Medusa: the design and development of a TCP/IP communications nexus for the MOOS Portable Ocean Observatory Program

Timothy Meese, MBARI
Wayne Radochonski, MBARI
Mark Chaffey, MBARI

Monterey Bay Aquarium Research Institute
7700 Sandholdt Road
Moss Landing, California 95039 USA

***Abstract*- One of the key capabilities envisioned of MOOS (MBARI Ocean Observing System) is to provide both power and communications to oceanographic instrumentation in the benthic environment. With many technical elements, such as low-power instrument control 'nodes' and EOM (electro optical mechanical) cabling being proven at sea, the need to manage TCP/IP communications over a fiber optic link to the benthic environment became apparent. This requirement led to the construction of a custom network switch with fiber optic capability named Medusa. Medusa provides the foundation for MOOS undersea networks with managed L2 TCP/IP packet switching including fault tolerance and enhanced diagnostic capabilities for gauging the health of the undersea network. This paper will describe the overall requirements, design, software capabilities, and features, of the Medusa platform.**

1. INTRODUCTION

MBARI has been actively developing, over the past 4 years, the technology, both marine architecture, and electronics & software infrastructure that enable scientific measurements not only at the ocean surface, but in the benthic environment as well by providing power and optical communications. This was made possible by the development of the EOM cable which combines needed mechanical substratum, while providing power and fiber optic communications connectivity to the seafloor. This has been well documented in [2], and is depicted in Figure 1. A surface expression provides power via a bank of lead-acid batteries charged by photovoltaic panels and a wind generator. Long range communication is also provided for telemetry transport. Subsurface BINs (benthic instrumentation nodes) are connected to the surface expression by means of the EOM cable and specialized benthic cabling. Testing of the surface expression and EOM cable design has taken place over the past two years and has been documented in [3], and now the software and electronic components to complete the overall system are under development and test. At the same time, much planning and development of science-driven experiments have informed the initial scientific deployment of this system, known as the MOOS Science Experiment or MSE.

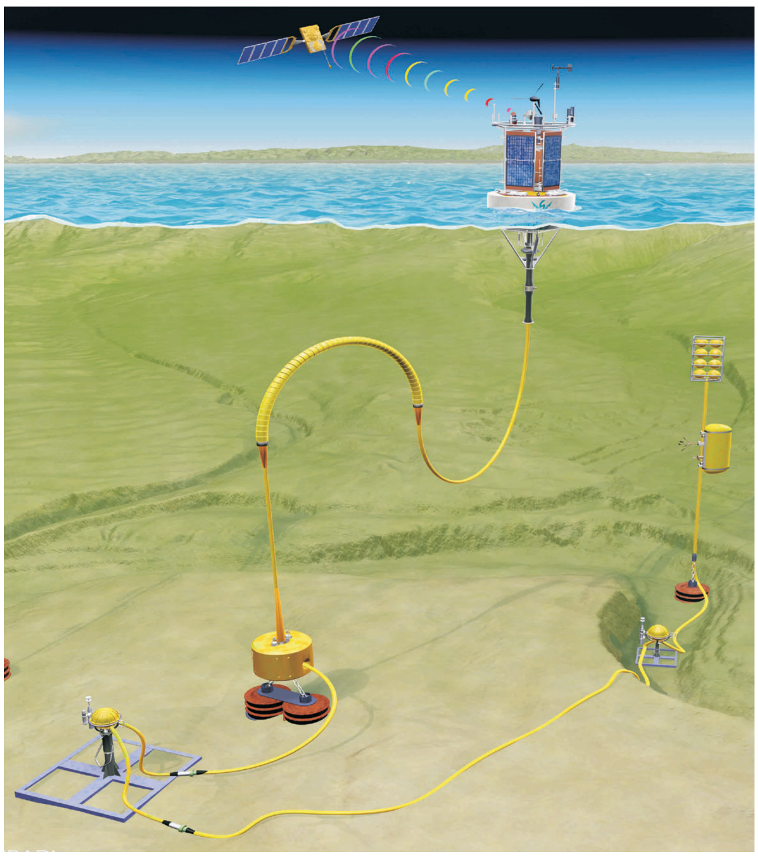

Figure 1: MOOS Science Experiment 2006 (Shepard Meander)

