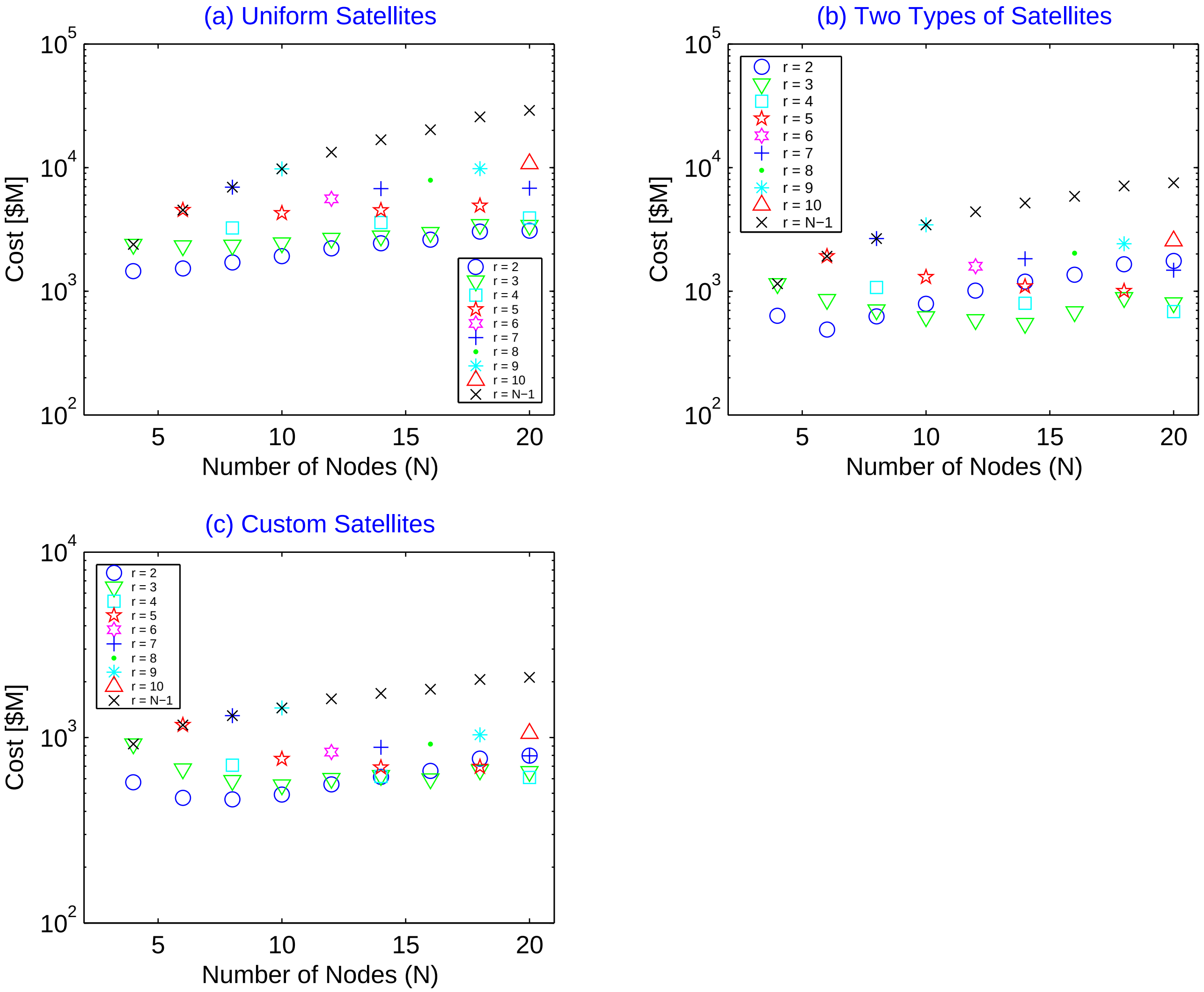


Figure 3-37: Connected circulant constellations with 2 hub nodes: communications cost MLM with Symmetric Modified Incremental Dijkstra's algorithm for uniform all-to-one traffic.

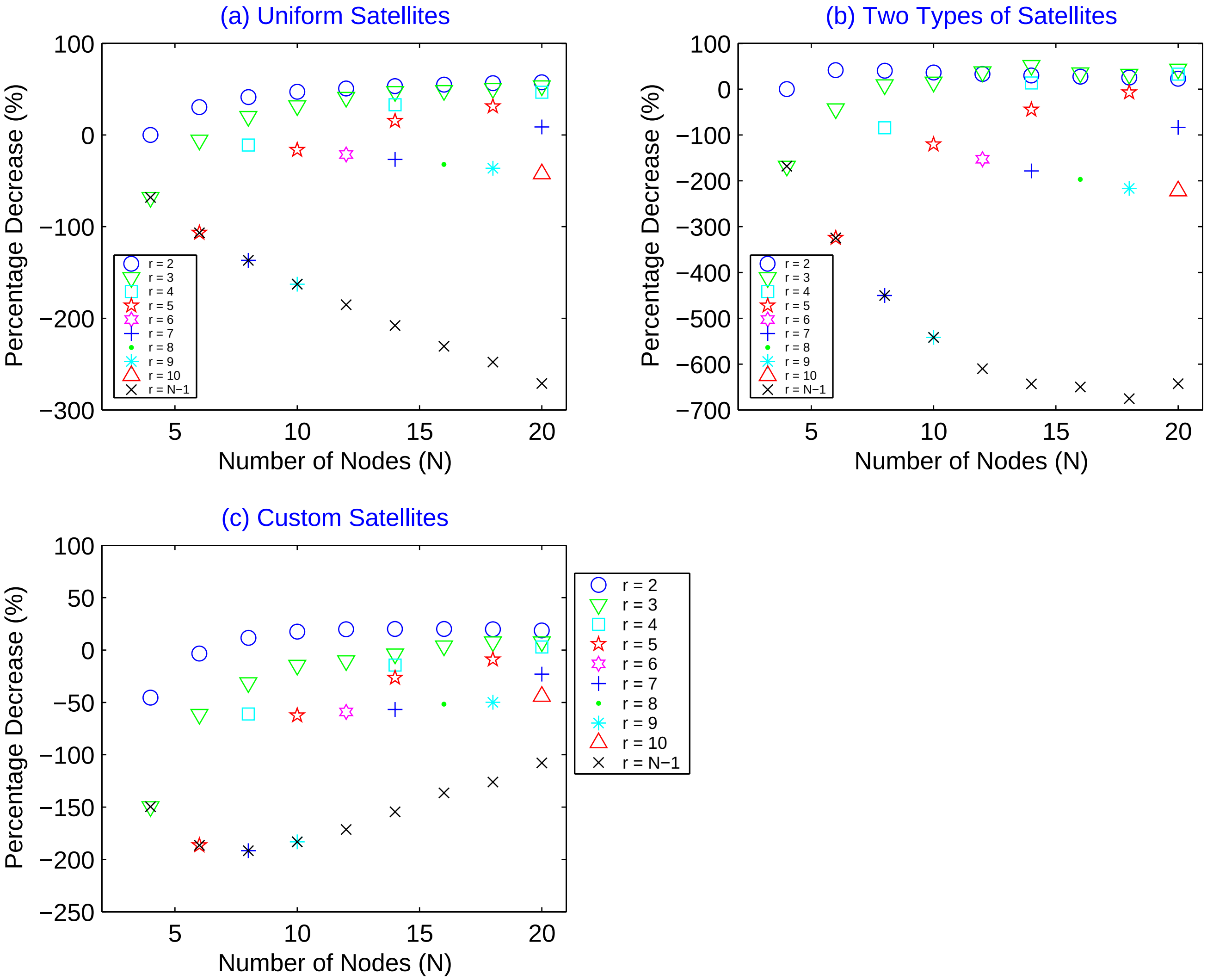


Figure 3-38: Connected circulant constellations: percentage decrease in communications cost MLM with Symmetric Modified Incremental Dijkstra's algorithm for uniform all-to-one traffic between constellations with 1 hub node and constellations with 2 hub nodes.

3.1.5.6 Mixed Traffic

In this section, the communications cost of mixed traffic on space-based network backbone constellations are evaluated. Traffic is separated into two classes: uniform all-to-all and uniform all-to-one. The amount of uniform all-to-one traffic for source-hub node pairs is denoted by ρ, where $0 \leq \rho \leq 1$. The remaining traffic, $1 - \rho$, is uniformly distributed between all remaining source-destination node pairs.

Hub constellations. Figures 3-39, 3-40, and 3-41 show results for mixed traffic on hub constellations for the MLM cost case. Figures A-13, A-14, and A-15 show results for mixed traffic on hub constellations for the HMH cost case. In each of the figures, the amount of ρ is given as a percentage of the total traffic volume and is labeled as Hub Traffic.

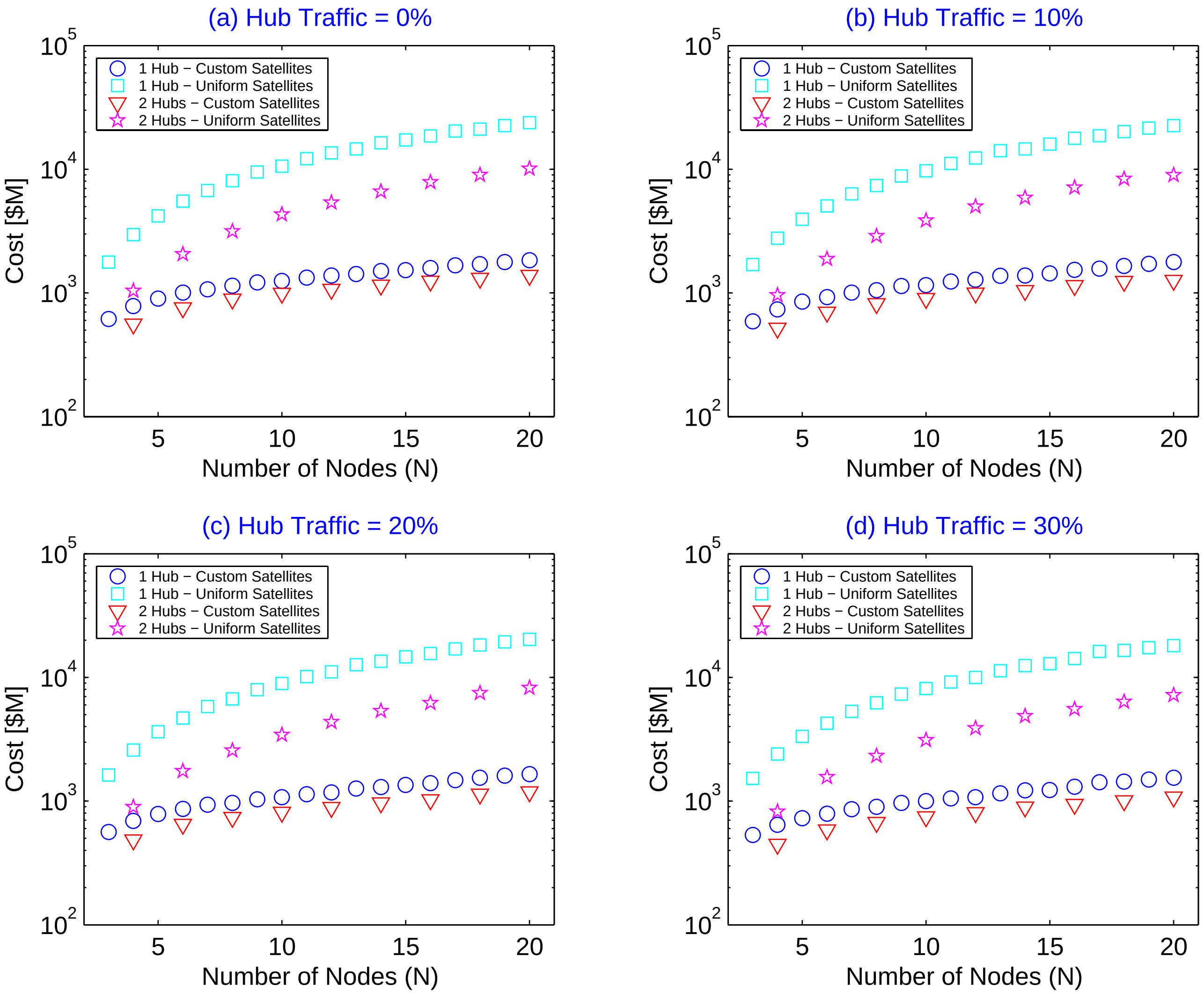


Figure 3-39: Hub constellations: communications cost MLM for mixed traffic I.

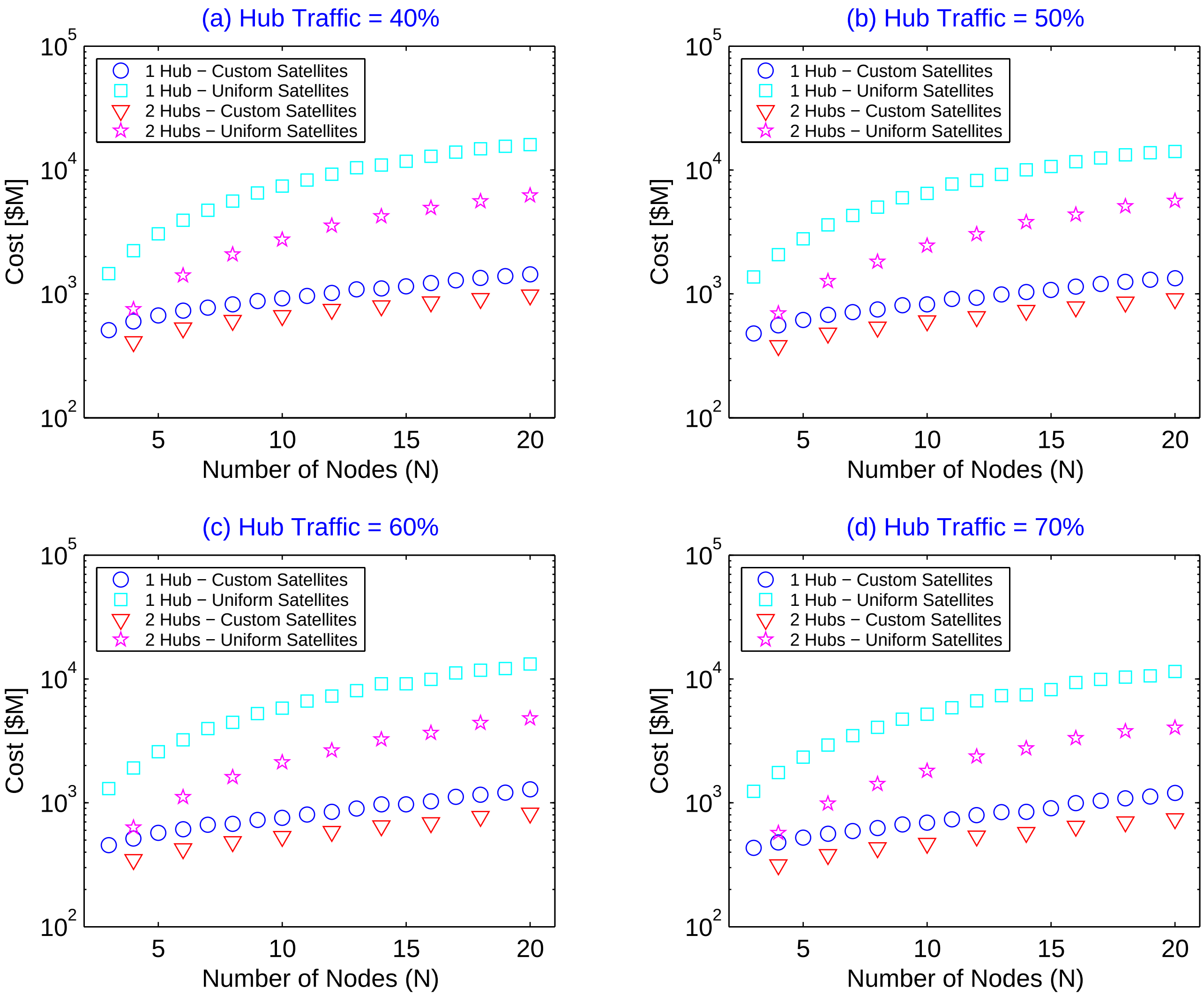


Figure 3-40: Hub constellations: communications cost MLM for mixed traffic II.

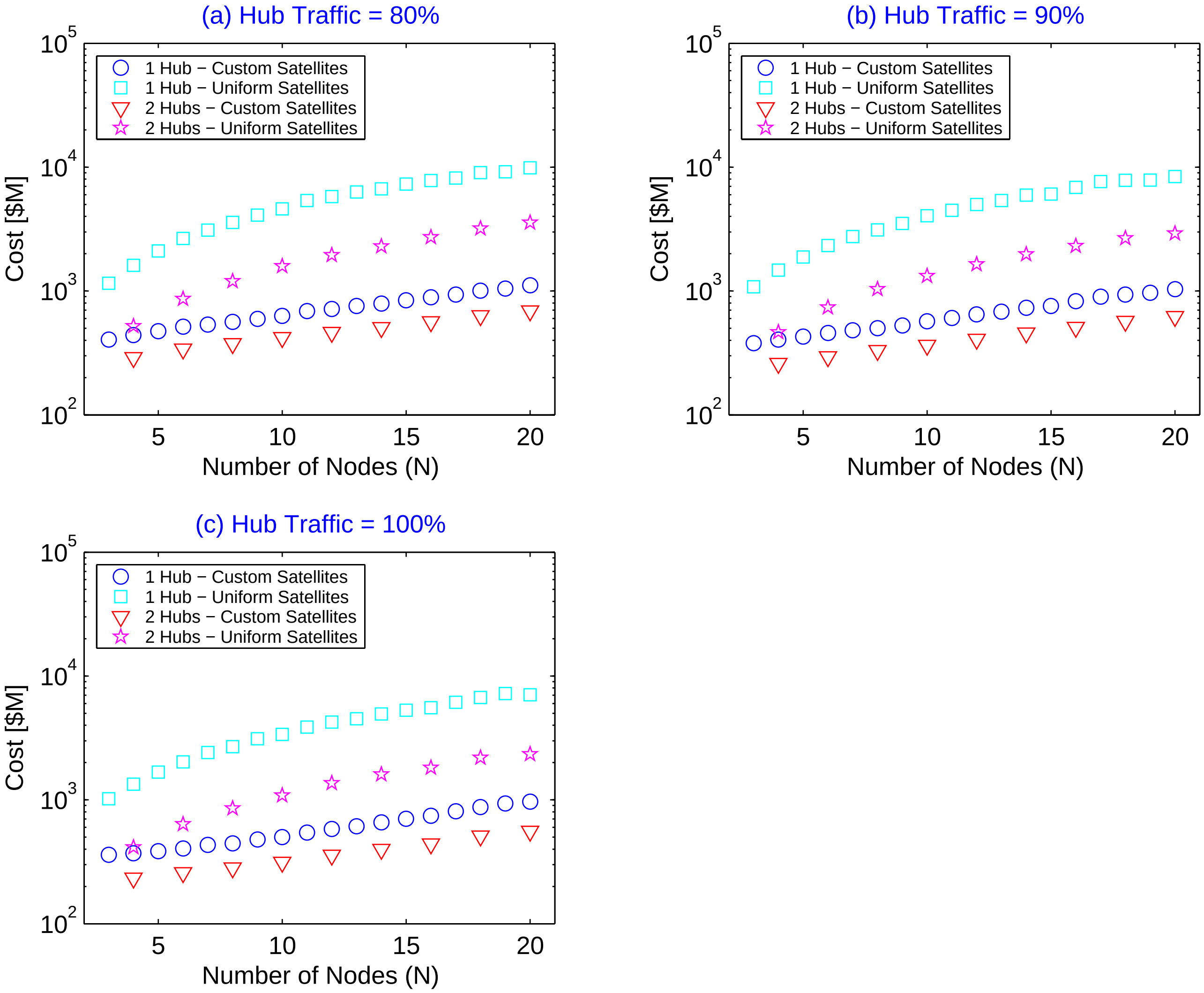


Figure 3-41: Hub constellations: communications cost MLM for mixed traffic III.

Connected circulant constellations. The effect of mixed traffic on the communication costs for connected circulant constellations with 1 hub node is studied. The amount of uniform all-to-one traffic is again represented by ρ where $0 \leq \rho \leq 1$ and denoted as Hub Traffic. The routing scheme chosen for the remaining uniform all-to-all traffic, $1 - \rho$, is the Symmetric Modified Incremental Dijkstra's algorithm. Figures 3-42, 3-43, and 3-44 show the results for the MLM cost case using uniform satellites. Figures 3-45, 3-46, and 3-47 show the results for the MLM cost case using 2 types of satellites. Figures 3-48, 3-49, and 3-50 show the results for the MLM case case using custom-built satellites. A similar set of results for the HMH cost case is shown in Appendix A.1.2.1.

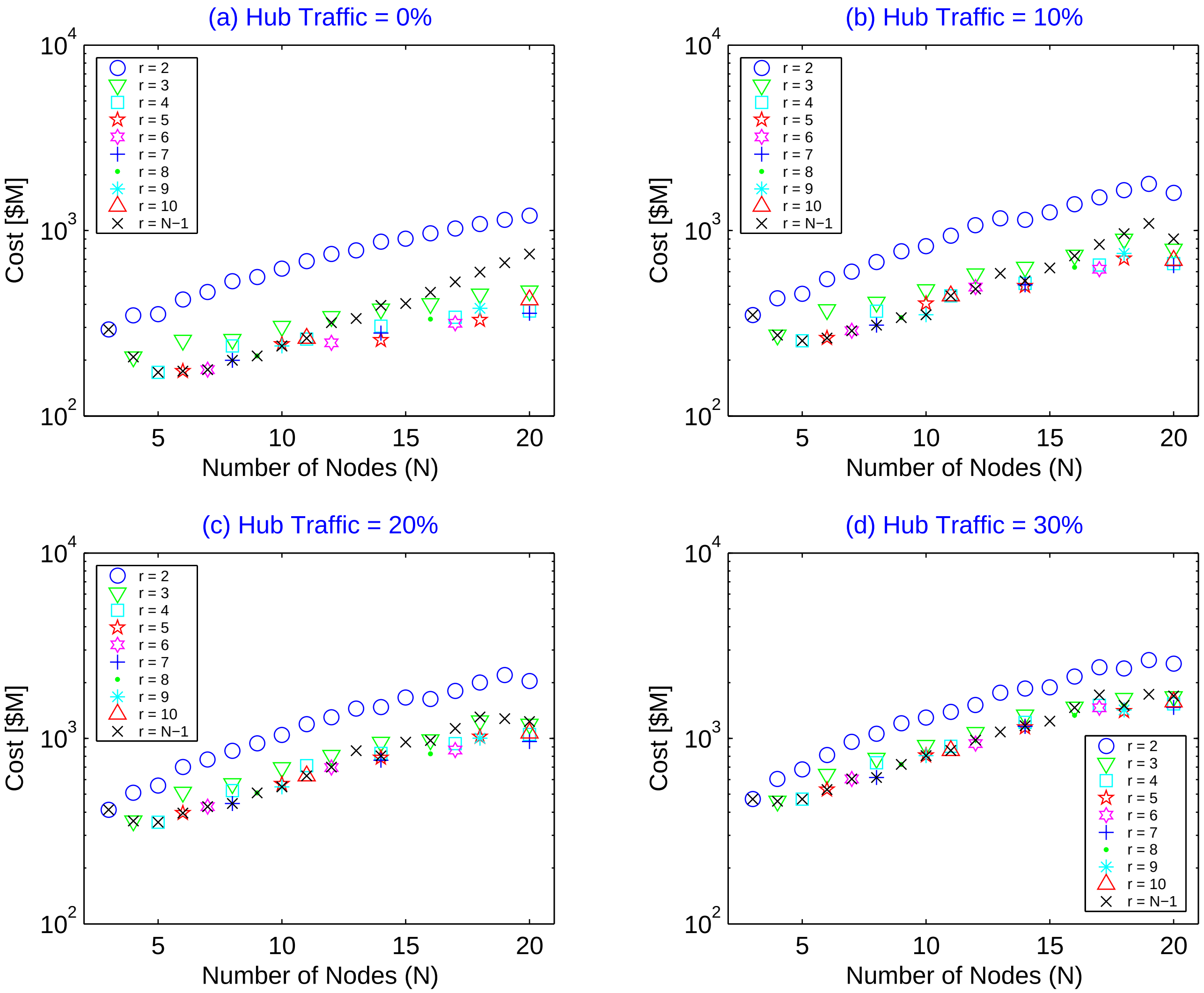


Figure 3-42: Connected circulant constellations with 1 hub node: communications cost MLM with uniform satellites for mixed traffic I.

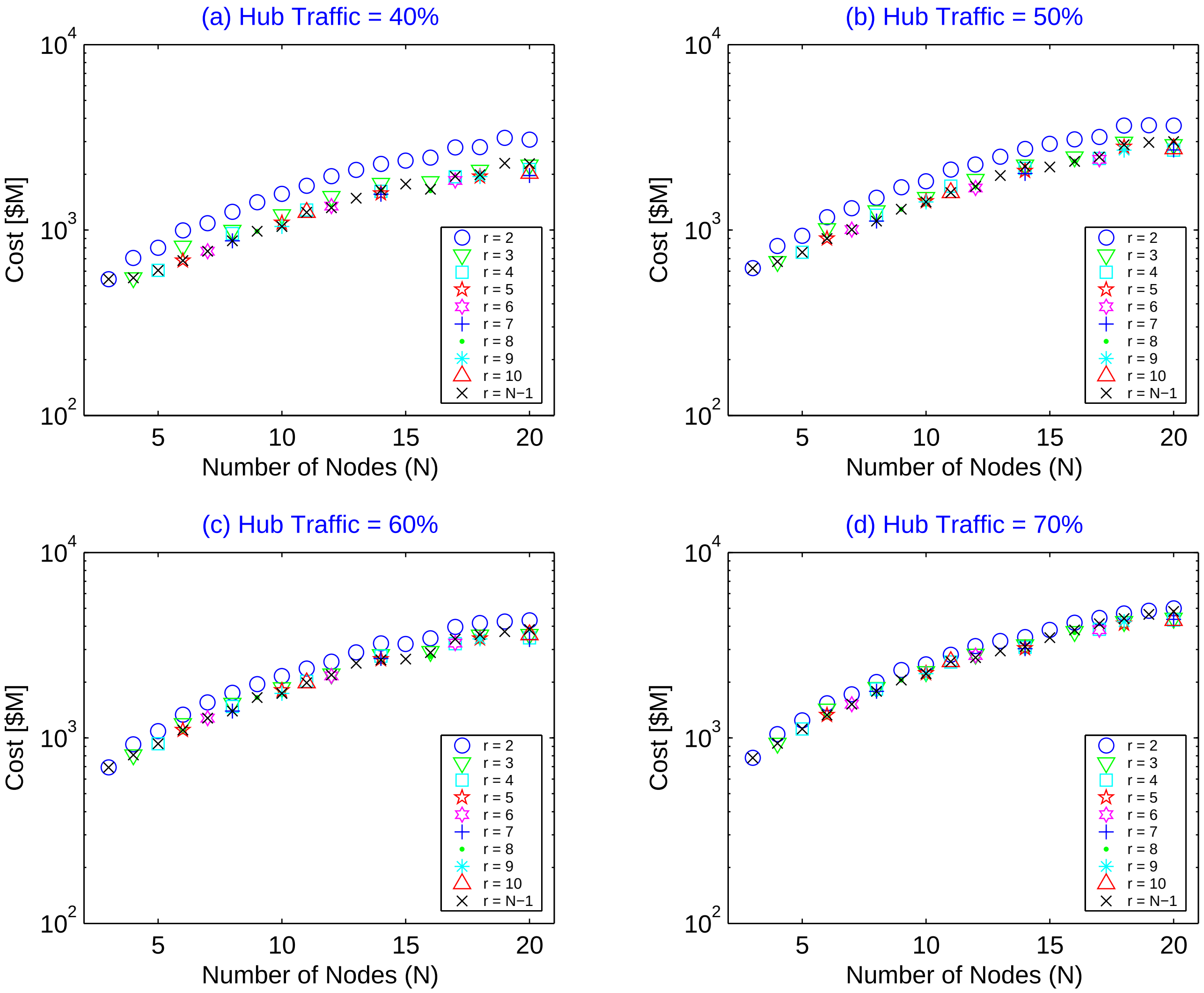


Figure 3-43: Connected circulant constellations with 1 hub node: communications cost MLM with uniform satellites for mixed traffic II.

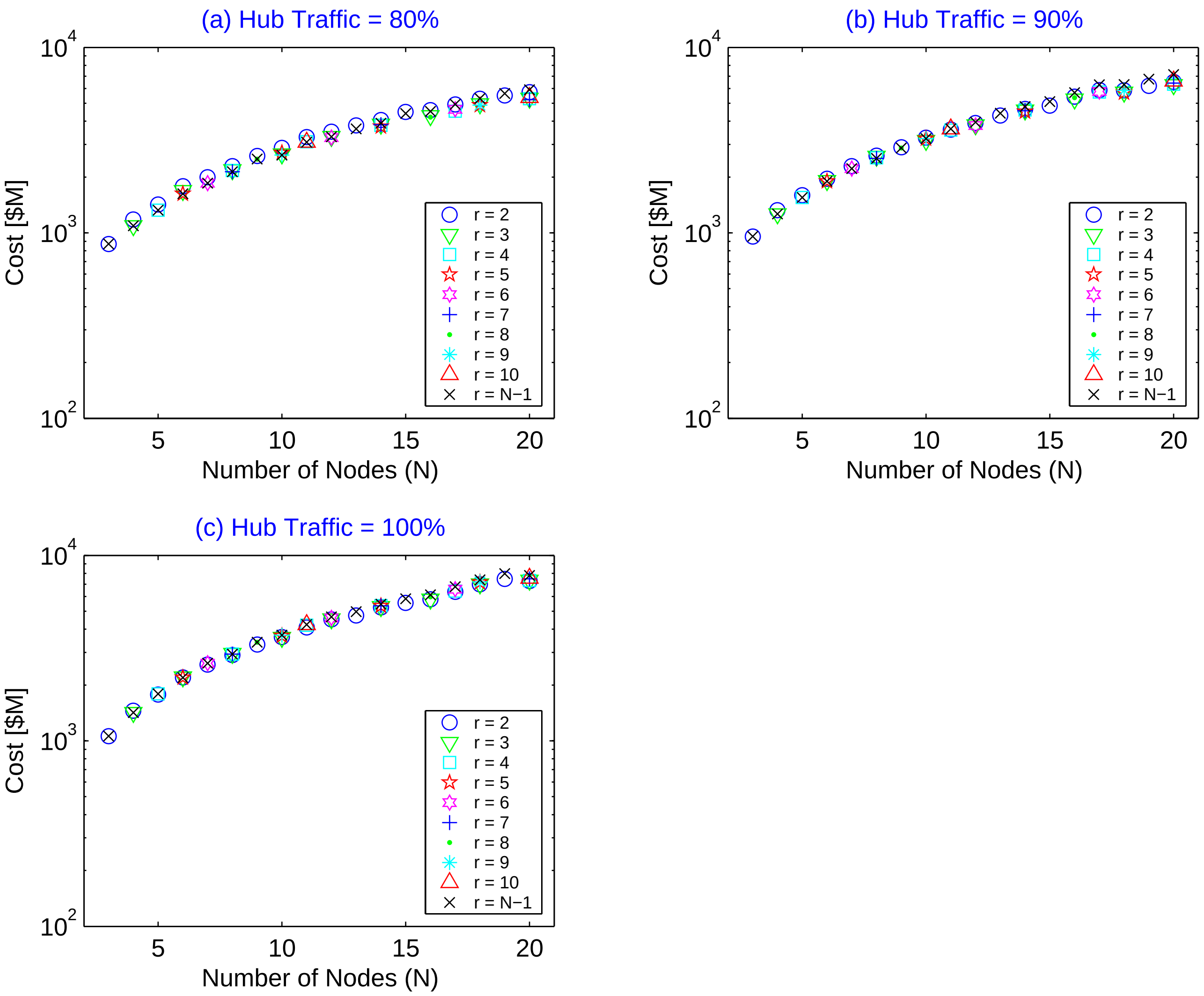


Figure 3-44: Connected circulant constellations with 1 hub node: communications cost MLM with uniform satellites for mixed traffic III.

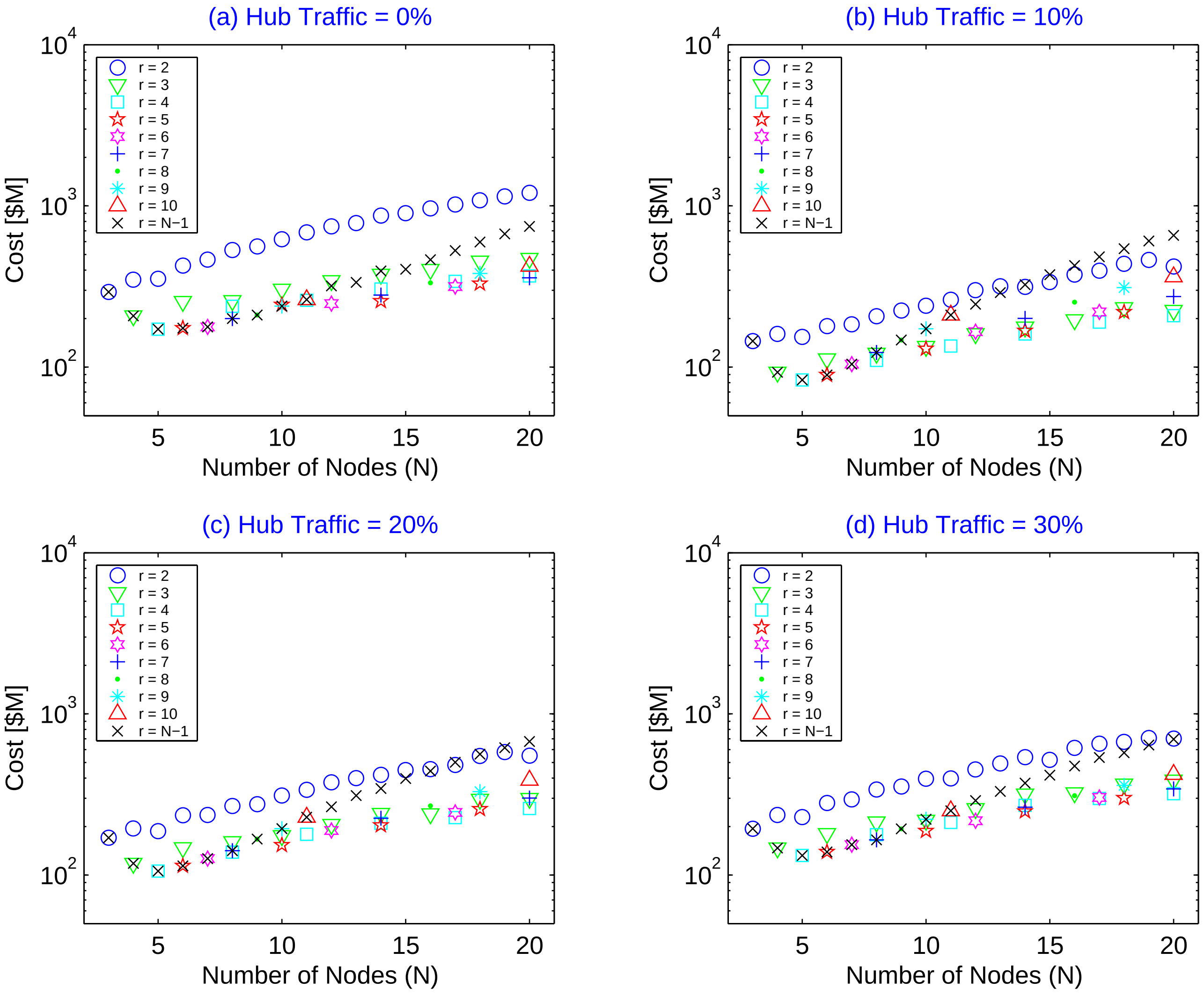


Figure 3-45: Connected circulant constellations with 1 hub node: communications cost MLM with 2 types of satellites for mixed traffic I.

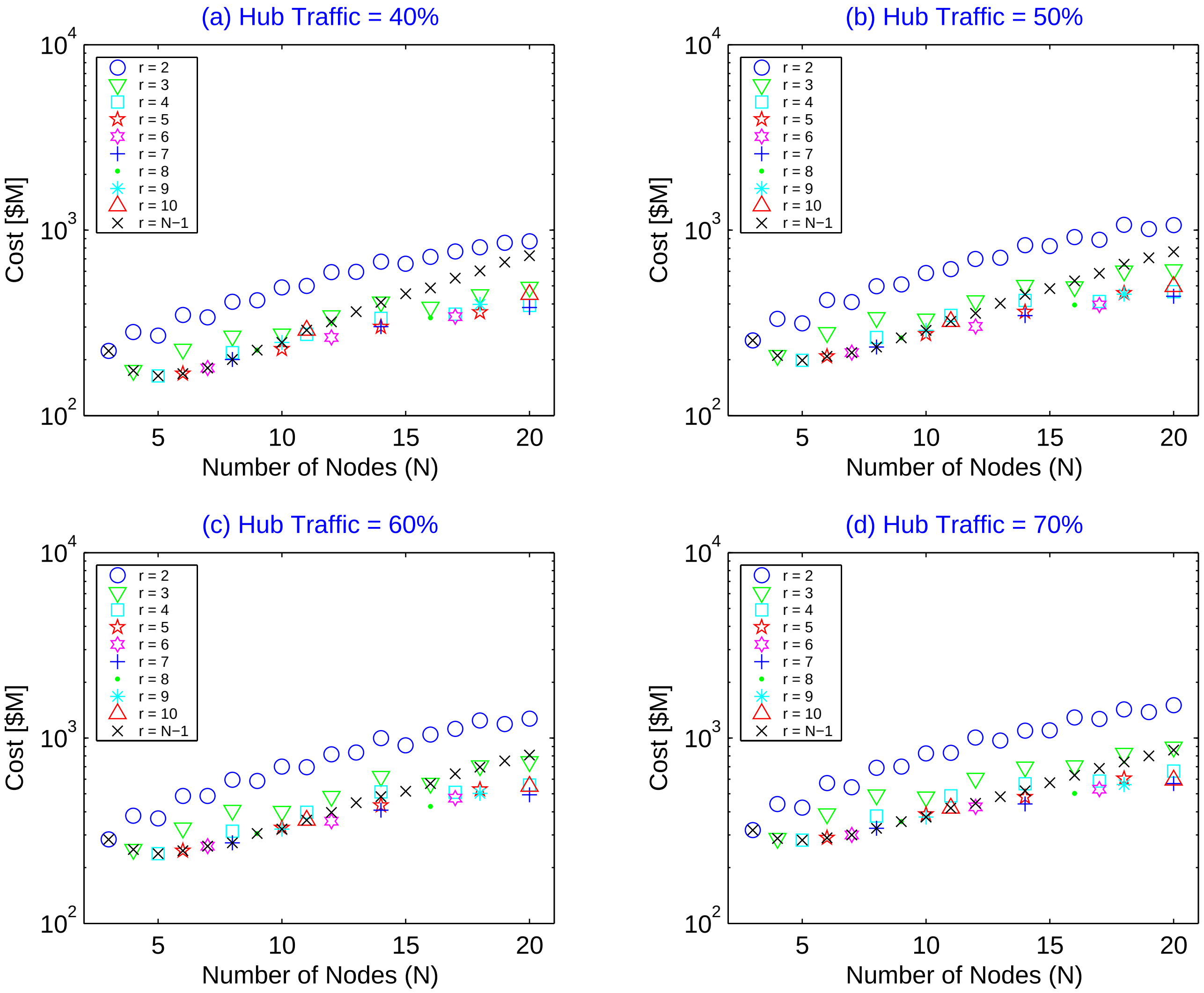


Figure 3-46: Connected circulant constellations with 1 hub node: communications cost MLM with 2 types of satellites for mixed traffic II.

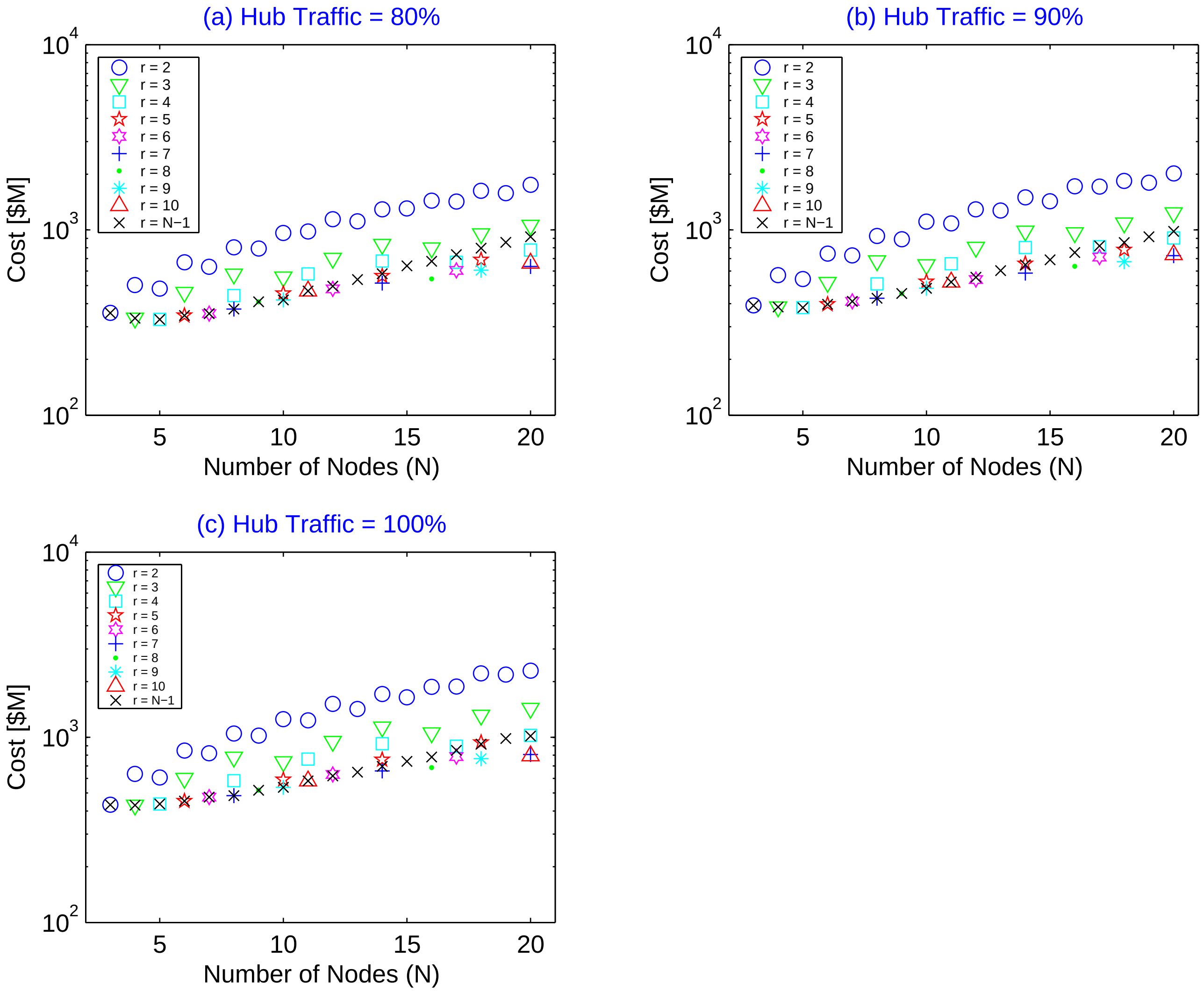
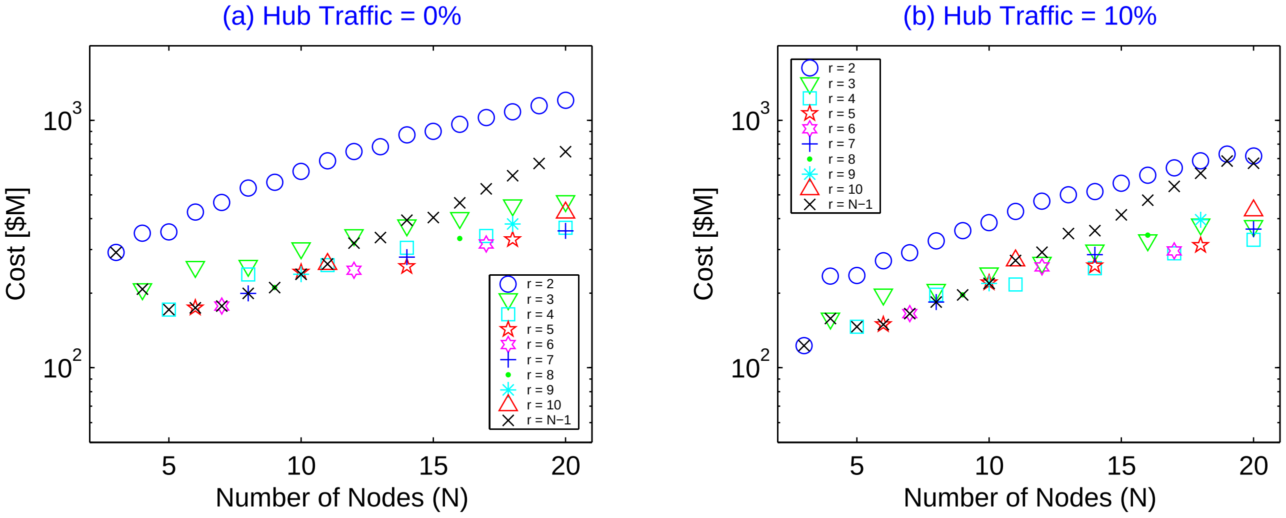
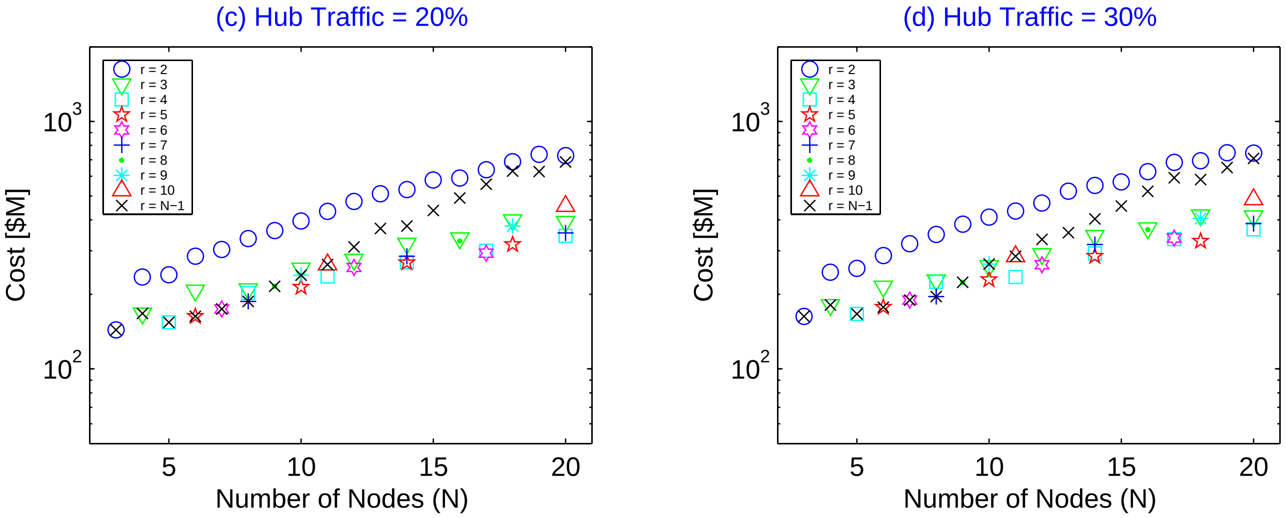


Figure 3-47: Connected circulant constellations with 1 hub node: communications cost MLM with 2 types of satellites for mixed traffic III.


(a) Hub Traffic = 0%

(b) Hub Traffic = 10%


(c) Hub Traffic = 20%

(d) Hub Traffic = 30%

Figure 3-48: Connected circulant constellations with 1 hub node: communications cost MLM with custom satellites for mixed traffic I.

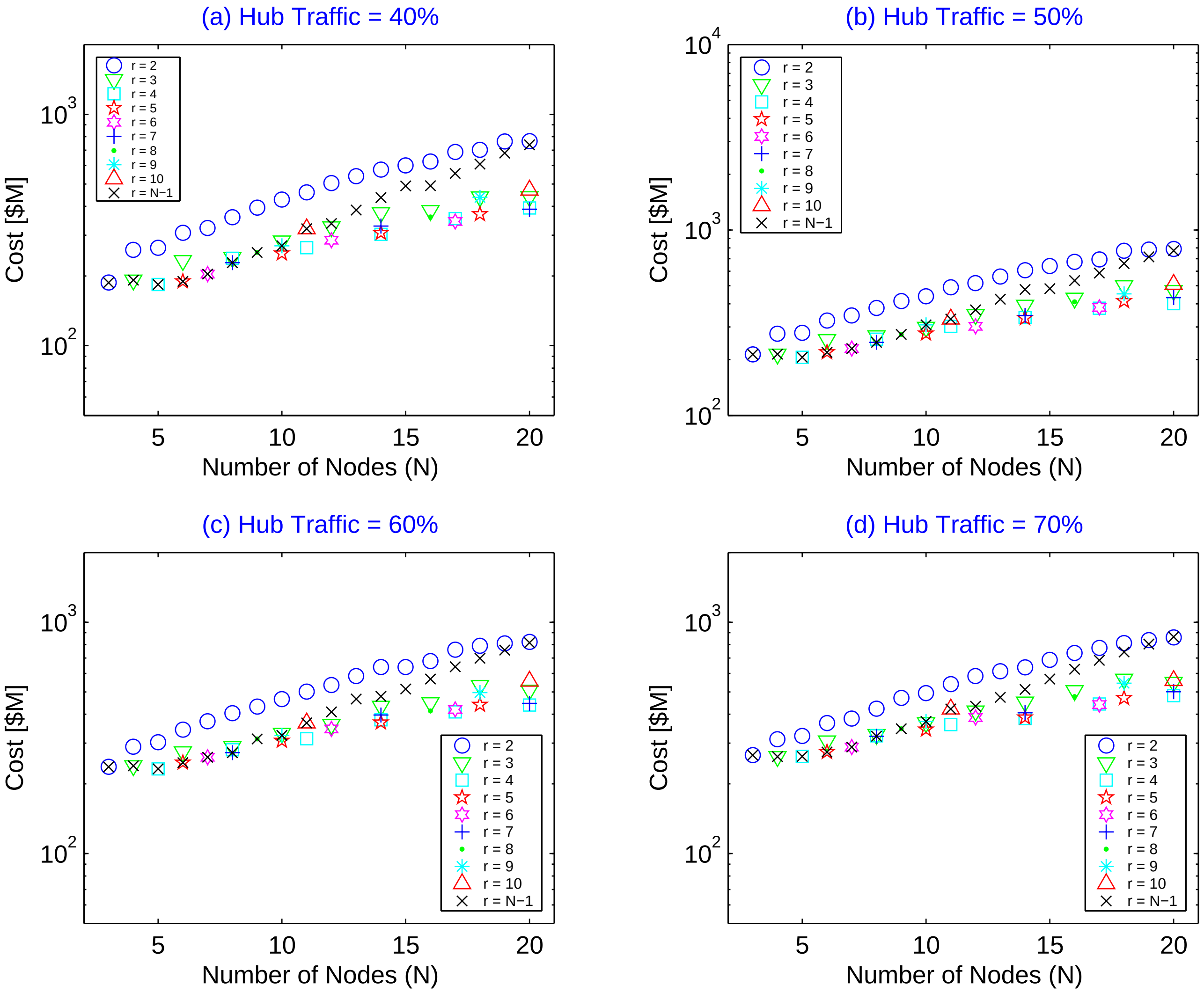


Figure 3-49: Connected circulant constellations with 1 hub node: communications cost MLM with custom satellites for mixed traffic II.

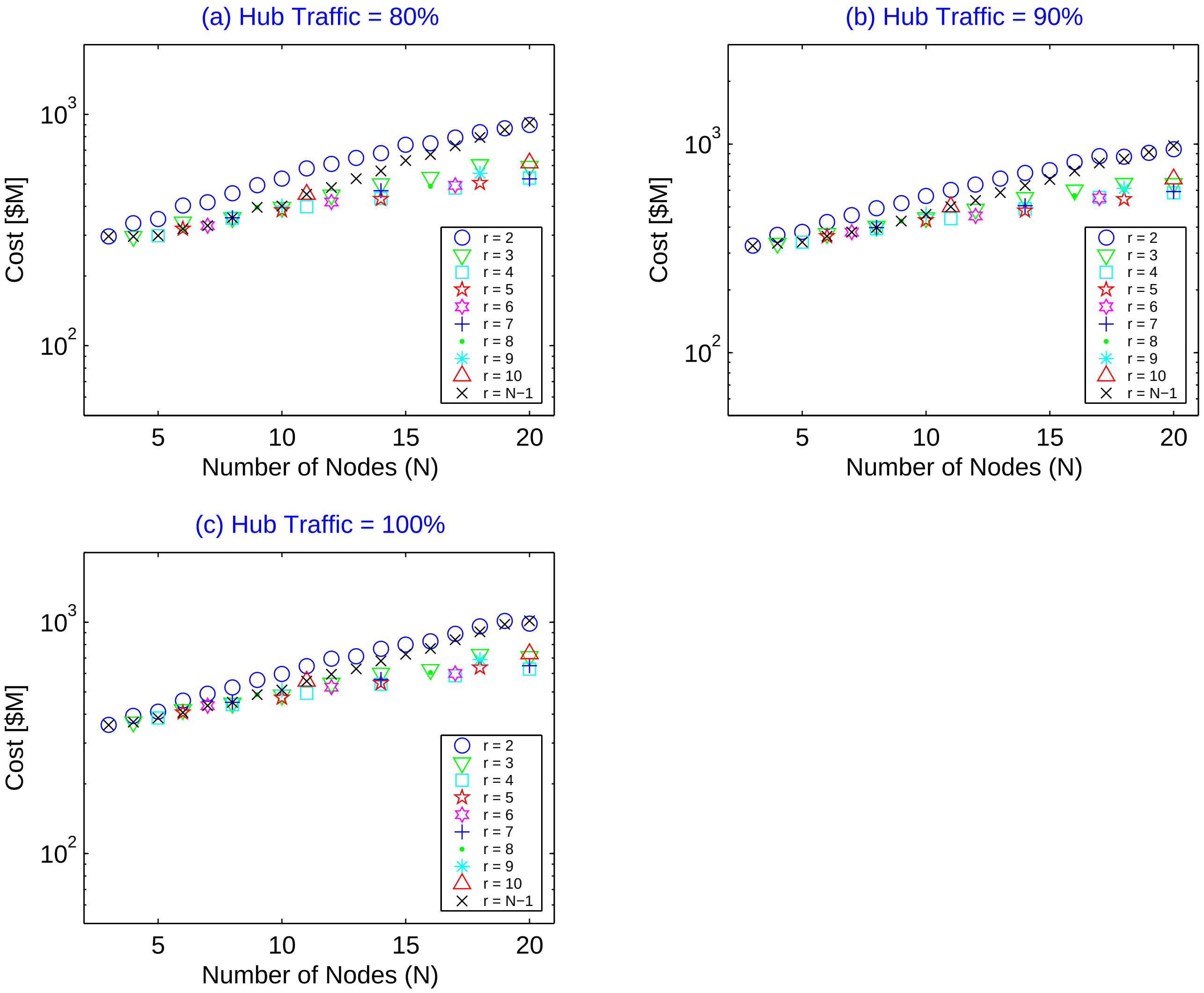


Figure 3-50: Connected circulant constellations with 1 hub node: communications cost MLM with custom satellites for mixed traffic III.

It is observed that the communications cost for mixed traffic on hub constellations are nearly insensitive to the different levels of mixed traffic. However, the communication cost of mixed traffic on connected circulant constellations substantially increases as the amount of uniform all-to-one traffic increases. Additionally, the connected circulant networks are under-utilized because not every intersatellite link has traffic. The communications cost is highest when the constellations are comprised of uniform satellites, i.e., each satellite has the ability to function as the central (or hub) node even if it does not handle as much traffic volume.

It can be shown that uniform all-to-one traffic is more cost-effective on hub constellations than on connected circulant constellations. The percentage increase in communications cost between uniform all-to-one traffic and uniform all-to-all traffic on the hub constellations for the MLM cost case and the HMH cost case are shown in Figure 3-51. Likewise, uniform all-to-all traffic is more cost-effective on connected circulant constellations than on hub constellations. The percentage increase in communications cost between uniform all-to-all traffic and uniform all-to-one traffic on the connected circulant constellations for the MLM cost scenario and the HMH cost scenario are shown in Figure 3-52. The comparisons are conducted for the case of building uniform satellites for the space-based network backbone constellation.

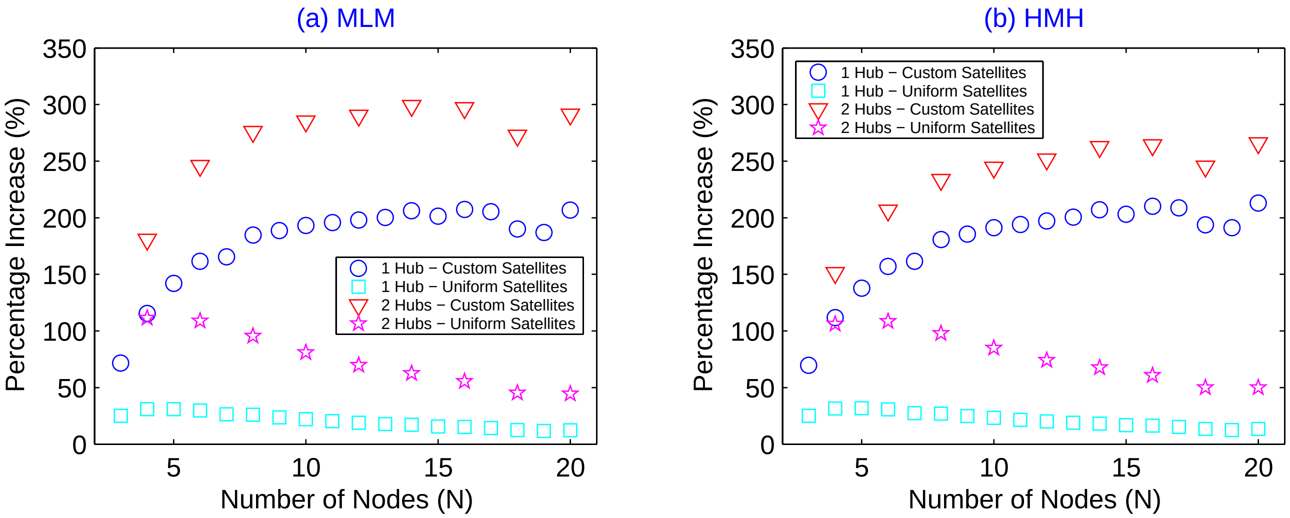


Figure 3-51: Percentage increase in communications costs for hub constellations between uniform all-to-one traffic and uniform all-to-all traffic.

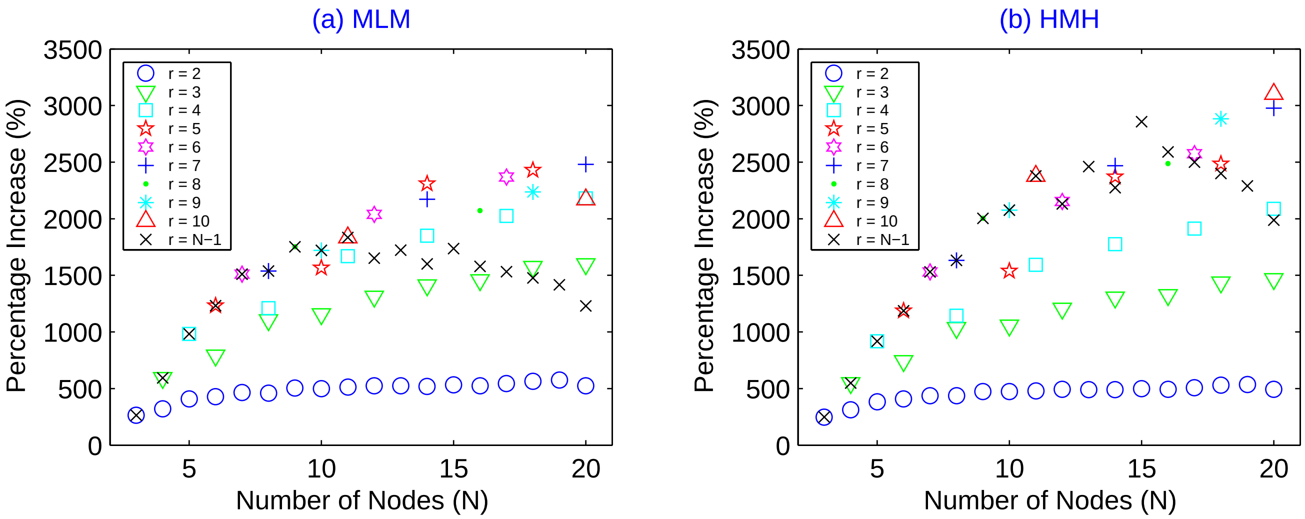


Figure 3-52: Percentage increase in communications costs for connected circulant constellations between uniform all-to-all traffic and uniform all-to-one traffic.

3.1.5.7 Two Disjoint Communities of Users with Mixed Traffic

Although the constellation tradespace includes node sizes of $3 \leq N \leq 20$, there is greater interest in studying smaller constellations of $4 \leq N \leq 6$. Generally, these constellation sizes have provided the lowest communication costs. The communications cost for both types of constellations of sizes $N = 4$, 5, and 6 under mixed traffic are shown in Figure 3-53 and Figure A-25 for the MLM cost case and the HMH cost case, respectively. The figure legend has been simplified: 1-H CSat refers to 1 Hub - Custom Satellites, 1-H USat refers to 1 Hub - Uniform Satellites, 2-H CSat refers to 2 Hub - Custom Satellites, and 2-H USat refers to 2 Hub - Uniform Satellites. At low levels of uniform all-to-one traffic ($\rho < 0.3$), connected circulant constellations with uniform jump spaces provide lower communications costs. At high levels of uniform all-to-one traffic ($\rho > 0.3$), hub constellations provide lower communications costs. This threshold seems to hold for all 27 permutations of cost values for the communications components (antennas, switches, and links). Results are shown in Appendix A.2.

In this section, two disjoint communities of users for small constellations are studied. Likely regions of interest occur when user group 1 has mostly uniform all-to-all traffic ($0.8 \leq \rho \leq 1.0$) while user group 2 has mostly uniform all-to-one traffic ($0 \leq \rho \leq 0.2$). The question of interest to examine then is whether it is more cost-effective to build one system to satisfy the traffic demands of both user communities or build two separate systems, each

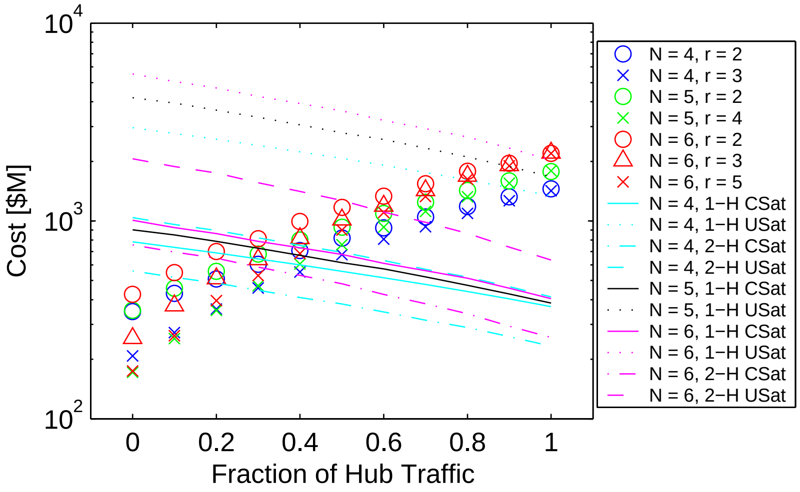


Figure 3-53: Comparing communications costs MLM of N=4,5,6 for a range of mixed traffic.

tailored to a specific user group. Two cases are considered: (1) when both user communities have equal traffic volume and (2) when one user community has a comparatively small traffic volume (approximately 10%) to that of the other user community.

Equal Traffic Volumes. In this situation, each user community has an equivalent amount of total traffic (i.e., $T = 400$ wavelengths). Using Figure 3-53, the constellation types of lowest communication cost for the various amounts of mixed traffic is re-plotted in Figure 3-54 for the MLM cost case. The total communication costs of building two separate systems for the various amounts of mixed traffic are shown in Figure 3-55. The contours lines separate the regions of different constellation types. In each region where two constellation systems are deployed, C_i indicates the constellation that is built for user group i, which is dependent on the amount of mixed traffic within each user group.

The costs of building one satellite constellation systems to handle the sum of the two user traffics (i.e., $T = 800$ wavelengths) are shown in Figure 3-56. Traffic among the two user communities does not mix within the constellation, i.e., each satellite node has two switches. Building two separate switches in one satellite node is less expensive than building one large switch to handle both types of traffic (i.e., $k_2[W_{1H} + W_{1U}]^2 + k_2[W_{2H} + W_{2U}]^2 < k_2[W_{1H} + W_{1U} + W_{2H} + W_{2U}]^2$) where W_{iH} is the number of wavelengths for uniform all-to-one traffic for user group i and W_{iU} is the number of wavelengths for uniform-all-to-all traffic for user group i. Likewise, the contour lines in Figure 3-56 separate the lowest cost constellation built within each region of varying mixed traffic between the two user groups.

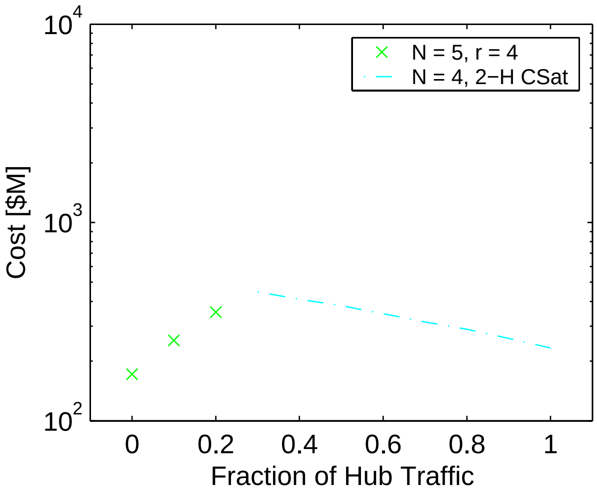


Figure 3-54: Lowest communications costs for MLM cost scenario with mixed traffic.

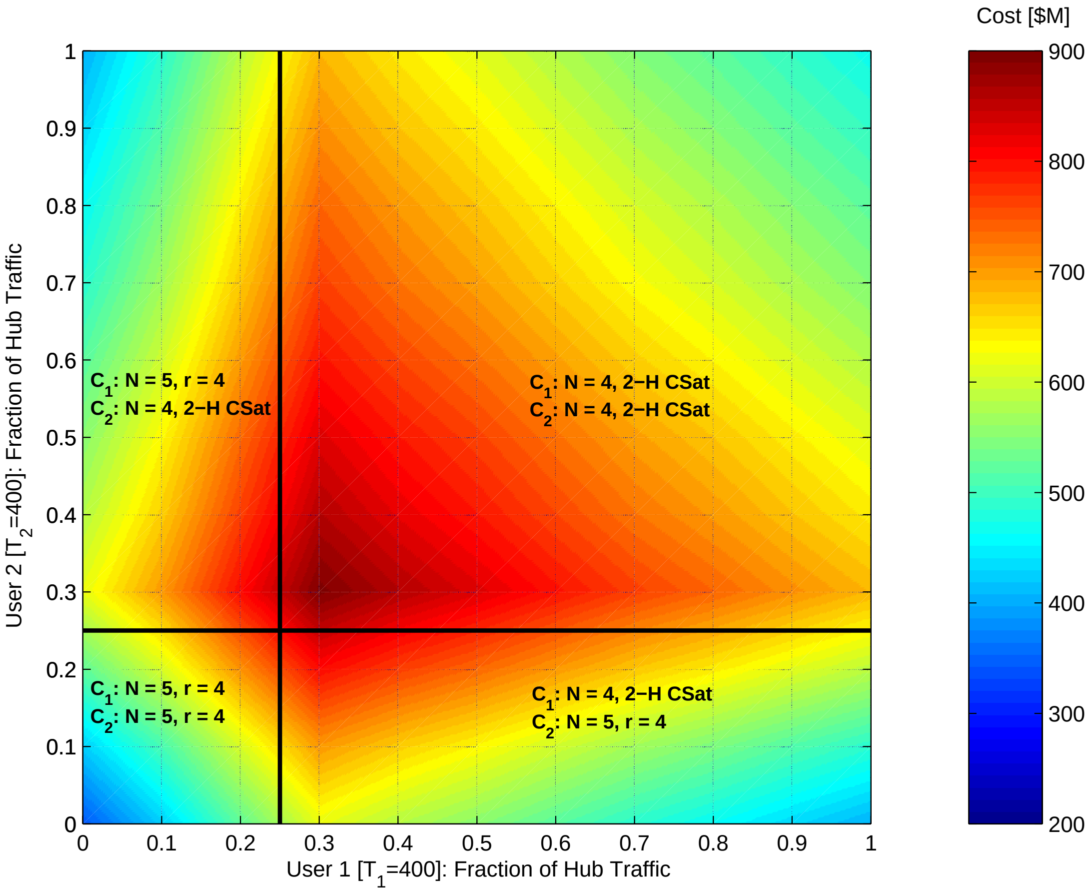


Figure 3-55: Lowest communications costs of 2 separate constellation systems for MLM cost scenario with mixed traffic.

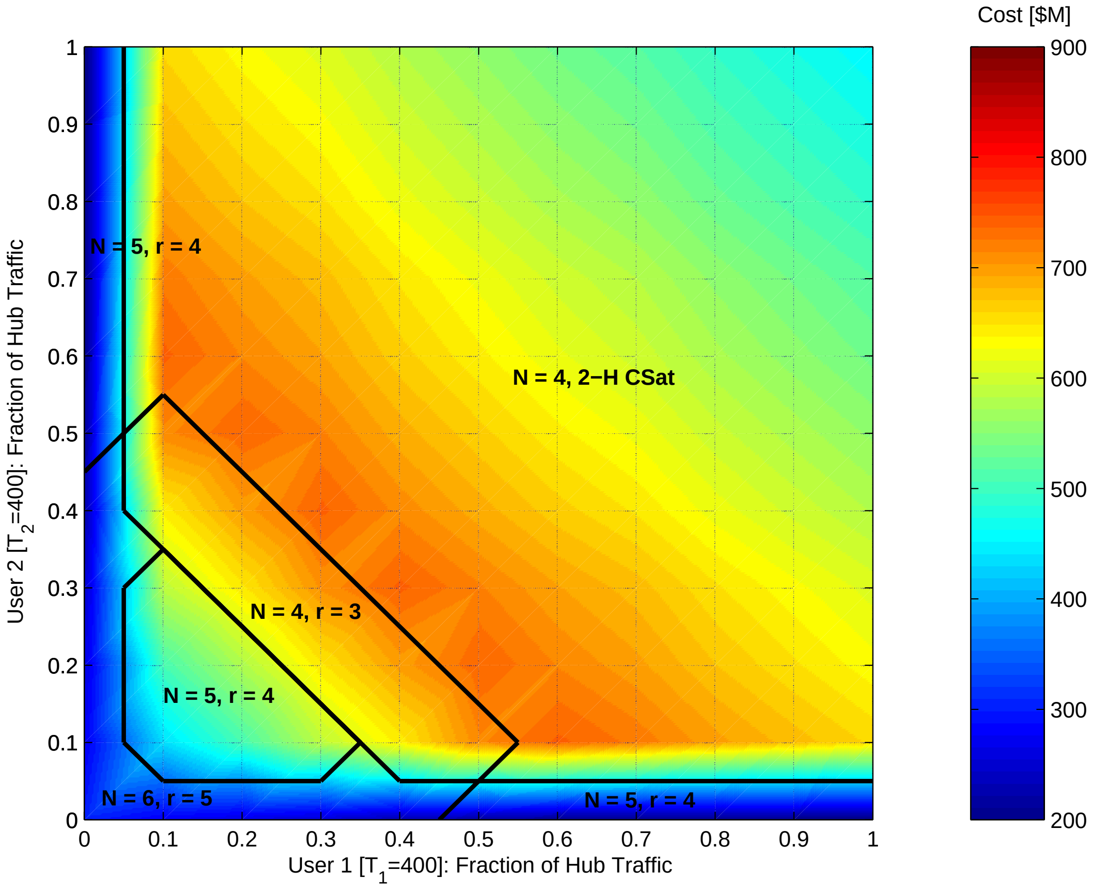


Figure 3-56: Lowest communications costs of 1 constellation systems for MLM cost scenario with mixed traffic.

A comparison of the costs between two separate satellite constellation systems and one satellite constellation systems for two disparate user groups is made. The results for the lowest cost systems are shown in Figure 3-57. Generally, a one satellite constellation system can satisfy most scenarios. Two separate satellite constellation systems are more cost-effective when the two user groups of traffic are very disparate (i.e., one user group has a high amount of uniform all-to-one traffic and a low amount of uniform all-to-all traffic while the other user group has a small amount of uniform all-to-one traffic and a high amount of uniform all-to-all traffic). This occurs in the MLM cost case when one user community has hub traffic greater than 40% while the other user community has hub traffic less than 30%. In regions where two constellation systems are deployed, C_i indicates the constellation that is built for user group i. A set of results for the HMH cost case is shown in Appendix A.1.3.1.

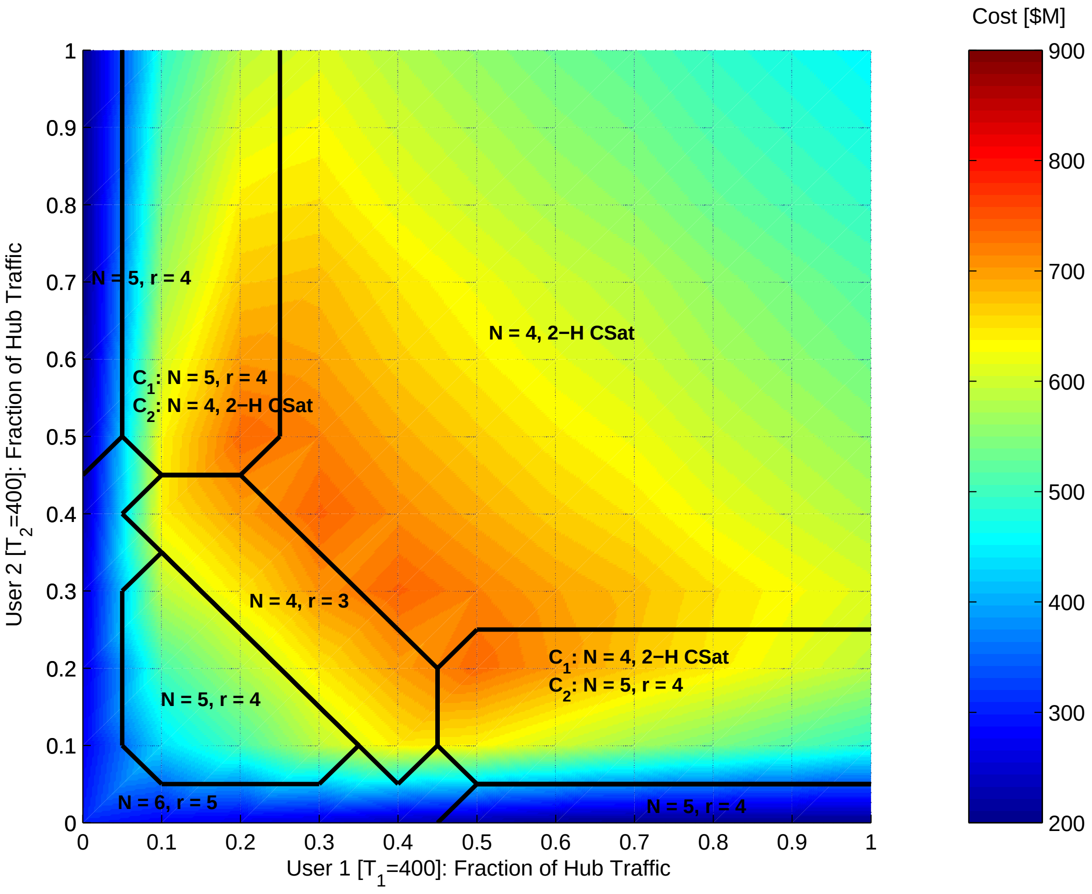


Figure 3-57: Comparison of lowest communications costs of 1 constellation systems vs. 2 separate constellation systems for MLM cost scenario with mixed traffic.

Different Traffic Volumes. In this situation, one user community has significantly less traffic volume than the other user community. Total traffic volume is kept constant at $T = 800$ wavelengths. User group 1 is the smaller community using 80 wavelengths while user group 2 uses 720 wavelengths (i.e., traffic volume of user group 1 is about 11% of traffic volume of user group 2). The communications cost for constellations tailored to each user group's traffic volume is shown Figure 3-58 and Figure 3-59 for the MLM case. The constellation types of lowest communications cost for the various amounts of mixed traffic is re-plotted in Figure 3-60 and Figure 3-61.

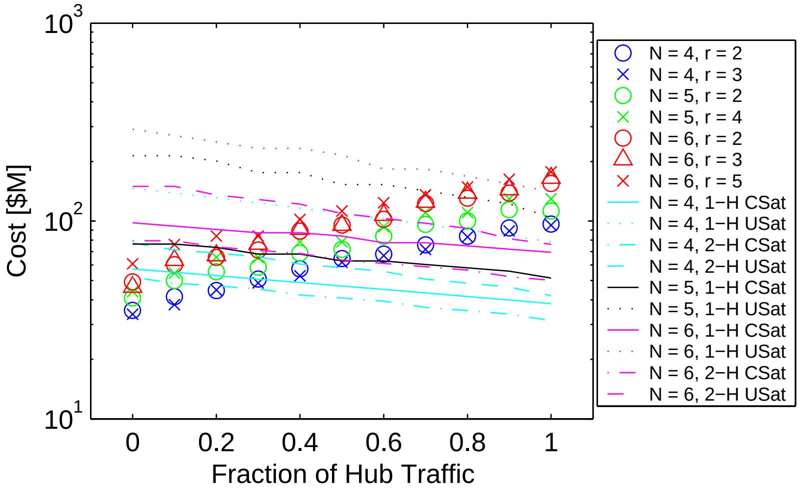


Figure 3-58: Comparing communications costs MLM of N=4,5,6 for a range of mixed traffic for user community 1 ($T_1 = 80$).

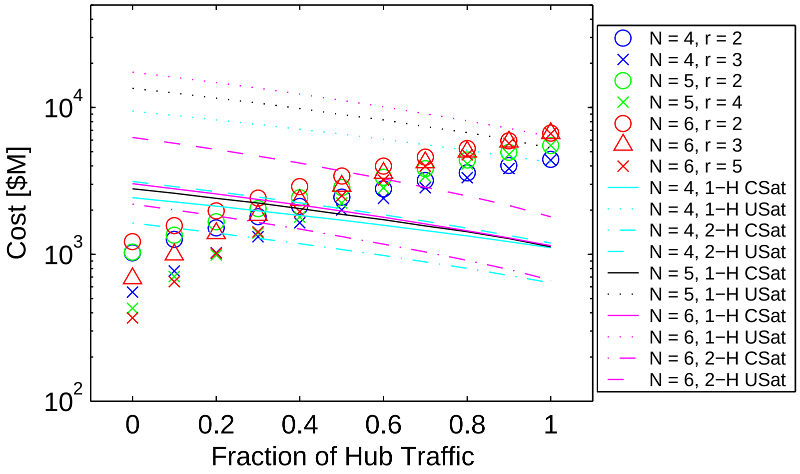


Figure 3-59: Comparing communications costs MLM of N=4,5,6 for a range of mixed traffic for user community 2 ($T_2 = 720$).

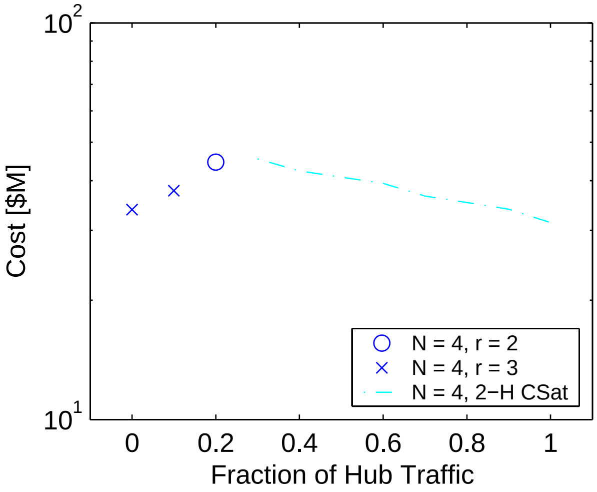


Figure 3-60: Lowest communications costs for MLM cost scenario with mixed traffic for user community 2 ($T_1 = 80$).

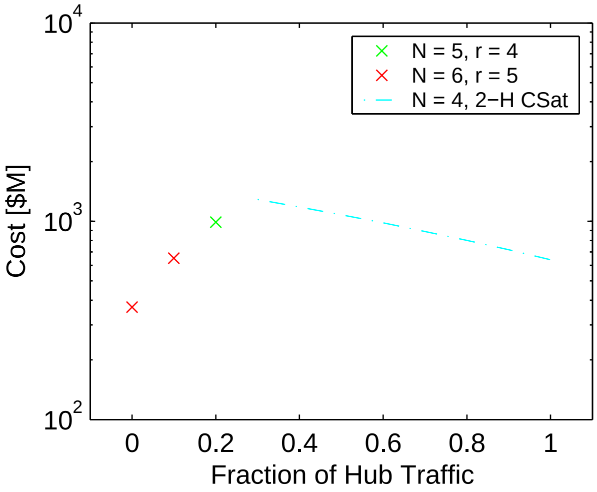


Figure 3-61: Lowest communications costs for MLM cost scenario with mixed traffic for user community 1 ($T_2 = 720$).

The total communications costs of building two separate systems for the various amounts of mixed traffic are shown in Figure 3-62. The costs of building one satellite constellation systems to handle the sum of the two user traffics (i.e., $T = 800$ wavelengths) are shown in Figure 3-63. Again, traffic among the two user communities does not mix within the constellation, i.e., each satellite node has two switches.

