



U.S. DEPARTMENT OF  
**ENERGY**

PNNL-22484

Prepared for the U.S. Department of Energy  
under Contract DE-AC05-76RL01830

# Sensor Characteristics Reference Guide

J Cree  
A Dasu  
P Fuhr  
S Lanzisera  
T McIntyre

R Muehleisen  
M Starke  
P Banerjee  
T Kuruganti  
C Castello

S McDonald, Editor

April 2013



**Pacific Northwest**  
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY  
*operated by*  
BATTELLE  
*for the*  
UNITED STATES DEPARTMENT OF ENERGY  
*under Contract DE-AC05-76RL01830*

**Printed in the United States of America**

**Available to DOE and DOE contractors from the  
Office of Scientific and Technical Information,  
P.O. Box 62, Oak Ridge, TN 37831-0062;  
ph: (865) 576-8401  
fax: (865) 576-5728  
email: [reports@adonis.osti.gov](mailto:reports@adonis.osti.gov)**

**Available to the public from the National Technical Information Service,  
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161  
ph: (800) 553-6847  
fax: (703) 605-6900  
email: [orders@ntis.fedworld.gov](mailto:orders@ntis.fedworld.gov)  
online ordering: <http://www.ntis.gov/ordering.htm>**



This document was printed on recycled paper.

(9/2003)

# Sensor Characteristics Reference Guide

J Cree	R Muehleisen****
A Dasu*	M Starke**
P Fuhr**	P Banerjee*
S Lanzisera***	T Kuruganti**
T McIntyre**	C Castello**

S McDonald, Editor

April 2013

\*EDL   \*\*ORNL   \*\*\*LBNL   \*\*\*\*ANL

Prepared for  
the U.S. Department of Energy  
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory  
Richland, Washington 99352

## Abstract

The Buildings Technologies Office (BTO), within the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy (EERE), is initiating a new program in Sensor and Controls. The vision of this program is:

- Buildings operating automatically and continuously at peak energy efficiency over their lifetimes and interoperating effectively with the electric power grid.
- Buildings that are self-configuring, self-commissioning, self-learning, self-diagnosing, self-healing, and self-transacting to enable continuous peak performance.
- Lower overall building operating costs and higher asset valuation.

The overarching goal is to capture 30% energy savings by enhanced management of energy consuming assets and systems through development of cost-effective sensors and controls.

One step in achieving this vision is the publication of this Sensor Characteristics Reference Guide. The purpose of the guide is to inform building owners and operators of the current status, capabilities, and limitations of sensor technologies. It is hoped that this guide will aid in the design and procurement process and result in successful implementation of building sensor and control systems. DOE will also use this guide to identify research priorities, develop future specifications for potential market adoption, and provide market clarity through unbiased information.



# Executive Summary

The purpose of the guide is to inform building owners and operators of the current status, capabilities, and limitations of sensor technologies. It is hoped that this guide will aid in the design and procurement process and result in successful implementation of building sensor and control systems. DOE will also use this guide to identify research priorities, develop future specifications for potential market adoption, and provide market clarity through unbiased information.

This Guide is organized into four major sections:

- Power Generation/Storage – Such as batteries, AC wired, or other power source.
- Communications and Networking – Both Wired and Wireless
- Signal Data Aggregator – Interfaces to Legacy and More Modern Networks
- Signal Conversion to Data – Analog Signal Conversion to Digital and Other Data Formats

Each section is divided into chapters that, in general, describe the technical detail of alternative technologies including current status, typical applications, and the advantages and disadvantages of these competing approaches. When available, information is also provided on relative costs. Finally, information is given on the interdependencies of these sensor characteristics such as range, energy requirements, and reliability.

This guide does not advocate for any particular approach to implementation of a building sensor or control system. Instead, the goal is to better inform building owners and operators as to the features, benefits, and limitations of these various approaches. It is subsequently hoped that a better informed consumer will make an informed choice given the particulars of their building and needs.

This guide will also help DOE to define future opportunities to extend capabilities or reduce costs of sensor technologies through research, technology specifications, or other means to advance the technology. In this manner, DOE hopes to move closer to achieving the vision whereby buildings operate at peak efficiency, are easier to maintain, and provide owners with lower operating costs and higher valuations.



## **Acknowledgments**

The authors would like to thank Alexis Abramson of the Building Technologies Office (BTO) for her support of this study and resultant publication. The authors would also like to thank George Hernandez of Pacific Northwest National Laboratory for his technical oversight and review of this guide, and Anne Wagner of Pacific Northwest National Laboratory for overall project management and a fresh eye. Finally, the authors would like to thank Cary Counts and Megan Peters for their help in making this publication a coherent whole from the sum of the parts contributed by each author.





## Acronyms and Abbreviations

AC	Alternating Current
ACK	Acknowledgments
ADC	Analog-to-digital Converters
AEG	Acoustic Energy Generators
API	Application Programming interface
CCA	Clear Channel Assessment
CCD	Charge-coupled Device
CDC	Capacitance-to-digital Converter
CMOS	Complementary Metal Oxide Semiconductor
CSMA	Carrier Sense Multiple Access
DAC	Digital-to-analog Convertor
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DOE	Department of Energy
DWT	Discrete Wavelet Transform
EIS	Energy Information System
FFT	Finite Fourier Transform
FIR	Finite Impulse Response
GPS	Global Positioning Satellite
HART	Highway Addressable Remote Transducer
HVAC	Heating, Ventilation, and Air Conditioning
IEC	International Electrotechnical Commission
IIR	Infinite Impulse Response
IP	Internet Protocol
IR	Infrared
LAN	Local Area Network
MAC	Media Access Control
MFC	Microbial Fuel Cell
NEC	National Electric Code
NRE	Non-recurring Engineering
OPC	Object Process Control
PAN	Personal Area Network
PEG	Piezoelectric Generators
PEMFC	Proton Exchange Membrane Fuel Cell
PHY	Physical Layer
PIFA	Planar Inverted F-Antenna

PLC	Power Line Carrier
PoE	Power over Ethernet
POF	Polymer Optical Fiber
PSCC	Power Systems Communication Committee
PV	Photovoltaic
REST	REpresentational State Transfer
RF	Radio Frequency
RFID	Radio Frequency Identification
SAR	Successive Approximation Register
SCADA	Supervisory Control and Data Acquisition
SDM	Sigma-delta Modulation
SHA	Sample-and-hold Amplifiers
SOAP	Simple Object Access Protocol
STP	Shielded Twisted Pair
TDMA	Time Division Multiple Access
TEG	Thermoelectric Generator
TIA	Telecommunications Industry Association
USB	Universal Serial Bus
UTP	Unshielded Twisted Pair
VFC	Voltage-to-frequency Converter
VPN	Virtual Private Network
WPT	Wireless Power Transfer
WSN	Wireless Sensor Network

# Contents

Abstract.....	iii
Executive Summary.....	v
Acknowledgments.....	vii
Acronyms and Abbreviations.....	ix
1.0 Power Generation: Main Wired.....	1.1
1.1 Introduction.....	1.1
1.2 Technical Details.....	1.2
1.3 Characteristic Interdependencies.....	1.6
1.4 Typical Applications.....	1.6
1.5 Relative Cost Impacts.....	1.6
1.6 Summary/Conclusions.....	1.7
1.7 References.....	1.7
2.0 Power Generation: Source Excited.....	2.1
2.1 Introduction.....	2.1
2.2 Technical Details.....	2.1
2.2.1 Characteristic Interdependencies.....	2.2
2.2.2 Typical Applications.....	2.3
2.3 References.....	2.3
3.0 Power Generation: Other Local.....	3.1
3.1 Introduction.....	3.1
3.2 Technical Details.....	3.1
3.2.1 Radioisotope Generator.....	3.1
3.3 Characteristic Interdependencies.....	3.6
3.4 Typical Applications.....	3.6
3.5 Relative Cost Impacts.....	3.6
3.6 Summary/Conclusions.....	3.7
3.7 References.....	3.7
4.0 Power Generation: Batteries.....	4.1
4.1 Introduction.....	4.1
4.2 Technical Details.....	4.2
4.2.1 Primary Batteries.....	4.2
4.2.2 Alkaline Batteries.....	4.4
4.2.3 Zinc/Air Batteries.....	4.4
4.2.4 Lithium Batteries.....	4.6
4.3 Secondary Batteries.....	4.11
4.3.1 Nickel-Cadmium Batteries.....	4.14

4.3.2	Lithium-Ion Batteries .....	4.16
4.4	Characteristic Interdependencies.....	4.17
4.5	Typical Applications .....	4.18
4.6	Relative Cost Impacts.....	4.18
4.7	Summary/Conclusions .....	4.19
4.8	References .....	4.20
5.0	Power Generation: Self-Harvesting Sources .....	5.1
5.1	Introduction .....	5.1
5.2	Technical Details.....	5.1
5.2.1	Thermal Power Harvesting.....	5.3
5.2.2	Vibration Power Harvesting.....	5.3
5.2.3	Photovoltaic Power Harvesting .....	5.5
5.2.4	Electromagnetic Power Harvesting .....	5.5
5.2.5	Wind Power Harvesting .....	5.5
5.2.6	Acoustic Power Harvesting.....	5.5
5.3	Characteristic Interdependencies.....	5.5
5.4	Typical Applications .....	5.6
5.5	Relative Cost Impacts.....	5.8
5.6	Summary/Conclusions .....	5.9
5.7	References .....	5.9
6.0	Power Generation: Other Local Storage.....	6.1
6.1	Introduction .....	6.1
6.2	Technical Details.....	6.1
6.2.1	Electrostatic Capacitors.....	6.2
6.2.2	Electrolytic Capacitors .....	6.2
6.2.3	Electrochemical Capacitors.....	6.3
6.3	Characteristic Interdependencies.....	6.4
6.4	Typical Applications .....	6.5
6.5	Relative Cost Impacts.....	6.5
6.6	Summary/Conclusions .....	6.5
6.7	References .....	6.5
7.0	Communications and Networking: Introduction .....	7.1
7.1	Network Topology .....	7.2
7.2	Communication Specifications and Requirements.....	7.2
7.2.1	Corruption .....	7.3
7.2.2	Bandwidth .....	7.3
7.2.3	Throughput.....	7.4
7.2.4	Delay .....	7.4
7.2.5	Reliability.....	7.5

7.2.6	Robustness.....	7.5
7.2.7	Scalability.....	7.5
7.2.8	Signal-to-Noise Ratio.....	7.5
7.2.9	Interference.....	7.5
7.2.10	Message Contention.....	7.6
7.2.11	Extensibility.....	7.6
7.3	System Requirements.....	7.6
8.0	Communications and Networking: Network Management.....	8.1
8.1	Configuration.....	8.1
8.2	Localization.....	8.2
8.3	Concepts of a Communication Protocol.....	8.3
8.4	Physical Connection.....	8.4
8.5	Direct Connection.....	8.5
8.6	Logical Connection.....	8.6
8.7	Data Communication.....	8.8
8.8	Interoperability.....	8.8
8.9	Security Concepts.....	8.10
8.10	Wired and Wireless Systems.....	8.12
9.0	Communications and Networking: Wired.....	9.1
9.1	Introduction.....	9.1
9.2	Power Line Carrier Communications.....	9.1
9.3	Dedicated Wiring for Building Sensor Communications.....	9.4
9.4	Network Architectures for Wired Building Sensors.....	9.7
9.5	Fiber Optic Communications for Wired Building Sensors.....	9.10
9.6	Summary/Conclusions.....	9.13
9.7	References.....	9.14
10.0	Communications and Networking: Wireless.....	10.1
10.1	Introduction.....	10.1
10.2	Wireless for Building Automation.....	10.1
10.3	Wireless Technologies.....	10.5
10.4	Distance between Transceivers.....	10.5
10.5	Building Automation-Strength Wireless Sensor Networks.....	10.6
10.6	Coexistence of 802.11 and 802.15.4.....	10.8
10.7	Network Topologies.....	10.10
10.7.1	Bus Networks.....	10.11
10.7.2	Ring Networks.....	10.12
10.7.3	Star Networks.....	10.13
10.7.4	Tree Networks.....	10.13
10.7.5	Mesh Networks.....	10.14

10.7.6 Mesh Networks Variants .....	10.15
10.7.7 Data Rates for Backhaul Network Technologies .....	10.18
10.8 Summary .....	10.19
10.9 References .....	10.21
11.0 Signal Data Aggregator: Legacy Building Interfaces.....	11.1
11.1 Introduction .....	11.1
11.2 Technical Details.....	11.4
11.2.1 Legacy Interfaces .....	11.4
11.2.2 Sensor Cabling Advantages and Disadvantages.....	11.6
11.2.3 Data Tagging with Metadata .....	11.7
11.3 Characteristic Interdependencies.....	11.7
11.4 Typical Applications .....	11.8
11.5 Relative Cost Impacts.....	11.9
11.5.1 Current Cost Impacts.....	11.9
11.5.2 Future Low-Cost Possibilities .....	11.10
11.6 Summary/Conclusions .....	11.10
11.7 References .....	11.11
12.0 Signal Data Aggregator: Building Network Interfaces.....	12.1
12.1 Introduction .....	12.1
12.2 Technical Details.....	12.3
12.2.1 Common Network Interfaces .....	12.3
12.2.2 Constraints Set by the Network Stack .....	12.4
12.2.3 Bidirectional Versus Unidirectional Data Transfer.....	12.6
12.2.4 Understanding Received Data .....	12.6
12.2.5 Device Drivers for Data Aggregators.....	12.6
12.3 Characteristic Interdependencies.....	12.7
12.4 Typical Applications .....	12.8
12.5 Relative Cost Impacts.....	12.9
12.5.1 Current Cost Impacts.....	12.9
12.5.2 Future Low-cost Possibilities .....	12.10
12.6 Summary/Conclusions .....	12.10
12.7 References .....	12.10
13.0 Signal Data Aggregator: External Network Interface.....	13.1
13.1 Introduction .....	13.1
13.2 Technical Details.....	13.2
13.2.1 Application Programming Interfaces .....	13.3
13.2.2 Archiver Structure .....	13.4
13.2.3 Additional IT Network Considerations .....	13.5
13.3 Characteristic Interdependencies.....	13.6

13.4 Typical Applications .....	13.6
13.5 Relative Cost Impacts.....	13.7
13.6 Summary/Conclusions .....	13.7
13.7 References .....	13.7
14.0 Signal Conversion to Data: Signal Form Conversion.....	14.1
14.1 Introduction .....	14.1
14.2 Analog-to-Digital Converters.....	14.1
14.2.1 Integrating ADC.....	14.2
14.2.2 Sigma-Delta Modulation ADC.....	14.4
14.2.3 Direct or Flash ADC .....	14.5
14.2.4 Successive Approximation Register ADC .....	14.6
14.2.5 ADC Summary .....	14.8
14.3 Current-to-Digital Conversion .....	14.8
14.4 Capacitance-to-Digital Conversion .....	14.8
14.4.1 Voltage-to-Frequency Converters (VFCs).....	14.9
14.5 Frequency-to-Digital Conversion and Frequency Counters.....	14.9
14.6 Optical Signal Conversion .....	14.10
14.7 Infrared Signal Conversion .....	14.10
14.8 RF Signal Conversion .....	14.11
14.9 Signal Conditioning.....	14.11
14.10 References .....	14.12
15.0 Signal Conversion to Data: Data Fusion .....	15.1
15.1 Introduction .....	15.1
15.1.1 Wider Spatial/Temporal Coverage.....	15.2
15.1.2 Better Robustness/Fault Tolerance.....	15.2
15.1.3 Better Estimation of Information from Data .....	15.2
15.2 Technical Details.....	15.3
15.2.1 Low-Level Fusion .....	15.3
15.2.2 Feature-Level Fusion.....	15.3
15.2.3 Decision-Level Fusion .....	15.4
15.2.4 Distributed versus Centralized Fusion .....	15.5
15.2.5 Data Fusion Process Model.....	15.7
15.3 Characteristic Interdependencies.....	15.9
15.4 Typical Applications .....	15.9
15.5 Relative Cost Impacts.....	15.10
15.6 Summary/Conclusions .....	15.11
15.7 References .....	15.11
16.0 Signal Conversion to Data: Signal Processing .....	16.1
16.1 Introduction .....	16.1



16.2 Definitions:.....	16.2
16.2.1 Technical Details.....	16.2
16.3 References.....	16.11

# Figures

1.1	Depiction of a Sensor Node.....	1.1
1.2	Step-Down Transformers for Residential Single-Phase Power Supply.....	1.3
1.3	Step-Down Transformers for Residential Multiple Phase and Commercial with Potential Single-Phase Connection.....	1.4
1.4	Electrical Interconnection to Sensor Node .....	1.5
1.5	Electrical Interconnection to Sensor Node with a Switch Installed.....	1.5
2.1	Schematic of a Planar Inverted F-Antenna.....	2.1
2.2	Example of Wireless Power Transfer.....	2.2
2.3	Wireless Power Transfer .....	2.2
3.1	Schematic Illustrating Various Radioisotope Energy Conversion Techniques.....	3.2
3.2	Schematic Illustrating the Radioisotope Actuation of Self-reciprocating Cantilevers, and the Generation of Electrical Energy when the Cantilever is a Piezoelectric Unimorph.....	3.3
3.3	Depiction of a Fuel Cell .....	3.4
3.4	Depiction of Fuel Cell Interconnections.....	3.5
4.1	Electrochemical Cell .....	4.2
4.2	Advances in Development of Primary Batteries .....	4.3
4.3	Typical Discharge Curves of Lithium/Solid-Cathode Batteries.....	4.10
4.4	Discharge Profiles of Conventional Secondary Battery Systems and Rechargeable Lithium-Ion Battery at Approximately C/5 Discharge Rate.....	4.13
5.1	Energy Harvesting With and Without Storage Architecture .....	5.1
5.2	Configuration Drawing of a Typical Thermoelectric Generator .....	5.3
5.3	Basic Piezoelectric Cantilever System for Generating Electricity Using a Cantilever and Diaphragm .....	5.4
6.1	Depiction of Electrostatic Capacitor.....	6.1
6.2	Various Electrostatic Capacitors .....	6.2
6.3	Depiction of Electrolytic Capacitor.....	6.3
6.4	Depiction of Electrochemical Capacitor.....	6.4
7.1	A Sensor must not only Measure but also Communicate the Measured Result such that an Action can be Performed for it to be Useful.....	7.1
7.2	Different Network Topologies have their own Benefits and Drawbacks.....	7.3
7.3	Throughput of a Connection: Shipping a Package .....	7.4
8.1	It is Important to Know Where each Device is Placed so that it is Possible to Correlate the Data from a Device with the Location of the Device .....	8.3
8.2	Different Physical Media.....	8.5
8.3	The Purple Computer in the Middle Acts as a Network Gateway or Bridge and Translates from the Protocol of one Network into the Protocol of Another and Vice Versa .....	8.7
9.1	A Typical Building Lighting, Switch and Electrical Socket Layout .....	9.1
9.2	PLC Communication System .....	9.3

9.3	Signal Attenuation with Frequency .....	9.5
9.4	Data Rate Versus Distance .....	9.5
9.5	A Building Riser with Coexisting Communication and Power Wiring .....	9.6
9.6	RS-485 Data Rate Versus Cable Length .....	9.7
9.7	Frequently used RS-485 Bus Architectures.....	9.8
9.8	Building Automation Network Riser Diagram.....	9.9
9.9	A Classic IT-Centric Network Design for Building Automation Systems.....	9.9
9.10	Single-Mode and Multi-Mode Optical Fibers .....	9.11
9.11	A Wiring Closet in a Commercial Building that Includes the Telephone Exchange Wiring, Coaxial-Cable Based Broadband Satellite TV Distribution Wiring, a Fiber Optic Video Distribution Network, and a 96 Port Ethernet Distribution Network.....	9.13
10.1	The Signal Level from a Transmitter Decreases in a Nonlinear Manner with Distance.....	10.2
10.2	Data Rate as a Function of Separation Distance for an 802.11g Access Point and Client.....	10.2
10.3	Throughput of Wireless Systems Changes as a Function of Distance .....	10.3
10.4	Separation Distance, or “Footprint,” Versus Data Throughput for a Wireless Technologies ...	10.4
10.5	The 100 m RF Footprint Associated with 802.15.4 and 802.11 Transceivers is Overlaid onto a Photograph of an Industrial Site .....	10.4
10.6	The 802.15.1 Frequency Channels in the 2450 MHz Range.....	10.7
10.7	The 802.15.4 Frequency Channels in the 2450 MHz Range.....	10.7
10.8	The Overlapping Channel Assignments for 802.11 Operation in the 2400 MHz Range .....	10.7
10.9	802.11n Dual Stream Occupies 44 MHz of Bandwidth. Dual stream 802.11n in the 2.4 GHz band is shown.....	10.8
10.10	Non-Overlapping RF Footprints.....	10.9
10.11	Dense Deployment of Wireless Field Transmitters at a Building Automation Site.....	10.10
10.12	Networking Topologies Take Many Forms with Associated Levels of Complexity Required for Robust Fault-Tolerant Data Transport .....	10.11
10.13	A Linear Chain of Sensors, with their Associated Connection Points, Form an Easy-to- Scale Network .....	10.12
10.14	Outside Access to a Ring Network Relies on an Information Gateway being Connected to the Ring .....	10.13
10.15	A Wired Star Network.....	10.13
10.16	Classic Tree Networking Structure.....	10.14
10.17	An 802.11 Mesh Network Relying on Dual Radios.....	10.15
10.18	List of a Few of the Mesh Networking Algorithms used in Wireless Sensors.....	10.15
10.19	Bluetooth Devices rely on a Localized Network-to-Network Topology that is Meant to Allow Message Bouncing.....	10.16
10.20	As the Range Increases, the SNR Decreases, which for Constant Error Rate Means that the Data Rate must Decrease.....	10.17
10.21	List of RF Technologies by Data Rate and RF Coverage .....	10.18
10.22	Integrated Wireless Environment for Building Automation Settings.....	10.19

10.23	Bandwidth Requirements for a Variety of Applications and the Associated Wireless Technology that can Support such Requirements .....	10.20
10.24	Structured Fabric Design of Layered Wireless for a Building Automation Facility .....	10.20
11.1	Common Data Aggregation Scenarios Showing Network-Based and Legacy Aggregation to a Primary Data Aggregator as well as a case where a Local Aggregator makes Control Decisions for a Lighting System and Provides Summary Data to the Primary Aggregator.....	11.2
11.2	Full Data Aggregation System .....	11.3
11.3	Three Types of Interfaces in a Data Aggregator .....	11.4
12.1	Data Aggregator with Three Types of Interfaces .....	12.1
12.2	Open Systems Interconnection Model for a Network Stack, Showing the Name and Number of Each Layer Along with Information about What Happens at each Layer.....	12.2
13.1	A Data Aggregator has Three Types of Interfaces .....	13.1
13.2	A Data Aggregator Located in a Facility is Connected via the Local IT Network Through a Firewall to External Services and Applications.....	13.2
13.3	A Full Data Aggregation System.....	13.2
14.1	Output of the Integrator in a Dual-Slope ADC.....	14.3
14.2	Diagram of a Flash ADC System .....	14.5
14.3	Basic SAR ADC Architecture .....	14.7
14.4	Analog-to-Digital Conversion using a Voltage-to-Frequency Converter and a Frequency Counter .....	14.9
15.1	A Sensor Fusion Scenario Inside a Building .....	15.1
15.2	Low-Level Data Fusion Scheme .....	15.3
15.3	Feature-Level Fusion Scheme .....	15.4
15.4	Decision-Level Fusion Scheme .....	15.4
15.5	Centralized Fusion Architecture .....	15.5
15.6	Hierarchical Data-Fusion Architectures .....	15.6
16.1	Signal Conditioning.....	16.2
16.2	Example of Anti-Aliasing Filter Implementation.....	16.3
16.3	Types of Filters used in Digital Filters .....	16.4
16.4	Different ADC Architectures versus Bits and Bandwidth.....	16.7

## Tables

1.1	Comparison of Cost between Wireless and Wired Technologies .....	1.7
3.1	Comparison of Radiation Properties of $^3\text{H}$ , $^{63}\text{Ni}$ , and $^{147}\text{Pm}$ Thin Films .....	3.1
3.2	Cost Comparison to Generate One Kilowatt of Energy .....	3.7
4.1	Major Characteristics and Applications of Primary Batteries .....	4.3
4.2	Characteristics of 1.5-V Standard Alkaline-Manganese Dioxide Batteries .....	4.4
4.3	Characteristics of Zinc/Air Button and Coin Batteries .....	4.5

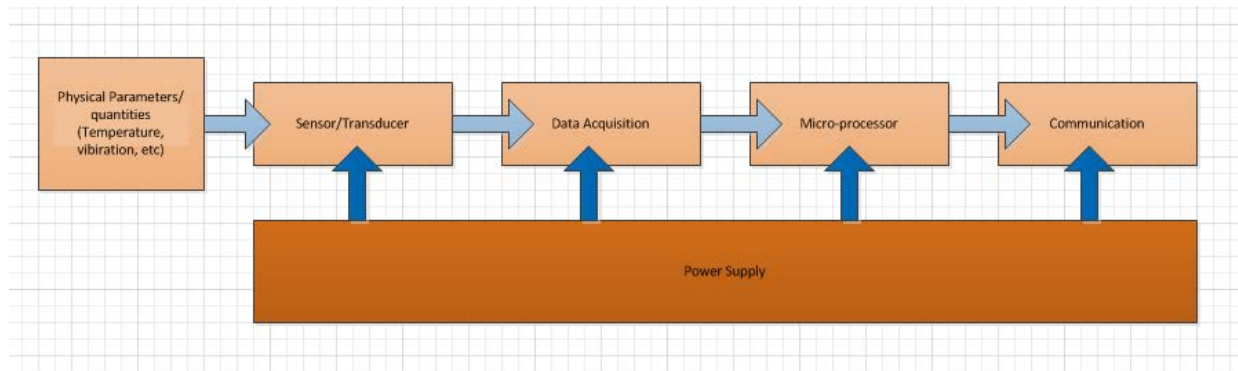
4.4	Characteristics of Lithium Primary Batteries .....	4.7
4.5	Characteristics of Typical Lithium/Solid-Cathode Batteries .....	4.9
4.6	Characteristics of Larger Li/MnO <sub>2</sub> Batteries.....	4.10
4.7	Specifications for Commercially Available Foil Cells and Batteries.....	4.11
4.8	Major Characteristics and Applications of Secondary Batteries .....	4.12
4.9	Comparison of Secondary Batteries .....	4.14
4.10	Specifications of Typical Sealed Nickel-Cadmium Single-Cell Batteries .....	4.15
4.11	General Performance Characteristics of Li-Ion Batteries .....	4.17
4.12	Cost of Energy Obtained by Primary Batteries .....	4.18
4.13	Cost of Energy Obtained by Secondary Batteries .....	4.19
4.14	Typical Energy Consumption of a WSN .....	4.19
5.1	Factors Affecting the Power Consumption of a Wireless Sensor Network .....	5.2
5.2	Power Density of Various Energy Harvesting Techniques .....	5.2
5.3	Amplitude of Vibration and Fundamental Frequency for Vibration Sources .....	5.6
6.1	Cost Estimation for 1 MJ or 277 Wh Storage .....	6.5
9.1	Data Rates for Various Cables.....	9.6
9.2	Symbol Rate Versus Length of Fiber for Various Optical Fibers .....	9.11
9.3	Gigabit Ethernet Distances for Optical Fiber .....	9.12
10.1	List of Building Automation Radio Technology Standards and Their Candidate Applications.....	10.5
10.2	Attenuation and Received Power Versus Range .....	10.6
10.3	Transmitter-to-Receiver Separation Distance as a Function of Data Rate .....	10.6
14.1	Summary of ADC Characteristics .....	14.8
16.1	Physical Parameters Measured and Conversion to Electrical Signals.....	16.1

# 1.0 Power Generation: Main Wired

## 1.1 Introduction

Sensors within building technologies represent the intelligence backbone that interconnects building equipment, subsystems, and analytical tools. The use of advanced sensors and controls will foster the future of energy-efficient building technologies with building energy-management systems showing significant expected growth of 17 percent compound annual rate for the next decade. This growth is anticipated because of advanced invasive, cost-effective, and easy-to-install sensor and control technologies such as wireless sensors switches, conductive media technologies, and energy-harvesting technologies <sup>[1]</sup>.

A key component to all sensor technologies is the supply power. There are many different approaches to providing power to a sensor node. A sensor node is defined in literature as a device that can sense, process, and communicate information. It usually consists of a sensor or transducer, a micro-processor, and data acquisition and communication systems <sup>[6],[7]</sup>. Hardware interconnections for a sensor node are shown in Figure 1.1.



**Figure 1.1.** Depiction of a Sensor Node

The basic power supply options consist of a direct interconnection to the building power system, batteries, or energy harvesting technologies. The decision on the power supply option to be used is governed by the power requirements. Depending on the design, purpose, interconnection, and the expected usage and lifetime of the sensor node, different levels of power, current, and voltage ratings can be anticipated. For example, consider a sensor node that performs environmental measurements such as temperature and atmospheric pressure. This type of sensor will typically perform periodic measurement on the order of minutes while a sensor being used for vibration detection within industrial applications will make measurements more frequently <sup>[3]</sup>. This difference in sampling periods will result in significantly different processing requirements and idling periods, which directly govern the energy consumption of a sensor node. Similarly, the communication methods employed will directly impact energy consumption. Commonly employed wireless communication approaches such as ZigBee™ or Bluetooth® have significantly different energy consumption rates. <sup>[4]</sup> In all aspects of the sensor design, the energy and source of power become a key issue.

<sup>1</sup> ZigBee™ is a standards-based wireless technology designed to address the unique needs of low-cost, low-power wireless sensor and control networks ([www.zigbee.org](http://www.zigbee.org)). Bluetooth® is a

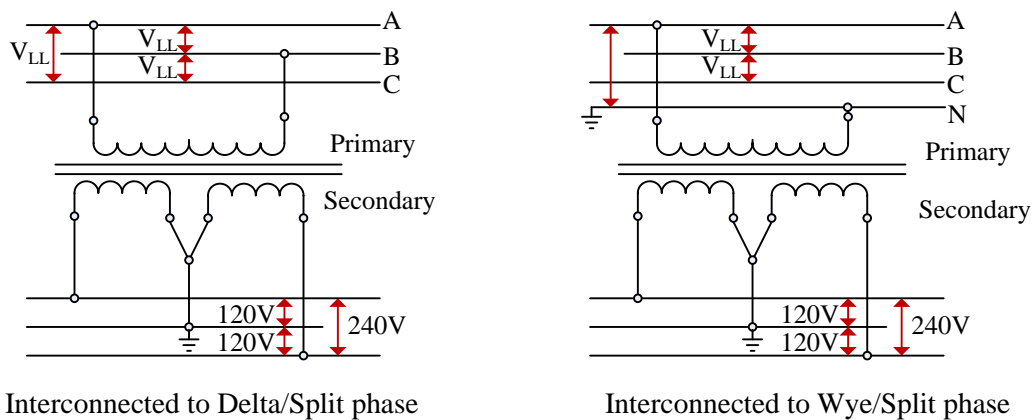
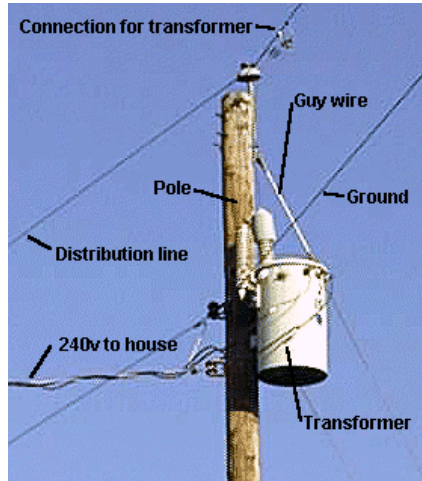
In the first six chapters of this document, the authors delve into the respective power generation and storage technologies typically used to provide the necessary power to various sensors in building applications. These chapters will focus on providing the technical details of the technology, describe interdependent relationships, typical applications, and provide insight into the associated costs. In this chapter, the topology of interconnecting the sensor directly into the main power delivery system of the building is examined. This is often described or named as a “wired” sensor.