II. MOOS SCIENCE EXPERIMENT 2006

The MOOS Science Experiment (MSE) is scheduled for deployment in 2006 at a site 50 nautical miles offshore in the Monterey Bay (California) at a site known as Shepard Meander. Among the Science questions that MSE endeavors to answer are to characterize the progression of concentrations of organic carbon from the surface to the benthos, and the interaction of other processes in the Monterey Canyon, such as sediment transport. In order to characterize these processes fully, both surface and benthic instrumentation is required. Benthic instrumentation includes an Acoustic Doppler Current Profiler to characterize flow gradients at depth.

III. OPERATIONAL REQUIREMENTS

With a standard electronics mechanical form factor [1] and network communications protocol (TCP/IP) already well established, the need for both low-power operation and enhanced network diagnostic visibility are central in this operating environment. The initial range for data communications was specified for 10 kilometers over a single-mode fiber. Support for 100 MBit/s TCP/IP was desired as well as a target power consumption of one watt per TCP/IP channel. Six channels were specified with four of the six channels implementing ethernet over single mode fiber (100BaseFX) for benthic-to-benthic as well as benthic-to-surface connections. The two remaining channels implement 100BaseTX ethernet for local connectivity to the 'node' controller and ethernet-enabled instrumentation.

Because the benthic installation will offer limited diagnostic accessibility, it is important that fiber optic and network operational parameters be made available for diagnosis of the health of the benthic network. Because the benthic instrumentation nodes will also be powered from the surface expression, it is important that power consumption be minimized, and that capabilities exist to create and manage power management schemes for the benthic fiber optic links.

IV. MOOS SYSTEM BACKGROUND

Before going into detail on the Medusa network controller and it's role in the MOOS benthic network, some background information on two other components of the MOOS system is in order. These are the MOOS Mooring Controller (MMC) which provides the main, low-power computing platform, and the Software Infrastructure, known as SIAM, that runs on the MMC to manage the network 'nodes' and their instrumentation.

A. MOOS Mooring Controller (MMC)

The MOOS Mooring Controller or MMC, is the main computing platform on a MOOS mooring. It is sometimes referred to as the 'node' controller. The MMC was custom designed to meet the needs of MBARI's next generation moored instrumentation platform with 16 serial ports and an MMU-equipped ARM processor (Intel StrongARM SA1110), capable of running the Linux operating system. While required to be low power, many resources are available to applications developers because the MMC runs Linux, such as Java and a full featured TCP/IP networking suite.

While providing these services, the ARM processor on the MMC can also be put into a low power state to conserve power. This state involves suspending the processor, and resumption of processing is based on the detection of timer- or I/O-based events. I/O-based events include General Purpose I/O (GPIO) edge transitions and reception of serial traffic if enabled. Modifications have been made to the standard Linux kernel and device drivers to accomplish this. The on-board ethernet controller on the MMC is also aware of when the processor is suspended and has added support to handle the transition back to a running state gracefully.

The Medusa utilizes one of the MMCs wakeup mechanisms to implement a system-wide power management scheme, which will be described in greater detail later in this paper.

B. Software Infrastructure & Applications for MOOS (SIAM)

As mentioned, the feature rich computing platform includes support for Java which has enabled development of a software infrastructure known as SIAM [3]. SIAM utilizes Java's native remoting technology, Remote Method Invocation (RMI), to manage the instruments on one or many MMCs remotely. RMI depends on a TCP transport layer, and has been successfully deployed using Point-to-Point Protocol, or PPP, over modem-like devices, such as satellite transceivers. The Medusa card also allows the extension of TCP/IP service over many kilometers via a fiber optic cable. The distributed-object implementation and service-oriented design of the SIAM framework allows the remote operation of platforms and instruments to be accomplished independent of the physical transport layer, as long as TCP/IP service is available.