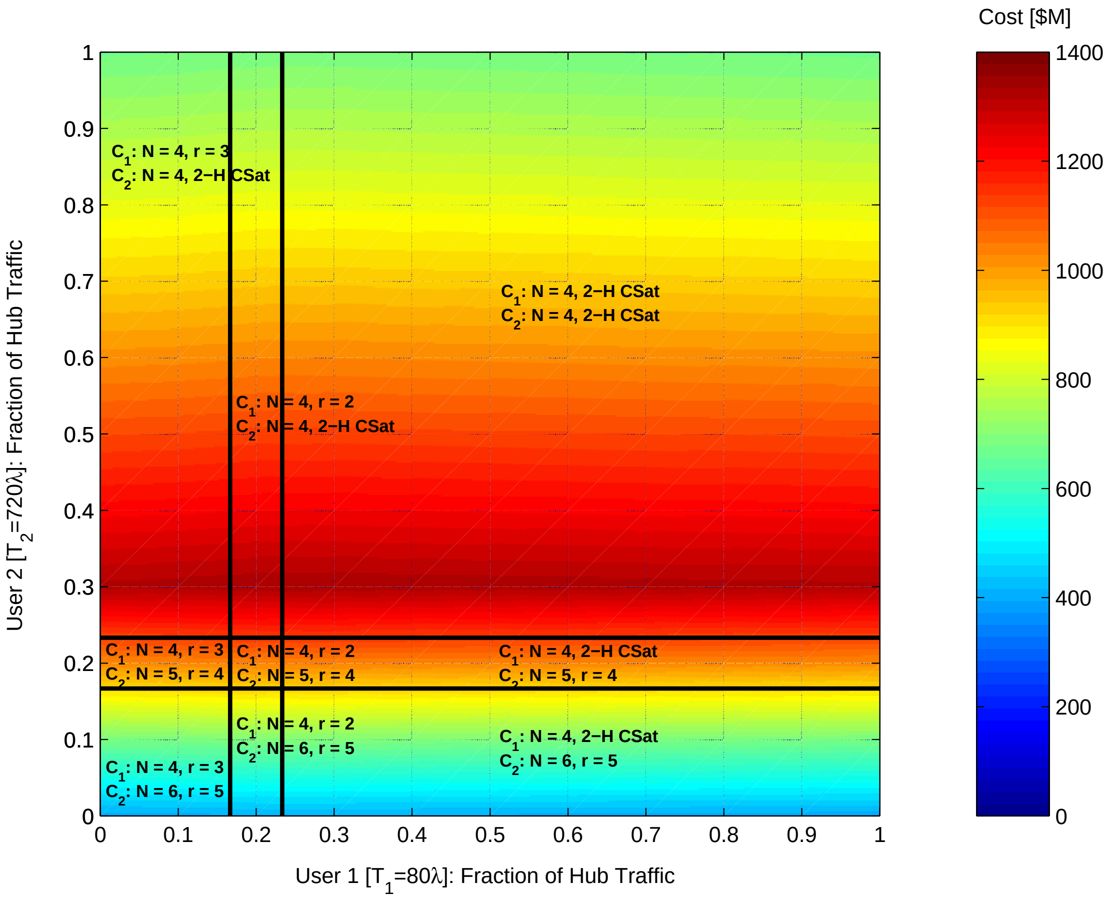


Figure 3-62: Lowest communications costs for MLM cost scenario with mixed traffic of unequal traffic volumes.

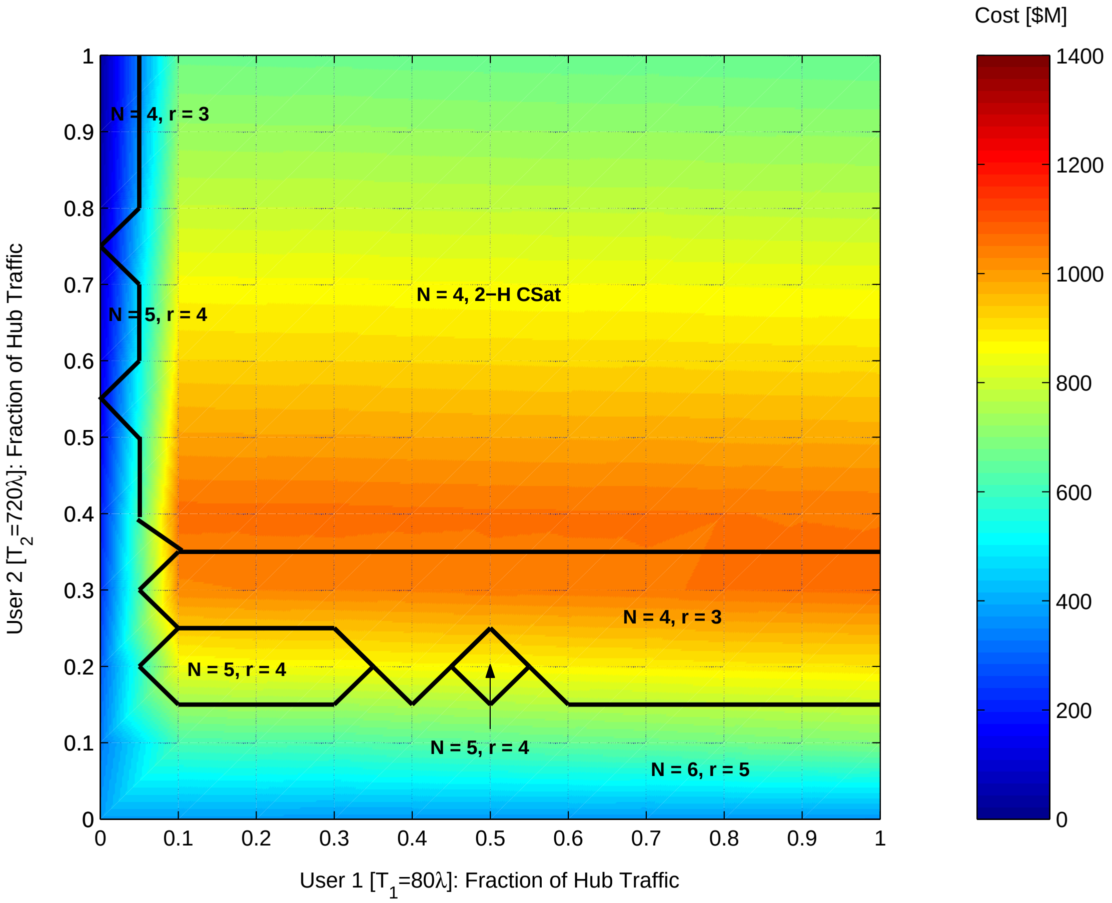


Figure 3-63: Lowest communications costs of 2 separate constellation systems for MLM cost scenario with mixed traffic of unequal traffic volumes using uniform satellites.

A comparison of the costs between two separate satellite constellation systems and one satellite constellation systems for two disparate user groups is made. The results for the lowest cost systems are shown in Figure 3-64. Generally, a one satellite constellation system can satisfy nearly all cases. Two separate satellite constellation systems are more cost-efficient when the two user groups of traffics are very disparate (i.e., one user group has a high amount of uniform all-to-one traffic and a low amount of uniform all-to-all traffic while the other user group has a small amount of uniform all-to-one traffic and a high amount of uniform all-to-all traffic). This occurs in the MLM cost scenario when user community 1 has hub traffic greater than 80% while the user community 2 has hub traffic less than 20%. In regions where two constellation systems are deployed, C_i indicates the constellation that is built for user group i. A set of results for the HMH cost case is shown in Appendix A.1.3.2.

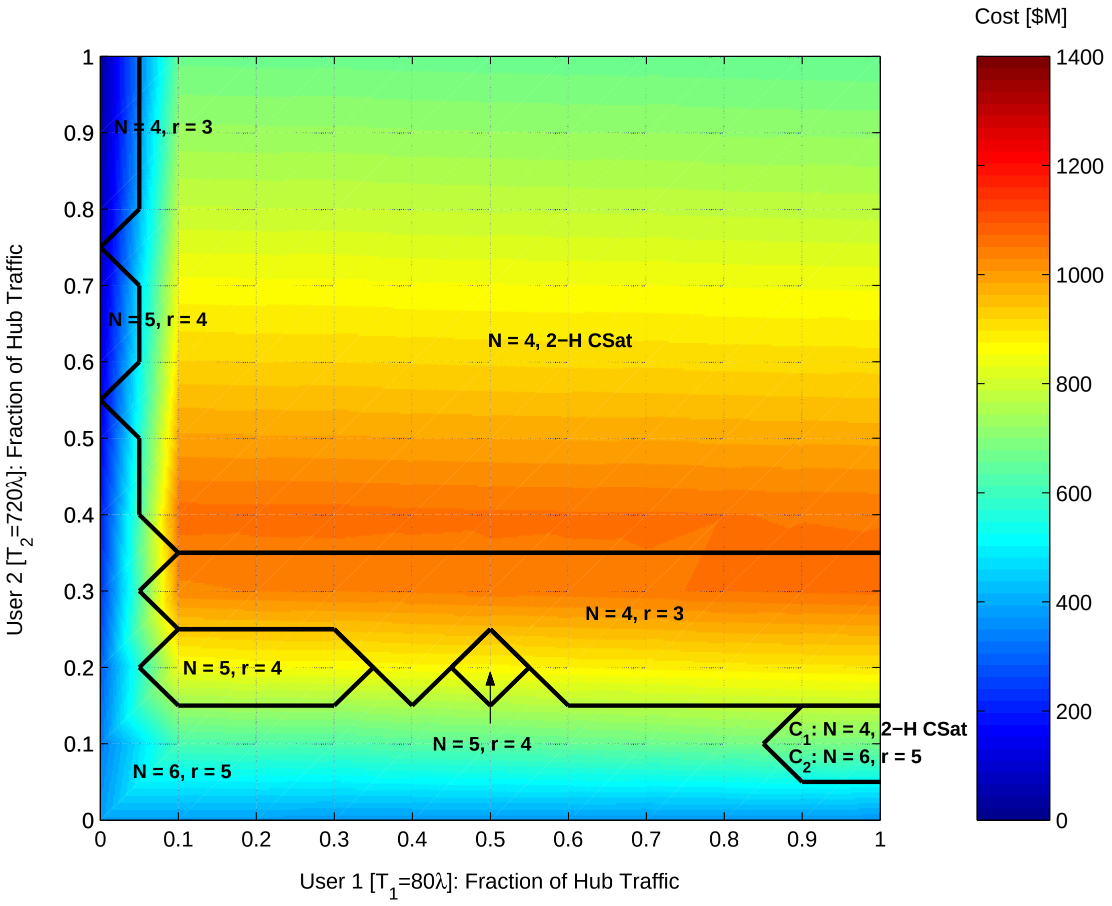


Figure 3-64: Lowest communications costs of 1 constellation systems for MLM cost scenario with mixed traffic of unequal traffic volumes using uniform satellites.

3.1.6 Launch Costs

The launch process can seriously limit the chosen satellite constellation, particularly if the launch cost is a significant fraction of the total costs of the satellites. A launch system should be chosen based on the following criteria: (1) the launch vehicle's capability to lift the required weight to the mission orbit, (2) spacecraft-to-launch-vehicle compatibility, (3) the required launch data versus vehicle availability, and (4) cost of the launch service [92].

To select the appropriate launch system, first ascertain the mission requirements and goals, because they determine the performance, trajectory, and the family of vehicles that can be used. A clear comprehension of the real mission need is particularly vital because it can affect the launch strategy. For example, the space-based processing satellites, with a design lifetime of 2-4 years, require periodic replenishment launches whereas the space-based network backbone satellites, with a design lifetime of 10-15 years, do not. These different requirements may demand different performance from the chosen launch system and its supporting infrastructure.

Satellite mass is one key factor in the choice of an appropriate launch vehicle. Figure 3-65 plots the weight of the communications payload per satellite in each type of network constellation. Results for connected circulant constellations are shown in Figure 3-65(a) and results for hub constellations are shown in Figure 3-65(b). For calculation of the launch weight of the communications payload, it is assumed that the weight of the laser communications package (e.g., antenna telescope and beam steering) is 100 pounds while

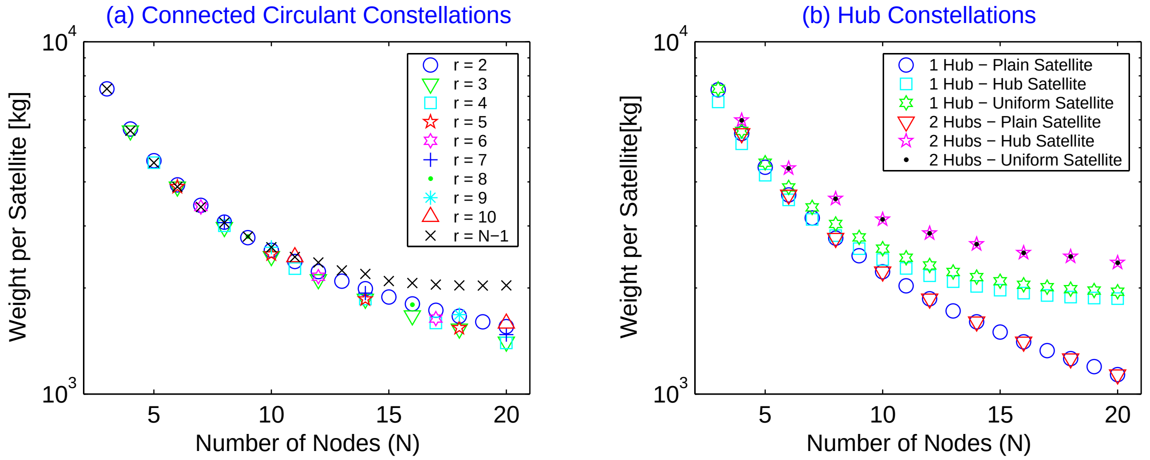
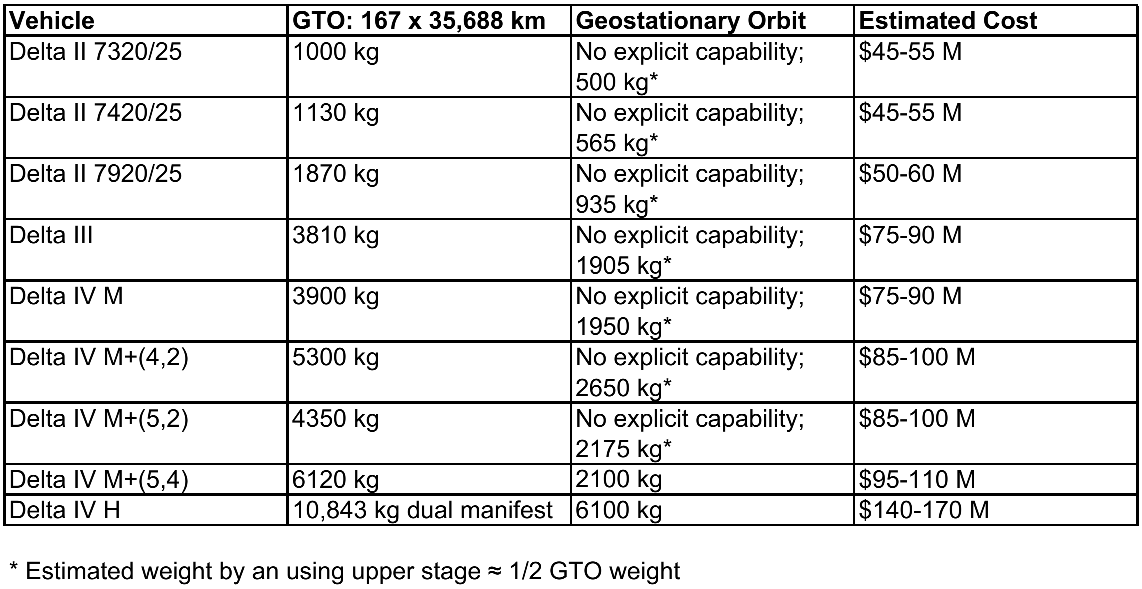


Figure 3-65: Launch weight of communications payload for 400 users with communications costs MLM.

each WDM wavelength link is 10 pounds. User entrance links are uniformly distributed among the number of nodes in the constellation; each receiver is 2 pounds. The number of users on the system is uniformly distributed among the number of nodes in the constellation; each transmitter is 8 pounds. Note that the unit of measure used in calculating the satellite weight is kilograms.

The family of Delta launch vehicles is taken into consideration for deploying the space-based network backbone satellites. Most launch vehicles have the ability to inject satellites into geosynchronous transfer orbits (GTOs) but the Delta IV Heavy configuration is the only launch vehicle that is expected to inject satellites directly into geostationary orbit. Table 3.9 provides information about the different available Delta configurations and their performance and cost [47]. To determine the appropriate launch vehicle for each variation of the space-based network backbone constellation, it is assumed that the launch weight of the communications payload makes up approximately 25% of the total launch weight of the backbone satellite. Assuming a launch to GTO, the remaining weight of the satellite is broken down as 25% for the spacecraft structure (e.g., solar arrays, propulsion system, etc.) and 50% for orbit insertion (i.e., fuel for orbit transfer to GEO). Recalibration of the launch weight per satellite in the backbone constellation is shown in Figure 3-66. Assuming that there are 400 users, each with one 40 Gb wavelength, Figure 3-66(a) shows the results for connected circulant

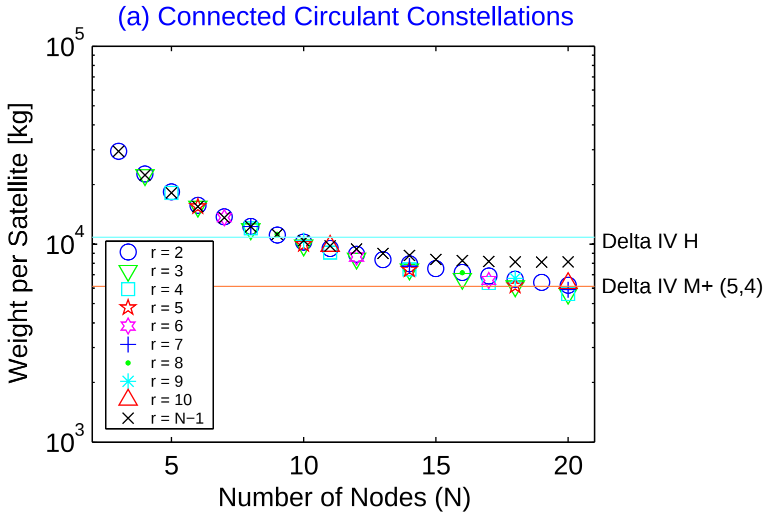


Vehicle	**GTO: 167 x 35,688 km**	**Geostationary Orbit**	**Estimated Cost**
Delta II 7320/25	1000 kg	No explicit capability; 500 kg*	$45-55 M
Delta II 7420/25	1130 kg	No explicit capability; 565 kg*	$45-55 M
Delta II 7920/25	1870 kg	No explicit capability; 935 kg*	$50-60 M
Delta III	3810 kg	No explicit capability; 1905 kg*	$75-90 M
Delta IV M	3900 kg	No explicit capability; 1950 kg*	$75-90 M
Delta IV M+(4,2)	5300 kg	No explicit capability; 2650 kg*	$85-100 M
Delta IV M+(5,2)	4350 kg	No explicit capability; 2175 kg*	$85-100 M
Delta IV M+(5,4)	6120 kg	2100 kg	$95-110 M
Delta IV H	10,843 kg dual manifest	6100 kg	$140-170 M

* Estimated weight by an using upper stage ≈ 1/2 GTO weight

Table 3.9: Delta launch vehicle performance and cost.

constellations and Figure 3-66(b) shows the results for hub constellations. In each subfigure, the asymptotes depict the maximum payload for each type of Delta vehicle. The main observation seen is that small constellation systems require heavier launchers than large constellations. Constellation sizes less than 10 are too heavy for the Delta IV H launch vehicle. The weight of the backbone satellites is driven by the large number of user entrance links. The Titan IV-B launch vehicle may be able to accommodate these heavy satellites; however the performance data for nonstandard reference orbits (e.g., GTO) are not available.


(a) Connected Circulant Constellations

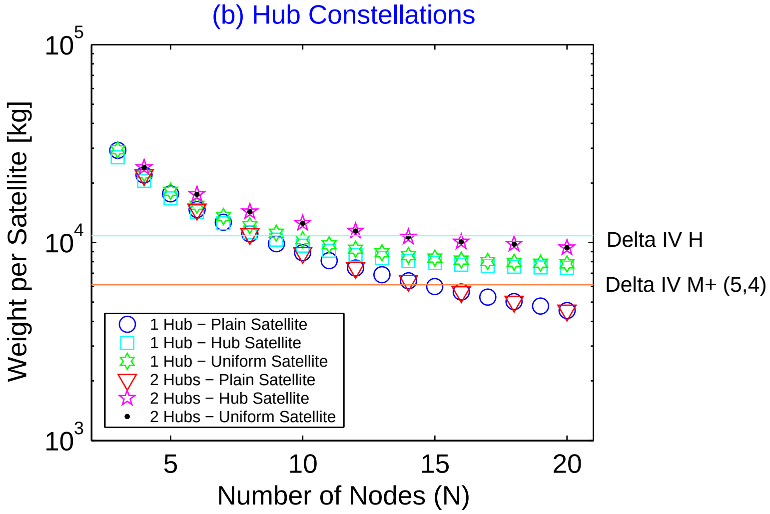

(b) Hub Constellations

Figure 3-66: Delta launch vehicle capability for space-based network backbone constellation for 400 users with communications costs MLM.

A system of 40 users may be more typical. Keeping all other parameters the same, each of the 40 users then has 400 Gb of capacity. The launch weight of the communications payload for 40 users is shown in Figure 3-67. Figure 3-68(a) shows the launch weight results for connected circulant constellations and Figure 3-68(b) shows the launch weight results for hub constellations. Here, all constellation sizes can be serviced with the family of Delta launch vehicles. Other space launch vehicles with payload sizes comparable to the Delta III include the following: Atlas III family, Ariane 4 and 5, Proton, and Sea Launch.

Given that nearly all the satellites for the various backbone constellations can be launched with currently available launch vehicles, of specific interest is the effect of launch costs on the constellation design choice. Figure 3-69 provides results for the connected circulant constellations with 400 users. Figure 3-70 provides results for the connected circulant constellations with 40 users. Figure 3-71 provides results for the hub constellations with 400 users. Figure 3-72 provides results for the hub constellations with 40 users. In each of these figures, there are 4 subfigures: (a) the communications cost, (b) the launch cost, (c) the sum of communications cost and launch cost, and (d) launch cost as a percentage of the total system cost. **Note that the launch cost for each of the Delta launch vehicles is calculated with the upper bound of the estimated cost range in Table 3.9.**

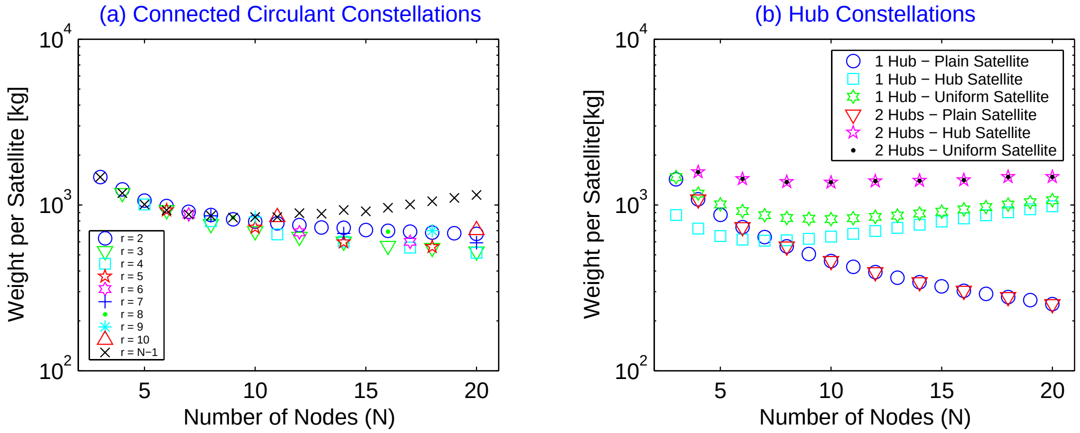
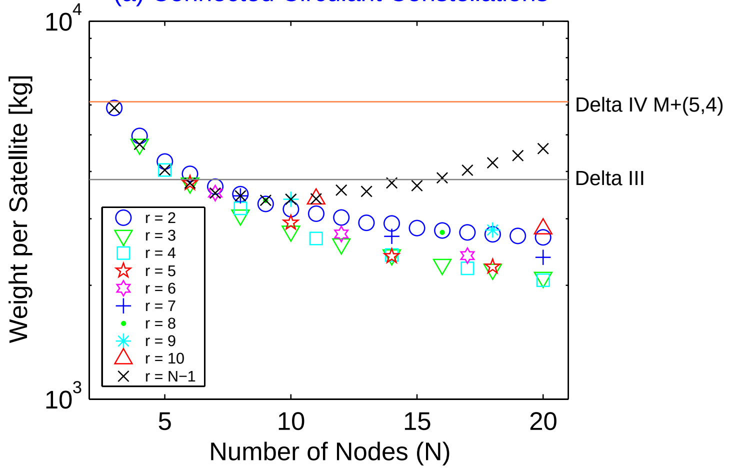


Figure 3-67: Launch weight of communications payload for 40 users with communications costs MLM.

(a) Connected Circulant Constellations



(b) Hub Constellations

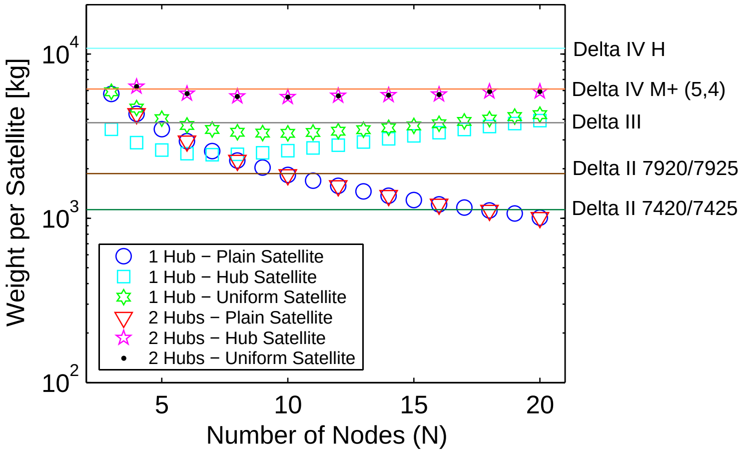


Figure 3-68: Delta launch vehicle capability for space-based network backbone constellation for 40 users with communications cost MLM.

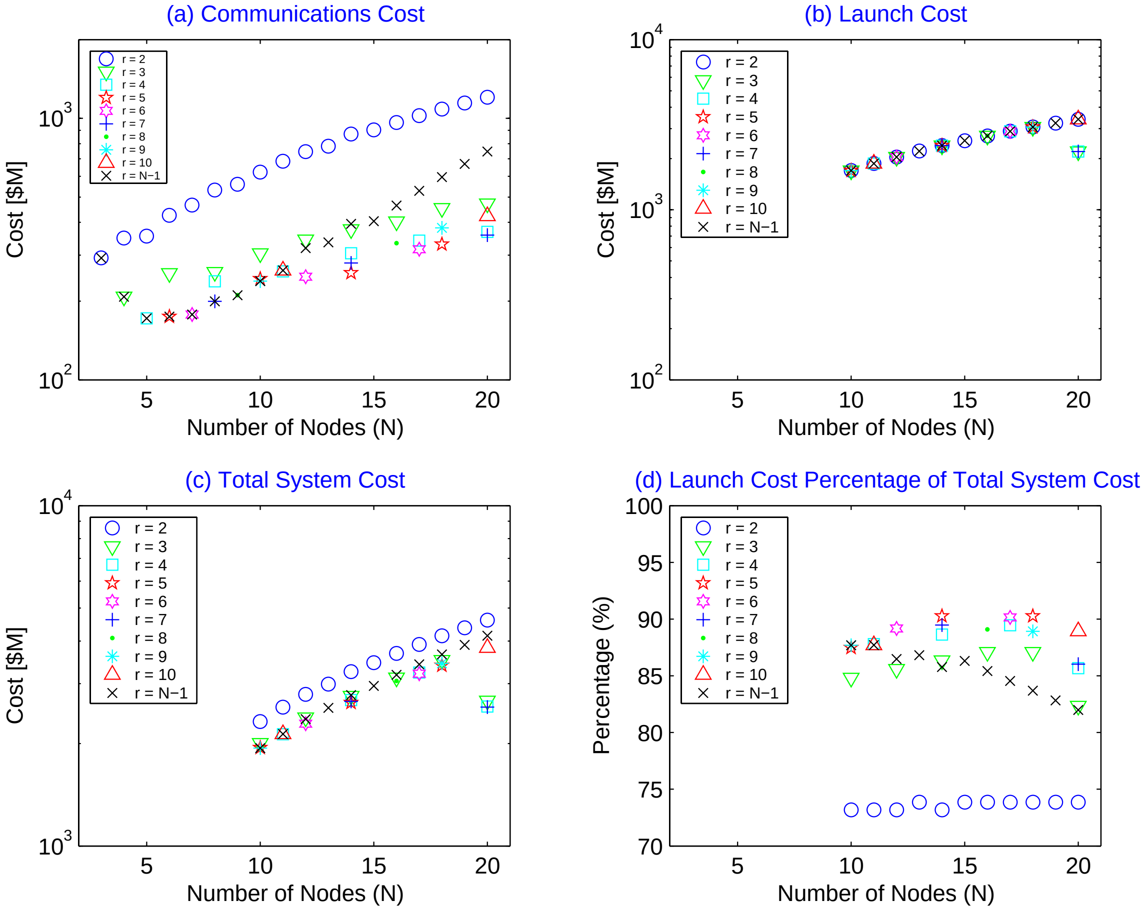


Figure 3-69: Costs for connected circulant constellations for 400 users with communications cost MLM.

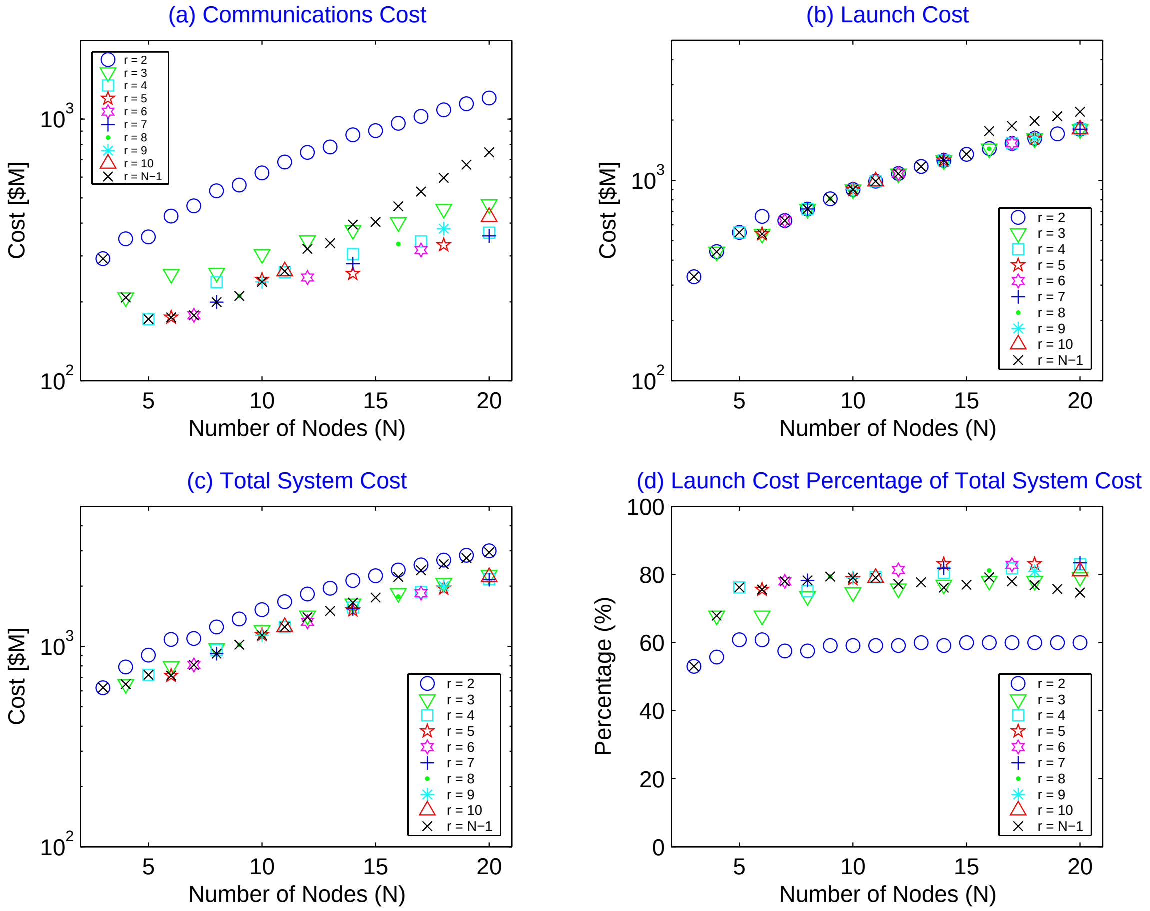


Figure 3-70: Costs for connected circulant constellations for 40 users with communications cost MLM.

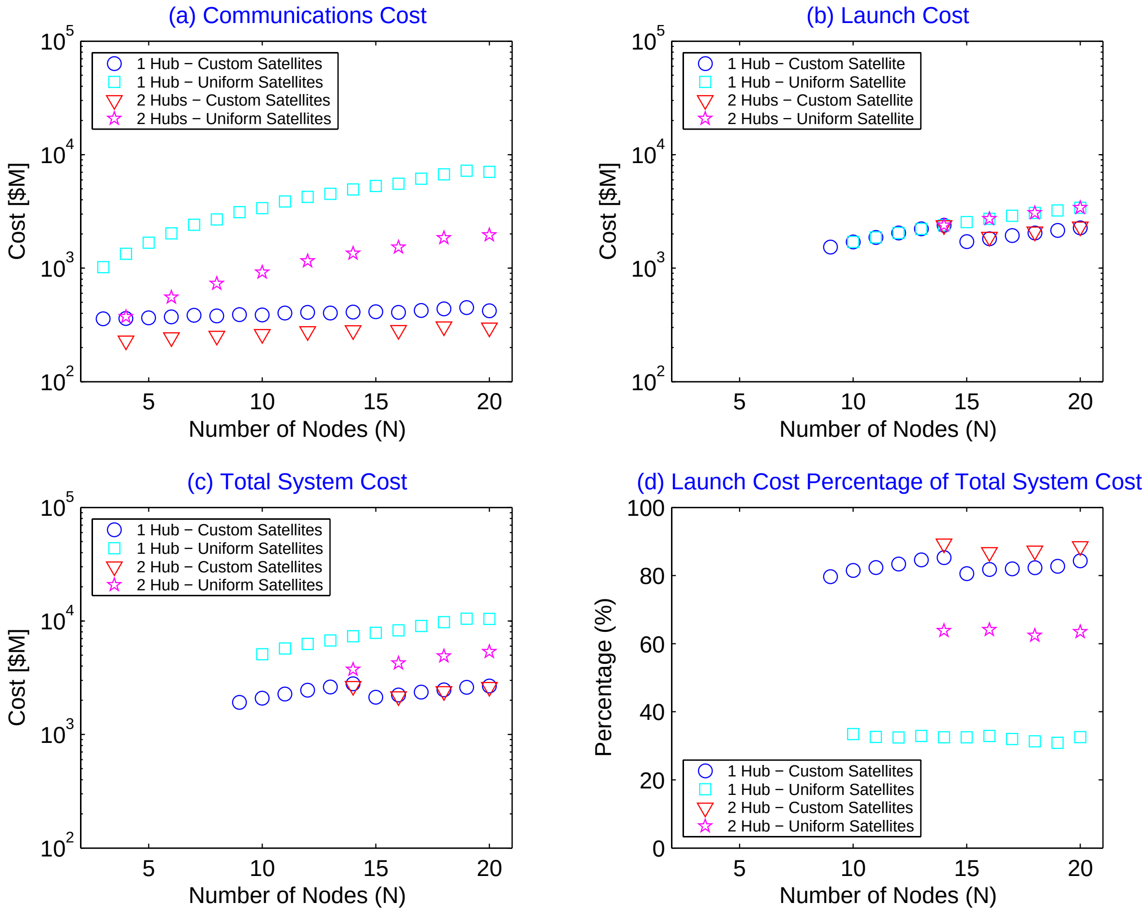


Figure 3-71: Costs for hub constellations for 400 users with communications cost MLM.

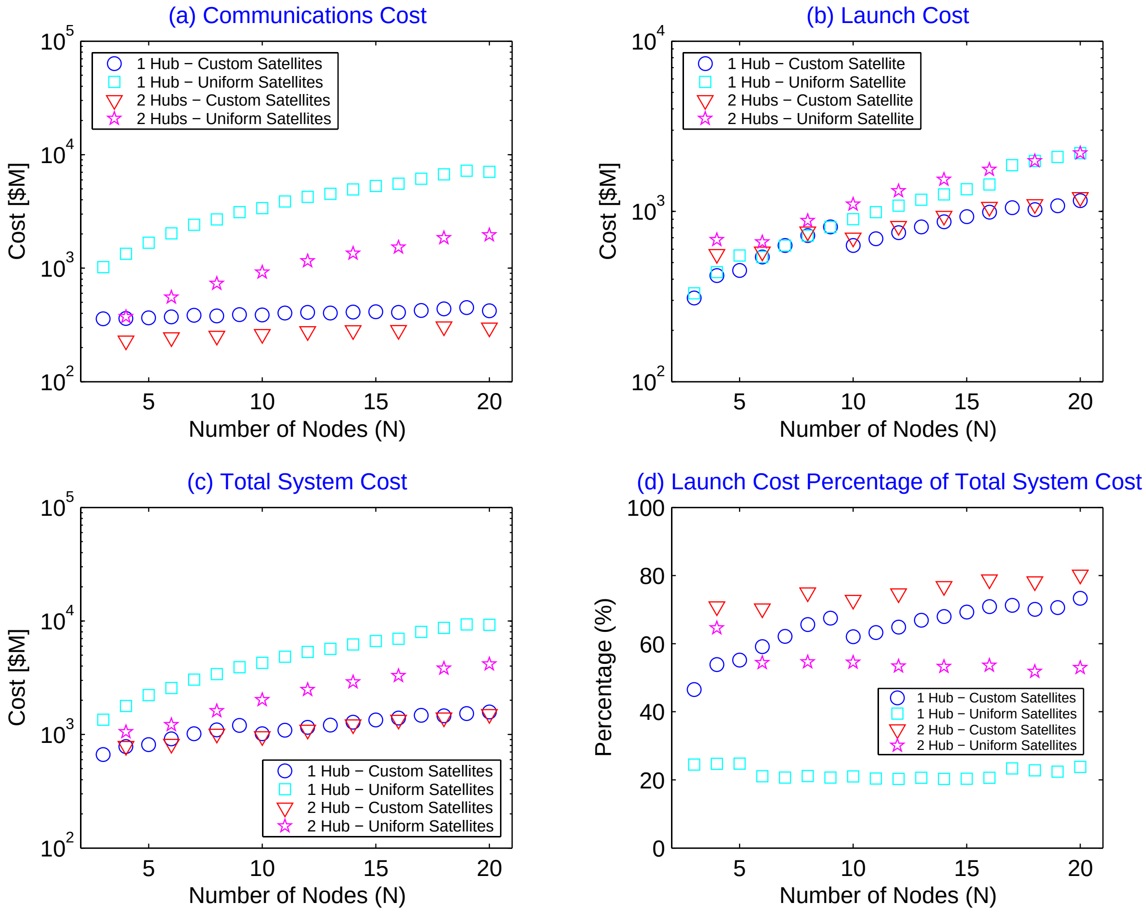


Figure 3-72: Costs for hub constellations for 40 users with communications cost MLM.

It can be seen from Figure 3-69 that the lowest total system cost for a connected circulant constellation with 400 users is $N = 10, r = 9$. The numerical values depicted in Figure 3-69 are listed in Table 3.10. The choice of a reliable low cost system is not as apparent for a connected circulant constellation with 40 users. Notice from Figure 3-70 that the constellation $N = 3, r = 2$ provides the lowest total system cost. This constellation provides no redundancy if there exists a launch failure or satellite failure. Losing one satellite breaks the space network's ability to provide near global coverage. Also note that without the effect of launch costs, the constellation with lowest communications cost is $N = 5, r = 4$. However, launch cost is a very significant factor that when taken into consideration, the constellation $N = 4, r = 3$ provides a lower total system cost than $N = 5, r = 4$. Table 3.11 lists the numerical values of the these connected circulant constellations. A careful trade must be made to balance redundancy, risk, and cost.

Choosing a reliable low cost hub constellation also requires a trade between redundancy, risk, and cost. As seen from Figure 3-71, the lowest total system cost for 400 users is $N = 9, 1$-Hub with custom satellites. A 1-hub constellation is susceptible to failure with the loss of the hub satellite. The lowest total system cost that provides redundancy is $N = 10, 1$-Hub with uniform satellites, where each satellite is built to handle the functions of either a hub or plain node. Table 3.12 lists the numerical values of these hub constellations. For 40 users, the lowest total system cost is $N = 3, 1$-Hub with custom satellites, as seen in Figure 3-72. Again, a constellation of 3 nodes does not provide redundancy in the event of a launch or satellite failure. The next lowest total system cost is $N = 4, 1$-Hub with custom satellites. This system is susceptible to a failure with the hub node. The next constellations to consider are $N = 4, 2-$Hub constellations. The percentage increase between using custom satellites and uniform satellites is 36.76%. Choosing between these two constellations is a trade between the degree of redundancy desired, risk, and cost. Table 3.13 lists the numerical values for these hub constellations.

	N = 10, r = 9
Communications Cost	$283.7103M
Launch Cost	$1.7B
Total System Cost	$1.9387B
Launch Cost Percentage of Total System Cost	87.69%

Table 3.10: Lowest costs for connected circulant constellations for 400 users with communications cost MLM.

	N = 3, r = 2	N = 4, r = 3	N = 5, r = 4
Communications Cost	$292.1867M	$207.9294M	$171.9356M
Launch Cost	$330M	$440M	$550M
Total System Cost	$622.1867M	$647.9294M	$721.9356M
Launch Cost Percentage of Total System Cost	53.04%	67.91%	76.18%

Table 3.11: Lowest costs for connected circulant constellations for 40 users with communications cost MLM.

	N = 9, 1-H Custom	N = 10, 1-H Uniform
Communications Cost	$390.3985M	$3.3837B
Launch Cost	$1.530B	$1.7B
Total System Cost	$1.9204B	$5.0837B
Launch Cost Percentage of Total System Cost	79.67%	33.44%

Table 3.12: Lowest costs for hub constellations for 400 users with communications cost MLM.

	N = 3, 1-H Custom	N = 4, 1-H Custom	N = 4, 2-H Custom	N = 4, 2-H Uniform
Communications Cost	$356.5951M	$359.8807M	$229.7807M	$372.7743M
Launch Cost	$310M	$420M	$560M	$680M
Total System Cost	$666.5951M	$779.8807M	$789.7807M	$1.0528B
Launch Cost Percentage of Total System Cost	46.51%	53.85%	70.91%	64.59%

Table 3.13: Lowest costs for hub constellations for 40 users with communications cost MLM.

This chapter has mainly focused on the future cost scenario of MLM. Today's cost scenario is probably HHH, i.e., High cost antennas, High cost switches, and High cost links. With such high component costs for the satellite nodes, launch costs is not as significant a factor as it is in the MLM cost scenario. Figure 3-73 illustrates the HHH results for connected circulants with 400 users. Figure 3-74 illustrates the HHH results for connected circulants with 40 users. Figure 3-75 illustrates the HHH results for hubs with 400 users. Figure 3-76 illustrates the HHH results for hubs with 40 users. Launch costs account for less than 50% of the total system cost for connected circulant constellations and less than 20% of the total system cost for hub constellations. Under the HHH cost scenario, choosing a reliable low cost connected circulant system is more difficult. The minimum cost structure is very flat so that there is no pronounced optimum choice. Several larger constellations provide lower total system cost, however there is current interest in smaller constellations, particularly of 4, 5, or 6 nodes. Once more, a trade must be made between the constellation size, risk, and cost.

Total launch costs for the space-based network backbone constellation may be reduced by configuring the launch vehicle to carry multiple satellites. The Delta II launch system, as an example, has been used for multiple launches of Iridium (five spacecraft per launch) and Globalstar (four spacecraft per launch) payloads. Although both Iridium and Globalstar are LEO constellation systems, the Delta launch system has demonstrated an ability to launch multiple satellites on one launch vehicle. The Delta III is generally used for single payloads or clusters of payloads for a single customer. The Delta IV-M and M+ versions can carry single payloads or multiple spacecrafts for a single customer. The Delta IV-H will have a comanifest capability for two spacecrafts [47].

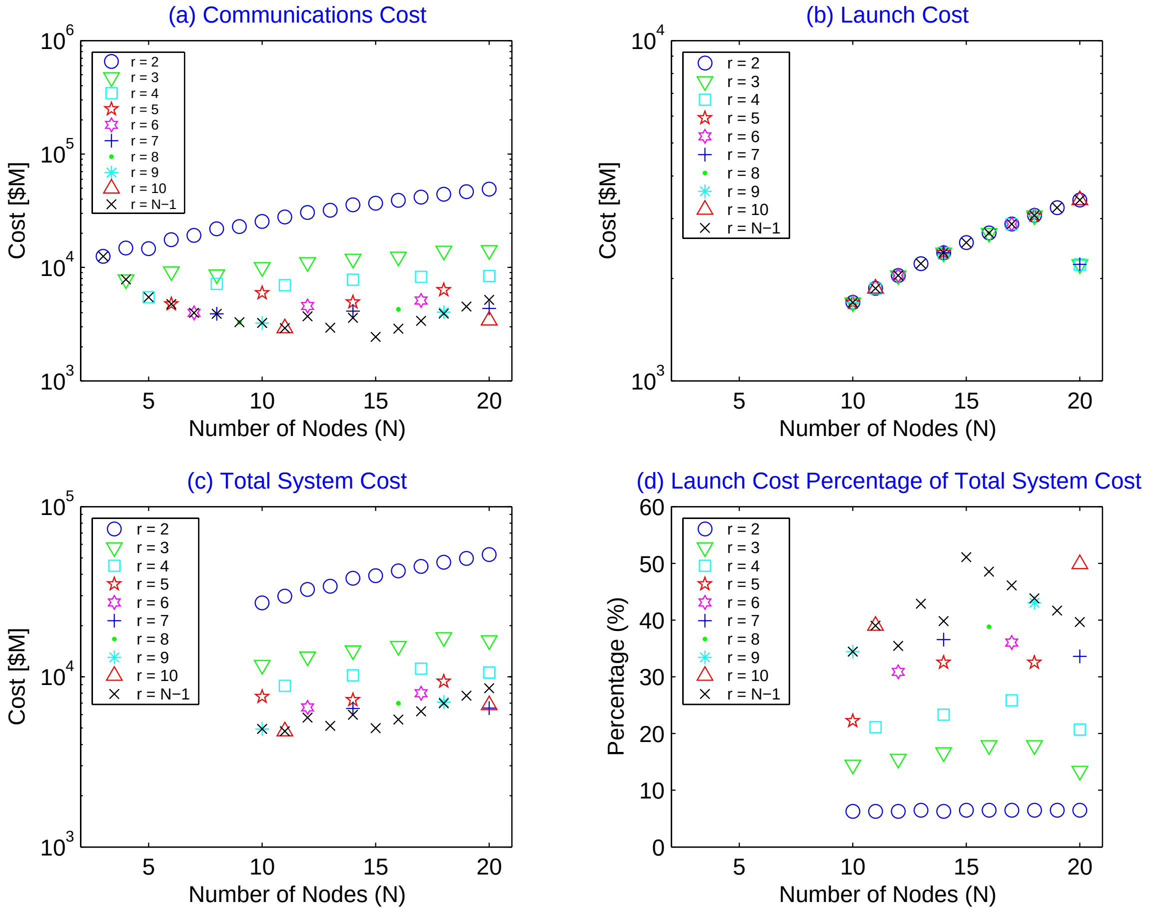


Figure 3-73: Costs for connected circulant constellations for 400 users with communications cost HHH.

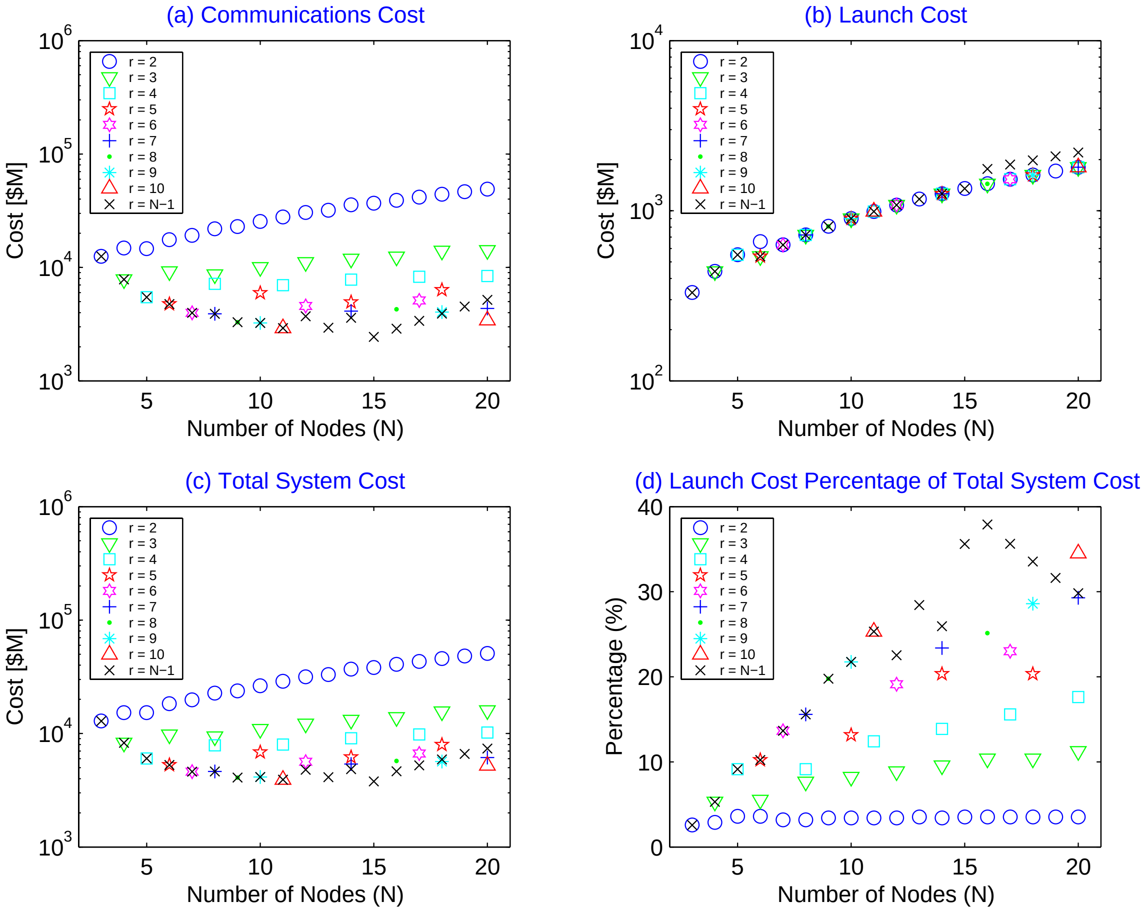


Figure 3-74: Costs for connected circulant constellations for 40 users with communications cost HHH.

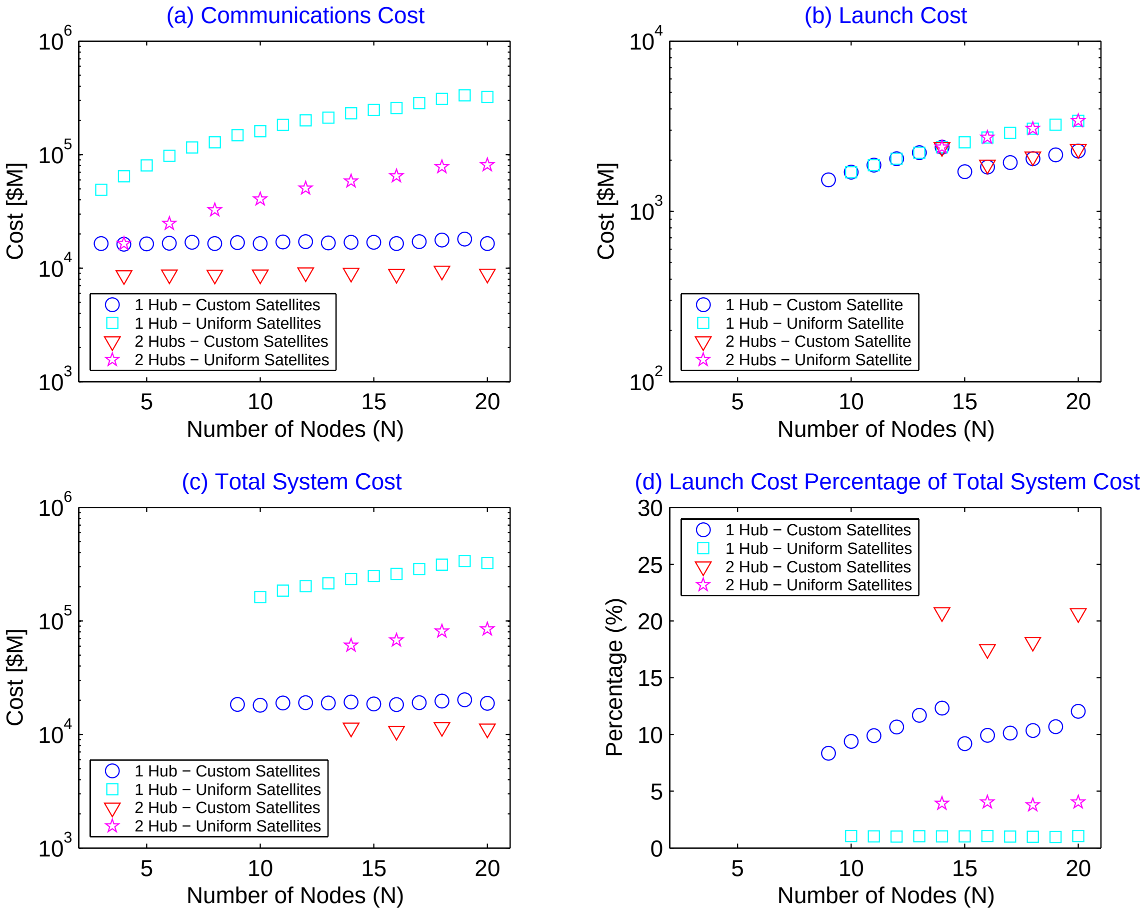


Figure 3-75: Costs for hub constellations for 400 users with communications cost HHH.

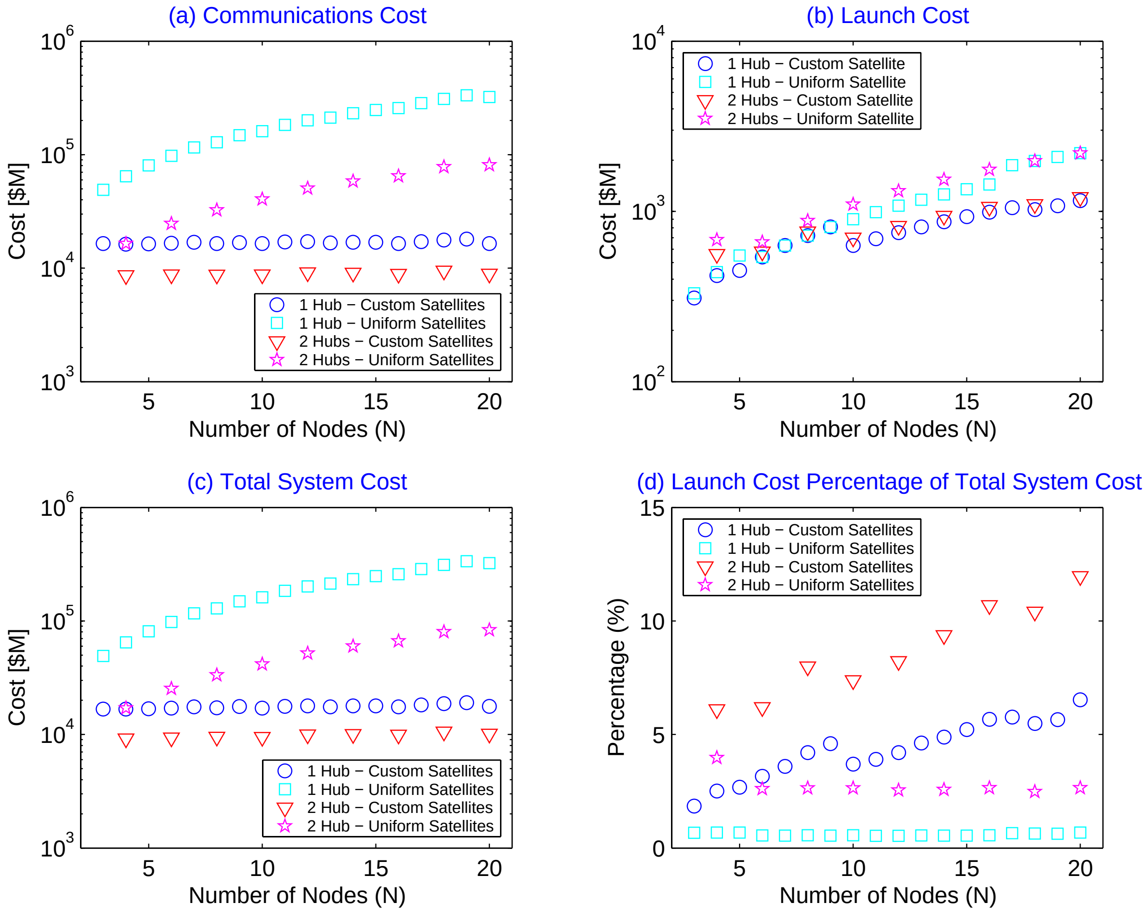


Figure 3-76: Costs for hub constellations for 40 users with communications cost HHH.

3.1.7 Discussion

The communications cost for various satellite constellations and various traffic models for the MLM case (medium cost antennas, low cost switches, and medium cost links) and HMH case (high cost antennas, medium cost switches, and high cost links) have been presented. For connected circulant constellations under uniform all-to-all traffic, ring constellations ($r = 2$) are the least cost-effective while full mesh topologies for $3 \leq N \leq 11$ are the most cost-effective. For hub constellations under uniform all-to-one traffic, all hub constellation types show relatively flat cost structures, with the 2-hub constellations being most cost-effective. Communications costs increase when uniform all-to-all traffic is placed on hub constellations and uniform all-to-one traffic is placed on connected circulant constellations. In these cases, cost increases for uniform all-to-all traffic are observed on the hub constellations because the average minimum hop distance for all source-destination pairs increases, thereby requiring additional wavelengths on the intersatellite links and larger switches in the satellite nodes. Cost increases are observed for the connected circulant constellations under uniform all-to-one traffic when any or all satellites are built identical to the hub node. In these instances, not all the intersatellite links in the network are fully utilized.

Communications costs for mixed traffic for the MLM and HMH cases have also been presented. At low levels of uniform all-to-one traffic, connected circulant constellations provide lower communications costs. At high levels of uniform all-to-one traffic, hub constellations provide lower systems costs. The question of how to satisfy two users communities with different make-ups of mixed traffic of equivalent and non-equivalent traffic volumes has been addressed by comparing whether two separate satellite constellation systems is more cost-effective than one satellite constellation system that is shared. Two separate satellite systems have been shown to be more cost-effective only when the traffic patterns of the two user groups are very disparate. Otherwise, in most cases, one satellite constellation system can be designed to serve both user groups at lower cost.

Of the three communications cost components, the cost of the switch is the largest driving factor. Using a switch with non-linear cost for 400 users, the full mesh constellations provide lowest communications cost. When the number of users is reduced to 100, full mesh constellations are no longer the most cost-effective. Furthermore, a comparison between the use of switches with linear cost and switches with non-linear cost show how the choice of the lowest communications cost small constellation system changes. When using a switch with linear cost, if there is any uniform all-to-one traffic ($\rho > 0\%$), the design choice is to use a hub constellation. When using a switch with non-linear cost, hub constellations are generally considered if uniform all-to-one traffic is approximately $\rho \geq 30\%$ of the total traffic mix.

Launch costs can play a significant role in the choice of constellation topology. Notice that with 400 users, constellations with less than 10 satellite nodes cannot be launched with the family of Delta launch vehicles. Launch cost for connected circulant constellations with communications cost MLM make up a substantial portion of the total system cost. With 400 users, launch cost accounts for more than 70% of the total system cost. With 40 users, launch cost accounts for more than 50% of the total system cost. The launch cost is less significant for constellations with communications cost HHH. The effect of launch costs on both MLM and HHH hub constellations are more spread out, depending on the type of satellite (e.g., hub, plain, or uniform) launched in the constellations.

3.2 Networked Shared On-Orbit Processing

Shared on-orbit processing resources can provide processing capabilities in space that include data and image compression. Recall Figure 2-4 which illustrates the high level data flow architecture for accessing a processing satellite from a mission satellite. Signals are collected by the mission satellite. Digitization may or may not occur on the mission satellite. If digitization occurs on the mission satellites, then the data is digitally transmitted to the backbone relay satellite, otherwise analog transmission is used. The backbone relay satellite then transmits the data to a networked processing satellite. If the analog data requires digitization, then the analog-to-digital conversion occurs first before any processing computations. Additional processing functions include compression, coding, and encryption. Depending on user needs, processed information can be sent to ground stations or space users via the backbone relay satellite or by the mission satellite. In general, communications data may be relayed back to the mission satellites while sensor data may be disseminated through the backbone relay satellites.

Processor connectivity and the data flow of processed and compressed information will have a significant impact on the design of the network backbone. If processed or compressed information is injected back into the backbone network (e.g., backflow to mission satellites), the network may have to be re-designed to handle the additional traffic (e.g., increase the number of wavelengths on the intersatellite links and use larger switches) which will increase total communications costs. This increase would have to be balanced out by a reduction in cost in other areas such as fuel, which will affect the lifetime of the satellite in space. The design changes will be dictated by the amount of traffic flowing back into the network (e.g., compression rates).

3.2.1 Data and Image Compression Applications

The process of transforming an input data stream (e.g., a source stream or original raw data) into another data stream that is smaller in size or lower in rate is known as *data compression*. Many techniques for data compression exist, based on different ideas, suitable for different types of data, and producing different results. The two main techniques for compression are: (1) redundancy reduction and (2) intelligent deletion of unusable or less important information [74].

In digital systems, the three motivations for utilizing data compression are: (1) transmission bandwidth conservation, (2) transmission time reduction, and (3) storage efficiency [46]. Although transmission bandwidth and storage capacity for digitized data have grown at extraordinary rates, the amount of data to be transmitted and stored grows even faster. System designers recognize that there is never an adequate amount of bandwidth, time, or storage, and all are too expensive to waste. Data compression allows the use of these commodities to be more efficient. Opportunities for new products and services for data compression processing can be realized through software or even hardware upgrades.

The cost of data compression is not free. Processing power and processing time must be traded for transmission bandwidth, transmission time, and storage capacity. Processing time for data compressing is a function of the data, the data compression algorithm, and the processor speed. Although data compression algorithms may be computationally intensive, they can be alleviated by high-performance microprocessors, including FPGAs. At least, reasonable increases in processing cost are typically offset by significant reductions in transmission rate and storage costs.

3.2.1.1 Lossy and Lossless Compression

Compression methods can be classified into two categories: lossy and lossless. Lossy compression can achieve better compression rates by losing some information; therefore the result of decompression is not identical to the original data source. Lossy compression methods are popular for compressing speech, audio, image, and video. Even though some information is lost, the auditory and visual limitations to human interpretation makes lossy data compression acceptable for these applications. Lossless data compression is applied when an exact, bit-identical replica of the original data is required. The loss of even a single bit in character text, numeric data, or computer programs is unacceptable for scientific, business, computer programming, database, and e-mail applications.

The compression rate is dependent on the property of the data, the data compression algorithm used, and the level of acceptable information loss. The compression ratio, a measure of the quantity of compression obtained, can be calculated by dividing the original number of bits or bytes by the number of bits or bytes remaining after applying data compression. For lossless compression, typical compression ratios are 2:1 or 3:1, whereas for lossy compression, compression ratios can reach up to 100:1 or more [46].

3.2.2 Processing Satellite Connectivity

A processing satellite can be connected to the space-based network backbone in two ways: (1) connected to hub nodes and (2) connected to plain nodes. An analysis of both methods and system implications is provided. Table 3.14 shows the eight cases of possible processor connectivity. In addition, a brief discussion about connecting multiple processing satellites is given.

		Processor Connectivity			
		Hub		Plain (Hub')	
		No Blackflow	Backflow to All Sources	Backflow to Hub	Backflow to All Sources
Constellation Type	Circulant	Case I	Case III	Case V	Case VII
	Hub	Case II	Case IV	Case VI	Case VIII

Table 3.14: Matrix of processor connectivity cases.

3.2.2.1 Connection to Hub Nodes

Processing resources can be connected to the space-based network backbone via the hub nodes, as shown in the examples in Figure 3-77. Figure 3-77(a) shows an example connected circulant constellation where a processing satellite is connected to a node that has been labeled a hub node to collect hub traffic on the network. Traffic on the network is a mixture of uniform all-to-all traffic and uniform all-to-one traffic (hub traffic). Figure 3-77(b) shows an example 1-hub constellation while Figure 3-77(c) shows an example 2-hub constellation. Hub nodes are assumed to be the nodes that downlink information to the ground stations.

If all processed data from the processing satellites flow to the hub node for transmission to the ground and/or direct to users, then the traffic models used previously in this chapter for designing the space-based network backbone constellation remains unchanged. These types of situations are labeled as Case I and Case II in the Table 3.14. As there is no backflow of information into the backbone network, the only additional change to the network design is the interconnection between the hub node and the processing satellite. Hence, these cases can be reduced to the previously solved problem of analyzing mixed traffic within the various backbone constellations for lowest communications cost.

If there is backflow of information to be relayed to their original sources (e.g., the mission satellites), then the shortest path between the hub node and all other nodes will have to carry more wavelengths and larger switches will be required, thereby driving up total communications costs and should be offset by some benefits as a consequence. These types of situations are labeled as Case III and Case IV in the Table 3.14. Information backflow is concentrated on the shortest path because the traffic model does not consider the scenario where information backflow is split among multiple source-destination paths. For simplicity, all uniform all-to-one traffic analyses have been modeled along a single shortest

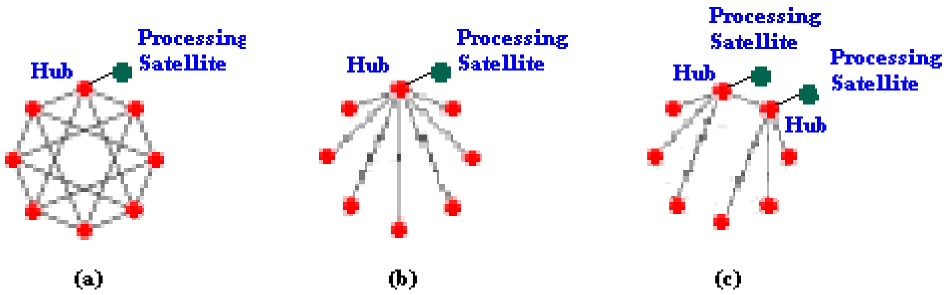
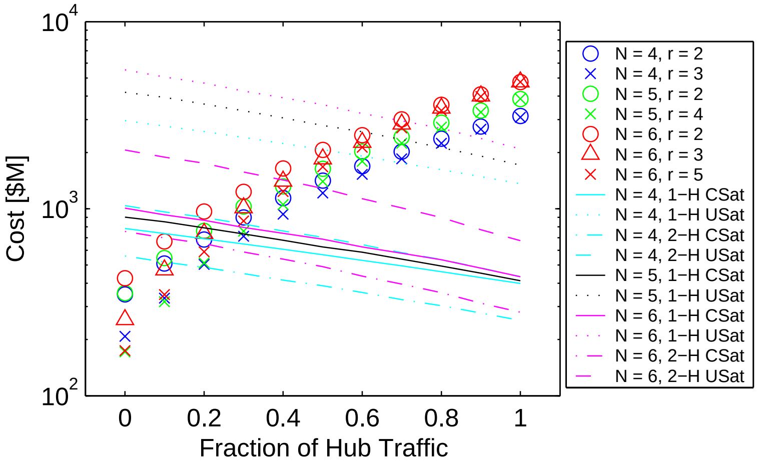


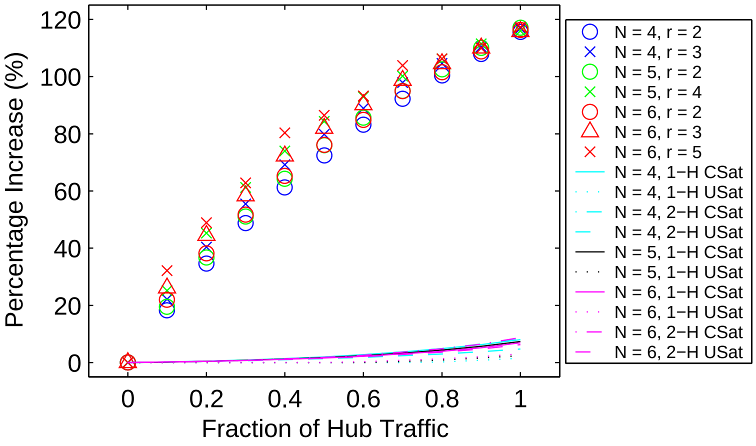
Figure 3-77: Example processor connectivity to a hub node.

source-destination path as well.

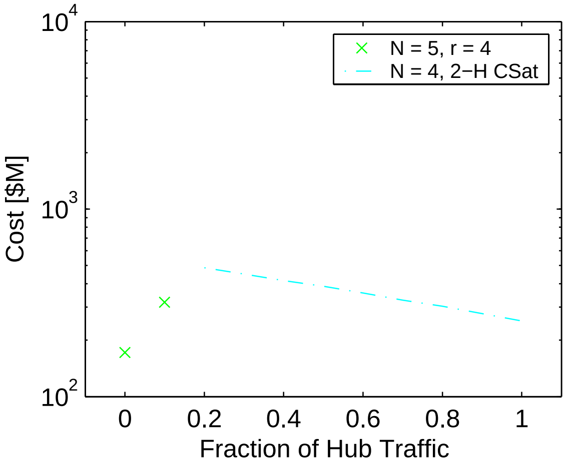
The effect of processed information backflow into the backbone network returning to the original data source nodes is examined for data compress rates of 2:1, 10:1, 25:1, 50:1, and 100:1. Results for Case III and Case IV under the MLM cost case are shown in Figures 3-78, 3-79, 3-80, 3-81, 3-82, respectively. For each compression rate case, three subfigures are shown: (a) the communications costs, (b) the percentage increase in communications costs when compared to Figure 3-53, and (c) the lowest communications cost constellation. Results under the HMH cost case are shown in Figures A-37, A-38, A-39, A-40, A-41, respectively. For each of these compression rate case, three subfigures are shown: (a) the communications costs, (b) the percentage increase in communications costs when compared to Figure A-25, and (c) the lowest communications cost constellation. As compression rates increase, the communications cost increase to the network backbone decreases.



(a) Communications costs MLM.

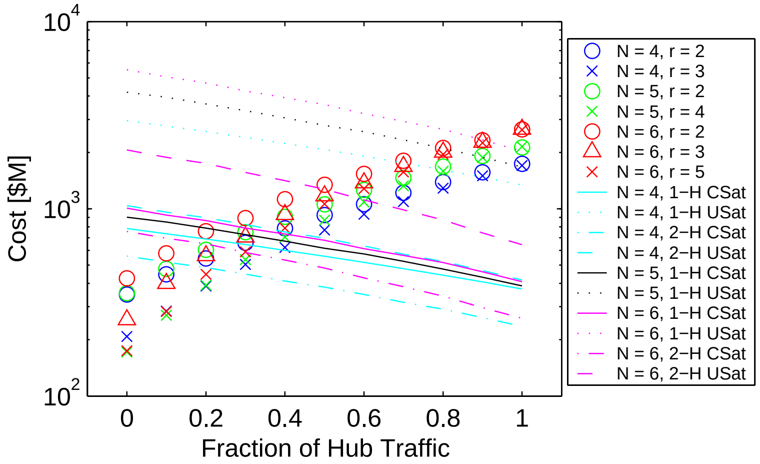


(b) Percentage increase in communications costs MLM.

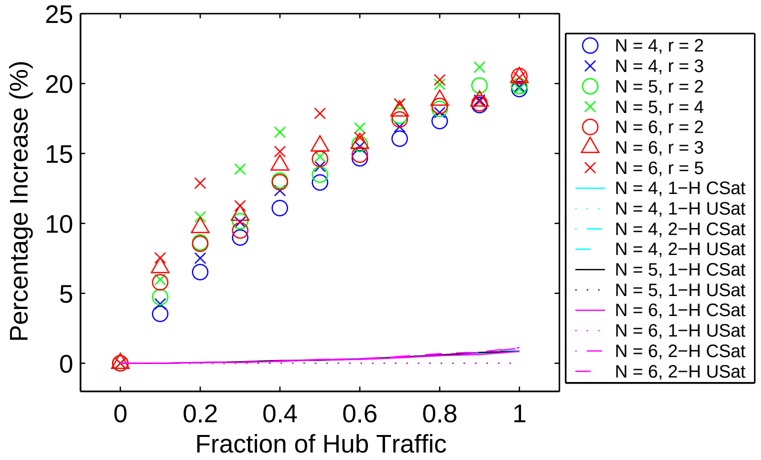


(c) Lowest communications cost constellations MLM.

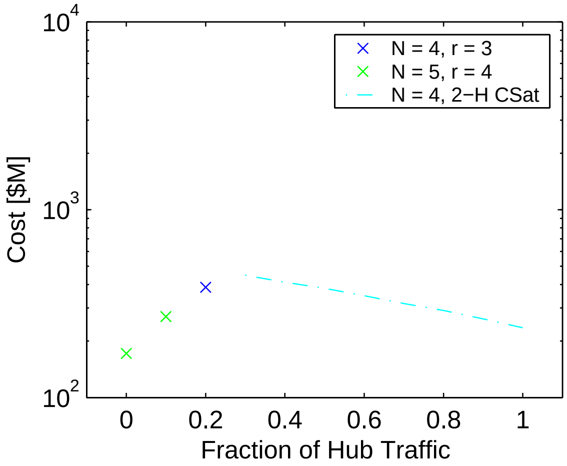
Figure 3-78: Processor-hub connectivity results for compression rate 2:1 MLM.



(a) Communication costs MLM.

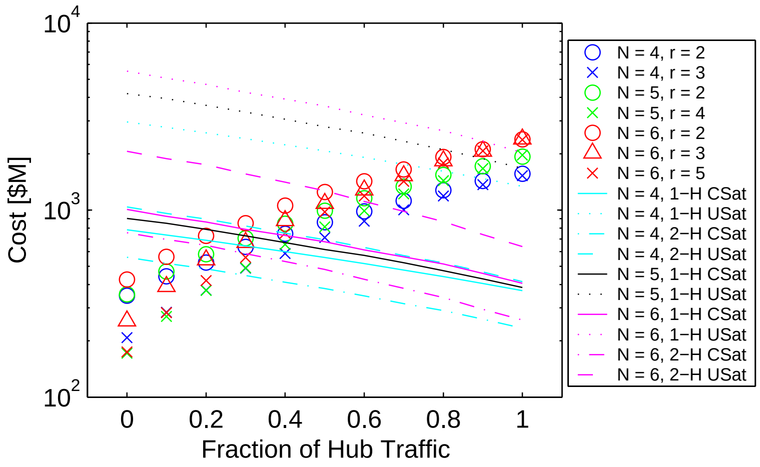


(b) Percentage increase in communications costs MLM.

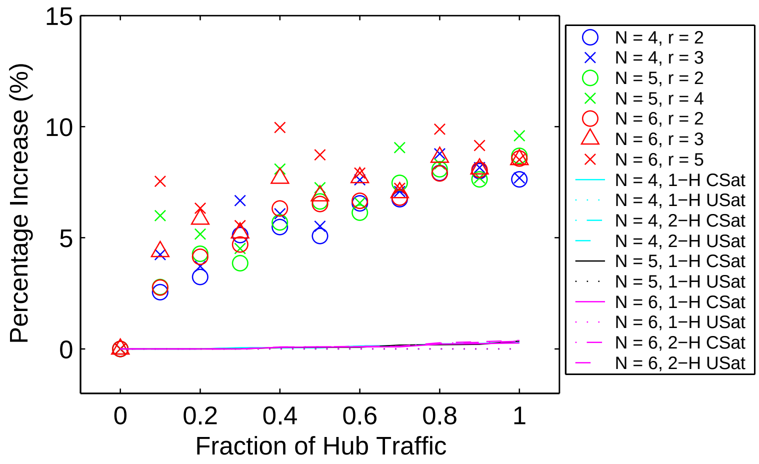


(c) Lowest communications cost constellations MLM.

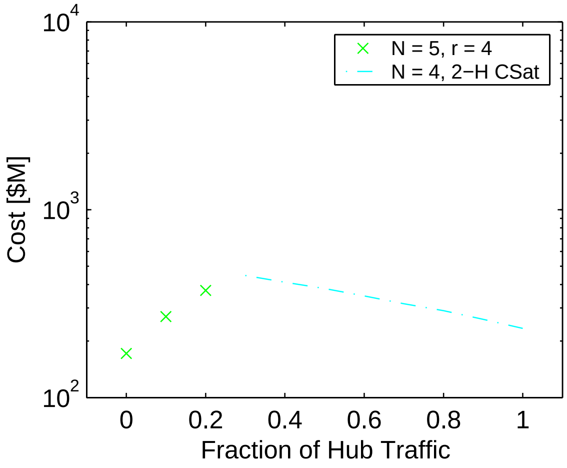
Figure 3-79: Processor-hub connectivity results for compression rate 10:1 MLM.



(a) Communication costs MLM.

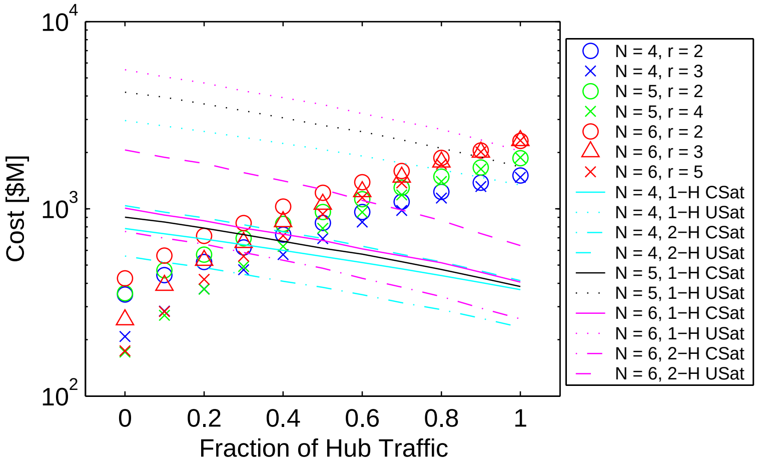


(b) Percentage increase in communications costs MLM.

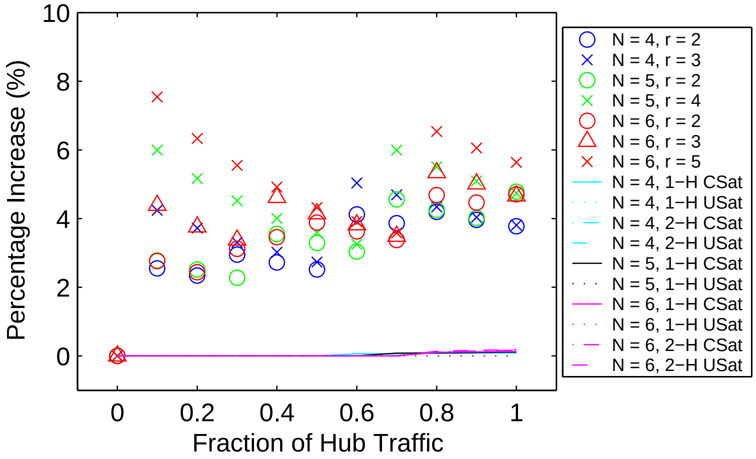


(c) Lowest communications cost constellations MLM.

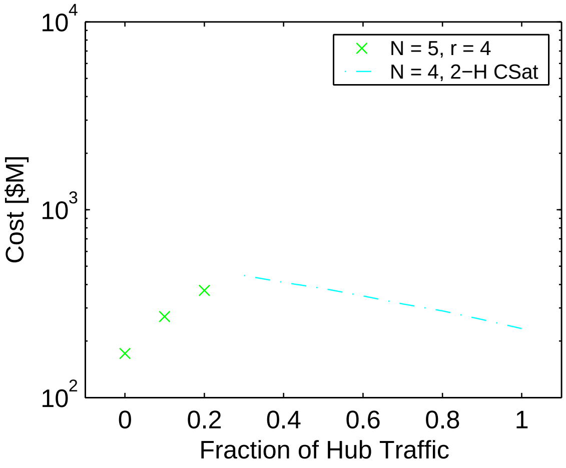
Figure 3-80: Processor-hub connectivity results for compression rate 25:1 MLM.



(a) Communication costs MLM.

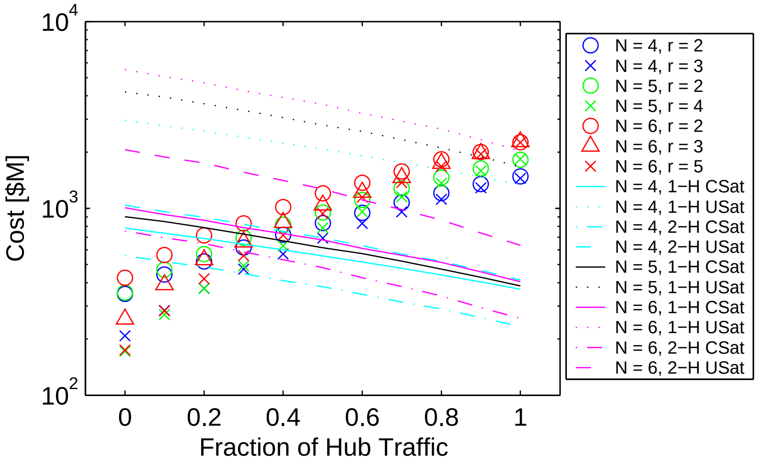


(b) Percentage increase in communications costs MLM.

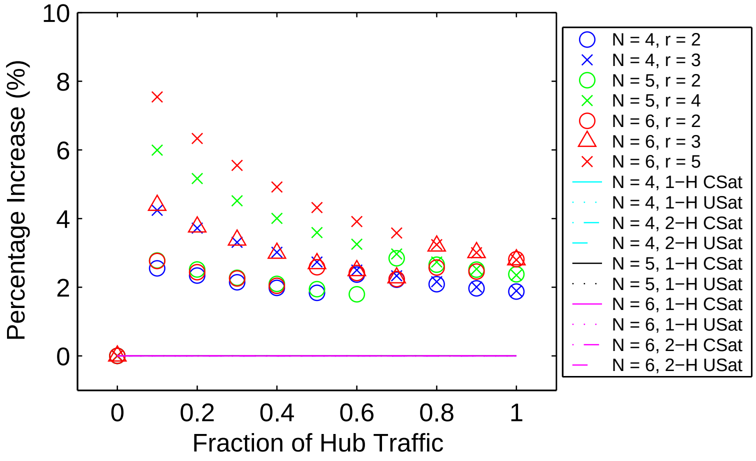


(c) Lowest communications cost constellations MLM.

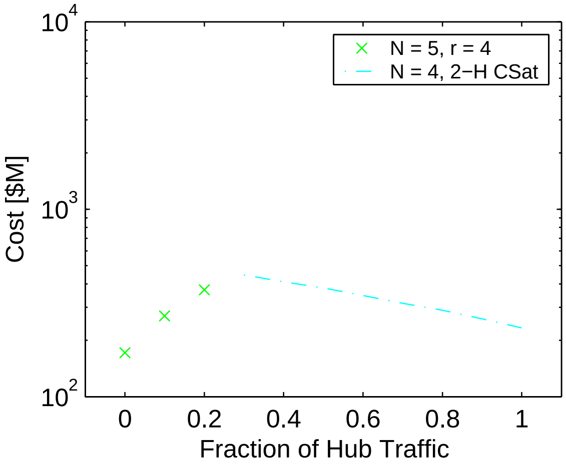
Figure 3-81: Processor-hub connectivity results for compression rate 50:1 MLM.