## 1.2 Technical Details

For residential homes and many commercial establishments, the interconnection to the distribution system can be achieved through a variety of approaches. In many locations, a set of poles with one phase of power (at 7200 or other distribution voltage) and a ground wire (although sometimes there will be two or three phases on the pole, depending on where the house is located in the distribution grid) deliver the interconnection to the power system. At more rural residential settings, there often is a transformer drum attached to the pole as shown in Figure 1.2. In this design, a ground wire runs down the pole and is stapled in a coil to the base of the pole and therefore is in direct contact with the earth, running 6 to 10 ft (1.8 to 3 m) underground. Two wires coming from the secondary side of the transformer are connected with the ground to deliver three wires to the house. The two wires from the transformer are insulated, and the third ground wire is bare. The two insulated wires each carry 120 V, but they are 180 degrees out of phase so the difference between them is 240 V. Thus, the arrangement allows power to be delivered for both 120-V and 240-V systems. Example interconnection configurations for both Delta and Wye connections on the high side are shown in Figure 1.2. In other locations, the distribution lines are installed underground, and green transformer boxes are used.

---

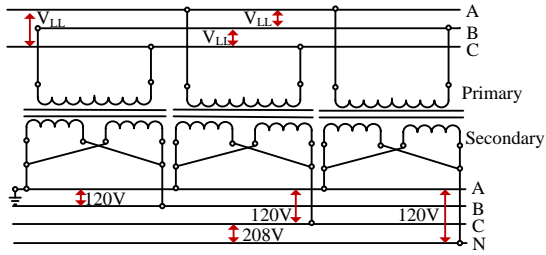
wireless technology standard for exchanging data over short distances (using short-wavelength radio transmissions from 2400–2480 MHz) from fixed and mobile devices ([www.bluetooth.com](http://www.bluetooth.com)).



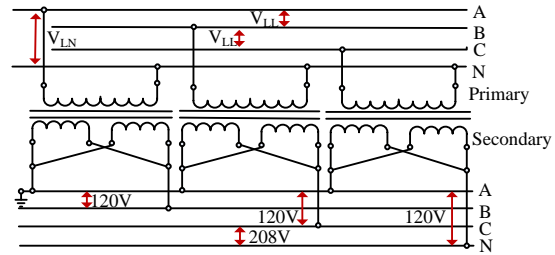
**Figure 1.2.** Step-Down Transformers for Residential Single-Phase Power Supply

Another configuration involves using three transformers to provide a three-phase Wye connection on the low voltage side. In this configuration, the phase-to-phase voltage is 208 V with a phase-to-ground voltage of 120 V for appliances. Figure 1.3 shows the Delta and Wye configurations on the high side. A three-transformer configuration provides better reliability as one transformer can fail and service will still be maintained.





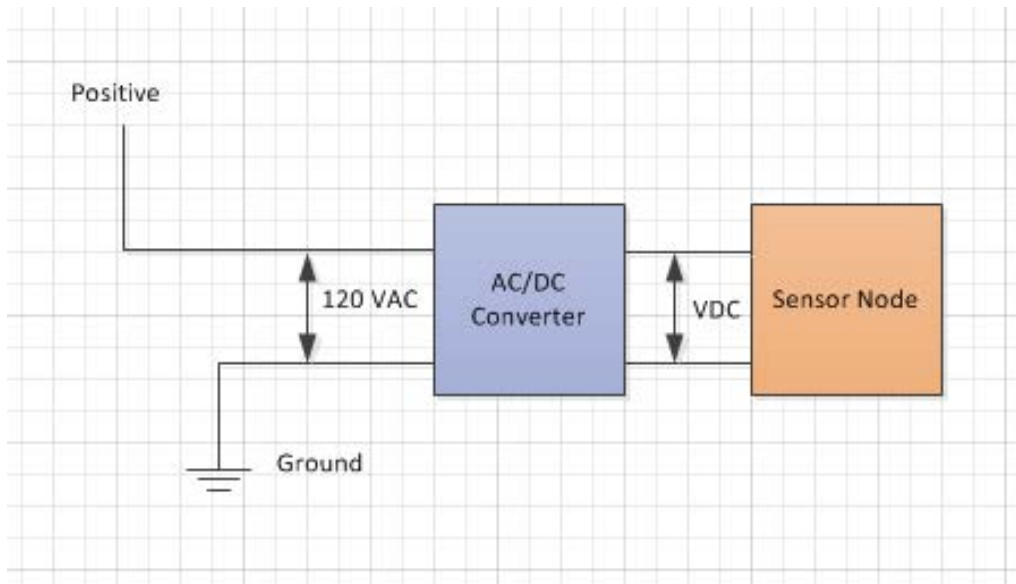
Interconnected to Delta/Three-Phase 208 V



Interconnected to Wye/Three-Phase 208 V

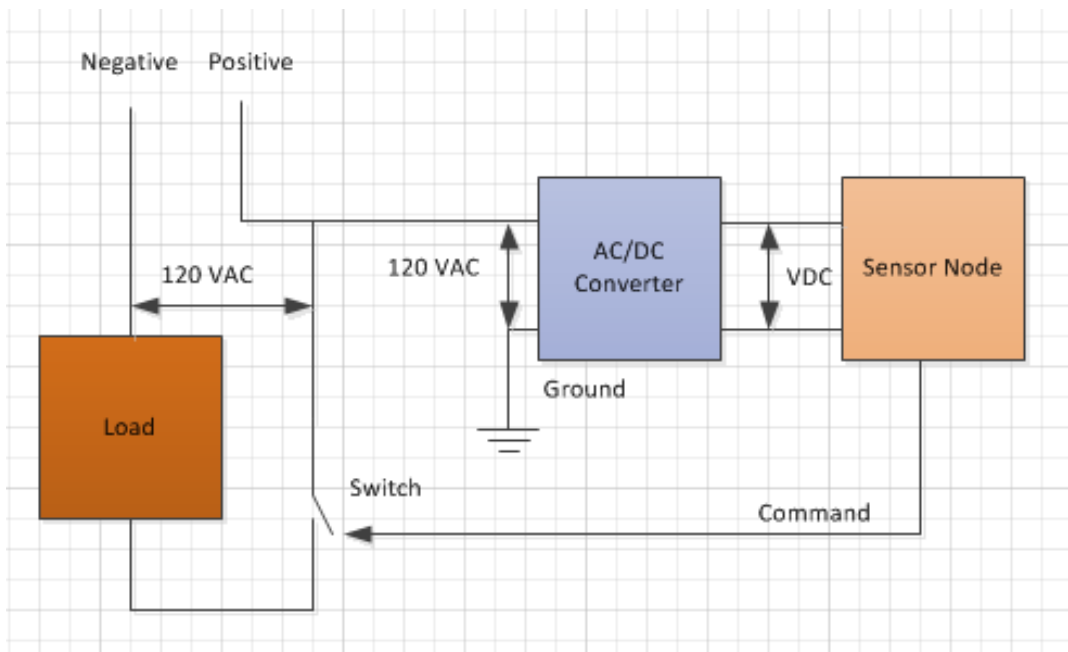
**Figure 1.3.** Step-Down Transformers for Residential Multiple Phase and Commercial with Potential Single-Phase Connection

Sensor nodes require direct current (DC) voltage to supply the electronic sensor and communication hardware. This voltage typically will range from several volts to 24 VDC depending on the design and needs of the hardware. The terminology of a “wired” sensor relates to the direct interconnection of the sensor node to the power delivery system of a building. As a result, the sensor node is typically interconnected to the single-phase 120 V alternating current (AC) source using a power adapter or an AC to DC converter. A 240 V interconnection also may be required. A depiction of this connection is shown in Figure 1.4



**Figure 1.4.** Electrical Interconnection to Sensor Node

In cases where a switch is combined with a sensor, the National Electric Code (NEC) may require that a neutral also be used as shown in Figure 1.3. The 2011 edition of the NEC introduced a provision 404.2.C that requires a grounded circuit conductor (neutral) to be included at all switch locations, with the following two exceptions: 1) when sufficiently sized conduit is used so that if the grounded circuit conductor were added the conduit would meet NEC code, or 2) when cable assemblies feeding the switch box are open through a framing cavity at the top, bottom, or same floor level, or through an unfinished floor, wall or ceiling.<sup>[8]</sup>



**Figure 1.5.** Electrical Interconnection to Sensor Node with a Switch Installed

Because the wired sensor type must be directly interconnected into the building power delivery system, there will be a limited number of available interconnection points in already constructed buildings. New interconnections in these buildings will require potentially long runs of wire and conduit for new sensor installations; therefore, the availability of sensors to particular locations may be limited. One option for interconnections includes using any existing outlets within an existing building although the availability may be limited<sup>[5]</sup>. Another approach is to plan the sensors into the original design of the building or adapt them into a restoration or remodeling project. Another potential option would be Power over Ethernet (PoE). The drive to use PoE started in the late 1990s with Voice over IP. Here, the power limit of 30 W for two pairs of wire has been suggested by IEEE 802.3<sup>[12]</sup>.

### 1.3 Characteristic Interdependencies

There are different relationships among the characteristics of a “wired” sensor. These relationships are described below:

- Significant sensing, processing, or communication will lead to higher energy requirements. As a result, the sizes of wired sensors typically are not limited.
- As this sensor node is attached to the direct supply of the building, building-wide outages also will deactivate the sensor node.
- Wired sensors will have limited reach to isolated locations; thus, they will have limited mobility. Hence, a wiring network may be limited with this approach.
- Wired sensors will be in the same location and rarely will be moved. Hence, the communication tends to be more robust. Wired sensors are not self-autonomous.
- Wired sensors typically are easy to maintain or will require very little maintenance.
- Wired sensors typically will not use wireless communication.
- Sensors requiring more than 30 W of continuous power will not use PoE.

### 1.4 Typical Applications

Applications where significant sensing, processing, or communication are required can expect significant energy requirements. As a result, a wired approach may be the most appropriate. Other applications include cases where a sensor controls a switch and can be used to replace existing infrastructure such as occupancy sensors.

### 1.5 Relative Cost Impacts

The cost of a wired system is primarily dependent on two key factors: 1) the degree of difficulty in routing the wire and meeting the appropriate electrical code requirements, shielding requirements, and wire support, and 2) the length of the cable<sup>[9]</sup>. As the first factor is directly related to retrofitting, new retrofits will see a significant cost reduction when non-wired solutions are adopted.

Wired solutions have been proven to be more expensive compared to wireless technology (see Table 1.1). For example, business operations can be disrupted when walls are penetrated to install wiring. These penetrations can require patching and repainting, thus increasing the time and labor costs required for the install <sup>[10]</sup>.

**Table 1.1. Comparison of Cost between Wireless and Wired Technologies**

Cost Component	Cost			
	In-Building Temperature Sensor Network		Monitoring System for Three Packaged HVAC Units	
	Wired Design	Wireless Design	Wired Design	Wireless Design
Sensors	\$1,800	\$3,000 <sup>1</sup>	\$636	\$636
Wiring	\$4,800 <sup>2</sup>	–	\$68 <sup>3</sup>	–
Communication and Signal-Conditioning Hardware	–	\$2,475	\$1,903	\$1,500 - \$5,900
Labor	– <sup>4</sup>	\$800	\$1,179 <sup>5</sup>	\$450
Total Cost	\$6,600	\$6,275	\$3,786	\$1,950 - \$7,000
Average Cost Per Sensor	\$220	\$209	\$316	\$163 - \$583

<sup>1</sup>Temperature sensors each with an integrated transmitter  
<sup>2</sup>Including labor for installation  
<sup>3</sup>Including conduit  
<sup>4</sup>Included in cost of wiring  
<sup>5</sup>Including installation of conduit

## 1.6 Summary/Conclusions

In this chapter, the interconnection of a sensor to a main power delivery system was discussed and interconnection through PoE was discussed.

## 1.7 References

1. R. Castilloux, “Sensors and Controls for BEMS: Providing the Neural Network to Net-Zero Energy,” Lux Research Report, March 2012.
2. B. Gholamzadeh, H. Nabovati “Concepts for Designing Low Power Wireless Sensor Network” Proceedings of World Academy of Science, Nov 2008, pp. 560.
3. M. Hemsptead, M. Lyons, D. Brroks, G. Wei, “Survey of Hardware Systems for Wireless Sensor Networks,” Journal of Low Power Electronics, Vol. 4, 1-10, 2008.
4. Roundy, Shad, et al., “Power Sources for Wireless Sensor Networks”, ”, Springer Berlin Heidelberg, 2004, pg 1-17
5. T. Hnat “The Hitchhiker’s Guide to Successful Residential Sensing Deployment” Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, 2011, pp. 232-245
6. Yen Kheng Tan and Sanjib Kumar Panda (2010). Review of Energy Harvesting Technologies for Sustainable WSN, Sustainable Wireless Sensor Networks, Yen Kheng Tan (Ed.), ISBN: 978-953-307-297-5, InTech, DOI: 10.5772/13062. Available from:

<http://www.intechopen.com/books/sustainable-wireless-sensor-networks/review-of-energy-harvesting-technologies-for-sustainable-wsn>

7. I. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "A Survey on Sensor Networks," IEEE Communications Magazine, Vol. 40, No. 8, Aug. 2002, pp. 102-114.
8. D. Behnke, "Impact of NEC 2011 Section 404.2(C) on Application of Occupancy Sensors" Available online at: [http://www.sensorswitch.com/Literature/WP\\_WSD\\_NGX.pdf](http://www.sensorswitch.com/Literature/WP_WSD_NGX.pdf)
9. M Kintner Meyer "Pros & Cons of Wireless" Ashrae Journal Nov. 2002, pp. 54-59.
10. EnOcean Alliance, "Whitepaper: Wireless Lighting Controls: A Total Cost Analysis," Available online.
11. D. Puccinelli, M. Haenggi, "Wireless Sensor Networks: Applications and Challenges of Ubiquitous Sensing," IEEE Circuits and Systems Magazine, Third Quarter 2005.

## 2.0 Power Generation: Source Excited

### 2.1 Introduction

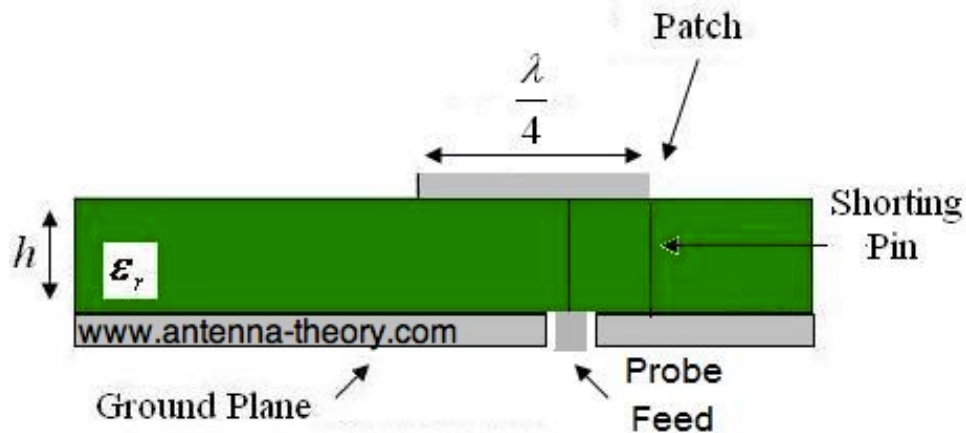
As a result of the extensive growth in portable communication devices, demand for antennas with multiple functionalities, low profiles, and smaller size in general has increased. As a result, there have been many advancements and variations in antenna design. Typical designs include patch antennas.

Wireless charging through inductive coupling is a technology that has been introduced.

### 2.2 Technical Details

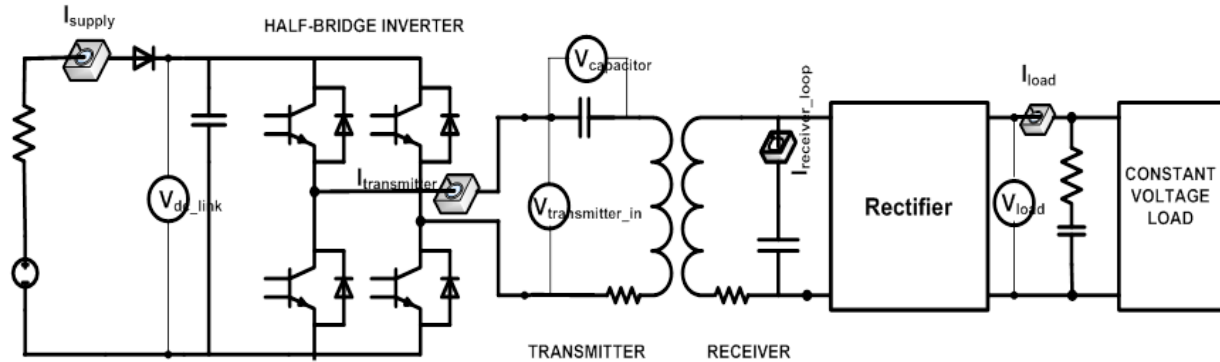
The Planar Inverted F-Antenna (PIFA) is composed of two basic components—a planar radiating element and a ground plane<sup>[1],[2]</sup>. The radiating element is parallel to the ground plane. This type of antenna is resonant at a quarter wavelength and resembles an “F”.

The PIFA is a variant of a monopole antenna where the radiating element has been made parallel to the ground plane. The resulting electric field that emerges after exciting the PIFA results from the radiating electric field and its ground plane. Highly favored for its smaller size and lower profile, the PIFA is highly favorable in wireless devices. PIFAs have many additional advantages such as being cost effective to manufacture, very easy to fabricate, higher bandwidth, and favorable electrical performance. PIFAs have multiple applications such as Bluetooth, Satellite Digital Multimedia Broadcasting (S-DMB), World Interoperability for Microwave (see Figure 2.1), Reconfigurable PIFA Antenna Access (WIMAX), and Wireless Local Area Network (WLAN). The reconfigurable PIFA used in this study was introduced in another research paper<sup>[1]</sup>. The advantage of this antenna is the presence of a ground plane. The ground plane absorbs the electric field incident on it and re-radiation is substantially decreased. A copper ground plane was used in this study. The switches used for this antenna<sup>[1]</sup> were modeled as perfect electric conductors. This antenna was chosen because the radiating part of the antenna is large and the presence of a ground plane allows less radiation of the power that it has already received.



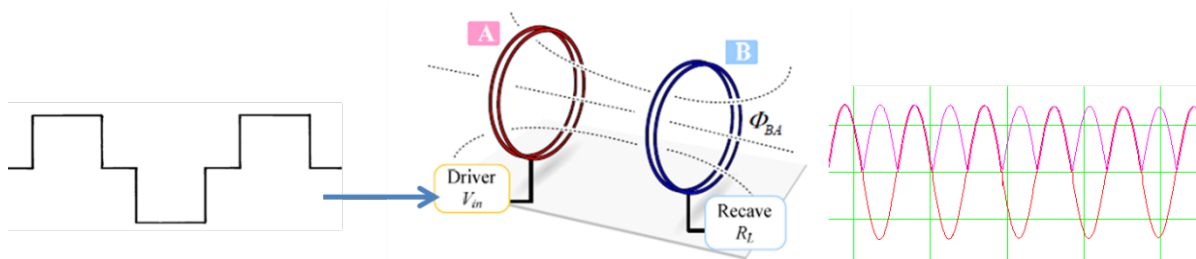
**Figure 2.1.** Schematic of a Planar Inverted F-Antenna

Wireless Power Transfer (WPT) is a flexible, convenient, safe, and autonomous charging method. The safety of this method stems beyond the inherent isolation of the source from the load via a large air gap between WPT transmit coil and sensor device. In addition, wireless charging can be done without wires. WPT charging is convenient and flexible not only because no cables and connectors are necessary but also because the charging becomes fully autonomous. Figure 2.2 represents an example of wireless charging system components.



**Figure 2.2.** Example of Wireless Power Transfer

In this system, the transmit side has a transmitter coil and a series tuning capacitor. The capacitor value is selected to create a resonance with the transmit coil inductor. The same approach is used for the receiver side, but the receiver side coil is tuned to the resonant frequency with a parallel tuning capacitor. Both transmit and receive sides are tuned to the same frequency. The inverter at the front side generates high frequency AC voltage, and the switching frequency of the inverter is set to the resonant frequency. Therefore, most of the power prefers to flow through the low impedance path provided by the resonant circuits. Thus, the right-side receiver coil captures most of the power transmitted by the left-side coil. A minimal amount of power that cannot be received is converted to the leakage field. After the receiver coil, a rectifier converts the induced voltage to DC power, which can be used to power the sensors. A depiction of the input and output waveforms are shown in Figure 2.3. Lee and co-workers<sup>[6]</sup> demonstrated 15 mW of power to six loads over a distance of 6.4 ft.



**Figure 2.3.** Wireless Power Transfer

### 2.2.1 Characteristic Interdependencies

- Wireless power transfer is inefficient over long distances.
- Wireless power transfer requires that the capacitor be chosen to meet a resonant frequency.

### 2.2.2 Typical Applications

The medical field has seen wireless power technology applied to satisfy the need for invasive power to electronic devices<sup>[5],[7]</sup>. Sensors in motion-based applications have used wireless power transfer as well<sup>[4]</sup>.

## 2.3 References

1. S. Punjala, K. Makki, "Wireless Powering of Sensors in Free Space and Concrete Medium," 13<sup>th</sup> International Symposium on Antenna Technology and Applied Electromagnetics and the Canadian Radio Sciences Meeting, 2009.
2. <http://www.antenna-theory.com/antennas/patches/pifa.php>
3. J. Albesa, M. Gasulla, T. Jager, L. Reindl, "Wireless Power Transmission for Autonomous Sensors in Removable Vehicle Seats,"
4. B. Griffin, C. Detweiler, "Resonant Wireless Power Transfer to Ground Sensors from a UAV
5. F. Zhang, S. Hackwoth, X. Liu, C. Li, M. Sun, "Wireless Power Delivery for Wearable Sensors and Implants in Body Sensor Networks
6. B. Lee, A. Hillenius, D. Ricketts, "Magnetic Resonant Wireless Power Delivery for Distributed Sensor and Wireless Systems,"
7. F. Zhang, S. Hackworth, X. Liu, H. Chen, R. Sciabassi, M. Sun, "Wireless Energy Transfer Platform for Medical Sensors and Implantable Devices."



## 3.0 Power Generation: Other Local

### 3.1 Introduction

Outside of typical source-excited generation technologies discussed in the previous chapter, other technologies for generating the appropriate power for sensor technologies have been explored. In many cases, these technologies are more experimental and typically are not applied in everyday sensor-based applications. The technologies discussed in this chapter include radioisotope generators and fuel cells.

### 3.2 Technical Details

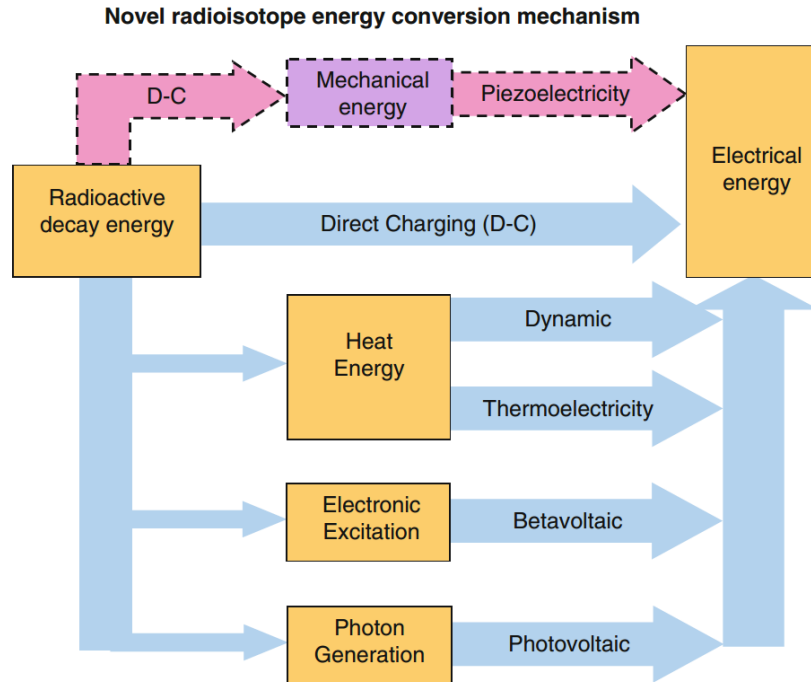
#### 3.2.1 Radioisotope Generator

A radioisotope generator produces electricity through radioactive decay. This process occurs in two phases: energy (e.g., heat) is derived from a radioisotope and then converted into electrical energy. There are a variety of different radioisotopes that can be used for this purpose:  $^3\text{H}$  (i.e., tritium),  $^{63}\text{Ni}$  (i.e., nickel-63), and  $^{147}\text{Pm}$  (Promethium-147). The characteristics of each of these isotopes are shown in Table 3.1<sup>[1]</sup>. The radiation energy emitted from radioisotopes listed in Table 3.1 are 1) alpha ( $\alpha$ ) radiation, which are doubly charged helium ions ( $E_{rad,avg} > 5 \text{ MeV}$ ); 2) beta ( $\beta$ ) radiation where  $\beta^-$  are electrons and  $\beta^+$  are positrons with  $E_{rad,avg}$  ranging from 5.3 keV to 2 MeV; and 3) gamma ( $\gamma$ ) radiation, which are photons with  $E_{rad,avg}$  being in the range of MeVs. These radioisotopes are currently being used to power autonomous sensor systems such as for environmental monitoring<sup>[2],[3]</sup>, civil infrastructure health monitoring<sup>[4]</sup>, and implantable medical devices<sup>[5],[6]</sup>. The energy densities of these radioisotopes range from 1 to 100 MJ/cc, and the radioisotope half-lives range from 1 to 100 years. When compared to conventional electrochemical (1 to 5 kJ/cc<sup>[7]</sup>) and hydrocarbon (10 to 20 kJ/cc) fuels, radioisotopes exhibit energy densities that are at most five times greater.

**Table 3.1.** Comparison of Radiation Properties of  $^3\text{H}$ ,  $^{63}\text{Ni}$ , and  $^{147}\text{Pm}$  Thin Films<sup>[1]</sup>

Parameters	$^3\text{H}$	$^{63}\text{Ni}$	$^{147}\text{Pm}$
Half-life (years)	12.4	100.3	2.64
Radiation type	$\beta^-$	$\beta^-$	$\beta^-$ , weak $\gamma$
$E_{rad,avg}$ (keV)	5.7	17.3	63.5
$E_{rad,max}$ (keV)	18.6	63	220
Highest available specific activity (curie/cc)	500	133	4300
Optimal layer thickness for $\eta_r > 75\%$ ( $\mu\text{m}$ )	0.1	0.8	6
Power density, ( $\mu\text{W}/\text{cm}^2$ )	0.7	0.45	827

Five primary techniques are used to convert radioisotope energy to electrical energy—thermoelectric engines; direct charge nuclear batteries; direct conversion batteries; indirect conversion batteries; and electro-mechanical batteries<sup>[1]</sup>. Figure 3.1 illustrates the different mediums before being converted to electrical energy<sup>[1]</sup>.



**Figure 3.1.** Schematic Illustrating Various Radioisotope Energy Conversion Techniques <sup>[1]</sup>

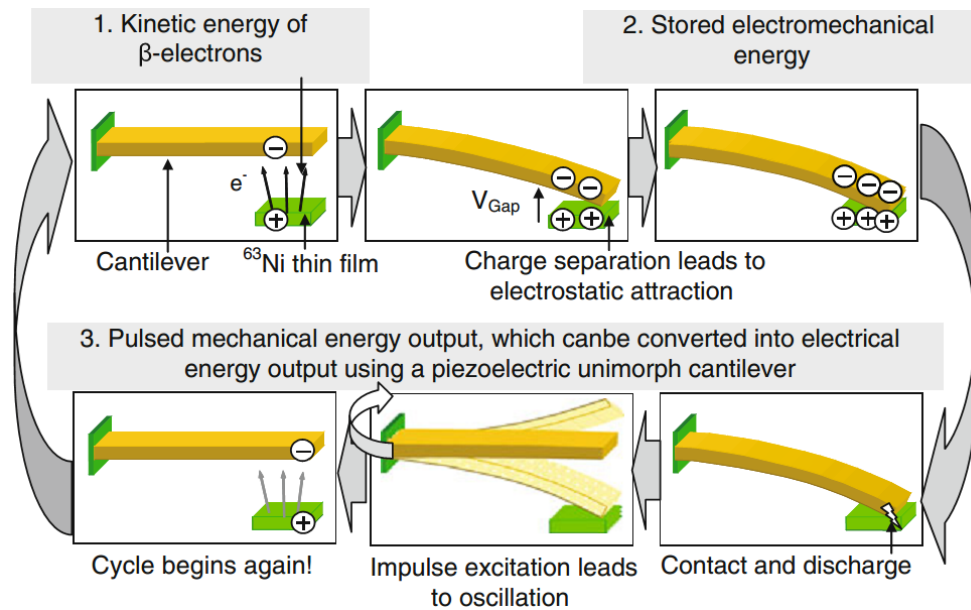
A thermoelectric engine converts the kinetic energy produced from radioisotopes into heat using collisions with target particles. Heat is then used to generate electricity using dynamic heat engines or direct conversion devices such as thermoelectric generators <sup>[8], [9]</sup>. However, thermoelectric generators are highly inefficient at the micro-scale because it is difficult to produce high temperatures or large heat gradients at that scale. Furthermore, to produce high temperatures requires large quantities of radioisotopes, thereby raising concerns about radiation safety. Therefore, sensor systems require other means of converting energy from radioisotopes to electrical energy besides using heat cycles.