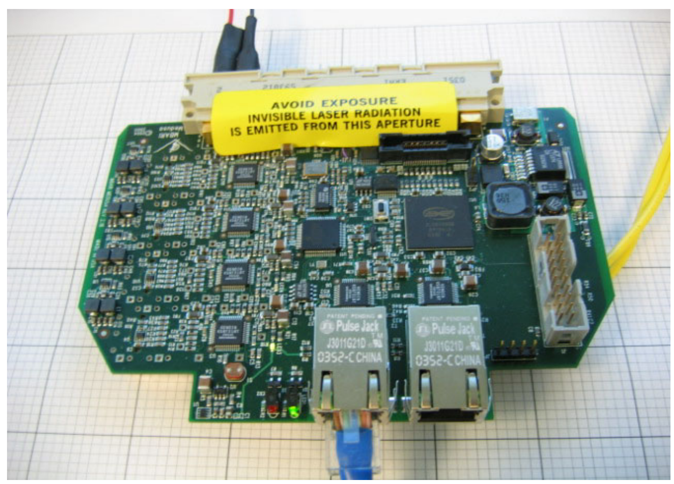

Figure 2: Top view of Medusa board on a 1 inch grid

V. HARDWARE DESIGN

A. Custom Requirements

Possibly the first question one might reasonably ask is why a custom hardware solution is justified in this instance. In the next few sections, an affirmative response to that question will be provided. The Medusa hardware is unique in terms of form-factor and supported functions, but more importantly provides 'managed switch' and fiber optic diagnostic capabilities that are an important component of a MOOS network.

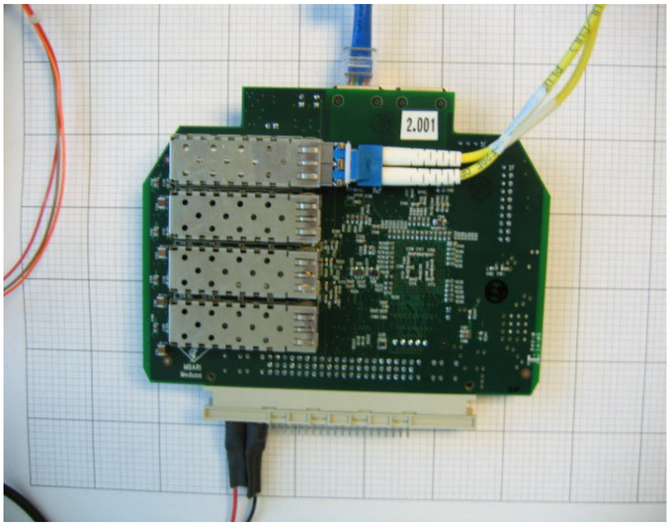

Figure 3: Medusa board reverse side with SFP module

Physically, the Medusa card is the same size as the other custom circuit boards, such as the MMC, in a MOOS backplane. The card size and backplane interfaces have been influenced largely by the IBC specification for printed circuit boards [1] designed to fit in a six-inch ID housing. Many COTS single-channel fiber/copper media converters and L2 switches were evaluated, but no product could be found that provided the remote diagnostic capability, channel density, or feature set that was required.

The Medusa board is complex, and is fabricated in 12 layers using state-of-the-art printed circuit board design and manufacturing techniques. The design is impedance controlled with a typical trace impedance of 50Ω. A 50 MHz clock is radially distributed to multiple components on the board using carefully length-matched stubs. The printed circuit board layout was challenging due to the small size of the board, the large number of components, and signal integrity requirements.