(a) Communication costs MLM.



(b) Percentage increase in communications costs MLM.



(c) Lowest communications cost constellations MLM.

Figure 3-82: Processor-hub connectivity results for compression rate 100:1 MLM.

3.2.2.2 Connection to Plain Nodes

Processing resources connected to the plain nodes, as shown in Figure 3-83, will require architectural changes in the network design. Figure 3-83(a) shows an example connected circulant constellation where a processing satellite is connected to a plain node on the network. Figure 3-83(b) shows an example 1-hub constellation while Figure 3-83(c) shows an example 2-hub constellation. Hub nodes are assumed to be the nodes that downlink information to the ground stations. The selected plain node, which can be denoted as hub', has the longest single hop intersatellite link distance from the hub node (i.e., $\theta = 180°$). This design choice will upper bound the communications cost of the network because the nodes in the constellation have the potential to relay the highest amount of pass-thru traffic (especially ring constellations), thereby incurring the largest cost increase.

Now, consider the hub node to be the ultimate destination for processed information. If a processing satellite is connected to a plain node, it is observed that data collection to the processor is isomorphic to the mixed traffic cases analyzed previously. For example, the plain node can be considered as the "pseudo-hub" node collecting all the necessary uniform-to-all traffic. Traffic on the network is a mixture of uniform all-to-all traffic and uniform all-to-one traffic (hub traffic). The plain node (hub′) then transmits the network traffic to the processing satellite and subsequently receives processed or compressed information. This data must then be re-directed to the true hub node for dissemination to the ground or to the original data sources.

If there is backflow of information to be relayed to their original sources, then the shortest path between the hub and all other nodes will have to carry more wavelengths and thus larger switches, driving up total communications costs. These types of situations are labeled

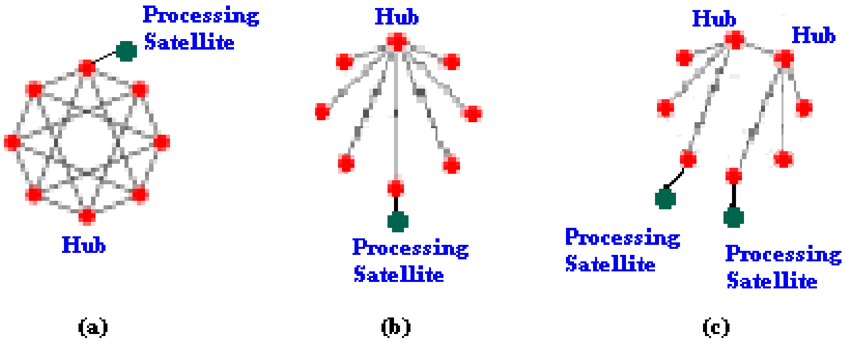
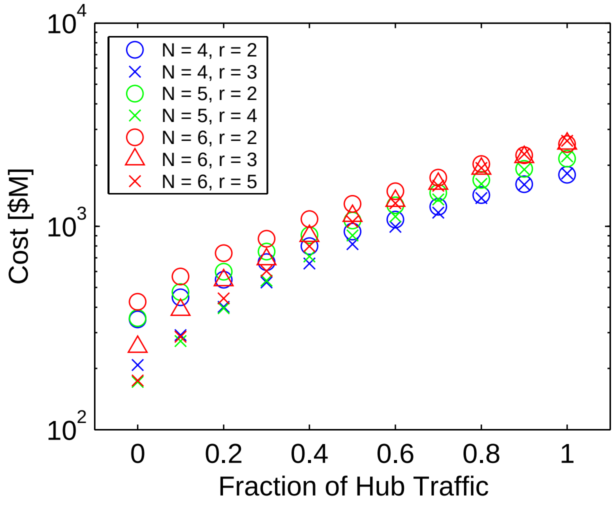


Figure 3-83: Processor connectivity to a plain node.

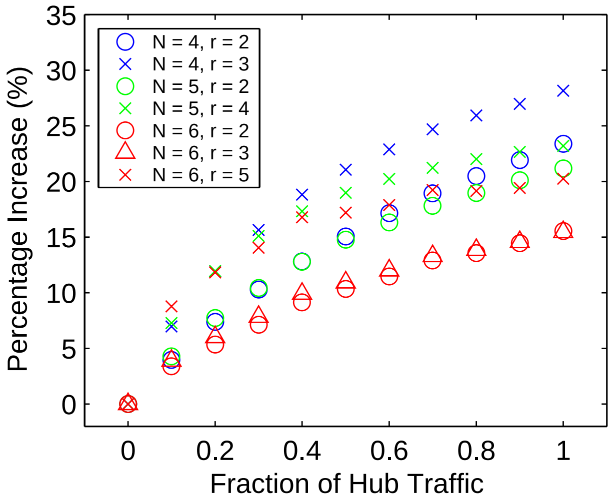
as Case VII and Case VIII in the Table 3.14. Again, for design simplicity, information backflow is concentrated on the shortest path because the traffic model does not consider the scenario where information backflow is split among multiple source-destination paths.

If there is backflow of traffic to the true hub node for data dissemination to the ground and/or directly to users, additional wavelengths on the shortest path between the plain node (hub′) and the hub node are also required. These types of situations are labeled as Case V and Case VI in the Table 3.14. Case V and Case VI will incur higher cost increases than Case VII and Case VIII as the total amount of backflow traffic is concentrated on a single path rather than being uniformly distributed back to the original source nodes. Hence, the cases of information backflow to the true hub node provide the upper bound of total communications cost especially if uniform satellites are used in the network backbone constellation.

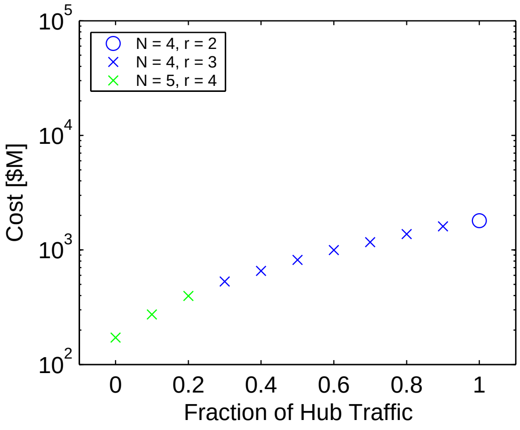
The effect of processed information backflow into the backbone network back to the true hub node for data dissemination to the ground is examined for information compress rates of 2:1, 10:1, 25:1, 50:1, and 100:1. Results for Case V and Case VI under the MLM cost case are shown in Figures 3-84, 3-85, 3-86, 3-87, 3-88, respectively. For each compression rate case, three subfigures are shown: (a) the communications costs, (b) the percentage increase in communications costs when compared to Figure 3-53, and (c) the lowest communications cost constellation. Results under the HMH cost case are shown in Figures A-42, A-43, A-44, A-45, A-46, respectively. For each of these compression rate case, three subfigures are shown: (a) the communications costs, (b) the percentage increase in communications costs when compared to Figure A-25, and (c) the lowest communications cost constellation. As compression rates improve, the communications cost increase to the network backbone decreases.



(a) Communications costs MLM.

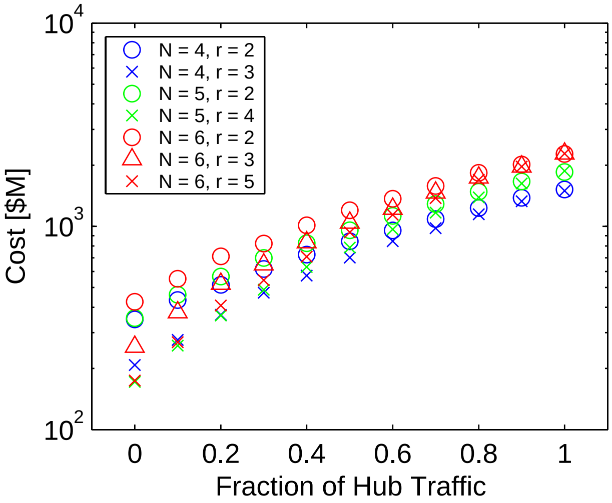


(b) Percentage increase in communications Costs MLM.

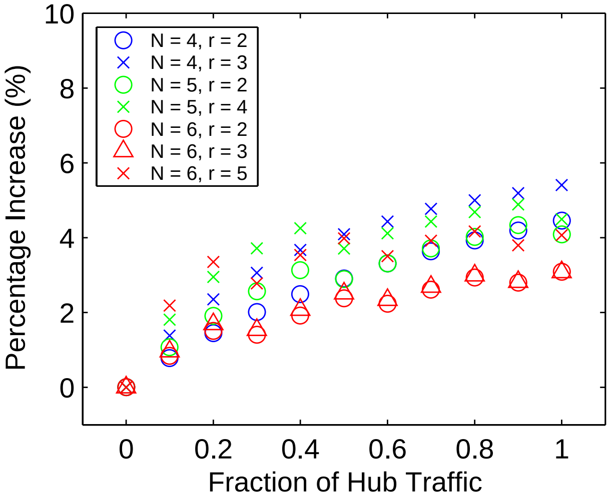


(c) Lowest communications cost constellations MLM.

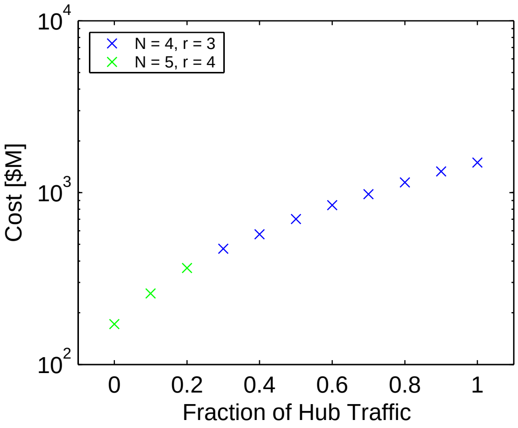
Figure 3-84: Processor-plain node connectivity results for compression rate 2:1 MLM.



(a) Communications costs MLM.

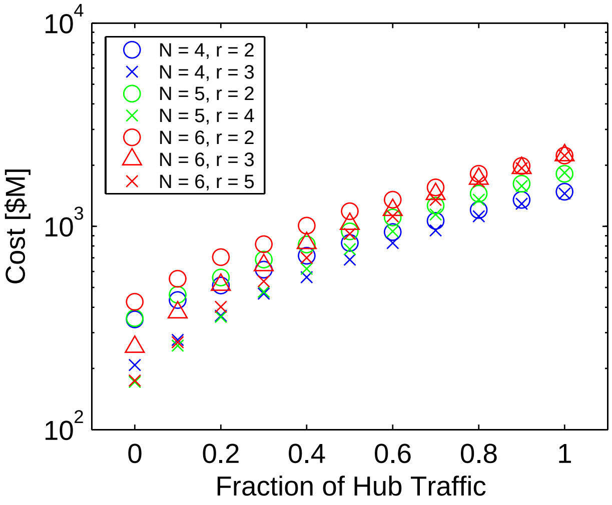


(b) Percentage increase in communications costs MLM.

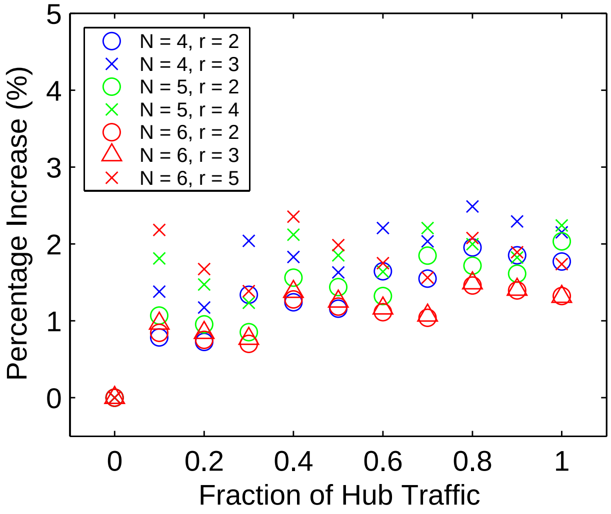


(c) Lowest communications cost constellations MLM.

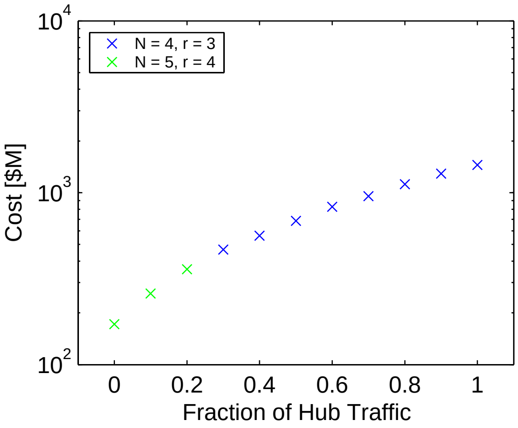
Figure 3-85: Processor-plain node connectivity results for compression rate 10:1 MLM.



(a) Communications costs MLM.

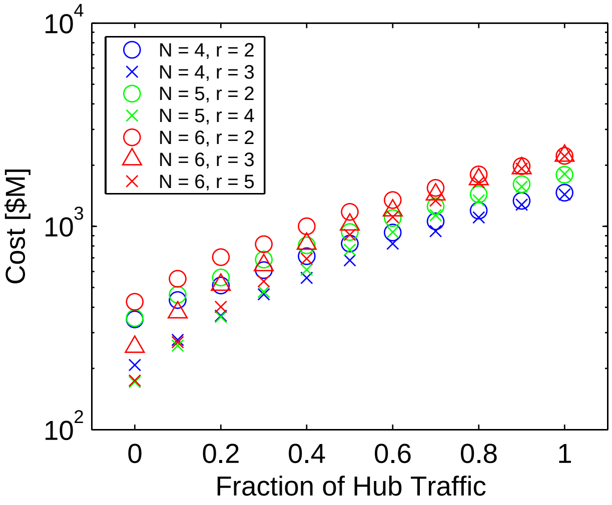


(b) Percentage increase in communications costs MLM.

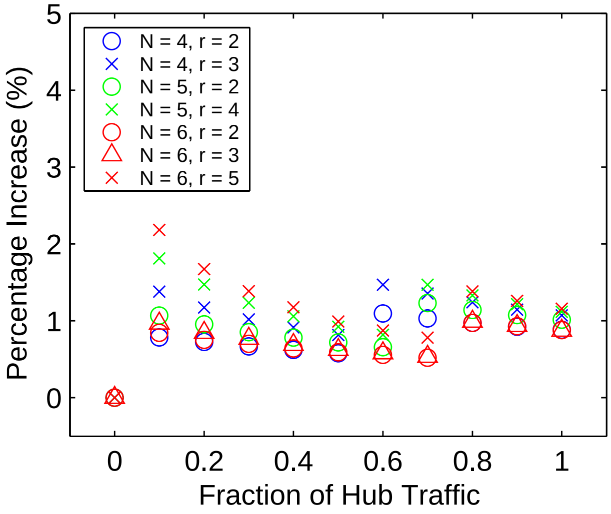


(c) Lowest communications cost constellations MLM.

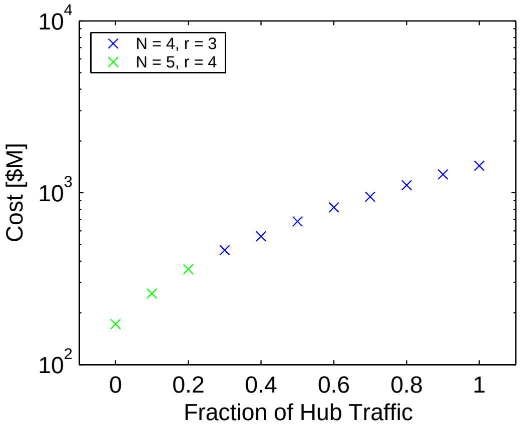
Figure 3-86: Processor-plain node connectivity results for compression rate 25:1 MLM.



(a) Communications costs MLM.

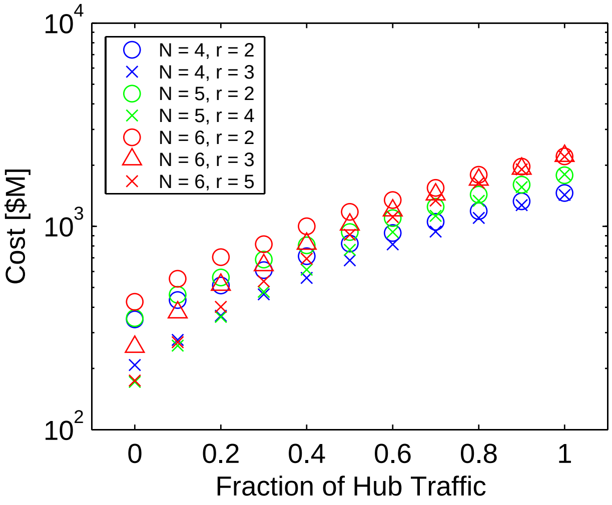


(b) Percentage increase in communications costs MLM.

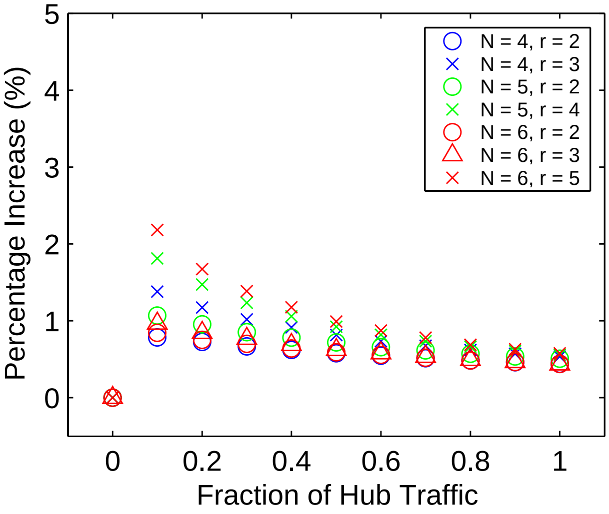


(c) Lowest communications cost constellations MLM.

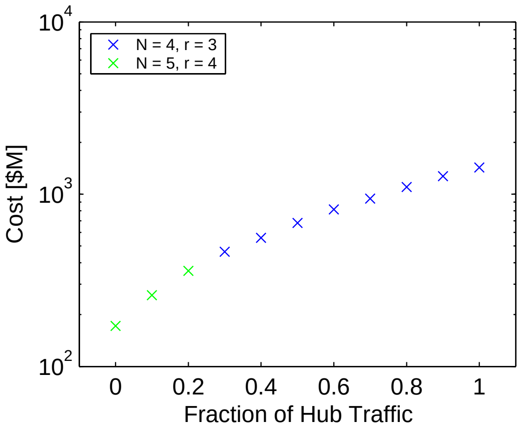
Figure 3-87: Processor-plain node connectivity results for compression rate 50:1 MLM.



(a) Communications costs MLM.



(b) Percentage increase in communications costs MLM.



(c) Lowest communications cost constellations MLM.

Figure 3-88: Processor-plain node connectivity results for compression rate 100:1 MLM.

Note that figures in Case VI and Case VIII do not include any hub constellations. Connecting processing resources to plain nodes in hub constellations will increase traffic routing complexity as data from source nodes must pass through the hub node in order to reach the processing resource. The modeling of this traffic is not isomorphic to the previous traffic models used in these sections. The average minimum hop distance for uniform all-to-one traffic to the plain node (hub$'$) will increase from 1 hop to 2 hops. Furthermore, due to this primary change in traffic pattern, communications costs will increase in both cases where information backflow is either destined for the true hub node or the original source nodes. Connecting processing resources to plain nodes will always require a higher cost than connected processing resources to the hub node.

3.2.2.3 Multiple Processing Resources

As the results have shown that it is more cost-effective to connect a processing satellite to the hub node that disseminates information to the ground, analyses of connecting multiple processing resources will consider processor-hub connectivity only. The question of interest then is how to connect multiple processing satellites to the hub node. Multiple processing resources can be connected to the network in two ways as shown in Figure 3-89. In the first case, each processing satellite requires its own intersatellite link to the hub node. This design is not scalable as the hub node must be designed with additional antennas to make each connection. The number of processing satellites cannot be increased in this case. For greater flexibility and scalability, a network of processing satellites can be connected to the hub relay node by using one of the processing satellites as the gateway to the backbone network. The choice of topology for the network of processing satellites is beyond the scope of this section. In any case, the advantage of a network of processing resources allows for the addition of new processing satellites without requiring any changes to the hub node.

While Figure 3-89 only shows processing connectivity to a single hub node in the network backbone constellation, distributed processing resources may also be considered. In scenarios of distributed processing resources, multiple backbone relay nodes may have processing satellites connected to them. These scenarios open up many techniques for traffic routing design of information that require data processing. Due to the increase in complexity of traffic routing, the analysis of these scenarios are considered to be beyond the scope of this section.

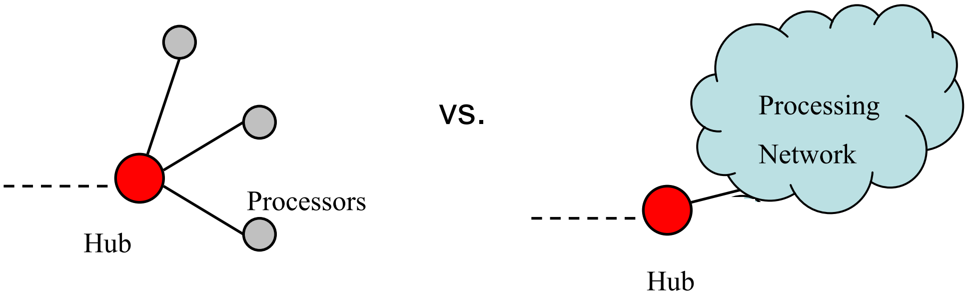


Figure 3-89: Processing Satellite Connectivity Options.

3.2.3 Discussion

Connectivity of processing resources to various space-based network backbone constellations has been examined in terms of communications costs. The traffic models developed in Section 3.1.5 remain valid for studying the traffic flows of transmitting information to the processing satellites and the various data dissemination options in Section 3.2.2. Communications costs on the backbone network increases as the total traffic volume increases due to backflow; the number of wavelengths on the intersatellite links must increase and larger switches must be used. The amount of additional wavelengths required is a function of the compression ratios that can be achieved by the processing satellites. Lower cost increases are seen when data compression rates are high. For minimal increase in the communications cost of the space-based network backbone, it is recommended that processing resources be connected to the hub node(s) that disseminates information to the ground.

3.3 Summary

This chapter has explored some architectural design considerations for building a space-based information network backbone. The backbone constellation is the essential building block to providing networking interconnections between space-based, distributed, and shared on-orbit processing resources. The objective of minimizing communications cost is used to analyze feasible satellite constellation solutions. Design choices (e.g., using uniform antennas and/or uniform satellites in the constellation) have been made with respect to their impact on the ability of the system to grow, in terms of usability, flexibility, scalability and cost.

Regardless of whether networked space-based information processing resources are available in the future, the concept of a space-based information network backbone remains central as networking is an efficient method of sharing communications among multiple users. The space-based information network backbone is designed to meet a number of mission requirements (e.g., high data rates, high connectivity, and low latency) at the least possible cost by modeling the communications costs and analyzing several constellation topologies under different traffic models. The topologies considered include connected circulant constellations and hub constellations with either 1-hub or 2-hubs. Constellation sizes are constrained to be $3 \leq N \leq 20$. A minimum of 3 satellites in GSO is required to provide near-global Earth coverage. Traffic models included uniform all-to-all, uniform all-to-one, and a mixture of both types of traffic.

The calculated communications cost results have indicated that with uniform all-to-all traffic, rings are the least cost-effective while for small constellation sizes ($3 \leq N \leq 11$) full mesh topologies are the most cost-effective. Communications costs for hub constellations under uniform all-to-one traffic have a relatively flat cost structure, with the least expensive topology type belonging to 2-hub constellations. Costs will increase when uniform all-to-one traffic is placed on connected circulant constellations because the network infrastructure that is built is under-utilized (e.g., it has the capacity to handle more traffic than given). Costs will also increase when uniform all-to-all traffic is place on hub constellations because the traffic routing in terms of hops as increased thereby requiring more wavelengths on the intersatellite links and larger switches. Analysis of communications costs for mixed traffic have indicated that for low levels of hub traffic, connected circulant constellations provide lower costs. Lower communications costs is obtained with hub constellations when there are high levels of hub traffic. Figure 3-53 indicates that the point of transition for the recommended architecture type occurs when the amount of hub (uniform all-to-one) traffic is greater than 30% of the total traffic.

The results presented hold when switches with a non-linear cost structure are used, i.e., when the technology constant of a switch is greater than 1. The switch with non-linear cost drives the connectivity of the recommended constellation architecture. Constellations with greater connectivity are more cost-effective. When a switch with linear cost is used, it can be observed that full mesh constellations are no longer the most cost-effective. Other cost factors (e.g., antenna costs) play a larger role in the cost-effectiveness of the constellations. In cases of mixed traffic and the use of a switch with linear cost, hub constellations are recommended whenever there is any amount of uniform all-to-one traffic.

The question of designing a space-based information network backbone to satisfy two user communications with different make-ups of traffic was analyzed next. If both communities have equal volumes of traffic, generally a single satellite backbone constellation system can be designed to accommodate them. If the two traffic patterns are very disparate (e.g., one group has very high levels of hub traffic while the other group has very low levels of uniform traffic), then two separate satellite backbone constellations is more cost-effective. From Figure 3-57, this occurs in the MLM cost scenario when one user community has hub traffic greater than 40% while the other user community has hub traffic less than 30%. The same phenomenon can be seen when the two user communities have unequal volumes of traffic. With a single satellite backbone constellation system, the recommended architecture can be interpreted as designing the backbone constellation mainly for the larger user community and having the small user community "piggyback" on the system.

Choosing a constellation based on the lowest communications cost may not be the optimal choice as launch costs can play a significant role. When launch costs make up a substantial portion of the total system cost, it may be necessary to choose a more expensive constellation (with more satellites) to provide redundancy in the case of launch or satellite failures. More detailed analyses can be performed to balance a trade between risk and cost.

With the design of the space-based information network backbone completed, the next step is to consider the connectivity of processing resources. Processing satellites can be connected to the designated hub node in the constellation or to any plain node. A minimal amount of communications costs is observed if processing satellites are connected to the hub node that disseminates processed or compressed data to the ground. A greater increase in communications costs is observed when processed or compressed data must be returned to the original source nodes or other source nodes requiring more wavelengths on the intersatellite links and larger switches in the backbone nodes. Processor connectivity to any plain node requires more complex traffic routing back to the hub node or source nodes for data dissemination and thus are not attractive candidates.

Chapter 4

Networked Space Processing Applications

An increasing number of space-based high-end engineering applications are anticipated. Examples of high-end engineering applications include space-based radar, adaptive beam forming of sparse arrays distributed over several satellites, optical remote sensing involving multispectral and hyperspectral imaging, and high-capacity data communications satellites. Many high-end engineering applications rely on the Fast Fourier Transform (FFT) algorithm. The FFT algorithm is useful for solving linear partial differential equations, convolution, time series and wave analysis, digital signal processing, and image filtering. As volumes of data generated in space are increasing at astronomical rates, providing shared on-orbit processing resources can help to increase computational capabilities of multiple mission satellites.

This chapter explores how decoupled, shared space-borne processing can be designed to support FFT-intensive applications for SIGINT and SAR. GPPs and FPGAs are used to demonstrate how the two different applications can be effectively supported. The flexibility to handle multiple applications with the same processing architecture demonstrates the value and cost-effectiveness of having shared processing resources. Shared space-borne processing assets thus should be seriously considered for future space missions. Additionally, the architectural implications of satellite networking and networked space-borne processing are explored. The use of optical links as a backbone or as high-speed entrance links is the first level of transformational space communications and networking. Optical satellite communications allows the concept of satellite networking to become economically viable. Further transformational applications and services can be created by using high fidelity analog transmissions and space-borne processing resources. Several suggestions for the new dimensions of space system architectures enabled by an optical satellite network include: on-orbit upgradeable network resources, interoperable distributed space communications, multiplatform distributed space communications, coherent distributed space sensing, multisensor data fusion, and restoration of disconnected global networks. Each of these applications is briefly examined.

4.1 Processing Architectures

Digital signal processing dates back to the 1960s with the use of mainframe digital computers for number-crunching applications using FFTs. Digital signal processing techniques became widespread with the development of the microprocessor in the late 1970s and early 1980s. The architecture design space of processors considered for the processing satellites include: GPPs, Digital Signal Processors (DSPs), FPGAs, and Application-Specific Integrated Circuits (ASICs). The characteristics of flexibility, performance, and energy efficiency are shown in relative terms in Figure 4-1. Note that processing speed is not the key feature. Factors that must be traded off against one another include minimizing system cost, power consumption, required memory capability, chip size, and the effort needed to develop hardware and software applications. A summary of the different characteristics, advantages, and disadvantages of each processor type is provided in Table 4.1.

GPPs have mature compilers and operating systems which make them easy to learn and easy to use. The ability to map an assortment of applications onto GPPs make them highly flexible, but they are not generally appropriate for the numerically-intensive requirements of digital signal processing. Because commercial GPPs depend on speculative execution and memory hierarchies with non-deterministic latency, the execution latency on individual processes/threads is hardly ever tightly bounded. Hard real-time requirements cannot be assured for high-performance applications. Additionally, the high power consumption of GPPs can restrict the system packing density, reduce systems reliability, and increase total system cost [21].

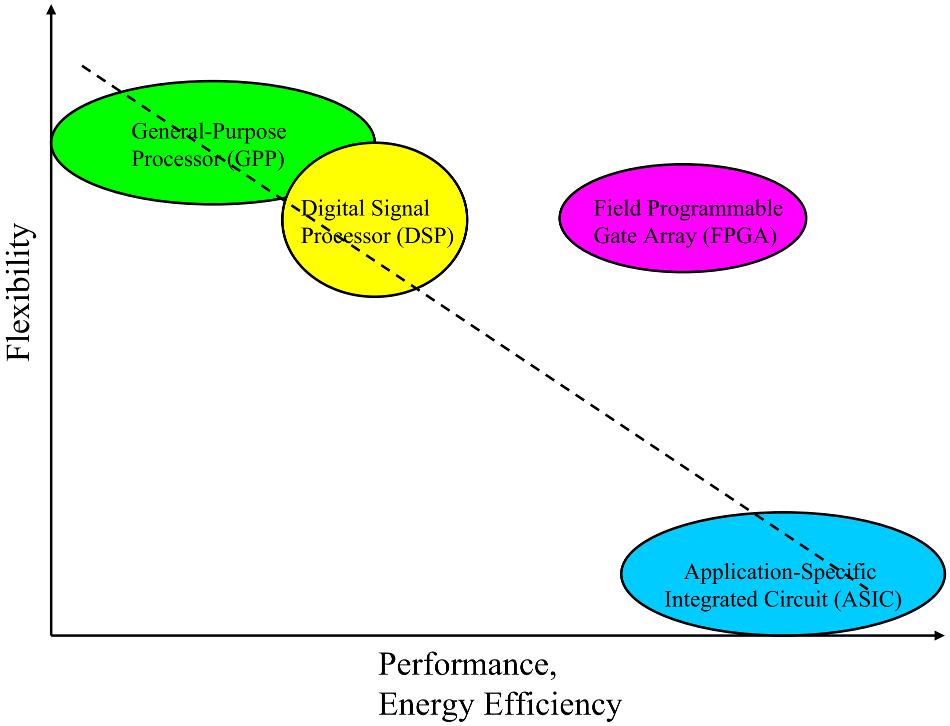
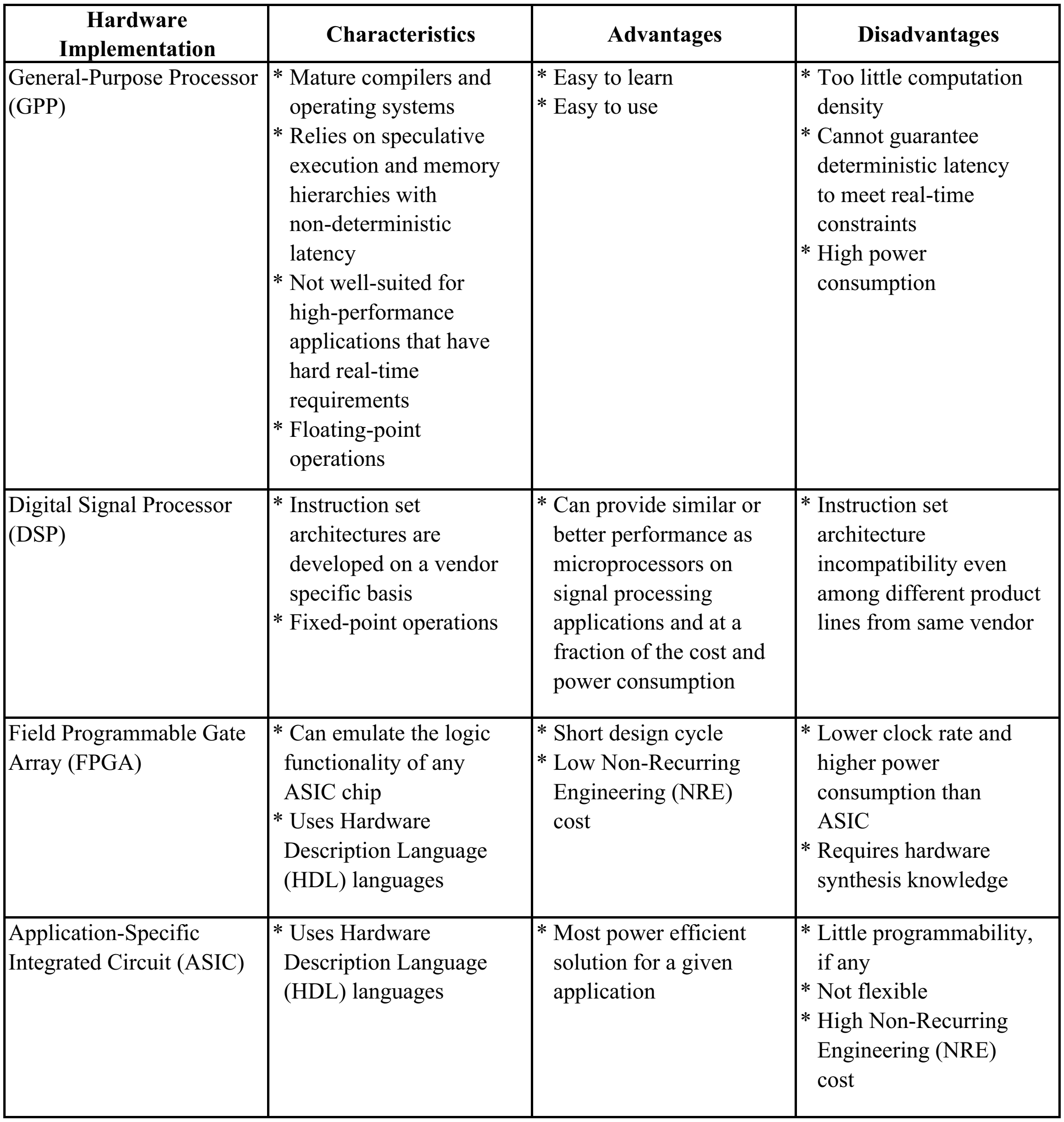


Figure 4-1: Design space of processing architectures.



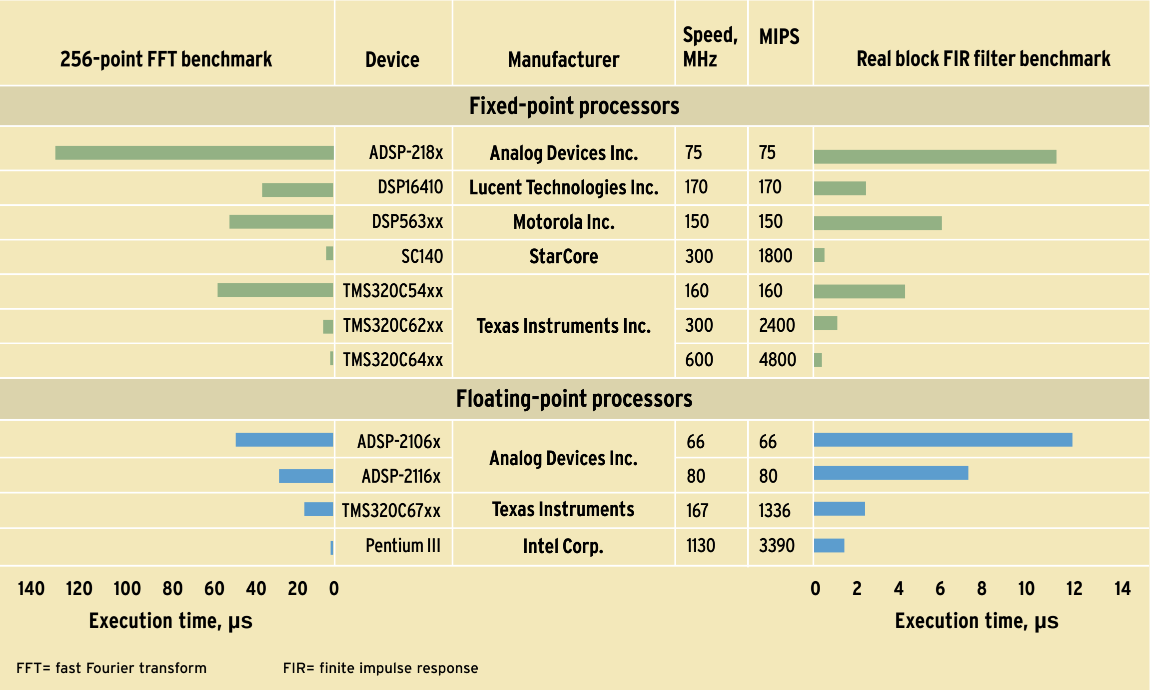
Hardware Implementation	Characteristics	Advantages	Disadvantages
General-Purpose Processor (GPP)	* Mature compilers and operating systems * Relies on speculative execution and memory hierarchies with non-deterministic latency * Not well-suited for high-performance applications that have hard real-time requirements * Floating-point operations	* Easy to learn * Easy to use	* Too little computation density * Cannot guarantee deterministic latency to meet real-time constraints * High power consumption
Digital Signal Processor (DSP)	* Instruction set architectures are developed on a vendor specific basis * Fixed-point operations	* Can provide similar or better performance as microprocessors on signal processing applications and at a fraction of the cost and power consumption	* Instruction set architecture incompatibility even among different product lines from same vendor
Field Programmable Gate Array (FPGA)	* Can emulate the logic functionality of any ASIC chip * Uses Hardware Description Language (HDL) languages	* Short design cycle * Low Non-Recurring Engineering (NRE) cost	* Lower clock rate and higher power consumption than ASIC * Requires hardware synthesis knowledge
Application-Specific Integrated Circuit (ASIC)	* Uses Hardware Description Language (HDL) languages	* Most power efficient solution for a given application	* Little programmability, if any * Not flexible * High Non-Recurring Engineering (NRE) cost

Table 4.1: Characteristics of processing architectures.

Special purpose DSPs can provide higher performance for specific algorithms than GPPs because DSPs have datapaths and instruction sets that are tuned to the computational requirements of signal processing. On the other hand, some GPPs now have the performance capability to compete with DSPs in digital signal processing applications. Benchmark tests developed by Berkeley Design Technology, Inc. gauge the execution time for several different processors running digital signal processing algorithms, as shown in Table 4.2. Performance is measured by the processor speed and the number of instructions per second. The DSP architecture types evaluated include: (1) conventional single-issue (e.g., Analog Devices' ADSP-218x and -2106x; Motorola DSP563xx; and Texas Instruments' TMS320C54xx), (2) enhanced-conventional single-issue (e.g., the ADSP-2116x and Lucent DSP164xx), (3) VLIW (Very Long Instruction Word) (e.g., StarCore's SC140, plus the TMS320C62xx, 'C64xx, and 'C67xx), and (4) high-performance superscalar GPP with SIMD (Single Instruction Multiple Data) enhancements (e.g., Intel Pentium III). Note that the digital signal processing benchmark results for the Intel Pentium III at 1.13 GHz are faster than the results for all but the fastest DSPs. This illustrates the point that enhanced GPPs can compete with DSPs. Although DSPs continue to offer advantages such as lower power consumption, peripherals oriented for digital signal processing, and execution time predictability, they lack good software development tools (e.g., compilers). GPP vendors tend to support software compatibility between processor generations whereas DSP vendors do not. Thus, DSP users are required to learn a new architecture, new tools, and to rewrite their software in order to upgrade to a newer, faster DSP.

At the opposite end of the hardware implementation spectrum from GPPs are ASICs. An ASIC chip is designed for a specific application, thus offering a fixed functionality. Through efficiently designed custom logic, an ASIC device delivers the best performance and energy efficiency possible for a given application. However, an ASIC solution lacks programmability because it is hardwired for one specific task. FPGAs can bridge the gap in performance, energy efficiency, and flexibility between ASICs and GPPs.

FPGAs are similar to ASICs but offer programmability which gives them the capability to handle a variety of different applications. The logic functionality of any ASIC device can basically be imitated by FPGAs which have bit-level logic and interconnect reconfigurability. Similar to ASICs, FPGA designs use HDL (Hardware Description Language) languages (e.g., Verilog, VHDL) as their main programming model which provides high



256-point FFT benchmark	Device	Manufacturer	Speed, MHz	MIPS	Real block FIR filter benchmark
		Fixed-point processors			
	ADSP-218x	Analog Devices Inc.	75	75	
	DSP16410	Lucent Technologies Inc.	170	170	
	DSP563xx	Motorola Inc.	150	150	
	SC140	StarCore	300	1800	
	TMS320C54xx	Texas Instruments Inc.	160	160	
	TMS320C62xx		300	2400	
	TMS320C64xx		600	4800	
		Floating-point processors			
	ADSP-2106x	Analog Devices Inc.	66	66	
	ADSP-2116x		80	80	
	TMS320C67xx	Texas Instruments	167	1336	
	Pentium III	Intel Corp.	1130	3390	

FFT= fast Fourier transform FIR= finite impulse response

Source: [31] Jane Eyre, "The digital signal processor derby," *IEEE Spectrum*, vol. 38, June 2001, p. 67.

Table 4.2: Digital signal processing benchmarks.

device efficiency. Although FPGAs typically have a lower clock rate and higher power consumption, they are a good alternative to ASICs for low or medium volume products that do not have strict low power requirements. The advantages of FPGAs include the in-system-programmability, short design cycle, and low non-recurring engineering (NRE) cost. NRE costs are the one-time engineering costs associated with a project. The disadvantage of using FPGAs is that they do not have generalized high-level programming models or standard systems architectures, thus requiring hardware synthesis knowledge to maximize performance. Additionally, FPGA-based systems have not been widely used thus far. Meanwhile, there is growing interest in using FPGA-based systems for high-end reconfigurable computing because high computational throughput can be realized with large parallel functional units [21].

4.1.1 Parallel Processing and Distributed Processing

The demand for increasing computing power is a classic reason for the use of parallel and distributed computing. Performance is important for scientific and engineering applications (e.g., climate and weather monitoring, surveillance, and astrophysical observations and models) as well as commercial applications (e.g., Internet multimedia and database applications). Parallel and distributed computing allow multiple computational activities (processes) to run at the same time and even cooperate with each other because multiple processors are used and the processors are interconnected by some network. The difference between parallel and distributed computing can be summarized as follows: *parallel computing* splits an application up into tasks that are executed *at the same time*, whereas *distributed computing* splits an application up into tasks that are executed *at different locations, using different resources* [54].

Using processors in large quantities in a parallel machine can be cost-effective. For the same level of performance, parallel computers are cheaper to build than sequential computers. Higher levels of performance can even be obtained by using older processors in a parallel computer rather than the fastest and most expensive processor available at a given time. Sequential architectures are limited in overall performance because of limits in access time to memory (the amount of time it takes to access data and instructions from the working memory). Parallel and distributed computing will help alleviate the access time problem of both main memory and disks. Increasing the number of processors can lead to an increase in cache and main memory capacities. Allowing data to be processed closer to the location where they are generated can reduce traffic to a system-wide memory.

4.1.2 Parallel Processing Architectures

The many levels of parallelism have been classified by Flynn in [33] into the following taxonomy for computers [26]:

- *Single instruction stream, single data stream* (SISD): serial computers (uniprocessor).
- *Single instruction stream, multiple data streams* (SIMD): involves multiple processors performing the same instruction on different data at the same time.
- *Multiple instruction streams, single data stream* (MISD): involves multiple processors performing different instructions to a single datum.
- *Multiple instruction streams, multiple data streams* (MIMD): involves multiple processors performing different instructions on different data in parallel.

Although various modern computers are hybrids and do not fit neatly into the above categories, the classification has endured because it is simple, easy to understand, and provides a good first-order approximation. To appreciate the breadth of parallel processing architectures, see Figure 4-2. Of specific interest, for the design of the processing satellite, are MIMD architectures. MIMD architectures naturally lend themselves to handling multiple applications simultaneously.

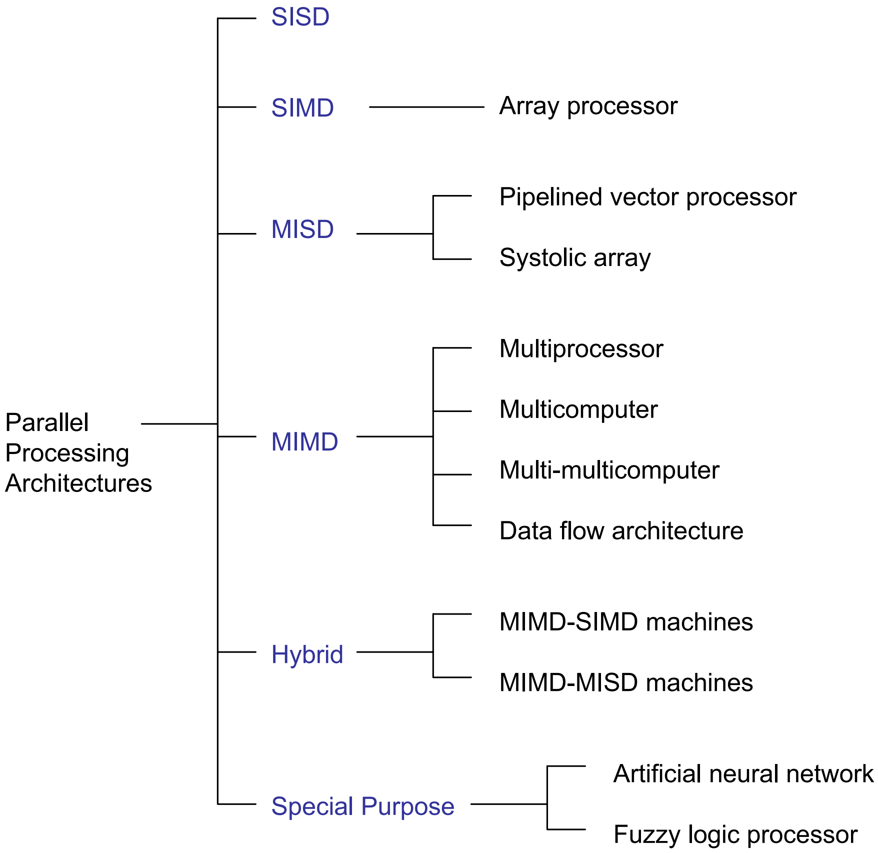
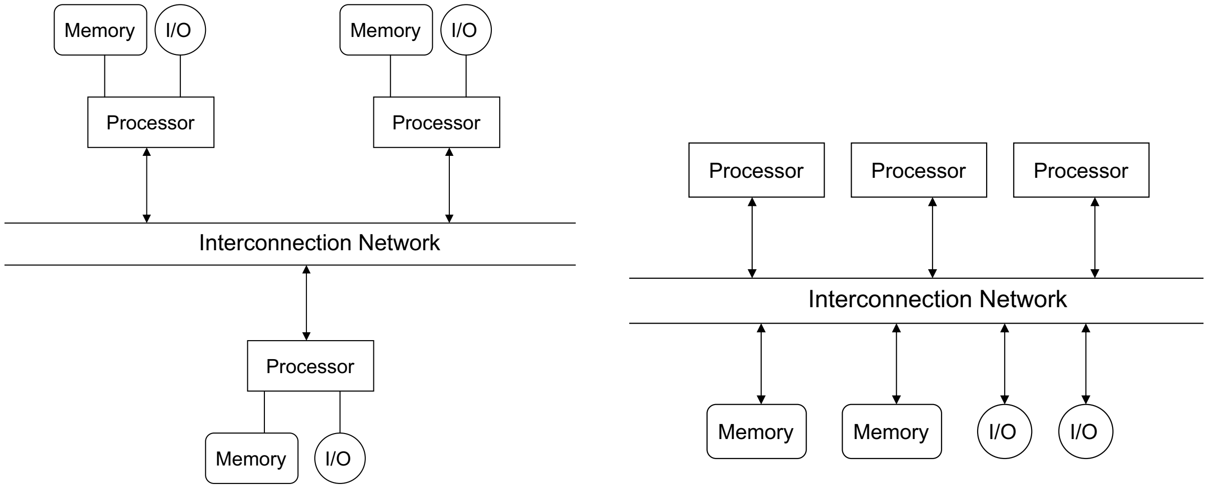


Figure 4-2: Parallel processing architectures [56].

4.1.2.1 MIMD Architecture

MIMD architectures make use of multiple processors that can perform independent instruction streams utilizing local data. MIMD architectures are synchronous computers, characterized by decentralized hardware control, where processes communicate by passing messages through an interconnection network or by accessing data in shared memory units. In MIMD architectures with no shared memory, processing nodes consisting of an autonomous processor and its local memory are connected with a processor-to-processor interconnection network. In MIMD architectures with shared memory, all processors have equal access to the memory. The general architecture of a distributed memory MIMD is illustrated in Figure 4-3(a) and the general architecture of a shared memory MIMD is illustrated in Figure 4-3(b).

The cost-effectiveness of an n-processor system over n single-processor systems has encouraged MIMD research and experimentation. MIMD computers can support parallel solutions that call for processors to work principally in an autonomous manner. Current research endeavors include designing multiprocessor architectures that will scale (accommodate a considerable increase in processors) and will fulfill the performance requirements of large scientific applications characterized by local data references. MIMD architectures with shared memory do not scale well. The disadvantages of shared memory is the high contention for the memory and the speed of dynamic random access memory (RAM), which is known to have high latency. Therefore, MIMD architectures with distributed memory are chosen for the processing satellites as the objective is to design processing satellites with the greatest flexibility to accommodate a variety of applications and algorithms.



(a) Distributed memory MIMD.

(b) Shared memory MIMD.

Figure 4-3: General MIMD architectures.

4.1.2.2 Interconnection Network

The interconnection network is the arrangement of pathways over which nodes communicate with each other. Nodes can be processors, memories, or switches. The interconnection network topology chosen should provide efficient performance for parallel programs with different interprocessor communication patterns. Networks are traditionally characterized by the following parameters [78]:

- *Node degree:* the number of edges adjoining a node. Networks with a small node degree are favored for their low cost. Networks can be scalable if the node degree is independent of the network size.

- *Network diameter:* the longest shortest path between any two nodes in the network. A small network diameter achieves low latency.

- *Bisection width:* the minimum number of links to cut that separates the network into two equal halves. The bisection width can be defined by the relationship:

  $$\text{bisection width} = \frac{\text{bisection bandwidth}}{\text{link width} \times \text{link rate}} \tag{4.1}$$

  where the bisection bandwidth is the rate at communication can take place between one half of the compute system and the other half, the link width is the number of wires in each link and the link rate is the transmission speed of each wire in the link. Equation 4.1 reflects the network bandwidth and the wiring density of the network.

- *Edge connectivity:* the minimum number of edges (or links) whose deletion will cause a connected graph to be disconnected. High connectivity is desired for better reliability and availability (less contention).

- *Cost:* the number of communication links in the network.

The interconnection network chosen should transmit a maximum number of messages in the shortest time with maximal reliability and minimum cost. Various tradeoffs must be made among these opposing requirements. Several common interconnection network topologies are characterized in Table 4.3, where N indicates the network size (the number of nodes in the network). Each interconnection network topology is briefly discussed.

Linear Array. The linear array (or bus) is simple to implement. It has the advantages of having a small node degree and low cost. On the other hand, the large network diameter and low bisection width make the linear array not scalable. Edge connectivity of 1 is undesirable. An example of a linear array is illustrated in Figure 4-4(a).

Ring. With one extra link, the ring divides the linear array network diameter by 2 and has twice as much bisection width and edge connectivity. Rings also have the advantages of a small node degree and low cost. An example of a ring is illustrated in Figure 4-4(b).

Star. A star network has a central hub node connected to a number of leaf nodes. The hub node can be complex because it can have a high node degree. Because all traffic must pass through the hub node, the network diameter is a constant. Like the linear array, a star has poor bisection width and edge connectivity. A star network is not scalable and fault tolerant. While defective leaf nodes may be removed with no disruption to the network, failure of the hub node halts the entire network. An example of a star is illustrated in Figure 4-4(c).

Binary Tree. In a binary tree, the node degree remains constant and the network diameter increases as logarithmically. Despite the fact that the tree network is scalable, it has low bisection width and edge connectivity. An example of a binary tree is illustrated in Figure 4-4(d).

Full Mesh. A full mesh is a completely connected network (i.e., any two nodes are directly connected), thus the network diameter is equal to one. The high node degree and cost have made this type of network unreasonable for constructing massively parallel machines.

2-D Mesh. The 2-D mesh can be connected in a number of ways. An example of a 4-connected 2-D mesh is illustrated in Figure 4-4(f). Each node is linked to its North, South, East, and West nearest neighbors. Compared to the binary tree, the 2-D mesh has a larger node degree, smaller network diameter, larger bisection width, and larger edge connectivity.

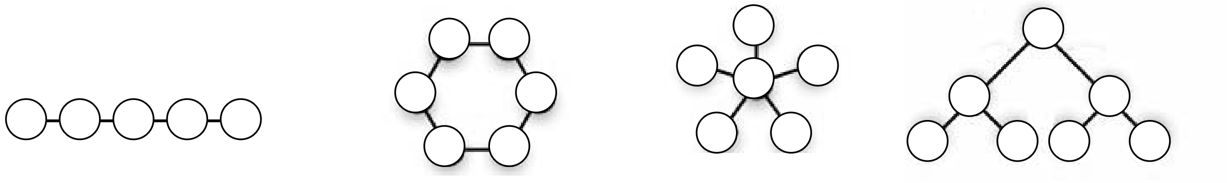
2-D Torus. The 2-D torus is a 2-D mesh with extra links connecting the ends of each row and column, as illustrated in Figure 4-4(g). The extra links make the node degree of all nodes equal, reduce network diameter, increase bisection width, increase edge connectivity, and increase cost.

3-D Mesh. Compared to the 2-D mesh, the 3-D mesh increases the node degree, reduces the network diameter, increases the bisection width, increases the edge connectivity, and increases the cost. A 3-D mesh is illustrated in Figure 4-4(h).

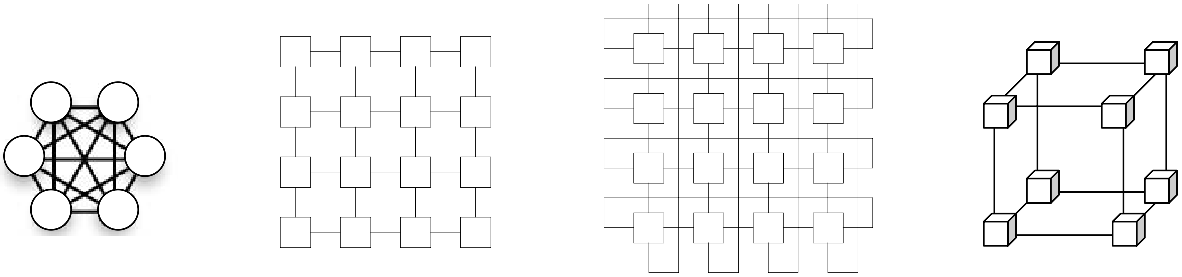
Hypercube. A hypercube (or Boolean n-cube) is a binary n-cube network that is comprised of $N = 2^n$ nodes where n is the number of dimensions. Each node has $n = \log_2 N$ bidirectional links to adjacent nodes. Thus, there are two nodes in each dimension. A linear array with 2 nodes is a 1-cube network. A 2-D mesh with four nodes can be considered as a 2-cube network. The 3-D mesh can be viewed as a 3-cube network. A 4-cube network is complex, but may be represented as two connected 3-cube networks. Hypercube topologies are shown in Figure 4-4(i). With n-bit numeric values ranging from 0 to $N - 1$, individual nodes can be distinctively labeled. The numeric values are assigned in such a way that the values of adjacent nodes differ by a single bit. The hypercube interconnection network has a non-constant node degree, small network diameter, large bisection width, and large edge connectivity. It, however, also does not scale well. A hypercube network with a large number of nodes can be very complicated and costly.

Network Topology	Node Degree	Network Diameter	Bisection Width	Edge Connectivity	Cost (Number of Links)
Linear Array (Bus)	1 (end nodes) 2 (other nodes)	$N-1$	1	1	$N-1$
Ring	2	$\frac{N}{2}$	2	2	$N-1$
Star	1 (leaf nodes) $N-1$ (hub node)	2	1	1	$N-1$
Binary Tree	1 (left nodes) 2 (root node) 3 (other nodes)	$2 \log \frac{N+1}{2}$	1	1	$N-1$
2-D Mesh (4-connected)	2 (at corners) 3 (at edges) 4 (others)	$2(\sqrt{N}-1)$ (square mesh)	$\sqrt{N}$	2	$2(N-\sqrt{N})$
Full Mesh	$N-1$	1	$\frac{N^2}{4}$	$N-1$	$\frac{N(N-1)}{2}$
2-D Torus	4	$\sqrt{N}$	$2\sqrt{N}$	4	$2N$
3-D Mesh (Cube)	3 (at corners) 4 (at edges) 5 (on faces) 6 (internal)	$3(N^{1/3}-1)$	$N^{2/3}$	3	$2(N-N^{2/3})$
Hypercube	$\log_2 N$	$\log_2 N$	$\frac{N}{2}$	$\log N$	$\frac{N \log_2 N}{2}$

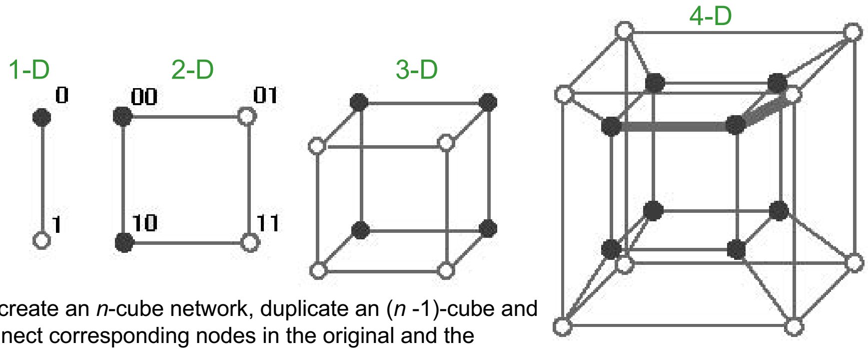
Table 4.3: Comparison of common interconnection network topologies [78].



(a) Linear Array [Bus] (b) Ring (c) Star (d) Binary Tree



(e) Full Mesh (f) 2-D Mesh [4-connected] (g) 2-D Torus (h) 3-D Mesh [Cube]



To create an *n*-cube network, duplicate an (*n* -1)-cube and connect corresponding nodes in the original and the duplicate. (Original nodes are shaded.)
Source: http://csep1.phy.ornl.gov/gif_figures/caf11.gif

(i) Hypercubes

Figure 4-4: Common interconnection topologies.

The high performance interconnection network should have the properties of small network diameter, symmetric topology, high bisection width, and routing simplicity. A small network diameter allows for low communication latency. A symmetric topology allows for constant node degree which subsequently allows for universality which can amortize design cost. Large bisection width increases the wiring density but implies high bandwidth. Furthermore, the interconnection network should have high connectivity and fault tolerance and support architectural expandability.

4.1.2.3 Bus Design

Processors can be connected to one another with a bus interface. In a computer system, the bus is a standard interconnection method that connects the central processing unit (CPU) to external memory and peripheral devices. The bus organization has the advantages of low cost and flexibility because new devices can be added. A bus has the disadvantage of creating a communication bottleneck, bounding the maximum input/output (I/O) throughput (otherwise known as I/O bandwidth). To satisfy the demands of high processing performance is a design challenge because the bus system must be capable of handling high I/O rates. The maximum bus speed is primarily limited by physical factors. The length of the bus and the number of devices prevent arbitrary bus speedup. It is desirable for the bus to have high I/O throughput and high I/O rates (low latency).

Design options for designing a bus system is summarized in Table 4.4. Decisions will depend on the trade between performance and cost. Bus systems with high performance have high cost, for example, use separate address and data lines, wider data lines, and multiple-word transfers. Bus designs that have been used in the personal computer include ISA (Industry Standard Architecture), EISA (Extended Industry Standard Architecture), Micro Channel, VESA (Video Electronics Standards Association) Local-Bus (also known as VL-Bus) and PCI (Peripheral Component Interconnect). Other peripheral busses include NuBus, TURBOchannel, VME (VersaModule Eurocard) bus, MULTIBUS and STD bus.

The clock speed and bus width are factors that will affect cost, power, and technology requirements. The peak transfer rate (potential I/O performance) is determined by the bus width and clock rate:

$$\text{I/O bandwidth} = \text{clock speed} \times \text{bus width} \tag{4.2}$$

Option	**High performance**	**Low cost**
Bus width	Separate address and data lines	Multiplex address and data lines
Data width	Wider is faster (e.g., 64 bits)	Narrower is cheaper (e.g., 8 bits)
Transfer size	Multiple words have less bus overhead	Single-word transfer is simpler
Bus masters	Multiple (requires arbitration)	Single master (no arbitration)
Split transaction?	Yes – separate request and reply packets get higher bandwidth (need multiple masters)	No – continuous connection is cheaper and has lower latency
Clocking	Synchronous	Asynchronous

Source: [44] John L. Hennessy and David A. Patterson, *Computer Organization & Design: The Hardware/Software Interface*, San Francisco, California: Morgan Kaufmann Publishers, Inc., 1998, p. 497.

Table 4.4: Options for a bus design.

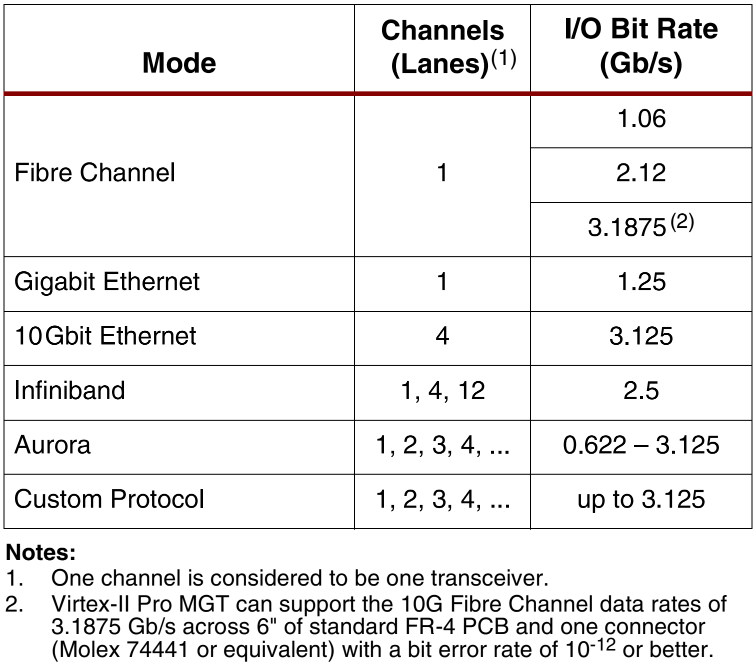
For example, the I/O bandwidth offered by a Intel Pentium III processor with a bus that operates at 133 MHz and a data I/O bus width of 64-bits (8 bytes) is 8.512 Gbps. Therefore, the peak I/O bandwidth required of the interconnection network for connecting N number of Intel Pentium III processors is $8.512N$ Gbps. A processing architecture requiring 250 Intel Pentium III processors would require an interconnection network with a burst rate capability of 2.128 Tbps. These numbers indicate an upper bound on the interconnection bandwidth requirement. It is unclear how much bandwidth is actually required by the processors, thus these numbers are an overestimate. Additionally, while the potential I/O performance is dictated by hardware, the operating system and processor command set control how much of that capability is delivered because the software that is involved on all interconnection networks requires some overhead. Network software is another aspect to consider and should be designed to provide resource management and control as well as security/protection measures. Regardless, development of an interconnection network capable of handling several terabits per second is important R&D for the future.

High bandwidth interconnections with FPGAs are made possible in the Virtex-II Pro and Virtex-II Pro X Platform FPGAs through the use of RocketIO and RrocketIO X Multi-Gigabit Transceivers (MGTs) which are flexible parallel-to-serial and serial-to-parallel embedded transceiver cores. The communications standards (protocols) and I/O bit rate supported by the RocketIO MGT and the RocketIO X MGT are provided in Tables 4.5 and

4.6 respectively. The RocketIO MGTs can operate at at any I/O bit rate in the range of 622 Mbps to 3.125 Gbps per channel. The RocketIO X MGTs can operate at any I/O bit rate in the range of 2.488 Gbps to 10.3125 Gbps per channel. The I/O bandwidth required for an FPGAs is:

$$\text{I/O bandwidth} = \text{I/O bit rate} \times \text{number of RocketIO ports} \tag{4.3}$$

The number of RocketIO ports available for the family of Virtex-II FPGAs is listed in Table 4.7. For example, the XC2VPX70 device has 20 RocketIO transceiver blocks, each capable of 10 Gbps SONET OC-192 connections. To interconnect N of these FPGA devices can require a total I/O bandwidth of $200N$ Gbps.



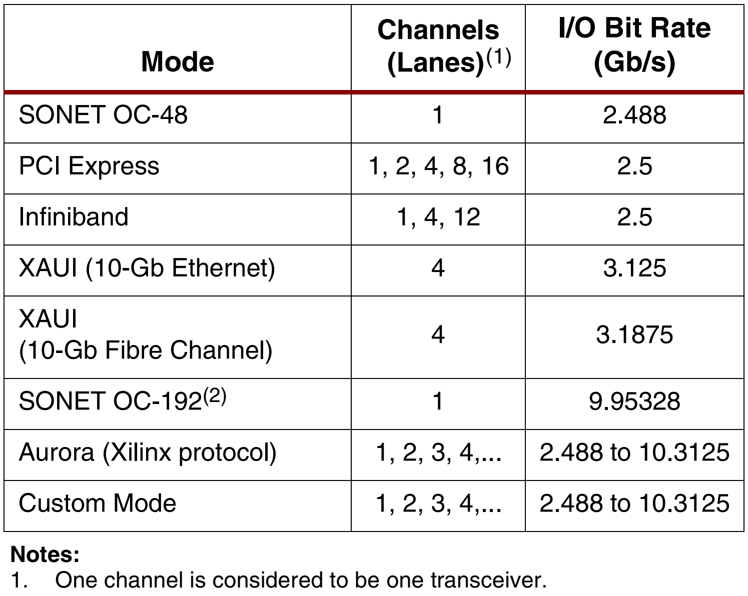
Mode	Channels (Lanes)(1)	I/O Bit Rate (Gb/s)
Fibre Channel	1	1.06
		2.12
		3.1875 (2)
Gigabit Ethernet	1	1.25
10Gbit Ethernet	4	3.125
Infiniband	1, 4, 12	2.5
Aurora	1, 2, 3, 4, ...	0.622 – 3.125
Custom Protocol	1, 2, 3, 4, ...	up to 3.125

Notes:

1. One channel is considered to be one transceiver.
2. Virtex-II Pro MGT can support the 10G Fibre Channel data rates of 3.1875 Gb/s across 6" of standard FR-4 PCB and one connector (Molex 74441 or equivalent) with a bit error rate of 10^{-12} or better.

Source: [93] Xilinx, *Virtex-II Pro and Virtex-II Pro X Platform FPGAs: Complete Data Sheet*, p. 20.

Table 4.5: RocketIO Transceiver.

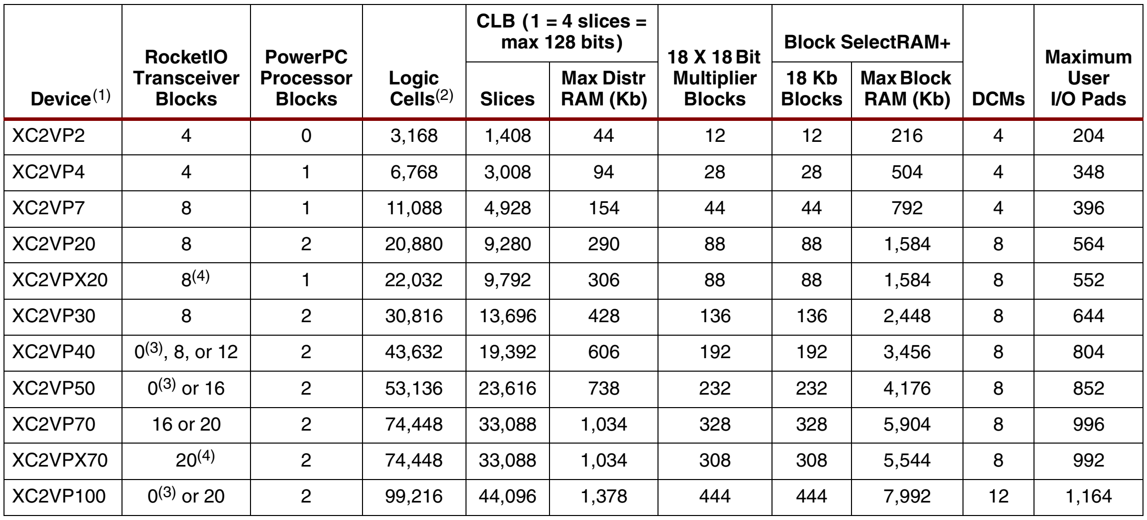


Mode	Channels (Lanes)(1)	I/O Bit Rate (Gb/s)
SONET OC-48	1	2.488
PCI Express	1, 2, 4, 8, 16	2.5
Infiniband	1, 4, 12	2.5
XAUI (10-Gb Ethernet)	4	3.125
XAUI (10-Gb Fibre Channel)	4	3.1875
SONET OC-192(2)	1	9.95328
Aurora (Xilinx protocol)	1, 2, 3, 4,...	2.488 to 10.3125
Custom Mode	1, 2, 3, 4,...	2.488 to 10.3125

Notes:
1. One channel is considered to be one transceiver.

<u>Source</u>: [93] Xilinx, *Virtex-II Pro and Virtex-II Pro X Platform FPGAs: Complete Data Sheet*, p. 12.

Table 4.6: RocketIOX Transceiver.



Device(1)	RocketIO Transceiver Blocks	PowerPC Processor Blocks	Logic Cells(2)	CLB (1 = 4 slices = max 128 bits)		18 X 18 Bit Multiplier Blocks	Block SelectRAM+		DCMs	Maximum User I/O Pads
				Slices	Max Distr RAM (Kb)		18 Kb Blocks	Max Block RAM (Kb)		
XC2VP2	4	0	3,168	1,408	44	12	12	216	4	204
XC2VP4	4	1	6,768	3,008	94	28	28	504	4	348
XC2VP7	8	1	11,088	4,928	154	44	44	792	4	396
XC2VP20	8	2	20,880	9,280	290	88	88	1,584	8	564
XC2VPX20	8(4)	1	22,032	9,792	306	88	88	1,584	8	552
XC2VP30	8	2	30,816	13,696	428	136	136	2,448	8	644
XC2VP40	0(3), 8, or 12	2	43,632	19,392	606	192	192	3,456	8	804
XC2VP50	0(3) or 16	2	53,136	23,616	738	232	232	4,176	8	852
XC2VP70	16 or 20	2	74,448	33,088	1,034	328	328	5,904	8	996
XC2VPX70	20(4)	2	74,448	33,088	1,034	308	308	5,544	8	992
XC2VP100	0(3) or 20	2	99,216	44,096	1,378	444	444	7,992	12	1,164

Notes:
1. -7 speed grade devices are not available in industrial grade.
2. Logic Cell = (1) 4-input LUT + (1) FF + Carry Logic
3. These devices can be ordered in a configuration without RocketIO transceivers.
4. Virtex-II Pro X devices equipped with RocketIO X transceiver cores.

<u>Source</u>: [93] Xilinx, *Virtex-II Pro and Virtex-II Pro X Platform FPGAs: Complete Data Sheet*, p. 2.

Table 4.7: Virtex-II Pro / Virtex-II Pro X FPGA Devices.

4.2 Signals Intelligence (SIGINT) Application

One of the most significant forms of intelligence gathering is acknowledged to be SIGINT. In addition to a nation's diplomatic and economic plans or events, interception of foreign signals can provide scientific information such as the characteristics of a nation's radars, spacecrafts, and weapons systems. SIGINT is currently classified into the following five main categories [72]:

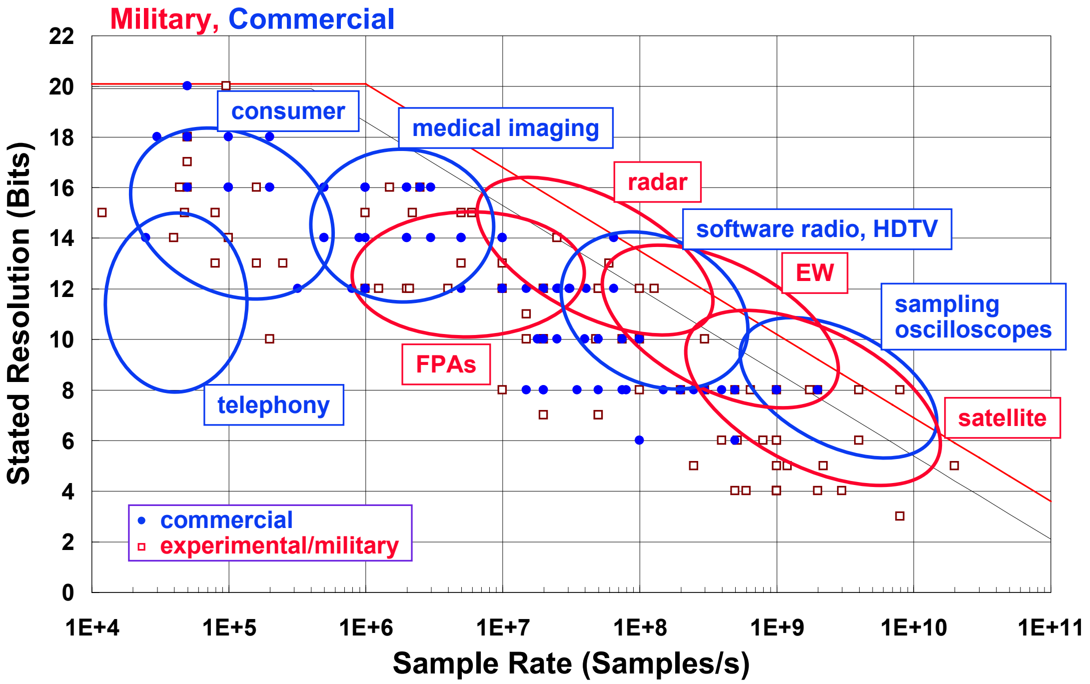
- Communications Intelligence (COMINT) - analysis of the source and content of message traffic, excluding radio and television broadcasts (e.g., voice, Morse code, radioteletype, or facsimile).
- Electronics Intelligence (ELINT) - analysis of non-communications electronic transmissions of military and civilian hardware.
  - Foreign Instrumentation Signals Intelligence (FISINT) - analysis of electromagnetic emissions associated with the testing and operational deployment of subsurface, surface, and aerospace systems (e.g., signals from video data links, beaconing, electronic interrogators, telemetry, and tracking-fusing-aiming/command systems).
  - Telemetry Intelligence (TELINT) - analysis of signals by which a missile warhead, missile, or missile stage, transmits about its performance during a test flight.
- Radar Intelligence (RADINT) - analysis of radar transmitters, not the electronic emanations from the radar, to obtain information such as flight paths, velocity, maneuvering, trajectory, and angle of descent.
- Laser Intelligence (LASINT) - interception of laser communications.
- Non-imaging Infrared - use of sensors that can detect the absence/presence and movement of an object via temperature.

Satellites designed for signals intelligence can detect transmissions from broadcast systems (e.g., radios), radars, and other electronic systems and provide information on the type and location of even low power transmitters, such as hand-held radios. According to [10], the U.S. has SIGINT satellites in LEO, GEO, and elliptical orbits. The first generation of these satellites was launched in the early 1970s. Known as Rhyolite, these satellites had a receiving antenna with a diameter on the order of 10 m. In the late 1970s, the next generation of these satellites was launched. Known as Chalet or Vortex, these satellites had an antenna diameter of several tens of meters. The most current SIGINT satellites, was launched in the mid-1980s. Known as Magnum, these satellites had very large deployable antennas with a diameter of nearly 100 m. Currently under development are satellites with even larger antenna designs. Increasing antenna diameter allows these satellites to identify lower power transmission and to locate a transmitter with greater accuracy.

4.2.1 Spectral Analysis Design Example

Consider a simple example of a satellite receiving or monitoring RF communications from the ground. In this situation, the sensor satellite receives an analog input signal with a bandwidth of X GHz. The aim is to identify frequencies within the X GHz bandwidth that contain signal and not noise, thereby reducing the downlink data requirement. One method of implementation is spectral analysis over the X GHz bandwidth. In order to process analog signals by digital means, it is first necessary to convert them into digital format. ADCs digitize analog signals into a sequence of numbers having finite precision. Figure 4-5 provides a sampling of available ADC technology in the military and commercial sectors. It also shows their capabilities in terms of sampling rate and output resolution in bits for both military and commercial applications.

In this design example, ADCs are implemented on the sensor satellite. Signals are collected and digitized on the sensor satellite and subsequently transmitted in digital format on intersatellite links to the processing satellite. Ideally, the most advanced ADC will be employed. Generally, a minimum number of ADCs onboard the sensor satellite should provide enough data to the processing satellite in an efficient manner while simultaneously minimizing the cost, power, and weight of the sensor satellite. Assuming that 8-bit resolution output and 1 picosecond clock jitter are adequate for example problem, Figure 4-5 indicates that the fastest commercial state-of-the-art ADC samples at approximately 1 GHz. Therefore, to handle an input of X GHz, the number of ADCs required is at least $2X$ as an analog signal of frequency f is digitized at the Nyquist rate $(2f)$ or higher. Figure 4-6 illustrates how the input analog signal is first analog-filtered into several sub-bands such that the ADCs can handle the signals.



Source: [90] Robert H. Walden, Analog-to-Digital Converter Survey and Analysis, Presentation Slides, 16 July 1999, p. 45.

Figure 4-5: Survey of commercial and experimental/military analog-to-digital converters and their applications.

The processing satellite receives the digitized signals from the sensor satellite via optical intersatellite links. A FFT frequency analysis is performed on the digital samples. The FFT algorithm allows a microprocessor system to act as a real-time spectrum analyzer, a digital filter, a digital signal correlator, or a deconvolution system. The objective in this design example is to resolve the signal with a resolution of approximately 1 kHz. The level of signal in each frequency band (or *bin*) is compared against a determined threshold level. Based on this comparison, the processing algorithm determines whether a given band at that particular instant contains signal or only noise. If a frequency bin is found to include elements of the desired signal, it is noted and will be sent on the downlink. Frequency bands that contain only noise will be discarded.

The following end-to-end design cases illustrate how the real-time spectral analysis problem can be solved with various commercially available devices (e.g., GPPs and FPGAs). The efficiency of the FFT analysis depends on the algorithm and its performance on a processor. There also exist bottlenecks with the real-time computational efficiency of the processors in the processing satellite. The main bottleneck is the system bus speed to trans-

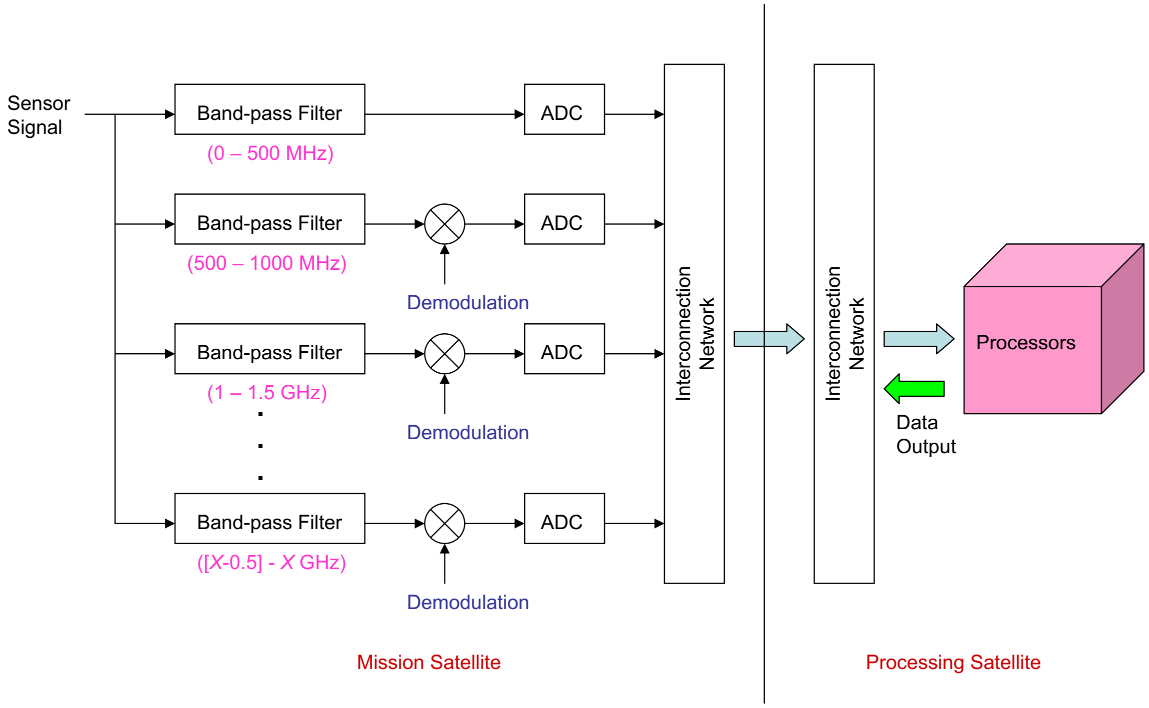


Figure 4-6: Example block diagram of data flow for signal intelligence application.

mit data to the processors and memory. These constraints and bottlenecks will dictate what is considered to be a reasonable end-to-end solution design.