A direct charging nuclear battery is composed of two electrodes made of electrically isolated concentric spheres, coaxial tubes, or parallel plates. Between the electrodes is space filled with vacuum or a dielectric material that does not absorb the radiation. The radioisotope is located on one of the electrodes as a thin-film coating. Charged particles emitted by the electrode coated with a radioisotope thin-film are collected to supply an electric load.

In a direct conversion battery, the radiation generated by the radioisotope is used to ionize atoms of a medium between two electrodes. A voltage gradient setup is used between both electrodes to separate and collect ionized charges that are supplied to any electrical load. For ionization of the working gas to be efficient, the size of the battery has to be large. Therefore, direct conversion batteries would not be good candidates for powering sensor systems.

An indirect conversion battery uses the kinetic energy of charged particles to irradiate radioluminescent materials, thereby generating light. Photovoltaic cells are used to convert the light into electrical energy. However, by using two processes to convert kinetic energy into electricity, the efficiency of both steps must be high to ensure good overall energy conversion efficiency. Another concern with this technique is using radiation sensitive photovoltaic components near radioisotopes.

Electro-mechanical conversion of sensor systems can be accomplished with  $^{147}\text{Pm}$ .  $^{147}\text{Pm}$  is needed because of its high power density ( $P_{g,dens}$  2.05 W/cc at 830 curie/g) <sup>[10]</sup> and long half-life ( $\tau_{1/2} = 2.6$  years). This method uses a radioisotope actuated reciprocating piezoelectric unimorph cantilever with integrated betavoltaics, <sup>[11]</sup> which is an extension of the radioisotope-powered, self-reciprocating cantilever <sup>[12], [13]</sup>. Figure 3.2 shows the operating principle of the self-reciprocating cantilever, <sup>[1]</sup> which uses a thin film of a radioisotope to electro-statistically actuate a cantilever. The pulsed mechanical energy produced by the cantilever can be converted into pulsed electrical energy by attaching a piezoelectric element to the cantilever.



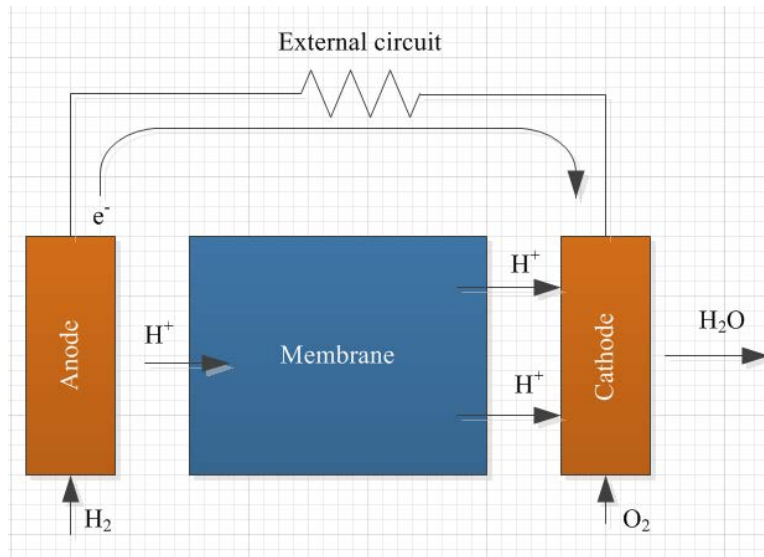
**Figure 3.2.** Schematic Illustrating the Radioisotope Actuation of Self-reciprocating Cantilevers, and the Generation of Electrical Energy when the Cantilever is a Piezoelectric Unimorph (Chapter 2 of Duggirala et.al. <sup>[1]</sup>)

The original invention of the fuel cell can be contributed to William Grove who, in 1839, discovered that electricity was produced when oxygen and hydrogen are combined in a particular configuration. Since that discovery, a variety of fuel cells have been developed <sup>[15]-[18]</sup>. The general classifications of fuel cells are based on the type of electrolyte used. The following is a list of the fuel cell types:

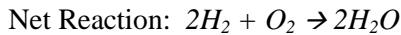
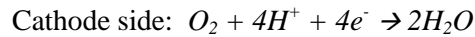
- Alkaline fuel cell (AFC)
- Phosphoric acid fuel cell (PAFC)
- Proton exchange membrane fuel cell (PEMFC)
- Molten carbonate fuel cell (MCFC)
- Solid oxide fuel cell (SOFC)
- Microbial fuel cell (MFC).

Figure 3.3 is a schematic showing the main components of a fuel cell—a cathode, an anode, and a bisecting membrane. At the anode, a catalyst separates the supplied hydrogen gas into electrons and positive hydrogen ions. The membrane permits the positive hydrogen ions to permeate from the anode to

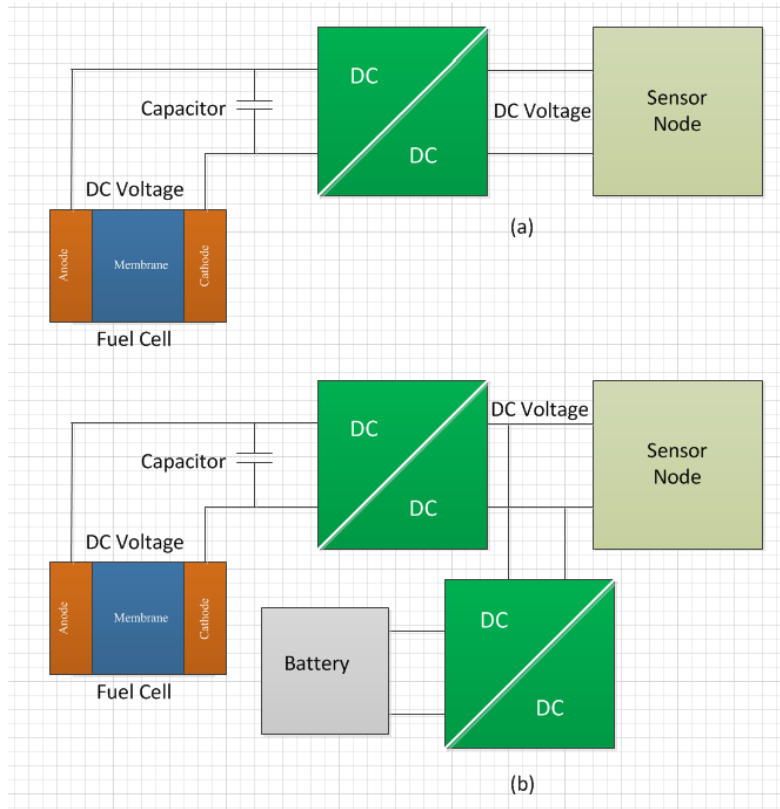
the cathode directly while rejecting the electrons to a provided electrical path. As the electrons and positive hydrogen ions reach the cathode, they recombine along with oxygen to produce water <sup>[18], [19]</sup>.



**Figure 3.3.** Depiction of a Fuel Cell



Fuel cells can further be classified based on operating temperature. High-temperature fuel cells, which include MCFCs and SOFCs, operate at temperatures above 500°C and react more readily and efficiently than their counterparts and do not require the use of a reformer. In addition, high-temperature fuel cells can function without an expensive noble metal catalyst such as platinum, thus reducing the overall cost of the device. Higher-temperature operation furthermore accommodates the use of different fuels and cogeneration. Nevertheless, high-temperature material and operational problems are severe and serve as deterrents for small-scale operations where quick startup is required. Thus, present high-temperature fuel cell applications have focused on small to large stationary power plants, where the efficiencies of internal reforming and co-generative capabilities outweigh the disadvantages of material breakdown and slow startup. Low-temperature fuel cells typically operate below 250°C and include AFC, PAFC, and PEMFC types. As these low temperatures do not permit internal reforming, an external source of almost pure hydrogen is essential. PEMFCs are the most likely fuel cells for typical sensor-based power supplies <sup>[16]</sup>. However, the difficulty in using a PEMFC for this type of application lies in the required size and power rating. Often these devices are termed “micro” fuel cells, and their sizes are in the millimeters. This requires the use of catalysts with high specific surface areas <sup>[22]</sup>.



**Figure 3.4.** Depiction of Fuel Cell Interconnections

Another type of fuel cell that appears to have potential application specifically in locations where water and microorganisms are available is the MFC <sup>[23]-[25]</sup>. These fuel cells extract bioenergy from the oxidation of microorganisms living in areas such as in sediments. The oxidation creates electrons and protons similar to a catalyst with hydrogen and uses the dissolved oxygen in water on the cathode side. The energy production of these fuel cells tends to be low (in mW levels) and often will require a storage capability to ensure the energy is available for sampling, processing, and communication. This should be done at a low frequency to ensure energy is available. This technology is not ideal for building technologies. With any fuel cell interconnection, a DC-DC boost converter will be necessary to convert the DC voltage from the fuel cell to the appropriate voltage. Numerous topologies exist and are discussed by Yu et al. <sup>[17]</sup>. In many cases, capacitors are recommended in the system configuration to increase the speed of response of the fuel cell. Several system interconnections are shown in Figure 3.4.

Because PEMFCs require pure hydrogen, hydrogen storage technologies are an important aspect. There are three ways to store hydrogen: 1) liquefied gas; 2) compressed gas; and 3) in metal hydrides. Each of these storage technologies has associated difficulties. The first technique is employed at large scales, but it requires extremely low temperatures. The second technology requires that the hydrogen be compressed to high pressures or be placed in extremely large containers (because hydrogen has a very low density). The third technique requires the combination of hydrogen gas with a chemical so when hydrogen is needed, a reverse chemical reaction is provided. This presents challenges associated with the type of material, cost, availability, and control of the recombination rates <sup>[18], [19]</sup>.

Another option is to generate hydrogen onsite through other generated sources that only capture energy intermittently such as photovoltaic or wind<sup>[18], [20]</sup>. This usually involves a medium stage of electrolysis, which will lead to additional losses. Electrolysis of water has been said to be 75 percent efficient<sup>[21]</sup>.

### 3.3 Characteristic Interdependencies

- Radioisotope Generators
  - Sensor systems need a high power density generation with long life.
  - The majority of sensor systems require 100  $\mu$ W to 100 mW.
  - Continuous power generation is needed for sensor and sensor system applications.
- Fuel Cells
  - Because of the low power production of MFCs, the frequency of measurements, processing, and communication should be low in the applied application.
  - Fuel cells produce DC, which must be controlled. Depending on the required output for the sensors, either a DC/DC converter or inverter will be needed.
  - As a result of the high cost of fuel cells, sensor-based applications are limited to applications where other energy sources will not work and the lifetime of the sensor is crucial.
  - Depending on the requisition approach for hydrogen gas, either a storage container or a reforming method such as electrolysis must be available to generate hydrogen. This will require another outside energy source.
  - In many cases, conversion stages and additional storage options are necessary, as fuel cells tend to have slow output response.
  - Fuel cells can have higher energy densities compared to batteries specifically when constructed at micro-sizes.

### 3.4 Typical Applications

Radioisotope generators are excellent sources of power for applications that need high power density and long-lasting power supplies in a sensor-system setting. Applications include environmental monitoring, civil infrastructure health monitoring, and implantable medical devices. For MFCs, applications are usually focused on areas where resources such as sunlight and other generation technologies are not available, but long lifetimes of sensors are required. Radioisotope generators and fuel cells also may become the choice to handle power needs when power density requirements are higher than in the case of batteries.

### 3.5 Relative Cost Impacts

It is difficult to gauge the cost of radioisotope generators because the few that are available for use in products such as heart pacemakers are not commercially available. When compared to other batteries, the

cost of radioisotope generators is significantly higher. This is because radioisotope generators are extremely long lasting and have high energy densities.

The cost of fuel cells is compared with other power generation technologies in Table 3.2 <sup>[26]</sup>. As shown in the table, fuel cells have the highest cost per kWh compared with nickel-cadmium (NiCd) batteries, gasoline and diesel-fueled generators, and electricity purchased from the electric grid. The price of fuel cells should decrease as research progresses and new manufacturing techniques are discovered <sup>[27]</sup>.

**Table 3.2.** Cost Comparison to Generate One Kilowatt of Energy <sup>[26]</sup>

Energy Source	Investment of Equipment to Generate 1kW	Lifespan of equipment before major overhaul or replacement	Cost of Fuel per kWh	Total Cost per kWh, including fuel, maintenance and equipment replacement
NiCd For portable use	\$7,000 based on 7.2V, 1000mAh at \$50/pack	1500h based on 1C discharge	\$0.15 electricity for charging	\$7.50
Gasoline Engine For mobile use	\$30 based on \$3,000/100kW (134HP)	4000 h	\$0.10	\$0.14
Diesel Engine For stationary use	\$40 based on \$4,000/100kW (134HP)	500 h	\$0.07	\$0.10
Fuel Cell For portable use	\$3,000 – 7,500	2000 h	\$0.35	\$1.85 – 4.10
For mobile use		4000 h	\$0.35	\$1.10 – 2.25
For stationary use		40,000 h	\$0.35	\$0.45 – 0.55
Electricity From electric grid	All inclusive	All inclusive	0.10	0.10

### 3.6 Summary/Conclusions

Other local generation technologies have been discussed in this chapter, specifically radioisotope generators and fuel cells. Advantages of these technologies include longevity of use (radioisotope generators) and high power densities (radioisotope generators and fuel cells). However, disadvantages of these technologies are radioisotope generators are not readily available and cost is high (radioisotope generators and fuel cells). Taking into consideration usability and cost for sensor and control applications in buildings, other power sources are more feasible. Further research is needed to make radioisotope generators and fuel cells viable options for application in buildings.

### 3.7 References

1. R. Duggirala, A. Lal, and S. Radhakrishnan, Radioisotope thin-film powered microsystems, Springer, ISBN: 978-1-4419-6762-6, 2010.
2. R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin, “Habitat monitoring with sensor networks,” Communications of the ACM, vol. 47, no. 6, pp.34-40, 2004.
3. K. c, J. Hart, and R. Ong, “Environmental sensor networks,” Computer, vol. 37, no. 8, pp. 50-56, 2004.

4. J. Lynch and K. Loh, "A summary review of wireless sensors and sensor networks for structural health monitoring," *Shock and Vibration Digest*, vol. 38, no. 2, pp. 91-128, 2006.
5. L. Olsen, "Beta voltaic energy conversion," *Energy Conversion*, vol. 13, no. 4, pp. 117-127, 1973.
6. G. Bloom, K. Murphy, and T. Cox, "An isotope-powered, indwelling intracranial pressure sensors," *Transactions of the American Nuclear Society*, vol. 13, no. 2, pp. 509-510, 1970.
7. D. Linden and T.B. Reddy, *Handbook of Batteries*. McGraw-Hill, 2002.
8. W.R. Corliss and D.G. Harvey, *Radioisotopic Power Generation*, Prentice-Hall, Englewood Cliffs, New Jersey, 1964.
9. K.E. Bower, Y.A. Barbanel, Y.G. Shreter, and G.W. Bohnert, *Polymers, Phosphors, and Voltaics for Radioisotope Microbatteries*. CRC Press, 2002.
10. H. Flicker, J. Loferski, and T. Elleman, "Construction of a promethium-147 atomic battery," *IEEE Transactions on Electron Devices*, vol. ED-11, no. 1, pp. 2-8, 1964.
11. A. Lal, R. Duggirala, and H. Li, "Pervasive power: a radioisotope-powered piezoelectric generator," *IEEE Pervasive Computing*, vol. 4, no. 1, pp. 53-61, 2005.
12. H. Li, A. Lal, J. Blanchard, and D. Henderson, "Self-reciprocating radioisotope-powered cantilever," *Journal of Applied Physics*, vol. 92, no. 2, pp. 1122-1127, 2002.
13. H. Li, "Radioisotope-powered self-reciprocating cantilever for micro power generation," Ph.D. dissertation, Cornell University, 2005.
14. DOE fuel cell handbook, EG&G Technical Services, Inc., DOE/NETL-2002/1179, 6th ed. pp. 2-9-2-14, 2002.
15. J. Larminie and A. Dicks, "Fuel cell systems explained," Wiley, UK, pp. 39-52, 2003.
16. G. Hoogers, "Fuel cell technology handbook," CRC Press, pp. 4-5-4-9, 2003.
17. X. Yu, M. Starke, L. Tolbert, and B. Ozipineci, "Fuel Cell Power Conditioning for Electric Power Applications: A Summary," *Electric Power Applications*, vol. 1, no. 5, pp. 643-656, September 2007.
18. F. Chraim and S. Karaki, "Fuel Cell Applications in Wireless Sensor Networks," *IEEE Instrumentation and Measurement Technology Conference*, pp. 1320-1325, May 2010.
19. R. Han, *Hydrogen Technology: Mobile and Portable Applications: Chapter 12 Development of Portable Systems*, Springer Berlin Heidelberg, ISBN 978-3-540-69925-4, 2008.
20. V. Devarja, S. Dinil, M. Diju, S. Vineeth, A. Jose, M. Jyothy, and M. Neethu, "PEM fuel cells as a power source for wireless sensor networks," *IEEE Symposium on Physics and Technology of Sensors*, March 7-10, 2012.
21. I. Papagiannakis, "Studying and Improving the Efficiency of Water Electrolysis using a Proton Exchange Membrane Electrolyser," Master thesis, Strathclyde University, 2005.
22. Z. Xiao, C. Feng, P. Chan, and I. Hsing, "Integrated proton exchange membrane micro fuel cells towards low power wireless sensor network applications," *IEEE International Conference on Portable Information Devices*, May 25-29, 2007.



23. F. Zhang, L. Tian, and Z. He, "Powering a wireless temperature sensor using sediment microbial fuel cells with vertical arrangements of electrodes," *Journal of Power Sources*, vol. 196, no. 22, pp. 9568-9573, November 2011.
24. B. Ringeisen, E. Henderson, P. Wu, J. Pietron, R. Ray, B. Little, J. Biffinger, and J. Jones-Meehan, "High power density from a miniature microbial fuel cell using shewanella oneidensis DSP 10," *Environmental Science and Technology*, vol. 40, pp. 2629-2634, 2006.
25. C. Donoval, A. Dewan, D. Heo, and H. Beyenal, "Batteryless, wireless sensor powered by a sediment microbial fuel cell," *Environmental Science and Technology*, vol. 42, no. 22, pp. 8591-8596, 2008.
26. I. Buchmann. (February 2001). Batteries in the Portable World. The fuel cell: is it ready? [website]. Available: <http://www.buchmann.ca/Article1-page1.asp>.
27. D. Quick. (October 9, 2009). gizmag. A nuclear battery the size and thickness of a penny [website]. Available: <http://www.gizmag.com/smaller-nuclear-battery/13076/>.

## 4.0 Power Generation: Batteries

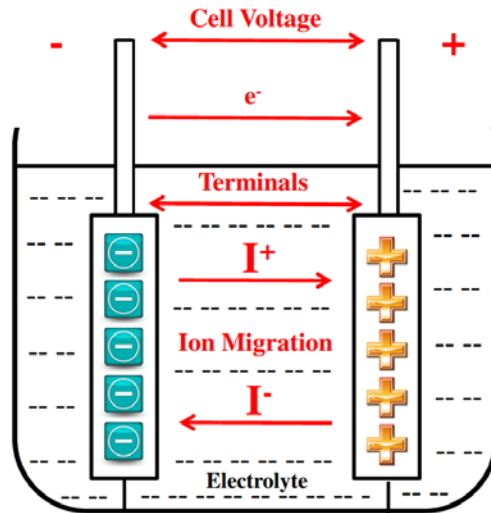
### 4.1 Introduction

The U.S. Department of Energy (DOE), through its Buildings Technology Office, is initiating a new program focused on sensors and controls for improving energy efficiency in buildings. Sensors are used to measure a physical property and generate a signal based on quantity. In a control system, sensors are used to relay environmental information to help in decision-making for management and command of a system. An important consideration when dealing with sensors and controls is power. This chapter aims to investigate storage batteries, which are widely used to power sensors and controls.

The storage battery, invented in 1800 by Italian physicist Alessandro Volta, was made of multiple electrochemical cells connected in series. Each cell produced electrical energy by converting stored chemical energy. Batteries were not adopted early because of their inconsistent current output. This changed in 1836 when British chemist John Daniell invented the Daniell cell, which used glass jars with liquid electrolytes to produce electrical energy. Throughout the years, batteries have evolved, replacing liquid electrolytes (i.e., liquid cell) with dry paste (i.e., dry cell) in cells. Because of their improved practicality, batteries are widely used in daily life; according to the Battery University, the battery market worldwide is approximately \$50 billion in U.S. dollars <sup>[1]</sup>.

There are two types of batteries that have the potential of being used in sensor and control systems for building applications—primary and secondary batteries. In primary batteries, the chemical reaction to produce electrical energy is irreversible, making these batteries disposable. This type of battery is generally used in devices with low current drain, typically away from alternative power sources. Common primary batteries found on the market are: zinc-carbon and alkaline batteries. Secondary batteries have the ability to be reused by supplying an external source of electrical energy (i.e., using a charger) to reverse chemical reactions during its normal use. An advantage with using primary batteries compared with secondary batteries is the higher energy density (energy stored per unit volume). A disadvantage of primary batteries is performance degrades with high-drain applications with loads under  $75 \Omega$  <sup>[2]</sup>. Secondary batteries also can be used in conjunction with energy harvesting and source excited technologies to ensure 24-hr energy delivery to the various sensor designs.

A diagram of an electrochemical cell is shown in Figure 4.1 <sup>[3]</sup>. When the cell is discharging, a negative current flows from the positively charged electrode (cathode) into the electrolyte. This reducing current flow is known as cathodic. When the cell is charging, positive current that flows in the opposite direction from the negatively charged electrode (anode) into the electrolyte is known as anodic. The term battery is made of one (monocell) or multiple (monobloc) electrochemical cells shown in Figure 4.1. When multiple cells are present, connections are in parallel and/or in series. Active materials in the cell are components responsible for cell reactions, including the electrodes and electrolyte. When chemical conversion occurs on active materials from charging and discharging, physical volume changes often occur. Non-active components of the battery cell include conducting components, separators, containers, terminal seals, vents, and valves.



**Figure 4.1.** Electrochemical Cell <sup>[3]</sup>

There are a variety of different parameters used to characterize a battery, specifically voltage, capacity, energy content, specific energy, energy density, internal resistance, self-discharge, and cycles. There are four types of voltage measurements for batteries—float, nominal, charge, and discharge voltages <sup>[4]</sup>. Float voltage is measured when the current is zero. The nominal voltage of a battery is the designated operating voltage. The charge and discharge voltages are voltages while the battery is charging and discharging, respectively. The capacity is the amount of electric charge in units of Ampere-hours (Ah) that the battery can produce. The energy content is the amount of energy in Watt-hours (Wh) the battery can produce. Specific energy and energy density are the relationships of energy content either to weight (e.g., Watt-hour/kilogram) or volume (e.g., Watt-hour/cubic centimeters) of a battery, respectively. The internal resistance is a battery characteristic that shows the capability to handle a certain load. Self-discharge is loss of charge from either electrode when a battery is idle. Cycles are the amount of times a battery is charged and discharged.

## 4.2 Technical Details

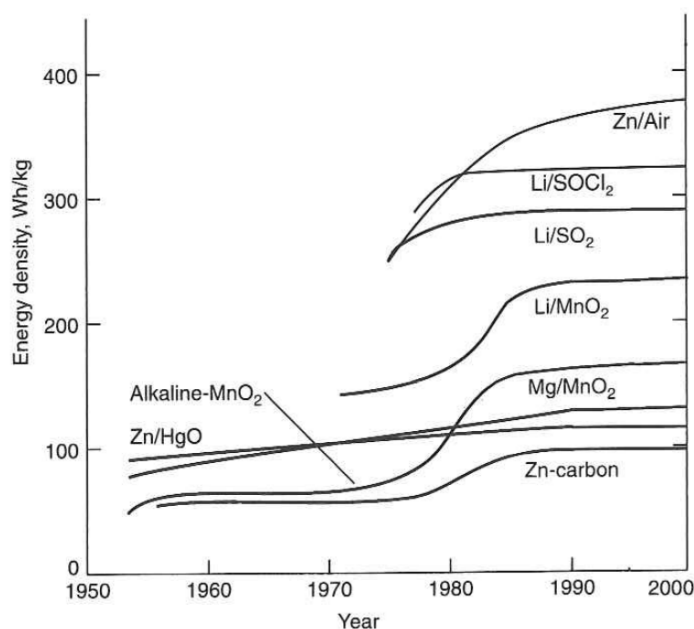
In this section, we identify batteries that are ideal for use with sensors and control systems in buildings. Batteries are currently the most common power source for wireless sensors. There are three different battery types identified for primary batteries and two for secondary batteries. Technical details are discussed for each type.

### 4.2.1 Primary Batteries

There are 10 types of primary batteries, and the characteristics and typical applications of each type are shown in Table 4.1 <sup>[5]</sup>. When dealing with sensors and controls applied to buildings, the most important characteristics for batteries include a high energy density, good shelf life (i.e., minimal self-discharge), and portability (i.e., size). Out of these 10 batteries listed in Table 4.1, alkaline, zinc-air; and lithium batteries were chosen for further review. These three battery types have been proven to have the highest energy densities as shown in Figure 4.2 <sup>[5]</sup>.

**Table 4.1.** Major Characteristics and Applications of Primary Batteries <sup>[5]</sup>

System	Characteristics	Applications
Zinc-carbon (Leclanché) Zinc/MnO <sub>2</sub>	Common, low-cost primary battery; available in a variety of sizes	Flashlight, portable radios, toys, novelties, instruments
Magnesium (Mg/MnO <sub>2</sub> )	High-capacity primary battery; long shelf life	Military receiver-transmitters, aircraft emergency transmitters
Mercury (Zn/HgO)	Highest capacity (by volume) of conventional types; flat discharge; good shelf life	Hearing aids, medical devices (pacemakers), photography, detectors, military equipment, but in limited use because of the environmental hazard of mercury
Mercad (Cd/HgO)	Long shelf life; good low- and high-temperature performance; low energy density	Special applications requiring operation under extreme temperature conditions and long life; in limited use
Alkaline (Zn/alkaline/MnO <sub>2</sub> )	Most popular general-purpose premium battery; good low-temperature and high-rate performance; moderate cost	Most popular primary battery; used in a variety of portable battery-operated equipment
Silver/zinc (Zn/Ag <sub>2</sub> O)	Highest capacity (by weight) of conventional types; flat discharge; good shelf life, expensive	Hearing aids, photography, electric watches, missiles, underwater and space application (larger sizes)
Zinc/air (Zn/O <sub>2</sub> )	Highest energy density, low cost; not independent of environmental conditions	Special applications, hearing aids, pagers, medical devices, portable electronics
Lithium/soluble cathode	High energy density; long shelf life; good performance over wide temperature range	Wide range of applications (capacity from 1 to 10,000 Ah) requiring high energy density, long shelf life (e.g., from utility meters to military power applications)
Lithium/solid cathode	High energy density; good rate capability and low-temperature performance; long shelf life; competitive cost	Replacement for conventional button and cylindrical cell applications
Lithium/solid electrolyte	Extremely long shelf life; low-power battery	Medical electronics, memory circuits, fusing



**Figure 4.2.** Advances in Development of Primary Batteries. Continuous discharge at 20°C; 40- to 60-hour rate; AA or similar size battery. <sup>[5]</sup>

## 4.2.2 Alkaline Batteries

The alkaline-manganese dioxide (zinc/KOH/MnO<sub>2</sub>) battery was introduced in the early 1960s. Since then, this battery type has become the dominant primary battery used for portable devices. This is mostly due to the advantages compared to its competitor the zinc-carbon (i.e., Leclanché) battery. These advantages include higher energy density, better service performance (e.g., continuous and intermittent, low and high rate, ambient and low temperature), lower internal resistance, longer shelf life, greater resistance to leakage, and better dimensional stability<sup>[5]</sup>.

The two types of alkaline-manganese dioxide batteries—cylindrical and button—are described in Table 4.2. For the cylindrical type of battery, there are the typical nomenclatures such as AA, AAA, C, D, etc. Manufacturers of button batteries define their own codes; however, there is nomenclature that has been defined by different standards agencies such as the International Electrotechnical Commission (IEC) and the American National Standards Institute (ANSI). The appropriate battery designs should be chosen based on the sensor and/or control system component application and power requirements.

**Table 4.2.** Characteristics of 1.5-V Standard Alkaline-Manganese Dioxide Batteries<sup>[5]</sup>

Size	IEC	ANSI	Capacity, Ah	Diameter, mm	Nominal dimensions		
					Height, mm	Weight, g	Volume, cm <sup>3</sup>
Cylindrical Types							
AAAA	LR61	25A	0.56	8	42	65	2.2
N	LR1	910A	0.8	12	29	9	3.3
	LR50		0.56	16	16	11	3.6
AAA	LR03	24A	1.1-1.25	10	44	11	3.8
AAA	LR61	15A	2.5-2.85	14	50	23	7.5
C	LR14	14A	7.1-8.4	26	50	66	26
D	LR20	13A	14.3-18	34	61	138	54.4
F	LR25		22	33	91	200	80
Button Types							
	LR41		0.035	7.9	3.6	0.6	0.2
	LR43	1167A	0.1	11.6	4.2	1.4	0.3
	LR44	1166A	0.145	11.6	5.4	2.3	0.5
	LR48		0.06	7.9	5.4	0.9	0.3
	LR53	1129AP	0.16	23	5.9	6.8	2.3
	LR54	1168A	0.072	11.6	3.1	1.1	0.3
	LR55	1169A	0.04	11.6	2.1	0.9	0.2

## 4.2.3 Zinc/Air Batteries

Heise and Schumacher commercialized the zinc/air battery in the early 1930s. Their battery was an alkaline electrolyte zinc/air battery with saturated porous carbon air cathodes to prevent flooding. Oxygen from the atmosphere is used to produce electrochemical energy. The air defuses into the battery

where it is used as the cathode reactant. The air comes into contact with the cell's electrolyte at the active surface within the interior cathode. This interaction promotes the reduction of oxygen in the presence of an aqueous alkaline electrolyte <sup>[5]</sup>.

Benefits of zinc/air batteries include high energy density, flat discharge voltage, long shelf life (sealed), no ecological issues, low cost, and their capacity is independent of load and temperature when within the design operating range. Disadvantages of this battery type include limited power output, short activated life, and dependence on environmental conditions <sup>[5]</sup>.

The primary forms of zinc/air batteries discussed in this chapter are button and coin batteries, which have capacities ranging from 42 mAh to 1100 mAh (see Table 4.3). However, larger forms of zinc/air batteries such as cylindrical round cells (e.g., AA and AAA cells) are available. Typical applications of smaller sized zinc/air batteries include hearing aids, while larger sized batteries have generally been designed for pagers or telemetry devices. In general, these batteries are used in continuous-drain applications.