Careful attention to mechanical detail was also required as well because the optical modules, required to occupy the back side of the board by size constraints, had to be easily and safely installed, as well as allow for a minimum bend radius of the optical cable exiting the optical module. A picture of the top and bottom side of the Medusa card is provided in Figures 2 and 3 respectively.

B. L2 Network Switch

The network switching functionality provided by the Medusa card consists primarily of two components. The first is a ZL-50409 L2 network switch made by Zarlink. The Zarlink switch is unique in that it is one of the few L2 switches available that allows the designer to utilize an external PHY for each network channel. This was a requirement for this design because of the mixture of fiber and copper channels. In this case, the Kendin/Micrel KS8721BT PHYs were chosen because of their ability to interface to either 10/100 BaseTX copper or 100BaseFX fiber. The Reduced Media Independent Interface or RMII is utilized to transfer data between the switch and the PHYs. This is a relatively unique capability not found on many COTS products, as the Medusa board provides four 100BaseFX fiber channels, with SFP Fiber optic transceivers. Two 10/100BaseTX copper channels are also provided, each with an industry standard RJ-45 connector.

C. Management Processor

The second fundamental component that provided a great deal of flexibility in the design is an ARM7TDMI processor, the LPC2138 from Philips [5]. The LPC2138, while highly integrated, is produced in a small TQFP-64 physical package and does not have an external bus. An 8-bit ISA-like bus, required to communicate with the switch, was emulated with general purpose I/O pins in software. The LPC2138, while having 32 KBytes of SRAM, has a plentiful 512 KBytes of flash-based storage for executable code and static data. This combined with the relatively low power consumption of 50 milliwatts at a 60 MHz clock rate made the LPC2138 a good choice for a management processor.

The management processor serves many purposes: it initially configures the Zarlink switch and other components on the board including the fiber optic modules. Another task of the management processor utilizes capabilities of the Zarlink switch to implement a ‘CPU’ channel which can receive and transmit traffic. This enables network applications as described below.

VI. MANAGEMENT PROCESSOR SOFTWARE

A. TCP/IP Network Stack

While the Zarlink switch performs packet switching independent of the management processor, once it is configured and started, there are many features of the L2 switch that are utilized during it’s operation by the management processor. The switch defines a ‘CPU’ port, to which all traffic addressed to it is forwarded. This addressing is accomplished by assigning the CPU port a unique MAC address in the switch. After this is done, the switch notifies the management processor when a packet has arrived for the ‘CPU’, and further processing takes place. This capability can be used like a network adapter, and offer TCP/IP services if a suitable TCP/IP stack is added. For this project, the uIP TCP/IP stack developed by Adam Dunkels [8] was initially utilized because of the relative ease of porting, and reasonable resource requirements. The Medusa software does not incorporate an RTOS, so the fact that uIP does not require an RTOS is advantageous also.

While uIP does not require an RTOS, it only processes one incoming packet at a time including replies to ARP requests or ICMP echo (‘ping’) requests if necessary. While this is not necessarily optimal for high throughput, the bandwidth needs of the Medusa are modest: for reporting of diagnostic data and implementing a custom UDP protocol for power management.

Another important way in which the management processor works with the Zarlink L2 switch is to implement link failover. Link failover is a mechanism for implementing link fault tolerance and will be discussed in the following section.

B. Fault Tolerance

The Zarlink L2 switch has the capability to detect link failures through the use of link 'heartbeat' packets. If one of these packets or normal network traffic is not received on a monitored link in a predefined amount of time, an interrupt indicating this situation is dispatched to the management processor. This allows the management processor to take action if redundant link resources are available. This redundancy is achieved through a capability known as 'trunking'.

If a logical network link has been aggregated into at least one active and standby physical network link, then the management processor can configure the network switch, such that traffic can be routed onto either one of these physical links. This capability is known as trunking, and the management processor includes software support to manage the setup and configuration of trunk groups on the switch. This configuration information is persisted between resets and power cycles via an I2C EEPROM that was included on the board.