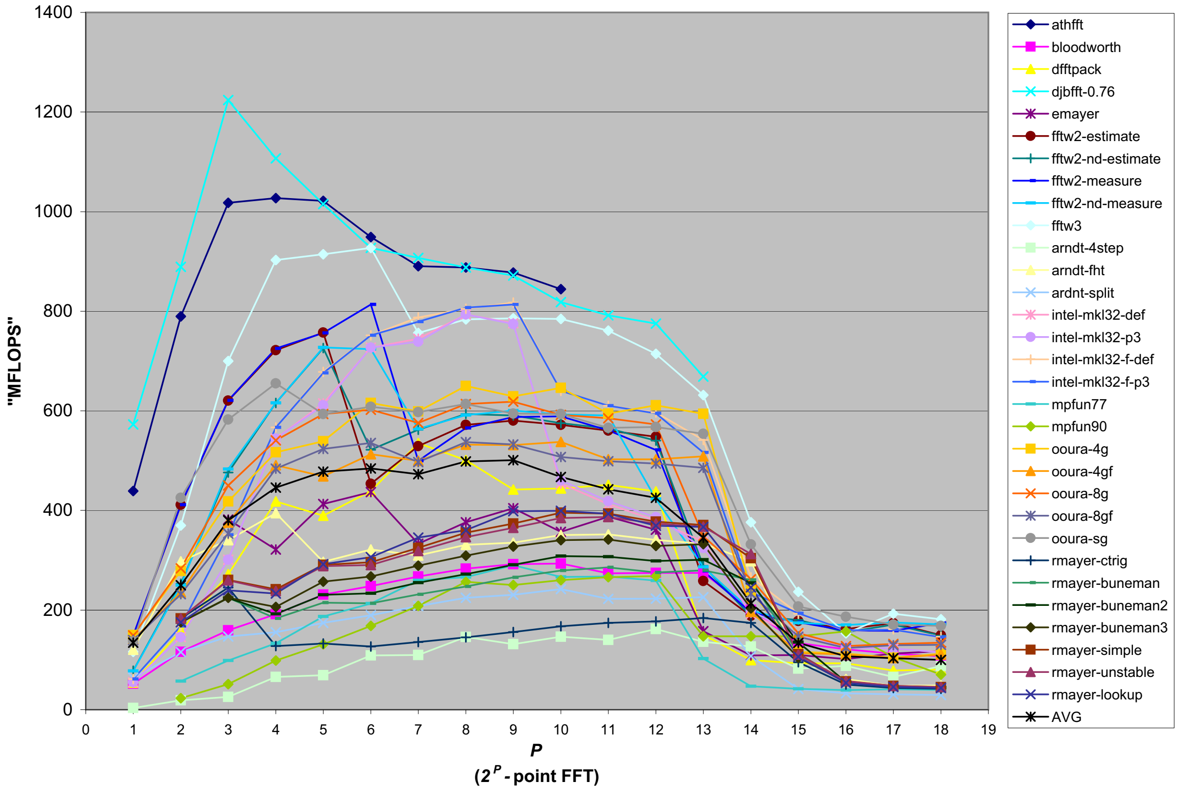
4.2.1.1 Problem Set-Up

To simplify the problem, assume that the input signal bandwidth is 1 GHz. The objective is to determine the number of processors required to obtain 1 kHz resolution in real-time. In digitization, frequency resolution, Δf, is defined as:

$$\Delta f = \frac{B}{S} \tag{4.4}$$

where B is the given bandwidth and S is the number of samples in FFT size. Sampling a 1 GHz band of signal produces 2 GSPS (giga samples per second). Given that $B = 2$ GSPS and $\Delta f = 1$ kHz, then the size of the FFT is 2×10^6 samples. Because the N-point FFT is a power of 2, S must be 2^{21} ($2^{21} = 2,097,152$). Figure 4-7 illustrates a performance benchmark of various FFT algorithm implementations on a Linux computer with a 1 GHz Intel Pentium III Coppermine chip. The results shown are for double-precision, complex, in-place, forward transforms where transform sizes are in powers of 2. An in-place transform is chosen because it is faster than an out-of-place transform. An in-place transform is one where the input data is overwritten by the output data and requires half as much memory. In Figure 4-7, the larger the MFLOPS (million floating point operations per second) , the better the performance obtained, thus the `fftw3` algorithm is chosen because it has the best performance overall for all transform sizes. MFLOPS is a common measurement for rating the speed of a processor.

The output data rate of an ADC is generally much higher than the data processing rate of digital signal processing, thus eliminating this bottleneck for efficient system integration becomes a challenging problem [96]. In order to handle a data rate of 2 GSPS (approximately 2^{31} samples per second), the 2^{21}-pt FFT must be computed in less than 1 msec because a 1 kHz modulation on a carrier requires a waiting time of 1 msec to obtain the samples. Notice that the 1 GHz Intel Pentium III Coppermine processor chip cannot handle a 2^{21}-pt FFT. On a 1 GHz Intel Pentium III Coopermine processor chip, the largest N-point FFT



Source: [32] Raw performance data obtained from Matteo Frigo, one of the FFTW authors.

Figure 4-7: FFT performance on a LINUX computer with a 1 GHz Intel Pentium III Coppermine chip.

that can be computed in less than 1 msec is $N = 2^{13}$. A 8192-point FFT can be computed in 0.843 msec. Therefore, for 1 kHz resolution, a maximum sample rate of 8 MSPS (mega samples per second) is possible.

4.2.1.2 Design Case I

Figure 4-8 illustrates a simple end-to-end design to resolve a 1 GHz band of signal with 1 kHz resolution. Onboard the mission satellite, the input signal is collected and subdivided into 250 streams of 4 MHz signal bands with band-pass filtering. The analog-to-digital conversion then produces 8 MSPS streams to be multiplexed and sent to the processing satellite. The received information on the processing satellite requires a demultiplexer to separate the 250 streams. Each stream is passed to a processor. Thus, this design case calls for 250 processors on the processing satellite. To include support for the system bus, Embedded Intel Pentium III processors at 1 GHz are used. Connecting 250 processors that operate with a bus speed of 133 MHz and a bus width of 64-bits requires an interconnection network capable of a peak I/O rate of 2.128 Tbps.

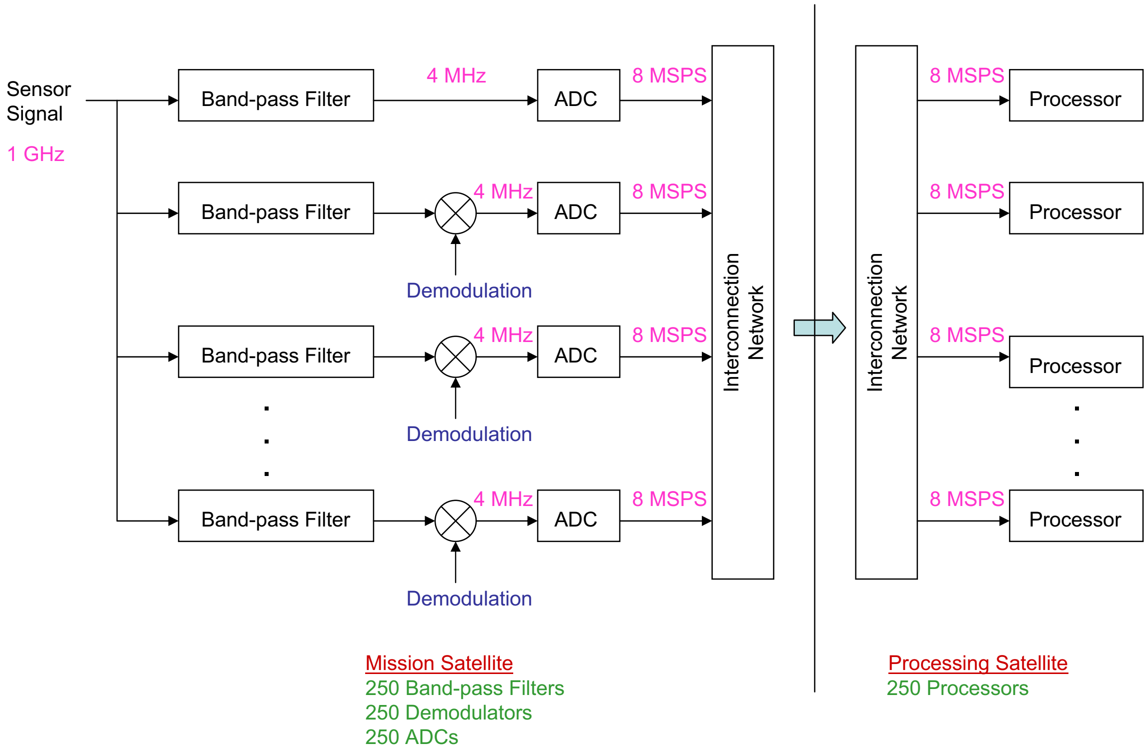


Figure 4-8: Spectral Analysis Design Case I.

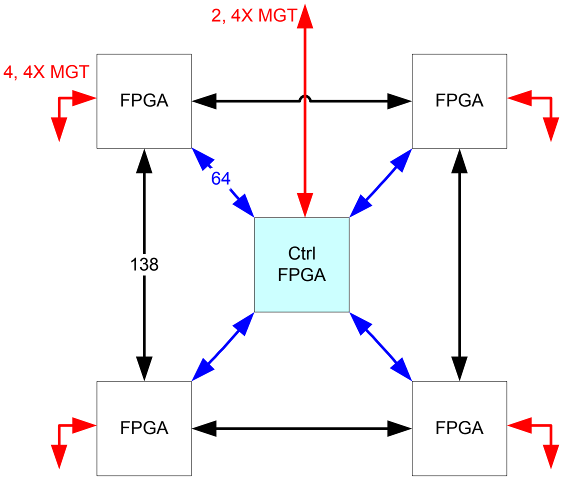
Notice that this design case requires a lot of hardware on the mission satellite: 250 band-pass filters, 250 modulators, and 250 ADCs. It also uses low-performance ADCs. The amount of equipment required on the mission satellite is unreasonable. There is a need to reduce the number of components (e.g., minimize size, weight, and power) while at the same time utilizing faster performing equipment.

4.2.1.3 Design Case II

For high-performance digital signal processing applications, neither GPPs nor ASICs are appropriate options because ASICs do not have the flexibility to handle a variety of applications and GPPs do not have enough computation density, consume too much power, and are unable to guarantee a deterministic latency to satisfy real-time constraints Without much hardware support for multiple digital signal processing communication protocols, DSPs are typically intended for single signal processor applications. When several DSPs are required, external ASICs or FPGAs are frequently used to collect/distribute the data to/from each of the DSPs attached [21]. The flexibility of the architecture is subsequently limited by ASICs. Any post-design optimizations and upgrades in features and algorithms cannot be obtained [16]. Moreover, if FPGAs are used to supplement the system, the next step forward is to process everything with FPGAs.

FPGAs can be designed to handle data as fast as any commercially available ADC can supply it [3]. Extremely high computational throughput can be achieved with FGPAs by using highly parallel architectures [95]. The spectral analysis problem can be solved with the BEE2 system developed by the Berkeley Wireless Research Center (BWRC) at University of California, Berkeley. The BEE2 system is built entirely with COTS components, as described in [21]. The basic computing element (processor) is a Xilinx Vertex-II Pro 70 FPGA chip with four DDR2 (Double Data Rate 2) 240-pin DRAM (Dynamic Random Access Memory) DIMMS where each DIMM (Dual In-line Memory Module) can sustain a maximum capacity of 1 GB.

A BEE2 compute module contains 5 FPGAs (4 basic computing elements and 1 control element). The control FPGA has extra global interconnect interfaces and controls signals to the secondary system elements. The connectivity on the compute module is shown in Figure 4-9. Interconnects include on-board LVCMOS (Low Voltage Complimentary Metal Oxide Semiconductor) connections and off-board MGT connections. The four compute FPGAs are

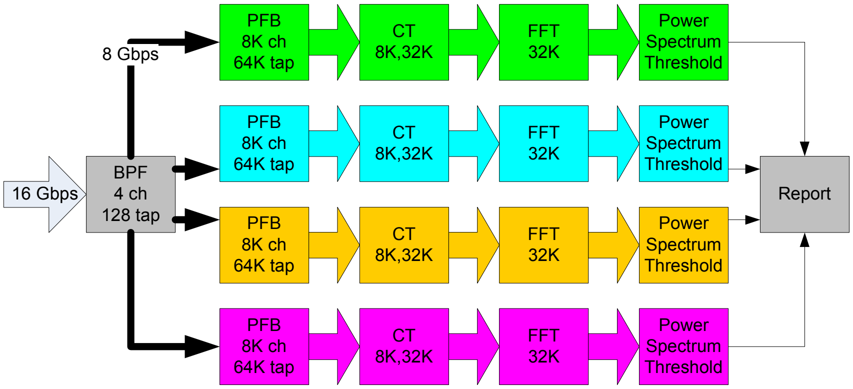


Source: [21] Chen Chang, John Wawrzynek, and Robert W. Brodersen, *Evaluation of Multi-FPGA Systems for High-Performance DSP Applications*, manuscript, December 2004, p. 5.

Figure 4-9: BEE2 compute node connectivity.

connected in a two-by-two 2-D local network. There are 138 physical single-ended circuits on each link between the neighboring FPGAs on the grid, designed to reach up to 150 MHz DDR using the LVCMOS signal standard, with a total bandwidth of 41.1 Gbps per link. From the control FPGA, there are four downlinks to each of the compute FPGA. There are 64 physical single-ended circuits on each downlink, designed to reach up to 150 MHz DDR, with an aggregate bandwidth of 19.2 Gbps per link. The MGTs on the FPGAs are used for all off-module connections. Every separate MGT channel is set in software to operate at 2.5 Gbps or 3.125 Gbps with 8B/10B encoding. To construct a 10 Gbps full duplex (20 Gbps total) interface, every 4 MGTs are channel-bonded into a physical InfiniBand 4X (10 Gbps) electrical connector [21].

The computational efficiency of the BEE2 system has been evaluated for a radio astronomy signal processing application, specifically a one billion channel spectrometer. A single BEE2 compute module with 5 XC2VP70 FPGAs can accommodate the entire billion channel spectrometer. The application objective is to have a spectral resolution of less than 1 Hz. The radio telescope antenna ADC provides 16 Gbps digital inputs that are split into four streams, one to each corner FPGA, as shown in Figure 4-10. Exact implementation details of the data processing of the spectrometer implementation can be found in [21]. Correspondences with the developers of the BEE2 system suggest that at least 2 FPGA devices are required for the described spectral analysis problem given in this dissertation.



Source: [21] Chen Chang, John Wawrzynek, and Robert W. Brodersen, *Evaluation of Multi-FPGA Systems for High-Performance DSP Applications*, manuscript, December 2004, p. 8.

Figure 4-10: One billion channel spectrometer data flow diagram.

The FFT design can be parameterized to almost any number of points with increasing resource utilization on the FPGA. The design is a fully streaming implementation, so it is matched to the input data rate.

4.2.1.4 Summary

Designing for the real-time spectral analysis example is mainly constrained by the processing capability of the processor and the interconnection speed. Because 8 MSPS streams are required with a GPP-based system, it is necessary to have at least 250 Intel Pentium III processors for every 1 GHz band of signal to analyze. The GPP-based system design solution illustrates the feasibility of using commercial processors to solve the problem. However, there may be more optimal designs that can be implemented, depending on the equipment chosen. Given limitations in size, weight, and power on the spacecraft, it is necessary to study the use of FPGAs. The BEE2 system, developed at BWRC, provides a more cost-effective solution. One computing module consisting of 5 Virtex-II Pro X FPGAs is more than sufficient to solve the specified spectral analysis problem given in this chapter. Assuming that 2 Virtex-II Pro X FPGAs can solve the problem, then for every X GHz band of signal to analyze, $2X$ Virtex-II Pro X FPGAs are required.

4.3 Space-Based Radar Application

Radar, an acronym for Radio Detection and Ranging, is a technology that can determine the distance (range) and velocity of an object from analyzing the echoes that the object reflects. A radar device transmits electromagnetic waves to an object which would then be reflected back off the object to the transmitter (or receiver). The received signal can then be analyzed. Most radar systems operate in the microwave region of the electromagnetic spectrum because these frequency bands allow for objects to be seen not only during the day, but at night, through clouds, fog, haze, rain, etc. These sensors, known as *real aperture radars* (RARs), allow for day, night, and all-weather imaging that is important for continuous and global monitoring of the Earth's surface. Radars are commonly flown on airborne and space-borne platforms (e.g., airplanes and satellites).

SBR is valuable for the defense sector because it provides observations of the Earth from orbit. Adversaries cannot hide from view behind obscuring terrain features. It allows military forces to observe more intensively into denied regions of interest, on a non-intrusive basis without risk to personnel or resources. The system would be available in wartime or peacetime. During peacetime, functions such as detailed mapping capabilities can be provided. A constellation of SBR satellites is being envisioned to cover the Earth on a continuous basis [79]. The SBR Program is managed by Air Force Space Command, Space and Missile Systems Center (AFSPC/SMC) and the National Reconnaissance Office (NRO), in cooperation with the Services and the National Imagery and Mapping Agency (NIMA) [80]. The SBR program is a transformational system for the DoD and the Intelligence Community (IC) that concentrates on maturing technology and developing an Intelligence, Surveillance, and Reconnaissance (ISR) system able of providing the following core capabilities: SAR, Ground Moving Target Identification (GMTI), and High-Resolution Terrain Information (HRTI) data.

4.3.1 Synthetic Aperture Radar (SAR)

SAR is a remote sensing technique that solves the limitation of poor resolution achieved by RARs. Doppler frequency is utilized to distinguish targets and pinpoint them in azimuth. Depending on the operating wavelength, in order to realize resolutions on the order of magnitude of meters, microwave sensors in RARs operating at hundreds of kilometers of altitude would call for antenna dimensions between several hundred meters to some kilometers [36]. Synthetic aperture (also known as synthetic antenna) is a very long antenna that is synthesized by moving a small one along a convenient path (the platform flight path) and then properly processing the received signals. Radar resolutions are improved in azimuth (i.e., in the direction of the velocity vector of the platform [e.g., airplane or satellite]). The attainable resolution is comparable to that which could be obtained by a very large physical antenna.

The potential applications of SAR data have widespread appeal to the defense and scientific community. SAR sensor systems built by various countries around the world are listed in Appendix B. A key mission objective of remote sensing SAR systems is to look for changes in measurements of surface details over long periods of time. SAR systems typically involve large data volumes with extensive processing to achieve the images with the required resolutions. Initially, these operations have been done through optical processing techniques. Digital SAR processing later superseded optical processing. Historically, high speed digital SAR processing systems have typically depended on custom hardware designs and have often been very complex and expensive.

SAR systems in the past used optical signal processing to manage the considerable quantity of data storage and computation required. Optical processing was commonly performed in non-real time on data that have been recorded on photographic film. The implemented processor employs an extremely complex lens system. Such optical processing systems have a number of disadvantages: (1) they can be expensive, (2) they are normally limited to producing a strip map while flying in a nearly straight path, and (3) motion compensation is difficult to implement given that the optical processor is not flexible. The desire for greater flexibility and real-time operation suggests digital processing. Nevertheless, digital processing also encounters the following bottlenecks: (1) high input signal bandwidth, (2) substantial storage requirement, and (3) large computation load for high-resolution mapping.

4.3.1.1 Digital Data Processing Algorithm

The oldest and most widely used SAR algorithm for digital SAR data processing is the range-Doppler algorithm. It was developed in 1979 for the processing of SEASAT data by MacDonald Dettwiler and Associates (MDA) and the Jet Propulsion Laboratory (JPL).

The main procedures in the algorithm, represented in Figure 4-11, are explained below [61]:

1. Range FFT: An FFT is carried out on the data in the range direction.
2. Range compression (RC) with Secondary Range Compression (SRC): RC and SRC are accomplished by using an array multiple in the range direction.
3. Range Inverse FFT (IFFT) : An IFFT is carried out in the range direction.
4. Azimuth FFT: An FFT is carried out in the azimuth direction.
5. Range Cell Migration Correction (RCMC): RCMC is accomplished by a shift and interpolation operation that assembles the target trajectories in memory.
6. Azimuth Compression (AC): AC is done with an array multiply in the azimuth direction.
7. Azimuth Inverse FFT: An IFFT in the azimuth direction finalizes the construction of the image.

The computational classification of the main procedures of range-Doppler algorithm is shown in Figure 4-12. The percentage of total operations utilized by each of the main procedures in the algorithm, in addition to azimuth matched filter generation (AMFG) procedure is shown. The basic operation types in the algorithm include FFTs, complex vector multiplications (CVM), filtering or interpolation operations, and scalar operations. The percentage of total operations by each basic operation type is shown in Figure 4-13. In the calculation of computational amounts, a single operation (add, multiple, etc.) on real operands is considered to be a base unit. From Figure 4-13, it is observed that FFTs make up the majority of the computation in the range-Doppler algorithm. Therefore, it is possible to use the same processing satellite for SAR processing as SIGINT processing.

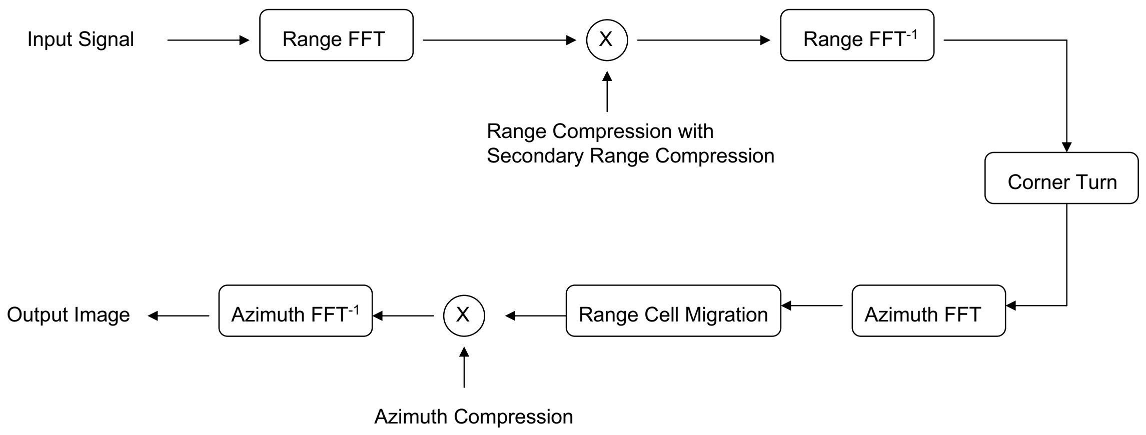


Figure 4-11: Block Diagram of the SAR Range-Doppler Algorithm

.

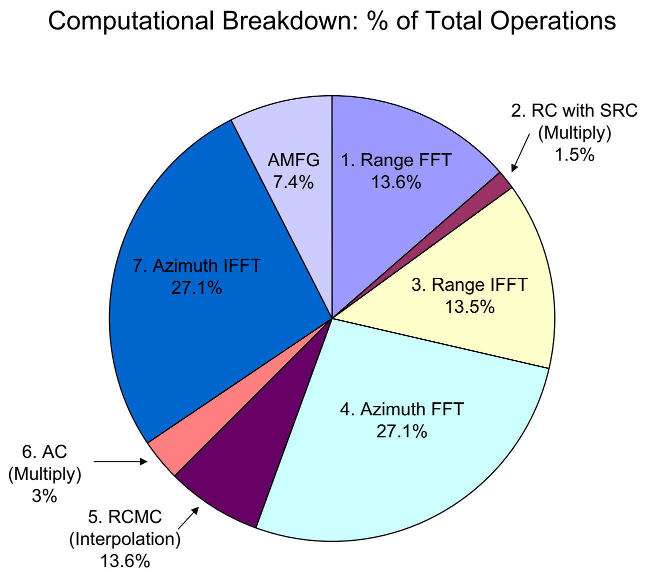


Figure 4-12: Computational Classification of the Range-Doppler Algorithm [61].

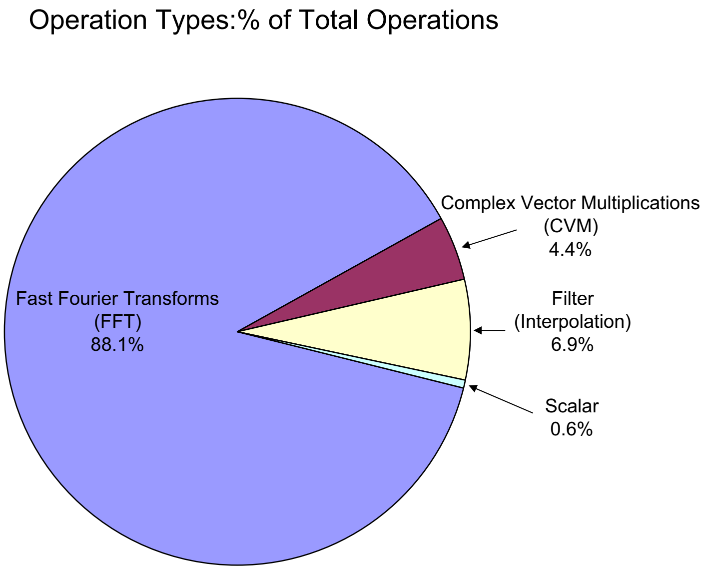


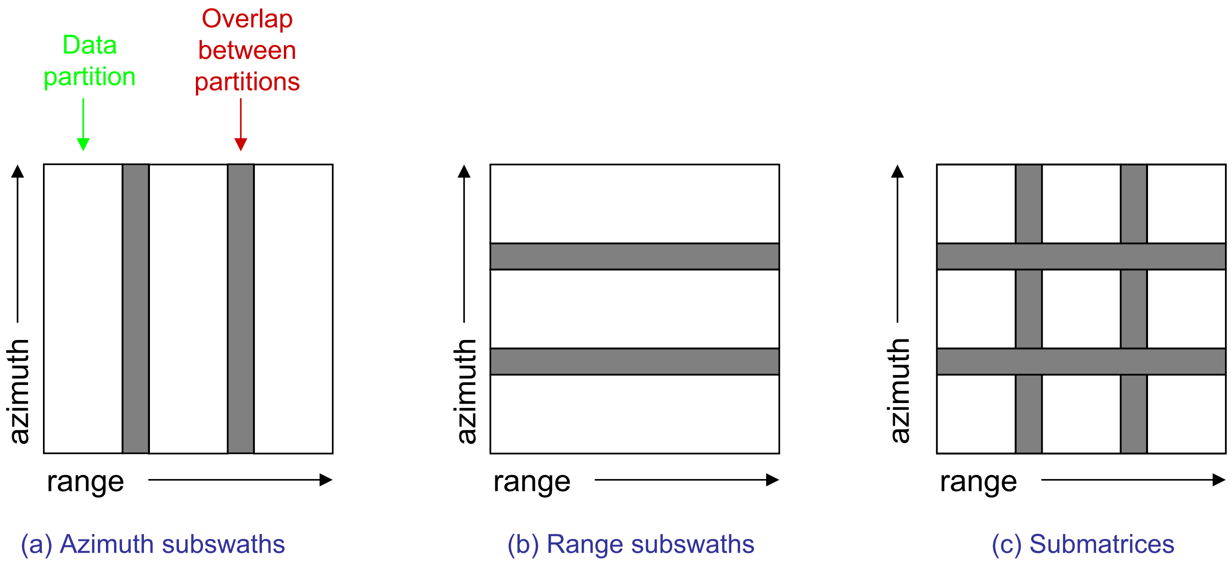
Figure 4-13: Basic Operation Types of the Range-Doppler Algorithm [61].

4.3.1.2 Parallel Processing

Because of the large computational requirements, parallel processing is required to achieve real-time high rate SAR images. Parallelism can be achieved in SAR processing with the range-Doppler algorithm because the structure of the processing flows can be broken down into large sequential steps with simple data dependencies and synchronization requirements. Additionally, more efficiency can be gained by partitioning the data array sizes in range and azimuth directions. The granularity of partitioning approaches for SAR has been examined in [61]. Coarse grain parallelism includes (1) vertical partitioning, also known as pipelining or temporal partitioning, (2) horizontal partitioning, also known as data parallelism, and (3) vertical-horizontal partitioning, which is a combination of the other two partitioning techniques. Vertical partitioning can be implemented due to the large number of sequential steps. As processor throughput increases, so does latency. Parallel implementation of the FFT algorithm on MIMD machines have been studied and published [8, 42]. In horizontal partitioning, the data set is divided among the processors and each processor performs identical operations on the subsets of data. The advantages of coarse grain parallelism include simpler scheduling, less synchronization problems, and the ability to leverage commercial system components. On the contrary, fine grain parallelism can offer greater parallelism capabilities and higher speed-ups but require custom VLSI (Very Large Scale Integrated) designs.

While there are various partitioning approaches, the most common approach appears to be data parallelism or horizontal partitioning. Figure 4-14 illustrates the data partitioning options available for horizontal partitioning: (a) azimuth subswaths or strips, (b) range subswaths or strips, and (c) submatrices. Sizing of the data partitions is important because smaller partitions allow for more parallelism but will result in decreased efficiency. The horizontal partitioning approach is seen in the parallel processing of SAR imaging data from the Spaceborne Shuttle Imaging Radar-C/X-Band SAR (SIR-C/X-SAR) in [63]. The SIR-C/X-SAR data is separated into eight different polarization channels, each to be processed independently of each other in the correlation phase. The first level of parallelization is to assign the data associated with each of the polarization channel to a different group of processors. This data decomposition allows for simultaneous processing of all eight polarization channels, supports good I/O scalability, and reduces the level of system-wide interprocessor communication. A second level of parallelization is implemented where the computation and I/O associated with the azimuth and range lines of each polarization channel are assigned to the processors within each of the polarization groups. The processing throughput of the SIR-C/X-SAR data is significantly increased via the two levels of parallelization. However, a balance between the workload and overhead must be maintained as performance can be limited by the overhead. The speedup and efficiency of a parallelized chirp scaling algorithm for SAR imaging has been examined in [94]. Collective communication (interprocessor communication) is the main overhead in most SAR algorithms.

The growth in performance capabilities of large-scale commercial parallel computers, along with their CPU memory and disk capabilities, has led to the recent trend of using commercial parallel computers for SAR missions. The designs of three parallel processing systems capable of processing Radarsat SAR data (4.5 GOPS [giga operations per second]) at a minimum of 1/10 the real-time rate have been examined in [60]. In order to process all the data obtained in one satellite SAR ground station with no backlog, 1/10 real-time processing (445 MOPS [mega operations per second]) is usually adequate. A summary of the findings are presented: (1) A special-purpose DSP system, implemented with 1 Sharp/Butterfly LH9124 vector DSP and 8 Analog Devices ADSP-21060 SHARC (scalar processor), offers the highest performance at the expense of less flexibility and more design work (high development cost). (2) A general-purpose DSP system, implemented with 16 Analog Devices ADSP-21060 SHARC processors, offers a balance between performance and



(a) Azimuth subswaths (b) Range subswaths (c) Submatrices

Figure 4-14: Data partitioning options in horizontal partitioning [61].

flexibility. General-purpose DSPs, optimized for DSP operations, can be flexible enough to handle a variety of algorithms. General-purpose DSP systems can have a lower chip count and are a lower cost solution than GPP architectures; however these systems are more specialized and require more design effort. (3) A GPP system, implemented as a configuration of 10 workstations (DEC Alpha 21064) with a 130 Mbps network, offers high performance, large economies of scale, and a high degree of flexibility as the same hardware can be used for applications other than SAR.

Notice that the system architecture of BEE2 is similar to NOW (network of workstation) clusters, with workstations replaced by BEE2 modules, and Ethernet replaced by InfiniBand. It is reasonable to consider using a number of BEE2 modules in the processing satellite to parallel process SAR data in real-time.

4.3.1.3 Processing Complexity

This subsection focuses on the processing complexity of a SAR system. The main problems with building a digital SAR processing system for a high resolution and large swath spaceborne imaging sensor satellite are related to the need for a large quantity of data memory needed to create a synthetic aperture and the very high speed arithmetic computation requirements. *Range resolution* is defined as the minimum distance between two points which are distinguishable and is proportional to the signal bandwidth [28]. Two objects

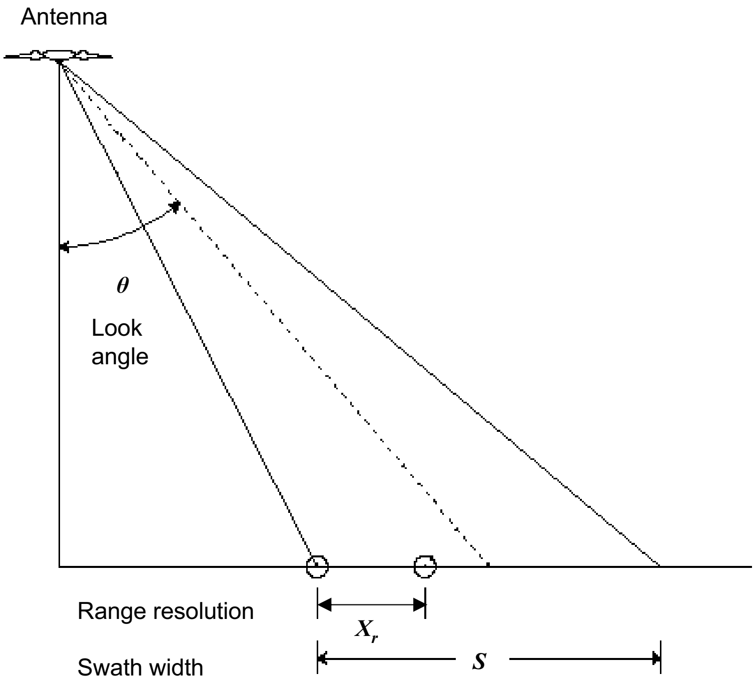


Figure 4-15: SAR geometry.

can be discerned if the trailing edge of the pulse echo from the nearer object arrives at the antenna before the leading edge of the pulse echo from the remote object. Figure 4-15 is a simple illustration of the geometry for calculating the range resolution, where S is the swath width. Mathematically, range resolution, X_r, is a function of the look angle, θ, and signal bandwidth, B, given as:

$$X_r = \frac{c}{2B \sin\theta} \tag{4.5}$$

where c is the speed of light. Note, as θ approaches zero (approaching the nadir line directly below the satellite), $\sin\theta$ also approaches zero, resulting in exceedingly poor ground range resolution. Objects near the nadir line are virtually impossible to differentiate because they are nearly the same distance from the antenna. Also note that the resolution of the SAR is independent of the altitude of the sensor. This is due to the fact that the imaging mechanism uses the Doppler shifts in the echo and the differential time delays between surface points, neither of which is a function of the distance between the sensor and the surface [28]. Needless to say, the altitude still plays a major factor in determining the power required to obtain a detectable echo and in determining the size of the antenna.

It can also be seen from Equation (4.5) that at constant look angles, improving range

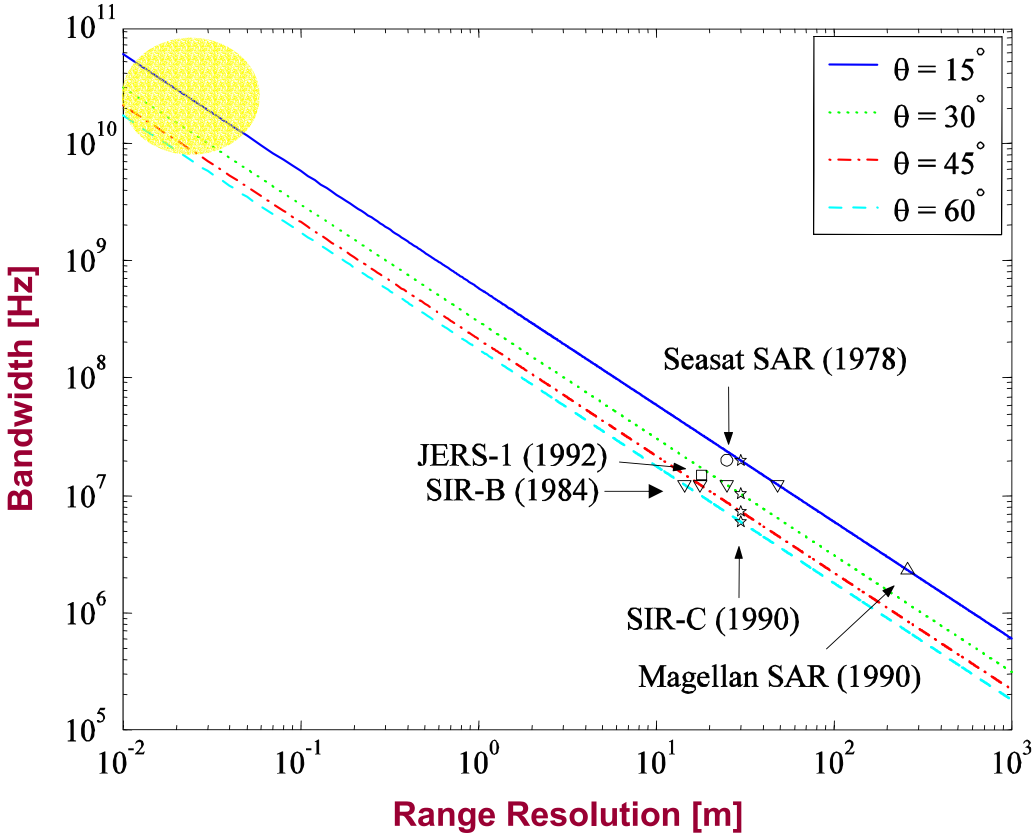


Figure 4-16: Range Resolution of Operational SAR Satellites.

resolution requires an increasing amount of bandwidth. Figure 4-16 illustrates this property with look angles of 15°, 30°, 45°, and 60° along with a sampling of several operational SAR satellite systems. With a look angle of 15° and a range resolution of 1 cm, a bandwidth of approximately 58 GHz is required. This region, highlighted in Figure 4-16, illustrates the region of future sensing satellites with very small range resolution capabilities that require large amounts of bandwidth.

The arithmetic complexity, as derived in [28], is a simplified quantification of the processing complexity and can be measured by the product of the output pixel rate and the number of arithmetic operations required per pixel. The relationships required to calculate the arithmetic complexity of space-borne SAR system are listed in Table 4.8. The pixel rate at the output of the processor K_0 is equal to the product of the number of image pixels

Relationship	Definition / Comment
$S = \frac{\lambda h}{W \cos\theta}$	Swath width
$X_r = \frac{c}{2B \sin\theta}$	Ground range resolution
$X_a = \frac{L}{2}$	One look highest azimuth resolution
$F_a = \frac{\lambda h}{L \cos\theta}$	Azimuth footprint is also equal to the length of the synthetic aperture
$T_i = \frac{F_a}{v}$	Maximum integration time
$PRF > \frac{2v}{L}$	Nyquist criterium
$PRF < \frac{cW \cos^2\theta}{2\lambda h \sin\theta} = \frac{c}{2S \sin\theta}$	Avoid overlap of successive echoes

Table 4.8: Main relationships for space-borne SAR systems [28].

across the swath multiplied by the number of azimuth pixels per second:

$$\begin{aligned} K_0 &= \left(\frac{S}{X_r}\right)\left(\frac{v}{X_a}\right) \\ &= \left(B\frac{2S\sin\theta}{c}\right)\left(\frac{2v}{L}\right) \\ &= B(T_e)(PRF) \end{aligned} \qquad (4.6)$$

where S is the swath width, X_r is the ground range resolution, X_a is the one look highest azimuth resolution, v is the velocity of the sensor, L is the length of the antenna, T_e is the total time spread of the echo, and PRF is the pulse repetition frequency. Because the product $T_e \times PRF$ is usually a little less than 1, Equation (4.6) can be simplified to:

$$K_0 \backsimeq B \qquad (4.7)$$

The number of arithmetic operations required to produce a pixel is at least equal to the number of echoes used in the generation of one pixel:

$$\begin{aligned} K_0' &= PRF \times \left(\frac{F_a}{v}\right) \\ &= \frac{2v}{L}\left(\frac{h\lambda}{vL\cos\theta}\right) \\ &= \frac{2\lambda h}{L^2\cos\theta} \end{aligned} \tag{4.8}$$

The product of Equation (4.7) and Equation (4.8) is the arithmetic complexity for real-time time domain processing, which can be simplified as follows:

$$\begin{aligned} K_a &= K_0K_0' \\ &= \frac{2\lambda hB}{L^2\cos\theta} \end{aligned} \tag{4.9}$$

A frequency domain implementation requires $O(\log K_0')$ computational complexity for each output value. When a real-time frequency domain approach is used to process the SAR data, the arithmetic complexity is then:

$$\begin{aligned} K_a' &= K_0\log K_0' \\ &= B\log\left(\frac{2\lambda h}{L^2\cos\theta}\right) \end{aligned} \tag{4.10}$$

Table 4.9 lists the the processing complexity for SEASAT SAR and the variables used for the calculation. This method of calculating the processor complexity does not take into account the control-function complexity which includes reference functions generation and updating, error corrections, etc.

	SEASAT SAR (1978)
S = swath width	100 km
h = sensor altitude	800 km
λ = operating wavelength	23.5 cm
X_r = range resolution	25 m
B = bandwidth	19 MHz
θ = look angle	18°
L = antenna length	10.7 m
K_0 = output pixel rate [per sec]	19×10^6
K_a = real-time time domain processing	66 GOPS
K_a' = real-time frequency domain processing	67 MOPS

Table 4.9: Processor complexity for SEASAT SAR.

While SAR systems currently can provide 8 m to 30 m resolution as seen in Figure 4-15, there is growing demand for very high resolution images of 1 m resolution or less. This demand is being met in part by new satellite systems operating in the visible field (e.g. IKONOS). Future SAR systems with less than 1 m resolution are expected through the use of an acquisition mode known as spotlight SAR. Spotlight mode offers finer azimuth resolution than what is achievable in strip map mode using the same physical antenna [19].

A calculation of the computational complexity for very high resolution space-borne SAR systems based on the worst-case scenario for the data rate required to map the entire Earth at 1 m^2 resolution is provided. The area rate is defined as the surface area per second of the satellite sensor sweep and can be written as:

$$\text{Area rate} = \text{Ground track velocity} \times \text{Swath width} \tag{4.11}$$

where the ground track velocity is given by:

$$\text{Ground track velocity} = \frac{\text{Earth circumference}}{\text{Orbital period}} \tag{4.12}$$

Given that the radius of the Earth is 6378.14 km, the circumference of the Earth is 40,075 km. Assuming that the sensor has an orbital period of 90 minutes, the ground track velocity is calculated to be 7.4213 km/sec. The area size of the strip is determined at the equator where the swath width is greatest. Assuming that the sensor completes a polar orbit around the Earth in 90 minutes, the swath width is 2504.7 km ($\frac{1.5}{24\text{ hours}} \times 40,075\text{ km} = 2504.7\text{ km}$). Thus, the area rate for the stated problem is $1.8588 \times 10^{10}\text{m}^2/\text{sec}$.

A pixel is an individual measurement in the satellite sensor sweep. Thus, for 1 pixel per 1 m^2, the total number of pixels is 18.588 billion/sec. SAR image formation requires computationally complex calculations as high as 1000 floating point operations per image pixel [19]. Three algorithms that can form fine-resolution digital imagery from spotlight SAR data have been shown to require less than 300 real operations per pixel [19]. For a 1-m resolution, air-to-ground spotlight SAR system, the polar format algorithm (PFA) uses 280 real operations per pixel, the range migration algorithm (RMA) uses 288 real operations per pixel, and the chirp scaling algorithm (CSA) uses 235 real operations per pixel. Assuming a 300 floating point operations per pixel algorithm, the computation complexity for 1 m resolution becomes 5.5764 trillion floating point operations per second. A system of 40

BEE2 modules (200 Virtex-II Pro X FPGAs) is capable of delivering up to 28.8 TOPS (tera operations per second) (16-bit integer) or 2 TFLOPS (tera floating point operations per second) [21]. Assuming linear scaling, 1 FPGA is capable of 10 GOPS. At this rate, 558 FPGAs are required for 1 m resolution (and 5.5764 million FPGAs are required for 1 cm resolution). The quantity of processors that can be flown is primarily limited by their power consumption, as discussed in Section 4.4.2. Given that a system of 200 Virtex-II Pro X FPGAs requires 12 kW of power [21], a system of 558 FPGAs requires 33.48 kW. This large amount of power consumption is not currently available on any commercial spacecraft (e.g., a Boeing 702 Plus satellite offers a power range up to 25 kW [15]).

This design problem illustrates the upper bound on the number of processors required because it assumes 100% duty cycle. Because the oceans cover about 70% of the Earth's surface, the sensor may not need to be operational at all times. Thus, the number of processors required may be substantially lowered. A strategy for obtaining high resolution imaging with a more reasonable number of processors is to have sensors take a wide area scan at a lower resolution and then conduct a spot scan for a very small area of interest at a higher resolution. For example, with a 30% duty cycle and a 10 m resolution for a wide area scan, the number of FPGAs required is 2. Assuming that an area of interest occurs about 1% of the time, a spot scan at 1 m resolution requires 6 FPGAs. This strategy thus requires a total of 8 FPGAs that can be flown on a satellite.

4.3.2 Summary

Constructing a SAR image from raw data requires an enormous amount of signal processing. High speed digital SAR processing systems have traditionally depended on custom hardware designs which can be very complex and expensive. This traditional drawback with SAR as a remote sensing instrument can be improved with high-end reconfigurable computer systems. The power of FPGA-based processing systems can meet the requirements put forth by SAR processing; launching new and innovative options for SAR processor architectures. Because of the large computational requirements, parallel processing is used extensively for SAR processing.

4.4 Processing Satellite

This section discusses the processing architecture and payload sizing of the processing satellite. The power budget is one of the major factors in determining the size of the processing payload. Processors can consume a lot of power, thus the maximum number allowable onboard is constrained by the power limitation of the spacecraft. Implementing a large number of processors requires an interconnection network and storage. Consequently, power requirements for the interconnection network and storage must also be taken into account.

4.4.1 Processing Architecture

The processing satellites should accommodate various digital signal processing applications and support a wide variety of global communication schemes. Figure 4-17 illustrates a generic connectivity architecture for the processing satellite. The processing elements (PEs) can be used to create various types of network topology, such as a tree or a 3-D mesh. Tree communication networks are useful for data aggregation or distribution. For applications that need high bisection bandwidth and random communication among many compute modules, the interconnection network processing can be designed to use a crossbar switch technology. Switches provide point-to-point communication that is faster than a shared medium. Ethernet is an example of a shared medium that can be used to build conventional networks but does not provide the performance or features required for high-performance and high-availability. Moreover, aggregate bandwidth of the switch is many times that of the single shared medium. Switches also allow the interconnection network to scale with a very large number of nodes. Nevertheless, it is quite common to have both a high speed and a low speed interconnection network. A lower speed interconnection network can support command and control communications.

Communications between the processing elements and other computing and storage devices can also be served by either the crossbar switch network or the Ethernet network. Options for storage devices include Network Attached Storage (NAS) or a Storage Area Network (SAN). A NAS acts as a storage unit to the network, like a file server. A SAN creates a separate back-end network designed specifically for storage-heavy traffic because standard networks cannot handle the bandwidth requirements of certain applications. This allows storage devices to have increased scalability, availability, and performance. Today,

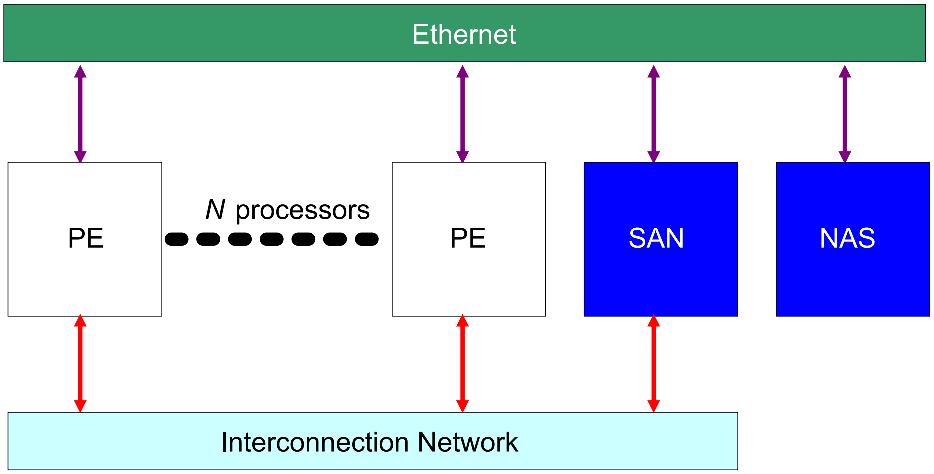
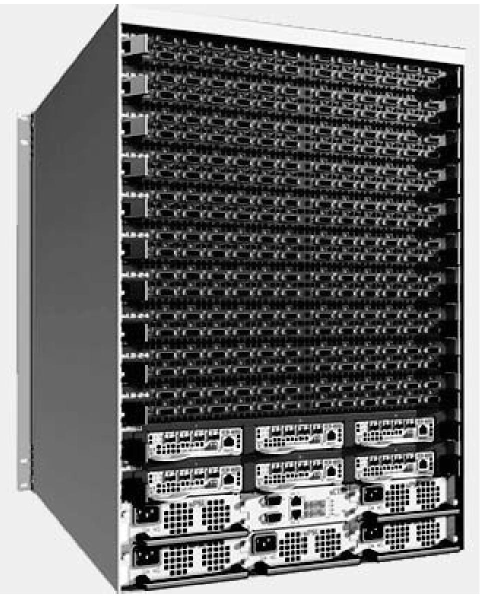


Figure 4-17: Generic connectivity for processing satellite.

the distinction between NAS and SAN has become fuzzy. The primary distinction between NAS and SAN products is the choice of network protocol.

InfiniBand is an industry-standard technology used to connect processor nodes and I/O nodes to form a system area network. The InfiniBand specification defines the raw bandwidth of the base 1X connection at 2.5 Gbps. It then specifies two additional bandwidths, referred to as 4X and 12X, as multipliers of the base link rate. Commercial providers of InfiniBand technology include Voltaire and Mellanox Technologies, Inc. Voltaire's ISR 9288 InfiniBand Switch Router, shown in Figure 4-18, supports up to 288 InfiniBand 4X ports or 96 InfiniBand 12X (30 Gbps) ports. It fits in a 19-inch rack mountable chassis. The dimensions (Height × Width × Depth) are 24.5 in (622 mm) × 17.5 in (444 mm) × 22.75 in (578 mm). Depending on the configuration, the weight of the switch ranges from 110 to 187.5 lbs (50 to 85 Kg). One such switch is less than 10% of the total payload that can be launched by the smallest Delta launch vehicle. Examining the state of the art ground-based system of a dense switch provides an estimate of the size and weight of the interconnection network that may be required for the processing satellite because additional engineering and packaging for space usage would have to be done. Alternatives to InfiniBand technology for the interconnection network include Myrinet from Myricom and QsNet from Quadrics. Regardless of which technology is chosen for the interconnection network, the interconnection network should provide high performance (e.g., high data rate and low latency interprocessor communication) and high availability (low rate of contention).

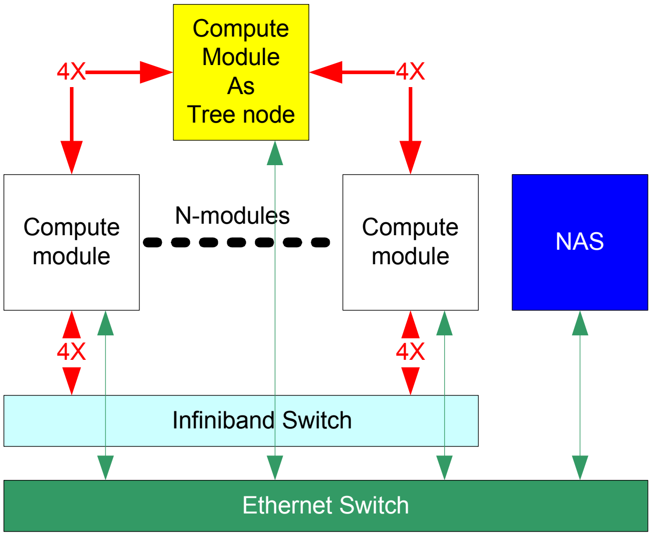


Source: [88] Voltaire ISR 9288 InfiniBand Switch Router Datasheet. http://www.voltaire.com/

Figure 4-18: Voltaire ISR 9288 InfiniBand Switch Router.

FPGAs are the key enabler and building block for high-end reconfigurable computing architectures because they consist of a matrix of logic blocks and an interconnection network that are programmable. By downloading bits of configuration data onto the hardware, FPGAs allow for a high degree of flexibility in the network [16]. For example, FPGAs can be programmed for circuit switched routing or dynamic packet switched message routing. Furthermore, the predictable memory and network latencies allow for static scheduling of memory access and data transfers in some real-time applications [21].

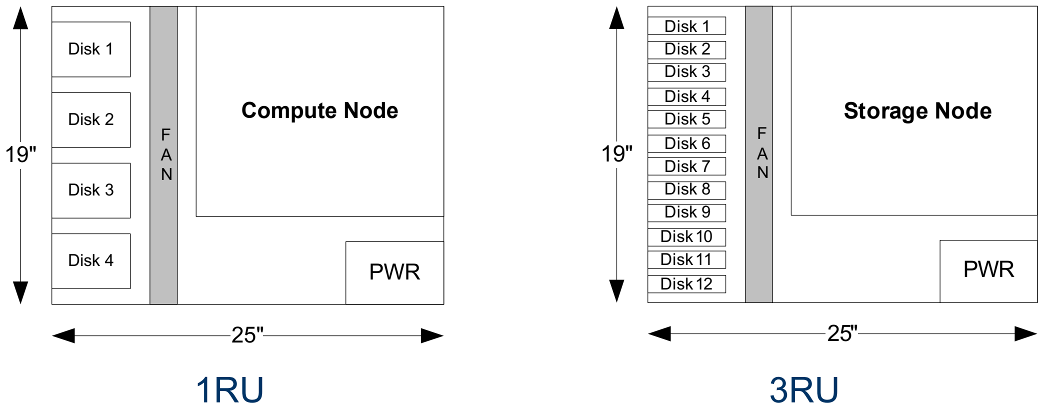
The BEE2 system, described earlier in Section 4.2.1.3, is an example of a multiple FPGA module that can be used as a building block to develop the processing architecture of the processing satellites. Figure 4-19 illustrates the types of global communication networks that the BEE2 system supports: (1) a low latency 4-ary global communication tree, (2) high-bandwidth non-blocking crossbar switch, and (3) a Gigabit Ethernet switch. Each compute module has up to 18 InifiBand 4X connectors (10 Gbps full duplex each) and can act as a global communication tree (GCT) node, connecting up to 2 independent parent nodes and up to 16 other compute modules as its leaves. The InfiniBand crossbar switch is used for the processing interconnection network. The control FPGA on each compute node connects to the storage devices and uses the regular 10/100Base-T Ethernet connection for low speed system control, monitoring, and data archiving functions [21].



Source: [21] Chen Chang, John Wawrzynek, and Robert W. Brodersen, *Evaluation of Multi-FPGA Systems for High-Performance DSP Applications*, manuscript, December 2004, p. 6.

Figure 4-19: Compute node connectivity for BEE2 system.

The possible chassis configuration for a BEE2 compute node and a storage node are shown in Figure 4-20. The unit of 1 RU is equivalent to 1.75 inches. U is the standard unit of measure for designating the vertical usable space, or height of racks (metal frames designed to hold hardware devices) and cabinets (enclosures with one or more doors). Once again, examination of a ground-based system provides an estimate for what is required on a satellite. At the compute node local level, there are 4 SATA (Serial Advanced Technology Attachment) disks using RAID 0 (Redundant Array of Independent Disks), offering 500 GB of capacity. First generation SATA have a bandwidth of 1.5 Gbps, with future interfaces planned to 600 Mbytes/sec [76]. Storage nodes that can be shared among multiple nodes can be implemented with up to 12 SATA disks using RAID 5, offering 3 TB of capacity. The different RAID levels provide a measure of fault tolerance and performance. RAID 0 provides no redundancy while RAID 5 is one of the most popular implementations because it provides excellent performance and good fault tolerance. Additional storage can be provided externally with either a SAN or a NAS. While the implementation discussed uses hard disk technology for a ground-based storage system, flash memory is an alternative device that should be considered for storage in space systems. The use of flash memory has dramatically increased in embedded systems because flash memory devices are high density, low cost, nonvolatile, fast, and electrically reprogrammable [39]. Designing the storage media for the



<u>Source</u>: [84] BEE2 Team, *Building BEE2: a Case for High-End Reconfigurable Computer (HERC)*, presentation, January 12, 2004.

Figure 4-20: Chassis Configurations.

processing satellite will require a trade between the key characteristics of access, speed, capacity, cost, and radiation tolerance.

4.4.2 Payload Sizing

One of the major budgets in spacecraft design is power. Power consumption of microprocessors is important characteristic to evaluate for payload sizing of the processing satellites. The *payload* of a spacecraft contains mission-specific equipment or instruments while the *spacecraft bus* carries the payload and provides the following housekeeping functions: support the payload mass; maintain the payload at the precise temperature; supply electric power, commands, and telemetry; place the payload in the correct orbit and stay there; point the payload properly; and provide data storage and communications, if necessary. Power consumption is mainly associated with heat generation, which is a major disadvantage in attaining increased performance. Keeping high performance processors cool is a key concern. Like all electronic equipment, processors have specified safe temperature ranges that correspond to their limits for normal operation. Overheated processors suffer from problems such as system crashes, lockups, and random reboots. Moreover, problems from overheated process can manifest itself through memory errors, application errors, or disk problems. In rare instances, a severely overheated processor can be permanently damaged.

The Boeing 702 satellite is used in the following examples as the spacecraft to house the processing payload. The Boeing 702 satellite offers system modularity in addition to a payload tailored to customer specifications. It also offers separate bus and payload thermal environments and substantially large heat radiations to provide a cool and stable thermal environment for both the bus and payload, which increases satellite reliability over lifetime service. Processing satellites built with the Boeing 702 satellite system for the space-based information network architecture can be deployed with the following space launch vehicles: Atlas III family, Delta III, Ariane 4 and 5, Proton, and Sea Launch [15]. The existence of multiple launch vehicles provides added flexibility in terms of launch date availability and low cost options.

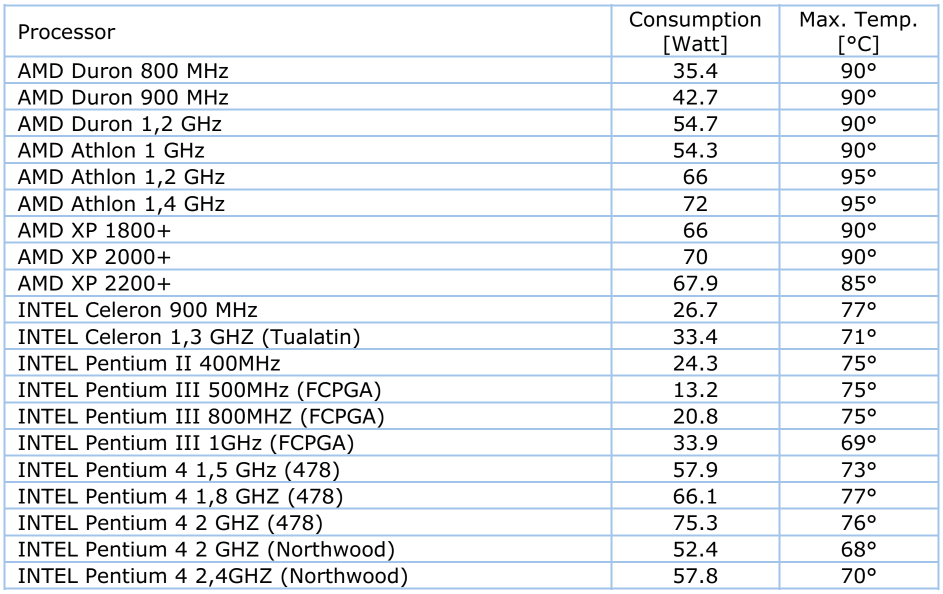
To determine the size of the processing payload (the number of processors that can be flown on a satellite), the power availability of the spacecraft must first be determined. The payload/bus integration design of a Boeing 702 satellite permits fast parallel bus and payload processing. The design of the power system allows the Boeing 702 satellite to offer power up to 18 kW [14]. The "Plus" version of the Boeing 702 satellite offers power up to 25 kW [15]. Given the power dissipation of a processor, the size of the processing payload is determined by the following simple relation:

$$\text{Size of Processing Payload} = \left\lfloor \frac{\text{Total Payload Power}}{\text{Power Dissipation of a Processor}} \right\rfloor \tag{4.13}$$

Note that this calculation does not take into account the power requirements of the interconnection network nor the storage devices that are also implemented. For example, the Voltaire ISR 9288 InfiniBand Switch Router has a maximum power consumption of 2.5 kW for a full configuration.

4.4.2.1 Pentium-based Payload

This section discusses a point design of using the Intel Pentium III processors at 1 GHz for the processing payload because there is interest in launching Pentium-based processing satellites into orbit. Additionally, data regarding total dose radiation, FTT computation performance, and power consumption are readily available. The power consumption and the maximum temperature of several current AMD and Intel processors are listed in Table 4.10. As newer processors attempt to include extra features and to operate at faster speeds, power consumption increases are more likely. Processor designers strive to compensate for this trade principally through technology, by means of lower-power semiconductor processes,



Processor	Consumption [Watt]	Max. Temp. [°C]
AMD Duron 800 MHz	35.4	90°
AMD Duron 900 MHz	42.7	90°
AMD Duron 1,2 GHz	54.7	90°
AMD Athlon 1 GHz	54.3	90°
AMD Athlon 1,2 GHz	66	95°
AMD Athlon 1,4 GHz	72	95°
AMD XP 1800+	66	90°
AMD XP 2000+	70	90°
AMD XP 2200+	67.9	85°
INTEL Celeron 900 MHz	26.7	77°
INTEL Celeron 1,3 GHZ (Tualatin)	33.4	71°
INTEL Pentium II 400MHz	24.3	75°
INTEL Pentium III 500MHz (FCPGA)	13.2	75°
INTEL Pentium III 800MHZ (FCPGA)	20.8	75°
INTEL Pentium III 1GHz (FCPGA)	33.9	69°
INTEL Pentium 4 1,5 GHz (478)	57.9	73°
INTEL Pentium 4 1,8 GHZ (478)	66.1	77°
INTEL Pentium 4 2 GHZ (478)	75.3	76°
INTEL Pentium 4 2 GHZ (Northwood)	52.4	68°
INTEL Pentium 4 2,4GHZ (Northwood)	57.8	70°

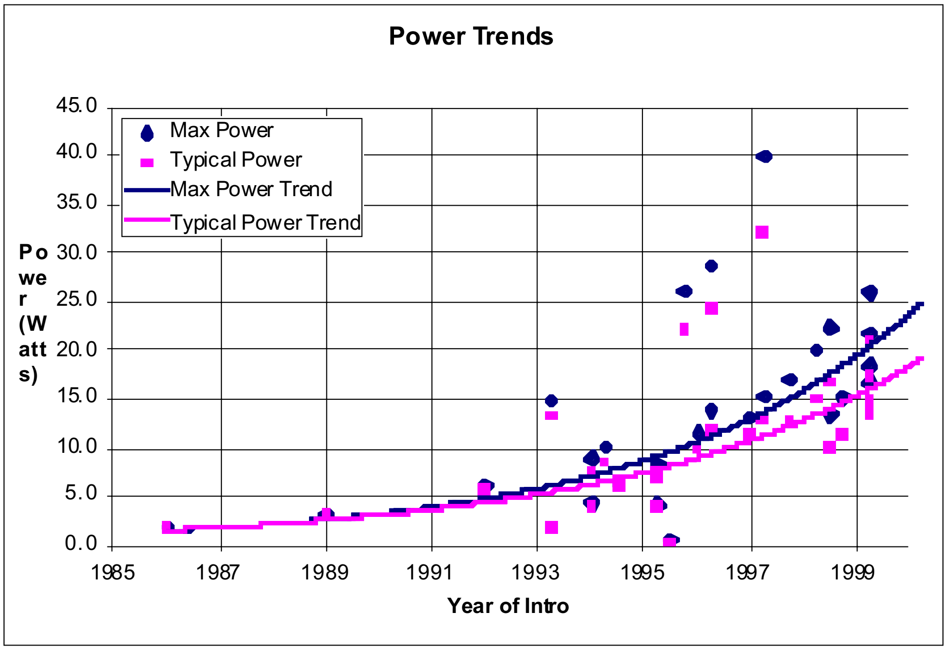
Source: [67] Processors heat development, http://www.pcsilent.de/en/tips/cpu.asp, 2004.

Table 4.10: Power consumption of AMD and Intel processors.

and reducing the circuit size and die size.

The power consumption trend of Intel processors since 1986 is shown in Figure 4-21. The general trend observed is that maximum processor power consumption grows every four years by a factor slightly greater than 2X. A second trend to be aware of is the discrepancy between maximum power consumption and typical power consumption. Since 1996, for a typical Intel Pentium processor, the power consumed when operating on a synthetic high-power workload is observed to be 20% higher than the power consumed when the same processor is operating on a high-power section of a real application [41]. The discrepancy between maximum power consumption and typical power consumption poses a difficult dilemma to the system designer.

Using Equation 4.13 and data from Table 4.10, given a total payload power of 18 kW, the Boeing 702 payload can accommodate at most 530 Intel Pentium III processors. Given a total payload power of 25 kW, the Boeing 702 Plus payload can accommodate at most 737 Intel Pentium III processors. Recall that this calculation does not take into account the power requirements of the interconnection network nor the storage devices that are required. To maximize the number of processors in the payload requires a design that balances the amount of necessary equipment (e.g., processors, interconnection network, and storage devices) and the power budget.



Source: [41]. Stephen H. Gunther, Frank Binns, Douglas M. Carmean, Jonathan C. Hall, "Managing the Impact of Increasing Microprocessor Power Consumption," *Intel Technology Journal*, 1st quarter, 2001, p. 2.

Figure 4-21: Power consumption trend of Intel processors.

Assuming that power limitation is the major constraint to sizing the processing payload, other high performance processors with lower power consumption than the Intel processor line should be considered. As other processors may have better characteristics, building satellites with the Intel processor chip provides a conservative design point in the tradespace.

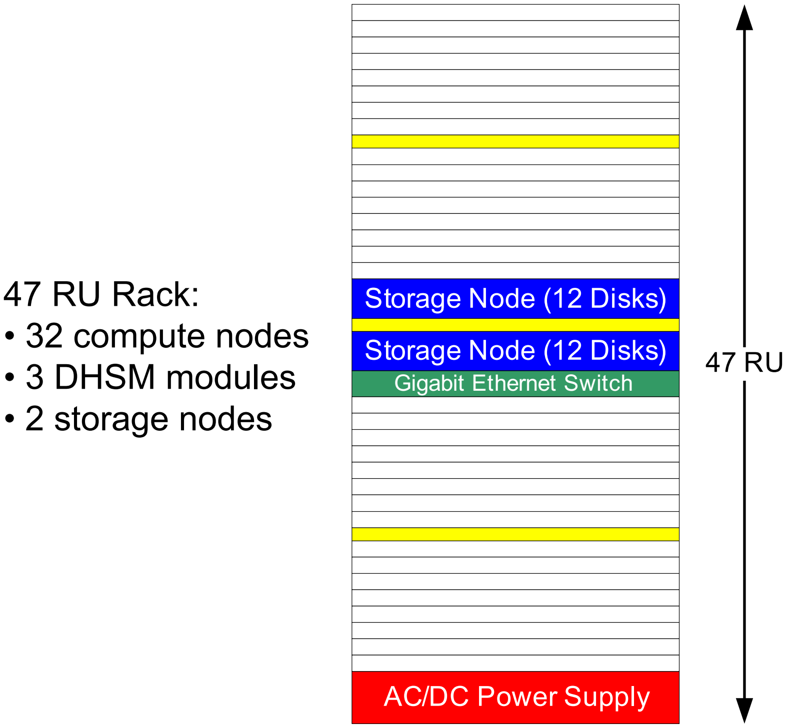
4.4.2.2 FPGA-based Payload

Increases in power consumption can also be seen in current technology trends in FPGA devices as they are being designed to operate at higher frequencies and to maximize device utilization. The power consumption of an FPGA device can be generally defined as:

$$P = \sum_{\text{at all nodes}} CV^2 f \tag{4.14}$$

where P is the total power consumption, C is the net capacitance of all interconnect and logic resources, V is the operating voltage, and f is the transition frequency. System design requirements usually dictate the operating voltage and external load capacitances, thus adjustments to the internal net capacitance and toggle frequency must be made in order to minimize power consumption for an FPGA device [86].

FPGAs can be powerful solutions as parallel architectures for digital signal processing. Examination of a state-of-the-art ground-based FPGA processing system used for high-end reconfigurable computing provides a good estimate of the size, weight, and power requirements for the processing satellite. A 47 RU rack, for example, can hold 32 BEE2 compute nodes, 3 DHSM (Distributed Hierarchical Storage Management) modules, and 2 storage nodes, as shown in Figure 4-22. The figure does not include the interconnection network chassis. DHSM provides efficient storage management, which includes pre-loading and pre-fetching. This configuration of 160 Virtex-II Pro X FPGAs provides a peak performance of 25 TOPS or 1 TFLOPS and 5 Tbps I/O bandwidth. It has 512 GB local memory and 128 GB of global shared memory. There are 16 TB of local disk and 6 TB of shared work storage. It has a total power consumption of 12 kW [84]. Notice that this implementation has plenty of power remaining in the power budget for other equipment (e.g., interconnection network, storage devices, and additional processors).



Source: [84] BEE2 Team, *Building BEE2: a Case for High-End Reconfigurable Computer (HERC)*, presentation, January 12, 2004.

Figure 4-22: Capacity of a 47 RU rack.

4.4.2.3 Hybrid Processing Payload

System designers should consider heterogeneous systems for the processing architecture. Xilinx provides such a solution. Xilinx offers embedded PowerPC chips in the fabric of the Virtex-II Pro / Virtex-II Pro X FPGA devices. A heterogeneous processing architecture can make use of GPPs at the back end for decision making and FPGAs at the front end for jobs like filtering and transforms. While the implementation of the Virtex-II Pro / Virtex-II Pro X FPGA devices makes the processing satellite inherently hybrid, system designers may consider interconnecting a separate block of GPPs (e.g., off-board Pentium chips). There may be space applications which do not need the parallel processing capability that the FPGA-based architecture provides but requires much faster processing speed than the PowerPC chips offer. The embedded PowerPCs on the Virtex-II Pro / Virtex-II Pro X FPGA devices are currently 5-10x slower than the fastest sequential processors [21]. Determining the payload sizing of the different types of processors in the architecture requires careful consideration of the number and types of applications to be serviced and balancing trades between raw speed, I/O capability, memory, and power efficiency. The spacecraft structure (e.g., thermal and cooling environment) may also affect the size and design of the processing payload that will be flown.

4.4.3 Summary

The paradigm of de-coupling processing units from traditionally-designed mission satellites allows processing resources to be shared across many different network users. Mission satellites are able to access these resources via the space-based information network backbone. This architectural concept also alleviates the need for individual high data rate downlinks and can be much more cost-effective overall. Designing the processing satellite requires an analysis of the numerous missions to be served and determining an architecture that can be flexible and generic enough to handle multiple different types of applications. The choice of processor depends not only on raw speed, but on maximizing on-chip memory, I/O bandwidth, and power efficiency. The goal is to maximize computational size in order to handle present and future applications.

The chosen processing architecture must be upgradeable to allow for insertion of the newest technology that will provide improved computational capacity within the same size, weight, and power constraints. The processing architecture should be stable and scalable in order to exploit technological advances. Technology improvements include enhancing GPPs to be capable of handling digital signal processing and lowering power consumption of processors. The next generation of Virtex FPGAs (Virtex-IV) will provide about 4-6x performance improvement and reduce power consumption by up to 50%. Specifications for the development of 100 Gbps InfiniBand and DDR2 memory up to 800 MHz, 4 GB per DIMM are underway. Not only will improvements be made to existing technology, but new technologies may emerge. For example, iSCSI (Internet Small Computer System Interface) is an Internet Protocol (IP)-based storage networking standard for linking data storage facilities currently being standardized by the Internet Engineering Task Force (IETF). Using IP-based technologies will further advance satellite data networks to accommodate IP-based applications and enhance interoperability. Regardless of the specific hardware implementation, the design of a stable and scalable processing architecture allows for the interchange of new components without changes to the application or network software, which subsequently allows for maximum software re-use.

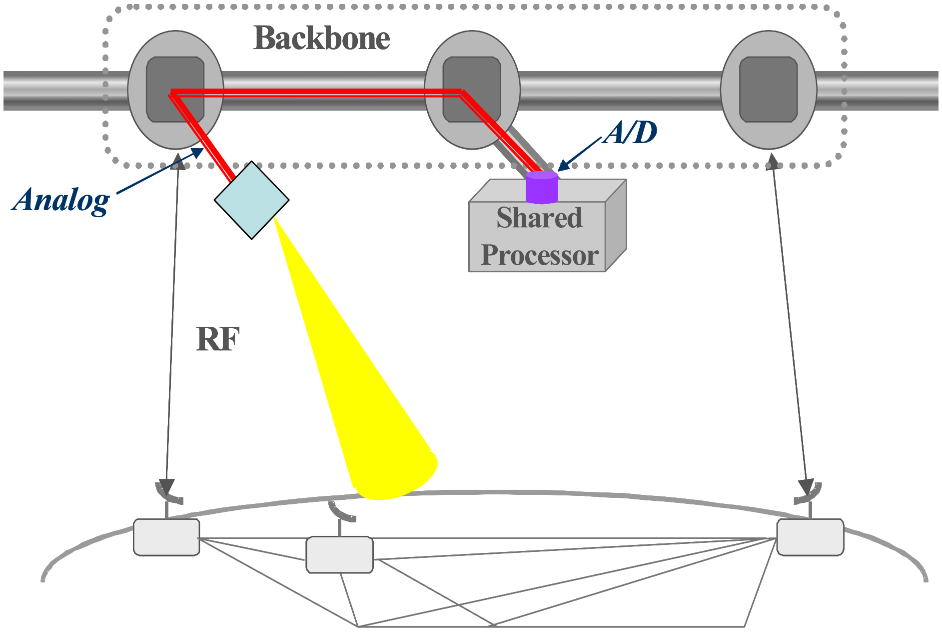
4.5 Other Space Applications

Optical intersatellite links can supply satellite networks with a significant increase in capacity and at a reduced cost than radio frequency intersatellite links. Not only can they allow the construct of a cost competitive space network but they can enable the design of new application architectures. This section discusses several innovative space architecture implications of a high speed optical satellite network. Networking allows for the sharing of connected resources (e.g., processors, data storage devices, etc.). Networking also allows for interoperability as it is an efficient method of sharing communications among multiple users. Novel ways of utilizing this network may transform satellite communication applications and many different space missions.

4.5.1 On-orbit Upgradeable Network Resources

As seen in the previous chapters, the space information network architecture emphasizes the placement of sensor equipment on one space asset, processing capability on another space asset, and the use of communications links to move the information around. The de-coupling of the different components highlights the different life cycle durations of the equipment. The antenna and RF front-end technologies for the sensor satellite do not advance as fast the processing elements. Figure 4-23 illustrates the concept of decoupling the processing elements into a separate satellite. This allows the raw RF analog signal or the digitized waveform to be transmitted to a processing satellite to perform the remainder of the receiver function via software. Both the processors and software on the processing satellite can be upgraded or reprogrammed to adopt new or better modulation, coding, media access control (MAC) protocols, and switching. Upgrades to the processing satellite can occur on a faster time-scale than the mission satellite.

As the amount of data collected for satellite applications and services continues to grow tremendously, on-orbit data storage is another network resource that should be considered. These dedicated network devices can provide easy access to data for many users via networking. Data can be stored in the form of files, such as e-mail boxes, Web content, remote system backups, weather measurements, mapping information, etc. It is highly desirable to make use of all satellite data in real-time. However, if the networked processing resources are not available due to other higher priority tasks, raw data can be stored to be retrieved and processed at a later time. Although memory devices are susceptible to SEU, techniques



Source: [20] Vincent W.S. Chan, "Optical Satellite Networks," *Journal of Lightwave Technology*, Vol. 21, No. 11, November 2003, p. 2825.

Figure 4-23: Reconfigurable and upgradeable RF satellite access network.

for designing reliable systems with certain levels of SEU protection has been shown in [51] for main memory, logic, and cache memory.

4.5.2 Distributed Computing in Space

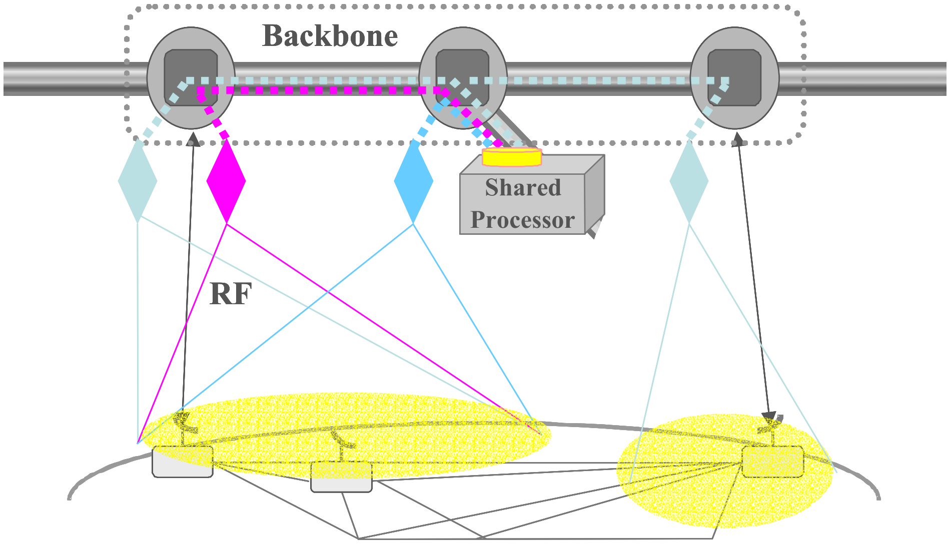
A distributed computing system in space is a collection of autonomous processing satellites that are interconnected with each other and cooperate to perform the processing for an individual task. The growth of distributed computing models has been limited by bandwidth bottlenecks, a lack of compelling applications, and security, management, and standardization challenges. There is an interest in distributed computing as processing power and communications bandwidth increases. Millions of desktops and desktop processing cycles are used for SETI@Home, the well-known worldwide distributed computing project whose objective is to locate intelligent life in the universe. The types of application tasks that can take advantage of distributed computing in space include [29] :

- Database searching against a massive database can be split across many processing satellites as a submitted query can be run concurrently against a section on each processing satellite.
- Complex modeling and simulation techniques that increase the accuracy of results by increasing the number of random trials, as trials could be run concurrently on many processing satellites, and combined to attain greater statistical significance (e.g., observing for scene changes in images [image subtraction]).
- Complex simulations for weather forecasting, geophysical exploration, geological changes, and moving platform (e.g., vehicle) changes.

The traditional paradigm of standalone computation on an individual satellite can be transformed into distributed computation on connected cooperative processing satellites. The computational burden and power consumption of an application can be evenly distributed across the network. A scheduler must determine the proper workload distribution as a function of bandwidth and available processing capability.

4.5.3 Interoperable Space Communications

Traditionally, satellite systems have not followed the paradigm of connecting disparate modalities to form a single network, which is common in terrestrial networks. The main obstacle has been that satellite systems have been designed as stove-pipes that are not interoperable. Connecting several existing satellite systems would require terrestrial gateways or teleports. Gateways or teleports are capable of switching voice, image, and data transmissions between satellites and terrestrial networks. However, they are costly and utilize a considerable fraction of the uplink and downlink resources for the connections. Furthermore, overall end-to-end network response would be noticeably delayed by the interconnections at the Application Layer. Figure 4-24 illustrates a high speed optical satellite backbone with processing resources that can be used to perform the conversion gateway function to connect different satellite communication systems in space rather than on the ground ("gateway in space"). Network management functions for the intersatellite links can also be provided. Connectivity and the resulting increase in satellite bandwidth are key to transforming the existing stove-piped satellite community into a data satellite network community serving considerably many more users. The processing required are modulation/demodulation of several different formats, coding/decoding, interleaving/de-interleaving, authentication, protocol conversion, switching, routing, and encryption/decryption. The data rates seen on the gateways are trivial in comparison to the data rates required for signal intelligence applications as far as demodulation and the other functions can also be supported if the aggregate data rates are less than or equal to 1 Gbps. Thus, the computational burden of



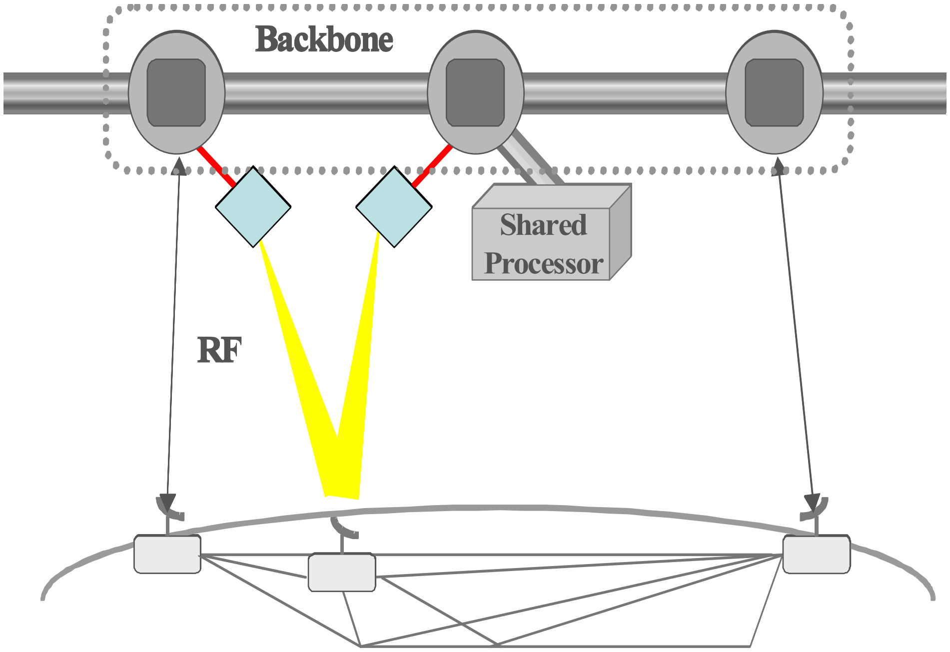
Source: [20] Vincent W.S. Chan, "Optical Satellite Networks," *Journal of Lightwave Technology*, Vol. 21, No. 11, November 2003, p. 2825.

Figure 4-24: Interoperable interconnected space communications.

gateway functionality on the processing satellites will not be very significant.

4.5.4 Multiplatform Distributed Satellite Communications

The realization of a multiplatform satellite communication system is made possible by the optical satellite network, as shown in Figure 4-25. In this system, a multielement antenna array is distributed over multiple satellites. This configuration allows for improved performance for small and low power terminals by creating a large gain electronic antenna pattern on the user and suppressing the signals of interference users by placing nulls on their signals. This technique can be done via a MAC protocol and computed dynamically in rapid response to bursty user demands. Additionally, parallelization allows for the simultaneous demodulation of many users. This can be implemented with minimum control overhead and computational complexity. The amount of bandwidth required for multiplatform distributed satellite communications is not expected to exceed the available network capacity on the optical intersatellite links. Recall that there are 400 wavelengths in the network system and each wavelength has a data rate of 40 Gbps. The processing load is largely the same as the previous example with some added for antenna processing and the use of reference clocks



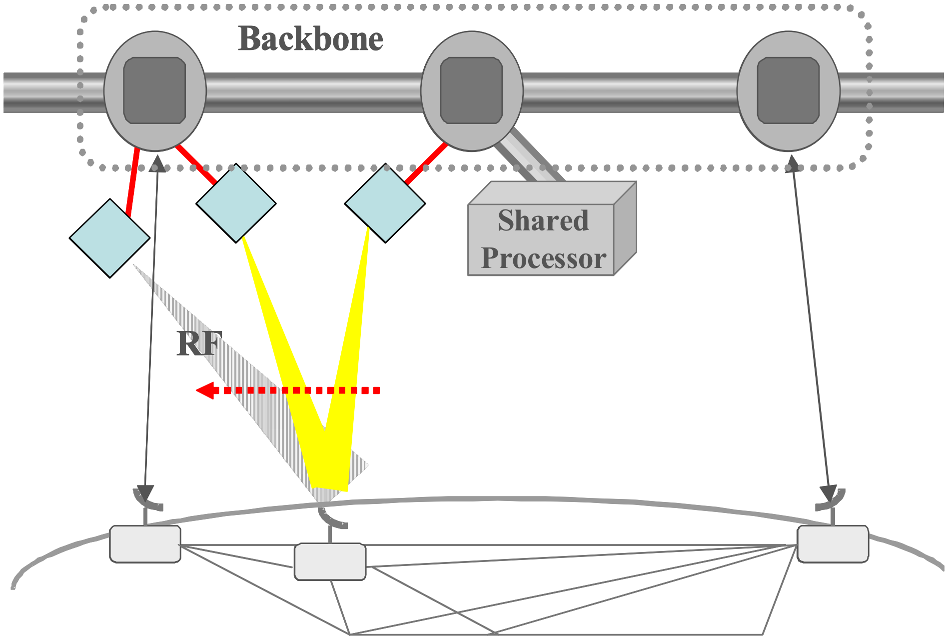
Source: [20] Vincent W.S. Chan, "Optical Satellite Networks," *Journal of Lightwave Technology*, Vol. 21, No. 11, November 2003, p. 2826.