**Table 4.3.** Characteristics of Zinc/Air Button and Coin Batteries <sup>[5]</sup>

<i>a. Standard Single-Cell Batteries, 1.4 Volts</i>									
Generic type	IEC No.	ANSI no.	Nominal diameter, (mm)	Nominal height, (mm)	Average weight, (g)	Rated capacity, (mAh)	Standard drain, (mA)	Limiting current, (mA)	Typical useful service life, (months)
5*	PR63		5.7	2		42	0.4		1-2
10*	PR70	7005ZD	5.7	3.5	0.3	70	0.4	2	1-2
312*	PR41	7002ZD	7.7	3.9	0.6	134	0.8	7	1-2
13*	PR48	7000ZD	7.7	5.2	0.9	260	0.8	12	1-2
675*	PR44	7003ZD	11.4	5.2	1.8	600	2	22	2-3
2330**	PR2330		23.2	3		960	4		1-2
630**	PR1662		16	6.2	3.5	1100	4		2-3

(\*) **Source:** Duracell, a Gillette Company). <sup>[5]</sup>

(\*\*) **Source:** Panasonic Matsushita) Product literature (coin batteries). <sup>[5]</sup>

<i>b. High Capacity Zinc/Air Single-cell Batteries, 1.4 Volts</i>				
Generic type	ANSI no.	Max. ANSI diameter, (mm)	Max. ANSI height, (mm)	Rated capacity, (mAh)
10	7005ZD	5.8	3.6	90
312	7002ZD	7.9	3.6	160
13	7000ZD	7.9	5.4	290
675	7003ZD	11.6	5.4	630

**Source:** Rayovac Corporation, Ultra Hearing Aid Batteries. <sup>[5]</sup>

#### 4.2.4 Lithium Batteries

The use of lithium as the anode in non-aqueous primary batteries began in the 1960s; the first use of this battery technology was in the 1970s in select military applications. Since then, lithium has become an attractive option for anode material because of characteristics such as low weight, high voltage, high electrochemical equivalence, and good conductivity. Other advantages of lithium primary batteries are the availability of a variety of energy storage capacities (ranging from less than 5 mAh to 10,000 Ah) and physical configurations (small coin cells, cylindrical cells, and large prismatic cells).

There are many advantages of using lithium anodes for primary batteries compared with conventional batteries. Lithium batteries have much higher cell voltages (i.e., up to 4 V) compared with other primary battery systems (i.e., 1.5 V). Lithium batteries also have higher specific energies and energy densities, where the energy output is two to four times larger than that of conventional zinc anode batteries. The temperature range (i.e.,  $-80^{\circ}\text{C}$  to  $150^{\circ}\text{C}$ ) tends to be much wider with lithium batteries as well. Other advantages of lithium battery technologies include higher power density, the ability to deliver high current and power levels, flat discharge characteristics (the battery gives a constant voltage and resistance through discharging), and superior shelf life (lithium batteries can be stored for 10 years at room temperature). Table 4.4 identifies the different lithium primary battery technologies and characteristics such as construction features, electrical characteristics, and available sizes.

**Table 4.4.** Characteristics of Lithium Primary Batteries <sup>[5]</sup>

Soluble cathode batteries												
System	Cathode	Electrolyte		Separator	Construction	Voltage, V		Specific energy† Wh/kg	Energy density† Wh/L	Power density	Discharge profile	Available sizes
		Solvent	Solute			Nominal	Working* (20°C)					
Lithium/sulfur dioxide (Li/SO <sub>2</sub> )	SO <sub>2</sub> with carbon and binder on Al screen	AN	LiBr	Microporous Polypropylene	Spiral "jelly-roll" cylindrical construction; glass-to-metal seal	3.0	2.9–2.7	260	415	High	Very flat	Cylindrical batteries up to 35 Ah
Lithium/thionyl chloride (Li/SOCl <sub>2</sub> ) Low rate	SOCl <sub>2</sub> with carbon and binder on Ni or SS	SOCl <sub>2</sub>	LiAlCl <sub>4</sub>	Glass non-woven	Wafer construction	3.6	3.6–3.4	275	630	Low	Flat	0.4–1.7 Ah
					"Bobbin" in cylindrical construction	3.6	3.5–3.3	590	1100	Medium	Flat	Cylindrical batteries 1.2–19
					Prismatic with flat plates	3.6	3.5–3.3	480	950	Medium	Flat	12–10,000 Ah
High capacity												
High rate				Glass mat	Spiral "jelly-roll" cylindrical construction or flat disk	3.6	3.5–3.2	380	725	Medium to high	Flat	Cylindrical: 5–23 Ah Flat disk: up to 320 Ah
					Spiral "jelly-roll" cylindrical construction	3.9	3.8–3.3	450	900	Medium	Flat	2–30 Ah
Lithium/sulfuryl chloride (Li/SO <sub>2</sub> Cl <sub>2</sub> )	SO <sub>2</sub> Cl <sub>2</sub> with carbon and binder SS screen	SO <sub>2</sub> Cl <sub>2</sub> (some with additives)	LiAlCl <sub>4</sub>	Glass	Spiral "jelly-roll" cylindrical construction; glass-to-metal seal	3.95	3.5–3.1	450	900	Medium to high	Flat	7–30 Ah



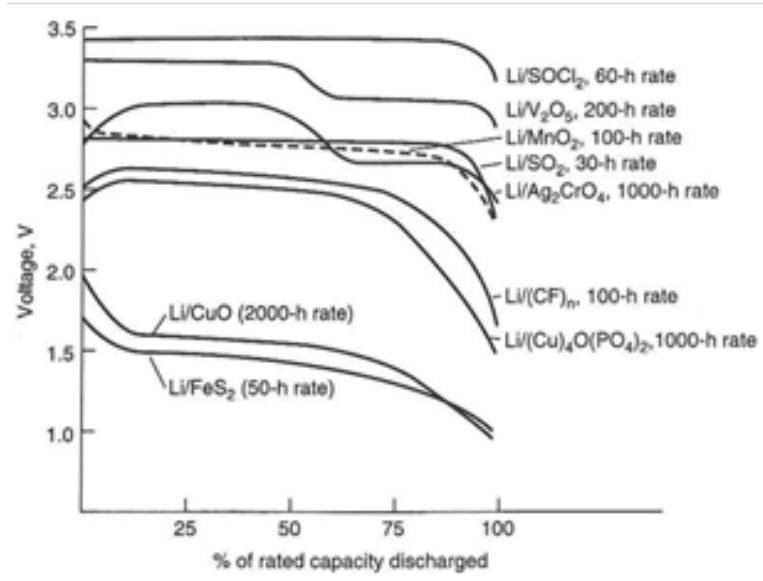
Solid cathode batteries												
Lithium/carbon monofluoride  (Li(CF) <sub>n</sub> )	CF with carbon and binder on nickel collector	PC + DME or BL	LiBF <sub>4</sub> or LiAsF <sub>6</sub>	Polypropylene	"Coin" construction crimped seal Pin type  Spiral "jelly-roll" cylindrical construction crimped or glass-to-metal seal Rectangular with flat plates	3.0	2.7–2.5	215	550	Low to medium Low	Moderately flat  Humped	Coin batteries to 500 mAh Small cylinders 25–50 mAh Cylindrical batteries to 5 Ah (commercial) and 1200 Ah (military)  Rectangular batteries to 40 Ah Cylindrical batteries 500–3500 mAh
Lithium/copper oxide (Li/CuO)	CuO pressed in cell can	1,3D	LiClO <sub>4</sub>	Nonwoven glass	"Bobbin" inside-out cylindrical construction: "Jelly-roll" cylindrical construction: crimped seal	1.5	1.5–1.4	280	650	Low	High initial voltage drop, then moderately flat	AA-size
Lithium/iron disulfide (LiFeS <sub>2</sub> )	FeS <sub>2</sub>					1.5	1.6–1.4	260	500	Medium to high	High initial drop, then flat	AA-size
Lithium/manganese dioxide (Li/MnO <sub>2</sub> )	MnO <sub>2</sub> with carbon and binder on supporting grid	PC + DME	Li salt	Polypropylene	"Coin construction with flat electrodes	3.0	3.0–2.7	230	545	Low to medium	Moderately flat	Coin batteries 65–1000 mAh
		Organic solvent	Li salt	Polypropylene	"Jelly-roll" cylindrical construction; crimped and hermetic seals	3.0	2.8–2.5	230	535	Medium to high	Moderately flat	$\frac{3}{4}$ A Cylindrical batteries typical, larger cells available to 33 Ah
		Organic solvent	Li salt	Polypropylene	"Bobbin" cylindrical construction	3.0	3.0–2.8	270	620	Low to medium	Moderately flat	Cylindrical batteries to 1.75 Ah
Lithium/silver vanadium oxide (Li/AgV <sub>2</sub> O <sub>11</sub> )	AgV <sub>2</sub> O <sub>5.5</sub> with graphite and carbon	PC, DME	LiAsF <sub>6</sub>	Microporous polypropylene	Rounded prismatic and D-shaped cross section	3.2	3.2–1.5	270	780	Low to medium	Multiple plateaus	Special sizes for implantable medical devices

The numerous types of lithium batteries that are available are listed in Table 4.5. The first commercially available and most popular lithium battery is lithium/manganese dioxide (Li/MnO<sub>2</sub>). Its popularity is a result of the relatively low cost, the excellent shelf life, good range of operating temperatures, and availability of coin and cylindrical cells. Another popular battery technology is the lithium/carbon monofluoride (Li{CF}<sub>n</sub>) battery, which is one of the early solid-cathode batteries. The attractiveness of this battery technology is a result of its high theoretical capacity and flat discharge rate. This battery type also is available in coin, cylindrical, and prismatic configurations.

**Table 4.5.** Characteristics of Typical Lithium/Solid-Cathode Batteries <sup>[5]</sup>

Type of Battery	Operating Voltage, V	Characteristics
Li/MnO <sub>2</sub>	3.0	High specific energy and energy density; wide operating temperature range (-20 to 55°C); performance at relatively high discharge rates; minimal voltage delay; relatively low cost; available in flat (coin) and cylindrical batteries (high and low rates)
Li/(CF) <sub>n</sub>	2.8	Highest theoretical specific energy, low- to moderate-rate capability; side operating temperature range (-20 to 60°C); flat discharge profile; available in flat (coin), cylindrical and prismatic designs.
Li/Cu <sub>4</sub> O(PO <sub>4</sub> ) <sub>2</sub>	2.5	High specific energy; long storage life; operating temperature range up to 175 °C; low- to moderate-rate capability; not currently available.
Li/CuO	1.5	Highest theoretical volumetric coulombic capacity (Ah/L); long storage life; low- to moderate-rate capability; operating temperature range up to 125 to 150 °C; no apparent voltage delay. Potential replacement for alkaline-manganese but not currently available.
Li/FeS <sub>2</sub>	1.5	Replacement for conventional zinc-carbon and alkaline-manganese dioxide batteries and better low-temperature performance and storability. Currently available in AA size as a direct replacement for alkaline-manganese
Li/Ag <sub>2</sub> CrO <sub>4</sub>	3.1	High voltage, high specific energy and energy density; low-rate capability; high reliability; used in low-rate, long-term applications; high cost
Li/AgV <sub>2</sub> O <sub>5.5</sub>	3.2	High specific energy and energy density multiple-step discharge; good rate capability; used in implantable and other medical devices
Li/V <sub>2</sub> O <sub>5</sub>	3.3	High energy density; two-step discharge; used in reserve cells (Chap. 20)

Typical discharge curves for solid cathode lithium batteries are shown in Figure 4.3. The lithium/silver chromate (Li/SOCl<sub>2</sub>) battery has a characteristic discharge curve that is relatively flat and at a high voltage. However, the high cost of this battery technology has limited its use to a select few applications. The lithium/vanadium pentoxide (LiV<sub>2</sub>O<sub>5</sub>) battery has a large volumetric energy density and a two-step discharge profile. The main application has been for reserve batteries. The lithium/silver vanadium oxide (Li/AgV<sub>2</sub>O<sub>5.5</sub>) battery has a high-rate of discharge and has been primarily used for medical applications where pulse load requirements exist. Other solid-cathode lithium batteries not mentioned operate at 1.5 V and were developed to replace traditional 1.5 V button or cylindrical cells. These battery types include the lithium/copper oxide (Li/CuO) battery, which has a high coulombic energy density and a higher capacity and lighter weight when compared with traditional cylindrical cells. Also, this technology is able to operate in high temperature environments and has a long shelf life even under extreme conditions. A similar battery type is the lithium/iron disulfide (Li/FeS<sub>2</sub>) battery.



**Figure 4.3.** Typical Discharge Curves of Lithium/Solid-Cathode Batteries <sup>[5]</sup>

**Table 4.6.** Characteristics of Larger Li/MnO<sub>2</sub> Batteries. Standard type: C/20 rate at room temperature to 2.0 V cutoff. High-Rate Type: C/10 rate at room temperature to a 2.0 V cutoff <sup>[5]</sup>

IEC cell size	Nominal capacity typical values		Dimensions		
	(Ah)	Max. continuous current* (mA)	Dia. (mm)	Ht. (mm)	Weight
M03	0.2	40	14.5	25	8
M04	0.3	60	14.5	25	9
M49	1.6	300	22.5	32	24
M52	4.5	800	26	51	57
M52 HR	4.0	1200	26	51	59
M56	5.6	900	26	60	69
M56 HR	5.5	1500	26	60	71
M19 HR	9.0	2000	33.5	58	103
M20	10.5	2000	34	61	115
M20 HR	10.0	2500	34	61	117
M24 HR	20.0	4000	33.5	111	201
M58	11.0	2000	42	51	125
M25	33.0	4000	42	125	355
M62	33.0	5000	42	133	355

Source: FRIWO Silberkraft data sheets.

**Table 4.7.** Specifications for Commercially Available Foil Cells and Batteries <sup>[5]</sup>

41	Dimension's thickness x L x W (mm)	Voltage (OCV) {volts}	Capacity {mAh}	Max. cont. current {mA}	Weight {g}
U3VF-A-T	0.8 x 38.6 x 30.0	3.2	27	20	1.1
U3VF-B-T	0.5 x 27.7 x 38.1	3.2	45	40	1.2
U3VF-G-T	1.1 x 36.3 x 25.7	3.2	120	12	1.1
U3VF-L-T	2.0 x 38.6 x 30.0	3.2	160	20	2.1
U6VF-K2	4.6 x 73.7 x 46.0	6.4	600	50	17.5
U3VF-K	2.3 x 73.7 x 46.0	3.2	800	55	8.5
U3VF-H	2.3 x 93.9 x 76.0	3.2	1800	60	18.0
U6VF-H2	3.9 x 93.2 x 78.2	6.4	1800	60	37.0
U3VF-D	2.3 x 92.0 x 92.0	3.2	2300	80	27.5

Source: Manufacturer's Data Sheets

### 4.3 Secondary Batteries

There are 15 types of secondary batteries shown in Table 4.8 <sup>[5]</sup>, and the characteristics and applications of each battery type are identified. When dealing with sensors and controls applied to buildings, the most important characteristics a battery must have are high energy density, good shelf life, and portability (i.e., size). Out of the 15 battery types, two were chosen for further review—nickel-cadmium and lithium-ion. These two batteries are shown to have the highest energy densities <sup>[5]</sup>.

**Table 4.8.** Major Characteristics and Applications of Secondary Batteries <sup>[5]</sup>

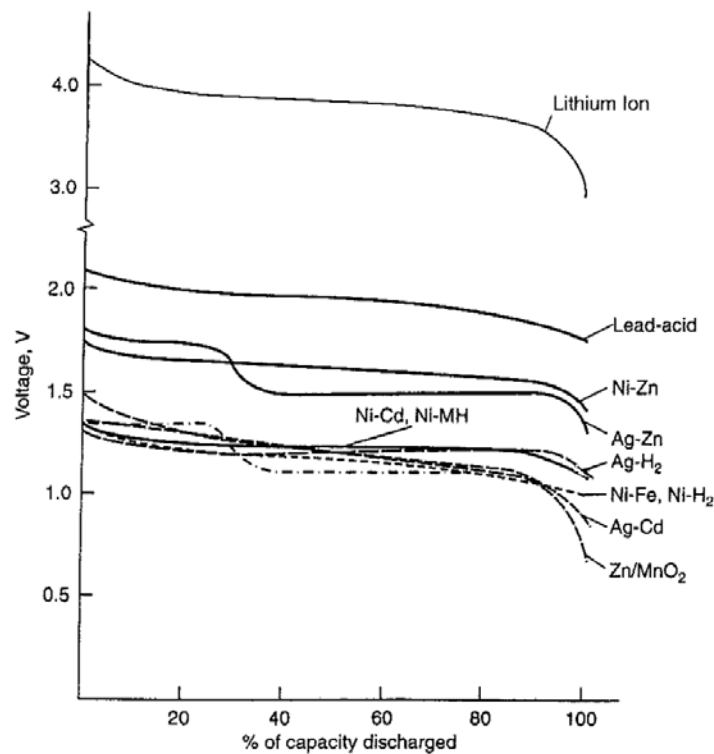
System	Characteristics	Applications
Lead-acid: Automotive	Popular, low-cost secondary battery, moderate specific energy, high-rate, and low-temperature performance; maintenance-free designs	Automotive SLI, golf carts, lawn mowers, tractors, aircraft, marine
Traction (motive power)	Designed for deep 6-9 h discharge, cycling service	Industrial trucks, materials handling, electric and hybrid electric vehicles, special types for submarine power
Stationary	Designed for standby float service, long life, VRLA designs	Emergency power, utilities, telephone, UPS, load leveling, energy storage, emergency lighting
Portable	Sealed, maintenance-free, low cost, good float capability, moderate cycle life	Portable tools, small appliances and devices, TV and portable electronic equipment
Nickel-cadmium: Industrial and FNC	Sealed, maintenance-free, good high-rate low-temperature performance, excellent cycle life	Aircraft batteries, industrial and emergency power applications, communication equipment
Portable	Sealed, maintenance-free, higher capacity than nickel-cadmium batteries	Railroad equipment, consumer electronics, portable tools, pagers, appliances, and photographic equipment, standby power, memory backup
Nickel-metal hydride	Durable, rugged construction, long life, low specific energy	Consumer electronics and other portable applications; electric and hybrid electric vehicles
Nickel-iron	High specific energy, extended cycle life and rate capability	Materials handling, stationary applications, railroad cars
Nickel-zinc	Highest specific energy, very good high-rate capability, low cycle life, high cost	Bicycles, scooters, trolling motors
Silver-zinc	Highest specific energy, very good high-rate capability, low cycle life, high cost	Lightweight portable electronic and other equipment; training targets, drones, submarines, other military equipment, launch vehicles and space probes
Silver-cadmium	High specific energy, good charge retention, moderate cycle life, high cost	Portable equipment requiring a lightweight, high-capacity battery; space satellites
Nickel-hydrogen	Long cycle life under shallow discharge, long life	Primarily for aerospace applications such as LEO and GEO satellites
Ambient-temperature rechargeable "primary" Types [Zn/MnO <sub>2</sub> ]	Low cost, good capacity retention, sealed and maintenance-free, limited cycle life and rate capability	Cylindrical cell applications, rechargeable replacement for zinc-carbon and alkaline primary batteries, consumer electronics (ambient-temperature systems)
Lithium-ion	High specific energy and energy density, long cycle life	Portable and consumer electronic equipment, electric vehicles, and space applications

Secondary battery technologies differ from primary batteries in the specific characteristics of reuse (secondary batteries can be recharged), power density (tends to be higher in secondary batteries), discharge profiles (secondary batteries typically are more sloped), and performance in low-temperature environments (secondary batteries perform better), specific energy and energy densities (usually higher

primary batteries), and charge retention (often larger in primary batteries). Lithium-ion batteries do not follow this trend as this technology has high energy densities, better charge retention, and other performance advantages because they use higher energy materials. Lithium-ion batteries were reviewed in greater detail in Section 4.3.4.

Secondary batteries have existed for over 100 years. The first secondary battery, the lead-acid battery developed by Planté in 1859, has been one of the most used secondary batteries to date. Edison introduced the nickel-iron alkaline battery in 1908. It was used as a power source for the early electric automobile, but because of its high cost, maintenance requirements, and lower specific energy, it lost its market share over time [6].

The discharge profiles of secondary batteries listed in Table 4.8 are shown in Figure 4.4 [5] at a C/5 rate. From the graph, it is apparent that lead-acid batteries have the highest cell voltage for aqueous systems (measuring at 2.2 V). All discharge curves shown for secondary batteries are flat except for silver oxide, which has a double plateau. The double plateau occurs because this battery technology has a two-stage discharge. The lithium-ion discharge curve, also shown, has the largest cell voltage, averaging 3.6 V. However, the lithium-ion discharge curve is not as flat as the other secondary batteries because of lower conductivity of the non-aqueous electrolytes and thermodynamics of intercalation [5]. It should be mentioned that the lithium-ion battery is the only non-aqueous secondary battery included in Table 4.8. Table 4.9 compares secondary batteries by rating their different characteristics (e.g., energy density, power density, efficiency, cost, etc...) where a rating of 1 is the best and a rating of 5 is the worst. An average has been calculated and the secondary battery with the lowest average is the lithium-ion.



**Figure 4.4.** Discharge Profiles of Conventional Secondary Battery Systems and Rechargeable Lithium-Ion Battery at Approximately C/5 Discharge Rate [5]

**Table 4.9.** Comparison of Secondary Batteries \* <sup>[5]</sup>

System	Energy Density	Power Density	Flat Discharge Profile	Low-temperature Operation	Charge Retention	Charge Acceptance	Efficiency	Life	Mechanical properties	Cost	Average
Lead-acid:											
Pasted	4	4	3	3	4	3	2	3	5	1	3.2
Tubular	4	5	4	3	3	3	2	2	3	2	3.1
Planté	5	5	4	3	3	3	2	2	4	2	3.3
Sealed	4	3	3	2	3	3	2	3	5	2	3
Lithium-metal	1	3	3	2	1	3	3	4	3	4	2.7
Lithium-ion	1	2	3	2	2	1	1	1	3	2	1.8
Nickel-cadmium:											
Pocket	5	3	2	1	2	1	4	2	1	3	2.4
Sintered	4	1	1	1	4	1	3	2	1	3	2.1
Sealed	4	1	2	1	4	2	3	3	2	2	2.4
Nickel-iron	5	5	4	5	5	2	5	1	1	3	3.6
Nickel-metal hydride											
Nickel-zinc	3	2	2	2	4	2	3	3	2	3	2.6
Nickel-zinc	2	3	2	3	4	3	3	4	3	3	3
Silver-zinc	1	1	4	3	1	3	2	5	2	4	2.6
Silver-cadmium	2	3	5	4	1	5	1	4	3	4	3.2
Nickel-hydrogen	2	3	3	4	5	3	5	2	3	5	3.5
Silver-hydrogen	2	3	4	4	5	3	5	2	3	5	3.6
Zinc-manganese dioxide	2	4	5	3	1	4	4	5	4	2	3.4

\* Rating: 1 to 5, best to worst respectively

### 4.3.1 Nickel-Cadmium Batteries

A major issue in secondary batteries is the buildup of pressure caused by gassing during overcharge. Using design features found in sealed nickel-cadmium batteries has eliminated this problem and the need for maintenance, leading to widespread acceptance and use of these batteries. Other advantages include long cycle life, good low-temperature and high-rate performance capability, long shelf life in any state of charge, and rapid recharge capability. Disadvantages include voltage depression or memory effect in certain applications, higher cost than sealed lead-acid batteries, poor charge retention, sealed lead acid batteries are better at high temperature and float service, environmental concern with the use of cadmium, and lower capacity than other competitive batteries <sup>[5]</sup>.

A listing of manufactured nickel-cadmium single-cell batteries with a variety of different physical and electrical characteristics is shown in Table 4.10 <sup>[5]</sup>. Types of batteries included in this list are standard, high-capacity, fast-charge, high-temperature, heat-resistant, and slim rectangular batteries. Multi-cell batteries with different output voltages, configurations, and sizes also are manufactured using these single cells.

**Table 4.10.** Specifications of Typical Sealed Nickel-Cadmium Single-Cell Batteries <sup>[5]</sup>

Battery size	Capacity at 0.2C rate, mAh	Dimensions, max., mm		
		Diameter	Height	Weight, g
Cylindrical Batteries				
Standard batteries: Charging – standard, 0.1C rate, 14–16 h; quick, 0.3C rate, 4–5 h				
N	170	12.0	29.3	9
AAAA	120	8.0	42.5	6
1/3 AAA	55	10.5	15.8	4
AAA	270	10.5	44.4	11
½ AA	300	14.5	30.3	14
AA	650	14.3	50.2	23
A	550	17.0	28.5	19
SC	1450	22.9	43.0	45
SC	1550	22.9	43.0	47
D	4400	33.0	59.5	160
D	4800	33.0	61.1	145
F	7700	33.2	91.0	230
M	12000	43.1	91.0	400
High-capacity batteries: Charging – standard, 0.1C rate, 14–16 h; quick, 0.3C rate 4–5h				
AA	880	14.3	50.3	23
AA	1150	14.3	50.3	24
A	650	17.0	28.5	18
A	1200	17.0	43.0	28
A	1550	17.0	43.0	31
SC	1900	22.9	43.0	47
SC	2400	22.9	50.0	58
D	5400	33.2	59.5	150
M	25500	43.1	146.1	700
Fast-charge batteries: Charging – standard, 0.1C rate, 14–16h; quick, 0.3C rate, 4–5 h; fast, 1.5C rate, 1 h				
A	550	17.0	28.5	19
4/5 SC	1250	22.9	34.0	43
SC	1400	22.9	43.0	52
SC	1850	22.9	43.0	54
SC	2000	22.9	42.9	56
C	3200	26.0	50.0	84
D	4300	33.0	59.5	160
High-temperature batteries: Charging – standard, 0.1C rate, 14–16 h				
AA	650	14.3	48.9	23
SC	1650	22.9	43.0	49
C	3100	26.0	50.5	78
D	4500	33.2	59.5	145
F	7700	33.2	91.0	230
M	12000	43.1	91.0	400
Heat-resistant batteries: Charging – standard, 0.1C rate, 14–16 h; quick, 0.3C rate, 4–5 h				
2/3 AA	300	14.5	30.3	14
AA	650	14.3	50.2	23
4/5 SC	1350	22.9	43.0	52
SC	1800	22.9	42.9	56
C	2200	26.0	50.0	80



**Table 4.10. (Contd)**

Slim Rectangular Batteries				
Capacity at 0.2C rate, mAh	Dimensions, max., mm			
	Height	Width	Thickness	Weight, g
Slim rectangular batteries: Charging – standard, 0.1C rate, 14–16 h; quick, 0.3C rate, 4–5 h; fast, 1.5C rate, 1 h				
450	48.0	17.2	6.3	17
650	48.0	17.2	8.5	22
650	67.0	17.2	6.3	24
900	67.0	17.2	8.5	30
1200	67.0	17.2	10.7	38

### 4.3.2 Lithium-Ion Batteries

Cells in a lithium-ion battery use lithium intercalation compounds as the positive and negative materials. As lithium-ion batteries cycle, lithium ions are exchanged between the positive and negative electrodes. Batteries that use this mechanism are referred to as rocking chair batteries because lithium ions “rock” back and forth between the positive and negative electrodes as the cell is charged and discharged <sup>[5]</sup>. The positive electrode is typically a metal oxide while the negative electrode is a graphitic carbon. During the charge or discharge processes, lithium ions are inserted or extracted within the active materials, particularly from interstitial space between atomic layers.

Lithium cobalt oxide (LiCoO<sub>2</sub>) was used on the first batteries to be marketed and the majority of currently available for the positive electrode material. LiCoO<sub>2</sub> has many advantages such as good electrical performance, easy preparation, good safety properties, and relative insensitivity to process variation and moisture. The introduction of high performance materials such as lithium manganese oxide (LiMn<sub>2</sub>O<sub>4</sub>) or lithium nickel cobalt oxide (LiNi<sub>1-x</sub>Co<sub>x</sub>O<sub>2</sub>) will improve cell and battery performance. As for the negative electrode, coke negative electrode materials were first used for commercialized battery products. As graphite materials improved, the battery industry started using graphitic carbons for negative electrodes because of their higher specific energy capacity and improved cycle life and rate capability.

Lithium-ion batteries have many advantages compared with other types of secondary batteries. These advantages include sealed cells with no maintenance requirements, long cycle life, a wide temperature range of operation, long shelf life, and low self-discharge rates. Other advantages include rapid charge capability, high rate and high power discharge capability, high coulombic and energy efficiency, high specific energy and energy density, and no memory effect. However, disadvantages include moderate initial cost, degradation at high temperature, need for protective circuitry, capacity loss or thermal runaway when overcharged, venting and possible thermal runaway when crushed, and cylindrical designs typically offer lower power density than nickel-cadmium or nickel-metal hydride batteries.

General performance characteristics of lithium-ion batteries are shown in Table 4.11 <sup>[5]</sup>. Typically, lithium-ion batteries have a high voltage, ranging from 2.5 V to 4.2 V, which is approximately three times more voltage than nickel-cadmium or nickel-metal hydride batteries can provide. Therefore, fewer cells are needed in a lithium-ion battery to reach a specific voltage. Lithium-ion batteries also offer larger specific energies and energy densities; lithium-ion batteries with specific energies over 150 Wh/kg and energy densities above 400 Wh/L are commercially available. These batteries also tolerate varying

degrees of temperature for the charging and discharging process, (i.e., -20°C to 60°C for charging and -40°C to 65°C for discharging). These characteristics make lithium-ion secondary battery systems robust for a variety of different applications.

**Table 4.11.** General Performance Characteristics of Li-Ion Batteries <sup>[5]</sup>

Characteristic	Performance Range
Operational cell voltage	4.2 to 2.5 V
Specific energy	100 to 158 Wh/kg
Energy density	245 to 430 Wh/L
Continuous rate capability	Typical: 1C High Rate: 5C
Pulse rate capability	Up to 25C
Cycle life at 100% DOD	Typically 3000
Cycle life at 20 to 40% DOD	Over 2000
Calendar life	Over 5 years
Self-discharge rate	2 to 10% / month
Operable temperature range	-40°C to 65°C
Memory effect	None
Power density	2000 to 3000 W/L
Specific power	700 to 1300 W/Kg

Thin-film lithium-ion batteries also are available. These batteries are similar to secondary lithium-ion batteries but use thin materials to produce a battery of millimeter thickness. Similarities include electrical characteristics such as voltage and current density. Advantages of thin-film lithium-ion batteries over typical secondary batteries are higher average output voltage, lighter weights that equate to higher energy density, and longer cycling life. The main difference is size, which allows the thin-film batteries to be used in thinner electronic devices and smaller medical devices. These batteries also can be fabricated in any shape, and can be stacked, and used in parallel, which reduces their space needs. Components of a thin-film battery include cathode, anode, substrate, electrolyte, current collector, and charge separator <sup>[7]</sup>.

There are a variety of research applications that use thin-film battery technologies. Yeo et al. <sup>[8]</sup> used a thin film battery for battery-assisted radio frequency identification (RFID) tags. Each tag consists of an RFID tag by produced by Power Paper Ltd. and a thin-film battery with dimensions of 29 mm x 23 mm. These batteries produce a nominal voltage of 1.5 V at a thickness of 0.7 mm. Carmo et al <sup>[9]</sup> used thin-film solid-state rechargeable batteries in thermoelectric energy-scavenging microsystems. In this device, the small thermal power generated by the human-body is converted to charge the thin film battery system.