The envisioned MOOS benthic network will utilize a "1+1" trunk configuration, where two fibers are allocated as active and standby links as described above. If it is detected that the active link has failed, traffic is switched to the standby link. If fast response is required, then the standby link is kept powered -- this is known as 'hot' failover. If the power savings of keeping the standby link powered down is more important than response time to failure of the primary link, and a robust network protocol such as TCP can retry lost packets, then the standby link can be kept powered down until needed -- this is known as 'cold' failover.

C. Diagnostic Capabilities

For fiber optic links, the Medusa utilizes SFP fiber modules. Most of these modules have a diagnostic interface, similar to an I2C EEPROM, specified by the "Specification for Digital Diagnostic Monitoring Interface for Optical Transceivers" authored by the SFF Committee [7]. This document specifies the physical interface, the data format, and the content of the diagnostic data for each SFP module that is comformant to this standard. This important operational information includes parameters such as optical transmit power, laser bias current, optical receive power, and laser temperature. A single byte of binary flags is also included to quickly gauge link integrity and status. The ability to monitor these parameters can provide early warning of possible developing network problems, and allow remote diagnosis and advance planning of interventionary actions.

While status and power levels can be returned for each fiber optic channel from the on-board web server, the primary means of user interface on the Medusa card is a robust CLI or command-line interface. This interface allows querying of the status of the SFP modules as well as access to the per-channel RMON statistics in the switch. All switch and MII registers on each PHY are writable from the CLI, which proved valuable during board bringup and software development. Lastly, the CLI provides a means of initially configuring the board and persisting that configuration.

The CLI is written in such a way that commands are easily added with just a rebuild, making it easy to test the system and add new capabilities quickly [10]. A SIAM instrument service (the component of SIAM that interacts directly with an instrument via a serial port) has already been implemented to take advantage of this interface, and report optical data to the shore side data system.

VII. NETWORK POWER MANAGEMENT

Another important role the management processor had to perform as the engineering of the MOOS benthic network progressed is an agent to implement network power management. Through discussions with hardware, software, and systems engineers, three 'levels' of managing the overall power consumption of the optical network were initially identified.

A. Level 0 Power Management

The default level of power management, level 0, is performed by the Medusa card as part of it's default startup. Any empty optical transceiver sites are identified by a lack of acknowledge to an I2C probe at the address of `0xA0`. The power to these sites is switched off, but more importantly, the PHY corresponding to this SFP site is put into powerdown mode. This can save a few hundred milliwatts of power per PHY.

B. Level 1 Power Management

The level 1 power management implementation includes level 0 and also a capability provided by the TCP/IP stack on the Medusa card described earlier in this document. In this case, a UDP packet can be unicast or multicast to a Medusa card.

The packet payload may contain an IP address; if so, the software checks this address against the address that has been configured as the MMC in the same chassis. If the addresses match, the MMC is awakened via a shared backplane interrupt mechanism. This is known as a selective wakeup packet. If the packet is sent to the same port, but has no IP address, the MMC is unconditionally awakened using the same mechanism. This is known as a generic wakeup.

C. Level 2 Power Management

Level 2 power management goes beyond level 1 and utilizes the ability of the Medusa card to detect received optical power to coordinate the powering and depowering of optical links. This is perhaps the most aggressive scheme for managing the power consumed by the optical links as it has the most possible gain in saved power, but can also isolate parts of the network if not implemented carefully. It is envisaged that a special out-of-band control protocol would have to be implemented to manage the mechanics of this scheme along with fiber optic signal detection. Design and use case development of this next level of power management is currently underway.

For MSE 2006, it is the current plan to use the Level 1 Power Management scheme described in section B. above.