Figure 4-25: Multiplatform distributed space communications.

for accuracy (e.g., use of timing markers for obtain accuracy of less than 1 degree of the RF carrier).

4.5.5 Coherent Distributed Space Sensing

Networks can replace single, high-cost, sensor assets with large arrays of distributed sensors for both security and surveillance applications. Distributed sensing has the advantages of being able to provide redundant and hence highly reliable information on threats as well as the ability to localize threats by both coherent and incoherent processing among the distributed sensor nodes. Thus, image-oriented sensing and object identification from space can be considerably enhanced via a distributed satellite system. Figure 4-26 illustrates a GEO-location application where two satellites can produce the arm of a long baseline interferometer as long as the two sensed signals can be assembled for coherent processing. To preserve phase information, it is necessary to have either fine quantization and significant data rate transmissions or coherent analog transmission at high fidelity (also known as transmission transparency). The baseline data rate of the VLBI (Very Long Baseline Interferometry) Space Observatory Program 2 (VSOP2) is 1.024 Gbps, based on 2 bits per



Source: [20] Vincent W.S. Chan, "Optical Satellite Networks," *Journal of Lightwave Technology*, Vol. 21, No. 11, November 2003, p. 2825.

Figure 4-26: High-resolution multiplatform distributed sensing application.

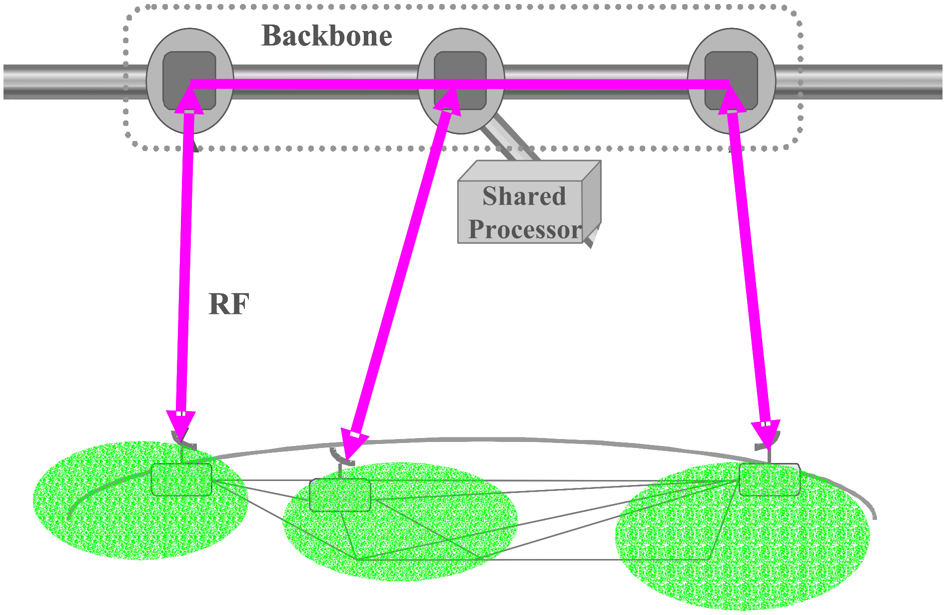
sample of a Nyquist (equivalent) 256 MHz bandwidth [81]. Instrument sensitivity can be raised by increasing bandwidth and changing the sampling rate or using analog transmission. Various practical considerations for increasing sensitivity have been shown in [81]. As the largest data rate required is 4.096 Gbps, it will not be a significant burden on the space-based network. For high-end usage of the network, consider an example of 2 SAR systems. Phase coherence can be achieved by bringing the data from each system to the processing satellite through separate buffers and using the same reference clock to gate the data out of each buffer. SAR phase coherence is attained from the use of an extremely stable reference clock to generate all RF and infrared (IR) frequencies. The maximum data rate is 2 times 1 SAR system output rate plus the referencing procedure.

4.5.6 Multisensor Data Fusion

There is an increasing interest in the use of networks for large-scale applications such as environmental monitoring, surveillance, and battlefield awareness. Traditionally, these applications relied on centralized sensor array processing where all processing occurs on a central processor. Distributed sensors require a network for collaboration between sensors. Multisensor data fusion is necessary for target detection, classification, identification, and tracking functions. Multisensor data fusion systems try to combine information from multiple sources and sensors in order to obtain inferences that cannot be realized with a single sensor or source. With the rapid evolution of computers, techniques for multisensor data fusion can be drawn from a diverse set of disciplines including signal and image processing, artificial intelligence, pattern recognition, statistical estimation. Algorithms that utilize the advantages of a network of spatially separate sensor nodes need to be developed. Cooperative data fusion techniques for analyzing the vast flow of data collected span high-level decision corroboration (e.g., voting), feature fusion, and full coherent beam formation. There is ongoing research in the implementation of efficient fusion architectures that support intelligent integrated processing of incoming data streams of large volumes of data arriving at very high rates without excessive computational complexities.

4.5.7 Restoration of Disconnected Global Networks

Today, there exists a terrestrial fiber network that serves as a backbone, connecting together subnets of different modalities to allow the Internet to operate as a single network. Parts of the terrestrial network may be disconnected as a result of natural or man-made disasters. A satellite network can act as a backup, restoring global connectivity, as shown in Figure 4-27. This architecture requires a sufficient number of gateways connecting the terrestrial network and satellite network, a network management and control strategy for discovery of surviving resources and connection using satellite network assets. The space-based network infrastructure will be able to accommodate the temporary increase in capacity demand. Spare capacity assignment for restoration of link failures can be designed with very little computational complexity. Given that restoration of IP-based terrestrial networks is well studied, the satellite network should adopt similar architectures and strategies for interoperability and internetworking with the terrestrial networks.



Source: [20] Vincent W.S. Chan, "Optical Satellite Networks," *Journal of Lightwave Technology*, Vol. 21, No. 11, November 2003, p. 2826.

Figure 4-27: Network for reconstitution, reconnection of disconnected terrestrial networks.

4.6 Summary

Providing a space-based information network backbone with shared on-orbit processing resources can help increase computational capabilities, levels of inter-spacecraft communications, and interconnect a multitude of space assets and users. Processing capabilities allow for the handling of the large volumes of data generated in space. Eliminating the need to transmit raw data to the ground for processing can allow for more rapid access of space-processed data to the end user. Processing in space can also reduce the amount of data that is disseminated. Ultimately, a processor acts as a bandwidth compression device.

The examples of spectral analysis and SAR data processing are used to highlight the possibility of using commercial processors for real-time data analysis. Both applications utilize algorithms that are largely made up of FFT calculations. This commonality allows for the sharing of a single processing satellite. Given the vast amount of raw data for each application, the technique of parallel processing and/or time-sharing on the same processing satellite affords efficiency and low cost. Onboard processing on each mission satellite is not required due to the availability of a shared processing resource.

FPGA-based processing systems are superior for real-time digital signal processing applications. GPP-based systems are more user-friendly and more suited for general applications (e.g., Internet multimedia applications, desktop publishing applications, etc.). However, the hardware implementation of the processing architecture for the processing satellites is not necessarily a choice between GPPs and FPGAs. It may be reasonable to develop a hybrid processing architecture with these two different types of processors. Although today's FPGA platform chips contain mid-range GPP processors, such as the PowerPC chip, including off-board GPP-based systems in the same satellite may be valuable to applications that do not require the parallel architectures of FPGAs but do require powerful raw speed. Additionally, considerations for the interconnection network, storage devices, and power limitation of the spacecraft may dictate the balance between the two types of processors in designing a hardware architecture that is flexible for many different algorithms and applications. To evolve with the technological advances of components such as processors, interconnection networks, and storage devices, the processing architecture must be stable and scalable.

The space information network can also enable innovative distributed space systems for sensing and data fusion. Processing is not necessarily the only shared on-orbit network resource. Data storage in space can also be considered. Data storage may be necessary for large amounts of non-critical raw data that can afford a small time delay. Other space applications enabled by high-end space-borne processing include distributed computing in space, interoperable space communications, multiplatform distributed satellite communications, coherent distributed space sensing. Additionally, connectivity to the space information backbone can serve as backup communications by providing protection and restoration to global network failures on the ground.

Because it is hard to predict what applications may be developed in the future, it is important that the space-based information network and processing satellites be designed not finely tuned to specific known applications but to support a broad range of applications with satellites in all viable orbits, reasonable data rates, processing power, and adequate memory. These attributes will be difficult to support in the old space paradigm where communications and processing hardware co-exist with the prime payload in the mission spacecraft. The decoupled, rapidly upgradeable paradigm that is proposed in this dissertation reduces the need to project future requirements from 10 to 2 years, which is a lot less risky and more accurate.

Chapter 5

Infrastructure Investment and Development

Satellite systems may be very expensive but they offer the opportunity to accomplish missions which cannot be conducted on the surface of the Earth, or otherwise may be performed more effectively or efficiently from space [75]. Space-based information collection and dissemination capabilities enable the American military to support strategic operations. For example, military forces can quickly communicate missile attack warnings, navigate through areas of conflict without encountering hostile forces, and strike identified targets of interest from land, sea, or air with precision. The difficulties involved in transmitting real-time actionable intelligence information directly to the lowest possible echelon of forces are numerous. Not only are there technical, bandwidth, and data fusion obstacles, but there also exists many disparate legacy systems that need to communicate with each other. Compounding these problems are the political, cultural, and organizational challenges related to advancing space systems technology and support [43]. A core contribution of this dissertation is the analysis of space technologies as being sufficiently mature and numerous to be best considered an infrastructure. An infrastructure is defined as the underlying foundation or basic framework for a system. This chapter examines direct U.S. government investment for a future space-based information network infrastructure that can support and serve a mixture of technical components (hardware, software, and communications facilities) and users.

Infrastructure investments are difficult to value because their benefits are elusive: spread across many areas, not easily recognized locally by individual users or missions and contingent upon subsequent investments. The traditional source of funds for infrastructure development has been public finance [62]. The common justifications for government involvement include: (1) serving government missions, (2) addressing private market failures, and (3) dealing with private underinvestment in research and development because public goods, such as national defense, will always be undersupplied by a free market economy [9, 82]. Today, some believe that investing in an infrastructure to sustain business in space is a necessary prerequisite for stimulating widespread commercial space development, while others believe that satellite communications have become commercially viable and that industry should be required to invest in future technology [69]. This chapter directly addresses the implications for the commercialization of the architecture that is proposed in this dissertation. Case studies regarding past direct government funded defense projects and global mobile satellite communications businesses are studied in order to recommend a technology policy regarding the development of a space-based information network infrastructure. A technology policy is defined as policies involving government participation in the economy with the objective of influencing the technological innovation process [82]. The lesson learned from the case studies is that choosing a generic infrastructure design offers the ability to choose commercialization pursuits. This chapter explicitly does not address the security issues in such a choice, as such an analysis is best informed by information reasonably assumed to be sensitive or classified.

5.1 Issues and Motivation

One of the most difficult questions to answer is that of what should be the role of the national government in the development of new satellite technologies and systems. The major organizational stakeholders in U.S. military space include the NRO, the IC, and the DoD, which includes the Army, the Navy, and the Air Force [43]. Military space operations today are faced with many budgetary constraints. Each stakeholder does not have the financial flexibility required to unilaterally support the capabilities it demands. These constraints result in a vital need for the space-based information network infrastructure as it is the common basis for connecting multiple users and providing communications and services among them. It may be necessary for the military to undertake cooperative arrangements with civil agencies and the private sector, which may call for the major organizational stakeholders to become active participants in policy debates concerning civil and commercial space systems and technologies [2].

In the U.S., the government has led the way into space but has not made the changeover to a private sector enterprise completely. In the past, the paradigm of leading the way and then stepping aside has been successful for the government in its role as a facilitator. For example, in settling the American West, the essential precursor infrastructure, military guarantees of protection and rights of way on post roads, enabled private industry to build railroads. Air travel, another example, began in America due to government investment of aeronautical research and development. The momentum for the development of private airlines was provided by the government through the Air Mail Act of February 1925 and amendments in June 1926, which gave Postmaster General the authority to contract with air carriers for U.S. mail delivery [49]. These contracts gave new airlines the financial incentives adequate to become operational, including continuing R&D. Since the Cold War, the aerospace industry has grown to be the defense industrial base of the U.S.. The survival of American defense contractors and the space industry is directly affected by U.S. defense projects.

The Cold War space paradigm is characterized by the use of space assets and activities to achieve national foreign policy goals of international power and prestige [49]. The U.S. is no longer faced with the same situation. A new paradigm must emerge for activities in space. The space paradigm has an opportunity to shift its focus to the future by determining the areas of technology convergence that aid in the construction of a common R&D infrastructure and increasing the role of the private sector in space development. With the inertia of and resistance (active and passive) by those benefiting from the status quo, government directed space policy will at best evolve incrementally [49]. Change may be disruptive but it allows for movement in new directions by pursuing opportunities created by change. The new paradigm introduced in this chapter involves coordinated investment that leverages optical satellite networking and commercially available processors and network protocols to build a generic space-based information network infrastructure.

5.2 Case Studies

It is useful to examine what lessons can be learned from past government participation in technology development. Examples of U.S. government intervention include the railroad, automobile, telephone, radio, and television. The outcomes are not at all times positive as support to one industry may be devastating for another. For example, the development of a federal highway system aided the automobile industry but led to the demise of railroads. Another negative is the unintended consequences of government regulation and standardization of an industry. In the case of telephony, American Telephone & Telegraph Company (AT&T) reaped more benefits than the consumer [49].

Large complex engineering projects funded primarily by the government to support military missions offer insights into the potential for a space-based infrastructure. The case studies selected here are GPS, the U.S. Interstate Highway System, and ARPANET, the precursor to the Internet. Each of these infrastructures required federal government funding. In each case, the net benefits to society from investment were larger than the benefits that private individuals or firms could offer. The government provided direct funds, mitigated risks, and served a critical coordination function. In addition to these case studies, an examination of the commercial market for global mobile satellite communications is included.

5.2.1 Global Positioning System

GPS, originally known as NAVSTAR (Navigation System with Timing and Ranging), was developed by the DoD and deployed by 1993 to provide military ground, sea and air forces with all-weather round-the-clock navigation ability. In its defense role, GPS has demonstrated excellent capabilities. Today, GPS has become an important asset in many civilian applications and industries around the world (e.g., corporate vehicle fleet tracking, surveying, boating, aircraft, travel directions). A number of the current applications were not envisioned in the initial development and operation of the system in the early 1970s [58]. GPS was made available to the commercial sector only after being pressured by the companies that built the equipment who saw the enormous potential market for it. Thus, GPS satellites broadcast two signals, one for civilian use and one that only the military can decode. The characteristics of the two GPS signals are shown in Table 5.1. The DoD

	Precision Positioning Service (PPS)	Standard Positioning Service (SPS)
Users	U.S. and Allied military; U.S. government agencies; Selected civil users	Civil users worldwide
Access	U.S. government approval; Requires cryptographic equipment and keys and specially equipped receivers	Free of charge; No restrictions
Horizontal Accuracy	22 meters	100 meters
Vertical Accuracy	27.7 meters	156 meters
Time Accuracy	200 nanoseconds	340 nanoseconds

Table 5.1: Global Positioning System signals.

intentionally degrades the Standard Positing Service (SPS) accuracy. Precision Positioning Service (PPS) signals are only available upon DoD authorization. Civilian demand for more accurate GPS data has resulted in a technique called Differential GPS (DGPS). DGPS uses a receiver at a fixed location to broadcast corrected GPS signals to nearly mobile receivers. Thus, accuracy to within 1 to 3 meters has been available to the public.

In 1996, the White House reaffirmed peaceful scientific, civil and commercial use of GPS services globally and at no cost. The policy change in open access by the U.S. government granted free use of GPS to the world. The U.S. policy on GPS is based on balancing the basic requirement of retaining military advantage of the technology with considerations of commercial and international policy. The U.S. government is committed to provide a stipulated level of service from GPS free of charge. This guarantee has allowed considerable investment to be made by industry in the development of hardware, software, and systems that, to be viable, depend upon the long-term availability of GPS signals. The private sector has successfully recognized and captured the value of GPS. The ground-based equipment market is currently dominated by commercial sales [2]. GPS-based services are delivering valuable products and services worldwide in areas such as civil aviation, travel directions, corporate vehicle fleet tracking, land surveying, and public safety in air navigation.

Presently, protection of U.S. military interests is done via DoD control of the space segment and global economic growth is promoted via commercial competition in ground-based GPS equipment. U.S. economic and military interests can be facilitated by the DoD through the quick incorporation of GPS into its own force structure and the acceleration of foreign military sales to ensure that GPS is adopted by allied forces [48]. However, many countries, predominantly Europe, have strong reservations about reliance, and consequently dependence, on a system controlled by the U.S. military. For them, the obstacle to the development of a separate radionavigation satellite system will be paying for the new system, as they may not be as motivated as the DoD, for they have been using the existing GPS system for free [49]. Nevertheless, the European Union (EU) decided in 1999 to explore plans to develop a satellite navigation system of their own, to be called Galileo [64]. The development and validation phase of Galileo is currently underway. Constellation deployment plans are scheduled for 2006 to provide service beginning in 2008 [38].

Private investments in GPS and remote sensing are bringing new technical capabilities to the world market and subsequently producing new opportunities (and risks) for military space operations. The U.S. military obtains significant value from secondary network effects. Components common to military and civilian applications are less expensive because of economies of scale. Individuals trained on commercial GPS are valuable service personnel due to reduced training costs. The lesson learned from GPS is that general purpose systems, when shared, can create value for the military both through secondary network effects and value to allies. The lesson not applicable to the space-based infrastructure is the zero marginal cost of shared broadcast. What is applicable is that a two-tiered system may be feasible.

5.2.2 U.S. Interstate Highway System

The primary purpose for the construction of the Dwight D. Eisenhower Interstate System of Interstate and Defense Highways, approved during one of the most unstable periods of the Cold War in 1956, was to support national defense. A well-organized national highway system could convey vast amounts of military equipment and supplies and large numbers of military personnel from one place to another. The national highway system continues to perform a vital national security function as the U.S. military's Strategic Highway Corridor Network (STAHNET) mainly relies on the interstate highway network [23]. America's strategic advantage in effective surface transportation is afforded from the accessibility and availability of a potential resource that could be reliably called upon in times of crisis.

The U.S. Interstate Highway System, the largest public works program in American history, was also created as a way to resolve the disparity between the insufficient level of highway facilities and the extraordinary demand for automobiles and automotive travel during the postwar period in the U.S.. Today, as traffic congestion continues to rise, surpassing the projected capacity growth, the urban interstate highway system has continued to operate efficiently. Additionally, the U.S. Interstate Highway System has had a tremendous effect on the country, drastically enhancing economic efficiency and productivity. Users benefited from increased mobility, reduced travel time, reduced operating costs, and expanded options for a higher quality of life, all of which contribute significantly to economic growth [23].

The lesson learned from the U.S. Interstate Highway System is that federal investment can create additional capacity that, when used, has tremendous economic benefits. The secondary effects of investing in the highway system include training in automotive systems design and maintenance used for tank warfare and ground warfare. The lessons not applicable to the space-based infrastructure is the step-wise incremental investment and the non-trivial amounts of financial contribution from the states. What is applicable is that highways are a generic infrastructure. The government did not try to enforce any but the most minimal standards on civilian automobiles.

5.2.3 ARPANET

ARPANET began as a low cost computer-to-computer network, developed in the early 1960s by the Defense Advanced Research Projects Agency (DARPA), the central research and development organization for the DoD. Built for military use and to test packet switching technology, ARPANET was a fully government funded network connecting the DoD and its contractors. The research institutions that were connected included Stanford Research Institute, the University of California at Los Angeles, the University of California at Santa Barbara, and the University of Utah [24].

ARPANET provided a technology development testbed as packet switching networks were a controversial concept in the 1960s. Early papers on the project addressed the technical goals of file transfer and remote login. The vision of collaborative communities was made possible through time-sharing systems that provided the ability to share programs, data, and hardware across the network [68]. The success of ARPANET then led to proposals to develop similar networks for non-defense uses.

The high degree of scalability, simplicity of design, and minimal centralized organization in distributed networking allowed for the explosive growth of the Internet. In the early 1980s, the National Science Foundation (NSF) funded NSFNET (NSF Network) to link supercomputer sites and networks around the world, thereby expanding the backbone of the Internet. Today, the Internet connects hundreds of millions of computer users and has become an example of a successful public private network. The Internet has grown beyond its initial role for universities and the research community, becoming less of a private network and more of a publicly-accessible network of networks [85].

The Internet can be characterized as a bottom-up collaborative organization. It is a generalized infrastructure supporting multiple functions and protocols that has benefited financially from federal and private investment in very high speed networking technologies, private investment in computers and internal networks, and carrier investment in local and long distance fiber optic networks motivated by the market for conventional voice and fax and special purpose data networks in addition to projected future demand for video services. A great deal of the technical design and progress has been the result of volunteers working through the ad hoc IETF[50]. Killer applications that emerged with the Internet include e-mail, web browsers, instant messaging, social networking sites, and e-commerce. A "killer" application is defined as an application program that intentionally or unintentionally gets a user to make the decision to buy the system on which the application runs. The classic example of a killer application is the spreadsheet program VisiCalc, later followed by Lotus 1-2-3, which introduced the value of desktop computers to the enterprise. Today, a killer application can refer to a generic type of application that has not existed before, to a particular product that first introduces a new application type, or to any application with wide appeal.

The lesson learned from ARPANET is that utilization of existing infrastructure can simplify rollout because the price of entry is lowered. Rollout is defined as the process of introducing a new product or service into the marketplace. Another key lesson learned is that expandable infrastructures of general-purpose information and communications technologies can expand for unforeseen uses. The Internet provides much more than file transfers and remote logins. What is applicable to the space-based information network is the sharing of the infrastructure between the military and civilian sectors. The Internet consists of overlay networks (networks built on top of one or more networks). The `.mil` domain is under strict military control. Even though it is connected to the worldwide Internet, `.mil` can set its own policies. Furthermore, the network is tightly connected, using the same hardware for the commercial sector and the military.

5.2.4 Commercial Global Mobile Satellite Communications

Satellite networking is an emerging market with enormous opportunities to provide many telecommunications services. Broadcast satellite to the home using digital video broadcast technology is the most visible satellite service to date. Broadcasting will continue to be attractive for video, audio, and data delivery purposes because the downlink signal is available everywhere within the footprint of a satellite. Global mobile satellite voice and data services are ideal for applications characterized by locations with unreliable infrastructure or no infrastructure, such as aviation, construction, disaster relief/emergency services, defense/military, maritime, mining, forestry, oil and gas, and leisure travel. Although satellite systems are great in their ability to reach many users and work in remote areas, they have not been able to compete in price with services provided by terrestrial technologies (e.g., fiber optical links and cellular technologies). Satellite technologies have not been able to replace these access mechanisms. The problem has not been that the technology did not work. Rather, the difficulty has been the ability to build a commercially viable global mobile satellite communications system, i.e., gaining market acceptance, as illustrated by the examples of Iridium and Globalstar. Note that there are satellite companies that are successful, such as INMARSAT (International Maritime Satellite Organization) and INTELSAT (International Telecommunications Satellite Consortium), but they both started as intergovernmental organizations. INTELSAT was formed in 1964 to establish the first commercial global satellite communications system (i.e., carrier services and broadcast and video services). It first began with 11 participating countries, but now has over 100 members and services over 149 countries [57]. INMARSAT was formed as a maritime-focused intergovernment organization in 1979 to provide mobile satellite communications worldwide.

5.2.4.1 Iridium

Iridium, began in 1989, is a satellite network system consisting of 66 LEO satellites providing wireless telecommunications service anywhere on the globe. Satellites communicated with each other via inter-satellite links. The FCC (Federal Communications Commission) application was filed in 1990 and clusters of satellites began to be launched in 1997. The venture cost $5 billion to construct and maintain and required 800,000 users within five years to be viable. The targeted market segment was business customers in remote areas where service would be valuable because wireline connections were limited. Customers had to be willing to invest thousands of dollars in the telephone handset in addition to paying for access at a rate beyond that of international calling rates (e.g., $3-$8 a minute) [10].

In terms of technology diffusion, Iridium had trouble crossing the chasm of early adopters to early majority, as shown in Figure 5-1. Diffusion, in the business sense, is defined as the process by which a new idea or new product is accepted by the market. Iridium had projected 1.824 million users by 2001 and 3.224 million users after 10 years, in 2006. By August 2000, Iridium had only delivered service to 10,000-20,000 users and was forced to file for bankruptcy protection. Iridium was originally designed as a voice-only network. Its technical design made it difficult to leverage the value of text messages, image transmission or other emergent applications. To date, Iridium is still operational and provides service to the DoD and commercial, rural, and mobile sectors.

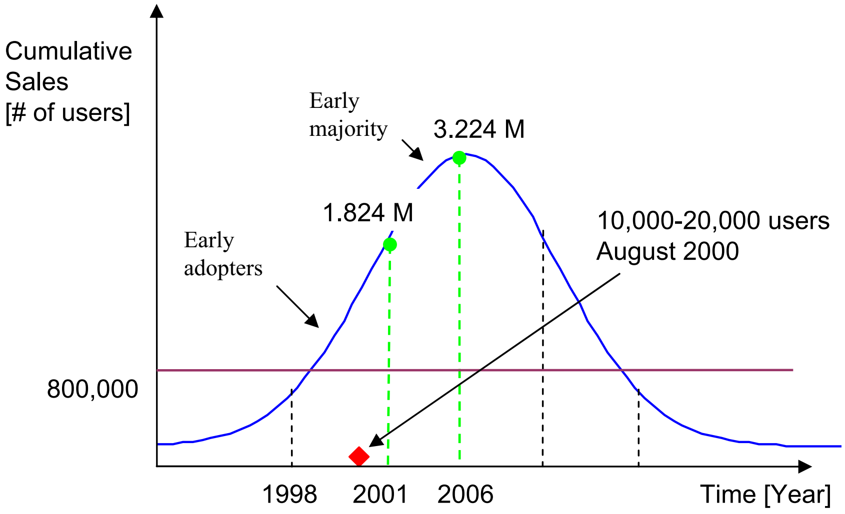


Figure 5-1: Iridium's projected diffusion curve. (Not drawn to scale).

5.2.4.2 Globalstar

Globalstar filed an FCC application in June 1991 for a $2 billion mobile communications system that consisted of 48 LEO satellites cooperating with existing public mobile networks, public switch telephone networks (PSTNs), government networks and private networks. The system aimed to provide low-cost, high-quality telephony and other digital telecommunications services such as data transmission, paging and facsimile. Users on the Globalstar communications network would make or receive calls using hand-held or vehicle-mounted terminals that were able to switch from conventional cellular telephony to satellite telephony as required. The large size and weight of the Globalstar hand-helds are similar to that of the cellular phones of the late 1980s. A subscriber's phone goes through one of 48 LEO satellites down through a gateway connected to a PSTN and on to the party called. If the call is destined for another portable Globalstar phone, the call will continue to a second gateway, up to a satellite and back down to the receiver. Globalstar's design architecture led the company to face many challenges and issues in obtaining license agreements for the location of international gateways. By 2001, Globalstar had to announce bankruptcy because the company had trouble convincing customers of the advantages of satellite service. To date, Globalstar is still operational and provides service for a variety of communications needs.

5.2.4.3 Failures to Capture and Deliver Value

The examples of Iridium and Globalstar have shown that private businesses, thus far, have not been able to gain a widely acceptable satellite telecommunications services consumer base for global mobile satellite voice and data services. While useful for the military, maritime and oil rig workers and outdoor adventurers, satellite phone service has not been accepted by the mainstream population. The problem is not the technology, but market acceptance (e.g., technology diffusion). Obstacles to market acceptance include clunky handsets (in the face of ever smaller cellular phones), inability of satellite phones to work indoors and the expensive price of a call (usually several dollars a minute).

In summary, the following factors are known to be general reasons for private business failure in satellite telecommunications:

- *Network externalities:* Unlike other infrastructures, this problem is specific to telecommunications because the value of the network increases as more users connect to it. The first subscribers to a network are usually those who highly value communications within their small community. Users with less to gain will postpone connecting to the network until prices drop. The concept of network externality arises as the network expands. The existing subscribers' average value of membership increases as more users join the network and connectivity spreads. While network externalities have applied to many telecommunications networks, it does not directly apply to the space-based infrastructure because the satellites systems are connected to the PSTN. Scaling and secondary or indirect effects exist here but direct network effects do not.

  - *Time to market is crucial in the presence of competing technologies.* Market forecasts for the growth of satellite communications did not adequately take into account the competition and rapid growth of terrestrial cellular service.
  - *Competitive advantages in niche markets are not exploited.* An independent global wireless network is an excellent tool for the military in areas of conflict, journalists, aid workers and refugees in remote areas. Other commercial niche markets include travel cruise ships. Providing to these vertical markets in the beginning might have increased the number of subscribers and revenues.

- *Lack of information to the public:* Users have to be informed about available telecommunications services. For example, Globalstar had found that its cellular network partners were not promoting the Globalstar satellite phone service due to the fear that satellite phones would cannibalize the terrestrial mobile service sales.

- *Price is too prohibitive to develop an adequately large market.* For example, the cost of an Iridium telephone handset was initially \$3,000 and later reduced to \$1,500. Potential customers are usually put off by large unit prices and often do not invest in new technology and unknown applications.

- *Lack of system flexibility or evolution:* Both Iridium and Globalstar were designed to carry voice traffic, competing with terrestrial cellular services. These systems were not able to evolve to meet the rising data demand in the market.

The lesson learned from analyzing failed business cases in global communication satellite systems is that government investment can span the gap between the early adopters and the early majority. Efforts include supporting users and stimulating adoption and diffusion of applications and innovative services. Legislative markets and regulatory approaches are vehicles to achieving equity of access and prices.

5.2.5 Summary

The examples of GPS, ARPANET, and the Dwight D. Eisenhower System of Interstate and Defense Highways derived their origins from military needs for national defense purposes. These infrastructures provide numerous applications for both the defense and commercial sectors. The GPS experience is one of the most positive manifestations of the commercial application of a military space system. The system being exploited already exists in orbit and was placed there for reasons that justify its existence without commercial involvement. Commercial applications are being built upon the opportunity presented, not demanding expensive and untried technologies to be placed in orbit. While GPS is a government launched and controlled system, it would be uneconomical for the private sector to not take advantage of the existing system and the existing technology.

However, applying the term “dual use technology” to GPS is a misnomer. Not only are other groups seeking the technology of satellite radionavigation, but access to the system itself. In this sense, GPS is similar to the Dwight D. Eisenhower System of Interstate and Defense Highways. The major commercial and private benefits of the interstate highways are derived from the direct use of the system rather than derived from indirect new road-building technology that could be sold or exported. Both systems are shared by the military due to their commercial and private utility. However, GPS is dissimilar to the Interstate Highway System in the sense that GPS resources can support an infinite number of users whereas highways have limited capacities.

ARPANET serves as a model of publicly supported basic research. Public funds have supported networking research in the interest of increasing communication and cooperation among academia, industry, and government researchers. The Internet has emerged as a decentralized quasi-public infrastructure of autonomous network domains, no longer limited to the military or the academic research community. Externalities of networking and a decentralized multidimensional market environment have promoted the growth of the Internet [50]. A great deal of the driving technology resulted from the development of the compute industry. None of the Internet killer applications existed when ARPANET was built for time-sharing. Similarly, new applications may emerge for a space-based information network infrastructure.

National defense missions cannot rely on the private sector to provide global mobile satellite communications. Even if global mobile satellite communications had been a viable market, the major disadvantages of military usage of commercial satellites include issues of performance, access and control, and security concerns. Commercial organizations may be unwilling to have their systems carry potentially life-threatening military communications in times of crisis. Additionally, the operation of these satellites would then no longer be under national control.

5.3 Development of a Space-Based Network Infrastructure

The economics of space (e.g., high cost, high risk, and low profits) provides a rationale for continued federal funding of R&D space programs. The general justification for government investment is the provision of a public good. National defense is the classic example of a public good. The term *public good* implies that it should be made available by the government. A public good is nonrivalrous and non-excludable in use. Nonrivalrous means that one person's benefit does not diminish another's opportunity to benefit from it as well. The reception of GPS signals is not affected by the number of users. A non-excludable good means that, once it exists, it is difficult or impossible to selectively deny the benefit to particular persons. For example, a user of GPS cannot deny another user from receiving GPS signals. Because GPS is a one-way broadcast, users need only receivers to pick up the signals. GPS could be made excludable by using many different encryptions. However, all non-encrypted broadcasts are non-excludable.

Network capacity in space, however, is a precious resource and is not yet infinitely abundant. Military reliance on space systems have grown tremendously throughout the last decades and will arguably increase more quickly in the future. The sharing of a space network infrastructure by the military with the private sector may decrease military utility due to a reduction in system availability. In the case of the U.S. Interstate Highway System, this reduction in military utility has not proven to be a problem because the military need for the interstate system has been relatively small compared to the total available capacity. The situation in space is not the same. Capacity constraints will dictate policies of access and control.

As military requirements increase and military budgets decrease, new policies in the development of space systems may emerge. The reduced budgets compromise the flexibility of the military to unilaterally sustain the capabilities it needs. Thus, the military might need to seek cooperative agreements with the private sector and civil agencies. All space applications, whether scientific, Earth observation, military or communications, either produce or communicate data. Opportunities exist for mutually beneficial cooperation to develop a space infrastructure that can effectively support activities whose end product is data. These opportunities arise because of the dual use nature for applications such as communications, satellite navigation and position location, remote sensing and environmental monitoring, and space launch.

Dual use systems present difficult problems in defining separate policy directions for military and civilian satellite systems. As spaceborne technical capabilities are becoming less of a dividing factor between civilian and military satellites, economic considerations may lead to the development of single satellite systems to serve both civilian and military missions. Today, most of the technologies and systems used in satellites for remote sensing are nearly indistinguishable, at least in kind though not in performance, to those used in reconnaissance satellites. Resolution requirements for civilian applications are starting to approach those of certain military applications. The obstacles to developing dual use satellite systems may be the political and security ramifications of such an endeavor. The dual use issue is complex because while civilian satellites would likely require upgrades in order to fulfill military objects, military satellites are inherently capable of dual use. Although dual use missions are becoming more common on *passive* military satellite systems such as GPS, the question remains of whether dual use is possible for *active* military satellite systems [2].

The question of applying dual use to every area of satellite applications is difficult to answer because a variety of new military and civilian uses for satellite systems have yet to be developed. For example, the characteristics and market potential of telemedicine are still unknown. Likewise, it is uncertain whether new technologies that will come up on the open market will have a fundamental effect on international security. For example, civilian applications for surveillance may compromise the positions of military forces in areas of conflict. In any case, the only areas where a separate military technology should be sustained are (1) technology that is unique to military application and (2) technology deemed critically sensitive [55]. Nevertheless, GPS has shown that even military systems can obtain value from secondary effects of a large, related market. The Internet is another example that illustrates that military control can cohabitate with civilian control. Today, any satellite designed for dual use would have to be negotiated between the government and the civilian sector.

The architecture presented in this dissertation allows for shared use of civilian and military satellites. The design of a spaced-based infrastructure enables a parallel and mutually beneficial investment on the basis of common standards. The policy for dual use can be designed into the system without requiring too much technical change or increased cost. For maximum interconnection and interoperability, open standards should be used. Because telecommunications infrastructure includes multiple networks with different functions, capabilities, patterns of ownership and use, an important role of government is assuring interconnection and interoperability so that society can gain the maximum value from both public- and private-sector investments. However, because many parts of the space-based information infrastructure have not reached the status associated with mature product markets, one can make a case for government regulation of standards. Government intervention is justified when minimal standardization is lacking and technical chaos is commonplace. Sometimes, the government may try to pick a winner. Other times, the government may emphasize research and development and support diffusion policies.

In general, standardization allows designers to foresee interconnection requirements and improve system parts. Standards allow consumers to make investments in assets (e.g., satellite phone handsets or terminals) and be assured that the assets' value will not depreciate due to loss of connectivity. Yet, this coordination comes at a price because standards restrict the options that users and vendors have. Both system users and vendors become locked-in to a set of technical products that they may change only at a high cost (e.g., switching costs). Lock-in is particularly costly when technical capabilities change quickly. Moreover, the government generally does not have the necessary competency to pick technology winners. Therefore, government is best advised to pick the minimal set of standards (e.g., lane width, IP) to ensure flexibility and interoperability as long as these standards are appropriate for efficient space applications.

The creation of shared standards for space-based information and communication technologies can allow government parties to engage in coordinated, effective investment. The vision of a convergence of many expensive, hardened, and isolated single purpose space systems into a more flexible and affordable infrastructure allows for non-duplication efforts by the many stakeholders involved in space activities. A coordinated effort among multiple organizations can help to reduce costs. Open standards allow for greater interoperability and the sharing of resources. The transformational policy changes for space-based systems go hand in hand with the transformational change in space system design.

Consequences of providing for commercial and private usage up-front may require technological changes in system design, thereby driving up costs. An open access policy for military systems also requires studying security issues. Although security is a significant issue to consider in the design, security concerns are beyond the scope of this dissertation. However, it is worth pointing out that even with a dual-use system, the space-based network infrastructure would remain under military control, similar to GPS.

Opening new markets and applications for space activities requires both cheaper access to space and shared dual use and more customer tailored space services for commercial and military users. The provision of advanced value-adding technologies (to widen the spectrum of applications), a more consistent data flow ("data on demand"), and data distribution capabilities should increase user demand and market size [25]; all of which can be met with a space-based information network using optical intersatellite links and providing large computational power.

Regarding the space-based information network infrastructure that has been presented in this dissertation, it is recommended that the space-based information network infrastructure be designed for the generic nature of data transport. The generic nature helps to satisfy the goals of integrability, interoperability, flexibility, scalability, and allows the system to be evolutionary. Iridium and Globalstar are examples of global mobile satellite communications that were initially designed for voice communications and were unable to evolve to meet various data demands. Additionally, it is recommended to design general purpose processing satellites. One question of interest yet to be answered is whether there is a market for generic spaceborne processing. GPS is a military space system which became a commercial success because geographic information became a commodity. The proposed space-based network architecture with shared on-orbit processing may make processing in space a commodity.

Presently, the government should move forward and design the space-based information network infrastructure to meet its current needs, keeping in mind the nature of a coherent infrastructure that can allow for commercial utilization, investment, and possible expansion. Designing military systems with a generic nature allows for expansion to accommodate civilian demand. Indeed, a generic generalizable design will offer flexibility to the government as well as the private sector. Market demand for space-based processing is uncertain and killer applications have yet to be designed. Nevertheless, the government's payment for the NRE cost of a space network will lower the price of entry for the commercial sector. Future innovations will make the system valuable.

Chapter 6

Conclusions

Space systems are used by the military and government organizations for a broad range of activities that include communications, remote sensing, imaging, navigation, positioning, surveillance, and reconnaissance. As military requirements increase and military budgets decrease, a new paradigm in the development of space systems has to emerge. Future space missions require a push in the state-of-the-art in critical technologies and the development of new technologies that can revolutionize space systems design and operations. There exists an opportunity to shift the space paradigm to the future by building a common space-based network infrastructure and increasing the role of the private sector in space development. This dissertation explored the architectural design of a space-based information network backbone that acts as the transport network for mission satellites as well as enables the concept of decoupled, shared, and perhaps distributed, space-borne processing resources for space-based assets.

As a result of increasing technology maturity and the convergence of computer technology and space technology, there is a growing interest in the use of onboard processing in space systems. For communications satellites, onboard processing provides increased efficiency and performance (e.g., signal regeneration). For space-based sensors, onboard processing can be used for data reduction (e.g., data processing and data compression). Data reduction is important because it eliminates the need for expensive high rate RF downlinks. Although optical frequencies can be used to substantially increase the downlink data rate, they are not well suited for space-to-ground communications because they require multiple downlink sites due to atmospheric effects (e.g., rain, cloud coverage, scintillation, and absorption/scattering). Multiple RF downlink sites are also possible but they are expensive (e.g., requires multiple transmitters). Without requiring expensive and expansive RF downlinks, onboard processing can significantly lower the overall system cost and provide increased performance (e.g., raise the resolution and coverage rate of space sensors). However, onboard processing technology for space systems is not widespread because it tends to lag behind current processing technology that is available for terrestrial systems. COTS processors are nearly 10 years more advanced than their radiation-hardened counterparts.

The concept of decoupled, distributed, and shared space-borne processing goes beyond onboard processing on individual satellites. The availability of a high speed optical space-based information network backbone provides connectivity for processing satellites. Optical intersatellite links are the enabling technology that provides enough data rate in the space-based information network backbone for the exchange of information between mission satellites and processing satellites. The ability to interconnect multiple platforms to utilize shared processing resources allows for much more efficient utilization of the processing systems with less redundancy and reduces the need for separate processing systems on individual mission satellites. These shared processing satellites can use the most advanced processors available at the time of launch. However, the cost for improved computational capabilities is the shorter lifetime due to radiation effects of the space environment. Although radiation tolerant, COTS processors cannot last the 10-15 year on-orbit life of a satellite and will need to be periodically replaced and replenished (as often as every year). Note that such a replacement and replenishment strategy is only reasonable when the space-borne processing is a shared resource. If not, replacement and replenishment of processing systems on individual mission satellites will be difficult to manage and expensive which negates any benefit of implementing more modern processors.

Regardless of whether networked space-based information processing resources are available in the future, the concept of a space-based information backbone is a key development for transforming the stove-piped satellite communications community to a satellite networking community serving multiple users. A horizontally organized space-based network backbone infrastructure can advance spacecraft interoperability and the levels of inter-spacecraft communications. The sharing of assets can reduce costs as duplication efforts by multiple organizations can be eliminated. The implementation of optical intersatellite links can go beyond the provisioning of large amounts of capacity and creating a network. Novel use of the space network can revolutionize satellite communications and space missions. Applications include distributed computing in space, interoperable space communications, multiplatform distributed satellite communications, coherent distributed space sensing, multisensor data fusion, and restoration of disconnected global terrestrial networks.

Space R&D is a costly, long-term investment. Several constellation topologies for the space-based information network backbone have been analyzed to provide high data rates, high connectivity, and low latency at a reasonable cost. Because of the high costs of space systems development and the astronomical cost of launch systems, most space activities have been initiated by the government. Opening the space domain to the commercial sector will require change via the development a new space paradigm. In addition to the need for a solution to the problem of high launch costs, space policies should emphasize increasing the role of the private sector in space and building a common space-based network infrastructure. The private sector will enter the space realm if it could find profit motives for space technology development and operation. Private sector investment in space allows the government to invest in other areas. However, the government will have to continue to invest in developing space systems for defense purpose where the profit potential is low or none (e.g., space science and national security programs). This is the classic public goods argument for government support of infrastructures.

Although military space systems are not necessarily designed for dual use, the government should invest in the development of a space-based information network infrastructure that is generic in nature. Other than the fact that it is very hard to predict the characteristics of future defense users of the network, it is desirable to develop a ‘generic’ as opposed to a ‘specialty’ network because it allows for expansion to accommodate civilian demand, lowers the price of entry for the commercial sector, and makes way for innovation to enhance and provide additional value to the system. Therefore, there exists the opportunity to shift the space paradigm toward identifying areas of technology convergence in order to build a common R&D infrastructure and increasing the role of the private sector in space development.

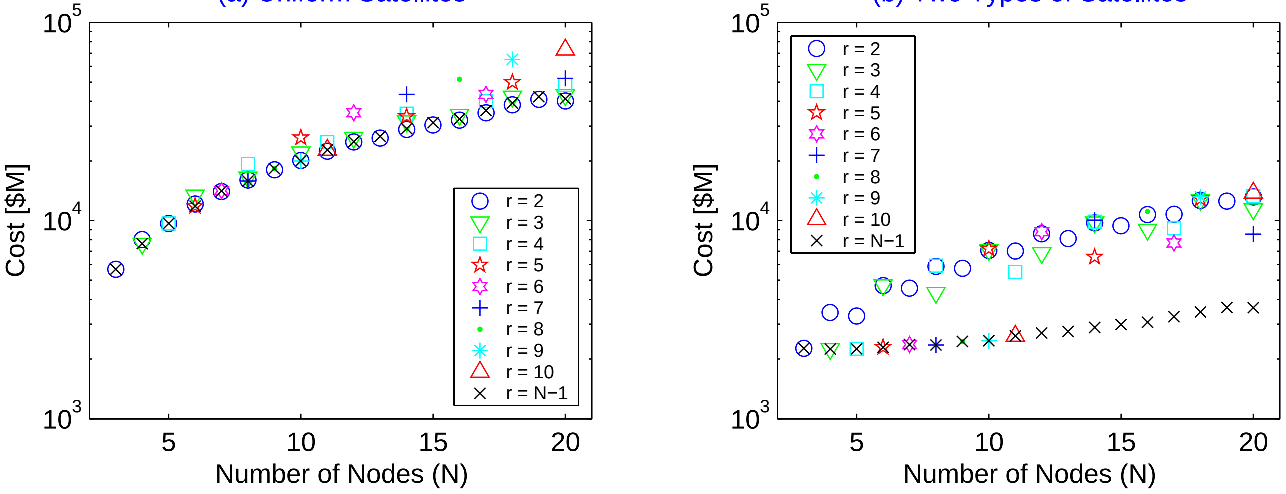
Appendix A

Network Architecture Cost Model Variations

This appendix provides supplementary results for the communications costs of the space-based information network backbone discussed in Chapter 3. In the first section, results for the HMH cost case are provided. These results are comparable to the results seen for the MLM cost case shown in Chapter 3. In the second section, cost results for small constellations ($N = 4$, 5, and 6) are shown for all cost permutations. It can be observed that the previously observed general trends remain the same across all cost permutations. Recall that Table 3.7 shows the 27 permutations of possible cost scenarios available for assessing the communications costs of the space-based information network backbone. The values for each cost level can be found in Table 3.6. In the third section, cost results for using linear switches on small constellations ($4 \leq N \leq 6$) are shown.

A.1 HMH Cost Model Results

A.1.1 Connected Circulant Constellations with Uniform Jump Spaces under Uniform All-to-All Traffic


(a) Uniform Satellites

(b) Two Types of Satellites

(c) Custom Satellites

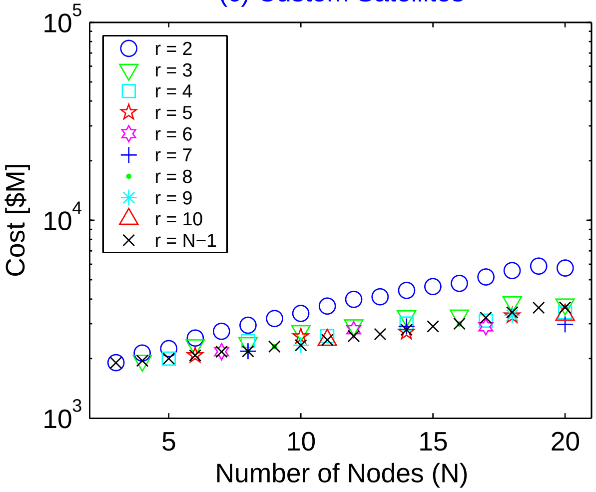


Figure A-1: Connected circulant constellations with uniform jump spaces: communications cost HMH for uniform all-to-one traffic with Dijkstra's routing.

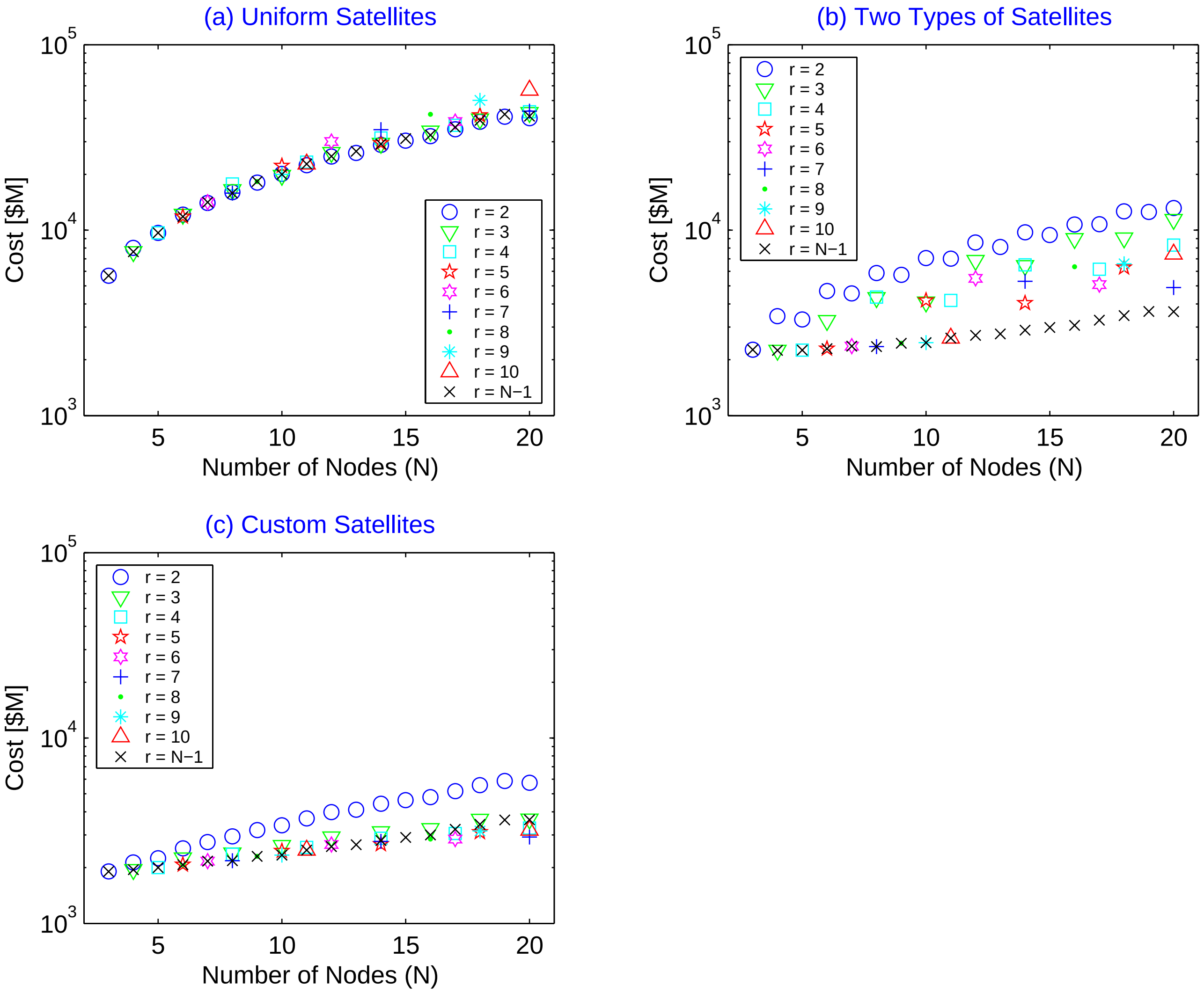
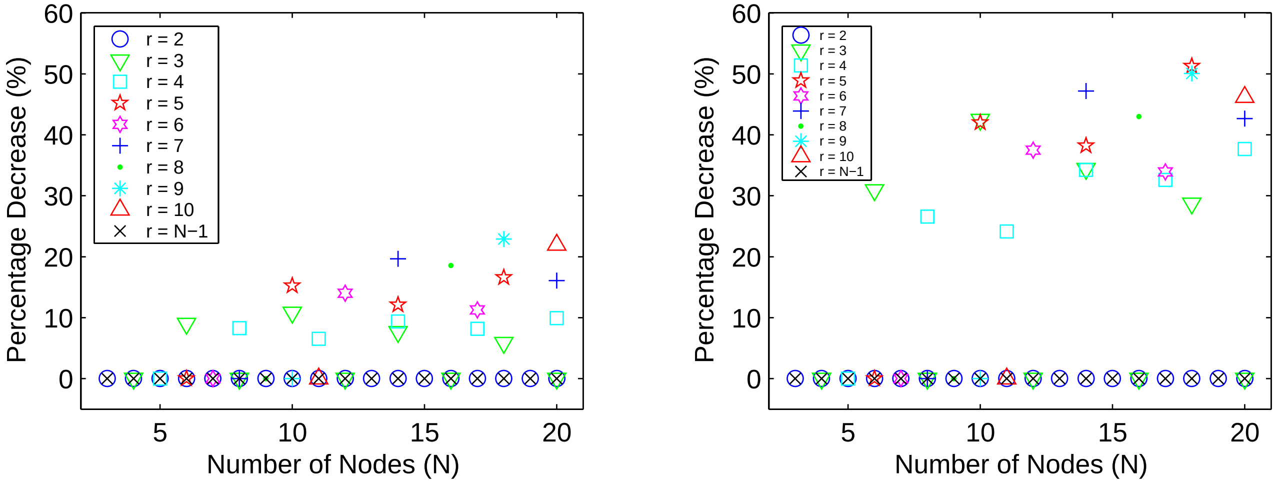


Figure A-2: Connected circulant constellations with uniform jump spaces: communications cost HMH for uniform all-to-one traffic with Symmetric Dijkstra's routing.

(a) Uniform Satellites

(b) Two Types of Satellites



(c) Custom Satellites

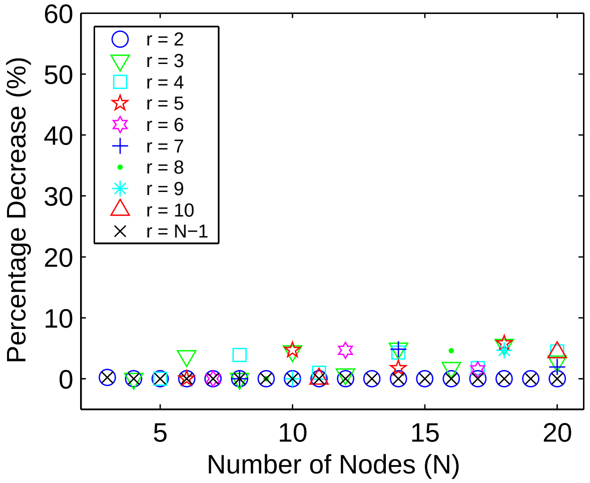


Figure A-3: Connected circulant constellations with uniform jump spaces: communications cost percentage decrease HMH between Dijkstra's routing and Symmetric Dijkstra's routing for uniform all-to-one traffic.

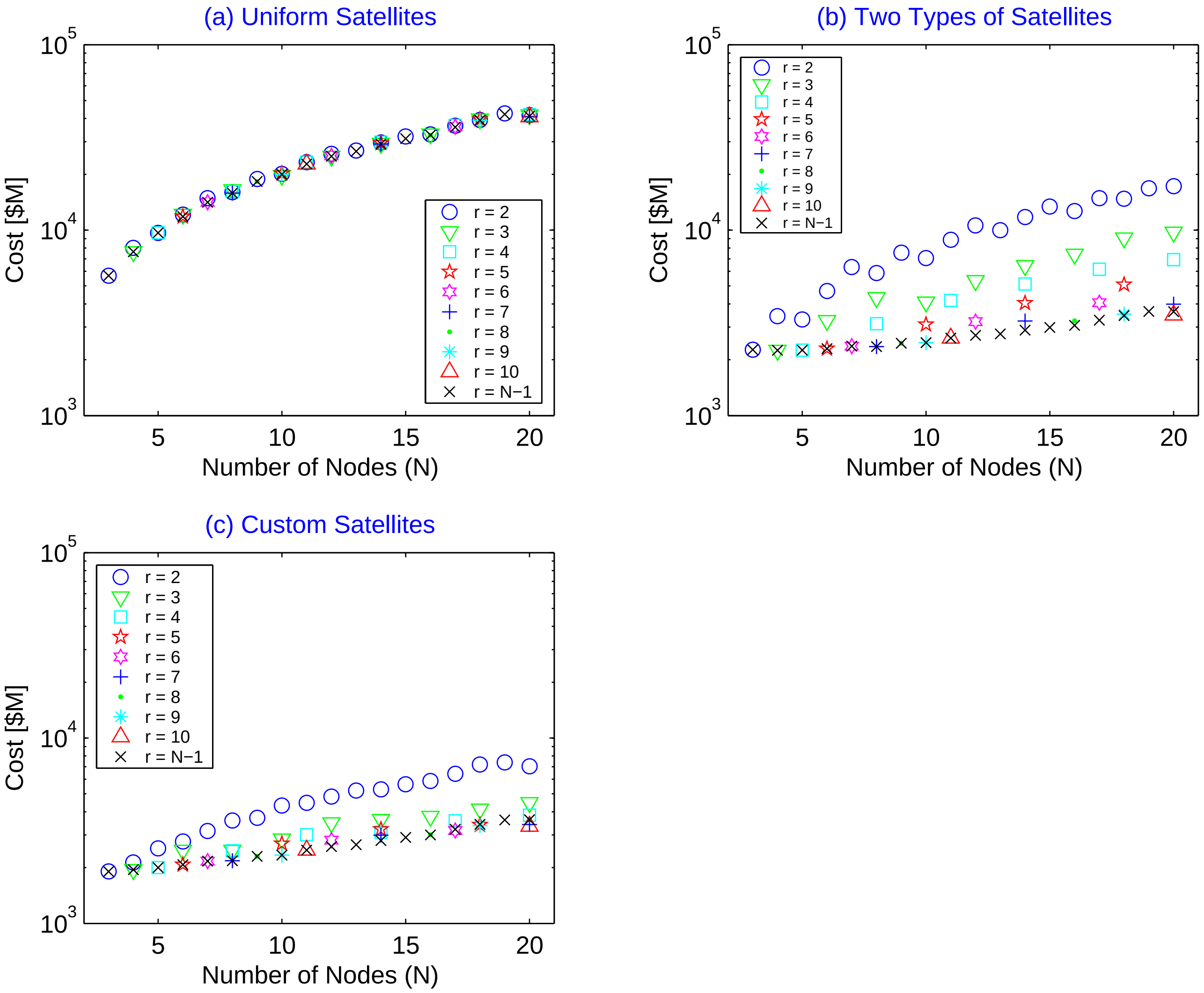
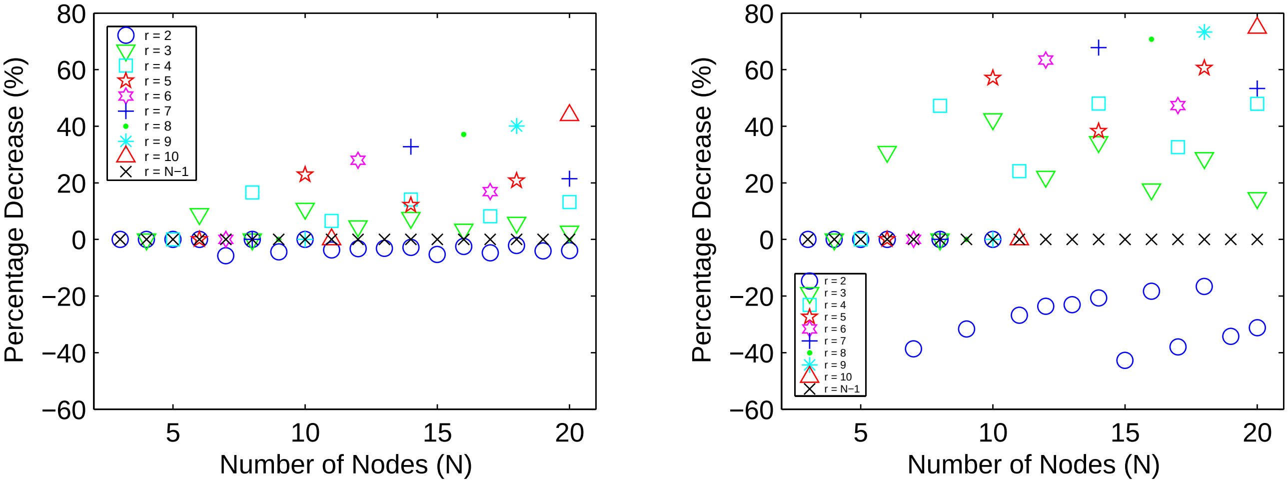


Figure A-4: Connected circulant constellations with uniform jump spaces: communications cost HMH for uniform all-to-one traffic with Incremental Dijkstra's routing.

(a) Uniform Satellites (b) Two Types of Satellites



(c) Custom Satellites

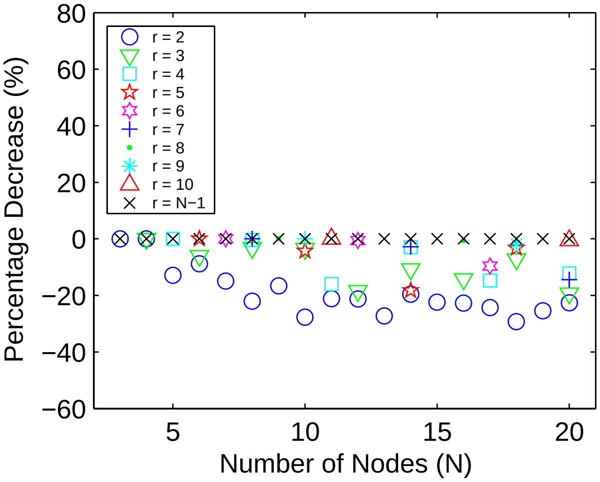
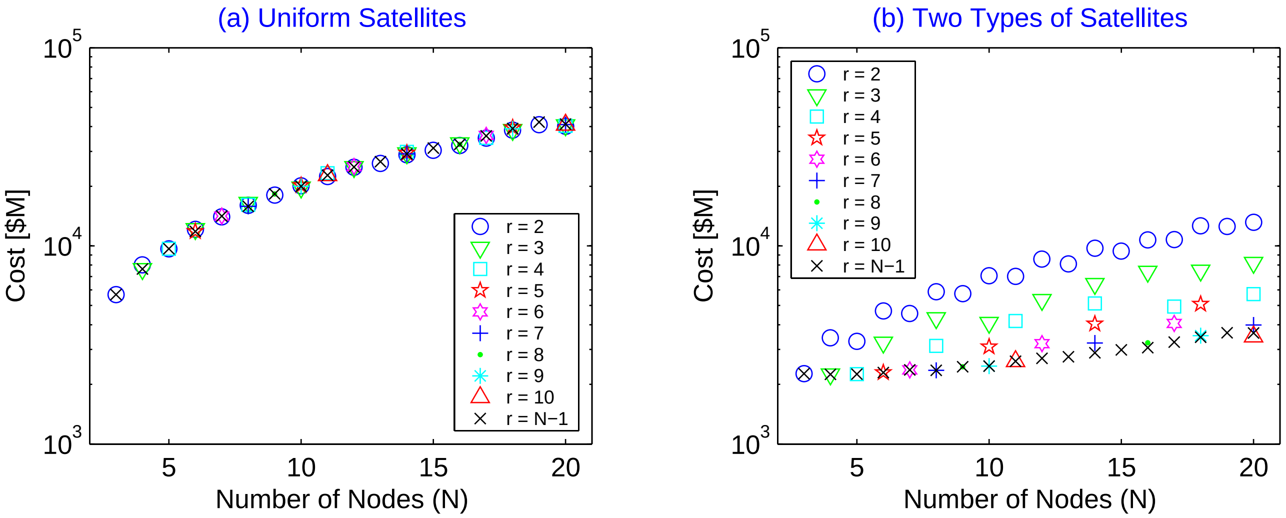


Figure A-5: Connected circulant constellations with uniform jump spaces: communications cost percentage decrease HMH between Dijkstra's routing and Incremental Dijkstra's routing for uniform all-to-one traffic.


(a) Uniform Satellites

(b) Two Types of Satellites

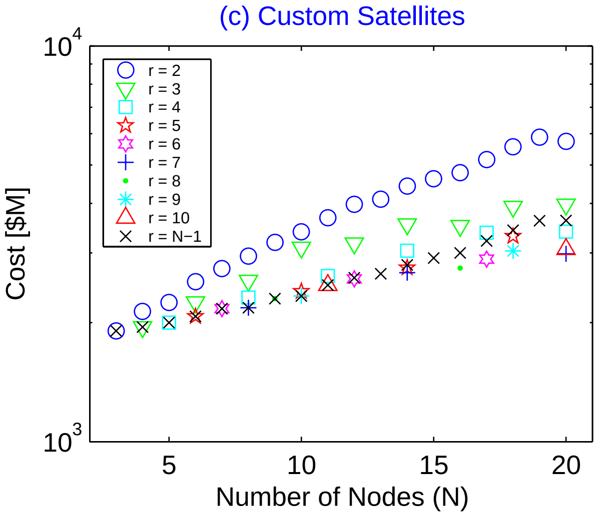
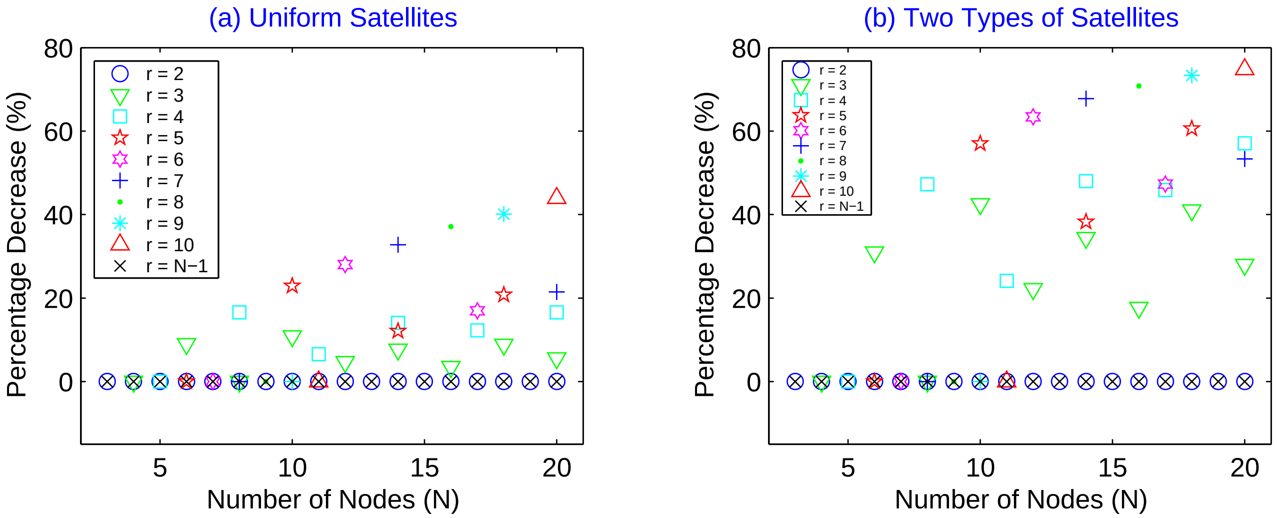

(c) Custom Satellites

Figure A-6: Connected circulant constellations with uniform jump spaces: communications cost HMH for uniform all-to-one traffic with Modified Incremental Dijkstra's routing.



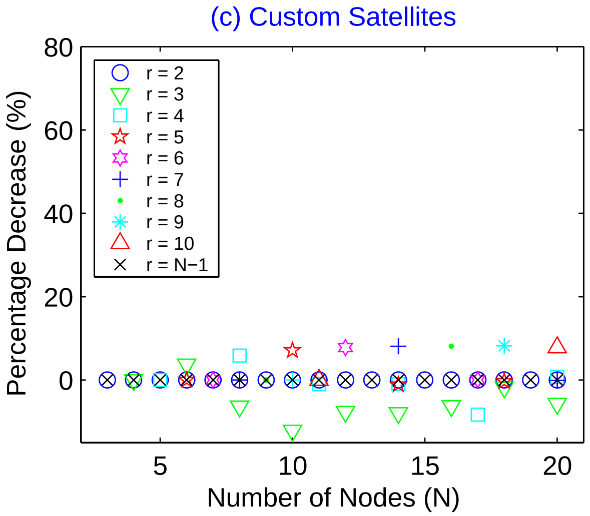
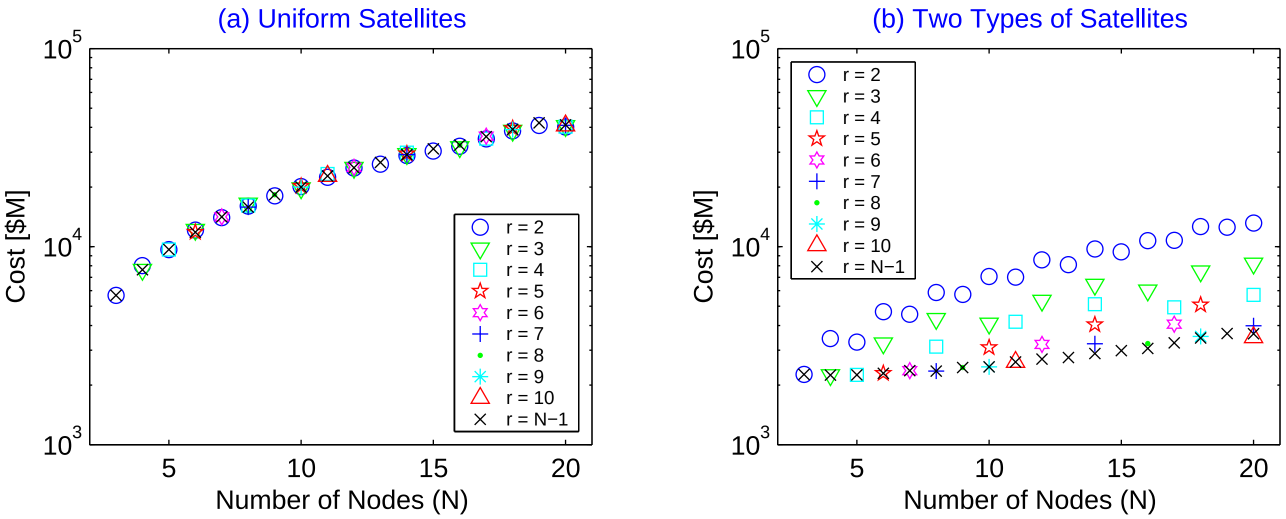


Figure A-7: Connected circulant constellations with uniform jump spaces: communications cost percentage decrease HMH between Dijkstra's routing and Modified Incremental Dijkstra's routing for uniform all-to-one traffic.



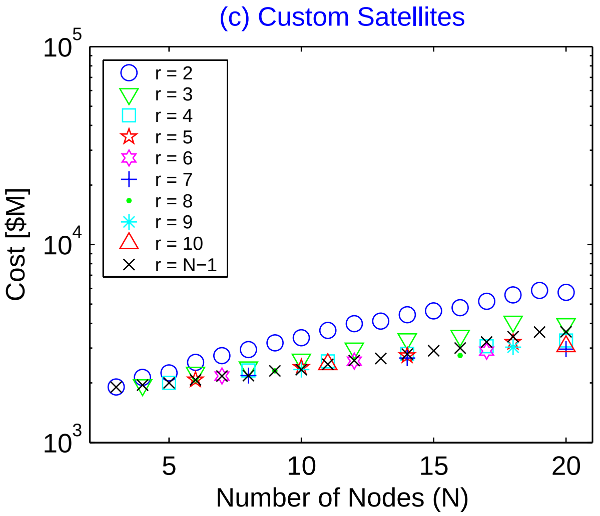
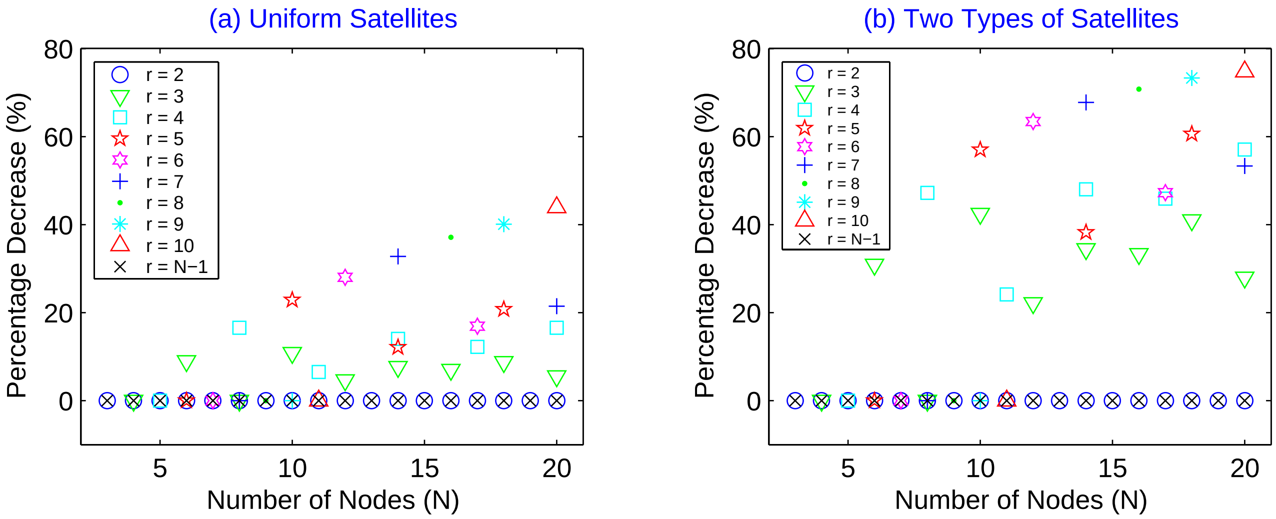


Figure A-8: Connected circulant constellations with uniform jump spaces: communications cost HMH for uniform all-to-one traffic with Symmetric Modified Incremental Dijkstra's routing.



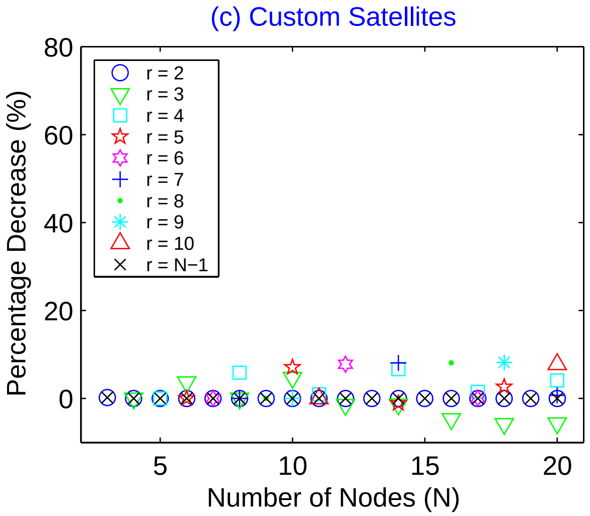
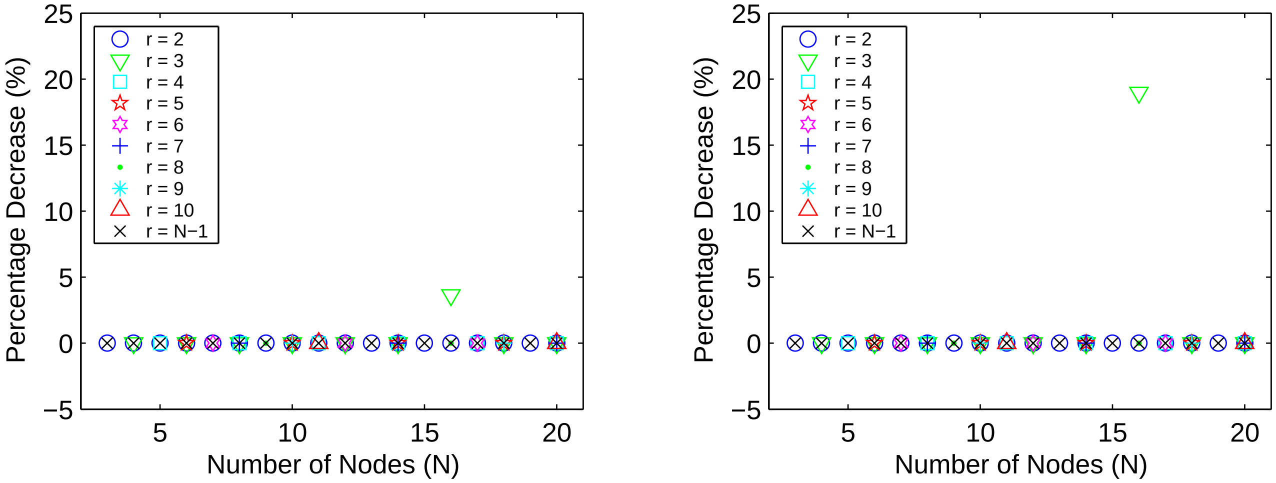


Figure A-9: Connected circulant constellations with uniform jump spaces: communications cost percentage decrease HMH between Dijkstra's routing and Symmetric Modified Incremental Dijkstra's routing for uniform all-to-one traffic.



(a) Uniform Satellites

(b) Two Types of Satellites

(c) Custom Satellites

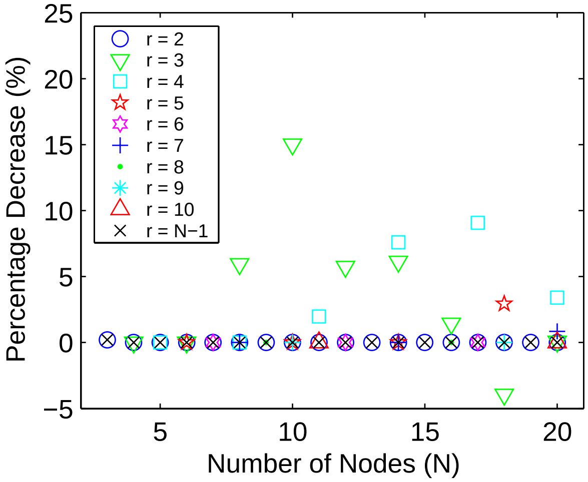
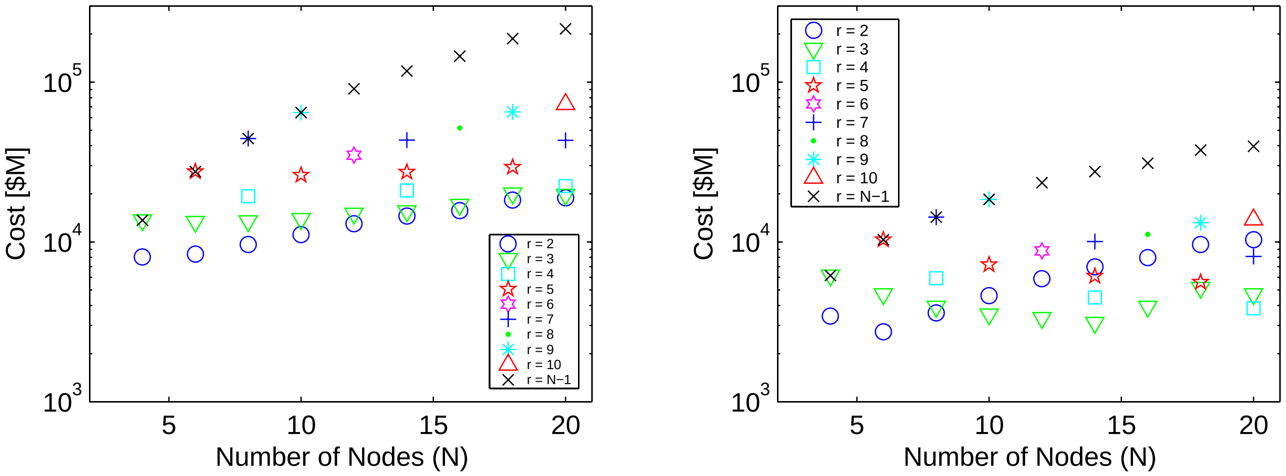
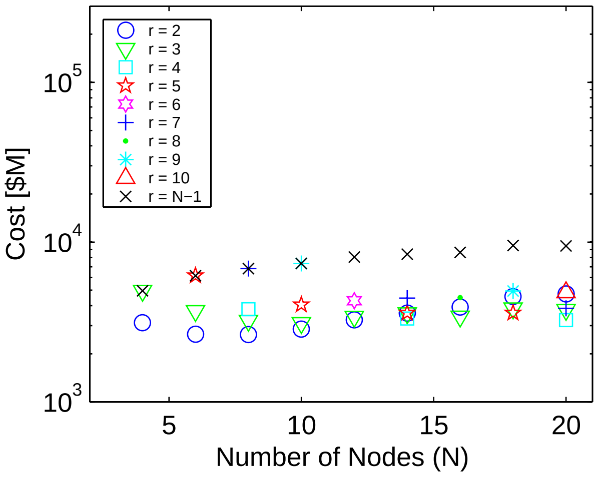


Figure A-10: Connected circulant constellations with uniform jump spaces: communications cost percentage decrease HMH between Modified Incremental Dijkstra's routing and Symmetric Modified Incremental Dijkstra's routing for uniform all-to-one traffic.



(a) Uniform Satellites

(b) Two Types of Satellites

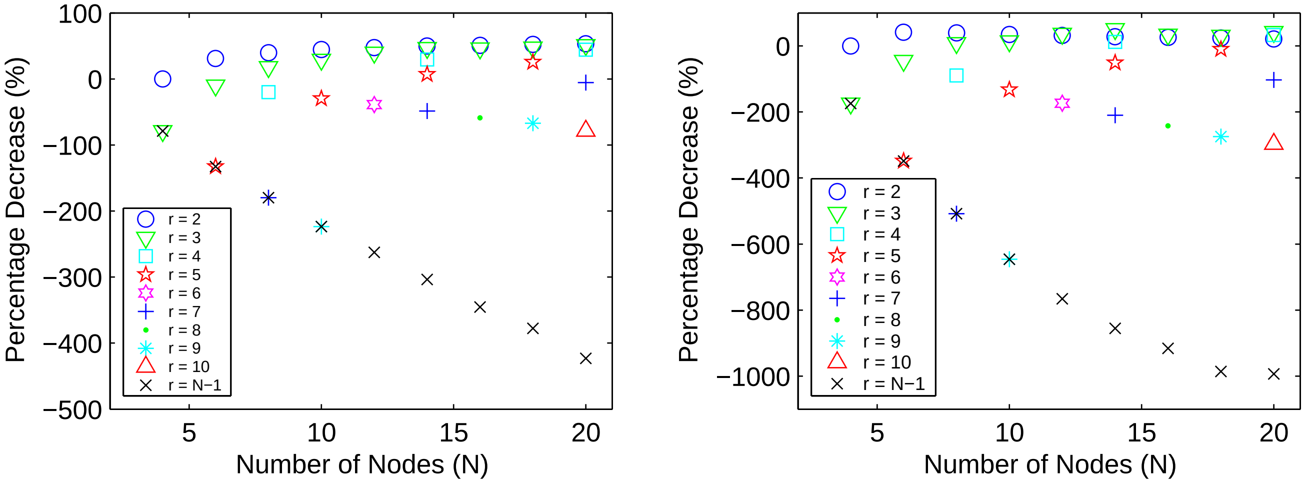


(c) Custom Satellites

Figure A-11: Connected circulant constellations with uniform jump spaces: communications cost HMH with Symmetric Modified Incremental Dijkstra's routing for uniform all-to-one traffic with 2 hubs.