## 4.4 Characteristic Interdependencies

Parameters used to characterize a battery are voltage, capacity, energy content, specific energy, energy density, internal resistance, self-discharge, temperature, and cycle. There are many dependencies between different parameters, a few of which are listed below:

- Capacity increases as voltage decreases
- Charge capacity decreases as cycles increase
- Discharge capacity decreases as cycles increase

- Discharge capacity increases as voltage decreases
- Capacity decreases over time
- Increased energy causes decreased cell voltage
- Voltage increases as charge time increases
- Increased charging time decreases the charging current.

## 4.5 Typical Applications

There has been much research dealing with remote sensing and control using battery systems for power that can be applied to sensors and control systems for building applications. Khan and Vemuri <sup>[10]</sup> developed a wireless sensor network (WSN) that can be dynamically reconfigured based on the remaining battery capacity. Lower power implementations will be used when the remaining battery capacity is low. Gatzianas et al. <sup>[11]</sup> demonstrated cross-layer resource allocation for WSNs using rechargeable batteries. A control and power allocation policy was proposed in this research to maximize total system utility. Ma et al. <sup>[12]</sup> constructed an energy efficient virtual backbone for WSNs. This backbone focused on reducing power dissipation and prolonging network lifetime by controlling routing, broadcasting, and data propagating. Qiuling et al. <sup>[13]</sup> developed a battery discharge and modulation scheme for WSNs. They studied an accurate analytical battery model with a focus on capacity effect and recovery effect—two key phenomena that affect battery lifetime. Ma and Yang <sup>[14]</sup> studied the effect of battery behavior for routing streaming data transmissions in WSNs. They developed a computable energy model to characterize battery discharge behavior, and then used it to develop a battery-aware routing protocol to schedule routing for a WSN based on the current state of the battery. Aziz et al. <sup>[15]</sup> compiled a survey that deals with energy efficiency and topology control techniques that extend battery life in WSNs.

## 4.6 Relative Cost Impacts

The cost of batteries for sensor and control systems applied to residential and commercial buildings is difficult to estimate because needs vary depending on the hardware being used, the numerous types of batteries, and different configurations. Table 4.12 and Table 4.13 <sup>[16]</sup> give the reader a rough idea of cost for a subset of popular primary and secondary batteries, respectively.

**Table 4.12.** Cost of Energy Obtained by Primary Batteries <sup>[16]</sup>

Cost/Wh/ x1000=\$	AAA Cell 1.5V Alkaline	AA Cell 1.5V Alkaline	C Cell 1.5V Alkaline	D Cell 1.5 Alkaline	9 Volt Alkaline	Camera battery 6V Lithium
Capacity (Ah)	1.1	2.8	7.8	17	0.6	1.4
Energy (Wh)	1.6	4.2	11.7	25	4.2	8.4
Battery Cost (estimated, U.S. \$)	0.65	0.5	0.8	2	1	10
Cost per kWh (U.S. \$)	406	120	68	80	238	1.2

**Table 4.13.** Cost of Energy Obtained by Secondary Batteries <sup>[16]</sup>

Cost/Wh/ x1000=\$	Nickel-cadmium 6 cells	Nickel-metal-hydride 6 cells	Lead-acid small 6 cell pack	Li-ion pack with two 18650 cells	Lead-acid for wheelchairs and scooters
Capacity (Ah)	0.6	1	2	2	33
Voltage (V)	7.2	7.2	12	7.2	12
Energy per discharge (Wh)	4.3	7.2	24	14.4	396
Cycle life (best cases)	1500	500	250	500	250
Battery cost (ref. only)	50	70	50	150	100
Cost per kWh (US\$)	7.75	19.5	8.5	20	1

## 4.7 Summary/Conclusions

A detailed discussion of batteries and application to sensors and controls for residential and commercial buildings is provided in this chapter. Different primary and secondary battery technologies, applications, and costs are included in the discussion. A subset of primary and secondary battery systems was chosen for further analysis based on characteristics important to sensors and control systems used in a building environment. These characteristics are high energy density, good shelf life (i.e., minimal self-discharge), and portability (i.e., size). The reviewed technologies for primary batteries were 1) alkaline; 2) zinc-air; and 3) lithium/thionyl chloride (Li/SOCl<sub>2</sub>). The reviewed technologies for secondary batteries were nickel-cadmium and lithium-ion. Electrical characteristics were given for each technology. These characteristics can be used to determine the most appropriate battery technology for chosen sensors and control hardware being used for building's applications.

It is difficult to recommend a specific battery type for use with sensors and control systems applied to building monitoring because of the wide range of possible hardware configurations. A general example can be given to make the reader aware of typical hardware needs and the batteries that would fit those needs. Healy <sup>[17]</sup> described the typical energy consumption of wireless sensors, and his findings are summarized in Table 4.14. Based on these values and the data communication scheme being used, power consumption can be calculated on a per time basis, and battery technologies can be targeted to meet these needs for a specific time period. A decision can then be made on a battery technology based on cost and size. Environmental considerations also must be considered. With all the information presented in this chapter, based on electrical characteristics, cost, and size, secondary lithium-ion batteries should be seriously considered as the power source for sensors and control systems used in buildings.

**Table 4.14.** Typical Energy Consumption of a WSN <sup>[17]</sup>

Radio mode	Power consumption (mW)
Transmit	15
Receive	13
Idle	12
Sleep	0.016

## 4.8 References

1. Battery University. (date unknown). Battery statistics [website]. Available: [http://batteryuniversity.com/learn/article/battery\\_statistics](http://batteryuniversity.com/learn/article/battery_statistics).
2. Battery University. (date unknown). Will secondary batteries replace primaries? [website]. Available: [http://batteryuniversity.com/learn/article/will\\_secondary\\_batteries\\_replace\\_primaries](http://batteryuniversity.com/learn/article/will_secondary_batteries_replace_primaries).
3. H.A. Kiehne, Battery Technology Handbook, 2<sup>nd</sup> Edition, Marcel Dekker, Inc., New York, NY, 2003.
4. Battery Education. (April 7, 2006). Battery voltage [website]. Available: [http://www.batteryeducation.com/2006/04/battery\\_voltage.html](http://www.batteryeducation.com/2006/04/battery_voltage.html).
5. D. Linden and T.B. Reddy, Handbook of Batteries, 3<sup>rd</sup> Edition, McGraw-Hill Companies, Inc., New York, NY, 2002.
6. A.J. Salkind, D.T. Ferrell, and A.J. Hedges, "Secondary Batteries 1952 – 1977," Journal of the Electrochemical Society, vol. 128, no. 8, August 1978.
7. N.J. Dudney, "Thin film micro-batteries," The Electrochemical Society Interface, pp. 44-48, Fall 2008.
8. J. Yeo, S.-G. Moon, and J.-Y. Jung, "Design of antennas for a battery-assisted RFID tag with a thin and flexible film battery," IEEE Antennas and Propagation Society International Symposium, pp. 5463-5466, June 9-15, 2007.
9. J.P. Carmo, R.P. Rocha, A.F. Silva, L.M. Goncalves, and J.H. Correia, "Integrated thin-film rechargeable battery in a thermoelectric scavenging microsystem," IEEE International Conference on Power Engineering, Energy, and Electrical Drives (POWERENG'09), pp. 359-362, March 18-20, 2009.
10. J. Khan and R. Vemuri, "Energy management in a battery-powered sensor networks with reconfigurable computing nodes," IEEE International Conference on Field Programmable Logic and Applications, pp. 543-546, August 24-26, 2005.
11. M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Control of wireless networks with rechargeable batteries," IEEE Transactions on Wireless Communications, vol. 9, no. 2, 2010.
12. C. Ma, Y. Yang, and Z. Zhang, "Constructing battery-aware virtual backbones in sensor networks," IEEE Proceedings of the 2005 International Conference on Parallel Processing (ICPP'05), 2005.
13. T. Qiuling, L. Ye, Q. Yongming, and W. Huan, "Joint scaling battery discharge and modulation scheme in wireless sensor networks," The 5<sup>th</sup> International Conference on Computer Science & Education, August, 24-27, 2010.
14. C. Ma and Y. Yang, "Battery-aware routing for streaming data transmissions in wireless sensor networks," IEEE 2<sup>nd</sup> International Conference on Broadband Networks, pp. 464-473, October 3-7, 2005.
15. A.A. Aziz, Y.A. Sekercioglu, P. Fitzpatrick, and M. Ivanovich, "A survey on distributed topology control techniques for extending the lifetime of battery powered wireless sensor networks," IEEE Communications Surveys & Tutorials, accepted for inclusion in a future issue.
16. Battery University. (date unknown). The cost of portable power [website]. Available: [http://batteryuniversity.com/learn/article/the\\_cost\\_of\\_portable\\_power](http://batteryuniversity.com/learn/article/the_cost_of_portable_power).

17. W. Healy. (April 26, 2010) Building sensors and energy monitoring systems [website]. Available: [http://www1.eere.energy.gov/buildings/building\\_america/pdfs/5\\_building\\_sensors\\_healy.pdf](http://www1.eere.energy.gov/buildings/building_america/pdfs/5_building_sensors_healy.pdf)

## 5.0 Power Generation: Self-Harvesting Sources

### 5.1 Introduction

Significant number of sensors in buildings operate at a low-duty cycle from four measurements an hour to as low as one measurement a day. While significant cost reductions can be achieved with harvesting energy, key engineering choices have to be made to match the functional requirements of the system under consideration. The sensor devices have to be operational at relatively low-power from several hundred microwatts to few tens of milliwatts. A relatively constant (from application perspective) source of energy is needed for power harvesting. An intermediate accumulator or energy storage mechanism with relatively long life cycle is often required for continual harvesting. These energy storage mechanisms typically include long-life, batteries, capacitors, and springs.

There are two different architectures for self-harvesting as described in [1]. The energy can be harvested and used immediately, called Harvest-Use architecture. Alternatively, the energy can be stored in an intermediate storage source (like batteries, capacitors, or springs) and used when needed, called Harvest-Store-Use architecture. Figure 5.1 shows the implementation of both architectures.

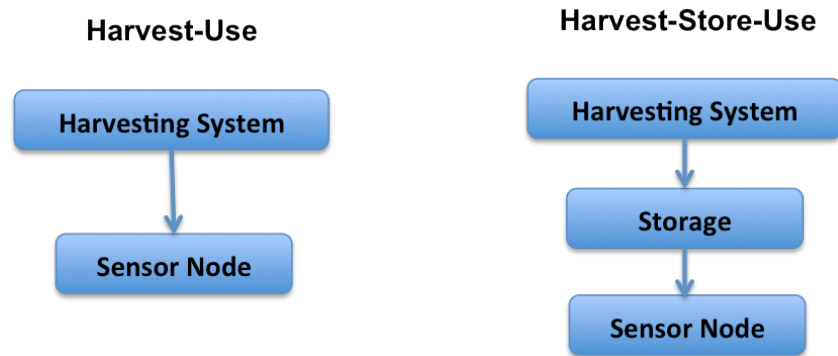


Figure 5.1. Energy Harvesting With and Without Storage Architecture <sup>[3]</sup>

### 5.2 Technical Details

Over the past decade power consumption of the electronic circuits, particularly complementary metal oxide semiconductor (CMOS) circuits, has reduced from several mW per transistor to less than 100nW per transistor. As the size and power requirements of the electronics (sensors, wireless transceivers etc.) decrease, the use cases for self-harvesting devices are emerging. Several low-data rate, low-power wireless network protocols are currently under development and deployment. These include IEEE 802.15.4, ISA 100.11a, Wireless HART, and ZigBee™. Several of these devices have shown to operate on commercial off the shelf (COTS) batteries for 2-3 years with a duty cycle of 5-10%. These devices prove to be good candidates for energy harvesting applications. The energy harvesting however depends on various functional and operational aspects of the system under consideration. Table 5.1 shows all the considerations for energy harvesting. Table 5.2 shows the power densities of various techniques.

**Table 5.1.** Factors Affecting the Power Consumption of a Wireless Sensor Network <sup>[5]</sup>

<b>Power Supply</b>	<ul style="list-style-type: none"> <li>• Discharge rate</li> <li>• Relaxation effect</li> <li>• Battery Dimensions</li> <li>• Supply Voltages</li> <li>• Type of Electrode Material Used</li> <li>• Diffusion Rate of the Active Materials in the Electrolyte</li> </ul>
<b>Sensors</b>	<ul style="list-style-type: none"> <li>• Physical to Electrical Signal Conversion</li> <li>• Complexity of Supporting Components</li> <li>• Signal Sampling</li> <li>• Signal Conditioning</li> </ul>
<b>ADC</b>	<ul style="list-style-type: none"> <li>• Sampling Rate</li> <li>• Aliasing</li> <li>• Dither</li> </ul>
<b>Microprocessor</b>	<ul style="list-style-type: none"> <li>• Core Operating Frequencies</li> <li>• Power Proportional to Process and Computational Load</li> <li>• Ambient Temperature</li> <li>• Application Code</li> <li>• Peripheral Utilization</li> </ul>
<b>Radio</b>	<ul style="list-style-type: none"> <li>• Modulation Scheme</li> <li>• Data Rate</li> <li>• Transmission Range</li> <li>• Operational Duty Cycle</li> </ul>

**Table 5.2.** Power Density of Various Energy Harvesting Techniques

	Power Density ( $\mu\text{W}/\text{cm}^3$ ) one year lifetime	Power Density ( $\mu\text{W}/\text{cm}^3$ ) 10 year lifetime
Solar (outdoors)	15,000 – Direct sun, 105 Cloudy day	15,000 – Direct sun, 105 Cloudy day
Solar (indoors)	6-Office desk	6-Office desk
Vibrations (piezoelectric conversion)	250	250
Vibrations (electrostatic conversion)	50	50
Acoustic noise	0.003 at 75dB, 0.96 at 100dB	0.003 at 75dB, 0.96 at 100dB
Temperature gradient	15 at 10C gradient	15 at 10C gradient
Shoe inserts	330	330
Batteries(non-rechargeable lithium)	45	3.5
Batteries(rechargeable lithium)	7	0
Hydrocarbon fuel (micro heat engine)	333	33
Fuel Cells (Methanol)	280	28

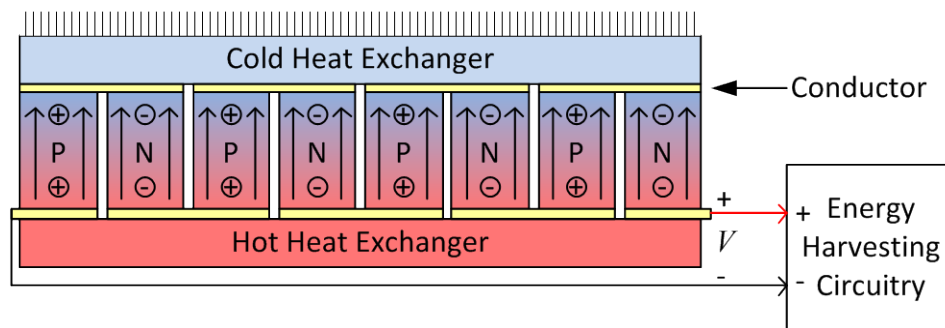


### 5.2.1 Thermal Power Harvesting

Thermal power harvesting is a concept where differences in temperature are used as a source to initiate a flow of electrons (or electron holes) resulting in energy. Certain materials produce a flow of electrical current when the material is subjected to a temperature difference (see Figure 5.2). This technique requires source of heat in a relatively accessible location from where the sensors are located (typically placed for best observable location of the thermal behavior of the building). Heat exchangers, compressors, and motors are locations where there is plenty of waste heat that can be used for this purpose.

Thermoelectric devices generate an electric potential when a thermal gradient is induced across them. Pyroelectric devices on the other hand, generate electric potentials in response to a rise or drop in the temperature of the devices. Thermophotovoltaic devices (TPV) are composite PV devices that consist of a narrowband thermal emitter and a tuned PV cell. Because TPV devices require very high temperatures to be effective, their use within buildings is limited, although there are possible applications in industrial situations where very high temperature energy sources exist.

For situations where there are significant temperature gradients, such as boiler casings, the hot water pipes leaving a boiler or entering a hot water heater, the chilled water pipes leaving a chiller, or other industrial process waste heat, thermoelectric devices can be effectively used to harvest energy. For situations where there is a high variability in temperature, such as on a rooftop, pyroelectric devices can be used.



**Figure 5.2.** Configuration Drawing of a Typical Thermoelectric Generator (TEG). This generator uses 4 P-N junctions to increase the device sensitivity.

### 5.2.2 Vibration Power Harvesting

Mechanical vibration can be a significant source for energy harvesting. Several components like motors, ducts, and compressors produce significant vibrations in buildings with acceleration amplitudes ranging 1 to 10 m/s<sup>2</sup> [2]. Thorough reviews of the literature can be found in both recent papers [3]-[13]. The extraction of vibration energy can come by a variety of mechanisms which can be roughly categorized as piezoelectric, electromagnetic, or electrostatic, and most recently through the properties of shape memory alloys. Shape memory alloys [14] are an extremely new technology and since they are far from commercialization they will not be covered in this review.

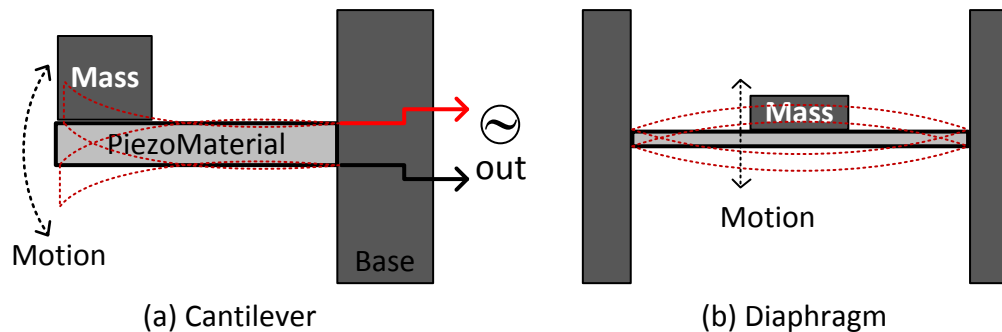
Piezoelectric energy harvesting systems use vibrations from a mechanical source to deform a piezoelectric material (see Figure 5.3). This deformation or strain on the material induces voltage across

the material due to the crystal structure, often called piezoelectric effect. The advantage of the piezoelectric transducer is the thin profile and wideband of vibrations that can induce voltage. Piezoelectric generators (PEG) utilize the piezoelectric effect, wherein a stress in a piezoelectric material causes a charge separation in the material atoms resulting in an electric potential. The electric potentials can be quite substantial, which means that DC-DC or AC-DC voltage converters may be unnecessary with a piezoelectric harvester. A very thorough review of the literature and description of the basic physics of standard piezoelectric harvesters is available in the recent book by Erturk and Inman<sup>[15]</sup> so another review will not be made here.

Electromagnetic transducers convert vibrational energy to electrical current through Faraday’s law of electromagnetic induction: the generation of current in a conductor in the presence of a changing magnetic field. In practice, the magnetic field is usually made to change by moving a conductor through a stationary magnetic field or by moving a magnet near a stationary conductor.

Electrostatic transducers convert vibrational energy to electric current through Coulomb’s law of Electrostatics: The force between two charged particles is inversely proportional to the distance of the particles. When charged particles are placed between a set of plates placed near each other but not in contact with one another and there is relative motion between the plates, the voltage across the two plates will vary with the motion.

Nearly all vibration harvesters, whether electromagnetic, electrostatic, or piezoelectric can be analyzed as damped mass-spring resonant systems<sup>[3]</sup>. The mass is a moving plate, coil, or magnet. The spring can be a physical spring or another restoring force such as additional magnets, and the damping can be mechanical, electrical, or viscous in nature. As resonant systems, the maximum power output occurs at the natural frequency of the system<sup>[3]</sup>. As a result, most vibration harvesters are designed to operate (or are “tuned”) to a specific frequency, which is selected based on the prominent frequencies of the vibration source. Much electromagnetic equipment will have vibration at the power line frequency or the 2nd harmonic of that frequency (60/120Hz or 50/100 Hz) and so many of the energy harvesters are designed to resonate near one of those frequencies, however, many pieces of heating, ventilation, and air conditioning (HVAC) equipment can have significant vibration at lower or higher frequencies usually related to the rotational rate of motors driving the vibrating fans, pumps, or compressors. Vibration harvesting devices can work wherever there are sufficient vibration levels and frequencies are fairly constant.



**Figure 5.3.** Basic Piezoelectric Cantilever System for Generating Electricity Using a (a) Cantilever and (b) Diaphragm. Motion of mass creates strain in piezo cantilever or diaphragm which generates electricity.

### 5.2.3 Photovoltaic Power Harvesting

Photovoltaics convert energy from light to electricity, commonly referred as photovoltaic effect. The usability of this source within buildings is practical with abundant sources of light. However, depending on the location of the sensor, the light source is not continuous over time and the PV source may have to be coupled with an energy storage mechanism like batteries or capacitors. The low temperature thin film solar cell technologies are advancing at a rapid pace. The integration of a PV cell even at a low efficiency of 5% would enable the realization of a self-powered smart sensor platform.

### 5.2.4 Electromagnetic Power Harvesting

Electromagnetic power harvesting is a mechanism where the ambient radio frequencies are used to harvest energy. A typical device called a rectenna is used for harvesting RF energy. A rectenna is a rectifying antenna, a special type of antenna that is used to convert microwave energy into direct current electricity. They are used in wireless power transmission systems that transmit power by radio waves.

### 5.2.5 Wind Power Harvesting

The basic wind turbine can have one of two possible configurations, vertical axis or horizontal axis. There is a significant amount of airflow in the HVAC ducts of buildings that can be used for power generation without affecting the building subsystems.

### 5.2.6 Acoustic Power Harvesting

Acoustic energy generators (AEG) are essentially microphones that have been designed for maximum energy extraction rather than maximum fidelity. In order to maximize efficiency, AEG are often designed as resonant systems, like vibration harvesters; however, AEG can be utilized as broadband harvesters at frequencies below the fundamental resonance frequency of the device, albeit at reduced efficiency. Further, the acoustic field itself can be amplified through the use of Helmholtz resonators (fluid mass-spring systems) and the AEG could then be placed within the Helmholtz resonator for increased effectiveness.

## 5.3 Characteristic Interdependencies

The characteristic interdependencies for energy harvesting are:

- The power budget of the sensor application has to match the power output of the energy harvester. This is a primary driver for architectural and technological selection for self-harvesting
- The reliability of sensor over time is dependent on the source of energy and careful consideration has to be taken for enhancing system-level reliability (for example, redundancy).
- Optimal sensor performance for various building monitoring locations depends on the sensor locations. Architectural choices have to be made to consider optimal performance of sensor and energy harvester.
- Performance degradation of harvesters and sub-systems (charge cycle of batteries or capacitors used for storage) has to be considered over time for successful self-harvesting applications.

- In the case of thermoelectric devices they are electrically connected with solder which limits the upper temperatures at which the device can be operated.

## 5.4 Typical Applications

Several building monitoring sensors can utilize the energy harvesting techniques described above.

Thermoelectric: Thermoelectric devices can work wherever significant temperature differences exist (> 10C) for significant times. Examples of current use within buildings include industrial process equipment, hot water and steam pipes, electrical power distribution systems (especially transmission lines, transformers, and electrical busses), motor casings, and waste heat from data centers. For the harvesting of micropower for wireless sensors, thin film technologies and MEMS construction methods are being used with TEG to shrink the size and cost of the basic TEG module. Such devices could harvest thermal energy from hot and chilled water lines, from the casings of HVAC ductwork, and from temperature gradients within building envelopes or within window glazing.

**Table 5.3.** Amplitude of Vibration and Fundamental Frequency for Vibration Sources

Vibration Source	Acceleration (m/s <sup>2</sup> )	Peak Frequency (Hz)
Blender casing	6.4	121
Clothes dryer	3.5	121
Person tapping their heel	3	1
Small microwave oven	2.5	121
HVAC vents	0.2-1.5	60
Windows near busy road	0.7	100
Second story floor of busy office	0.2	100

Source: Heat Exchangers, Motors, Compressors

Applications: Equipment condition monitoring

1. Vibration: Vibration harvesting devices can work wherever there are sufficient vibration levels and frequencies are fairly constant. Table 5.1 contains information about the amplitude of vibration and fundamental frequency for vibration sources commonly found in buildings.

Looking at some of the entries in the table, door frames, HVAC vents, windows, and floors, have ambient vibration in excess of 0.1 m/s<sup>2</sup> with resonance frequencies in the 60 – 100 Hz range.

PEG tuned to lower frequencies (60 Hz – 100 Hz) are best used for motor and pump based HVAC vibration or structural vibration generated from rail or automobile traffic. PEG tuned to higher frequencies (100 Hz – 2 khz) are best used for vibration of equipment with significant higher

harmonics, equipment with inherently high fundamental frequencies (turbines and compressors), and for turbulence induced vibration on HVAC ducts, diffusers, and piping. PEG tuned to higher resonant frequencies are also useful for harvesting vibration from transient events such as button pressing, or ambient building vibration. The high voltage output of the piezo makes a PEG an obvious choice for this use case. In this case, the devices are excited into resonance by the transient event, but the resonance excitation is not maintained. If the sensor/transmitter being powered by the transient event only needs to respond to the transient event (as with a button press) then such a PEG system is acceptable. If the sensor is supposed to have constant power, then a different technology should probably be considered.

Source: Motors, ducts, fans

Applications: Equipment condition monitoring, airflow monitoring, fault detection and diagnosis

2. Acoustic: Acoustic energy harvesters are used in locations where high sound levels, especially tonal sound levels exist. Since nearly all sound in buildings is generated from vibrating equipment and there are inefficiencies in converting vibration energy to acoustic energy, direct harvesting of vibration equipment vibration energy is always more efficient than harvesting the acoustic energy from the sound generated by the same equipment. In some cases however, it is impractical to place vibration harvesters on the vibrating equipment in question because it is moving (fan blade), is too hot (engine exhaust manifold or combustion chambers), however in these cases flow harvesters might be appropriate and more efficient. In other situations there may be a large number of pieces of equipment all adding to the acoustic field and the combined acoustic emissions can then be most economically harvested with a single acoustic harvesting device. Within buildings the most promising use of acoustic energy harvesters is probably within HVAC air handling units and in ductwork, where high fan and flow noise levels exist, but only after efficiency gains have been made.

Source: Air handling unit, Ducts

Application: Equipment condition monitoring, flow measurement

3. Photovoltaic:
  - a. Source: Inside/outside building,
  - b. Applications: Indoor air quality monitoring, building comfort sensing (temperature, relative humidity)
4. Electromagnetic:
  - a. Source: Inside/outside buildings, motors
  - b. Application: Indoor air quality monitoring, building comfort sensing (temperature, relative humidity)
5. Wind:
  - a. Source: Inside air ducts, diffusers
  - b. Applications: Airflow monitoring, In-duct temperature monitoring

## 5.5 Relative Cost Impacts

TEG devices are available in three scales: large high power devices designed for harvesting large amounts of energy (tens to hundreds of W) from high temperatures ( $\Delta T > 50K$ ), smaller scale devices designed to harvest small powers (mW to W) from lower temperature devices ( $\Delta T = 20K-100K$ ) and, just coming to market are the first microscale devices designed for generating very small powers (uW to mW) with lower temperatures (5K – 50K). The smaller scale TEG devices are appropriate for use in building sensor networks and are available from many manufacturers including Evergen, Laird, Marlow, Micropelt, Perpetuum, TECTEG and Thermolife to name a few. These devices typically cost \$10 - \$100 per mW of generated electricity, but those prices often include heat sinks and additional voltage regulation circuitry to generate output directly capable of powering sensors, computers, and communication equipment. Microscale devices by Nextreme, Micropelt, and Thermolife are just being released at costs of \$60-\$200 each, but these costs are expected to drop dramatically as production scales up. Pyroelectric devices are not yet being commercially developed so cost estimates are not available.

Larger scale EMG is available from several manufacturers including Microstrain, Ferro Solutions, Perpetuum. Ferro and Perpetuum are both developing microscale MEMS based EMG. Pricing has not been released. Enocean has recently released ECO 200, a special purpose EMG for harvesting energy from the push of a button or flip of a switch, primarily for use in with their ultra-low power wireless transmitter. These devices sell for under \$7 in quantity 1000+.

No ESGs are being commercially manufactured, but since the methodology is no different from other MEMS devices, it should not take industry long to begin manufacturing ESGs once the commercial ready devices have been developed. MEMS electret microphones which are based on very similar technology are being manufactured by a wide variety of manufacturers including Analog Devices, ST Micro, Knowles, and Memstech. These MEMS microphones have dropped in price to under \$2 in quantity 1000+ suggesting that the price of ESG could be similar.

PEG are currently considered a smaller market than EMG, but because of the ability for the PEG devices to scale down and integrate with standard Integrated Circuit (IC) design, the market for PEG is expected to grow significantly in the next decade [ Harrop 2012].

Commercially available macrosized PEG are available from a variety of manufacturers including Microstrain, Mide, Piezo Systems, Advanced Ceramics, KCF Technologies. The cost of the macroscale commercial products currently ranges from \$20 - \$100, for the raw piezoelectric devices but \$50 to \$200 when the rest of the energy harvesting components are included. However, the cost of similar sized piezoelectric microphones and speakers is significantly lower (under \$1 in bulk) and so as the designs stabilize and demand increases, costs for macroscale PEG are expected to drop.

Microstrain and Mide are both developing microscale piezodevices which will be less expensive than the macroscale devices. Piezo based MEMS microphones are now available at costs similar to MEMS electret microphones (under \$2 each in bulk), so MEMS piezo energy harvesters should be able to be reduced to a similar price when purchased in bulk.

Acoustic energy harvesting devices are not generally commercially available, but several patents have been issued. The construction of diaphragmatic devices is similar to vibration energy harvesters and

electret microphones so the cost of these types of acoustic harvesters should be similar to vibration energy harvesters or electret microphones of a similar size.

Flexensional and hydrostatic are constructed significantly different than diaphragmatic devices and the implementation of these designs has not yet been attempted at the micro or nanoscale and the cost is expected to be significantly higher. Given the low energy conversion of these devices, other conversion technologies should be applied first.

## 5.6 Summary/Conclusions

A wide variety of energy harvesting techniques suitable for application in buildings is surveyed in this chapter. Buildings and subsystems consist of variety of equipment that can be used as virtually continuous source of energy. Advances in materials like piezoelectric materials and dye sensitized solar cells, combined with advances in low-power semiconductor electronics is providing exciting opportunities for energy harvesting in buildings applications. Advanced sensors and controls have proven to save 20-30% energy consumption in buildings. However, these energy savings can only be realized with ubiquitous deployment of sensors across buildings. Energy harvesting is required for realizing this potential to avoid labor costs in wire-powered and battery operated devices.