VIII. OPTICAL SUBSYSTEM

A. Fiber Optic Components

The Medusa card utilizes up to four SFP (Small Form Factor Pluggable) optical transceiver modules, utilizing wavelengths of 1310 nm and 1550 nm at the OC-3 transmission rate. Some representative data transmission rates are shown in Table 1 below. Specification of the SFP modules, their mechanical dimensions and electrical and digital interfaces are specified in the SFF Multisource Agreement, or MSA [6].

Optical Carrier Level	Data Transmmission Rate
OC-1	51.84 Mbit/s (basic SONET rate)
OC-3	**155.52 Mbit/s**
OC-12	622.08 Mbit/s

Table 1: Three Synchronous Digital Hierarchy (SDH) Systems and Data Rates

These SFP modules are readily available commercially from multiple sources in different operating wavelengths and emissive power. In this case, both 1310 nm and 1550 nm wave lengths are utilized over a single fiber by means of an optical duplexer. Each SFP module consumes approximately 750 milliwatts of power while in operation.

Power to each fiber optic module is externally switched and under software control by the processor. Each SFP module implements an I2C bus interface for retrieval of module properties and diagnostic information. While this feature is crucial, it is also a design issue that needed to be addressed, in that the SFP modules do not offer the hardware designer a mechanism to change the I2C address on a per-module basis. Hence they cannot all be on one I2C bus. We resolved this issue with a four-to-one I2C bus multiplexer, made by Philips. This allows the management processor tochoose which module occupies it's I2C bus at a given time.

This I2C capability, along with SFP fiber modules that implement the diagnostic data specification [7] allows various optical parameters, including supply voltage, transmitted and received optical power, and optical link integrity to be monitored in real time. In addition to monitoring the health of the fiber optic link, While formal BERT characterization of the optical modules and fibers was not able to be performed before the publication of this paper, much lab testing was done. Figure 4 below shows a typical lab test setup with three systems, each separated by two spools of fiber optic cable of approximately 5km in length. Network link integrity was not adversely affected without the use of an inline optical attenuator. Link integrity was maintained up until approximately 34 dB of additional attenuation was added. This gave us confidence that the hardware could work reliably within the budgeted 15 dB of attenuation anticipated with the actual deployed EOM cable. While formal BERT characterization of the optical modules and fibers was not able to be performed before the publication of this paper, much lab testing was done. Figure 5 shows a typical lab test setup with three systems, each separated by two spools of fiber optic cable of approximately 5km in length. Network link integrity was not adversely affected without the use of an inline optical attenuator. Link integrity was maintained up until approximately 34 dB of additional attenuation was added. This gave us confidence that the hardware could work reliably within the budgeted 15 dB of attenuation anticipated with the actual deployed EOM cable. the ability to sense the amount of received optical power can be effectively utilized in more sophisticated link power management schemes.