(a) Uniform Satellites

(b) Two Types of Satellites



(c) Custom Satellites

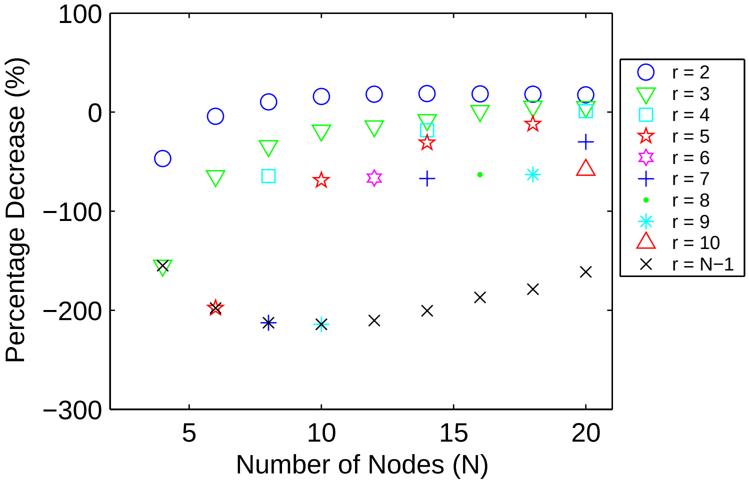


Figure A-12: Connected circulant constellations with uniform jump spaces: communications cost percentage decrease HMH with Symmetric Modified Incremental Dijkstra's routing for uniform all-to-one traffic between 1-hub and 2-hub constellations.

A.1.2 Mixed Traffic

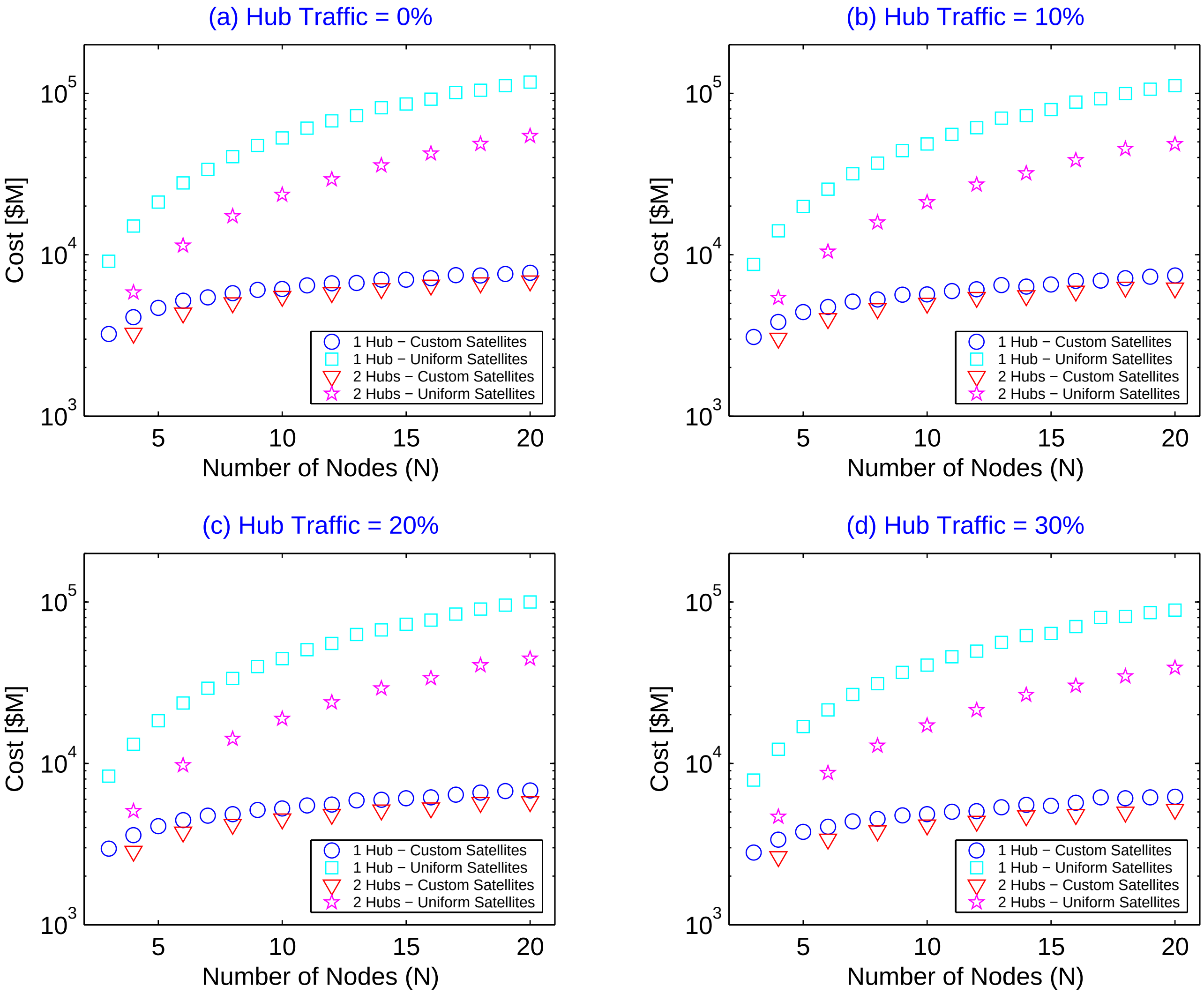


Figure A-13: Hub constellations: communications cost HMH for mixed traffic I.

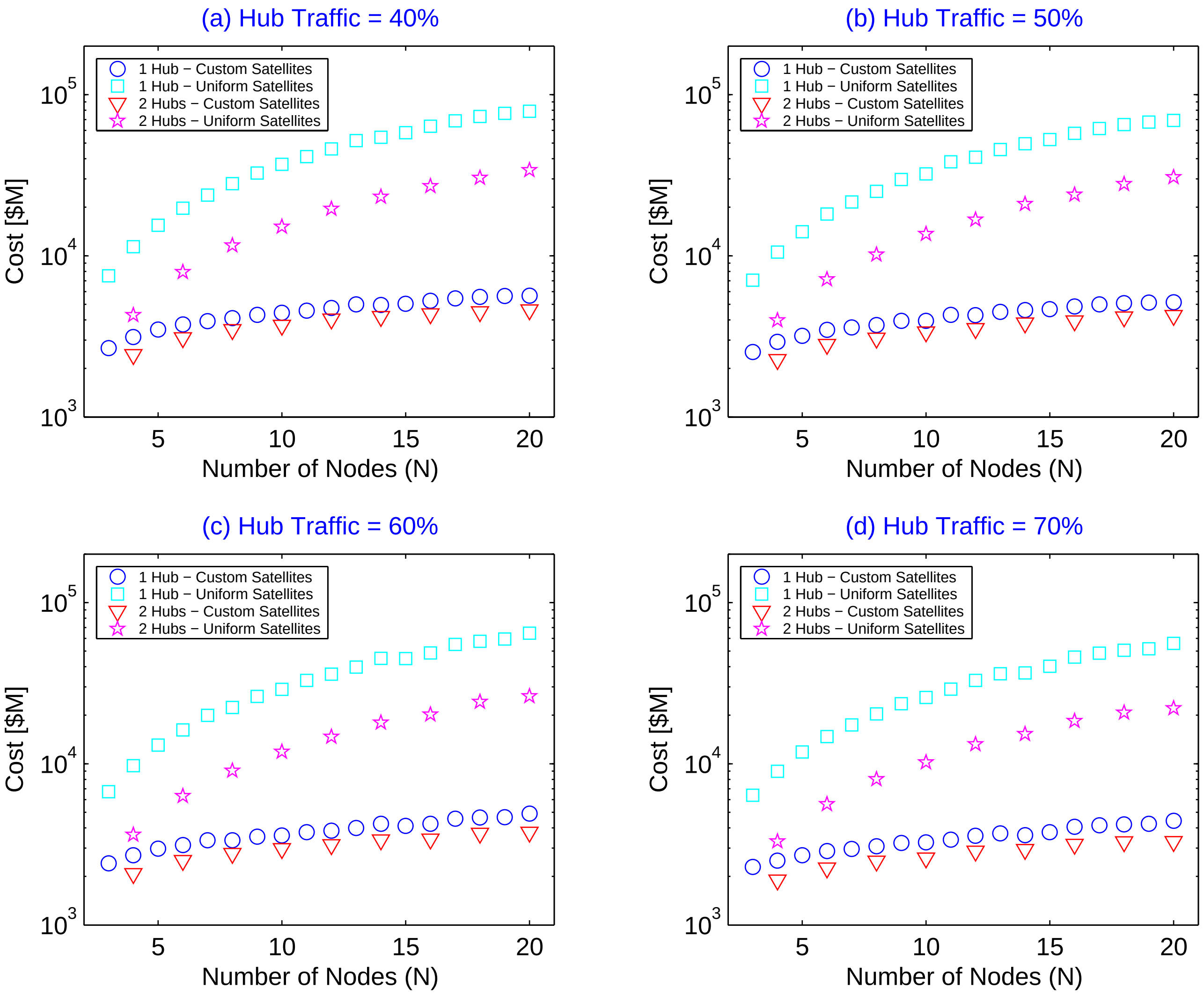


Figure A-14: Hub constellations: communications cost HMH for mixed traffic II.

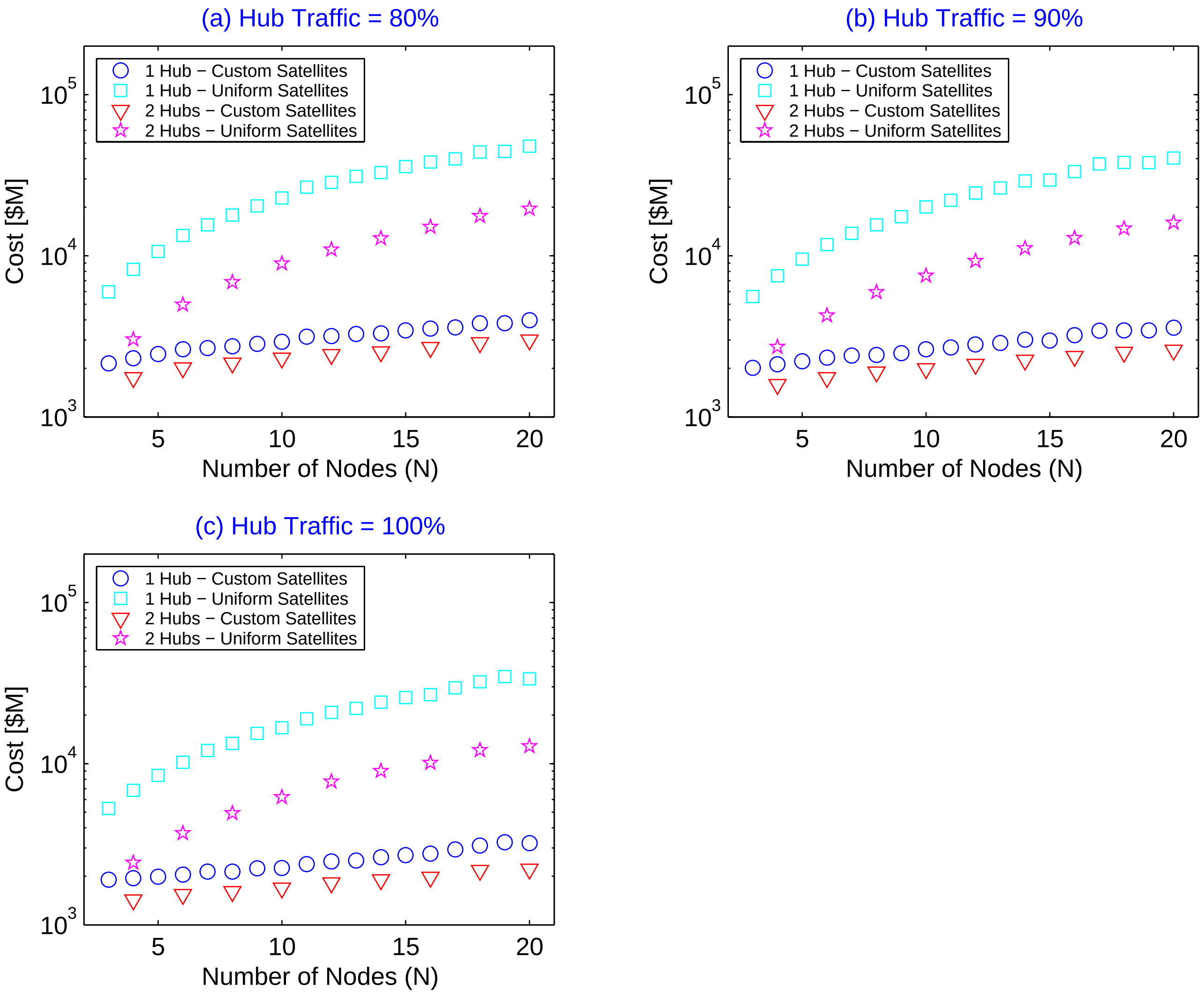


Figure A-15: Hub constellations: communications cost HMH for mixed traffic III.

A.1.2.1 Connected Circulant Constellations with Uniform Jump Spaces

Figures A-16, A-17, and A-18 show the results for the HMH case using uniform satellites. Figures A-19, A-20, and A-21 show the results for the HMH case using 2 types of satellites. Figures A-22, A-23, and A-24 show the results for the HMH case using custom satellites.

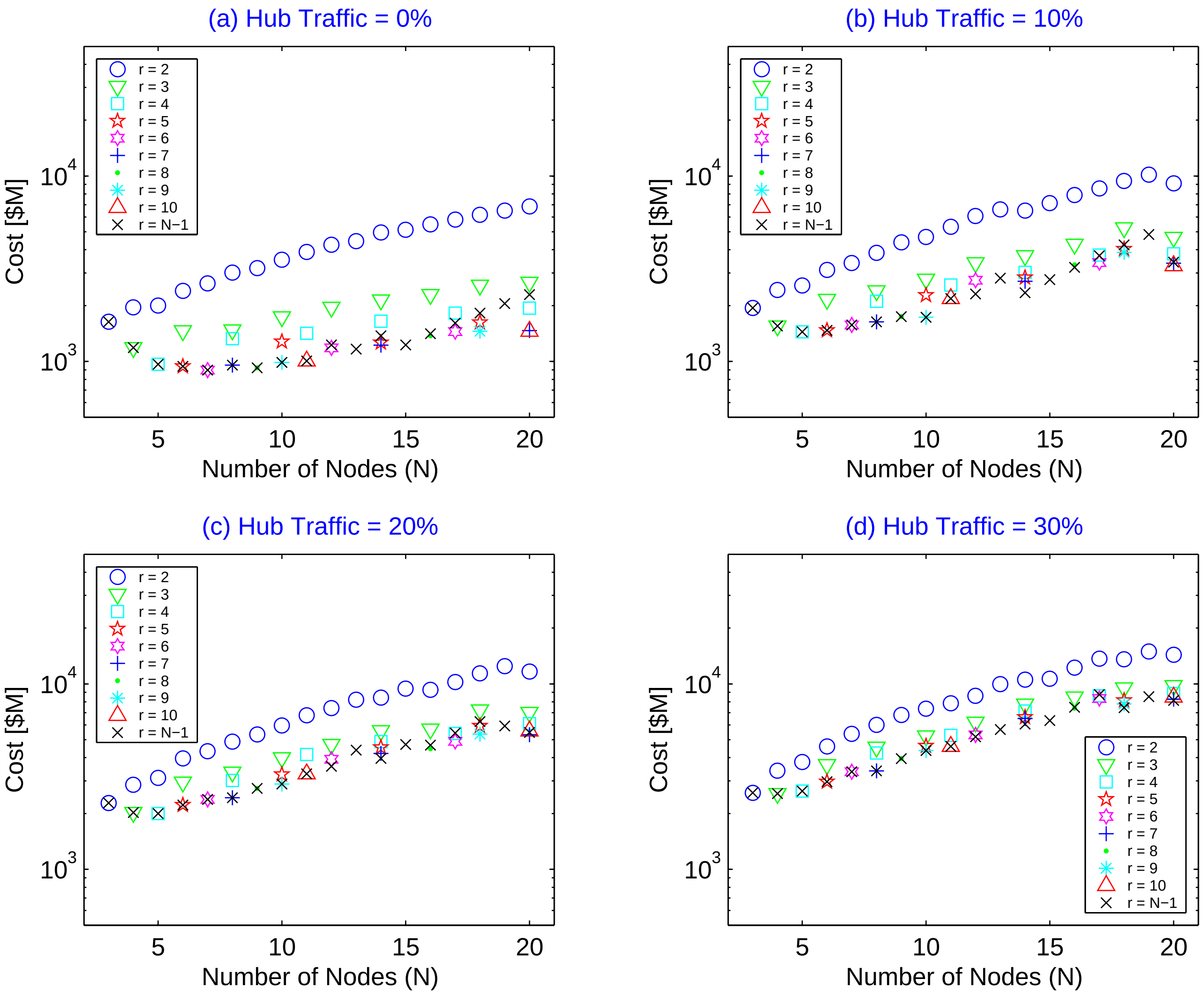


Figure A-16: Connected circulant constellations with uniform jump spaces: communications cost HMH with uniform satellites for mixed traffic I.

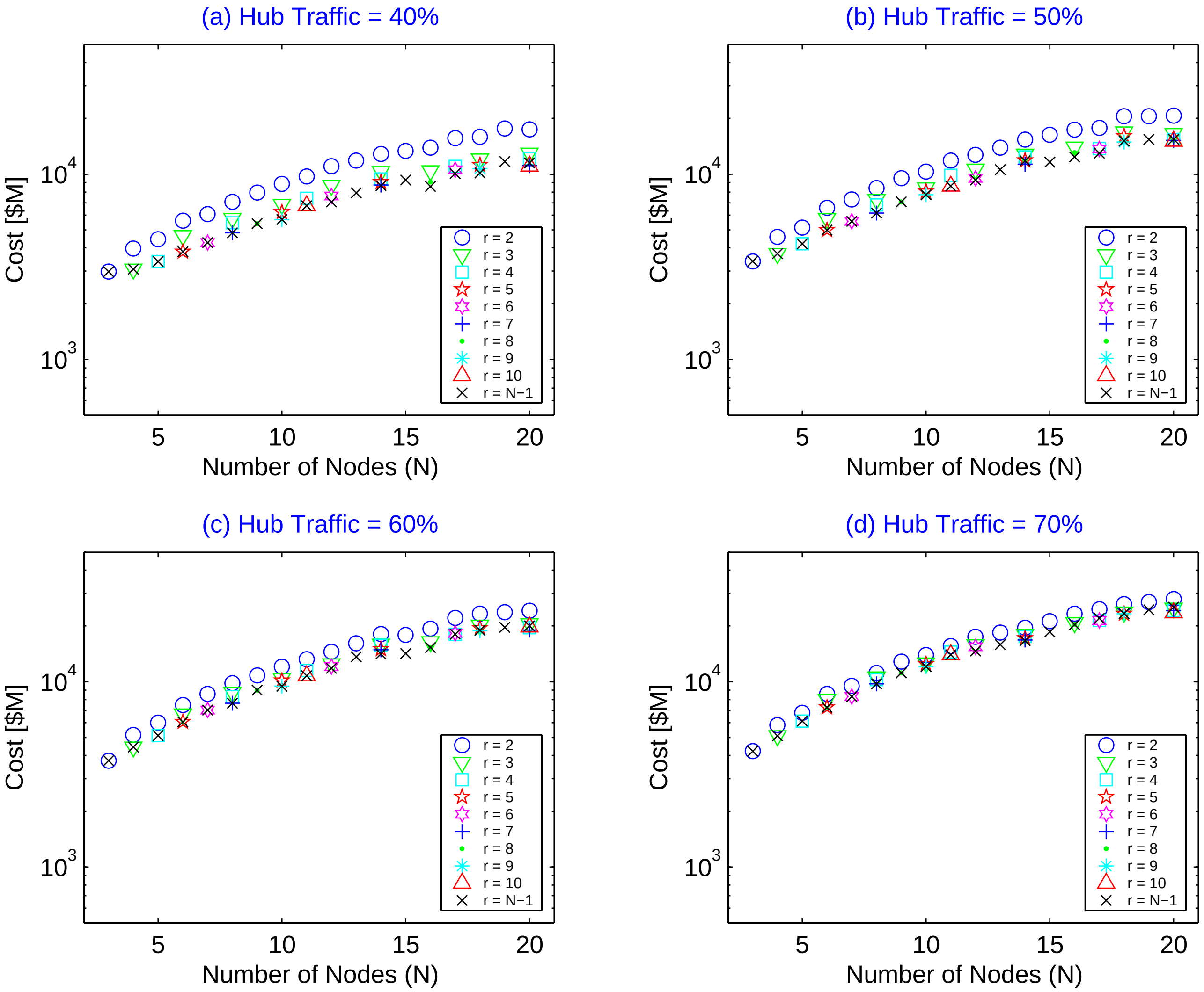
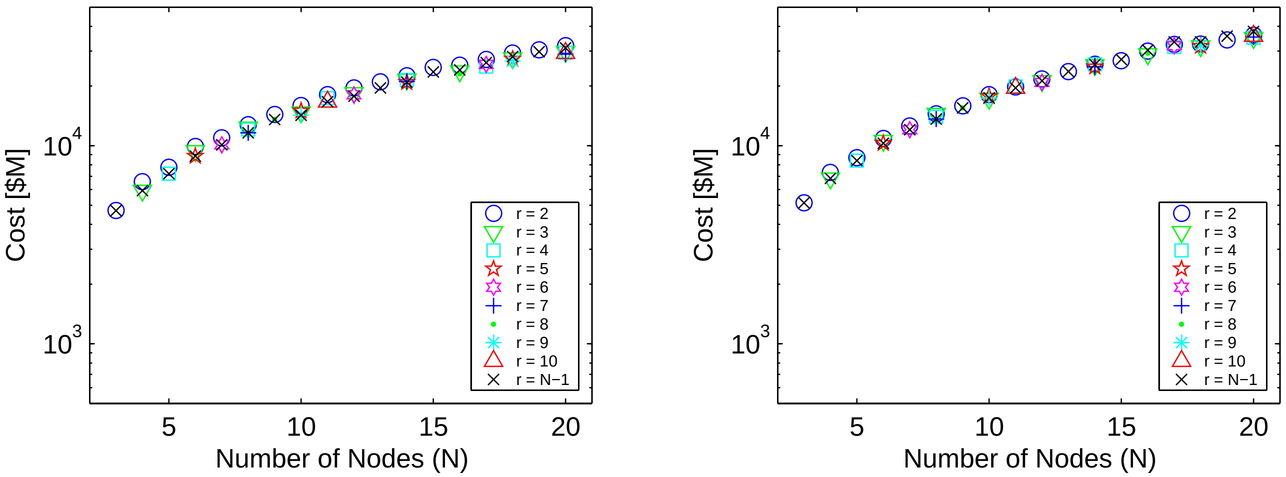


Figure A-17: Connected circulant constellations with uniform jump spaces: communications cost HMH with uniform satellites for mixed traffic II.



(a) Hub Traffic = 80%

(b) Hub Traffic = 90%

(c) Hub Traffic = 100%

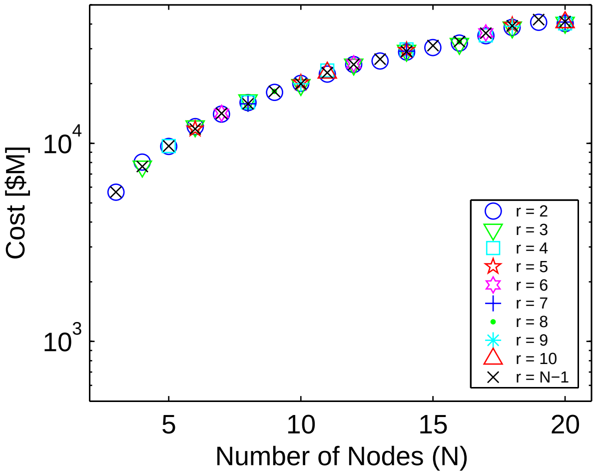


Figure A-18: Connected circulant constellations with uniform jump spaces: communications cost HMH with uniform satellites for mixed traffic III.

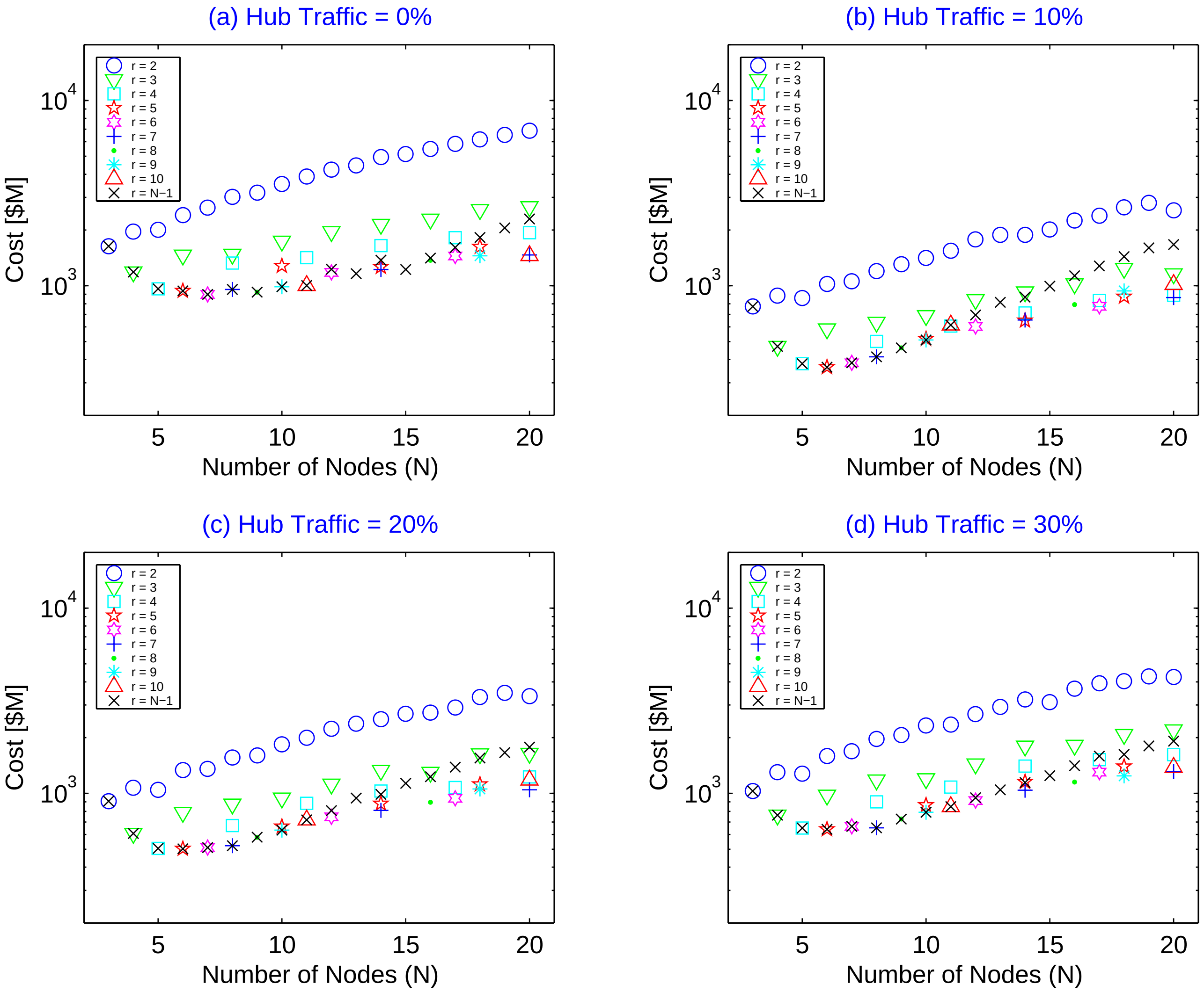


Figure A-19: Connected circulant constellations with uniform jump spaces: communications cost HMH with 2 types of satellites for mixed traffic I.

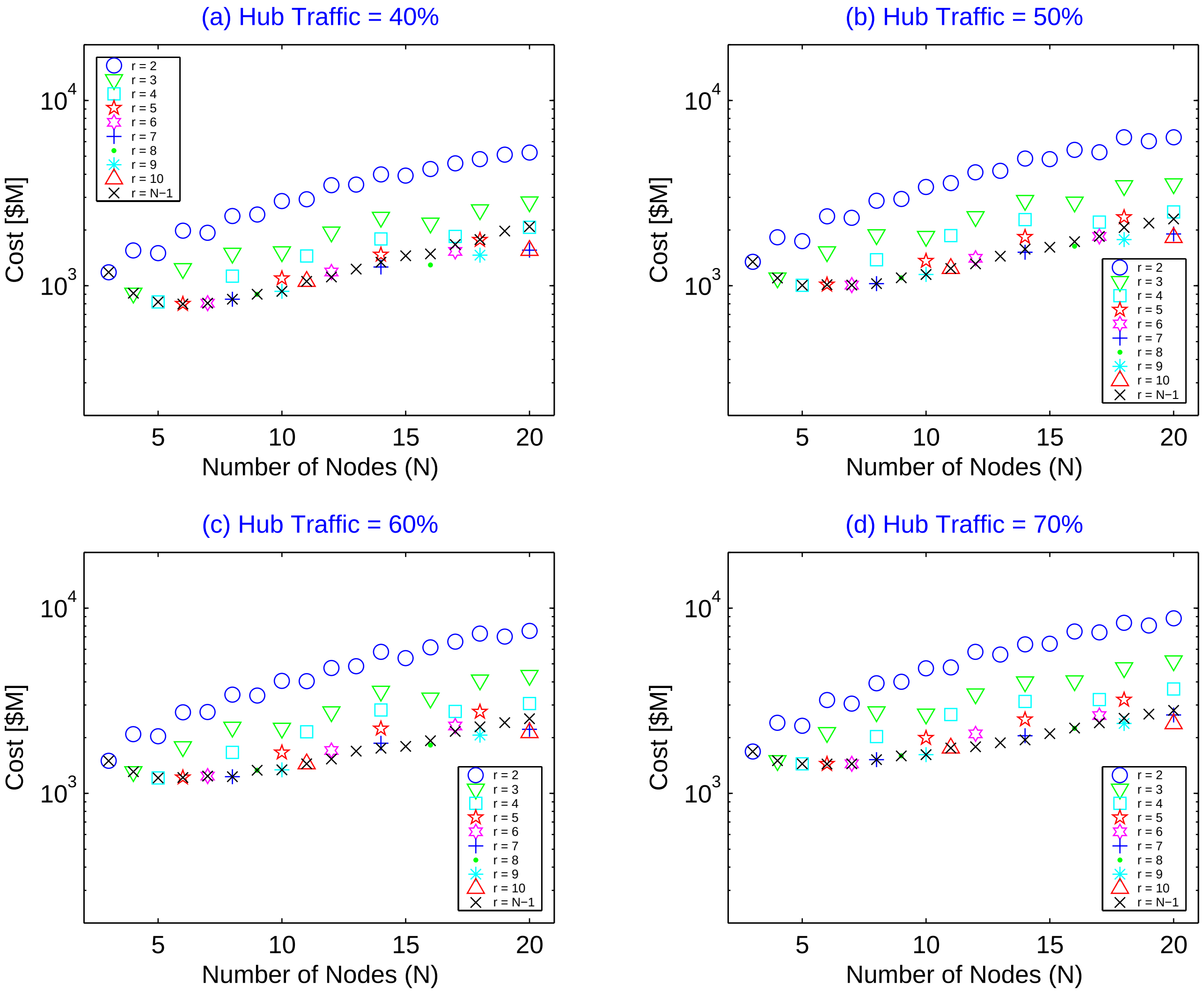
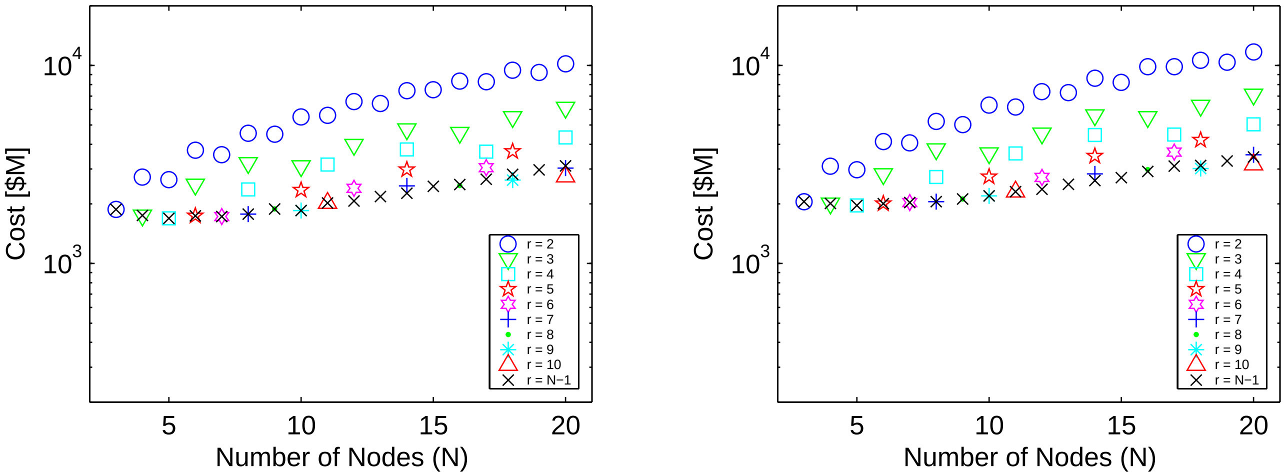
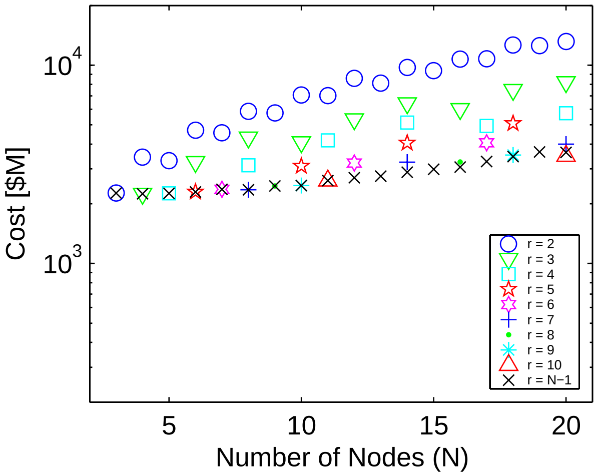


Figure A-20: Connected circulant constellations with uniform jump spaces: communications cost HMH with 2 types of satellites for mixed traffic II.



(a) Hub Traffic = 80%

(b) Hub Traffic = 90%



(c) Hub Traffic = 100%

Figure A-21: Connected circulant constellations with uniform jump spaces: communications cost HMH with 2 types of satellites for mixed traffic III.

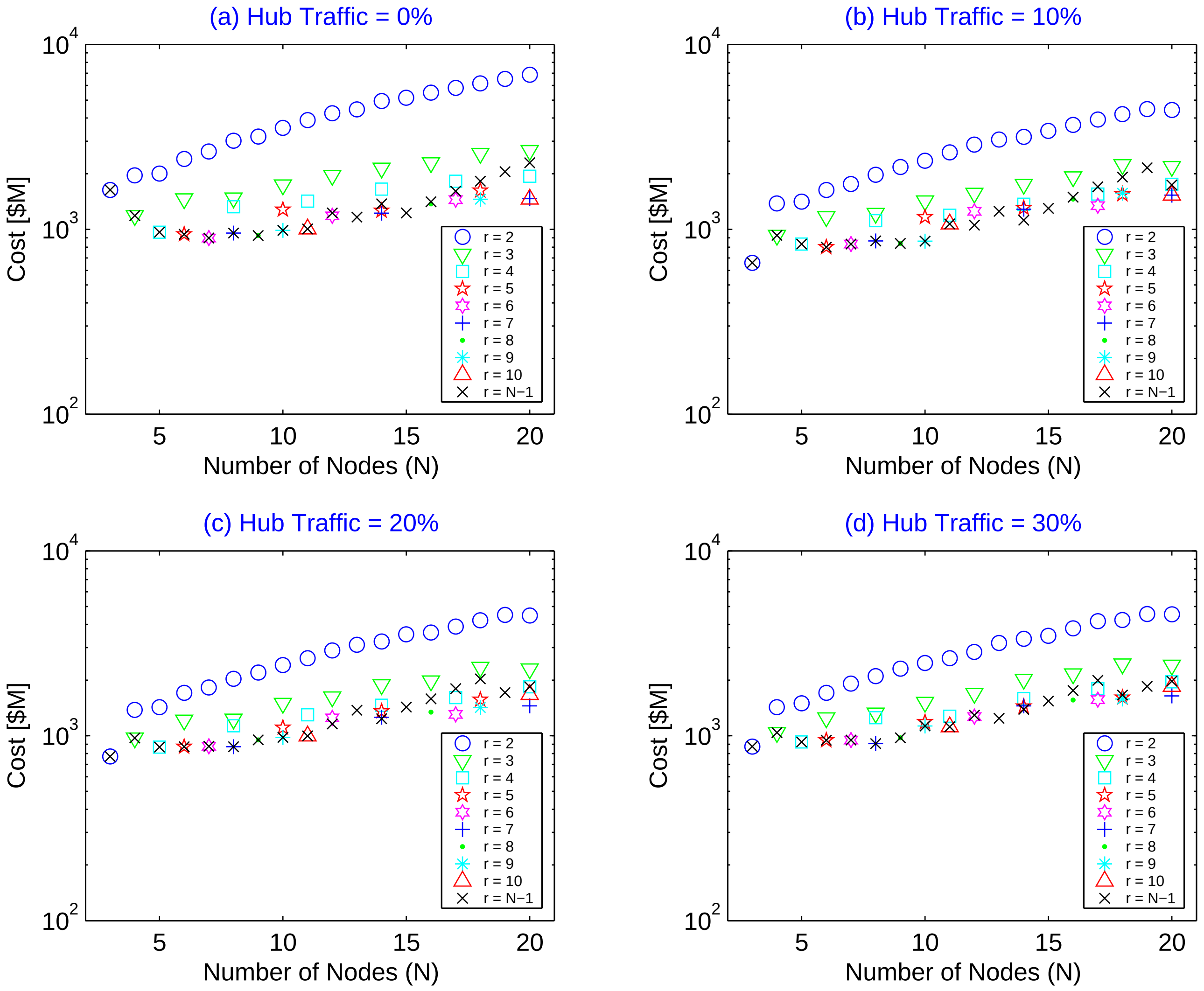


Figure A-22: Connected circulant constellations with uniform jump spaces: communications cost HMH with custom satellites for mixed traffic I.

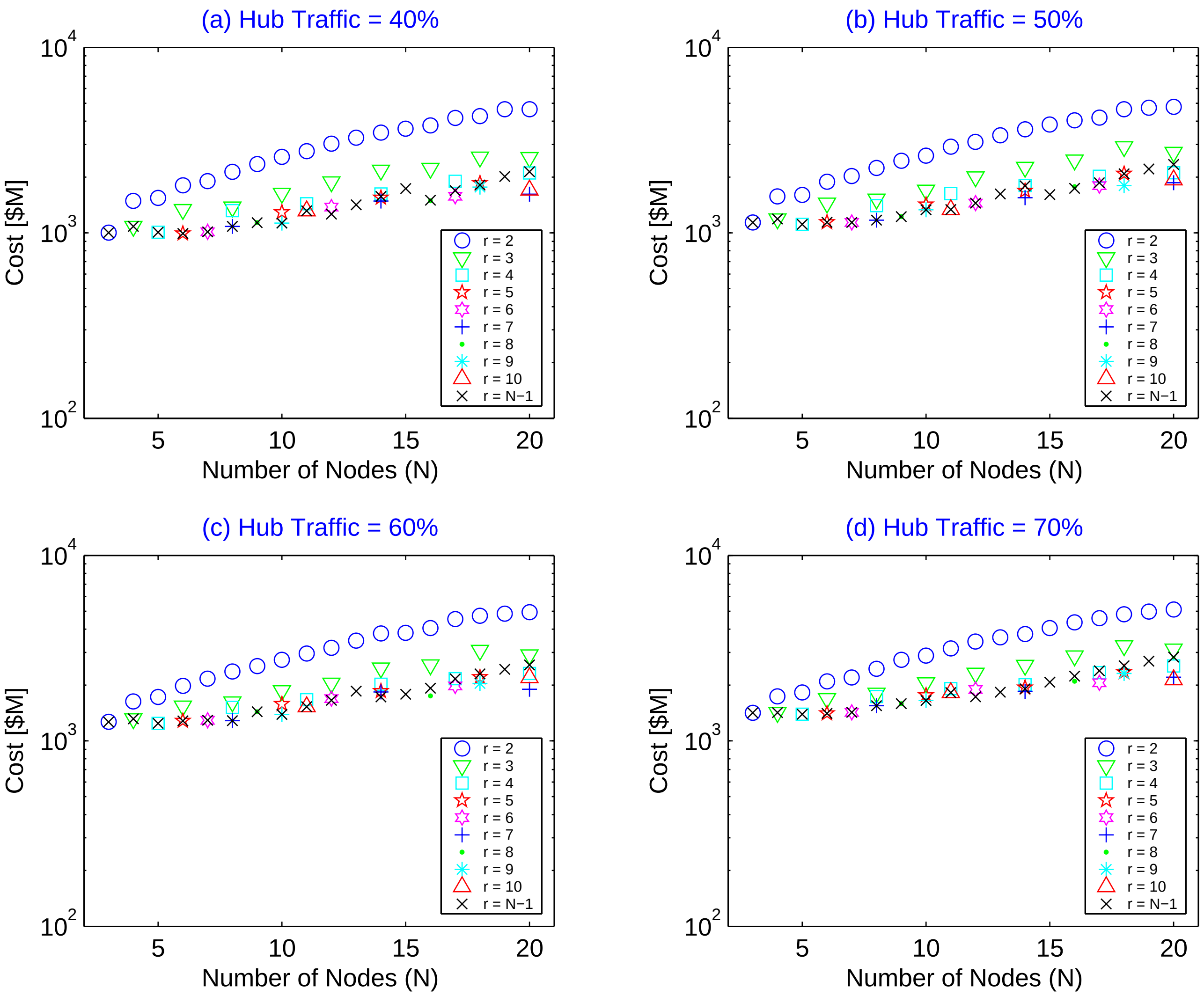
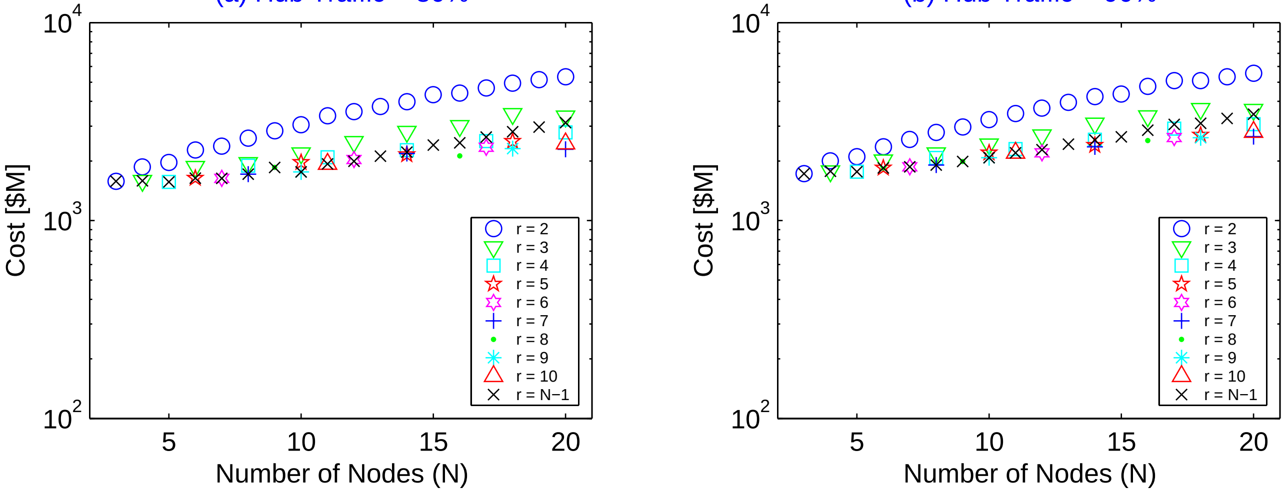


Figure A-23: Connected circulant constellations with uniform jump spaces: communications cost HMH with custom satellites for mixed traffic II.


(a) Hub Traffic = 80%

(b) Hub Traffic = 90%

(c) Hub Traffic = 100%

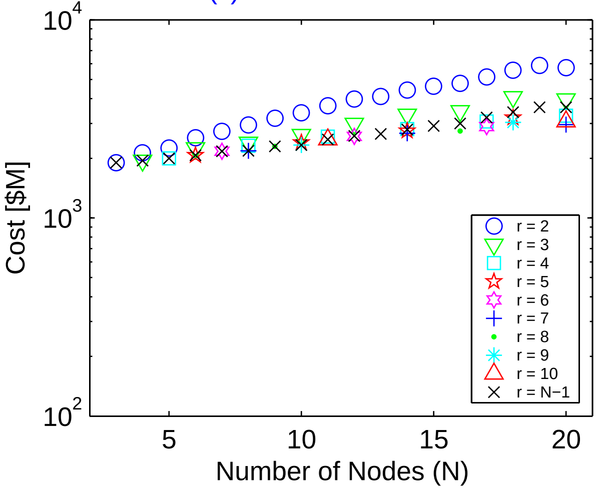


Figure A-24: Connected circulant constellations with uniform jump spaces: communications cost HMH with custom satellites for mixed traffic III.

A.1.3 Two Communities of Users with Mixed Traffic

A.1.3.1 Equal Traffic Volumes

Using Figure A-25, the constellation types of lowest system cost for the various amounts of mixed traffic is re-plotted in Figure A-26 for the HMH cost case. The total communication costs of building two separate systems for the various amounts of mixed traffic are shown in Figure A-27. Again, the contours lines separate the regions of different constellation types. In each regions where two constellation systems are deployed, C_i indicates the constellation that is built for user group i, which is dependent on the amount of mixed traffic within each user group.

The costs of building one satellite constellation systems to handle the sum of the two user traffics (i.e., $T = 800$ wavelengths) are shown in Figure A-28. Again, traffic among the two user communities does not mix within the constellation, i.e., each satellite node has two switches. Likewise, the contour lines in Figure A-28 separate the lowest cost constellation built within each region of varying mixed traffic among the two user groups.

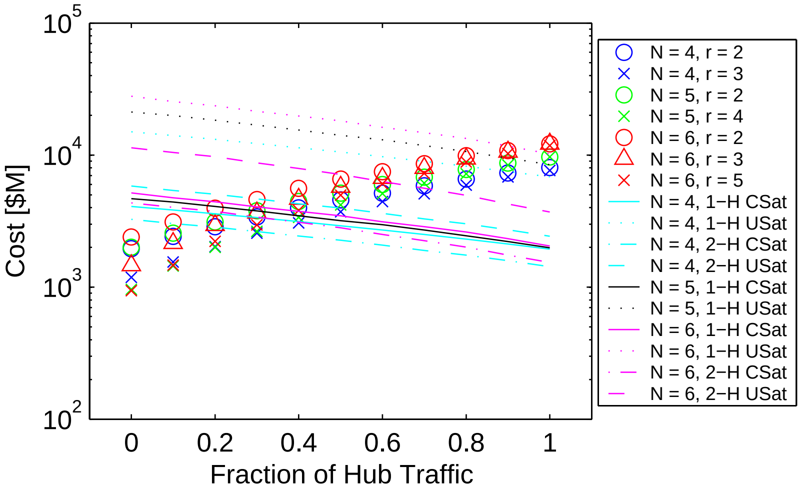


Figure A-25: Comparing communications costs HMH of N=4,5,6 for range of mixed traffic.

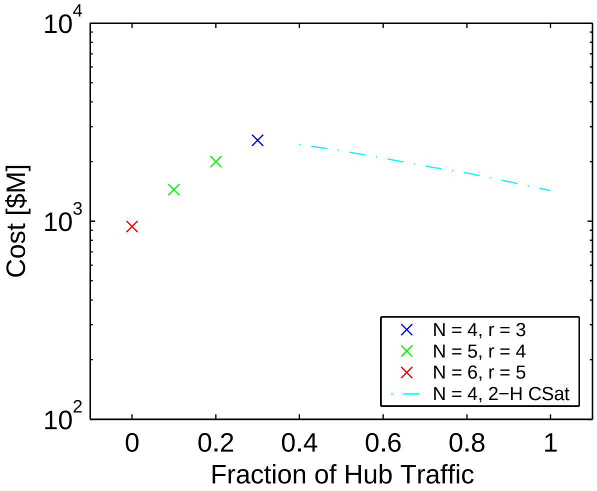


Figure A-26: Lowest communications costs for HMH cost scenario with mixed traffic.

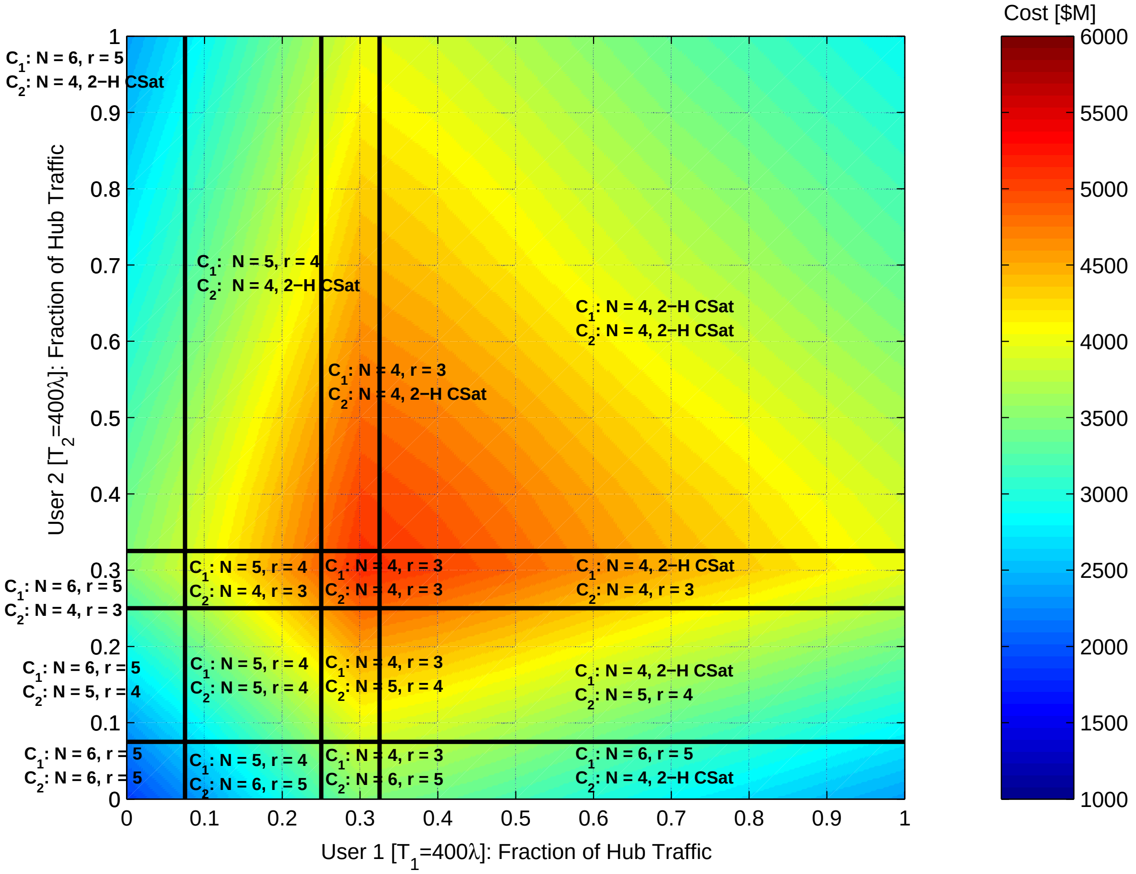


Figure A-27: Lowest communications costs of 2 separate constellation systems for HMH cost scenario with mixed traffic using uniform satellites.

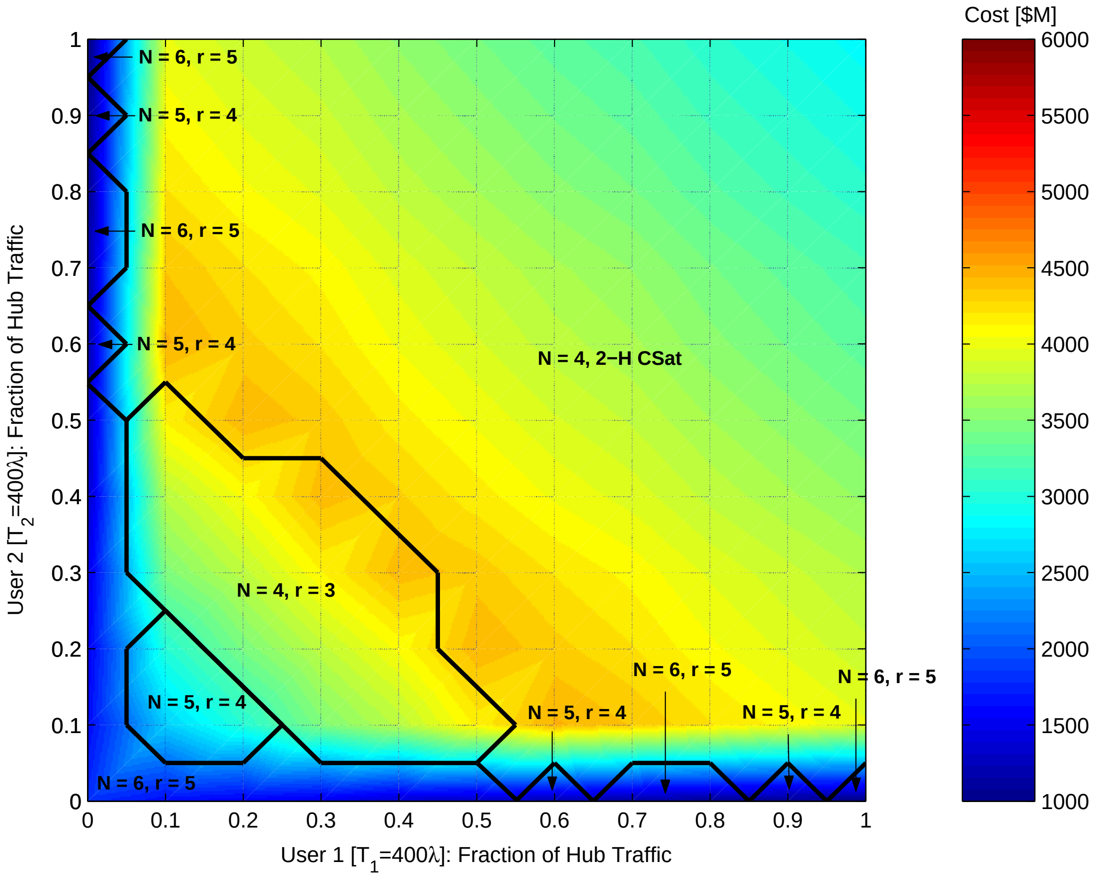


Figure A-28: Lowest communications costs of 1 constellation systems for HMH cost scenario with mixed traffic using uniform satellites.

A comparison of the costs between two separate satellite constellation systems and one satellite constellation systems for two disparate user groups is made. The results for the lowest cost systems are shown in Figure A-29. Generally, a one satellite constellation system can satisfy most cases. Two separate satellite constellation systems are more cost-effective when the two user groups of traffics are very disparate (i.e., one user group has a high amount of uniform all-to-one traffic and a low amount of uniform all-to-all traffic while the other user group has a small amount of uniform all-to-one traffic and a high amount of uniform all-to-all traffic). This occurs in the HMH cost case when one user community has hub traffic greater than 40% while the other user community has hub traffic less than 30%. In regions where two constellation systems are deployed, C_i indicates the constellation that is built for user group i.

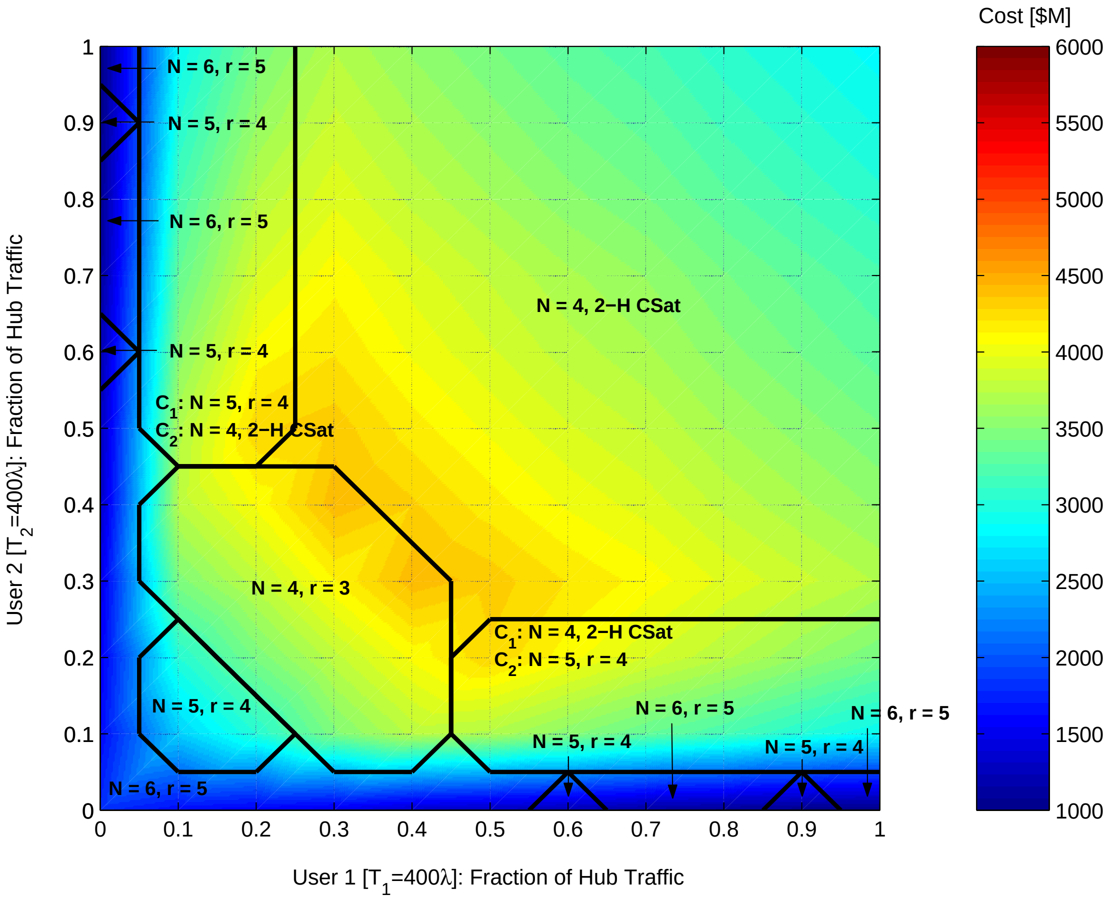


Figure A-29: Comparison of lowest communications costs of 1 constellation systems vs. 2 separate constellation systems for HMH cost scenario with mixed traffic.

A.1.3.2 Different Traffic Volumes

Next, the HMH cost scenario is considered for two user groups with different traffic volumes. Total traffic volume is kept constant at $T = 800$ wavelengths. User group 1 is the smaller community using 80 wavelengths while user group 2 uses 720 wavelengths (i.e., traffic volume of user group 1 is about 11% of traffic volume of user group 2). The communications costs for constellations tailored to each user group's traffic volume is shown Figure A-30 and Figure A-31 for the HMH case. The constellation types of lowest system cost for the various amounts of mixed traffic is re-plotted in Figure A-32 and Figure A-33.

The total communications costs of building two separate systems for the various amounts of mixed traffic are shown in Figure A-34. Here, the contours lines separate the regions of different constellation types. In each region where two constellation systems are deployed, C_i indicates the constellation that is built for user group i, which is dependent on the amount of mixed traffic within each user group.

The costs of building one satellite constellation systems to handle the sum of the two user traffics (i.e., T = 800 wavelengths) are shown in Figure A-35. Again, traffic among the two user communities does not mix within the constellation, i.e., each satellite node has two switches. Likewise, the contour lines in Figure A-35 separate the lowest cost constellation built within each region of varying mixed traffic among the two user groups.

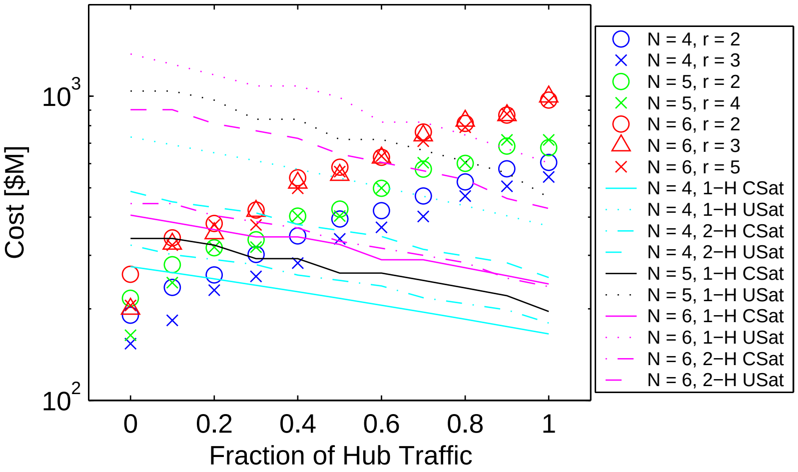


Figure A-30: Comparing communications costs HMH of N=4,5,6 for a range of mixed traffic for user community 1 ($T_1 = 80$).

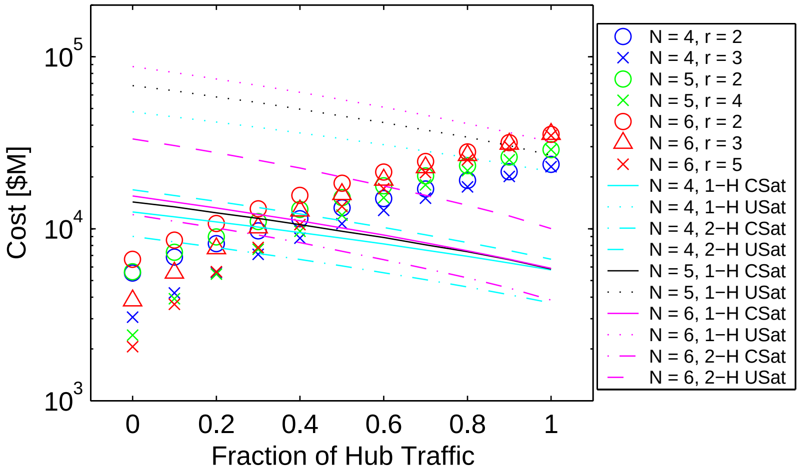


Figure A-31: Comparing communications costs HMH of N=4,5,6 for a range of mixed traffic for user community 2 ($T_2 = 720$).

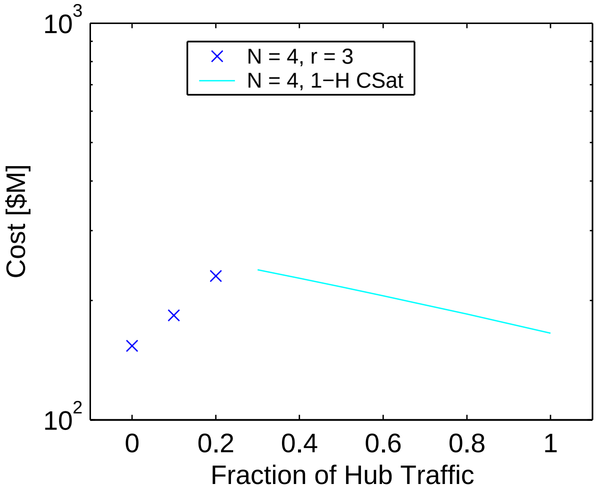


Figure A-32: Lowest communications costs for HMH cost scenario with mixed traffic for user community 1 ($T_1 = 80$).

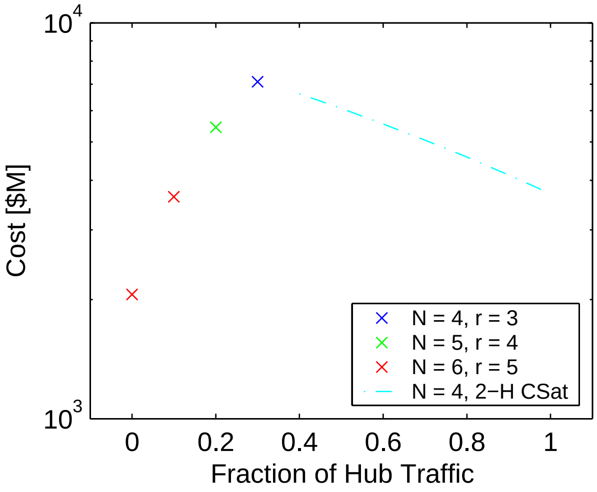


Figure A-33: Lowest communications costs for HMH cost scenario with mixed traffic for user community 2 ($T_2 = 720$).

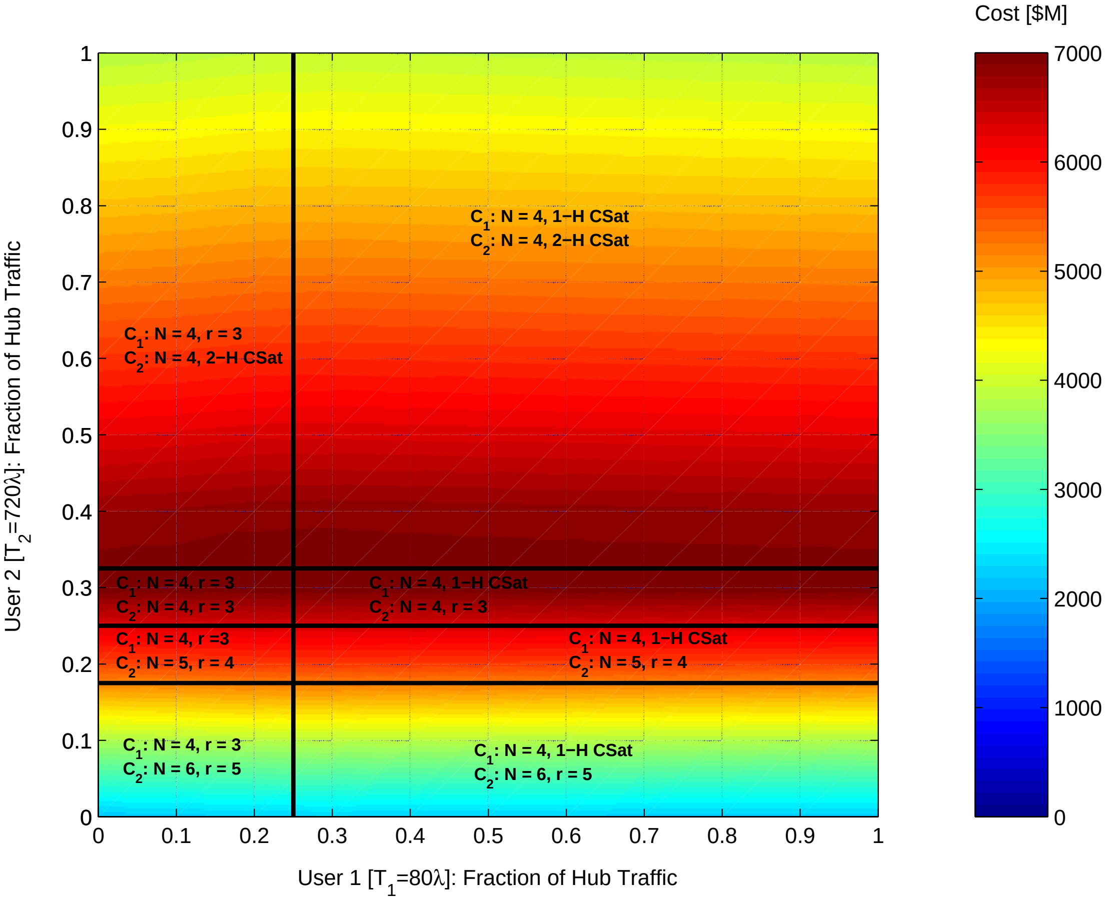


Figure A-34: Lowest communications costs for HMH cost scenario with mixed traffic of unequal traffic volumes.

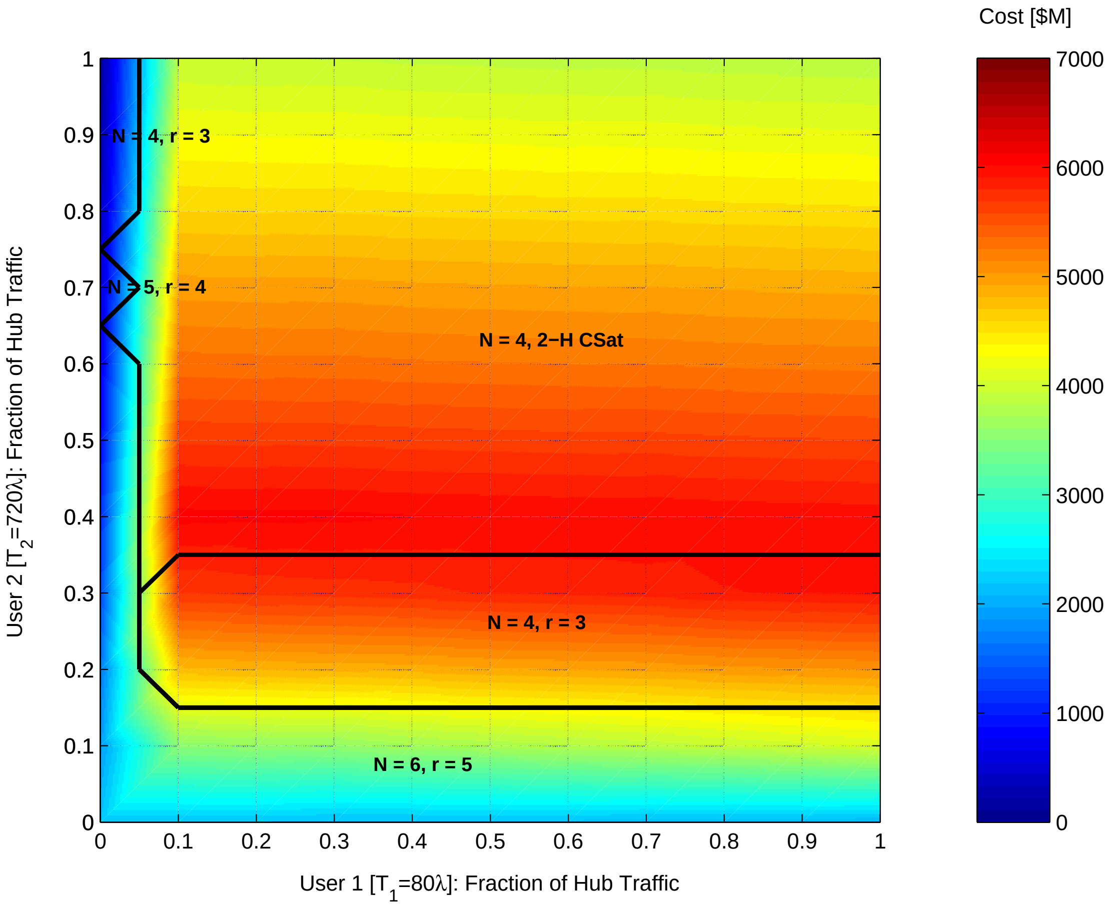


Figure A-35: Lowest communications costs of 2 separate constellation systems for HMH cost scenario with mixed traffic of unequal traffic volumes using uniform satellites.

A comparison of the costs between two separate satellite constellation systems and one satellite constellation systems for two disparate user groups is made. The results for the lowest cost systems are shown in Figure A-36. Generally, a one satellite constellation system can satisfy nearly all cases. Two separate satellite constellation systems are more cost-efficient when the two user groups of traffics are very disparate (i.e., one user group has a high amount of uniform all-to-one traffic and a low amount of uniform all-to-all traffic while the other user group has a small amount of uniform all-to-one traffic and a high amount of uniform all-to-all traffic). This occurs in the HMH cost scenario for the following situations: (1) user community 1 has hub traffic greater than 60% while user community 2 has hub traffic less than 20% and (2) user community 2 has hub traffic greater than 90% and user community 1 has hub traffic between approximately 10% and 30%. In regions where two constellation systems are deployed, C_i indicates the constellation that is built for user group i.

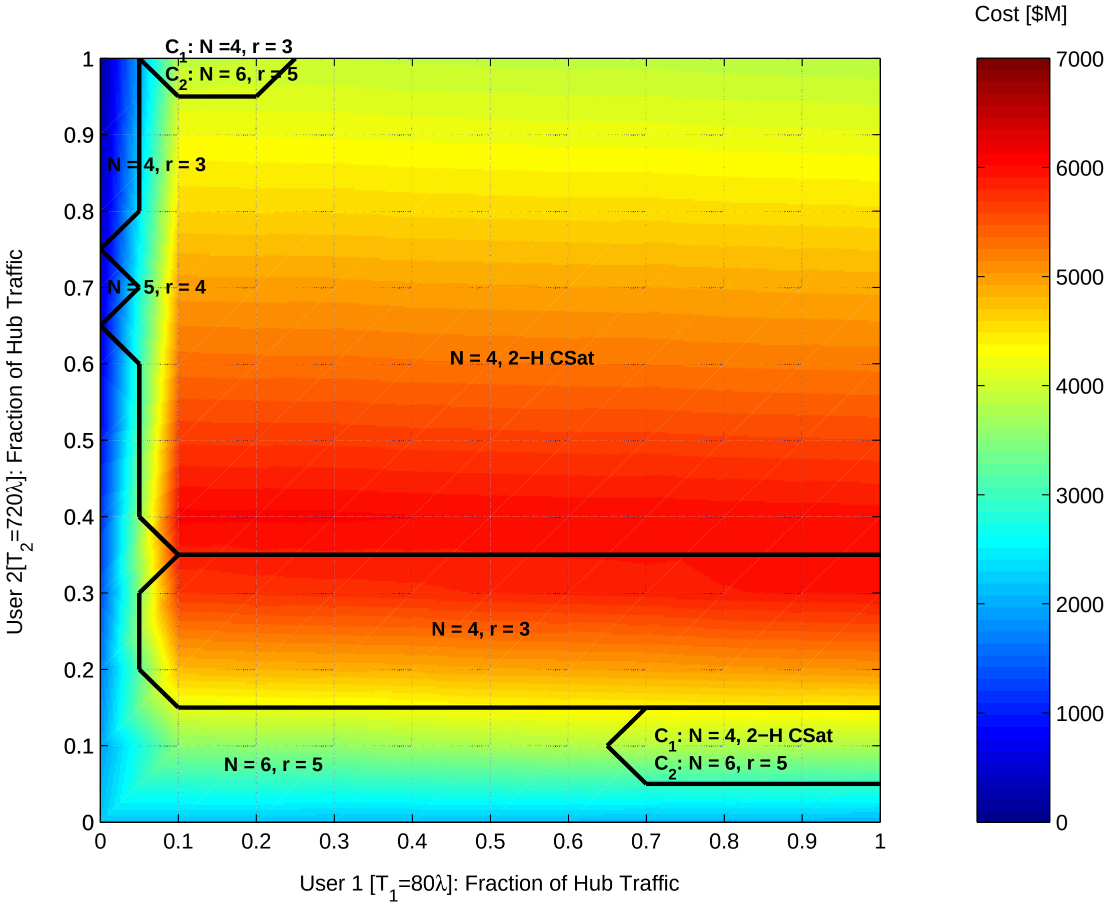
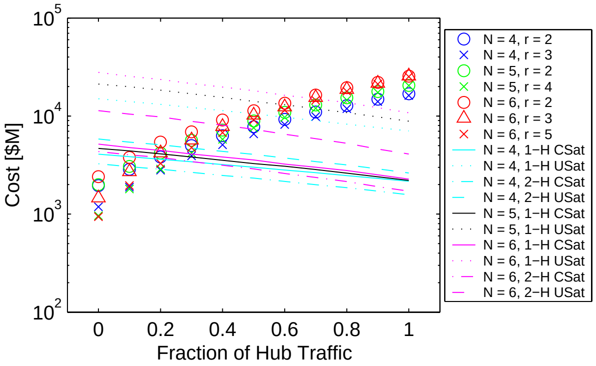


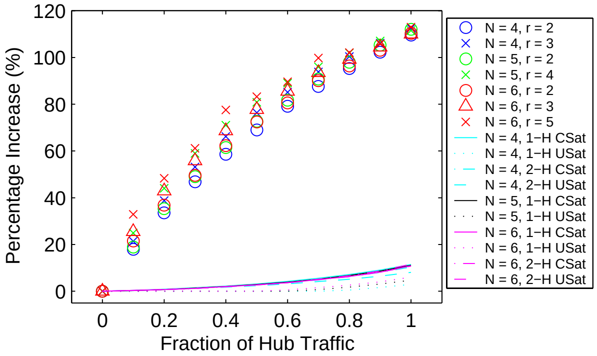
Figure A-36: Lowest communications costs of 1 constellation systems for HMH cost scenario with mixed traffic of unequal traffic volumes using uniform satellites.

A.1.4 Processor Connectivity Cases

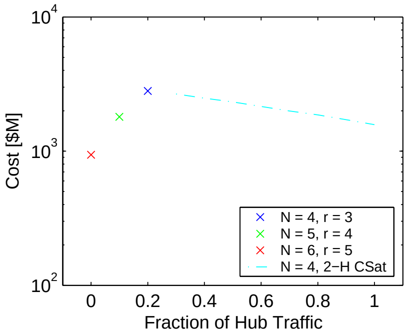
A.1.4.1 Connection to Hub Nodes



(a) Communication costs HMH.

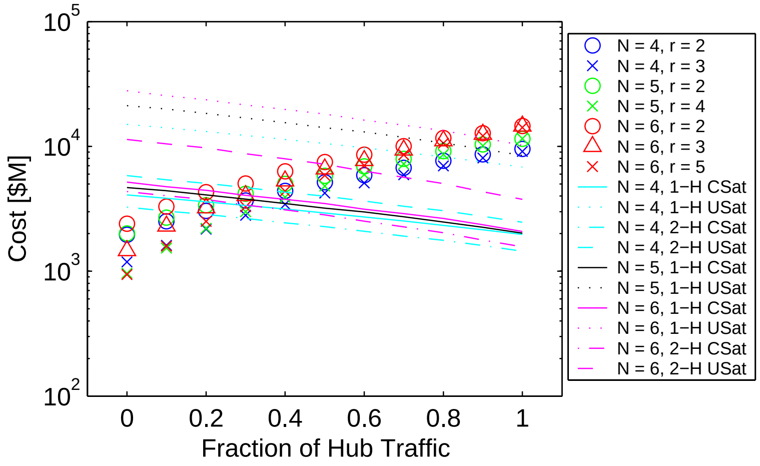


(b) Percentage increase in communications costs HMH.

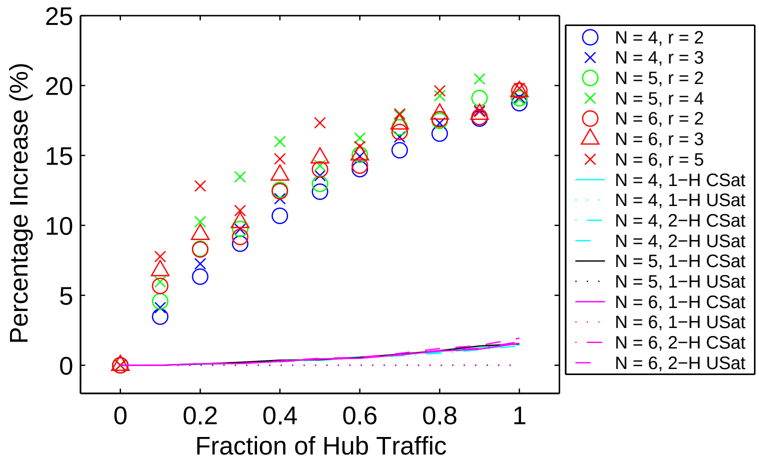


(c) Lowest communications cost constellations HMH.

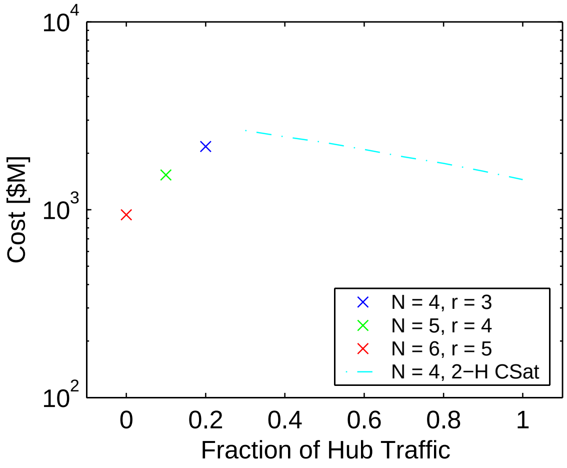
Figure A-37: Processor-hub connectivity results for compression rate 2:1 HMH.



(a) Communications costs HMH.

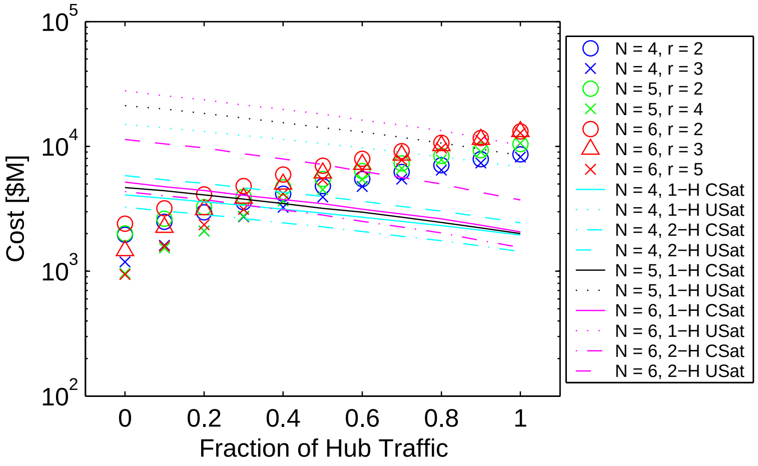


(b) Percentage increase in communications costs HMH.

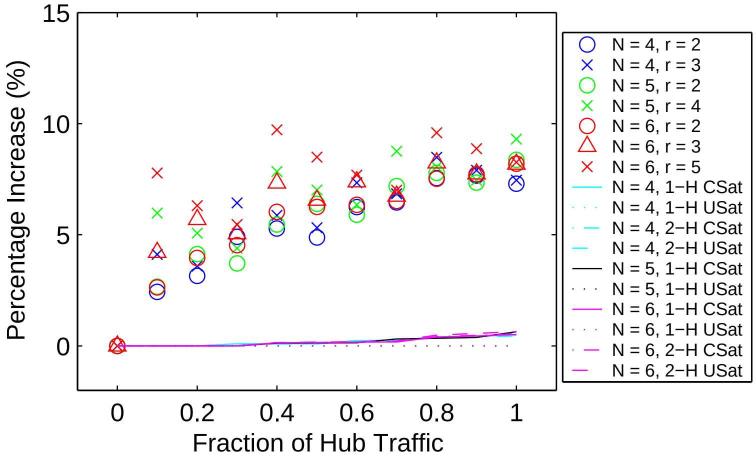


(c) Lowest communications cost constellations HMH.