## 5.7 References

1. Energy Impact of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults and the Energy Savings Potential, TIAX Report D0180, for U.S. DOE Contract 030400101, Nov 2005.
2. S. Katipamula and M.R. Brambley, "Methods for fault detection, diagnostics, and prognostics for building systems – a review, part I," HVAC&R Research, vol. 11, no. 1, pp. 3-25, Jan 2005.
3. Sujesha Sudevalayam Purushottam Kulkarni, "Energy Harvesting Sensor Nodes: Survey and Implications", IEEE Communications Surveys & Tutorials, Vol. 13, No. 3, Third Quarter 2011
4. <http://web.mit.edu/newsoffice/2010/energy-harvesting.html>
5. Costis Kompis, Simon Aliwell (editors), "Energy Harvesting Technologies to Enable Remote and Wireless Sensing, Knowledge Transfer Network, Sensors and Instrumentation, June 2008
6. Roundy, Shad, et al., "Power Sources for Wireless Sensor Networks", ", Springer Berlin Heidelberg, 2004, pg 1-17
- 7.
8. Arnold, D. P. (2007), Review of Microscale Magnetic Power Generation, Magnetics, IEEE Transactions on, 43(11), 3940–3951.
9. Arnold, D. P., and N. Wang (2009), Permanent Magnets for MEMS, Microelectromechanical Systems, Journal of, 18(6), 1255–1266.
10. Beeby, S. P., M. J. Tudor, N. M. White, and S. P. B. and M. J. T. and N. M. White (2006), Energy harvesting vibration sources for microsystems applications, Measurement Science and Technology, 17(12), R175–R195.

11. Chalasani, S., and J. M. Conrad (2008), A survey of energy harvesting sources for embedded systems, Southeastcon, 2008. IEEE, 442–447.
12. Chye, W. C., Z. Dahari, O. Sidek, and M. A. Miskam (2010), Electromagnetic micro power generator — A comprehensive survey, 2010 IEEE Symposium on Industrial Electronics and Applications (ISIEA), (Isiea), 376–382.
13. Harb, A. (2011), Energy harvesting: State-of-the-art, Renewable Energy, 36(10), 2641–2654.
14. Galchev, T., H. Kim, and K. Najafi (2011), Micro power generator for harvesting low-frequency and nonperiodic vibrations, Journal of Microelectromechanical Systems, 20(4), 852–866.
15. Karaman, I., B. Basaran, H. E. Karaca, a. I. Karsilayan, and Y. I. Chumlyakov (2007), Energy harvesting using martensite variant reorientation mechanism in a NiMnGa magnetic shape memory alloy, Applied Physics Letters, 90(17), 172503-172505.
16. Mateu, L., and F. B. Moll Echeto (2011), Review of energy harvesting techniques and applications for microelectronics, in Proc. of SPIE (2011), International Society for Optical Engineering.
17. Mitcheson, P., and E. Yeatman (2008), Energy harvesting from human and machine motion for wireless electronic devices, Proceedings of the IEEE, 96(9), 1457–1486.
18. Naruse, Y., N. Matsubara, K. Mabuchi, M. Izumi, and S. Suzuki (2009), Electrostatic micro power generation from low-frequency vibration such as human motion, Journal of Micromechanics and Microengineering, 19(9), 094002.
19. Pereyma, M. (2007), Overview of the Modern State of the Vibration Energy Harvesting Devices, Perspective Technologies and Methods in MEMS Design, 2007. MEMSTECH 2007. International Conference on, 107–112.
20. Priya, S., and D. J. Inman (2009), Energy Harvesting Technologies, edited by S. Priya and D. J. Inman, Springer.
21. Roundy, S., P. K. Wright, K. S. J. Pister, M. V. Converters, P. K. Wright, and C. Hall (2002), Micro-electrostatic vibration-to-electricity converters, Fuel Cells (methanol), 220, 22.
22. Roundy, S., P. K. Wright, and J. Rabaey (2003), A study of low level vibrations as a power source for wireless sensor nodes, Computer Communications, 26(11), 1131–1144, doi:10.1016/S0140-3664(02)00248-7.
23. Sari, I., T. Balkan, and H. Kulah (2008), An electromagnetic micro power generator for wideband environmental vibrations, Sensors and Actuators A: Physical, 145-146, 405–413.
24. Spreemann, D., and Y. Manoli (2012), Electromagnetic vibration energy harvesting devices, Springer.
25. Torres, E. O., and G. A. Rincon-Mora (2010), A 0.7- m BiCMOS Electrostatic Energy-Harvesting System IC, Solid-State Circuits, IEEE Journal of, 45(2), 483–496.
26. Vullers, R. J. M., R. Van Schaijk, I. Doms, C. Van Hoof, and R. Mertens (2009), Micropower energy harvesting, SolidState Electronics, 53(7), 684–693.



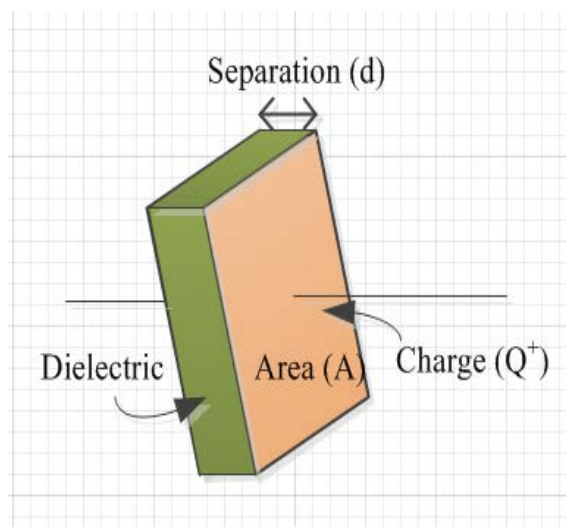
## 6.0 Power Generation: Other Local Storage

### 6.1 Introduction

Capacitors are a fundamental electrical circuit element that is used for smoothing circuits in power supplies, providing backup for microcomputers, to store energy, to tune circuit inductance in resonance circuits, to provide reactive power compensation, and for many other applications. This chapter will examine the different types of capacitors and configurations for sensor applications.

### 6.2 Technical Details

Capacitor technologies exist in many forms, shapes, sizes, and material construction. They are similar to batteries in construction in that there is an anode and cathode. However, capacitors and batteries operate in fundamentally different ways; batteries store energy electrochemically, while capacitors store energy electrostatically (although some newer capacitors do both.) Capacitors use a dielectric between two charged surfaces to create a potential or voltage as shown in Figure 6.1. The different types of capacitors—electrostatic, electrolytic, and electrochemical<sup>[1]-[3]</sup>—are often classified by the configuration and dielectric material employed.

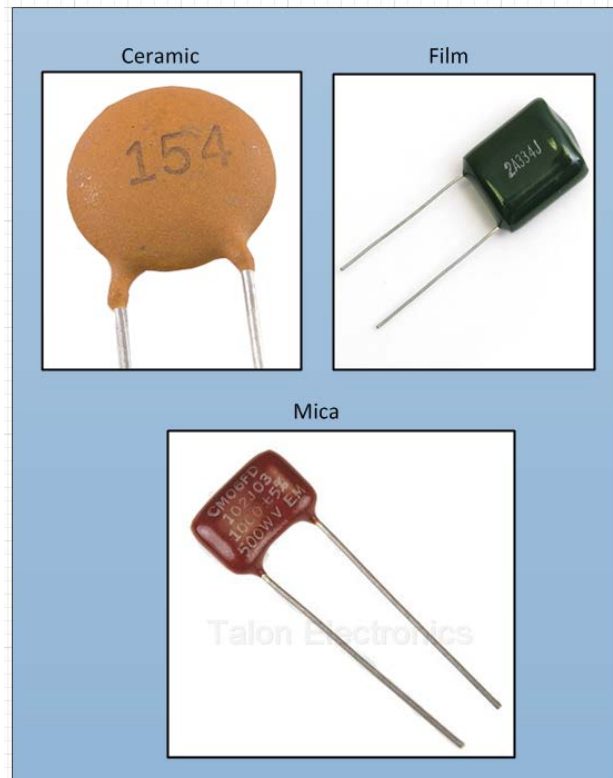


**Figure 6.1.** Depiction of Electrostatic Capacitor

The amount of energy stored in a capacitor is directly proportional to the capacitance and voltage squared. The terminology of capacitance is provided to represent the geometric property of the capacitor design and the material property of dielectric. The equation for capacitance is directly proportional to the area and permittivity of the material and inversely proportional to the separation<sup>[1]-[3]</sup>. Hence, to obtain a high-energy capacitor, the separation must either be very small or the area must be large. In the following sub-sections, different capacitor types are discussed.

## 6.2.1 Electrostatic Capacitors

Electrostatic capacitors include the air, mica, film, and ceramic types, examples of which are shown in Figure 6.2. When a continuous variation of capacitance is required such as in the application of tuning transmitters, receivers and transistor radios, air capacitors are usually used. These variable dielectric capacitors are constructed of multiple plates—one set of fixed plates and a second set of movable plates (the rotor vanes) that move between the fixed plates. The overall capacitance is governed by the position of the moving plates with respect to the fixed plates. Maximum capacitance is achieved when the two sets of plates are fully meshed together.



**Figure 6.2.** Various Electrostatic Capacitors

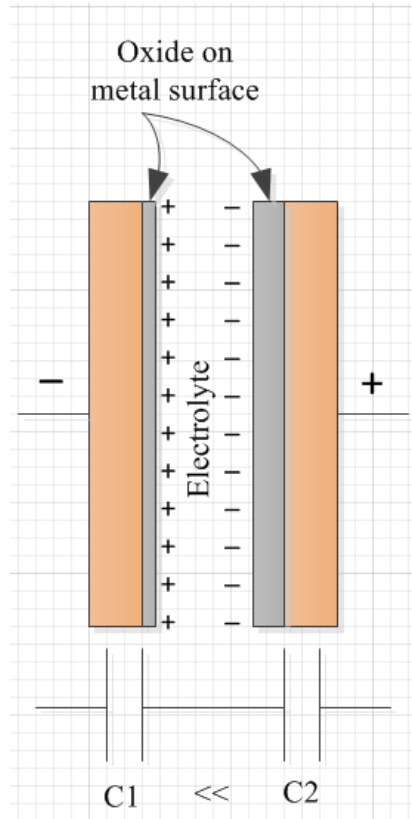
The most common type of available capacitor is the film capacitor. This type has a relatively large family with the primary differences associated with the dielectric properties of the materials used in their construction. Some example dielectric materials commonly applied to film capacitors include polyester (Mylar), polystyrene, polypropylene, polycarbonate, metallized paper, and Teflon. Ceramic capacitors are manufactured by double-sided coating a ceramic disc and stacking the discs. These capacitors typically exhibit large non-linear changes in capacitance compared to changes in temperature and are used as de-coupling capacitors as they also are non-polarized devices.

## 6.2.2 Electrolytic Capacitors

The most common types of electrolytic capacitors include those constructed of aluminum and tantalum (see Figure 6.3). These capacitors generally are used when very large capacitance values are required (i.e., DC link capacitors in inverter or rectifier applications). A semi-liquid electrolyte solution

in a jelly or paste form is inserted between the plates as shown in Figure 6.3. The dielectric is a very thin layer of oxide that is grown electrochemically in production. Because the insulating layer is extremely thin, very small electrolytic capacitors can have large capacitance values.

Most electrolytic capacitors are polarized and can only have DC voltage applied in a single direction. Electrolytic capacitors are generally used in DC power supply circuits because of their large capacitances in relation to size to help reduce the ripple voltage or for coupling and decoupling applications. The disadvantage to these capacitors is their relatively low voltage rating, and because of the polarization requirement, these capacitors cannot be connected to AC supplies.

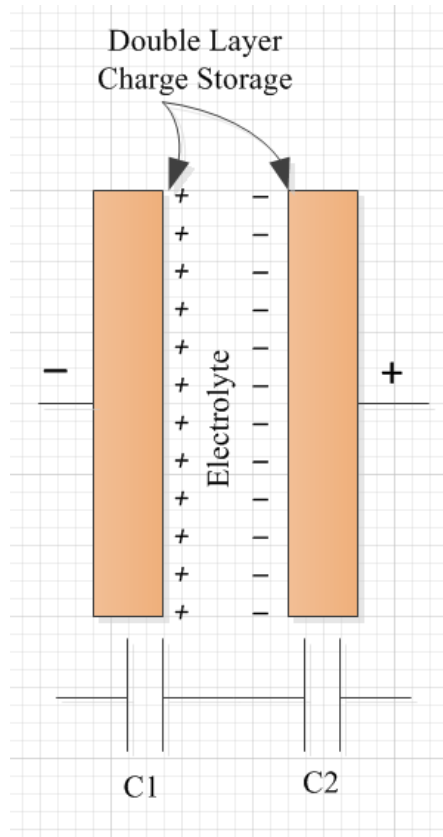


**Figure 6.3.** Depiction of Electrolytic Capacitor

### 6.2.3 Electrochemical Capacitors

Electrochemical capacitors are commonly referred to as “super-capacitors” or “ultra-capacitors” by manufacturers and in the literature when they have a double layer. These capacitors store electrical charge in an electric double layer at the interface between a high-surface-area carbon electrode and a liquid electrolyte as shown in Figure 6.4.

The ultra-capacitors represent a compromise in design between rechargeable batteries and standard capacitors. Capacitors can provide significantly higher power densities (or current densities) than batteries; however, their energy density is lower by about two to three orders of magnitude. Ultra-capacitors achieve significantly higher energy densities than standard capacitors, but retain many of the favorable characteristics of capacitors, such as long life and short charging time.



**Figure 6.4.** Depiction of Electrochemical Capacitor

Ultra-capacitors store ionic charge in an electric double layer to increase their effective capacitance compared to only storing charge across a dielectric material, as is the case with traditional capacitor technologies. Through the addition of an electrolyte, the resulting energy density of commercially available ultra-capacitors is about one order of magnitude higher than standard capacitors and about one to two orders of magnitude lower than rechargeable batteries and, as a result, appear to be an attractive replacement for batteries in wireless sensor applications (see Table 6.1).<sup>[5]</sup>

### 6.3 Characteristic Interdependencies

- Capacitors typically have high power density, but lower energy density compared to batteries.
- Because of the higher energy density of super-capacitors, super-capacitors and ultra-capacitors have been examined as energy storage options in conjunction with renewable resources, fuel cells, and other generation technologies<sup>[7]-[9]</sup>.
- In energy-storage functions, capacitors should be connected to DC.
- Where a wireless sensor does not need significant energy (mW), ultra-capacitors can be sufficient.
- In applications, where energy is not the necessary driving factor, capacitors with faster response times, such as electrostatic and electrolytic capacitors, are more appropriate.

## 6.4 Typical Applications

Capacitors are typically found in applications where a buffering agent is needed. This is a result of the high power density and low energy density of these devices.

## 6.5 Relative Cost Impacts

These devices can be extremely expensive because of the low energy density compared to other technologies. Table 6.1 shows a range of prices for various types of capacitors.

**Table 6.1.** Cost Estimation for 1 MJ or 277 Wh Storage

Capacitor Type	Mass (kg)	Volume (m <sup>3</sup> )	Cost (\$k)	Response Time (s)
Electrostatic	200,000	140	700	10 <sup>-9</sup>
Electrolytic	10,000	2.2	300	10 <sup>-4</sup>
Electrochemical	30-100	.02-0.1	2-20	~1

## 6.6 Summary/Conclusions

Capacitors can be an option for sensor power supplies, but only if the amount of energy required for an extended period is extremely low. These devices are best suited in combination with an energy-harvesting or source-excited technology.

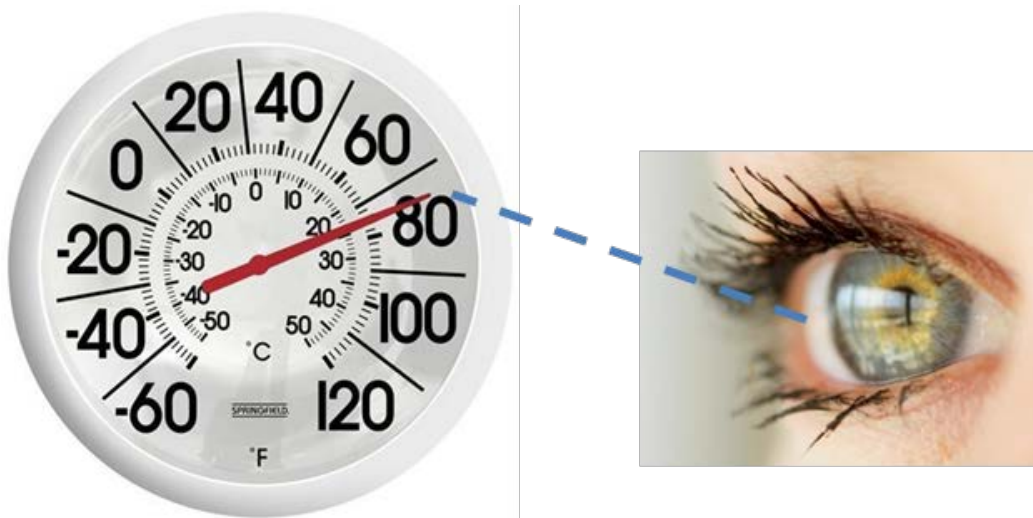
## 6.7 References

1. S. Atcitty, Electrochemical Capacitor Characterization for Electric Utility Applications, PHD Dissertation, Virginia Polytechnic Institute and State University, November 2006.
2. J.M. Miller, Ultracapacitor Applications, IET 2011.
3. J. R. Miller, "Electrochemical Capacitors (ECs): Technology, Applications, and Needs," Presentation Electrical Energy Storage Workshop April 2007.
4. B. Somanathan Nair, Electronic Devices and Applications, Prentice-Hill 2002.
5. S. Roundy, M. Strasser, P. Wright, "Powering Ambient Intelligent Networks,"
6. J. Miller, P. Simon, "Fundamentals of Electrochemical Capacitor Design and Operation,"
7. F. Simjee, P. Chou, "Everlast: Long-life, Supercapacitor-operated Wireless Sensor Node"
8. C. Park, P. Chou, "AmbiMax: Autonomous Energy Harvesting Platform for Multi-Supply Wireless Sensor Nodes"
9. S. Kim, K. No, P. Chou, "Design and Performance Analysis of Supercapacitor Charging Circuits for Wireless Sensor Nodes"

## 7.0 Communications and Networking: Introduction

A sensor is a device that observes and quantifies one or more specific physical characteristics and communicates that information to something or someone who can act upon the information. The sensor must not only measure (or sense) the physical characteristic, it must be able to communicate the measured result. A sensor is not useful unless the measurement is made available for some action or observation.

A common outdoor dial thermometer, for example (Figure 7.1), translates the thermal energy content of the outdoor air into mechanical displacement of a pointer around the face of the thermometer. The measured temperature is communicated by the pointer's location relative to the thermometer's face. A person observing the thermometer can then decide if a jacket might be needed based upon the temperature shown.



**Figure 7.1.** A Sensor must not only Measure but also Communicate the Measured Result such that an Action can be Performed for it to be Useful

The same holds true for an electronic sensor that measures a characteristic and communicates the measurement data electronically to an entity that acts on that data. The action can range from displaying or recording the data to an autonomously driven system that uses the data in a feedback loop to modify the outputs (e.g., a temperature controller). The data transfer can use a single technology at all communication points or a mix of many different communication techniques depending on the system requirements. Furthermore, the response timing can range from near real-time (as fast as possible) to a post-processing system that delays any action until a specific time or event.

The required responsiveness of a system and the complexity of the response are just two system requirements that need to be defined before the network technologies can be selected. However, selection is not a simple process because the available communication technologies can restrict a system's requirements/capabilities; the communication link can provide strict limits to a system because of, but not limited to, the following factors: data complexity, necessary security, available power, monetary cost, size constraints, acceptable delay, and/or the number of devices in the field.

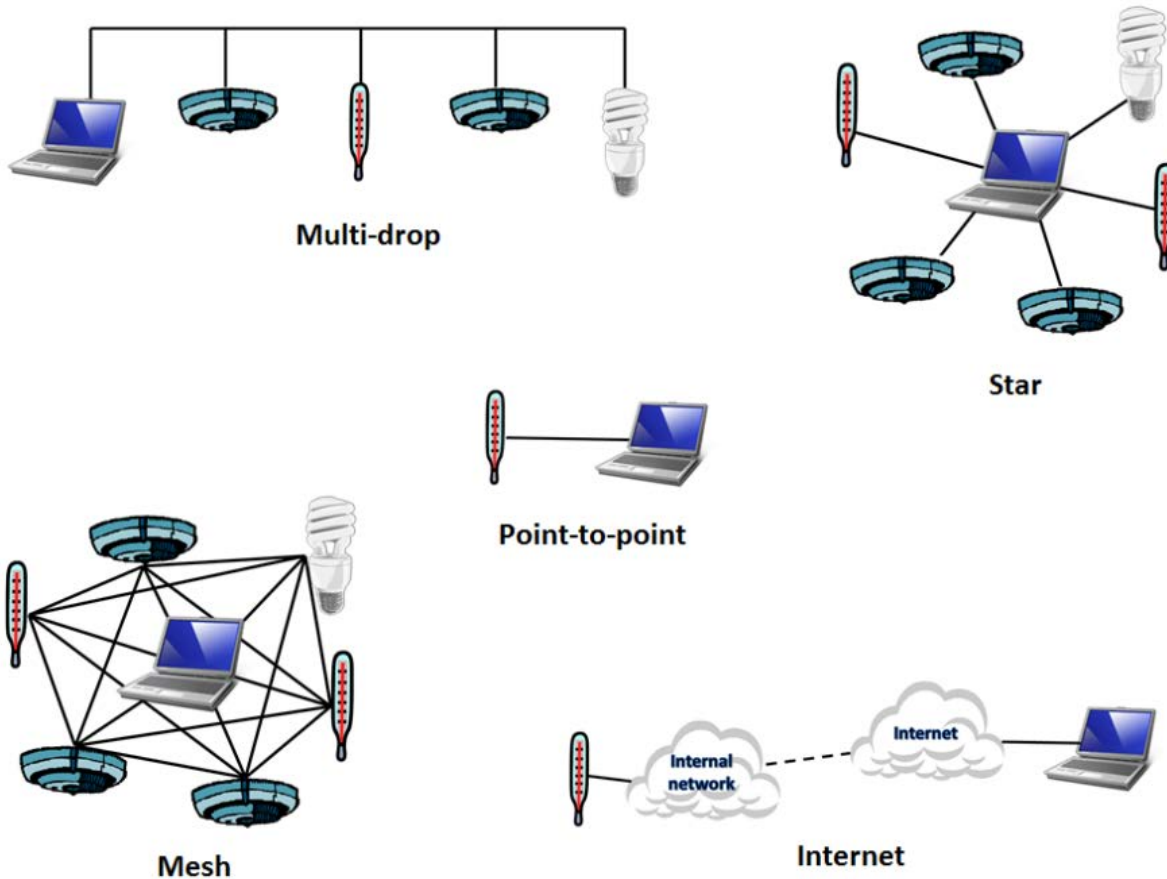
## 7.1 Network Topology

In general terms, network topology is the interconnection of networked devices. A network topology defines which devices are connected to which other devices either through a direct connection, a logical connection, or both. In building sensor networks, it is common (but not necessary) to also incorporate the physical locations of the devices with the network topology. For the most part, a network topology is a map describing how everything is connected. When routing a message across a network, this map or pieces of the map are stored at each device along the route. From the partial network topology stored on each device, a device determines how to route a message. The device determines where to send the message next, such that it will eventually reach the final destination. For example, when routing a package through the postal system, it is necessary to know the destination address and what facilities are directly connected to the facility that currently has the package. The sorting process then determines which facility to send the package to next, using the local network's topology to determine which facility would be the best next intermediate destination based on the final address such that the package arrives to the final destination in a timely manner.

Network topologies come in a number of different varieties. Some common topology types include star, mesh, multi-drop, and point-to-point networks. Each of these different topologies has its own pros and cons depending on the system and communication requirements. Point-to-point networks are the simplest type because there are only two devices that need to communicate with each other. Examples of this would be a device that is connected via Universal Serial Bus (USB), Bluetooth, or a serial connection. The multi-drop network is more complex because there are multiple devices that all share the same communication line. However, this type of connection provides a simple solution for extending a network because the next device only has to connect to the previous device. Star networks provide a simple addressing and routing mechanism, but limit the number of hops to a single hop between a device and some centralized device. A common star network topology is a router, such as the wireless routers found in many homes, where all of the devices communicate with each other through the router. Therefore, if the router were to fail, the entire network would fail. Mesh networks require additional routing and addressing schemes but eliminate the single point-of-failure issue of a star network. Therefore, if any one device fails in a mesh network, it is possible that the rest of the devices connected to the network can still communicate (see Figure 7.2).

## 7.2 Communication Specifications and Requirements

Communication requirements are specified by the required bandwidth, allowable delay, allowable error, communication reliability and robustness, and scalability of the network. Communication requirements also take into account environmental concerns that need to be addressed, such as communicating environments with large metal objects, tight spaces, high background noise levels, required low signal levels, and interference by other devices, as well as message corruption.



**Figure 7.2.** Different Network Topologies have their own Benefits and Drawbacks

### 7.2.1 Corruption

Corruption of a message can occur when one or more bits<sup>1</sup> in a message are flipped during transmission. Corruption of a message can also occur when the message is truncated because the received signal does not meet the protocol's specification.

### 7.2.2 Bandwidth

Bandwidth defines the amount of data that can be sent at any instant with units of bits per second (bps) or bytes<sup>2</sup> per second (Bps). The larger the bandwidth of a link, the more data the link can send in a given time period. Think of it this way; the size of a truck determines the available bandwidth when shipping a package, where a package represents a message. The truck can only hold a certain volume of packages at a time, and the larger the packages are, the fewer number of packages that can be shipped on the truck. If a second truck has to retransmit a package because the first one was broken, then there is overhead associated with reshipping the same package. Likewise, it takes longer to load many small packages than to load a single large package because of the overhead associated with moving each individual package.

<sup>1</sup> A *bit* is a single binary 1 or 0.

<sup>2</sup> A *byte* is a string of 8 bits. For example, 00010110, which is 22 in decimal format.



### 7.2.3 Throughput

Throughput refers to the amount of data that is sent through a link in a specified amount of time and is expressed in units of bits per second (bps) or bytes per second (Bps). As an example, when shipping bulk items such as bananas, the throughput of the shipping company is determined based on the quantity of bananas that arrive at the destination in good condition in a specific time period. Extra packaging can help ensure that the bananas arrive unharmed. However, extra packaging adds overhead because the packaging takes up extra space and time and, as such, reduces the throughput of the system. Likewise, if bananas arrive damaged, then the throughput of the system is also decreased because the bad bananas have to be thrown out, and it was a waste of effort to ship them in the first place (Figure 7.3).

### 7.2.4 Delay

Total delay is the sum of the amount of time it takes to push data onto the physical medium (transmission delay), the time it takes for a message to get from one point to the next (propagation delay), as well as the amount of time during which a message is being processed (processing delay) and waiting to be sent because of high data traffic (queuing delay). For example, when shipping a package, the further the destination is from the origin, the longer the delivery of the package takes. At the same time, if it is around a major holiday and a very large total number of packages is being shipped, there could be an increased delay while a package sits at a distribution center waiting to be processed, which can increase the total delay of the package.

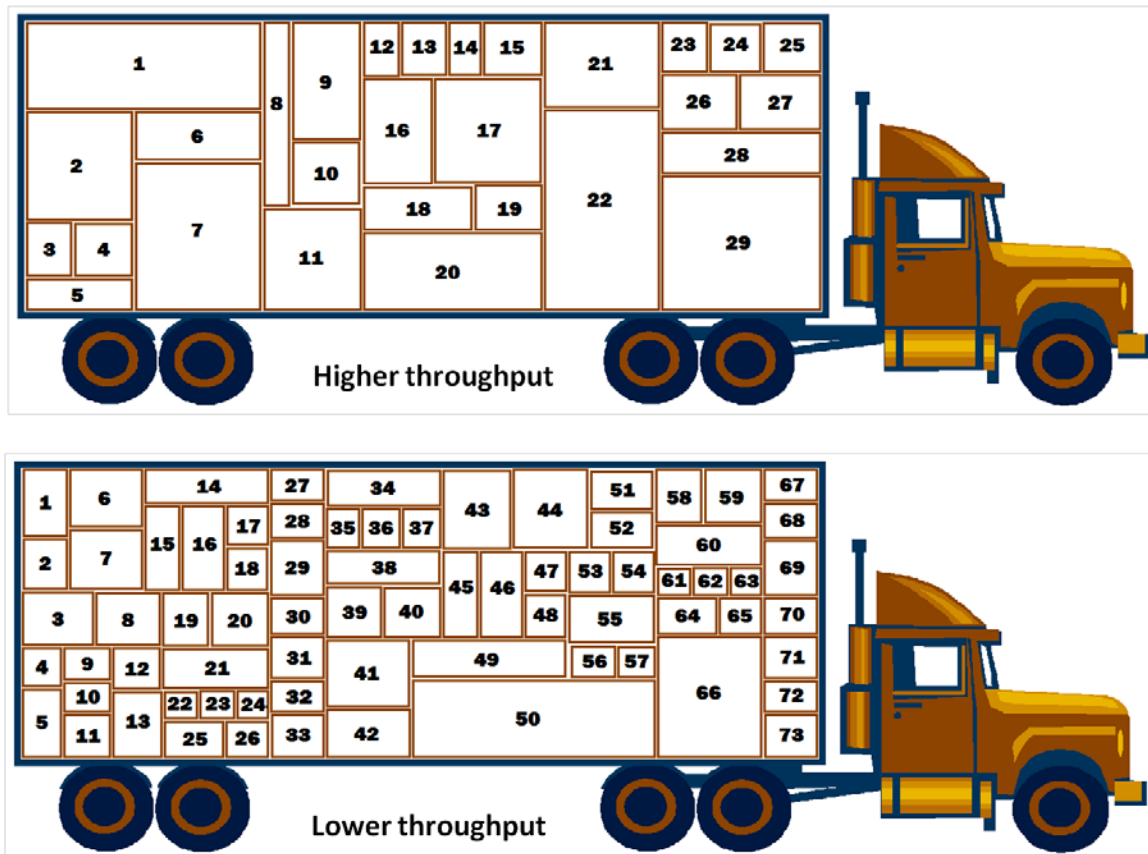


Figure 7.3. Throughput of a Connection: Shipping a Package

### **7.2.5 Reliability**

A data link's reliability is determined by how many errors the link generates including both message error and lost messages. A link is 95 percent reliable if the message is received at its destination completely and uncorrupted at least 95 percent of the time. A link's reliability affects the delay of a message as well as the required bandwidth of a link because bad messages might have to be retransmitted.