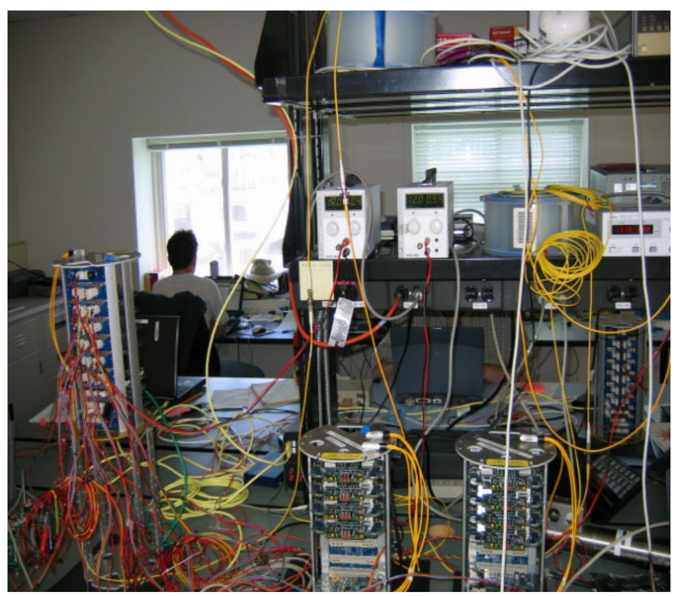

Figure 4: Three 'node' test system with two 5km fiber spools

IX. FIRMWARE DEVELOPMENT

Much of the flexibility in functions that can be accommo dated by the Medusa board are due to the recent advances in both microcontroller hardware and software development environments. Philips was one of the first companies on the market with a highly integrated ARM7TDMI-based microcontroller with on-board flash and SRAM, as well as many on-board peripherals, such as serial ports, I2C, SPI, and timers. In our case, the solid Philips I2C implementation provided easy interfacing to the SFP modules, and aided our development.

Software development solutions range from GNU toolchains that may be downloaded for free, to integrated development environments from IAR and Rowley & Associates. While dedicated JTAG solutions for the ARM have been costly in the past, parallel port or USB-based solutions are now available at very reasonable cost. The combination of an integrated development environment (Rowley Crossworks in this case) and USB JTAG interface greatly facilitated the development process. This enabled rapid prototyping, implementation, and debug of required features as well as new requirements needed for implementation.

X. CONCLUSION

In conclusion, COTS solutions always merit consideration as the decision to commit to designing and building custom hardware and developing custom software is never one to be taken lightly. In this instance, the ability to integrate a sophisticated network switch easily into existing mechanical and electrical hardware, along with the flexibility to implement enhanced diagnostic, fault tolerance, and power management features convinced us to develop this component of the MOOS infrastructure. We were also able to utilize to our advantage recent developments in switching silicon, low-power embedded processors, and standardized, modular fiber-optic transceivers.

While the benefits have yet to be fully demonstrated at the time of publication of this paper, initial lab testing and feature development necessary for the upcoming MSE 2006 deployment has shown the Medusa card to be a valuable element in the planned MOOS deployments with benthic instrumentation.

Acknowledgments

The authors gratefully acknowledge the support of this work by The David and Lucille Packard Foundation (http://www.packard.org).

References

[1] E. C. Mellinger, K. E. Prada, R. L. Koehler, K. W. Doherty, "Instrument Bus, An Electronic System Architecture for Oceanographic Instrumentation," *Technical Report WHOI-86-30*, Woods Hole Oceanographic Institution, August 1996

[2] M. Chaffey, E. Mellinger, W. Paul, "Communication and power to the seafloor: MBARIs ocean observing system mooring concept", *MTS/IEEE Oceans 2001 Conference Proceedings*, November 2001.

[3] A. Hamilton, M. Chaffey, E. Mellinger, J. Erickson, L. McBride, "Dynamic Modeling and Actual Performance of the MOOS Test Mooring", *MTS/IEEE Oceans 2003 Conference Proceedings*, November 2003.

[4] T. C. O'Reilly, et al, "Software Infrastructure and Applications for the Monterey Ocean Observing System: Design and Implementation", *MTS/IEEE Oceans 2004 Conference Proceedings*, November 2004.

[5] LPC2131/2132/2138 User Manual, Preliminary Release, Philips Semiconductors, November 22, 2004.

[6] SFF Committee, Small Form-Factor Pluggable (SFP) Multi-source Agreement (MSA), September 14, 2000.

[7] SFF Committee, SFF-8472 Specification for Digital Diagnostic Monitoring Interface for Optical Transceivers, Revision 9.3, August 1, 2002.

[8] ZL-50409, Managed 9-Port 10/100M Ethernet Switch, Data Sheet, Zarlink Semiconductor, January 2005.

[9] A. Dunkels, "Full TCP/IP for 8-bit Architectures", *Proceedings of the first international conference on mobile applications, systems and services (MOBISYS 2003)*, San Francisco, May 2003.

[10] B. Dupire, E. B. Fernandez, "The Command Dispatcher Pattern", *Proceedings of Conference on Pattern Languages of Programs*, Monticello, Illinois, September 2001.