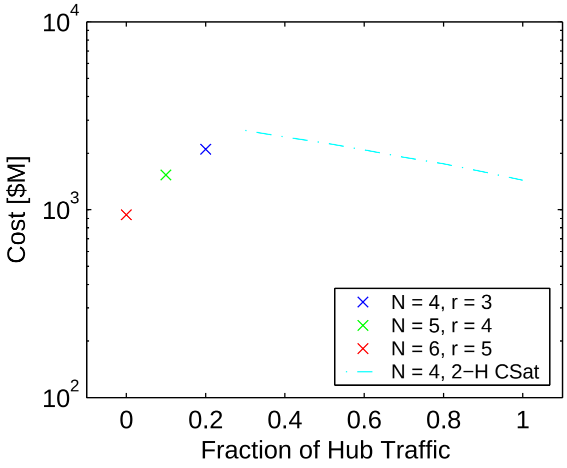
Figure A-38: Processor-hub connectivity results for compression rate 10:1 HMH.



(a) Communications costs HMH.

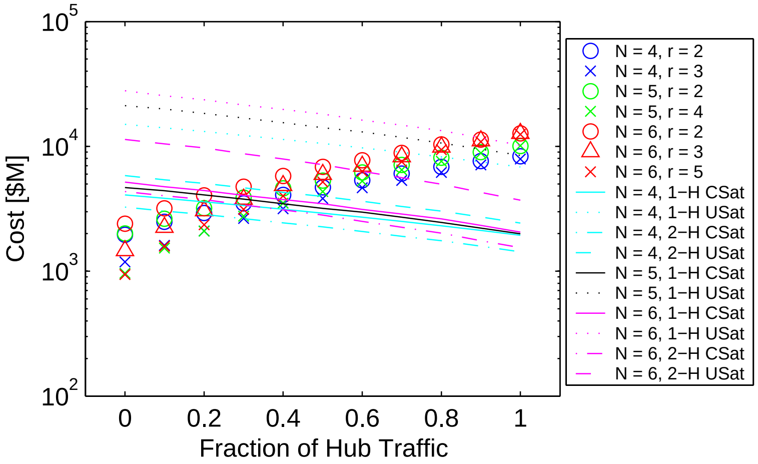


(b) Percentage increase in communications Costs HMH.

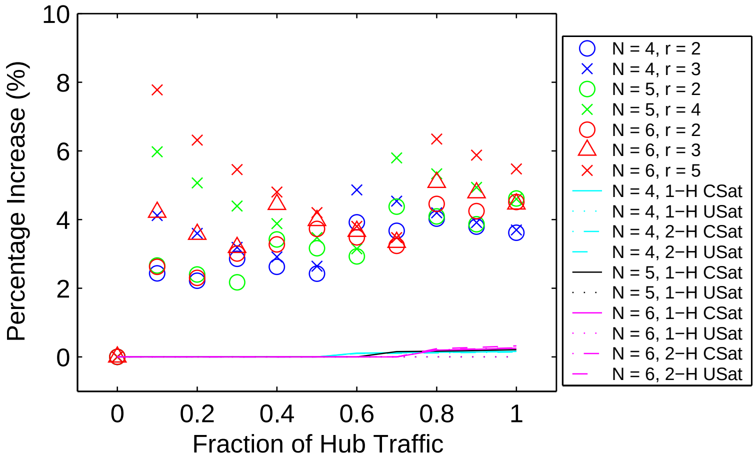


(c) Lowest communications cost constellations HMH.

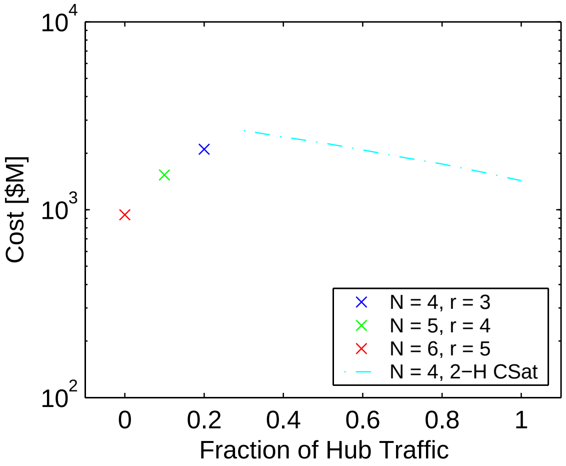
Figure A-39: Processor-hub connectivity results for compression rate 25:1 HMH.



(a) Communications costs HMH.

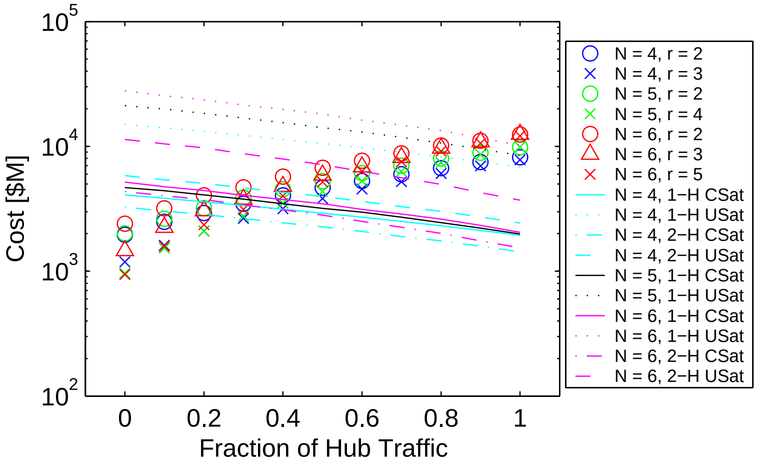


(b) Percentage increase in communications Costs HMH.

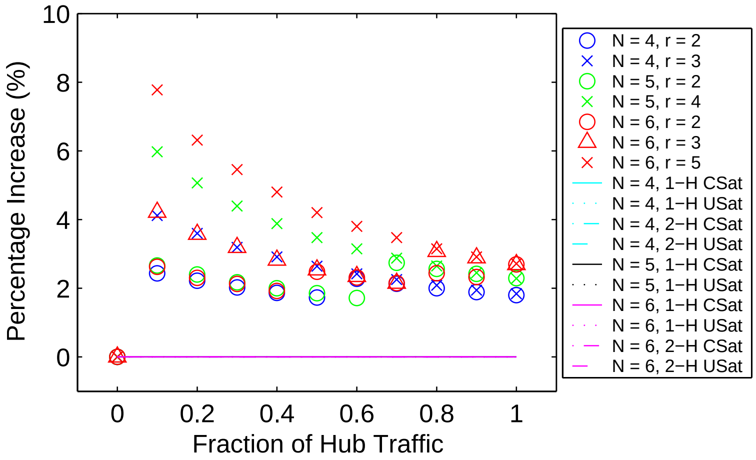


(c) Lowest communications cost constellations HMH.

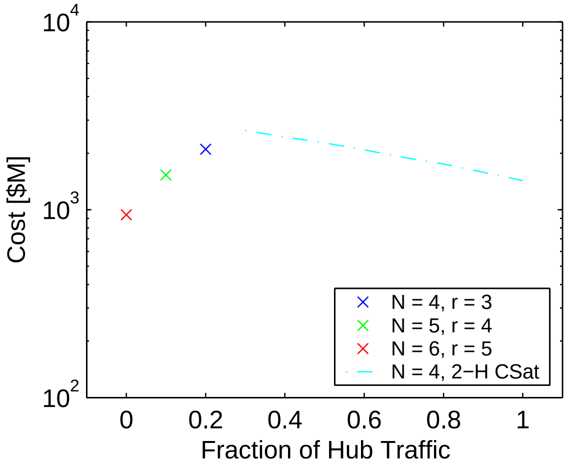
Figure A-40: Processor-hub connectivity results for compression rate 50:1 HMH.



(a) Communications costs HMH.



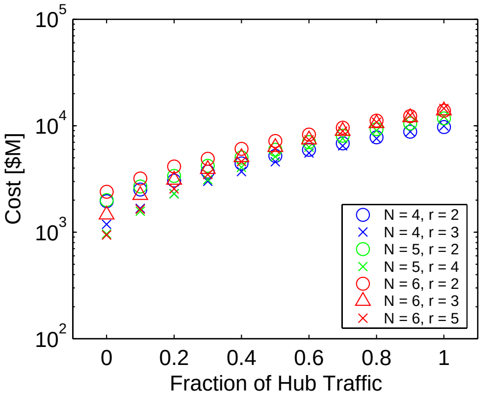
(b) Percentage increase in communications Costs HMH.



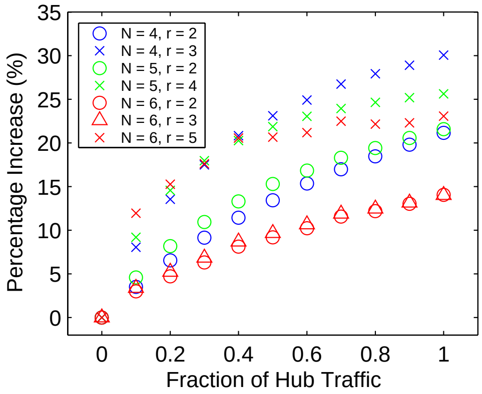
(c) Lowest communications cost constellations HMH.

Figure A-41: Processor-hub connectivity results for compression rate 100:1 HMH.

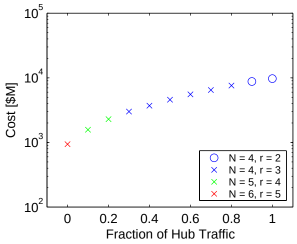
A.1.4.2 Connection to Plain Nodes



(a) Communications costs HMH.

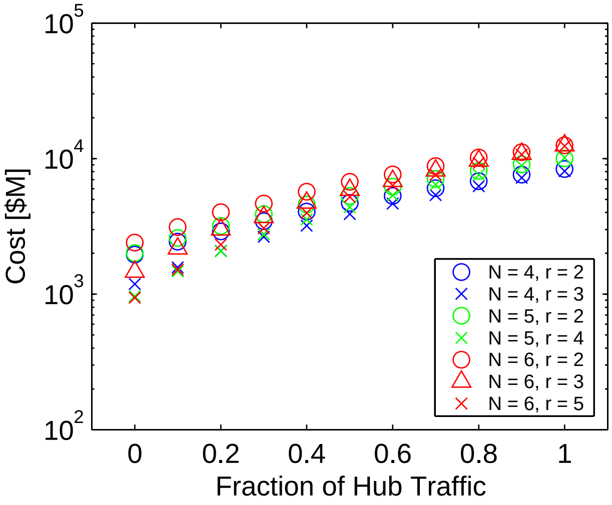


(b) Percentage increase in communications costs HMH.

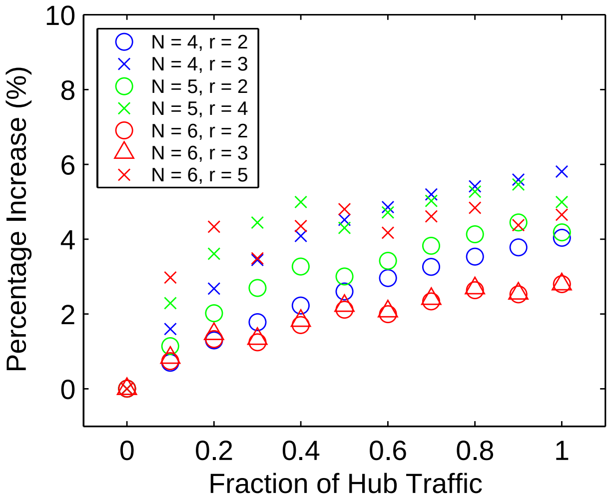


(c) Lowest communications cost constellations HMH.

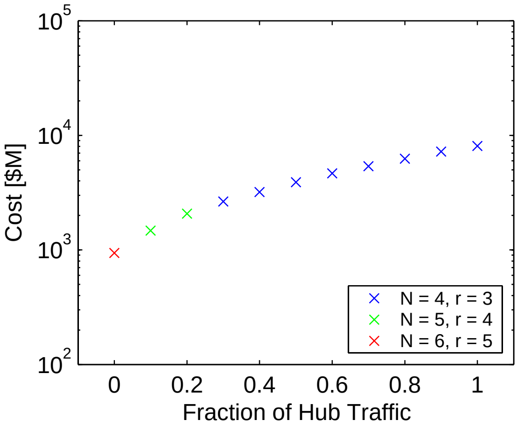
Figure A-42: Processor-plain node connectivity results for compression rate 2:1 HMH.



(a) Communications costs HMH.

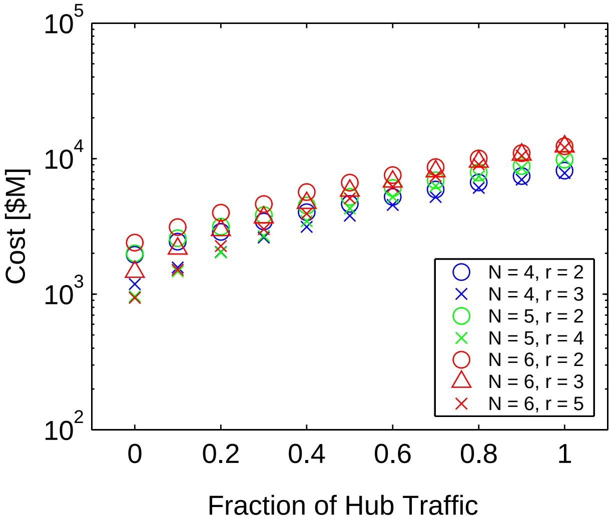


(b) Percentage increase in communications costs HMH.

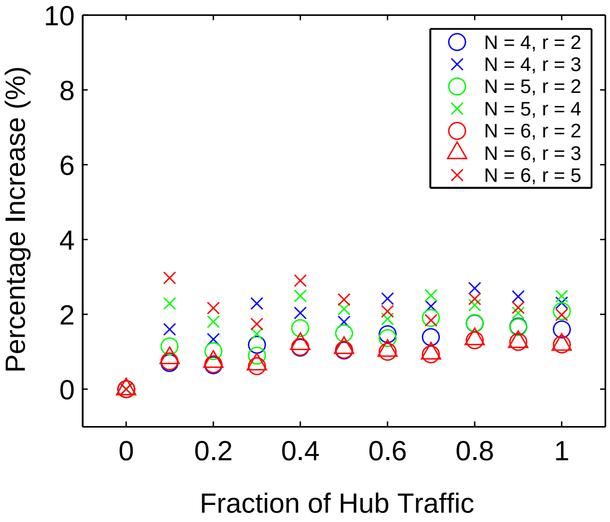


(c) Lowest communications cost constellations HMH.

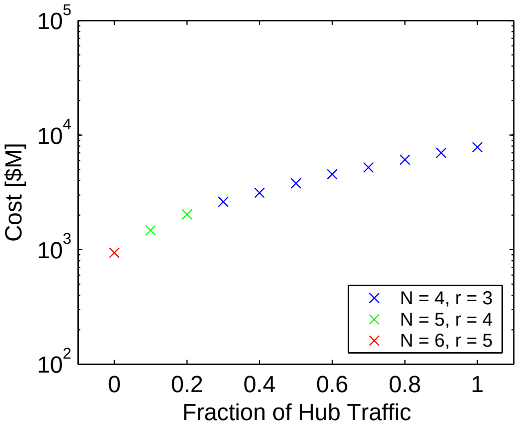
Figure A-43: Processor-plain node connectivity results for compression rate 10:1 HMH.



(a) Communications costs HMH.

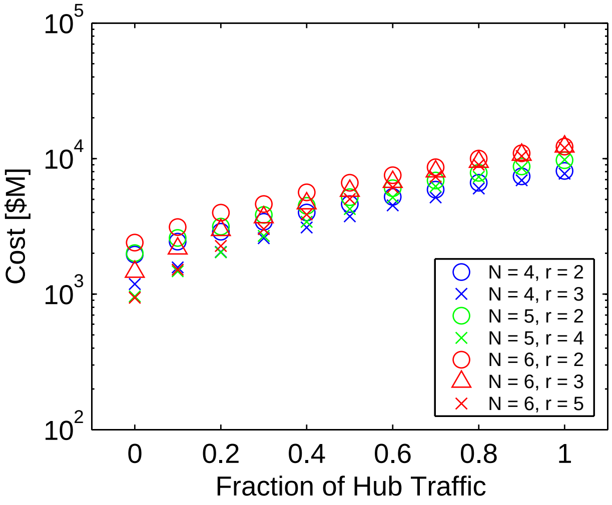


(b) Percentage increase in communications costs HMH.

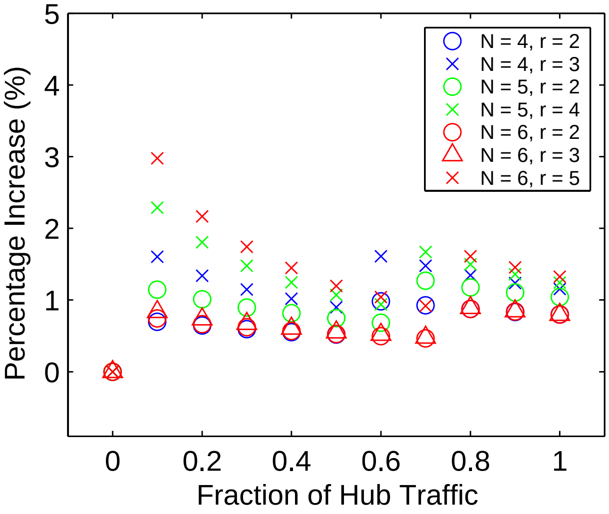


(c) Lowest communications cost constellations HMH.

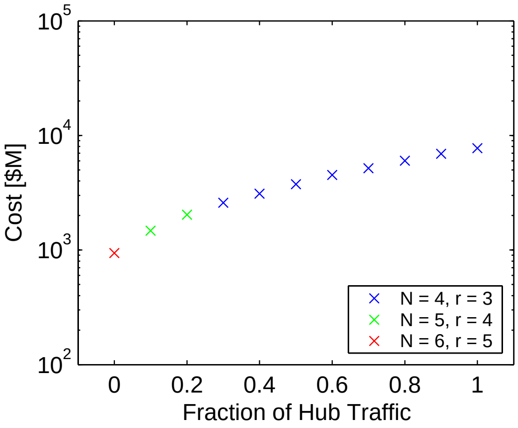
Figure A-44: Processor-plain node connectivity results for compression rate 25:1 HMH.



(a) Communications costs HMH.

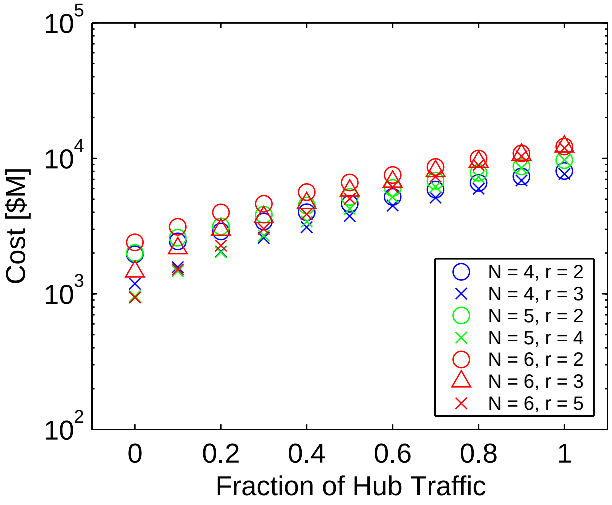


(b) Percentage increase in communications costs HMH.

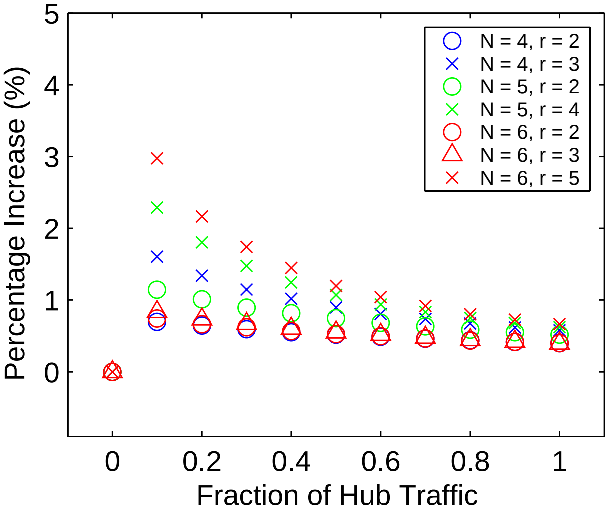


(c) Lowest communications cost constellations HMH.

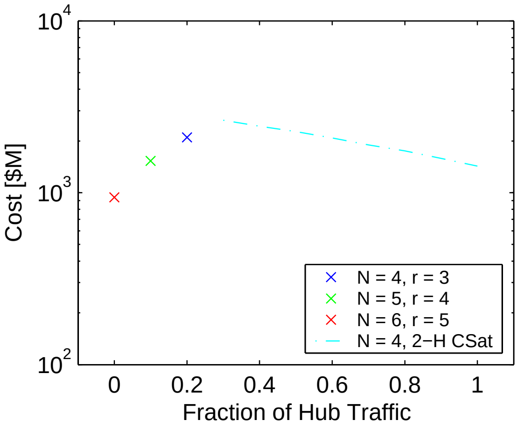
Figure A-45: Processor-plain node connectivity results for compression rate 50:1 HMH.



(a) Communications costs HMH.



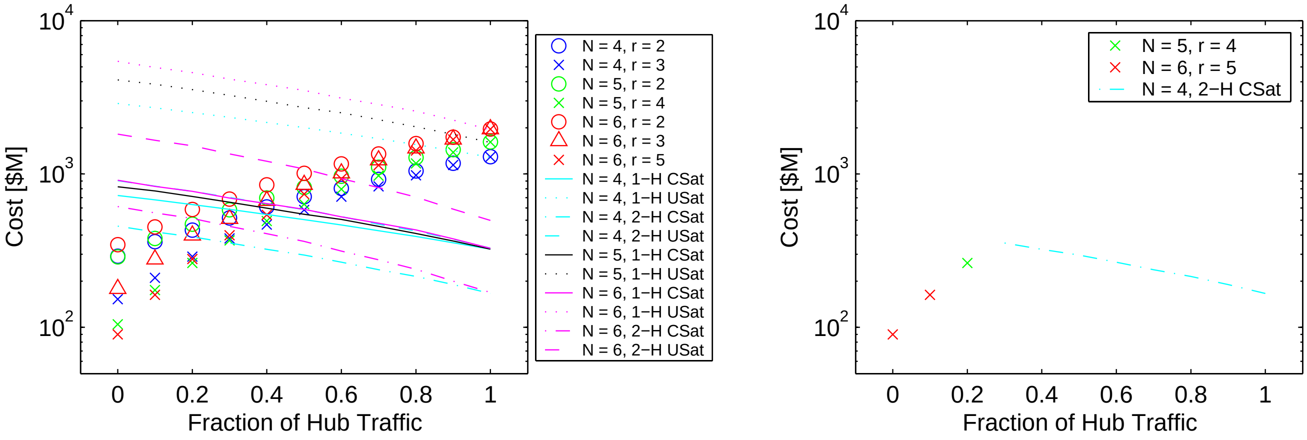
(b) Percentage increase in communications costs HMH.



(c) Lowest communications cost constellations HMH.

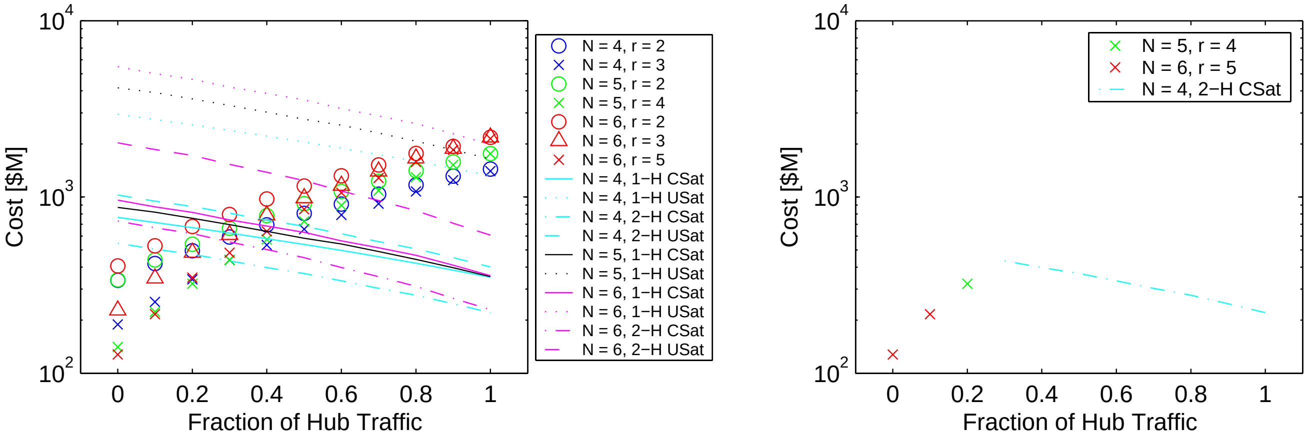
Figure A-46: Processor-plain node connectivity results for compression rate 100:1 HMH.

A.2 Cost Results for Small Constellations: All Cost Permutations



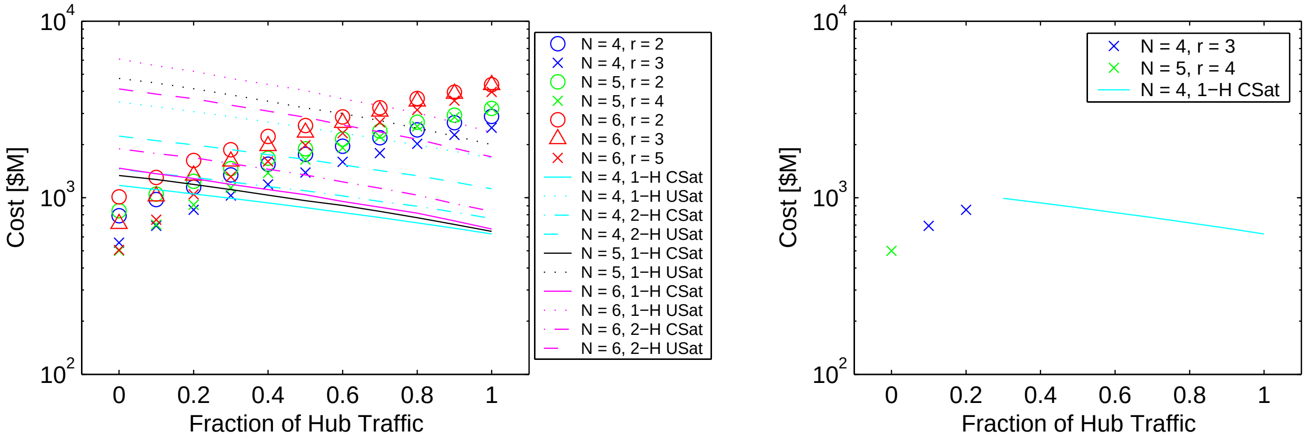
(a) Communications costs for N=4,5,6.
(b) Lowest communications costs for N=4,5,6.

Figure A-47: Communications costs LLL with mixed traffic.



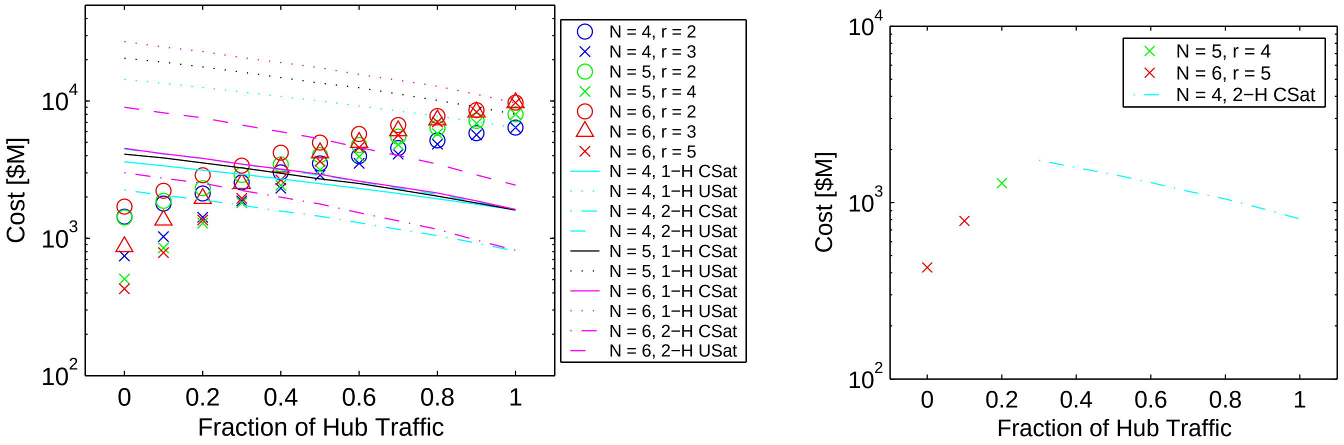
(a) Communications costs for N=4,5,6.
(b) Lowest communications costs for N=4,5,6.

Figure A-48: Communications costs LLM with mixed traffic.



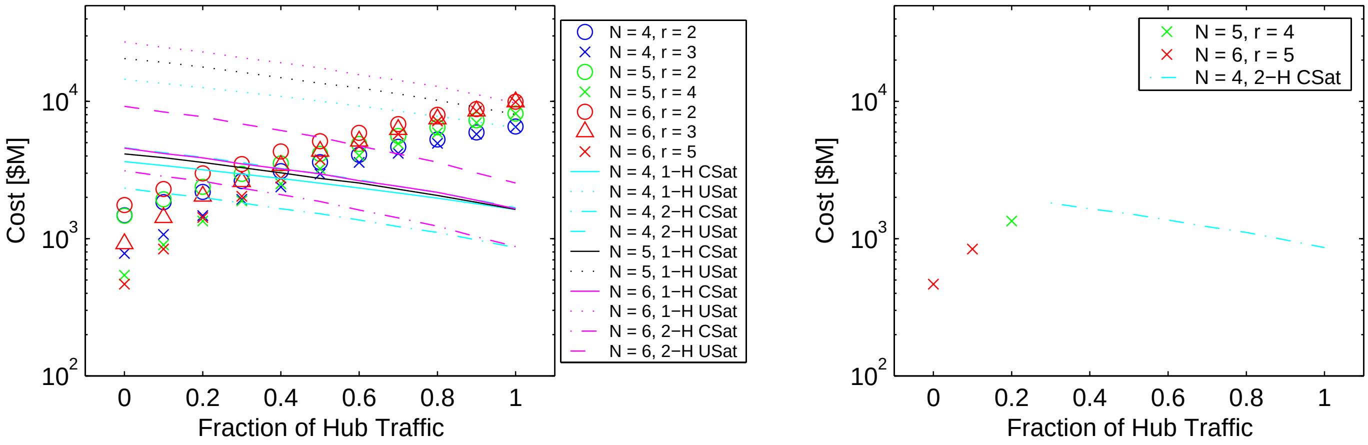
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-49: Communications costs LLH with mixed traffic.



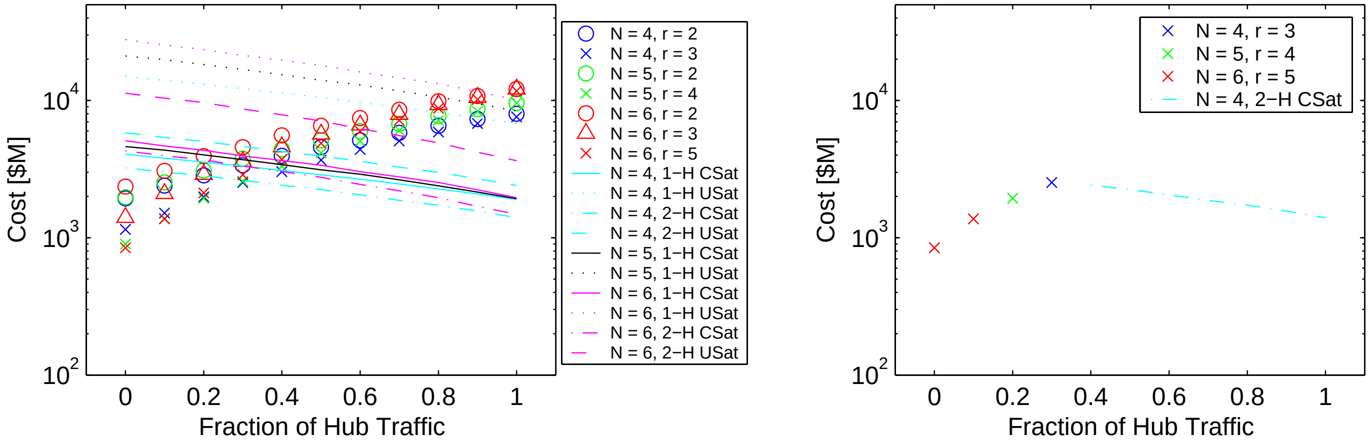
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-50: Communications costs LML with mixed traffic.



(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

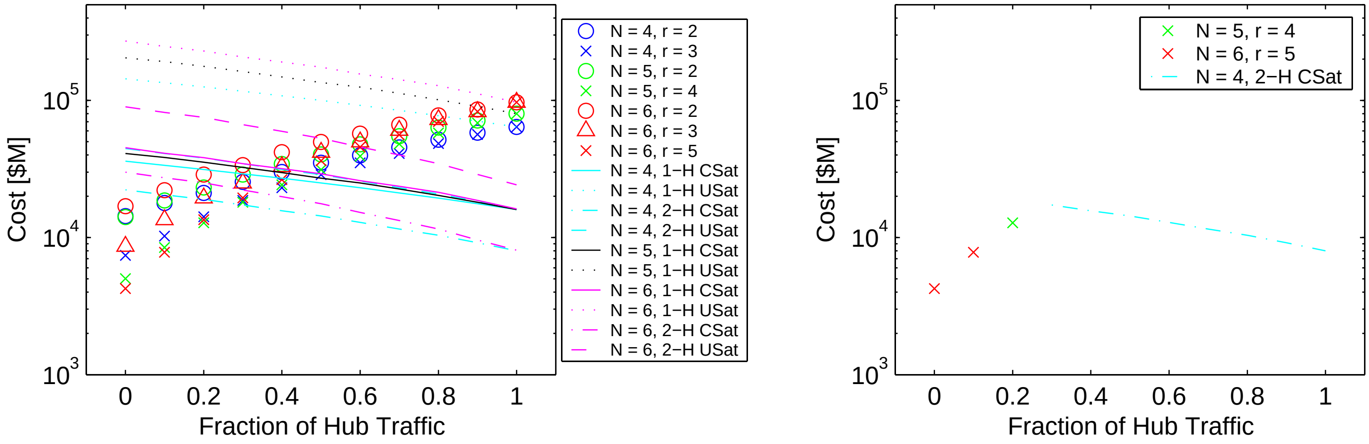
Figure A-51: Communications costs LMM with mixed traffic.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

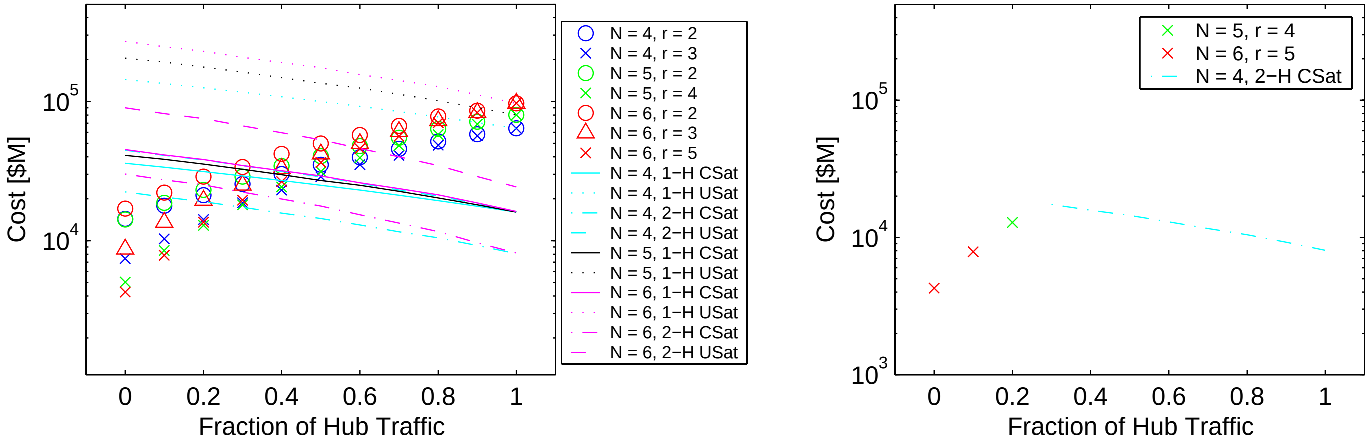
Figure A-52: Communications costs LMH with mixed traffic.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

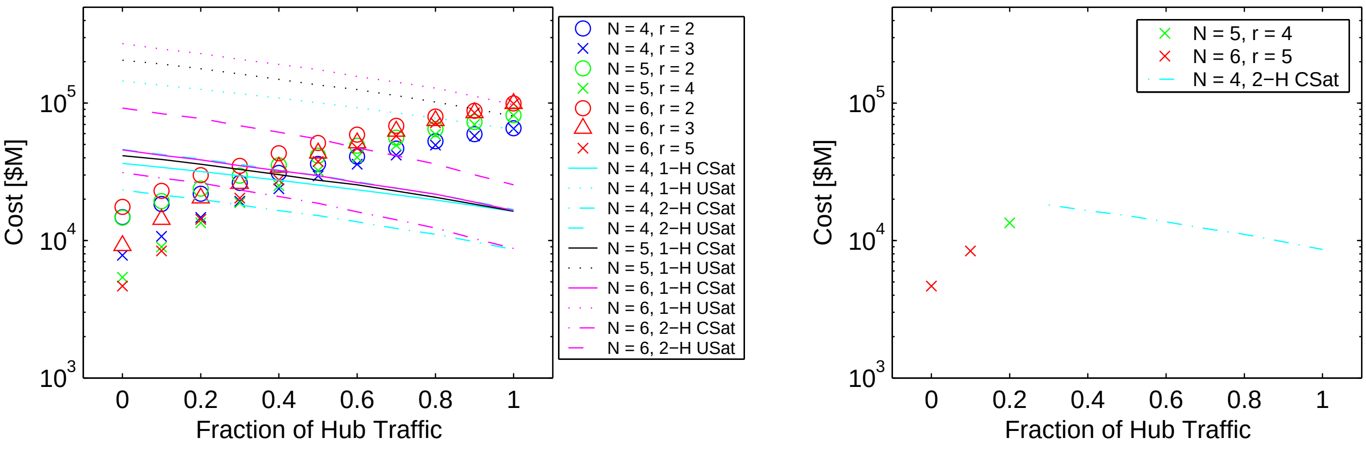
Figure A-53: Communications costs LHL with mixed traffic.



(a) Communications costs for N=4,5,6.

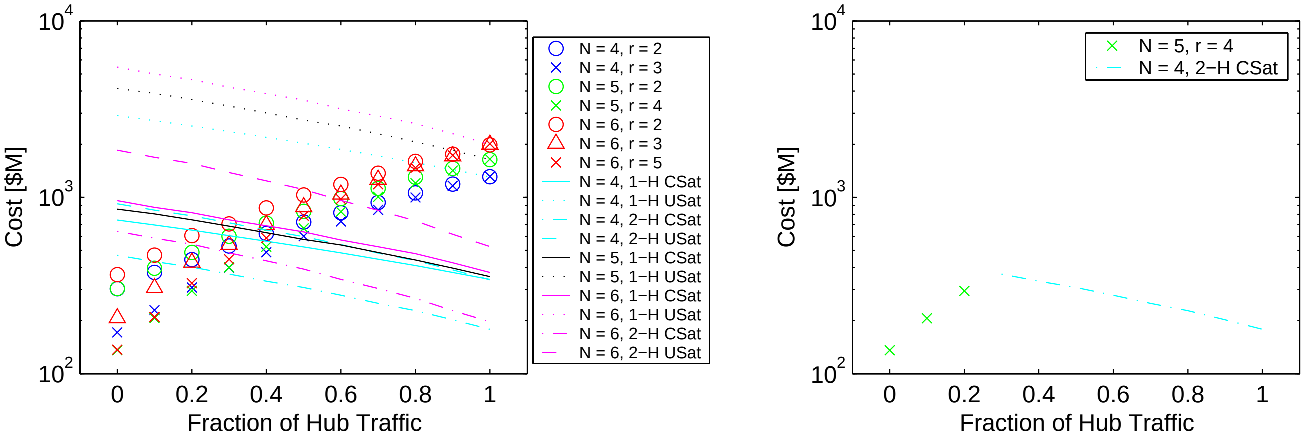
(b) Lowest communications costs for N=4,5,6.

Figure A-54: Communications costs LHM with mixed traffic.



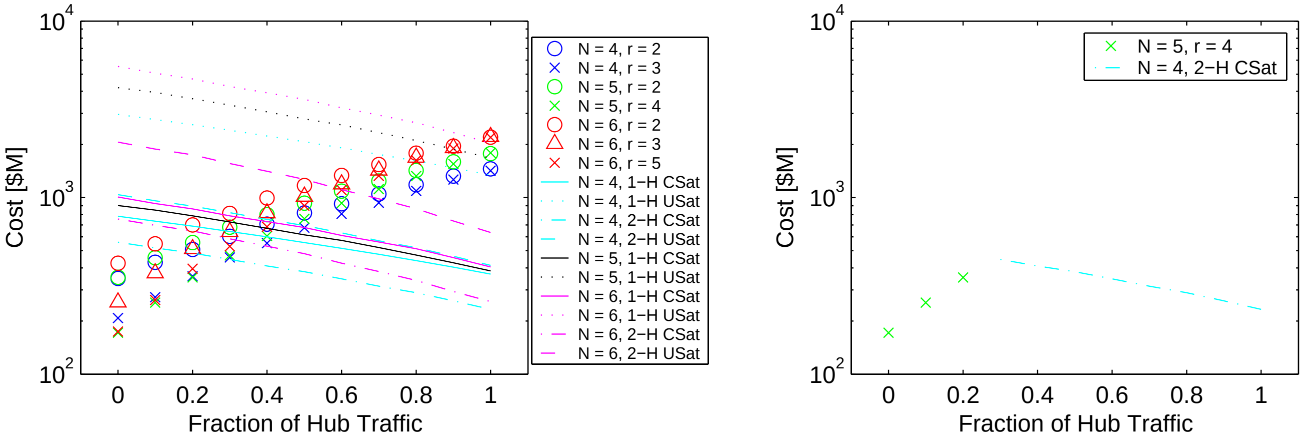
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-55: Communications costs LHH with mixed traffic.



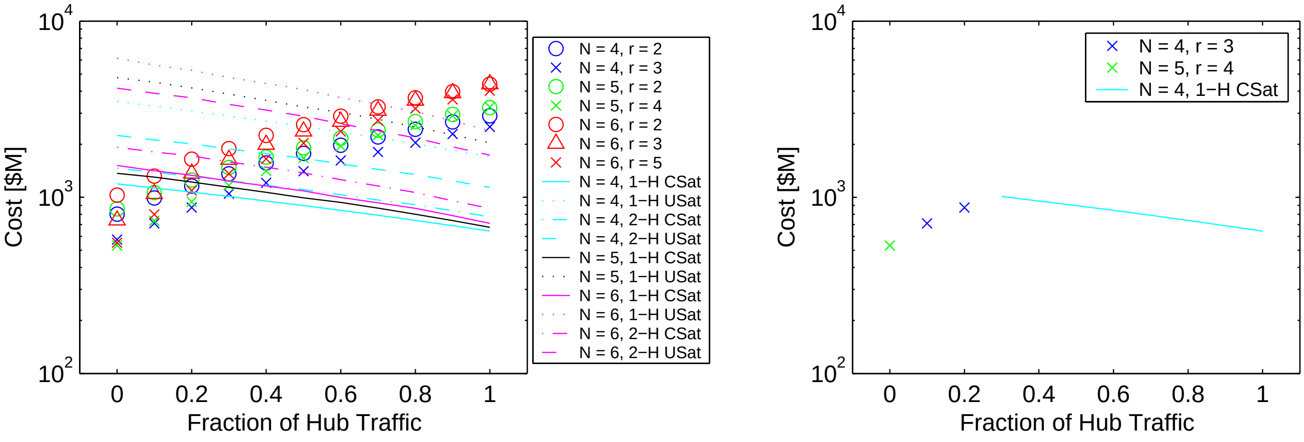
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-56: Communications costs MLL with mixed traffic.



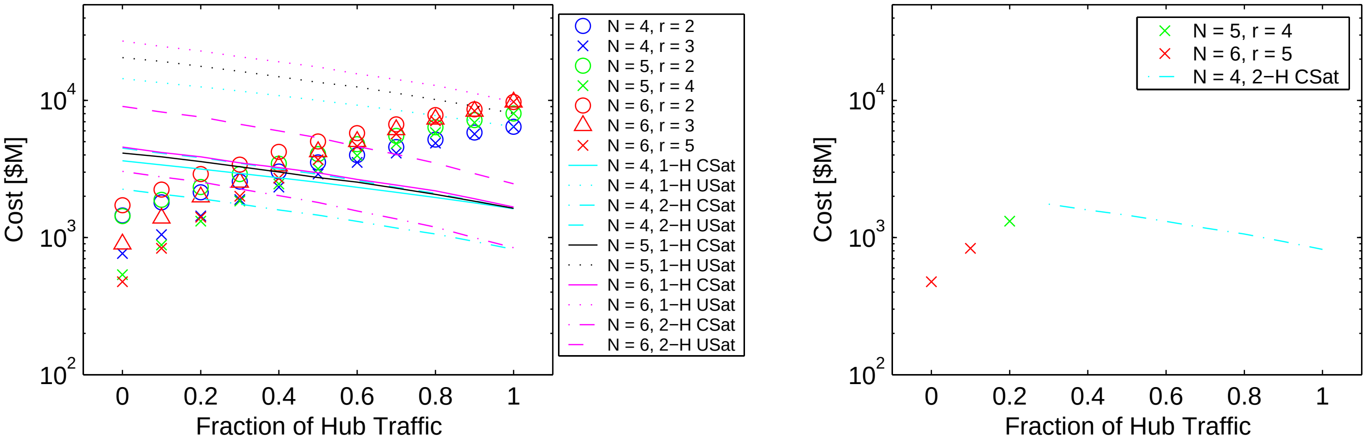
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-57: Communications costs MLM with mixed traffic.



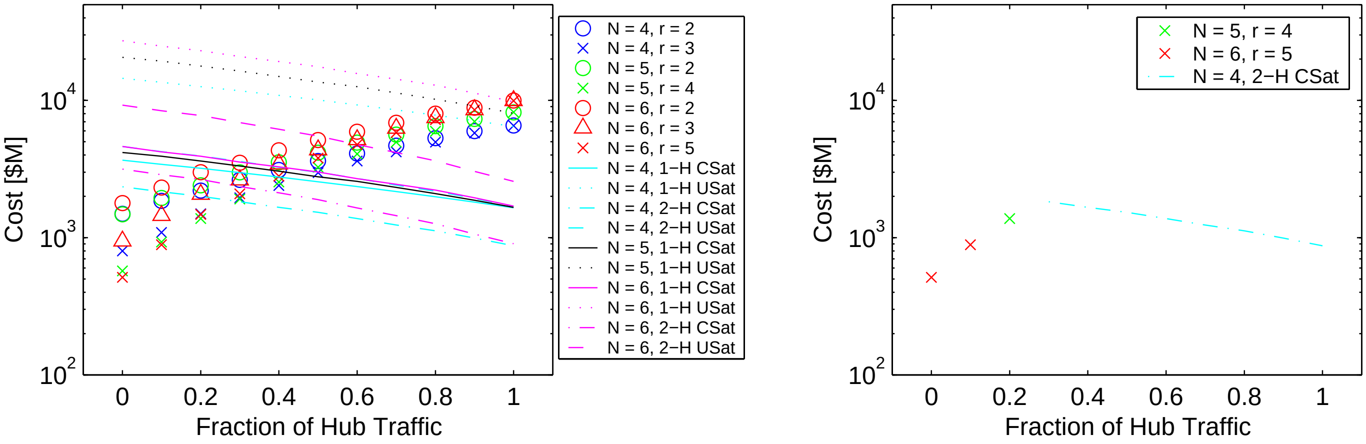
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-58: Communications costs MLH with mixed traffic.



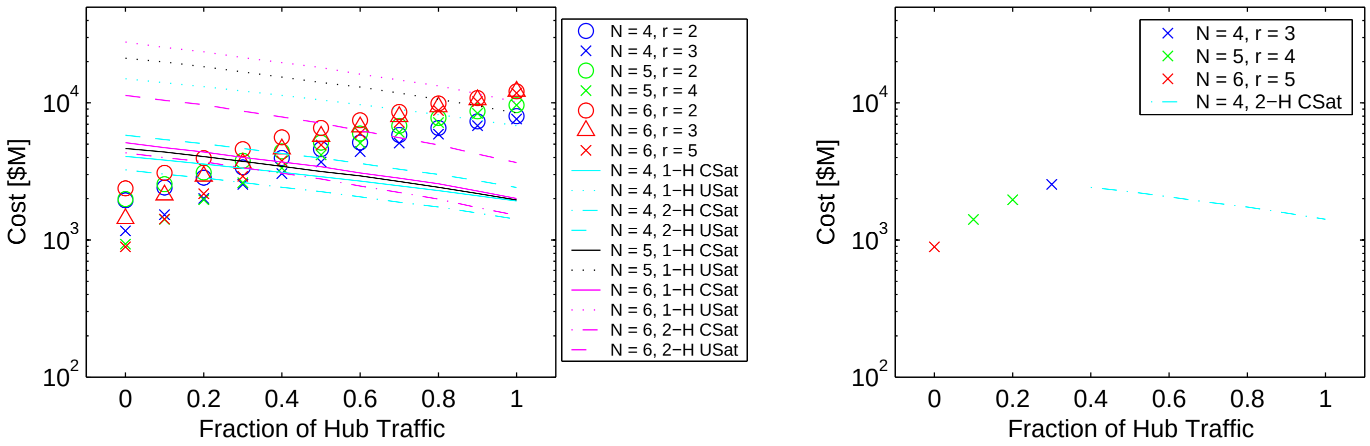
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-59: Communications costs MML with mixed traffic.



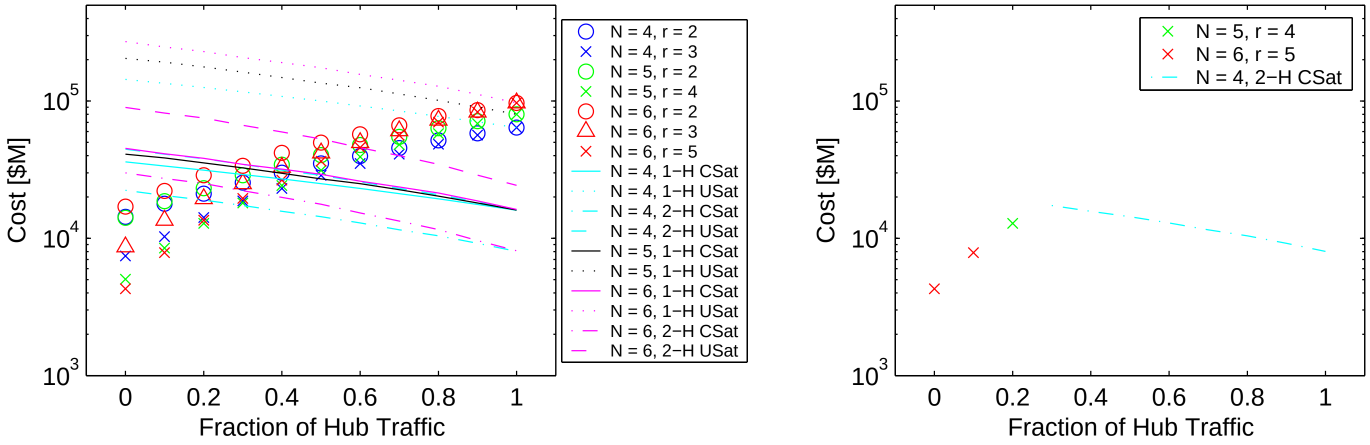
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-60: Communications costs MMM with mixed traffic.



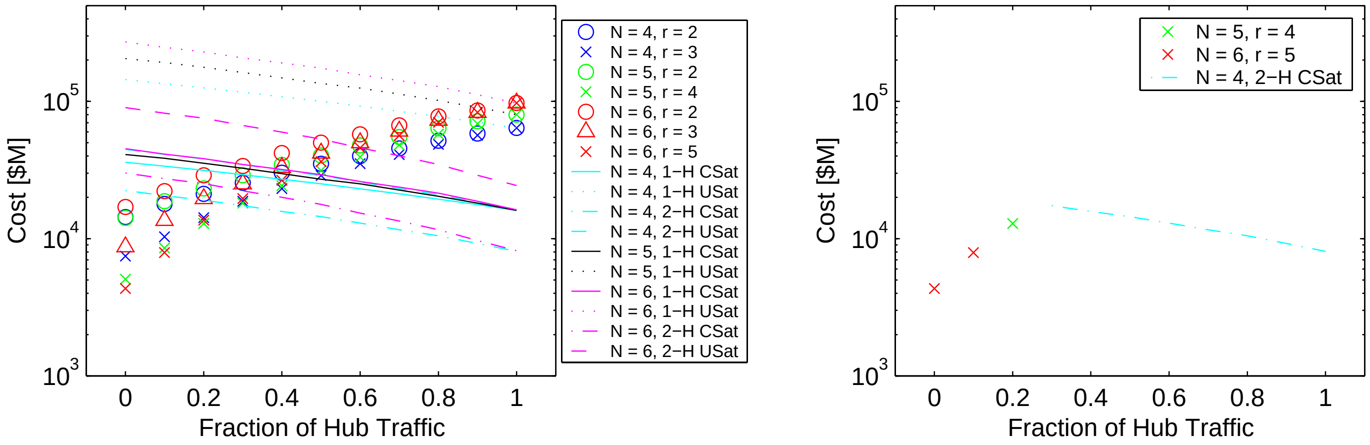
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-61: Communications costs MMH with mixed traffic.



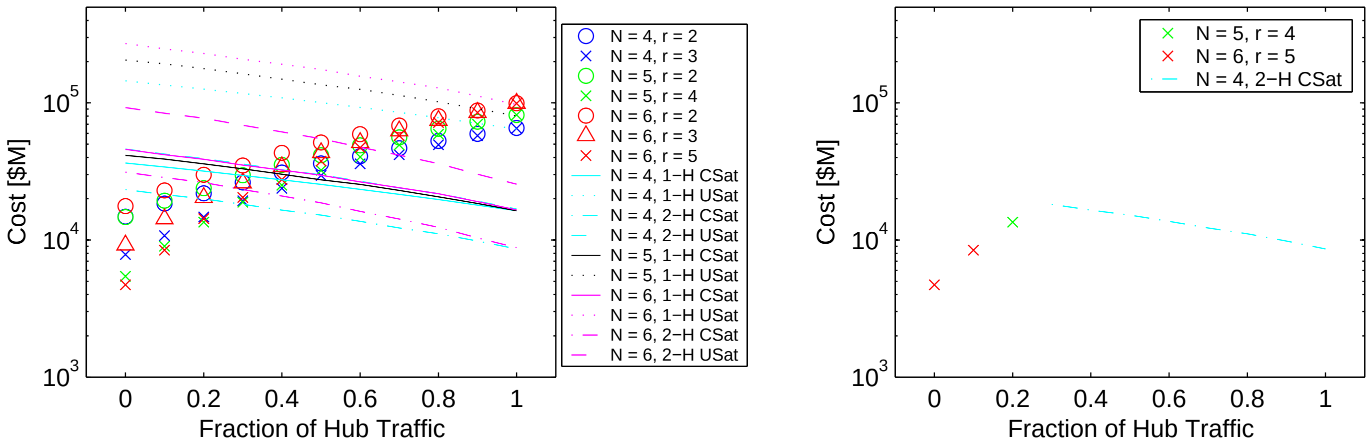
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-62: Communications costs MHL with mixed traffic.



(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

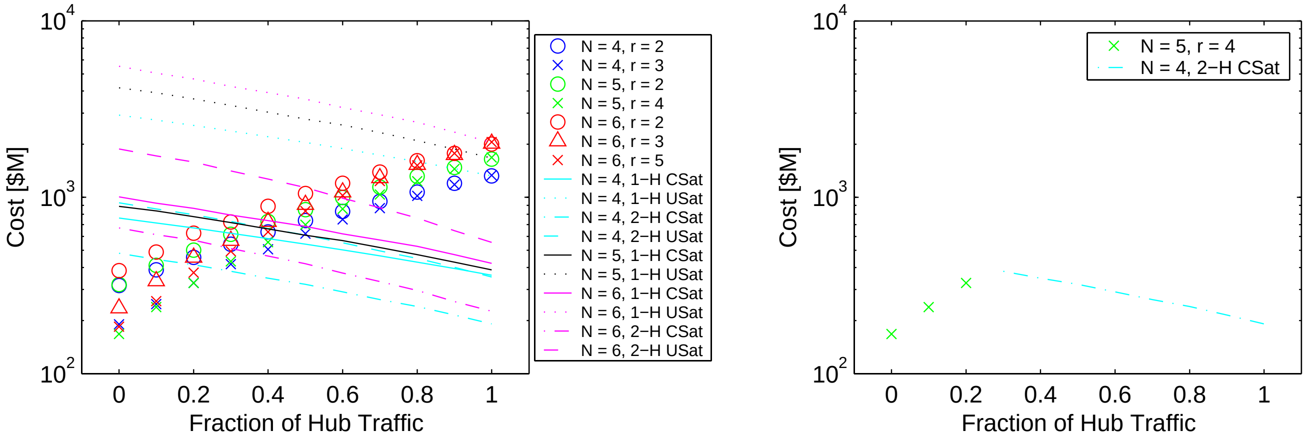
Figure A-63: Communications costs MHM with mixed traffic.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

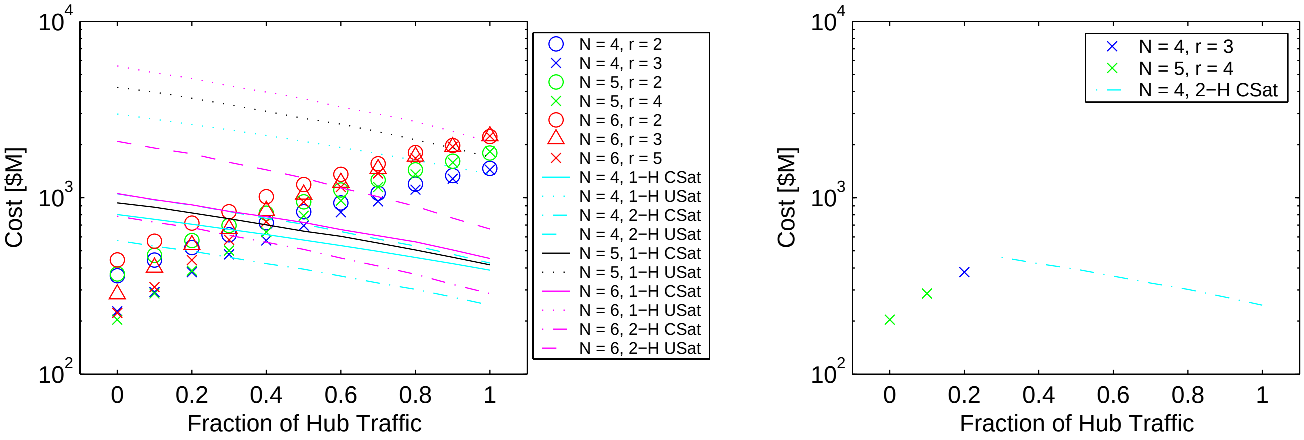
Figure A-64: Communications costs MHH with mixed traffic.



(a) Communications costs for N=4,5,6.

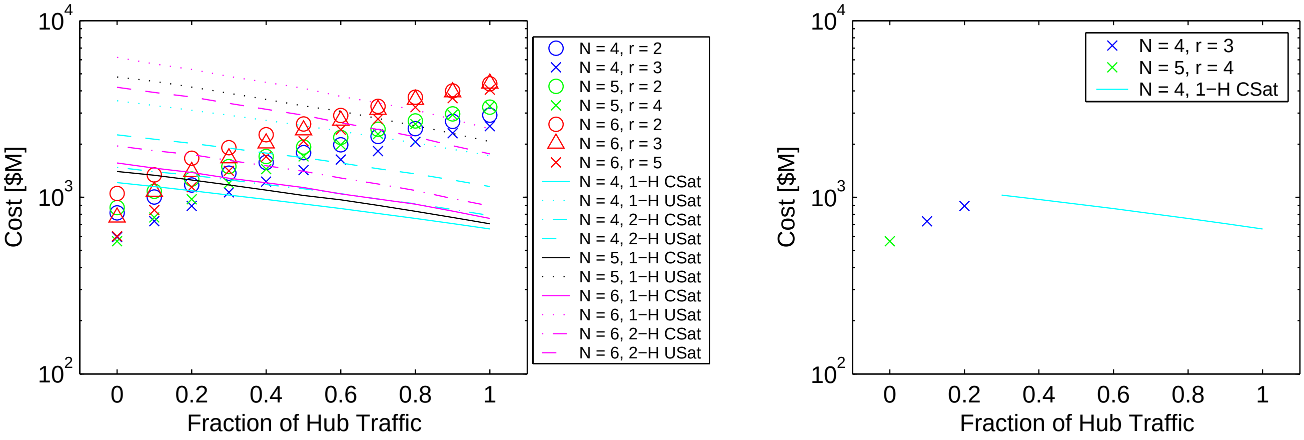
(b) Lowest communications costs for N=4,5,6.

Figure A-65: Communications costs HLL with mixed traffic.



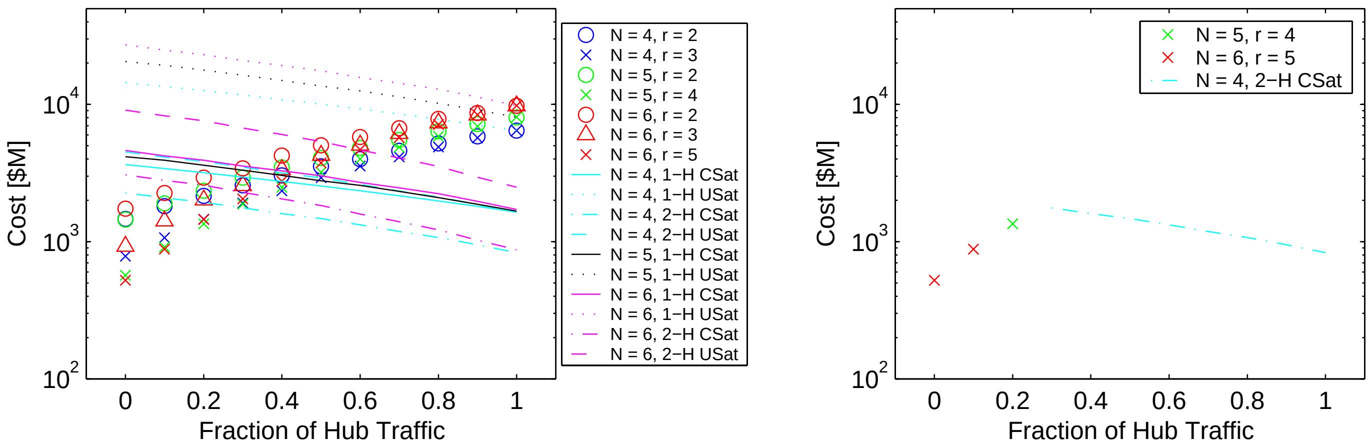
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-66: Communications costs HLM with mixed traffic.



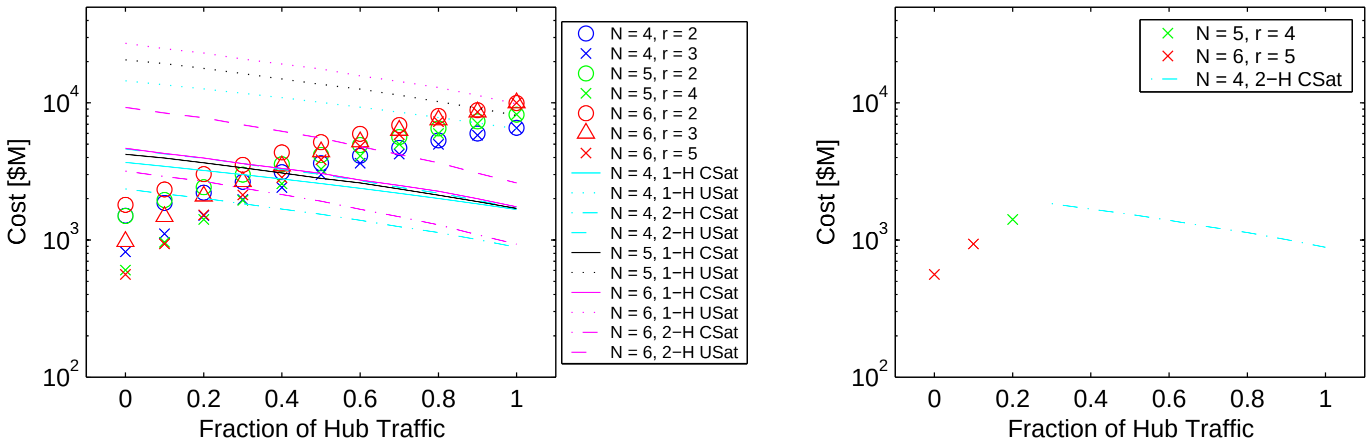
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-67: Communications costs HLH with mixed traffic.



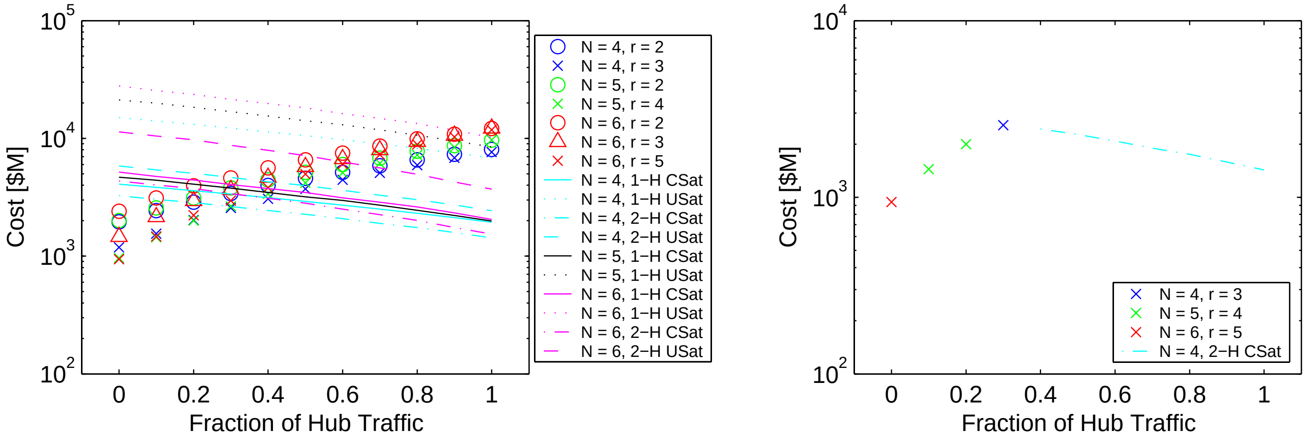
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-68: Communications costs HML with mixed traffic.



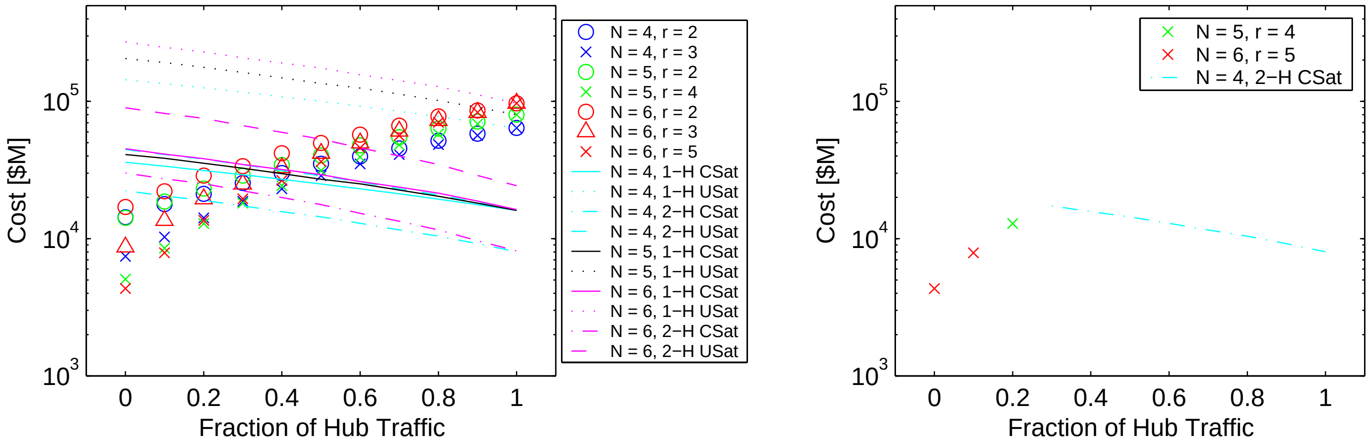
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-69: Communications costs HMM with mixed traffic.



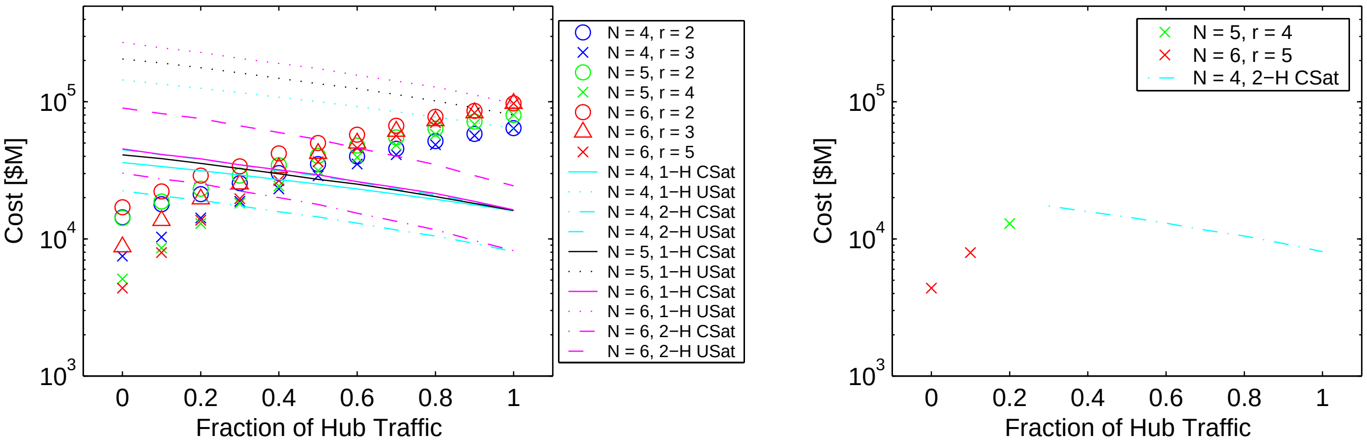
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-70: Communications costs HMH with mixed traffic.



(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

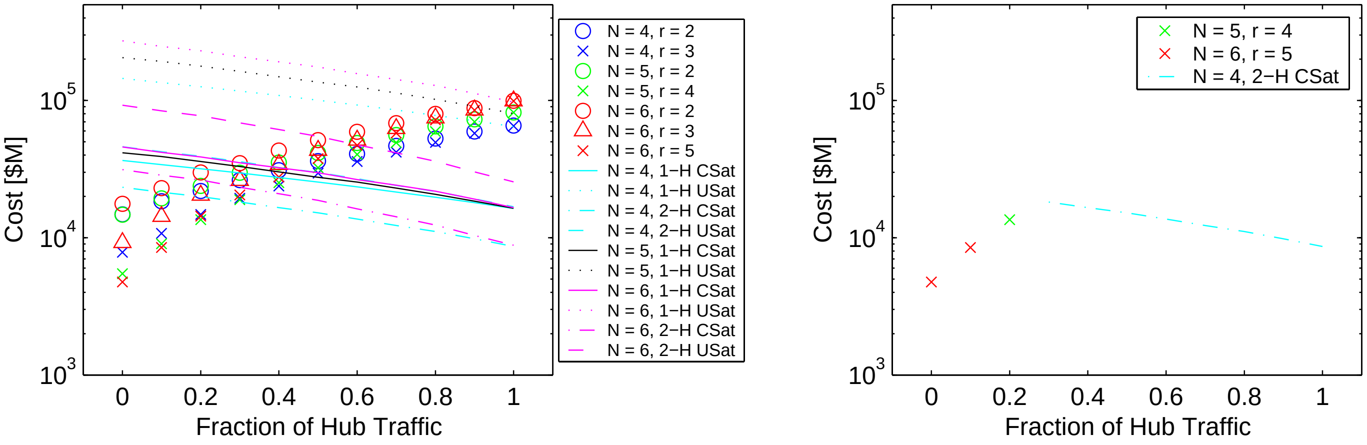
Figure A-71: Communications costs HHL with mixed traffic.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

Figure A-72: Communications costs HHM with mixed traffic.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

Figure A-73: Communications costs HHH with mixed traffic.

A.3 Cost Results using Linear Switches

The communications cost for uniform all-to-all traffic on connected circulant constellations for the MLM and the HMH cost case are shown in Figure A-74 and Figure A-75, respectively. For each cost case, two design examples are shown. The first design example uses custom antennas on the satellites, i.e., the antenna apertures are sized to the intersatellite link distance required to make the network connection. In the second design example, every antenna on the satellites is uniform, i.e., the antenna apertures are sized to the largest theoretical intersatellite link distance ($\theta = 180°$). Subfigures (a) and (b) show the communications cost for connected circulant networks with uniform jump spaces as a function of N nodes and node degree r. Subfigures (c) and (d) show which node degree r provides the lowest communications cost for every N number of nodes. The percentage increase in communications cost between using custom antennas and uniform antennas for both the MLM and the HMH cost case are shown in Figure A-76. The percentage increase in communications cost for using uniform antennas is small ($< 30\%$ overall).

The communications cost for uniform all-to-one traffic on hub constellations for the MLM and the HMH cost case are shown in Figure A-77. Note that the results shown in this section highlight the same cost trends as the results shown for the non-linear switch.

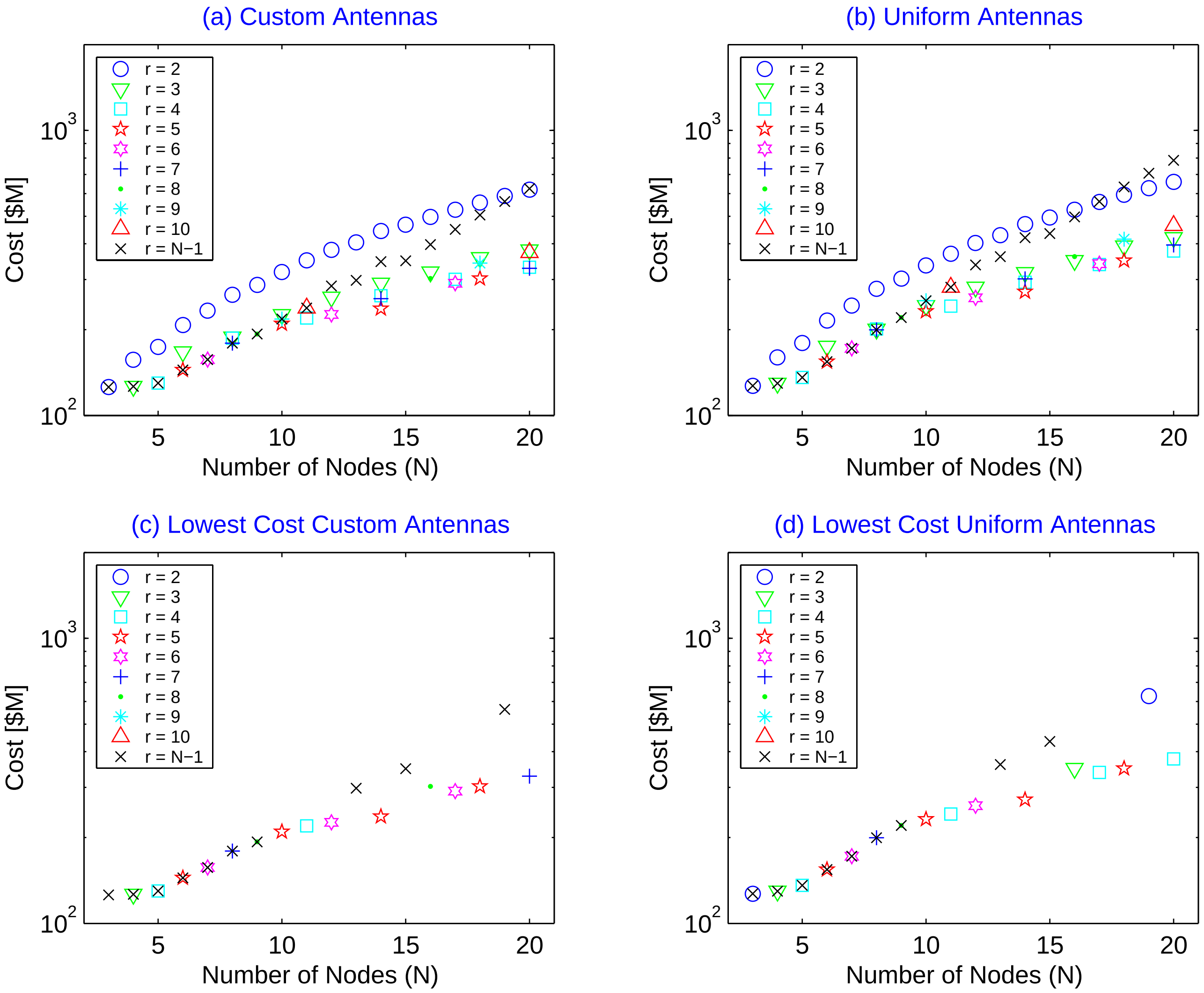


Figure A-74: Connected circulant constellations with uniform jump spaces: communications cost MLM for uniform all-to-all traffic.

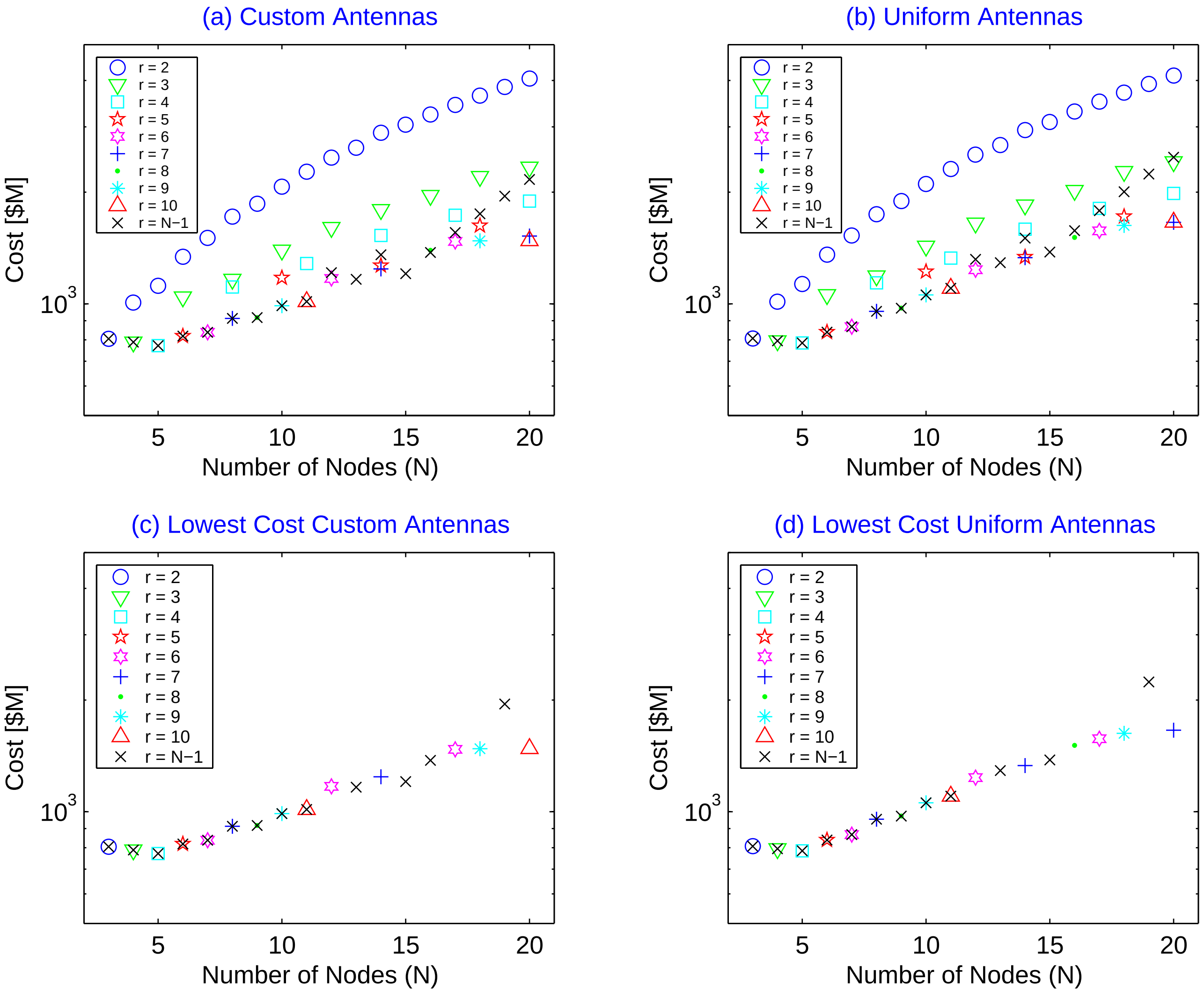


Figure A-75: Connected circulant constellations with uniform jump spaces: communications cost HMH for uniform all-to-all traffic.

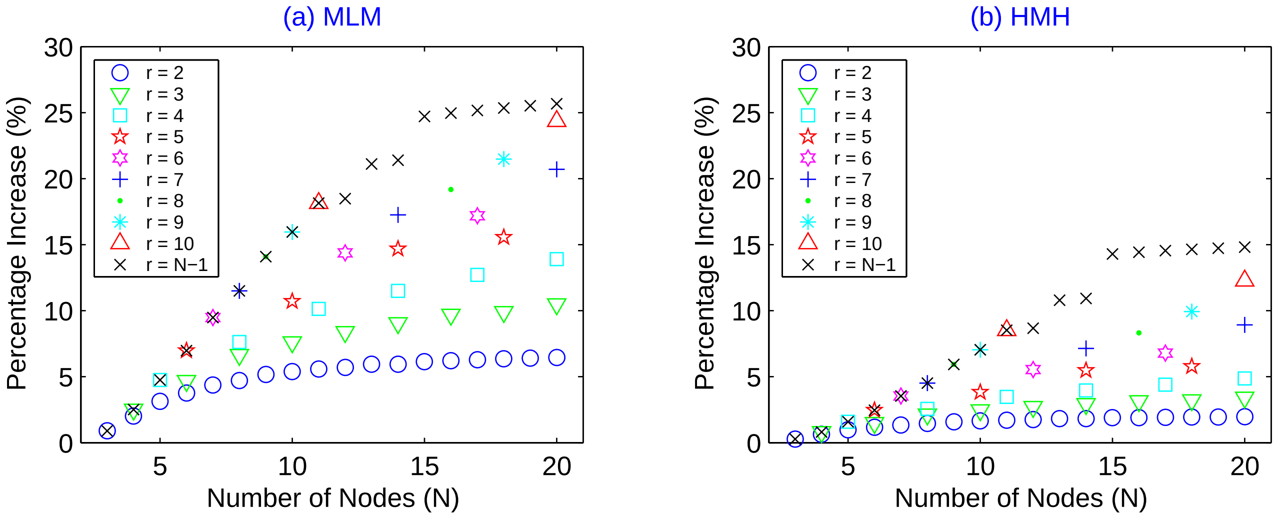


Figure A-76: Connected circulant constellations with uniform jump spaces: percentage increase in communications cost between custom antennas and uniform antennas for uniform all-to-all traffic.