### **7.2.6 Robustness**

The robustness of a network is a qualitative property defined by how well the network continues to operate in the event of a failure. A failure can be caused by an unreliable link, a link that was damaged, or a device that failed. A device failure could correspond to some component on a device (e.g., its sensor, processor, radio, etc.) failing or being damaged, or it might just be that the device batteries need to be charged or replaced. In a robust system, losing a device or receiving a corrupt message should not create problems. In the case of a critical device, backups would take the place of a failed device or a mechanism would be in place to ensure that any critical message was delivered. For example, when sending an important letter, using certified mail enables the sender to receive notification regarding whether the letter was received and, if it was not, the sender can resend it. A robust system also provides alternative routes for delivery of data or a message. For example, a robust postal system would have mechanisms in place to ensure that a letter reaches its destination, even when a post office or mail facility is shut down, by routing mail around the problem. The required robustness of a system depends heavily on its requirements. Safety critical systems, for example, require a high level of robustness to ensure that if a failure occurs, backups and procedures are in place to recover.

### **7.2.7 Scalability**

Scalability refers to how well a system can handle and provide service to an increased number of devices without impacting the rest of the system. For example, around the Christmas season, a shipping company has to scale up to provide service to more customers without negatively impacting normal service to its customers. To do this, the shipping company hires temporary workers to handle the extra work load, which allows them to scale down later in the year and save money.

### **7.2.8 Signal-to-Noise Ratio**

The signal-to-noise ratio indicates how "loud" a signal is compared to background noise. It can be used to specify how loud the signal must be for it to be received. Think of how loud someone must talk to be heard by the person sitting next to them in a quiet office compared to at a rock concert. In the absence of raising their voice, the signal-to-noise ratio would be much smaller at the rock concert than in a quiet office.

### **7.2.9 Interference**

Interference is caused when other devices transmit at the same time using the same or a similar physical medium (but not necessarily the same protocol). For example, there is interference when talking with someone else in a crowded room, even if everyone else is speaking a different language.

### 7.2.10 Message Contention

A message contention occurs when two or more devices transmit at or around the same time. For example, there is message contention when two or more people start to talk at the same time. Sometimes the loudest person is understood; other times the closest person or no one at all is understood.

### 7.2.11 Extensibility

Extensibility is a qualitative property that defines how easily a system can be extended by incorporating additional functionality or modifying existing functionality.

## 7.3 System Requirements

An important piece to the networking puzzle is to determine how networking fits into overall system requirements. Realistic requirement gathering is the key to building a successful system at a minimal cost. For example, if the goal is to turn off a light when no one is in a room, a complex system (where there are occupancy sensors in each room that report to a centralized data acquisition device with the ability to control each light in the building) might be overkill. A more appropriate and cost-effective solution might be to install special light switches that have occupant motion detectors and provide simple, local control.

Many different characteristics of a system are used to determine network requirements including, but not limited to, the data complexity, sampling rate, number of devices, reliability, robustness, guarantee of service, system lifetime, available power, monetary cost, allowable delay, and allowable error. When trying to define a network parameter, such as the required bandwidth, many, if not all, of the different system characteristics must be reviewed to determine the impact it has on the network parameter. This review also needs to ensure that the system is capable of meeting the requirements while not being exuberantly over specified, leading to an unnecessary increase in cost.

As an example, to determine the bandwidth requirements of a system designed to monitor the temperature of a room, the sampling rate of the temperature sensor and the number of sensors need to be considered. The sampling rate determines the speed at which data are generated at a single device. Coupled with the number of devices in a system, it is possible to determine how much data the entire system generates in a specific time period. This information, along with other parameters such as system reliability, is used to determine the bandwidth requirements for each part of the system. Before calculating the bandwidth requirements, however, the specified sampling rate, as well as density of devices, must be reviewed to ensure that they meet the objectives without being excessive.

Consider, for example, a sampling rate of once per second (which might be a little extreme because it is unlikely the temperature of a room changes significantly after only 1 second). A system using such a rapid sampling rate would require additional bandwidth to communicate the significant amount of redundant data. Each temperature reading would most likely be very close to the previous reading. This could drive up the cost and power consumption of the system. A more realistic update interval might be 15 minutes, the specific sampling rate depending on how quickly the room's temperature changes, the accuracy required, and in some cases, how quickly the system needs to be able to react to a change in the environment. It also is possible to design a system that only communicates data when a certain criterion

is met. For a temperature sensor, a requirement could be made such that the current temperature is sent once per hour, or if the temperature changes by 3°F or more from the last reported temperature.

It also is important to note that different sensors have different required update intervals and data complexity. Data complexity relates to the intended use of sensor data and not the sensor alone. The complexity of data from a temperature sensor when measuring the temperature of a room is low because the sensor typically outputs a single value (the current temperature) for each sample. An accelerometer used to analyze the frequencies of vibrations, in contrast, has a higher data complexity because this application can require sample rates of thousands of samples per second. A single data point from the accelerometer indicates the current acceleration at a specific time only. Vibrations are mechanical oscillations about a point of equilibrium with the acceleration constantly changing. A list of data points from an accelerometer captured at a fast sample rate can be analyzed to determine if a specific frequency of vibration was a component of the recorded data. Analyzing the current acceleration can be used to determine whether an object has received a large shock (e.g., from being dropped) but by itself, not the frequency of vibrations. Analysis of the frequency of vibrations is more complex and can be used to determine, for example, whether the structural integrity of a bridge is intact or a motor or fan is running. Thus, data complexity is a factor that needs to be considered when determining the required network bandwidth.

Other system parameters, such as the number of devices, can affect the required bandwidth. Having more devices creates a need to communicate more data as the number of sensing devices increases. The reliability, robustness, and quality of service specifications provide additional bandwidth concerns. The system might have to deal with retransmission of messages for reliability and multi-path routing for robustness or quality of service. Multi-path routing involves a message being forwarded through two or more paths in the network with the goal that at least one copy of the messages arrives at the desired destination. Finally, the lifetime of battery-powered devices places strict constraints on the available bandwidth of a system. An increase in message traffic (bandwidth usage) increases the power consumption, reducing the lifetime of batteries and, thus, decreasing the system lifetime.

This discussion only starts to define all the factors needed to determine the bandwidth required for communications. When defining any network requirement, every other system requirement needs to be reviewed to determine how and if a change in a system requirement directly affects the communication requirements.

Critical systems are special in that reliability, robustness, security, and quality of service are especially important and are given precedence over all other requirements. When dealing with a critical system, such as a fire alarm, it is imperative that the system is able to detect and report a fire even when the power in a building is out. As such, smoke detectors, even when they are wired into a building, have a battery that enables the system to operate even when there is a loss of line power.

## 8.0 Communications and Networking: Network Management

Managing a network includes managing all of the different parts of a network and ensuring that the system is working as specified and meeting (at least) the minimum requirements. This involves keeping records of the network configuration and maintaining the system by performing tasks such as changing batteries, replacing failed components, and reconfiguring devices that have corrupt or outdated configurations. It is important that maintenance is performed on a system.

When a network is expanded by adding functionality and sensors or by increasing the sampling rates, the stress and strain on the network also increase. Therefore, the network manager needs to determine how scalable a system is before expanding it. Some systems have a maximum permissible number of devices that is dependent on the number of device identifications (IDs) available. Other systems are limited by the amount of data that can be transmitted, amount of space available for the device and its components, and the number of hops between a sensor and the point where the sensed information is viewed, analyzed, or stored. When incorporating new features, the extensibility of a system (how easy is it to add another feature) must be reviewed and evaluated before components are added.

### 8.1 Configuration

Configuration of a device is different for each device. It includes, but is not limited to, installing the device, connecting any cables, starting the device, and performing any initialization that is required. The initialization can include setting the current time, sample rate, location, device ID, and security tokens as well as informing the device of any neighboring devices, end points that recorded data need to be sent to (e.g., data analyzers or storage), and possibly routing paths that devices should use for data to reach the desired destinations. The configuration required by each device depends on the device itself, its functionality, communication protocols, and the network capabilities and requirements.

Some configuration tasks can be performed autonomously by a system (i.e., without user input), but the degree of autonomy depends on the communication protocols, network capabilities, and network requirements. Autonomous configuration helps to reduce the amount of human involvement but can require additional components that increase the system cost or require additional power, making them infeasible for some systems.

System requirements drive the level of configuration. Consider two example systems. The first utilizes data loggers and post processing to store and analyze data. Devices (data loggers) are placed in a building and locally record data. After a period of time, a person collects the data from each device. In the second case, data are recorded and analyzed at a networked data analyzer in near real time. Therefore, as soon as the sensor device gets data, it sends it to the central location for analysis.

- **Case 1:** In a basic example, configuration tasks could include placing the logging devices throughout the building and recording a start time for each device. Then, at a later time, the desired data are extracted from each device, and the time and location of the device are recorded.
- **Case 2:** For a more complex system that records data in real time through a wireless link, the data handling process assigns a unique identifier to each device and makes it available to the communication process (this requires that the unique ID be available at the code level to tie it to each message). Each device is installed in the building, and the location is recorded. The device location

is associated with the device's unique identifier so the networked data analyzer can determine the location of each device. Communications with the other devices (in its range) are then verified.

A system's capabilities determine the resources available to the system to provide for autonomous configuration. For example, if each sensor device has a component with a unique identifier and a localization mechanism (e.g., global positioning satellite [GPS]), the device can autonomously associate its unique ID and location to each message, as is required in Case 2 above. However, these capabilities increase the cost of a device and draw additional electric current for operation. Furthermore, if the system is used indoors, a GPS signal will most likely not be available for localization because use of a GPS typically requires a clear view of the sky.

The communication protocols affect the level of configuration because each protocol can have a different set of requirements. Use of an USB link, for example, can be as easy as plugging in a cable to the correct port. Use of an Ethernet cable, on the other hand, can require that the Internet Protocol (IP) address of the computer be changed and firewalls be opened to enable communication. Furthermore, configuration tasks for the communication protocol can include setting up each link (e.g., plugging in a cable or selecting the correct channel) and setting up a routing path for each device so it can communicate across multiple hops.

Autonomously configured devices are convenient, and today everyday cell phone users rely on this feature, but it can be costly and increase power requirements. For cell phones, user behavior has shown that spending the extra money for the automatic network configuration feature even with the additional burden of needing to charge a phone every night is an acceptable tradeoff. However, sensor devices are expected to last for months or possibly many years without user intervention (e.g., having to replace or charge a battery).

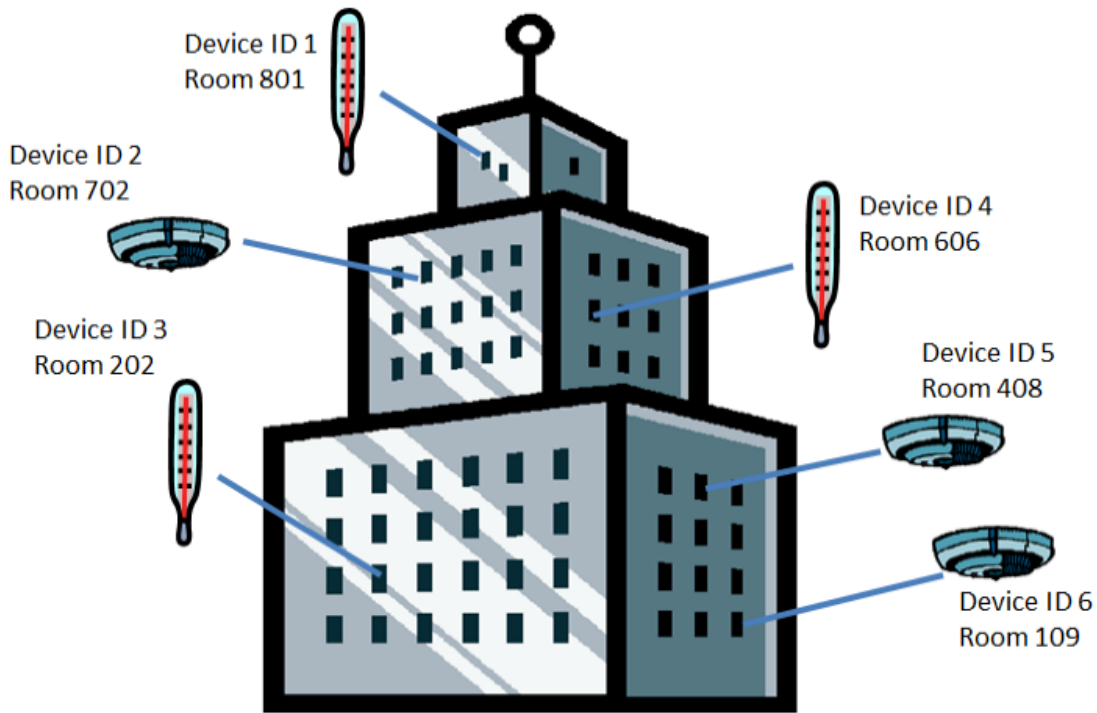
Manual configuration, on the other hand, can increase the work required to set up the network. Unless stringent records are kept, trying to track down all of these locations can be difficult, and mistakes will probably be made. Therefore, when dealing with manual configuration of electronic devices, a documented procedure should be followed to minimize mistakes, and stringent records should be kept so that later the state and configuration of a system are known.

## 8.2 Localization

Localization deals with determining the physical location of a device and associating that location with the sensor's data. In sensor networks, the location of the data can be as important as the sensor data itself. If a system has an occupancy sensor in every room, but there is no location associated with each sensor, then the system could report how many rooms had people in them but not which rooms were occupied. Such a system might not be helpful to a building operator who is trying to minimize electric power consumption in unoccupied rooms because the location of the rooms with no occupants would not be known.

Localization is a difficult problem that provides significant benefits to a system. The manual solution is to record the location of each device when it is installed, replaced, or moved and store it in a database. This solution works well for static/fixed devices (i.e., devices that do not move). For a system to

associate the location to the data received, a unique ID needs to be recorded with the received data, and the system needs to have access to the database that ties the unique ID of a device to its location (see Figure 8.1).



**Figure 8.1.** It is Important to Know Where each Device is Placed so that it is Possible to Correlate the Data from a Device with the Location of the Device. Without this correlation, a system can only determine that an event is occurring but cannot inform the user where the event is occurring, which can be as important as knowing that the event occurred.

For sensors that are moved frequently, manual configuration of the device's location typically leads to errors and additional workload because the location of the device must be updated every time it is moved. In these cases, automated localization can be used to ensure accurate location information is continuously available. GPS positions or cellular triangulation are common localization mechanisms for wireless systems located outdoors (provided these solutions do not make the system too expensive). For indoor localization, a number of solutions exist, but determining which solution is best is highly dependent on the environment and system requirements. Continual determination and updating of the device's location increase its power consumption. Power consumption issues can be mitigated somewhat by reducing the frequency of localization to the extent compatible with the application (i.e., a device that updates its location once per second uses more power than a device that updates once per day).

### 8.3 Concepts of a Communication Protocol

A communication protocol is a set of messages and rules that defines the syntax, semantics, and synchronization of the communication so two or more devices can communicate. Beyond merely describing message formats, order, and timing, communication protocols can incorporate capabilities such as signaling, error detection and correction, authentication, and encryption.

The spoken English language, for example, is a communication protocol that utilizes the air as a physical medium and sound waves as a signal. In its most basic form, the English language is composed of different sounds. Each sound alone conveys little to no meaning but when organized using the syntax and semantics of the English language, a set of sounds can be organized to form a word, which conveys more meaning than if the sounds used to compose that word were uttered in random order.

Even though the English language is thoroughly defined, communication can still be a challenge. When talking with another person, signaling mechanisms are used to control the conversation. For example, visual cues help to determine when the other person is about to speak, which helps to reduce the number of times both people talk at the same time. However, both people talking at the same time still occurs, which typically results in both people stopping, then a pause followed by one person trying to start again or informing the other person that they can talk first.

These problems are uncommon when speaking face to face but when talking to someone over a radio, a delay occurs and interference is introduced that can create many problems. To solve these problems, some form of synchronization and error detection/correction needs to be used. Thus, voice procedures can be used to standardize spoken communication. In this case, there are special words and phrases that have a certain meaning. “Loud and clear” is used to signal that the person heard and understands what was just said, whereas “say again” is used to signal that the message was not understood and please try again. The term “over” informs the other person that they can talk and “over and out” indicates that a person is leaving the conversation.

Similar to the voice procedure example, electronic communication relies on communication protocols to define the communication between devices. However, where humans can make a guess at what was said based on the context of the conversation and sort out many mistakes, electronics do not have such luxuries. If a signal isn't quite right or a pause is too short or long, the communications can have problems. The electronic communication link or connection is composed of many different parts: the physical connection between devices, the transmission link between two physically connected devices, the logical connection between any two devices that are connected through any number of devices that act as forwarders, and at the highest level, the data that is communicated across a network. Problems can occur at all of the levels of the communication from a sound being not quite right to a sentence being out of order, and each level needs to be able to fix the issues to continue a conversation.

## **8.4 Physical Connection**

Communication between two or more devices requires a physical connection that carries a signal, enabling the devices to communicate. Any medium capable of carrying a signal can be used as a physical medium with the most common being copper wire, optical fiber, air, and water (see Figure 8.2). The signal can be analog or digital, and can be implemented using a number of different technologies, including the state and/or intensity of light, the current, capacitance, or voltage of an electrical signal, and RF electromagnetic radiation.

The physical connection determines the connector, cable type and maximum cable length, as well as properties of the connection, such as the signal transmission level, signal strength, bandwidth, impedance, frequency range, and bit modulation. Depending on the physical medium functionality, such as carrier



sense (if the medium is in use) and collision detection (detecting whether two messages collided), information can be provided to higher-level protocols.



**Figure 8.2.** Different Physical Media

## 8.5 Direct Connection

The direct connection is responsible for ensuring that a sent message is received in a complete and uncorrupted state. This connection deals with the communication between devices that are physically connected to each other (i.e., a one-hop connection). To determine if a message is complete, some type of signaling is required to inform the receiver when a message starts and ends. This signaling protocol can be handled in many ways depending on the physical connection.

When using an asynchronous serial connection in which messages can be sent at any time, a simple way to signal the start and end of a message is to include special start and stop identifiers in the message. When the special symbols identifying the start and end of a message are received, the receiver knows that the message started or ended depending on which symbol was received. However, this solution requires that the special symbols only be used to signal the start and end of a message, which is not always possible. For example, when sending binary data, special symbols cannot be used because the binary data could contain the same symbols as those used for these purposes. In this case, the receiver would end up incorrectly starting or ending a message whenever it received the special symbol, which could be in the middle of a message. A well-known solution is “padding,” in which an extra copy of the special character is inserted before every instance of that character, except for where it has special meaning. When decoding the message, if there is only one special character, it designates the start or end of a message, but if there are two of the special character in a row, it designates that only one of the duplicates should be kept and passed through as data. Padding is a sufficient solution to this problem, but it increases the complexity of sending and receiving data and, therefore, the overall cost of the system.

One way to determine if the message sent is actually received is to use acknowledgments (ACKs). When the receiver gets a message, protocol requires that the receiver send a short reply indicating that the message was received. The word “over” when talking on radios is an example of an acknowledgement. If the sender receives the acknowledgement, it continues normally, but if it does not receive an ACK, then after a period of time, it resends the lost message. However, if the message was received and actually the ACK was lost, the receiver will get the same message twice and needs to discard the duplicate. To do this, a simple sequence number identifying a message can be used to easily designate when a message is a duplicate.

For battery-powered systems, acknowledgements may not be used because increased communications cause greater power consumption leading to shorter battery lifetimes. Therefore, typically wireless systems have no guarantee that a message was received at the next hop. The thinking behind this is that it is better to send more messages (e.g., increase the sampling rate of the sensor) and hope that some of them get through than it is to send a reply to an already short message. Also, when dealing with streaming data such as video, acknowledgements are not always possible because data are generated and consumed at a high rate. In such applications, by the time a message would be resent, the data are useless because it is outdated.

Error detection is used to determine whether a message was corrupted during communication by sending an encoded/compressed copy of the message along with the original message. At the receiver, a message is validated by performing the same encoding and then comparing it to the sender-encoded version. If the sender and receiver do not agree on the encoding, then the receiver assumes that the message is corrupt.

Some form of error correction is needed to correct a corrupt message. One solution is to use a negative acknowledgement (NACK) that requests that the sender resend the bad message. This solution encounters the same problems as described above for ACKs with resending data when dealing with streaming data or battery-powered applications. Forward error correction might be able to solve this problem. Forward error correction would use the encoded version of the message to identify and fix the errors in a message. However, forward error correction has an increased communications overhead because it requires the encoded version of the message to be longer than when performing simple error detection.

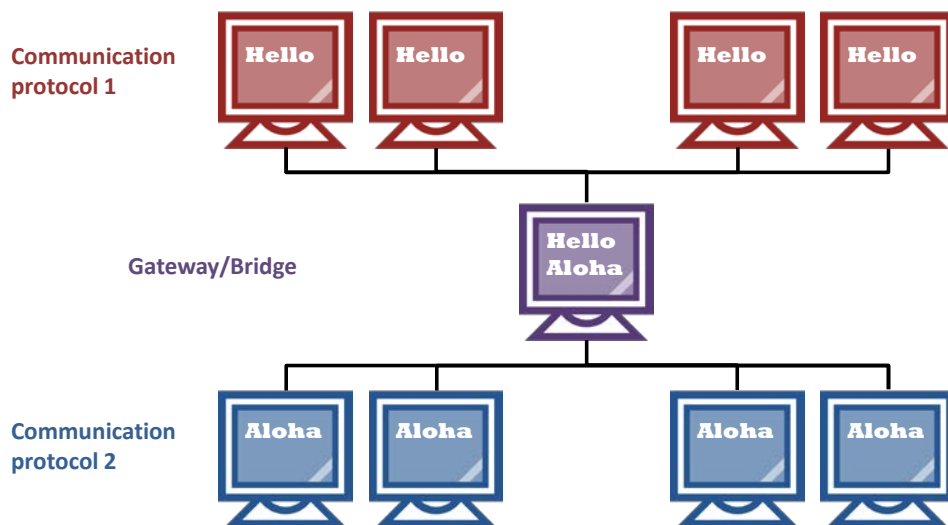
Each system has its own unique set of requirements that dictate the communication requirements. The additional cost associated with features that provide some level of confidence that a message is received in a complete and uncorrupted state need to be weighed against the unique set of requirements for that system. While it is desirable to have a guaranteed level of reliability, each added feature increases the system cost and power consumption, factors that need to be considered relative to the value of the added feature.

## 8.6 Logical Connection

A logical connection between two devices exists when the devices can communicate with each other. The logical connection can be a single-hop connection or a multi-hop connection. The single-hop connection is a direct connection where the two devices communicate directly with each other and are connected to opposite ends of the same communication medium. A multi-hop connection is a series of

direct connections where each device in the series forwards the message to the next device until the message is received at the destination. A single-hop connection can be thought of as handing a letter directly to the recipient of the letter, whereas a multi-hop connection would be analogous to sending a letter through the postal service (which may involve several intermediate stops before reaching the destination). In the post office example, a logical connection between the sender and recipient exists because there is a set of direct connections (sender to mailbox, mailbox to mail deliverer, mail deliverer to sorting facility, etc.) that together are capable of forwarding the message from the sender to the recipient. As with the postal service, a message sent across a logical connection between two devices might take different paths. When sending two letters from Seattle to New York, the first letter might go through Chicago and the second letter through Denver. The sender and recipient care only that the letter makes it to the final destination in an uncorrupted state and in a timely manner.

Different communication technologies can be used to create a logical connection between devices provided there is a gateway between the two networks that knows both communication protocols and is capable of translating from one protocol to the other (see Figure 8.3). For example, if a person receives a letter in a language that they cannot read, they need to find an interpreter that understands the content and can rewrite or read it in a language that the original recipient understands. The interpreter acts as a gateway or network bridge allowing the participants to communicate even though they use different languages. This type of logical connection is common in sensor networks because sensors frequently communicate with some type of “reader” device using a low power protocol while the reader sends the data to another device using a more reliable protocol with additional capabilities, such as the ability to connect and transmit data through the Internet.



**Figure 8.3.** The Purple Computer in the Middle Acts as a Network Gateway or Bridge and Translates from the Protocol of one Network into the Protocol of Another and Vice Versa

As with the direct link, acknowledgements and error detection/correction methods are used to ensure that a sent message is received completely and uncorrupted. Even if the direct connection is performing these checks at each segment (hop), it is possible to lose or corrupt messages across a multi-hop path so an end-to-end check of the message integrity might be necessary.

Logical connections also can have an additional requirement to ensure that messages are delivered in the order in which they were sent. Each message can be routed by forwarding devices along different

paths to the destination. Furthermore, each segment along a path from source to destination introduces a different delay (e.g., based on the physical links and its current state). As a result, a set of messages originating from the same device can be received in a different order than sent. Therefore, the logical link might need to make sure that it can put the messages back into the same order in which they were sent. Keeping track of message order can be accomplished using a sequence number to identify the order in which messages are sent.

## 8.7 Data Communication

With the logical link in place, any two logically connected devices can communicate with each other. However, even though they “speak the same language,” this doesn’t mean that they can conceptually communicate data between each other. For example, when a simple temperature sensor sends a temperature reading, it will most likely include other information along with the reading (e.g., device ID, timestamp, and location of the reading). The format of the data has yet to be specified by any of the underlying protocols, and therefore, another communication protocol is needed. Typically, data decoding is done by an application that runs on the analyzer. The analysis application must be able to determine the capabilities of each sensor device it is connected to, as well as how to decode the data. Vendors commonly supply an analysis application, but these applications typically only support that vendor’s devices. If other vendor devices are added to a system, the analysis application must be able to support all of the vendors’ devices. This can be achieved in a number of ways including using third-party software capable of supporting all of the devices or creating custom software and tailoring it to a set of specific needs.

## 8.8 Interoperability

Interoperability is the ability of a set of devices to communicate and interact with one another. Usually this refers to devices from different vendors. Interoperability is complex because there are different levels of interoperability in general. Each device in a system requires different levels of interoperability for the whole system to work. For example, if a system has a temperature sensor and a separate light sensor, the temperature sensor does not need to know how to read/decode the light sensor data and vice versa, but if the data from these two sensors are sent to the same data analyzer, it needs to be able to understand the protocols used by both sensors. In contrast, any device that simply forwards information, such as a router, only needs to be able to decode the portion of the messages dealing with sending the data from one point to the next. A device that only transmits data does not need to understand the data it is forwarding for the system to work. For example, a scribe only needs to write down what the speaker says, not conceptually understand the content. However, for the information to be successfully communicated and useful, the intended reader must understand the content.

Interoperability is further complicated when some vendors comply with a specific protocol, such as 802.11 (WiFi), but implement special capabilities beyond those defined in the standard protocol. This may provide for higher performance or the support of special features above and beyond the standard protocol. In this case, devices that meet only the standard protocol may interoperate with the devices with the special capabilities but all of the special features would likely not be available in them. Full interoperability depends on compliance with standard protocols for all features, especially for devices from different vendors.

To determine whether devices are interoperable, a user needs to make sure that the devices that have to directly communicate with each other use the same protocols for every level of communication that is required at the specific device. For example, if a device receives a message that it needs to forward, that device needs to use the same protocols for the physical connections and direct connections as the device that sent the message and the device that the message will be forwarded to. It also needs a way to determine which device the message needs to be forwarded to and how to convert a message to another protocol if it is acting as a gateway or bridge between two networks. It does not need to know the protocol that defines the format of the data. However, when dealing with logically connected devices that are sending and receiving data, not just forwarding messages, it is imperative that the devices use the same protocols to define the data formats.

Some protocols are implemented using different physical mediums. The wireless ZigBee™ protocol, for example, can operate at 2.4 GHz and at 916 MHz, but a ZigBee™ device does not have to work at both frequencies. As such, if one device only uses ZigBee™ at 2.4 GHz and the other runs only at 916 MHz, the two devices are not interoperable even though they both use the ZigBee™ protocol. To make matters more complex, just because a set of devices implements the same protocol does not guarantee that the devices are interoperable. Different vendors sometimes interpret a protocol specification slightly differently. This difference can create problems ranging from no communications to communications via a limited set of operations.

To ensure that the standards are interpreted in a common way and ensure multivendor interoperability, many protocol standards define certification processes and test centers. These test centers, such as the centers set up by the WiFi Alliance, subject devices to rigorous testing prior to allowing the devices to use the special “WiFi Certified” logo. WiFi Certified “guarantees” that devices meet the WiFi communication standard, essentially ensuring that devices can communicate with each other (they read and write using the same characters). The WiFi verification process does not verify that anything beyond the devices is interoperable (e.g., one device could read and write English and the other Dutch, but they could not communicate even though both languages use the same alphabet). For WiFi certified devices to interoperate, additional protocols need to exist above the WiFi protocol (e.g., IP, TCP, and HTTP). As with the WiFi certification process, these upper level protocols need to be tested to guarantee that two devices implementing them according to the standard are interoperable at the next level (e.g., the devices read and write English, specifically American English as opposed to British English, etc.).

Protocols used by vendors can be any combination of proprietary, open, or standard. For proprietary protocols, the vendor does not necessarily publically release any information about the protocol. To obtain the details or use a proprietary protocol, special licensing or the use of special hardware may be needed. In some cases proprietary protocols are open, but depending on the protocol, its use can be limited or require licensing. Extending proprietary protocols that are not open typically requires the vendor to be involved. Open protocols provide the details of a protocol to the public. Even though the details are publically available, special licensing or hardware can be required to use the protocol. Standard protocols are protocols that have been reviewed and accepted by a standards organization (e.g., the International Standards Organization [ISO]). A standard protocol provides a format that can be interpreted by multiple vendors with minimal variance in the interpretation. Standardized protocols are not necessarily free to use because there could be intellectual property tied to them or royalties might have to be paid to use them. However, standard protocols help to alleviate some interoperability concerns. Even with a standard protocol, it is possible that two vendors could interpret the standard

differently, and as mentioned before, without a required verification process for products, interoperability is dependent on every vendor interpreting the standard protocol in the same way. Vendors can accomplish this by communicating with each other regarding details of the protocol while developing compliant product lines.

## 8.9 Security Concepts

While it is obvious why a consumer might want a secure connection when performing an online transaction with a credit card, it might not be obvious why a wireless thermostat needs to be secured. The ability to control a thermostat from a cell phone allows users to change the temperature, for example, of their home remotely. This enables a user to make their home warmer or cooler than would be comfortable after they leave and change the temperature back to a comfortable range before they return home. As a result, energy use and costs can be decreased without the inconvenience of arriving at an uncomfortable home when returning. While remotely controlling a thermostat is a nice feature, without secure communication, it could prove to be a major inconvenience.

If an attacker were to intercept the communication between the thermostat and the cell phone, it would be relatively easy to determine if someone was home and if they had left recently (i.e., the thermostat was changed to allow the house to warm up to 85°F but the current temperature is 74°F). With this information, the attacker could determine the opportune time to burglarize the home. An attacker only wanting to create chaos might just manipulate the communications and set the thermostat temperature to 20°F in the middle of winter. But even this could cause pipes to freeze, which would end up damaging the home and possibly harming plants or animals. The thermostat temperature could then be increased to thaw the pipes and create flooding. Another attack could involve rerouting all of the data packets such that the user would not be able to access the current temperature or remotely change the thermostat. Attacks on commercial buildings or industrial facilities could have much greater consequences. Therefore, the security of control networks in buildings of all types is important to preventing harm from intruders. This section introduces basic network security concepts.