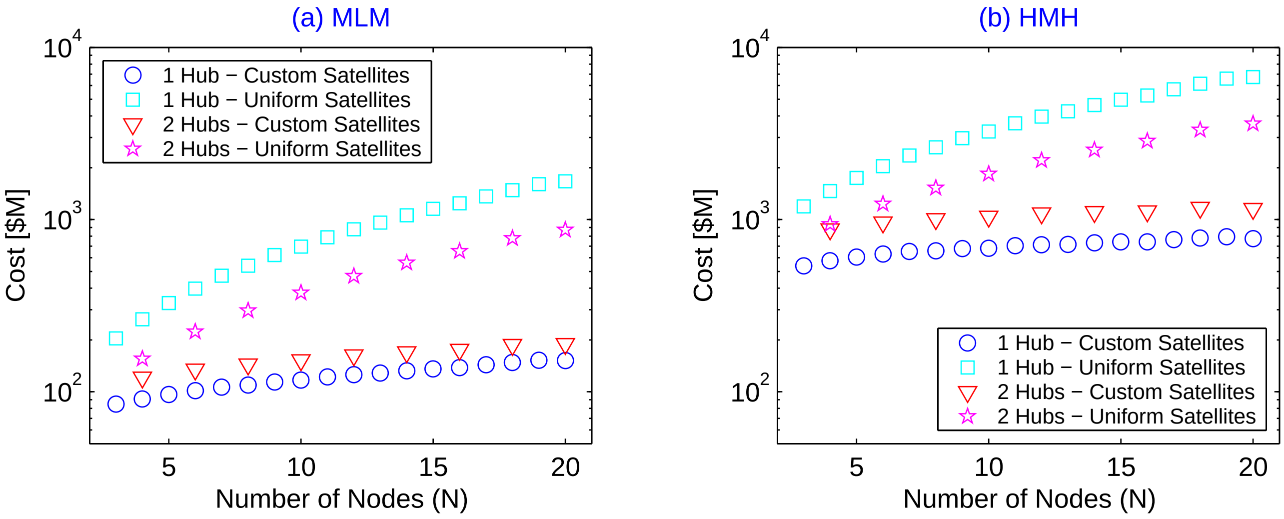
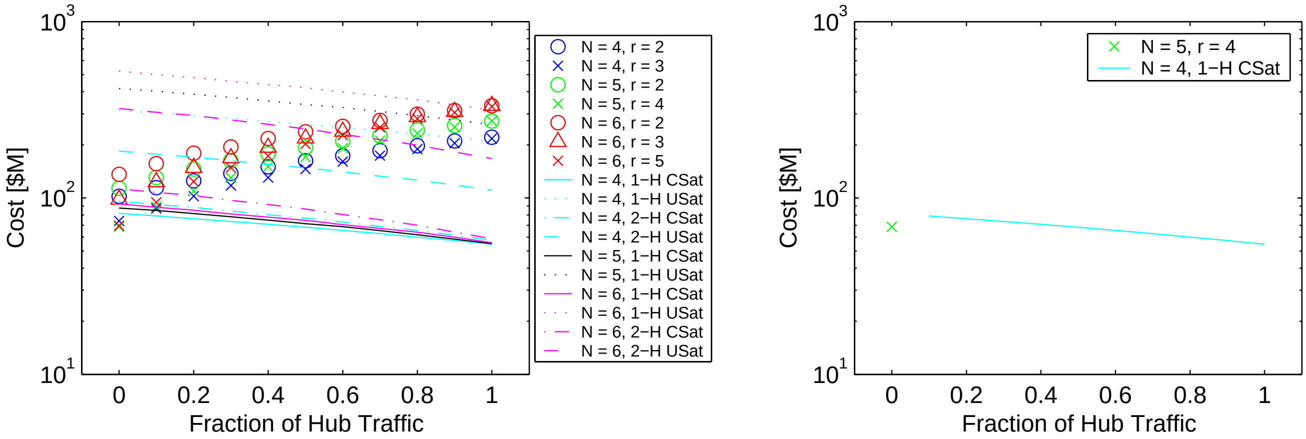


Figure A-77: Hub constellations: communications cost with linear switch.

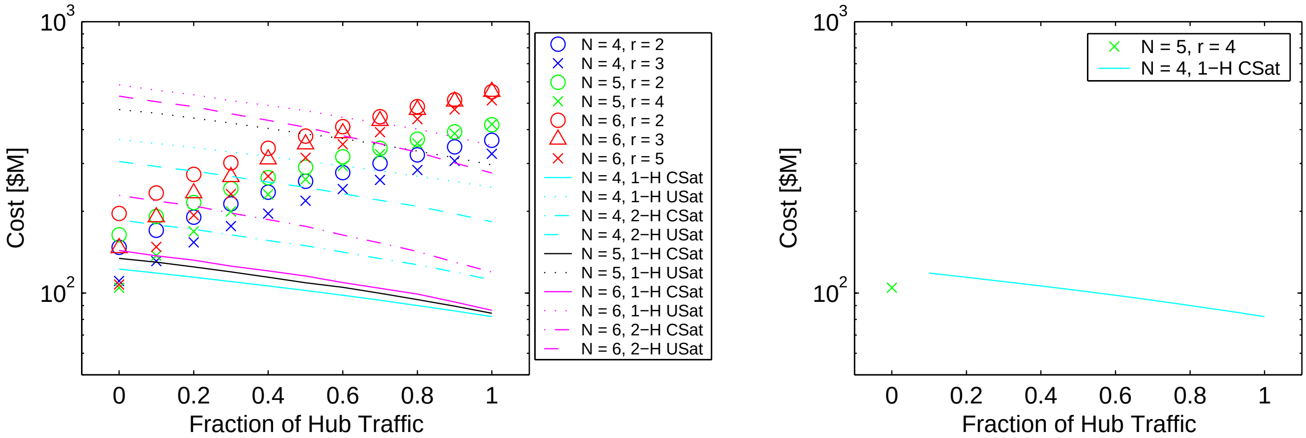
A.3.1 Small Constellations

As there is interest in studying small constellations of $4 \leq N \leq 6$ under mixed levels of traffic, the set of communications cost for all 27 permutations of cost cases using the linear switch is provided in Appendix A.3. The main observation that can be made from the results is that whenever there is any uniform all-to-one traffic, a hub constellation provides lower communications costs. That is, connected circulant constellations with uniform jump spaces should be chosen only when the traffic is 100% uniform all-to-all traffic. Lower communications costs are seen for hub constellations because the use of one large linear switch is less expensive than multiple smaller linear switches.



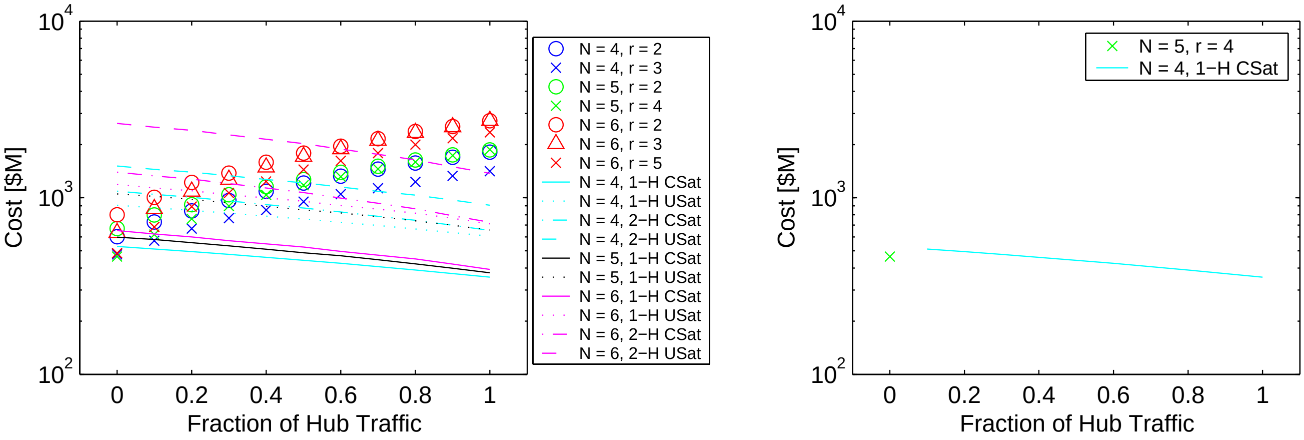
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-78: Communications costs LLL with mixed traffic and linear switches.



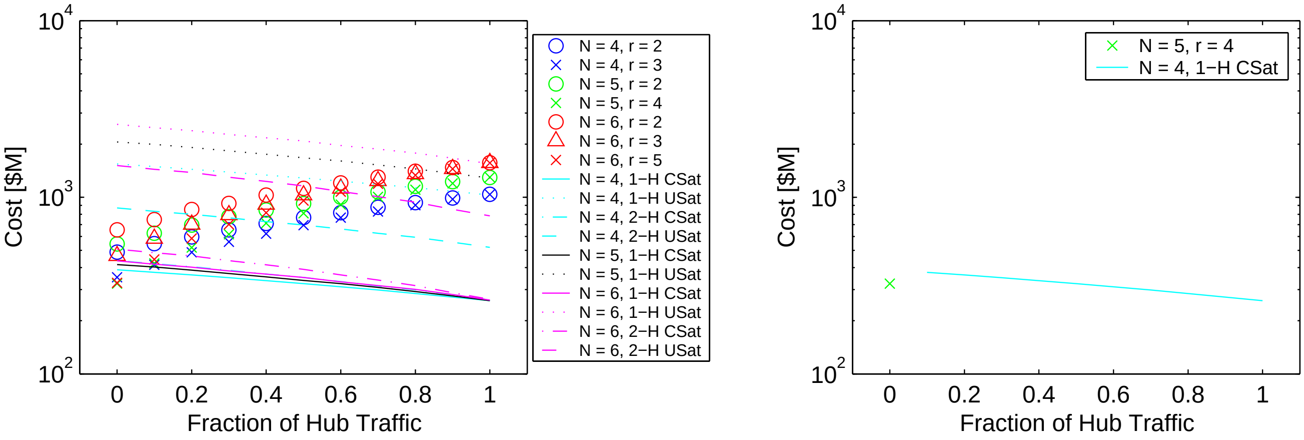
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-79: Communications costs LLM with mixed traffic and linear switches.



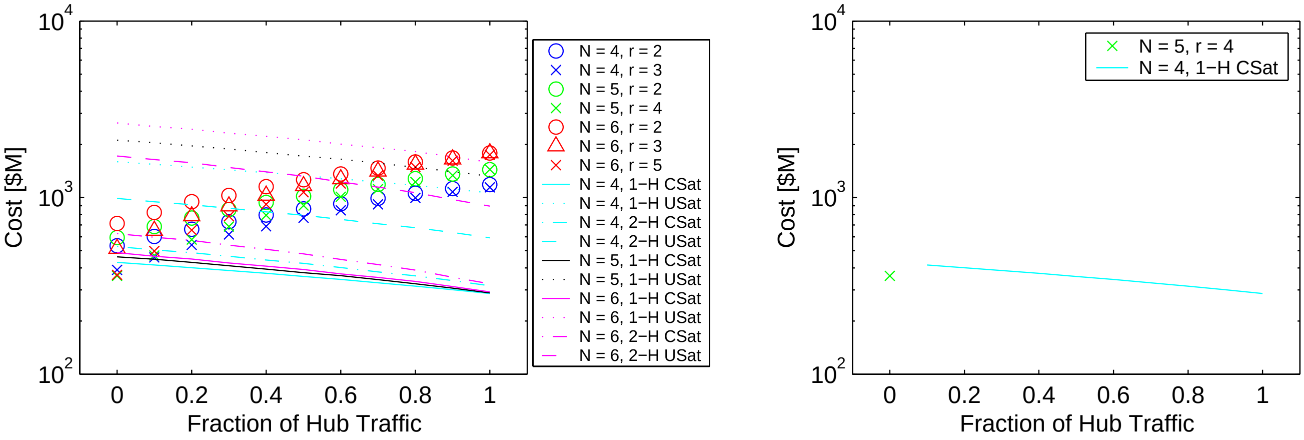
(a) Communications costs for N=4,5,6.
(b) Lowest communications costs for N=4,5,6.

Figure A-80: Communications costs LLH with mixed traffic and linear switches.



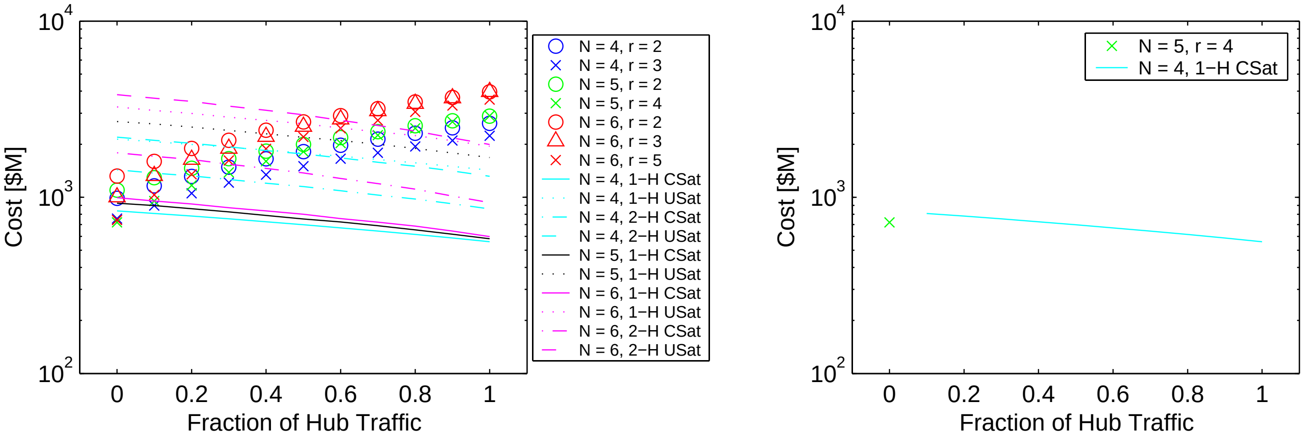
(a) Communications costs for N=4,5,6.
(b) Lowest communications costs for N=4,5,6.

Figure A-81: Communications costs LML with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.
(b) Lowest communications costs for N=4,5,6.

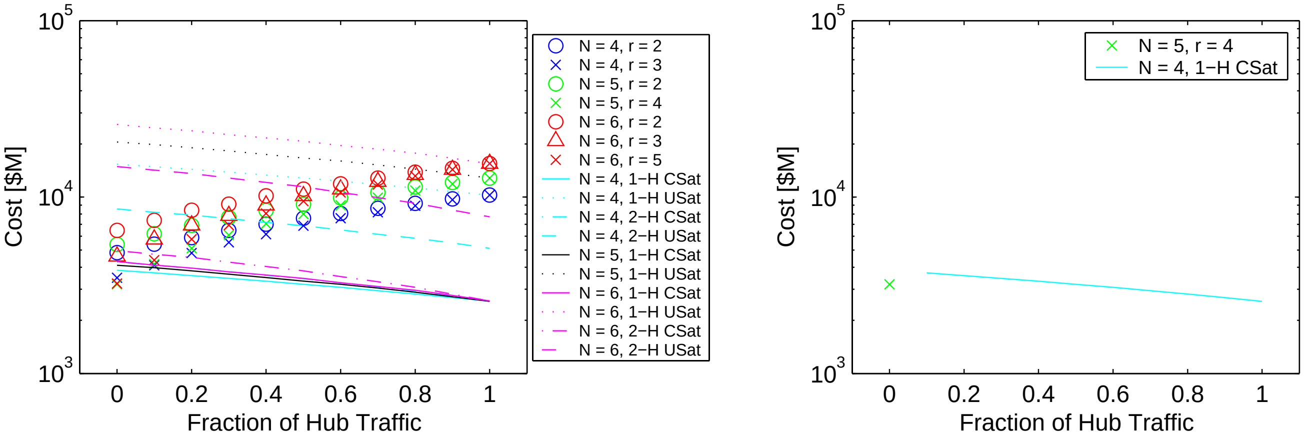
Figure A-82: Communications costs LMM with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

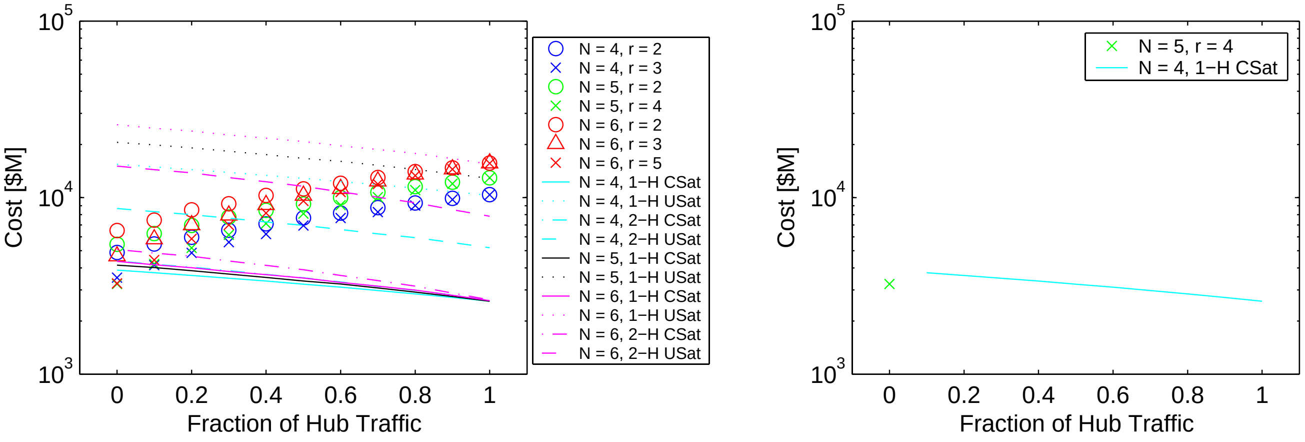
Figure A-83: Communications costs LMH with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

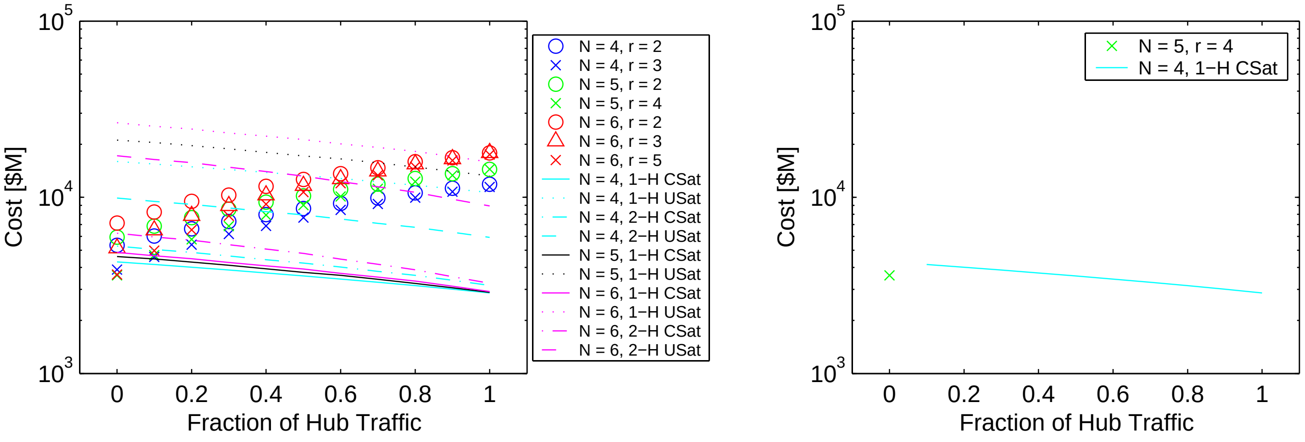
Figure A-84: Communications costs LHL with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

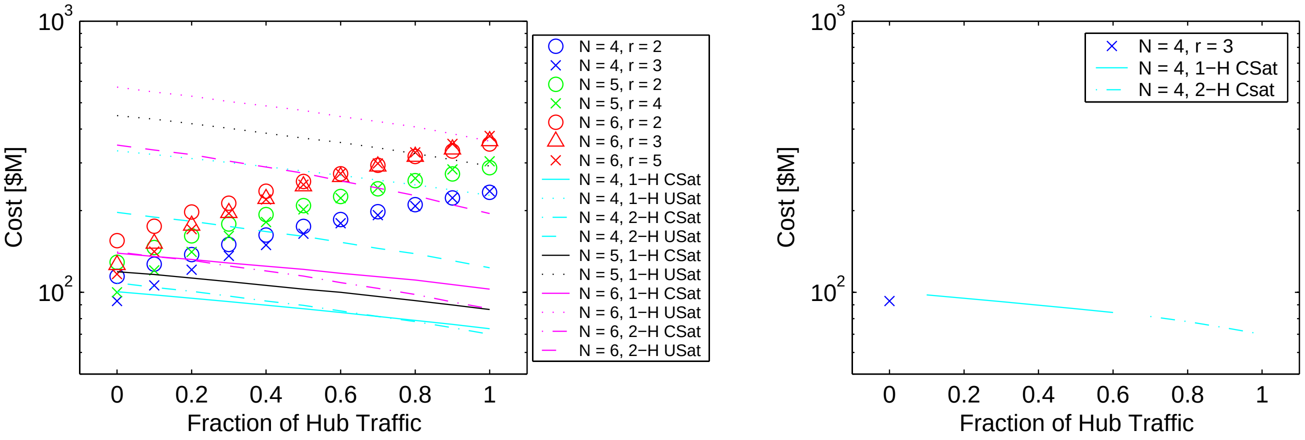
(b) Lowest communications costs for N=4,5,6.

Figure A-85: Communications costs LHM with mixed traffic and linear switches.



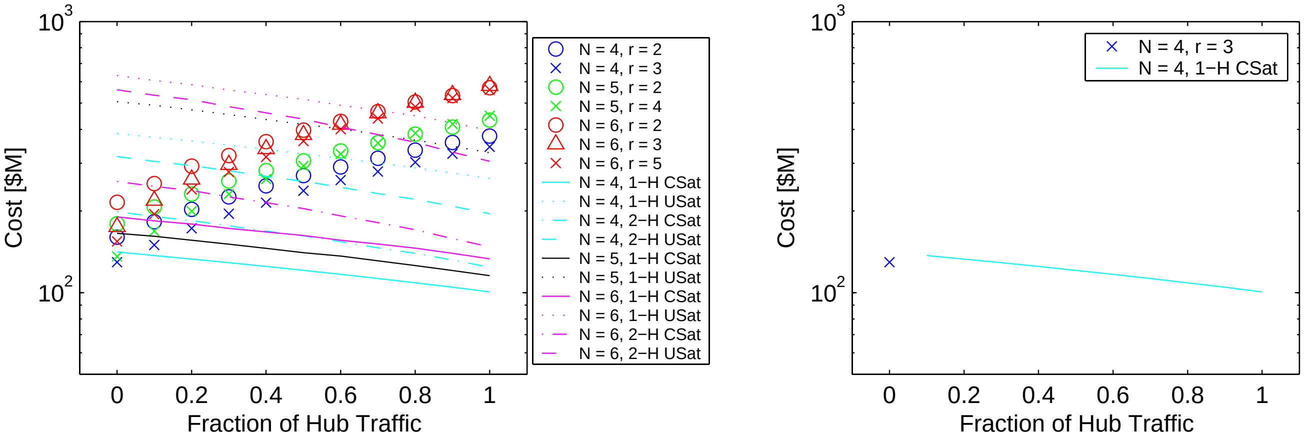
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-86: Communications costs LHH with mixed traffic and linear switches.



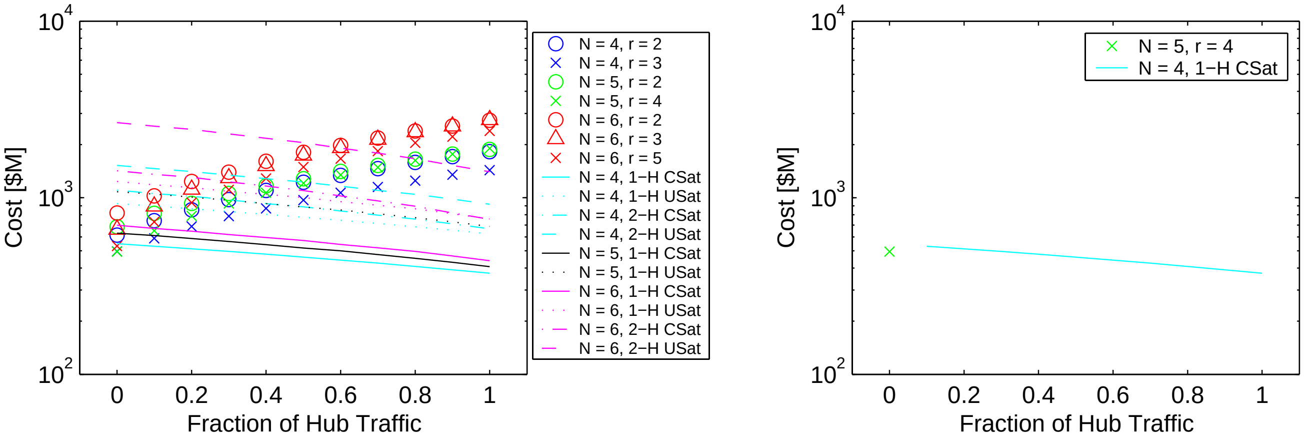
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-87: Communications costs MLL with mixed traffic and linear switches.



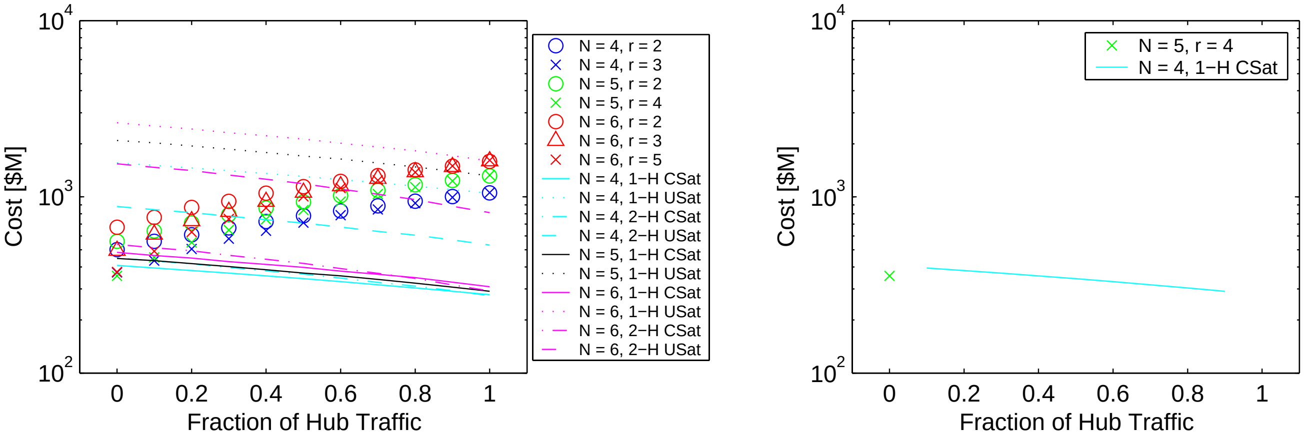
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-88: Communications costs MLM with mixed traffic and linear switches.



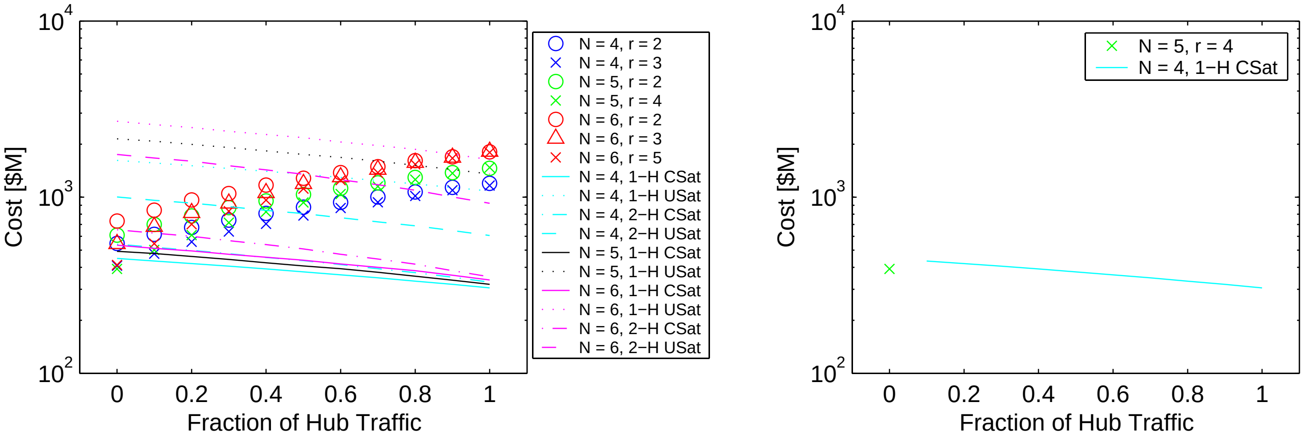
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-89: Communications costs MLH with mixed traffic and linear switches.



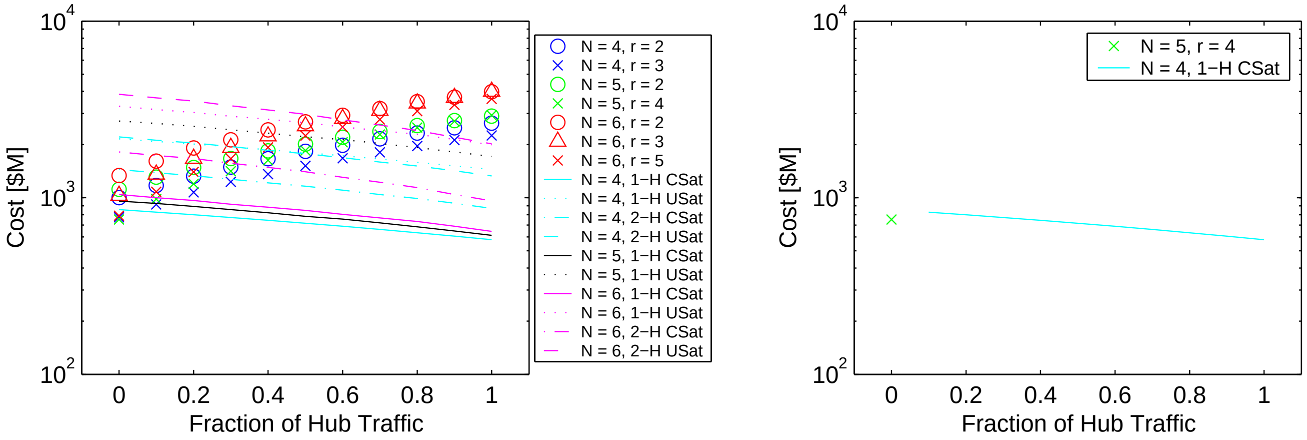
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-90: Communications costs MML with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

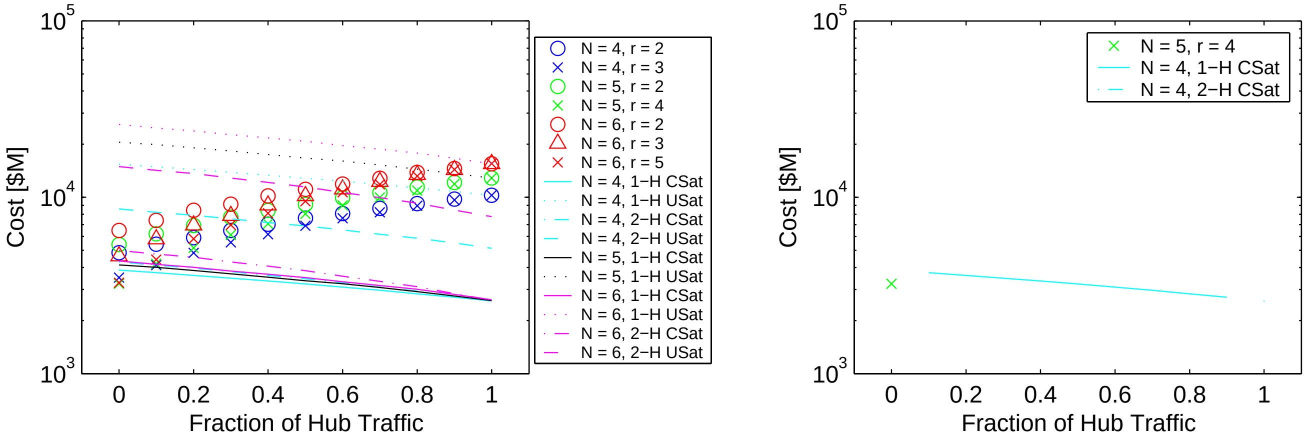
Figure A-91: Communications costs MMM with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

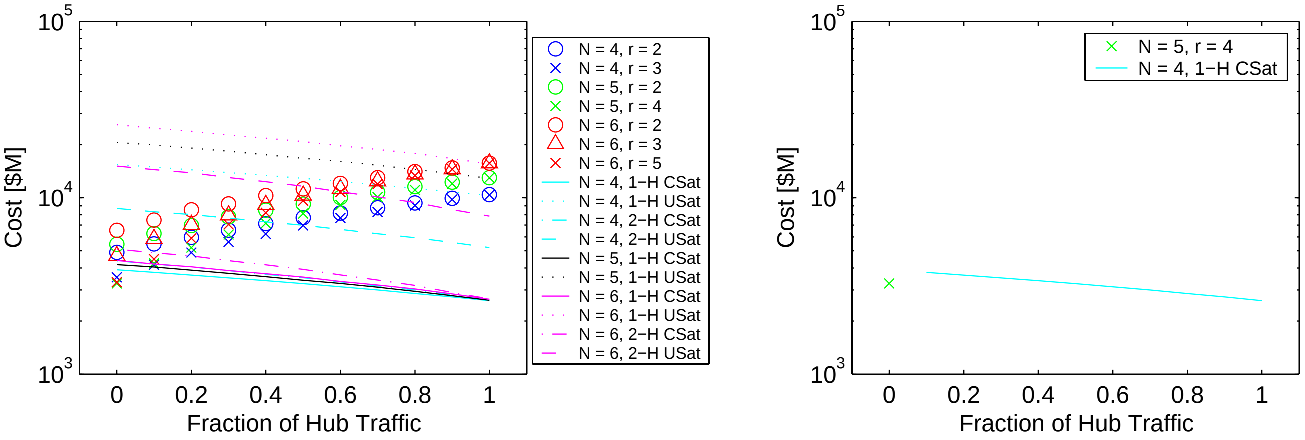
Figure A-92: Communications costs MMH with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

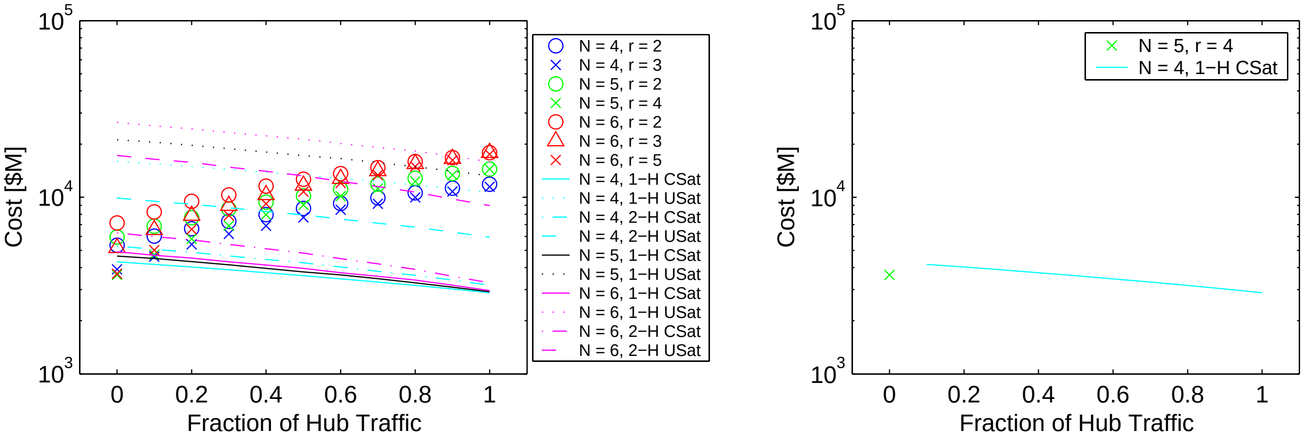
(b) Lowest communications costs for N=4,5,6.

Figure A-93: Communications costs MHL with mixed traffic and linear switches.



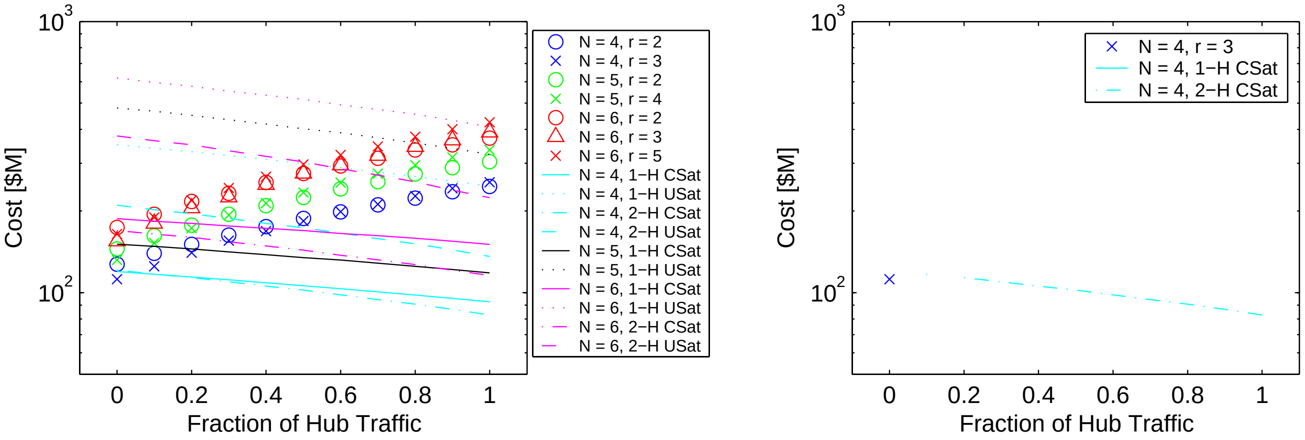
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-94: Communications costs MHM with mixed traffic and linear switches.



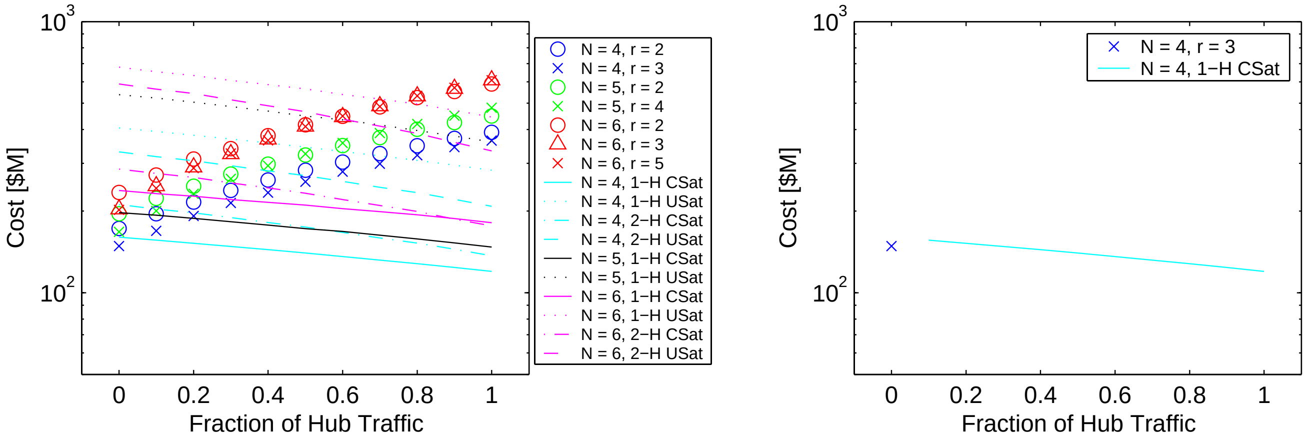
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-95: Communications costs MHH with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

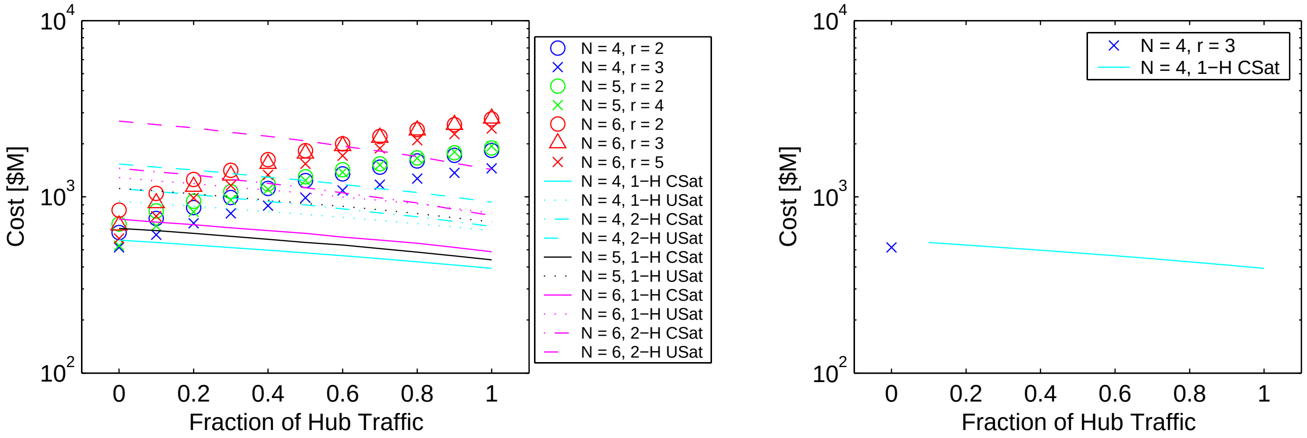
Figure A-96: Communications costs HLL with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

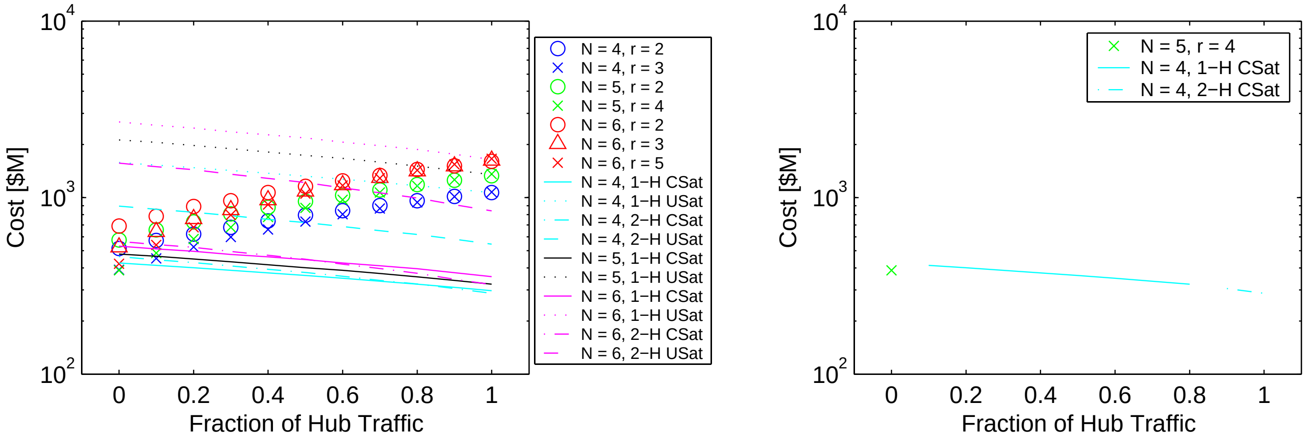
Figure A-97: Communications costs HLM with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

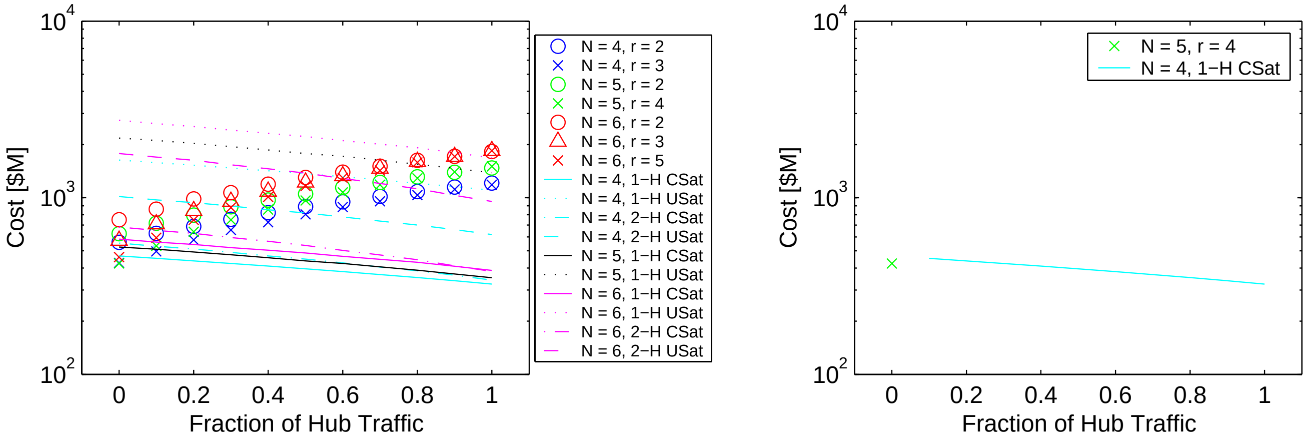
Figure A-98: Communications costs HLH with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

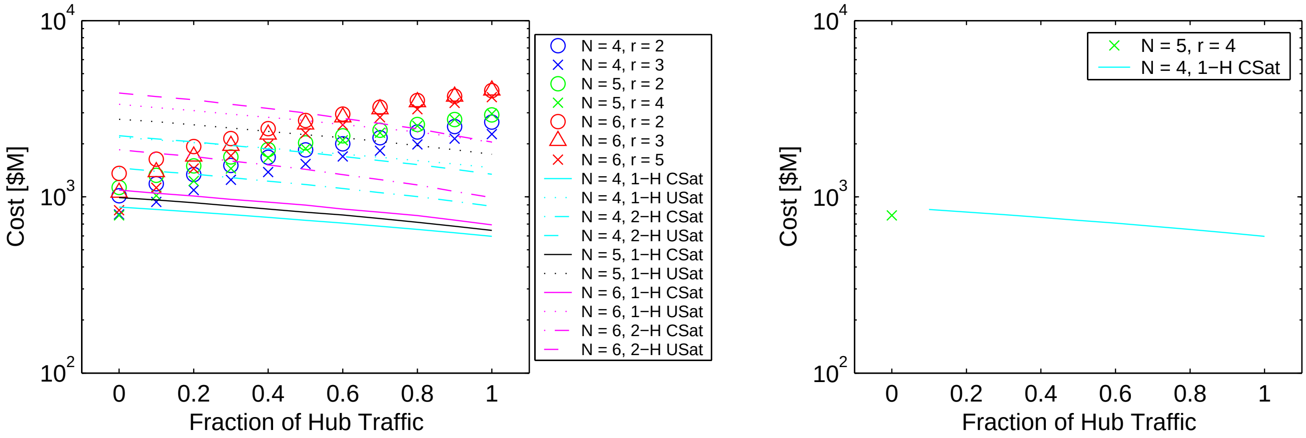
(b) Lowest communications costs for N=4,5,6.

Figure A-99: Communications costs HML with mixed traffic and linear switches.



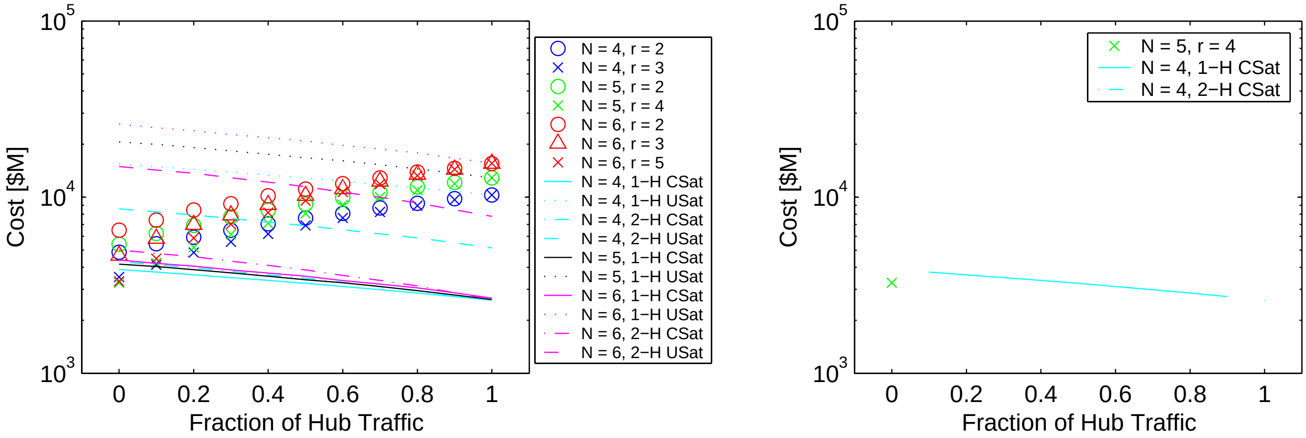
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-100: Communications costs HMM with mixed traffic and linear switches.



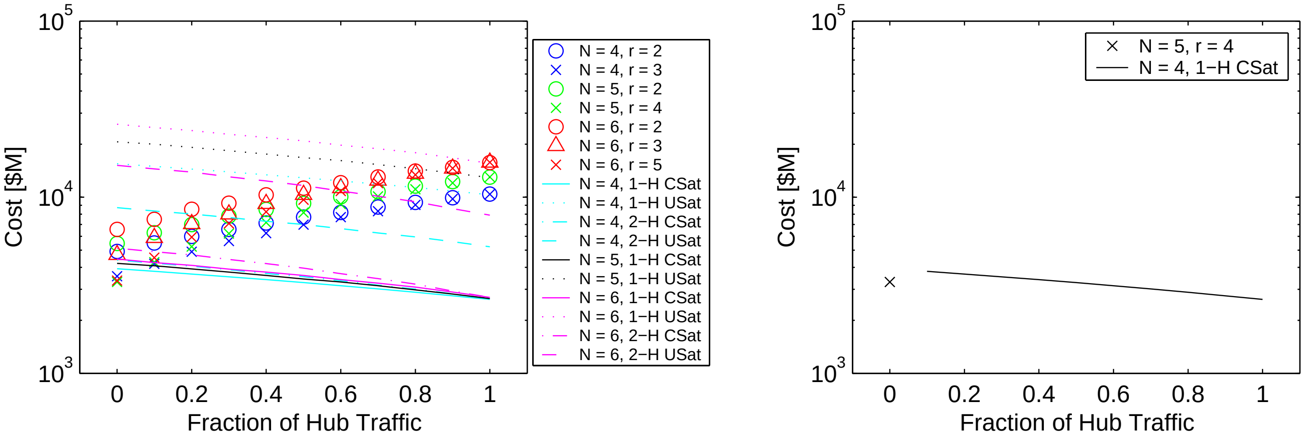
(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

Figure A-101: Communications costs HMH with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6. (b) Lowest communications costs for N=4,5,6.

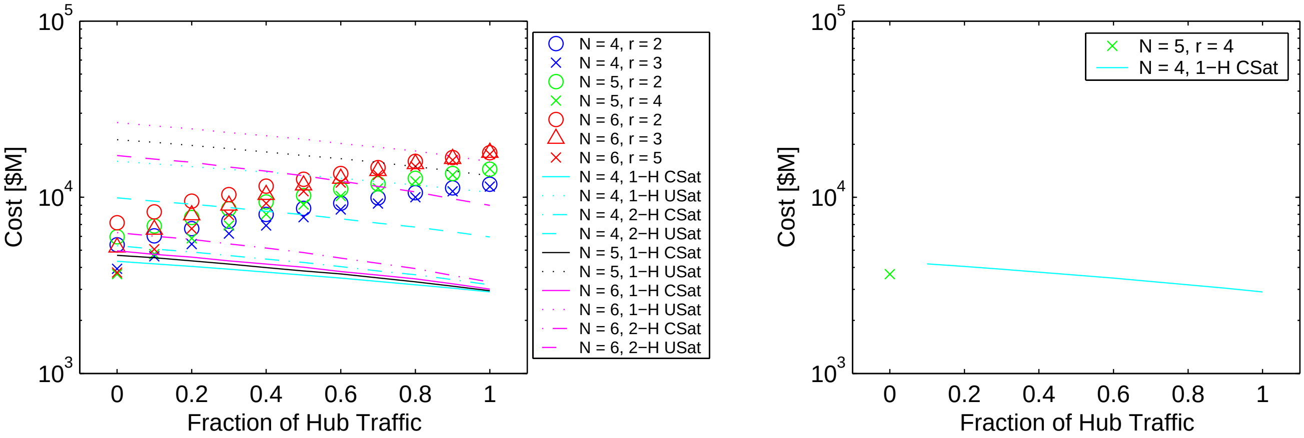
Figure A-102: Communications costs HHL with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

Figure A-103: Communications costs HHM with mixed traffic and linear switches.



(a) Communications costs for N=4,5,6.

(b) Lowest communications costs for N=4,5,6.

Figure A-104: Communications costs HHH with mixed traffic and linear switches.

Appendix B

SAR Satellites

	SEASAT-A	**SIR-A**	**SIR-B**
Country	United States	United States	United States
Platform	Satellite	Space shuttle	Space shuttle
Launch date	Jun-1978	Nov-1981	Oct-1984
Life time [days]	105	2.5	8.3
Frequency [GHz]	1.3 (L-band)	1.3 (L-band)	1.3 (L-band)
Polarization	HH	HH	HH
Orbit altitude [km]	795	260	224, 257, 360
Orbit inclination [deg]	108	38	51
Look angle [deg]	20	47	15-60
Swath width [km]	100	50	29-40
Antenna dimensions [m]	10.8 x 2.2	9.4 x 2.2	10.8 x 2.2
Pulse duration [μs]	33.4	30.4	30.4
Pulse bandwith [MHz]	19	6	12
Pulse repetition frequency [Hz]	1463-1640	1464-1824	1248-1824
Transmitted peak power [kW]	1	1	1.1
Data rate [Mb/s]	110 (5 b/sample)	Optical recording	30.4 (3-6 b/sample)

Source: [36] Giorgio Franceschetti and Riccardo Lanari, *Synthetic Aperture Radar Processing*, Electronic Engineering Systems Series, CRC Press LLC, Boca Raton, Florida, 1999, p. 5.

Table B.1: SAR Sensor Parameters for SEASAT-A, SIR-A and SIR-B.

	SIR-C (L-Band)	SIR-C (C-Band)	X-SAR
Country	United States	United States	Germany/Italy
Platform	Space shuttle	Space shuttle	Space shuttle
Launch date	Apr-1994	Apr-1994	Oct-1984
Life time [days]	11	11	11
Frequency [GHz]	1.3	5.3	9.6 (X-band)
Polarization	HH, HV, VH, VV	HH, HV, VH, VV	VV
Orbit altitude [km]	225	225	225
Orbit inclination [deg]	57	57	57
Look angle [deg]	20-55	20-55	20-55
Swath width [km]	15-90[a]	15-90[a]	15-60
Antenna dimensions [m]	12 x 2.9	12 x 0.7	12 x 0.4
Pulse duration [μs]	8.5, 33.2	8.5, 33.2	40
Pulse bandwith [MHz]	10, 20	10, 20	10, 20
Pulse repetition frequency [Hz]	1240-1736	1240-1736	1240-1736
Transmitted peak power [kW]	4.4	1.2	1.4
Data rate [Mb/s]	90 (4-8[b] b/sample)	90 (4-8[b] b/sample)	45 (4-6 b/sample, I/Q)

[a] In the experimental ScanSAR mode the sensor has been operated with a 225 km swath width.

[b] A block floating point quantization (BFPQ) can be applied.

Source: [36] Giorgio Franceschetti and Riccardo Lanari, *Synthetic Aperture Radar Processing*, Electronic Engineering Systems Series, CRC Press LLC, Boca Raton, Florida, 1999, p. 8.

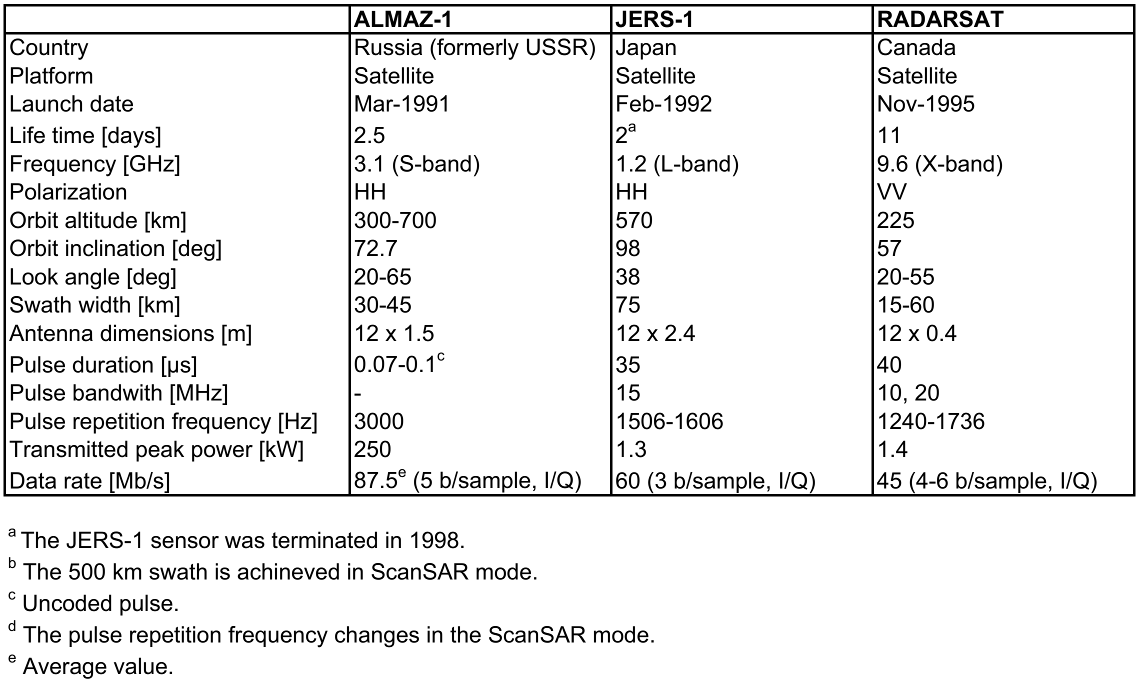
Table B.2: SAR Sensor Parameters for SIR-C and X-SAR.

	ERS-1, ERS-2
Country	European Union
Platform	Satellite
Launch date	Jul-1991, Apr-1995
Life time [days]	3[a]
Frequency [GHz]	5.3 (C-band)
Polarization	VV
Orbit altitude [km]	780
Orbit inclination [deg]	98.5
Look angle [deg]	23
Swath width [km]	100
Antenna dimensions [m]	10 x 1
Pulse duration [μs]	37.1
Pulse bandwith [MHz]	15.5
Pulse repetition frequency [Hz]	1640-1720
Transmitted peak power [kW]	4.8
Data rate [Mb/s]	105 (5 b/sample, I/Q)

[a] Both sensors are still operating.

Source: [36] Giorgio Franceschetti and Riccardo Lanari, *Synthetic Aperture Radar Processing*, Electronic Engineering Systems Series, CRC Press LLC, Boca Raton, Florida, 1999, p. 11.

Table B.3: SAR Sensor Parameters for ERS-1 and ERS-2.



	ALMAZ-1	**JERS-1**	**RADARSAT**
Country	Russia (formerly USSR)	Japan	Canada
Platform	Satellite	Satellite	Satellite
Launch date	Mar-1991	Feb-1992	Nov-1995
Life time [days]	2.5	2[a]	11
Frequency [GHz]	3.1 (S-band)	1.2 (L-band)	9.6 (X-band)
Polarization	HH	HH	VV
Orbit altitude [km]	300-700	570	225
Orbit inclination [deg]	72.7	98	57
Look angle [deg]	20-65	38	20-55
Swath width [km]	30-45	75	15-60
Antenna dimensions [m]	12 x 1.5	12 x 2.4	12 x 0.4
Pulse duration [µs]	0.07-0.1[c]	35	40
Pulse bandwith [MHz]	-	15	10, 20
Pulse repetition frequency [Hz]	3000	1506-1606	1240-1736
Transmitted peak power [kW]	250	1.3	1.4
Data rate [Mb/s]	87.5[e] (5 b/sample, I/Q)	60 (3 b/sample, I/Q)	45 (4-6 b/sample, I/Q)

[a] The JERS-1 sensor was terminated in 1998.

[b] The 500 km swath is achineved in ScanSAR mode.

[c] Uncoded pulse.

[d] The pulse repetition frequency changes in the ScanSAR mode.

[e] Average value.

Source: [36] Giorgio Franceschetti and Riccardo Lanari, *Synthetic Aperture Radar Processing*, Electronic Engineering Systems Series, CRC Press LLC, Boca Raton, Florida, 1999, p. 13.

Table B.4: SAR Sensor Parameters for ALMAZ-1, JERS-1 and RADARSAT.

Appendix C

List of Acronyms and Abbreviations

°	Degree
2-D	2-Dimensional
3-D	3-Dimensional
AC	Azimuth Compression
ADC	Analog-to-Digital Converter
AEHF	Advanced Extremely High Frequency
AFSATCOM	Air Force Satellite Communications
AFSPC/SMC	Air Force Space Command, Space and Missile Systems Center
AMFG	Azimuth Matched Filter Generation
APS	Advanced Polar Satellites
ARPANET	Advanced Research Projects Agency Network
ASIC	Application-Specific Integrated Circuit
AT&T	American Telephone & Telegraph Company
ATO	Air Tasking Order
AWS	Advanced Wideband System
bps	Bits per second
BWRC	Berkeley Wireless Research Center
C2	Command and Control

C4ISR	Command, Control, Communications, Computer, Intelligence, Surveillance, and Reconnaissance
CEC	Cooperative Engagement Capability
CMOS	Complementary Metal Oxide Semiconductor
CMV	Complex Vector Multiplication
Cf	Californium
Co	Cobolt
COMINT	Communications Intelligence
COTS	Commercial-Off-The-Shelf
CPU	Central Processing Unit
CSA	Chirp Scaling Algorithm
DARPA	Defense Advanced Research Projects Agency
dB	Decibel
DGPS	Differential GPS
DHSM	Distributed Hierarchical Storage Management
DIMM	Dual In-line Memory Module
DDR	Double Data Rate
DDR2	Double Data Rate 2
DoD	Department of Defense
DRAM	Dynamic Random Access Memory
DSCS	Defense Satellite Communications System
DSP	Digital Signal Processor
EHF	Extremely High Frequency
EISA	Extended Industry Standard Architecture
ELINT	Electronics Intelligence
EU	European Union
FCC	Federal Communications Commission
FFT	Fast Fourier Transform
FIA	Future Imagery Architecture
FISINT	Foreign Instrumentation Signals Intelligence
FLTSATCOM	Fleet Satellite Communications System
Gb	Gigabit

Gbps	Gigabits per second
GB	GigaByte
GCCS	Global Command and Control System
GCT	Global Communication Tree
GEO	Geosynchronous Earth Orbit
GOPS	Giga operations per second
GSO	Geostationary Orbit
GHz	GigaHertz
GMTI	Ground Moving Target Identification
GPP	General-Purpose Processor
GPS	Global Positioning System
GSPS	Giga samples per second
GTO	Geosynchronous Transfer orbit
HRTI	High-Resolution Terrain Information
Hz	Hertz
IC	Intelligence Community
IETF	Internet Engineering Task Force
IFFT	Inverse FFT
IMINT	Imagery Intelligence
in	Inch
INMARSAT	International Maritime Satellite Organization
INTELSAT	International Telecommunications Satellite Consortium
I/O	Input/Output
IOSA	Integrated Overhead SIGINT Architecture
IP	Internet Protocol
IR	Infrared
ISA	Industry Standard Architecture
iSCSI	Internet Small Computer System Interface
ISR	Intelligence, Surveillance, and Reconnaissance
J	Joule
JPL	Jet Propulsion Laboratory
kbps	Kilobits per second

kg	Kilogram
km	Kilometer
krad	Kilorad
LASINT	Laser Intelligence
LEASAT	Leased Satellite
LEO	Low Earth Orbit
LES	Lincoln Experimental Satellite
LVCMOS	Low Voltage Complimentary Metal Oxide Semiconductor
m	Meter
mm	Millimeter
Mbps	Megabits per second
MDA	MacDonald Dettwiler and Associates
MEMS	Micro-Electro-Mechanical Systems
MEO	Medium Earth Orbit
MFLOP	Million floating point operations per second
MGT	Multi-Gigabit Transceiver
MHz	MegaHertz
MIT	Massachusetts Institute of Technology
MILSTAR	Military Strategic and Tactical Relay
MIMD	Multiple Instruction Multiple Data
MIPS	Million Instructions per Second
MISD	Multiple Instruction Single Data
MOPS	Mega operations per second
MOSFET	Metal Oxide Silicon Field-Effect Transistor
msec	Millisecond
MSPS	Mega samples per second
MUOS	Mobile Users Objective System
NAS	Network Attached Storage
NASA	National Aeronautics and Space Administration
NATO	North Atlantic Treaty Organization
NAVSTAR	Navigation System with Timing and Ranging
NIMA	National Imagery and Mapping Agency

NPOESS	National Polar-orbiting Operational Environmental Satellite System
NRE	Non-Recurring Engineering
NRO	National Reconnaissance Office
NSF	National Science Foundation
NSFNET	National Science Foundation Network
PCI	Peripheral Component Interconnect
PFA	Polar Format Algorithm
PPS	Precision Positioning Service
PRF	Pulse Repetition Frequency
PSTN	Public Switch Telephone Network
QoS	Quality of Service
R&D	Research and Development
Rad	Radiation absorbed dose
RADAR	Radio Detection and Ranging
RADINT	Radar Intelligence
RAM	Random Access Memory
RAR	Real Aperture Radar
RC	Range compression
RCMC	Range Cell Migration Correction
RF	Radio Frequency
RMA	Range Migration Algorithm
s, sec	Second
SAN	Storage Area Network
SAR	Synthetic Aperture Radar
SBIR	Space-Based Infrared System
SBR	Space-Based Radar
SBSS	Space-Based Space Surveillance System
SEB	Single Event Burnout
SEL	Single Event Latchup
SEU	Single Event Upset
SHF	Super High Frequency
Si	Silicon

SIGINT	Signals Intelligence
SIMD	Single Instruction Multiple Data
SIR-C/X-SAR	Spaceborne Shuttle Imaging Radar-C/X-Band SAR
SISD	Single Instruction Single Data
SNR	Signal-to-Noise Ratio
SPS	Standard Positing Service
SRC	Secondary Range Compression
STAHNET	Strategic Highway Corridor Network
STSS	Space Tracking and Surveillance System
TB	Terabytes
Tbps	Terabits per second
TCA	Transformational Communications Architecture
TCS	Transformational Communications Satellites
TDRS	Tracking and Data Relay Satellite
TELINT	Telemetry Intelligence
TFLOPS	Tera floating point operations per second
THz	Tera Hertz
TID	Total Ionizing Dose
TOPS	Tera operations per second
UHF	Ultra High Frequency
U.S.	United States
VESA	Video Electronics Standards Association
VL-Bus	VESA Local-Bus
VLBI	Very Long Baseline Interferometry
VLIW	Very Long Instruction Word
VLSI	Very Large Scale Integrated
VME	VersaModule Eurocard
VSOP2	VLBI Space Observatory Program 2
W	Watt
WDM	Wavelength Division Multiplexing
WGS	Wideband Gapfiller Satellite

References

[1] Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows*. Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1993.

[2] Péricles Gasparini Alves, editor. *Evolving Trends in the Dual Use of Satellites*. United Nations, New York, 1993.

[3] Ray Andraka. High performance digital down-converters for FPGAs. *Xilinx Xcell Journal*, pages 48–51, 4th quarter, 2000.

[4] Army space reference text. http://www-tradoc.army.mil/dcscd/spaceweb/chap07b.htm.

[5] AstroExpo. COTS in space. http://www.astroexpo.com//news/articlesdetail.asp?id=233 &csort=1, March 23, 2004.

[6] AstroExpo. FPGAs designed for space. http://www.astroexpo.com//news/articlesdetail.asp? id=234&csort=1, March 23, 2004.

[7] Astroexpo. Radiation hardened electronics marketplace - the story of the decline. http://www.astroexpo.com//news/articlesdetail.asp?ID=235, March 23, 2004.

[8] A. Averbunch, E. Gabber, B. Gordissky, and Y. Medan. A parallel FFT on an MIMD machine. *Parallel Computing*, 15(1-3):61–74, September 1990.

[9] Walter S. Baer. Government investment in telecommunications infrastructure. In National Research Council, editor, *The Changing Nature of Telecommunications/Information Infrastructure*, pages 179–194. National Academy of Sciences, Washington, DC, 1995.

[10] David Baker, editor. *Jane's Space Directory*. Jane's Information Group Limited, Surrey, UK, seventeenth edition, 2001.

[11] Jason H. Bau. Topologies for satellite constellations in a cross-linked space backbone network. Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, May 2002.

[12] Joshua Boehm, Craig Baker, Stanley Chan, and Mel Sakazaki. A history of united states national security space management and organization. http://www.fas.org/spp/eprint/article03.html, 2000.

[13] Boeing satellite systems: MILSTAR II. http://www.boeing.com/defense-space/space/bss/factsheets/government/milstar_ii/milstar_ii.html.

[14] Boeing 702 Fleet. http://www.boeing.com/defense-space/space/bss/factsheets/702/702fleet.html.

[15] Boeing Satellite Systems 702 Satellite, USA. http://www.aerospace-technology.com/projects/boeing702/.

[16] Kiran Bondalapati and Viktor K. Prasanna. Reconfigurable computing systems. *Proceedings of the IEEE*, 90:1201–1217, July 2002.

[17] Fred Buckley and Frank Harary. *Distance in Graphs*. Addison-Wesley Publishing Company, Redwood City, California, 1990.

[18] John R. Budenske, Kevin S. Millikin, Jordan C. Bonney, Ranga S. Ramanujan, and O. Scott Sands. Space network architecture technologies. In *IEEE Aerospace Conference Proceedings*, volume 3, pages 1061–1069, Big Sky, Montana, March 2002.

[19] Walter G. Carrara, Ron S. Goodman, and Ronald M. Majewski. *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*. Artech House, Inc., Boston, Massachusetts, 1995.

[20] Vincent W.S. Chan. Optical satellite networks. *Journal of Lightwave Technology*, 21:2811–2817, November 2003.

[21] Chen Chang, John Wawrzynek, and Robert W. Brodersen. Evaluation of multi-FPGA systems for high-performance DSP applications. Manuscript submitted to International Symposium on Computer Architecture 2005, December 2004.

[22] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Algorithms*. The MIT Press, Cambridge, Massachusetts, 1996.

[23] Wendell Cox and Jean Love. 40 years of the US interstate highway system: An analysis - the best investment a nation ever made. http://www.publicpurpose.com/freeway1.htm, June 1996.

[24] Dorothy E. Denning. *Inforpmation Warfare and Security*. Addison Wesley Longman, Inc, Reading, Massachusetts, 1999.

[25] Andreas Diekmann and Hans-Peter Richarz. Future role and significance of space activities in reflection of global social, technological and economic trends. *Acta Astronautica*, 45(11):697–703, 1999.

[26] Ralph Duncan. A survey of parallel computer architectures. *Computer*, pages 5–16, February 1990.

[27] David D. Ehrhard. Standing in the strategic bandwidth gap: A view of military communications in 2012. http://www.airpower.maxwell.af.mil/airchronicles/cc/ehrhard.html, March 16, 2004.

[28] Charles Elachi. *Spaceborne Radar Remote Sensing: Applications and Techniques*. Institute of Electrical and Electronics Engineers, New York, New York, 1988.

[29] Leon Erlanger. Distributed computing: An introduction. http://www.extremetech.com/article2/0%2C1558%2C11769%2C00.asp, April 4, 2002.

[30] B. G. Evans, editor. *Satellite Communications Systems*. Institution of Electrical Engineers, London, The United Kingdom, third edition, 1999.

[31] Jennifer Eyre. The digital signal processor derby. *IEEE Spectrum*, 38(6):62–68, June 2001.

[32] FFTW. http://www.fftw.org/.

[33] M.J. Flynn. Very high spped computing systems. *Proc. IEEE*, 54(12):1901–1909, December 1966.

[34] Air Force. Milstar satellite communications system. http://www.af.mil/factsheets/factsheet.asp?fsID=118.

[35] Peter Fortescue and John Stark, editors. *Spacecraft Systems Engineering.* John Wiley & Sons Ltd., West Sussex, England, second edition, 1995.

[36] Giorgio Franceschetti and Riccardo Lanari. *Synthetic Aperture Radar Processing.* CRC Press LLC, Boca Raton, Florida, 1999.

[37] Robert M. Gagliardi and Sherman Karp. *Optical Communications.* John Wiley & Sons, Inc., New York, New York, 1995.

[38] Galileo: European satellite navigation system. http://europa.eu.int/.

[39] Jack Ganssle and Michael Barr. *Embedded Systems Dictionary.* CMP Books, San Francisco, California, 2003.

[40] Gary D. Gordon and Walter L. Morgan. *Principles of Communications Satellites.* John Wiley & Sons, Inc., New York, New York, 1993.

[41] Stephen H. Gunther, Frank Binns, Douglas M. Carmean, and Jonathan C. Hall. Managing the impact of increasing microprocessor power consumption. *Intel Technology Journal*, pages 1–9, 1st quarter, 2001.

[42] Anshul Gupta and Vipin Kumar. The scalability of FFT on parallel computers. *IEEE Transactions on Parallel and Distributed Systems*, 4(8):922–932, August 1993.

[43] Peter L. Hays. Space and the military. In Eligar Sadeh, editor, *Space Politics and Policy: An Evolutionary Perspective*, pages 335–369. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[44] John L. Hennessy and David A. Patterson. *Computer Organization & Design: The Hardware/Software Interface.* Morgan Kaufmann Publishers, Inc., San Francisco, California, second edition, 1998.

[45] James W. Howard, Jr., Martin A. Carts, Ronald Stattel, Charles E. Rogers, Timothy L. Irwin, Curtis Dunsmore, J. Anthony Sciarini, and Kenneth A. LaBel. Total dose and single event effects testing of the Intel Pentium III (P3) and AMD K7 microprocessors.

In *IEEE Radiation Effects Data Workshop*, pages 38–47, Vancouver, BC, Canada, July 2001.

[46] Roy Hoffman. *Data Compression in Digital Systems.* Chapman & Hill, New York, New York, 1997.

[47] Steven J. Isakowitz, Joseph P. Hopkins Jr., and Joshua B. Hopkins, editors. *International Reference Guide to Space Launch Systems.* American Institute of Aeronautics and Astronautics, Reston, Virgina, third edition, 1999.

[48] Dana J. Johnson, Scott Pace, and C. Bryan Gabbard. *Space: Emerging Options for National Power.* RAND, Santa Monica, California, 1998.

[49] Joan Johnson-Freese and Roger Handberg. *Space, The Dormant Frontier.* Praeger Publishers, Westport, Connecticut, 1997.

[50] Brian Kahin and Bruce McConnell. Towards a public metanetwork: Interconnection, leveraging, and privatization of government-funded networks in the united states. http://www.columbia.edu/dlc/wp/citi/citi488.html.

[51] Sherman Karp and Barry K. Gilbert. Digital system design in the presence of single event upsets. *IEEE Transactions on Aerospace and Electronic Systems*, 29(2):310–316, April 1993.

[52] Morris Katzman, editor. *Laser Satellite Communications.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

[53] Stephen G. Lambert and William L. Casey. *Laser Communications in Space.* Artech House, Inc., Norwood, Massachusetts, 1995.

[54] Claudia Leopold. *Parallel and Distributed Computing: A Survey of Models, Paradigms, and Approaches.* John Wiley & Sons, Inc., New York, New York, 2001.

[55] John Lesko, Phillip Nicolai, and Michael Steve. *Data Compression: The Complete Reference.* Springer-Verlag New York, Inc., New York, New York, second edition, 2000.

[56] Michael Manzke. Parellel processing architectures. http://www.cs.tcd.ie/Michael.Manzke/3ba5/3ba5_third_lecture.pdf.

[57] JPL The Mission and Spacecraft Library. Intelsat. http://msl.jpl.nasa.gov/Programs/intelsat.html.

[58] Keith D. McDonald. Technology, implementation and policy issues for the modernisation of GPS and its role in a GNSS. *The Journal of Navigation*, 51(3):281–293, September 1998.

[59] Hamish Meikle. *Modern Radar Systems*. Artech House, Inc., Norwood, Massachusetts, 2001.

[60] Peter G. Meisl, Mabo R. Ito, and Ian G. Cumming. Parallel processors for synthetic aperture radar imaging. In *Proceedings of The 1996 International Conference on Parallel Processing*, volume 2, pages 124–131, August 1996.

[61] Peter G. Meisl, Mabo R. Ito, and Ian G. Cumming. Parallel synthetic aperture radar processing on workstation networks. In *Proceedings of The 10th International Parallel Processing Symposium*, pages 716–723, April 1996.

[62] Tony Merna and Cyrus Njiru. *Financing Infrastructure Projects*. Thomas Terlford Publishing, London, England, 2002.

[63] Craig Miller, David G. Payne, Thanh N. Phung, Herb Siegel, and Roy Williams. Parallel processing of spaceborne imaging radar data. In *Proceedings of The 1995 ACM/IEEE Conference on Supercomputing*, San Diego, California, December 1995.

[64] P. Misra and P. Enge. *Global Positioning System: Signals, Measurements, and Performance*. Ganga-Jamuna Press, Lincoln, Massachusetts, 2001.

[65] A brief history of NASA. http://www.hq.nasa.gov/office/pao/History/factsheet.htm.

[66] John Parkinson. Take a hybrid approach to component obsolescence. *COTS Journal*, pages 20–27, January 2001.

[67] Processors heat development. http://www.pcsilent.de/en/tips/cpu.asp.

[68] Larry Press. Seeding networks: The federal role. *Communications of the ACM*, 39(10):11–18, October 1996.

[69] Andrew L. Quiat. Financing infrastructure for follow-on space business development. *Acta Astronautica*, 41(4-10):707–721, 1997.

[70] Rajiv Ramaswami and Kumar Sivarajan. *Optical Networks: A Practical Perspective*. Morgan Kaufmann Publishers, Inc., San Francisco, California, second edition, 2001.

[71] King Space Research. Worldwide military and intelligence systems in the pipeline. http://www.astroexpo.com/News/articlesdetail.asp?id=341&csort=1, November 30, 2004.

[72] Jeffrey T. Richelson. *The U.S. Intelligence Community*. Westview Press, Boulder, Colorado, 1999.

[73] William C. Ripley. A building-block approach to COTS technology: Adapting the configuration of a basic COTS system helps to meet specific requirements. *NASA Tech Briefs*, 28(3):40–41, March 2004.

[74] David Salomon. *Data Compression: The Complete Reference*. Springer-Verlag New York, Inc., New York, New York, second edition, 2000.

[75] Michiel Schwarz and Paul Stares, editors. *The Exploitation of Space: Policy trends in the military and commercial uses of outer space*. Butterworth & Co (Publishers) Ltd., London, United Kingdom, 1985.

[76] SATA technology. http://www.serialata.org/satatechnology.asp.

[77] James Shoemaker. Orbital express: A better way of doing business in space. Presentation Slides, 2001.

[78] Deszõ Sima, Terence Fountain, and Péter Kacsuk. *Advanced computer architectures: a design space approach*. Addison-Wesley, Harlow, England, 1997.

[79] Space based radar (SBR). http://www.losangeles.af.mil/smc/pa/fact_sheets/sbr.htm.

[80] Space based radar concept development. http://www.cbd-net.com/index.php/search/show/459180.

[81] J. C. Springett. Increasing space very long baseline interfermoetry sensitivity through higher data rates. The interplanetary network progress report 42-149, Jet Propulsion Laboratory, California Institute of Technology, May 15, 2002.

[82] Paul Stoneman. *The Economic Analysis of Technology Policy.* Oxford University Press, New York, 1987.

[83] Andrew S. Tanenbaum. *Computer Networks.* Prentice-Hall, Inc., Upper Saddle River, New Jersey, third edition, 1996.

[84] BEE2 Team. Building BEE2: a case for high-end reconfigurable computer (HERC). presentation.

[85] Paul Teske. When the public goes private. In Eli Noam and Aine Níshúilleabháin, editors, *Private Networks Public Objectives*, pages 83–92. Elsevier Science B.V., Amsterdam, The Netherlands, 1996.

[86] Sanjay Thatte and John Blaine. How to manage power consumption in advanced FPGAs. *Xcell Journal Online*, 2002. http://www.xilinx.com/publications/xcellonline/partners/xc_synplicity44.htm.

[87] Alan C. Tribble. *The Space Environment: Implications for Spacecraft Design.* Princeton University Press, Princeton, New Jersey, second edition, 2003.

[88] Voltaire. Voltaire ISR 9288 InfiniBand switch router datasheet. http://www.voltaire.com/.

[89] Robert H. Walden. Analog-to-digital converter survey and analysis. *IEEE Journal on Selected Areas in Communications*, 17(4):539–550, April 1999.

[90] Robert H. Walden. Analog-to-digital converter survey and analysis. Presentation Slides, 16 July 1999.

[91] Donald M. Waltz. *On-Orbit Servicing of Space Systems.* Krieger Publishing Company, Malabar, Florida, 1993.

[92] James R. Wertz and Wiley J. Larson. *Space Mission Analysis and Design.* Microcosm Press & Kluwer Academic Publishers, El Segundo, California & Dordrecht, The Netherlands, third edition, 1999.

[93] Xilinx. Virtex-II Pro and Virtex-II Pro X platform FPGAs: Complete data sheet, 2004.

[94] Xiaobo Yang, Hui Long, and Yiming Pi. Parallel implementation of SAR imaging on DAWNING3000. In *Proceedings of SPIE Algorithms for Synthetic Aperture Radar Imagery X*, volume 5095, Orlando, Florida, April 2003.

[95] Steve Zack and Suhel Dhanani. DSP co-processing in FPGAs: Embedding high-performance, low-cost DSP functions. *Xilinx White Paper: Spartan-3 FPGAs*, WP212 (v1.0):1–8, March 18, 2004.

[96] Rong Zhang, Xian-Ci Xiao, and Heng-Ming Tai. An efficient digital down conversion method for multiple wideband signals. In *The 2002 45th Midwest Symposium on Circuits and Systems*, volume 3, pages III–583–III–586, Tulsa, Oklahoma, August 2002.