Confidentiality, integrity, and availability are the three basic security concerns for electronic systems. Confidentiality concerns ensure that only authorized users (computer or human) can access information. Availability refers to the fraction of time that the resources of a system are available in a functioning state. A sensor network, for example, is available when any sensor can communicate with the system as needed. Availability can be decreased by system downtime (e.g., for repairs) or by inappropriate use or an intrusion. When a system or a system resource is not available for an unknown reason, it could be reporting confidential information to unauthorized users. Integrity refers to the trustworthiness of information. Information is trustworthy if it was sent from an authenticated source (authentication) and has not been accidentally or intentionally manipulated (data integrity) in any way before it reaches the destination. Manipulation goes beyond merely manipulating a single message as it is sent to the destination and also includes message spoofing (in which a compromised device generates fake messages) and message replay (in which a compromised device replays a message from another device repeatedly).

Electronic systems have multiple layers of security from physical security (e.g., protecting a system with locks and doors) to more computationally complex electronic security. Electronic security ranges from password protection of systems to data encryption using security keys. In building sensor networks,

the level of physical security depends on many aspects, such as building access, wireless versus wired systems, intranet versus Internet, and the locations of system components. The building access determines who can access the building (e.g., public, private, or limited access). If a data network is wireless, all of the signals are broadcast and anyone can “listen” to them if no measures are taken to prevent this. The signals, however, can be encrypted to prevent unauthorized users from understanding them. Before the encrypted data is meaningful to the intended user though, the packets must be decoded (although an attacker could still potentially replay or manipulate them). On the other hand, a wired system requires either physical access to a device or a wire to compromise the system, with the simplest attack involving cutting a wire or destroying a system. If a system is connected to the Internet, an attacker could try to compromise the system from anywhere in the world; whereas, if it is only on an internal network, the attacker must gain access to the internal network before they can attack it.

Because access to a device and/or wire makes it easier to compromise a system, when wires are placed in the walls instead of on the surfaces of walls, finding and tapping into or cutting a wire is much more difficult. Likewise, if network devices are behind closed and locked doors or in the walls/ceilings, it is harder to access them than if readily visible and accessible.

Systems that are protected with a strong password, security cards, and/or biometrics are harder to access and manipulate than those open to public use. Security keys used for data encryption and integrity, as well as authentication, are like passwords. A person without the right key cannot decode a secured message without a significant effort. Such keys can provide confidentiality by encrypting the data in such a way that only authorized users can access the data. It also is possible to achieve communication integrity by using the security keys to verify that the sender is an authentic sender and that the information has not been manipulated in anyway before it reached the destination. While all of this is possible, the complexity of the system is increased, typically resulting in an increase in monetary cost and power consumption, making some of these measures infeasible for battery-operated systems. Security keys come in a few different varieties with solutions ranging in cost depending on complexity. Typically, the more reliable and robust a security systems is, the higher its cost.

Security keys are either asymmetric (public/private keys) or symmetric (shared keys). A sender and a receiver share a symmetric key (i.e., they both know the same key). This allows the information to be encoded and decoded using the same key, where encoding is done by the sender and decoding by the receiver using the reverse process. Therefore, the computational requirements are lower for a symmetric key than for an asymmetric key solution. However, key management becomes an issue because each sender/receiver pair could require their own unique key, leading to a maximum of  $2^n$  keys, where  $n$  is the number of devices. The key management problem has led to many systems only using a single key known by the entire network or a small set of keys known by different groups of devices in the network. This type of solution allows any device that knows the security key to decode and encode messages. This could lead to confidentiality and integrity concerns because any message can be read by any of the devices, allowing more devices access to the encrypted information, as well as allowing any device with the key to potentially spoof, replay or modify a message. Authentication also is limited because a receiver can only check to see if a message came from a node with the same security key rather than being able to uniquely authenticate each device. Therefore, once a single device is compromised, all of the devices that have the same key also are compromised because there is no way to determine whether a message was sent from the compromised device or not.

Asymmetric keys, also known as public/private keys, are used in such a way that anyone can encode information using the public key, but only the authorized individual can decode the information with the private key. This also works in the other direction, where information can be encoded with the private key and decoded with the public key to uniquely authenticate a device. The algorithms for asymmetric key encryption require greater computational resources than algorithms for symmetric key encryption. Each device has two keys, limiting the number of keys to  $2n$ , where each device has to know its private key as well as the public key of any device it needs to communicate with. Typically, public keys are stored on a server that acts as a phonebook and allows each device to get the public key of a device as it needs it. However, it is important that the public key server is authentic and trustworthy; otherwise, it could lead to false authentication of devices because the server could provide fake public keys for false verification of a device's authenticity.

Security keys come in different sizes (bit-lengths), and when dealing with the same algorithm, a longer key provides a higher level of security because it provides a larger set of keys that have to be searched before the correct key can be found. Because of the ways different algorithms handle their keys, a direct comparison of key size from one algorithm to the next is not possible (some algorithms provide a higher level of security with a shorter key).

Key maintenance deals with updating the security keys as needed to make it harder for an attacker to find the right key. The idea is similar to changing a password every few months because if someone does gain access, their access is terminated when the password is changed unless or until they learn the new password.

## 8.10 Wired and Wireless Systems

This section introduces networking and communication for wired and wireless systems, which are covered more extensively in the next two chapters.

When browsing the Internet on a home network, the user sees little difference between using a wired and wireless router. The networking icon is different and the wireless connection allows for mobility, but the differences between the two technologies are invisible when browsing the Internet. Therefore, typical users do not think about any special issues pertaining to one technology over the other. Also, when dealing with consumer electronics, it is only a minor inconvenience to pay a little more, move to a certain spot to get service, or have to charge a battery every night.

In building sensor networks, the implementation details can have a large impact on the final cost, required maintenance, as well as the performance of a system. Understanding the concerns associated with a technology and selecting a suitable solution for the unique constraints and requirements is important. In many cases, the cost associated with cabling is prohibitive, leading to the use of wireless solutions for communication. Wireless solutions can provide a perfectly acceptable solution for a building sensor system, provided some special concerns are addressed. When dealing with wireless systems, it is important to know that it is a broadcast system that emits a signal into the air or water. As such, any other device that uses and or generates a similar signal can impact or be impacted by the introduction of another system. Interference can be generated by a number of sources including devices that don't communicate wirelessly. A microwave oven, for example, operates around 2.4 GHz and, if its shielding is inadequate, it can create interference for any device operating at the same frequency.



Wireless systems also are impacted by the environment and changes to the environment. Something as simple as the position of a person in a room could cause a wireless device to lose connectivity.

Special security concerns are associated with wireless systems because any device in range of a signal (i.e., the signal-to-noise ratio is high enough) can receive a transmission. While the transmission might be encoded, simply sending a message might leak information that was confidential. For example, a wireless silent alarm system that only generates messages when an alarm is activated can inform someone listening to the correct frequency if an alarm is triggered simply because a message was sent.

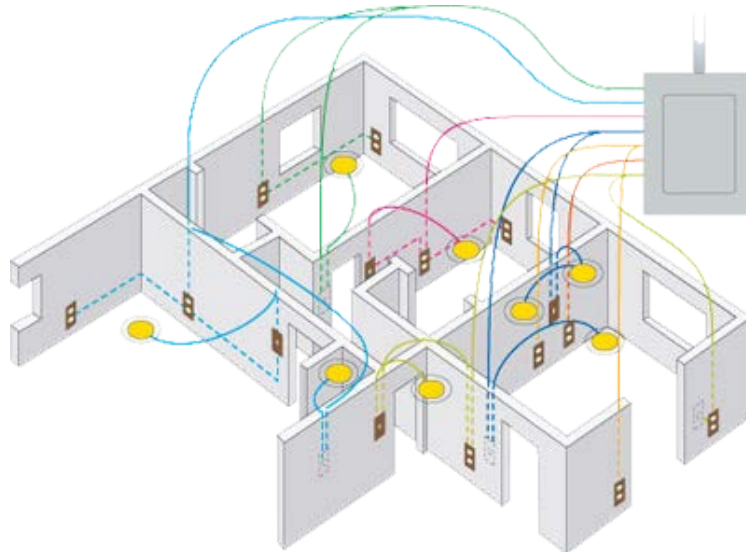
## 9.0 Communications and Networking: Wired

### 9.1 Introduction

When it comes to wired communications for sensors within commercial buildings, there are initially just two questions to be answered: 1) is it new construction (with wiring having to be installed), or 2) is it a structure with existing wiring that, perhaps, may be repurposed. While there are variants to these simplistic groupings, viewing the situation in this manner allows for a logical review of the relevant details associated with using a wired<sup>1</sup> backbone for the sensor communications network. The chapter organization begins with a brief review of the salient details associated with power line carrier communications (PLC, PLCC, also sometimes referred to as broadband over power lines, BPL), dedicated communications wiring (serial, Ethernet), and fiber optic communications. References are provided for each topic for those individuals who would like to delve more deeply into the technical details. The chapter concludes with a comparison of the covered technologies.

### 9.2 Power Line Carrier Communications

Buildings require an electrical power infrastructure to be designed and then installed. The typical process begins with an architect consulting with others and deciding where the building lighting and electrical power distribution plugs should be located. A diagram showing the plug and light locations, such as that shown in Figure 9.1, would be generated, and construction of the building would proceed, typically with an electrician installing the power distribution network wiring.



**Figure 9.1.** A Typical Building Lighting, Switch and Electrical Socket Layout

---

<sup>1</sup> In the context of this chapter, “wired” includes electrical (i.e., coax cable, CAT5/5e/etc Ethernet cabling, twisted pair, electrical power distribution wiring) and optical (i.e. multi-mode and single-mode optical fibers).

The realm of building wiring is well documented and covered by a number of regulations, building codes, and best practices<sup>2</sup>. Historically, national wiring standards have varied considerably. This situation is being addressed in European countries where an attempt has been made to harmonize national wiring standards via IEC Standard 60364 *Electrical Installations for Buildings*<sup>[7]</sup>.

Dual-use “power and communications” wiring was introduced over 100 years ago with power-distribution wiring beginning to be used as a communications medium. In 1919, the first system for voice-communication over power distribution wiring was placed into service. By the late 1920s, power line carrier (PLC) was used for pilot protective relaying with this application continuing ever since<sup>[8-11]</sup>. Over the decades, numerous vendors developed a wide array of proprietary communications methodologies. The IEEE established the Power Systems Communication Committee (PSCC) to address this situation resulting in a number of standards that are guiding the PLC situation<sup>[12]</sup>. As stated below, PSCC’s mission is purposefully broad and is not limited by operational constraints such as bandwidth/data rate nor is it bounded by “type” of physical media used (e.g., fiber, cable, wireless).

“The IEEE PES Power System Communications Committee (PSCC) develops communication standards, covering the treatment of all matters in which the dominant factors are the application, design, construction and operation of communications systems, including all matters necessary to the functioning of such communications systems employed in the generation, transmission, distribution, and utilization of electrical energy and their effects on system operation.”

PSCC is divided into a number of subcommittees, each with its own focus area. For the use of PLC for sensors in buildings, the Power Line Carrier Subcommittee (SC3), working with the IEEE Electromagnetic Compatibility Working Group (P956) has issued *P643 Guide for Power Line Carrier Applications*.

The P643 document clearly defines what is and is not within the guide.

“To study and document the relationships between power-line carrier (PLC) and low-frequency radio and radio navigation systems. PLC is defined as a communications medium used on the transmission lines of electric power utilities. Carrier on distribution lines or other conduction media is not included. The frequency range will include all frequencies up to and including 490 kHz; however, efforts will be made to avoid duplication of work by other groups dealing with specific segments of this band. The relationships studied will include interference mechanisms in both directions, i.e., the receiving characteristics and vulnerability of PLC systems to interfering signals, and the radiating characteristics of PLC systems with normal transmitted power levels.”

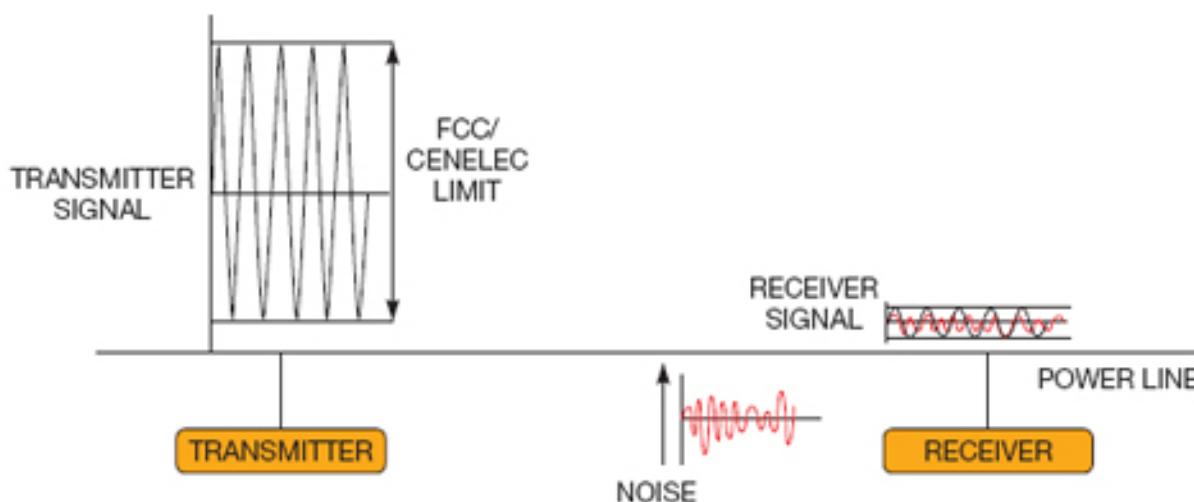
In general, PLC technology uses power lines as its transport medium. The data travels over the same power line that provides electricity, thus allowing the infrastructure in the building to also transport data, hence no additional communication infrastructure wiring.

Dual-use wiring presents a range of issues for the power system and the communication system to address. As with any communication system, there is a transmitter, a receiver, the transport medium, and

---

<sup>2</sup> Individuals interested in delving deeper into the details may find it beneficial to examine references [1-6].

a signal. In a generic PLC system, the transmitter modulates and injects the signal into the power line, and a receiver at the opposite end of the link extracts and demodulates the signal to retrieve the data (see Figure 9.2). The impedance of the power line attenuates the signal as it travels from the transmitter to the receiver. Any noise in the medium also corrupts the signal as it moves through the power line. The factors that affect the performance and reliability of a PLC system include the transmit-signal strength, noise on the power line, the impedance of the power line network, the protocol in use, and the receiver's sensitivity.



**Figure 9.2.** PLC Communication System

The performance of the communication system is dictated by the signal-to noise ratio (SNR). Given that, in PLC, the first priority of the transport medium is to distribute and deliver electrical power, it is not optimized for transporting communication signals. The physical wire and connected loads present a nonlinear transport medium with the amplitude and phase of the communication signal changing as demand for power changes (e.g., as appliances and other electrical equipment with their associated impedances cycle on and off<sup>3</sup>). The noise in such a transport medium may be quite high, which would lead the communication system developer to increase the signal strength of the transmitter until the best performance and power consumption over the power line is achieved. However, organizations such as the Federal Communications Commission (FCC) in the United States and the European Committee for Electrotechnical Standardization (CENELEC) in Europe tightly control transmit-to-signal strength. The FCC and CENELEC also regulate the harmonics that the main transmitter signal can inject into the power line. These regulations prevent signals on different frequency bands from corrupting one another. The line noise has two primary components—impulse noise and continuous noise. Both vary according to the devices (loads) connected to the electrical circuit and impact the overall SNR. The net result is a situation where the data communication transport medium properties vary in a combined predictable/unpredictable manner that leads to a requirement for a noise-robust media access control (MAC) protocol<sup>4</sup>.

<sup>3</sup> The situation is further complicated by the changing variety of devices and appliances that may be connected to the power distribution network. Reference [13] presents a plethora of measurements associated with impedance variations found in real-world networks.

<sup>4</sup> This situation is substantially different from PLC, PLCC, BPL of the past. Individuals interested in a more technical detailed description may wish to see reference [14-18].

In-building wiring on one side of a distribution transformer <sup>5</sup> has evolved to adopt IEEE Standard 1901. IEEE Standard 1901 defines the physical configuration for systems to use, and also the protocol that allows computer networks to send data over AC electrical power lines (in contrast to approaches such as power over Ethernet that sends DC power over computer network wires). As of Q3/2012, IEEE Standard 1901 has become the dominant power-line communication specification, <sup>[19]</sup> replacing a dozen previous power-line specifications <sup>[20]</sup> and including a mandatory coexistence Inter-System Protocol to deal with legacy PLC approaches <sup>[21]</sup>.

While ostensibly targeting the high bandwidth home information/entertainment market, IEEE Standard 1901 may be used for the transport of sensor information within commercial buildings. To that end, most devices that comply with IEEE Standard 1901 use AES128 encryption as required by National Institute of Technology (NIST)-led smart buildings initiatives with P1901.2 Working Group driving this application.

In addition, with the acknowledgement of IP addressable devices in the IEEE 802.x market, IEEE 1901 is working with IEEE P1905.1 to establish the guidelines for a 1-gigabit (1000-Mbps) product by 2013 <sup>[22, 23]</sup>. Networking products that fully comply with IEEE Standard 1901 will deliver data rates in excess of 500 Mbps in Local Area Network (LAN) applications and will achieve ranges of up to 1500 meters. The technology specified by IEEE Standard 1901 uses sophisticated modulation techniques to transmit data over standard AC power lines of any voltage at transmission frequencies of less than 100 MHz.

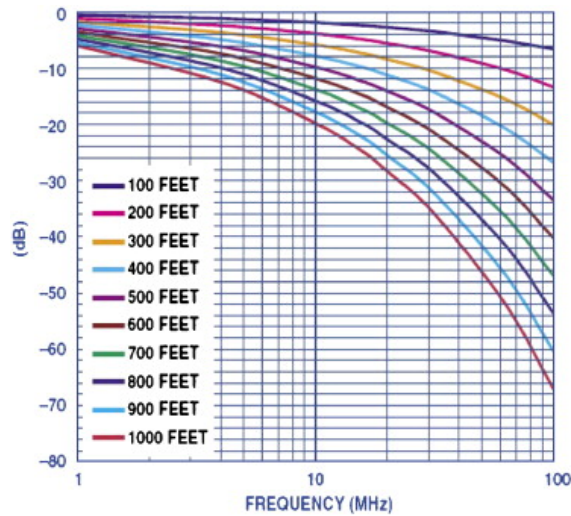
### 9.3 Dedicated Wiring for Building Sensor Communications

Given the nonlinearities and system vagueness associated with PLC and regardless of IEEE Standard 1901, the vast majority of buildings have dedicated wiring for communications signal transport (e.g., wiring or multiple wires bundled together into a cable). In such an instance, the communication system design presented as Figure 9.2 is still valid, although the transport (impedance) characteristics of the physical media and network are more predictable than in the PLC case. In a wire, electricity flows through a conductive material from one device to another. As it does so, it produces a slight magnetic field and some degree of electrical noise that radiates from the wire. The signal in the wire also is susceptible to electromagnetic fields present in its environment (i.e., interference). A situation in which other wires in the cable cause interference is referred to as crosstalk.

Attenuation associated with the wire causes the amplitude of the communication signal being transported along the wire to decrease, thereby limiting the effective length of the cable. A secondary signal amplitude (i.e., the attenuation effect) arises when the wire does not present a uniform attenuation level with respect to frequency. Both attenuation phenomena are illustrated in Figure 9.3. The data in the figure shows that 1) the higher the frequency of the signal the larger the attenuation that the signal “sees,” and 2) the longer the distance the higher the overall attenuation level that all frequencies are subjected to.

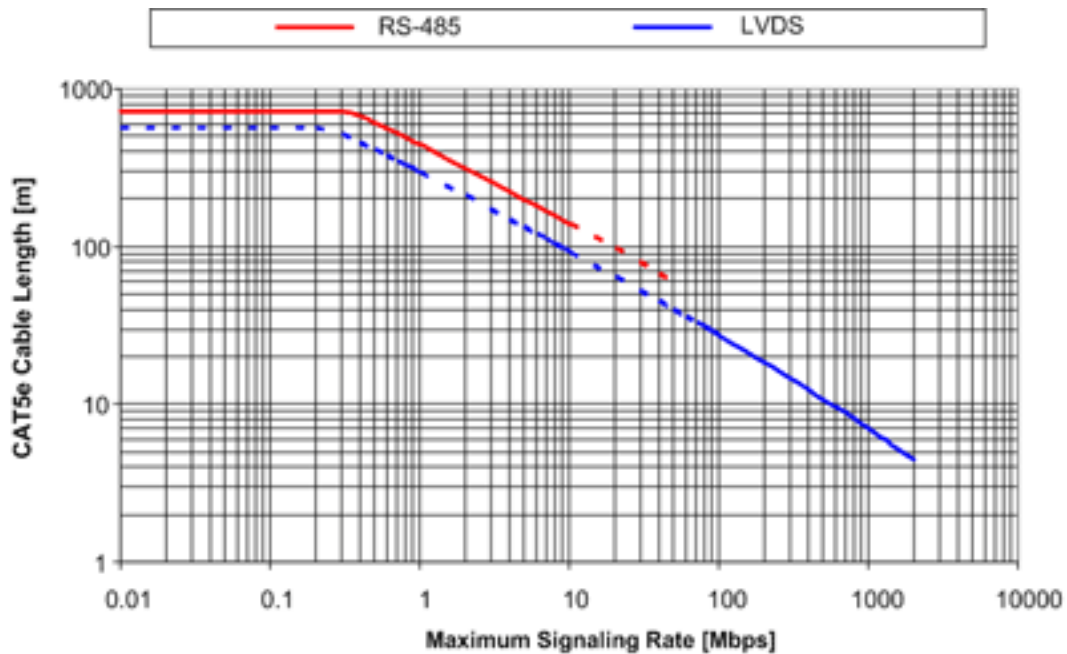
---

<sup>5</sup> Power distribution uses step-down transformers to reduce the voltage for use by customers. PLC communication signals cannot pass readily through transformers, as the transformer high inductance makes them act as low-pass filters, thereby blocking the communication high-frequency signals.



**Figure 9.3.** Signal Attenuation with Frequency

The net result of the data shown in Figure 9.3 is the longer the distance the higher the attenuation and the higher the frequency component the more the attenuation. Knowing that, as the data rate increases, the frequency components also increase leads to the situation shown in Figure 9.4 where the effective data rate versus cable length for a CAT5e cable are shown <sup>6</sup>.



**Figure 9.4.** Data Rate Versus Distance

Techniques have been developed to limit the attenuation suffered by cables. The most widely used technique is to twist a pair of wires around each other. This technique is known as twisted-pair, and nearly every copper networking cable uses it. Because of the susceptibility of networking cables to

<sup>6</sup> There are many signal modulation techniques that are used to minimize this issue in an attempt to extend the usable distance between transmitter and receiver.

accept interference, particularly from electrical power lines, whenever network cables need to be run through an exceptionally noisy area, such as a building riser that has both power and communications wiring (see Figure 9.5), a special type of network cable can be used. The normal twisted pairs are wrapped in a metallic foil that is grounded at both ends. This configuration dampens any external electrical noise that may have otherwise interfered with the signals on the network. These cables are referred to as shielded twisted pair (STP). It is used rarely, and only in environments that otherwise are unable to use unshielded twisted pair (UTP), as the more common network cable is known<sup>7</sup>.



**Figure 9.5.** A Building Riser with Coexisting Communication and Power Wiring

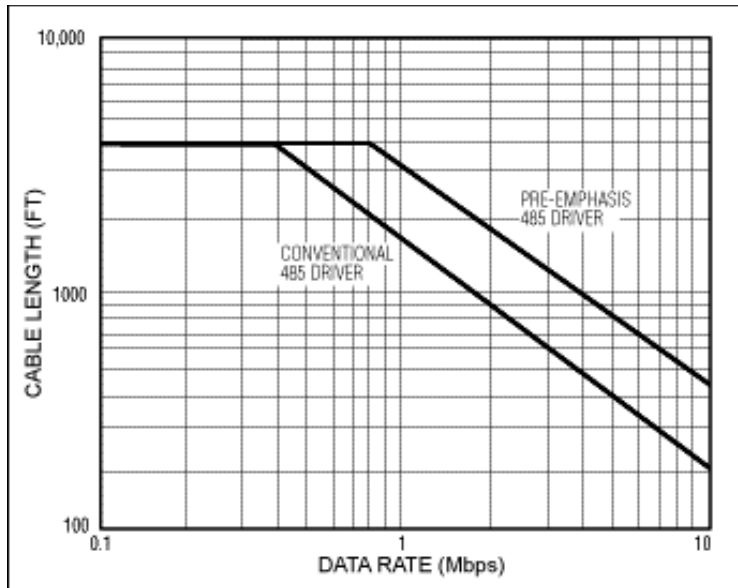
Most networking cable used in building applications is classified as either Category 5 (Cat-5) or Category 5 enhanced (Cat-5e), both of which are capable of transmitting up to gigabit speeds. Most new installations specify Category 6 (Cat-6) cable, which provides better insulation from cross talk and supports gigabit speed, but is designed to lower latency. Both Cat-5(e) and Cat-6 wire can be either UTP or STP, but UTP is far more common. Cat-5(e) and Cat-6 both have eight wires consisting of four twisted pairs of matching color, however the difference becomes apparent over distance. The twisting of the wires is what protects the data integrity as it traverses the cable. Cat-5e wire has a higher twist frequency than Cat-5 wire, and as such, the signal degradation is less intense over a longer length of cable. Table 9.1 presents the realistic usable data rate for a fixed length of each of the cable types discussed above.

**Table 9.1.** Data Rates for Various Cables (Fixed Length)

Type	Data Rate (Mbps)	Distance (m)
Category 5	100	100
Category 5(e)	1,000	100
Category 6	3,000	100
Category 6(a)	10,000	100

Serial communications using RS-485 signaling may also use Cat-5/Cat-6 cables. Data rate as a function of cable length for RS-485 is presented in Figure 9.6.

<sup>7</sup> The reason STP is not used as frequently as UTP is the additional expense of the metallic shielding, the extra work of grounding the shielding at each end, and because if installed improperly, the entire length of shielding can serve as an antenna, effectively doing more harm than good.



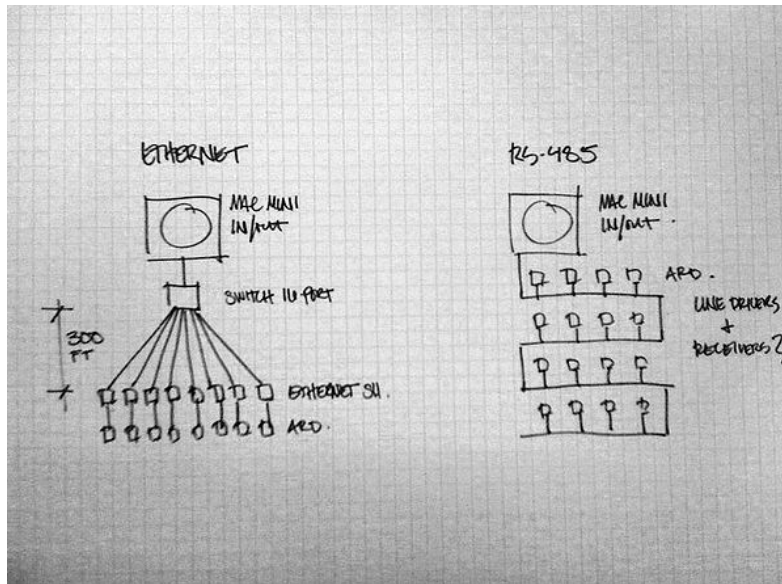
**Figure 9.6.** RS-485 Data Rate Versus Cable Length

## 9.4 Network Architectures for Wired Building Sensors

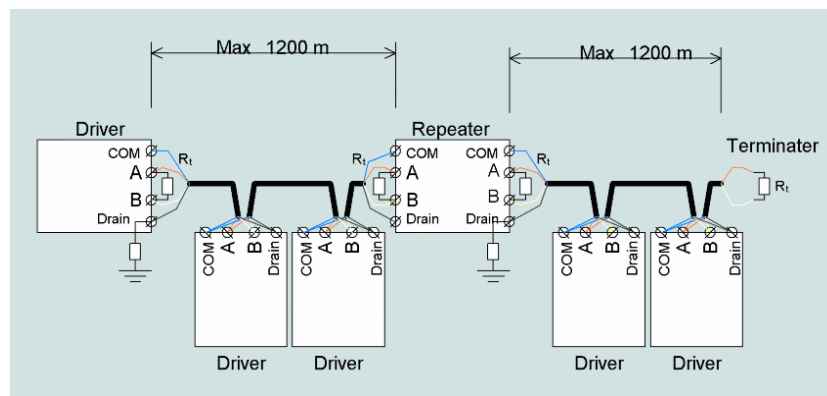
The connection of a building sensor system with a building management system (i.e., software) relies on the wiring of the sensors. In the RS-485 signaling scheme shown in Figure 9.6, which is a scheme used widely, a variety of bus architectures may be used, with the most prevalent being those shown in Figure 9.7<sup>8</sup>.

<sup>8</sup> See reference [24] for a realistic description of RS485.





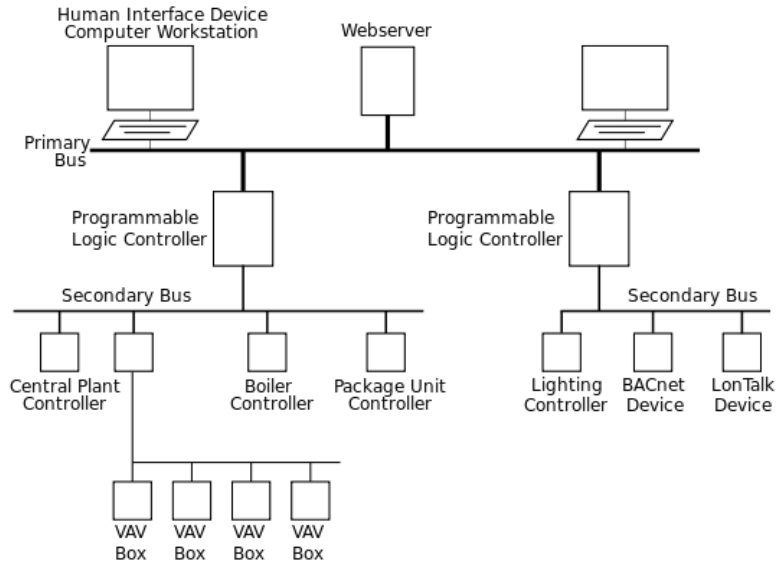
(a)



(b)

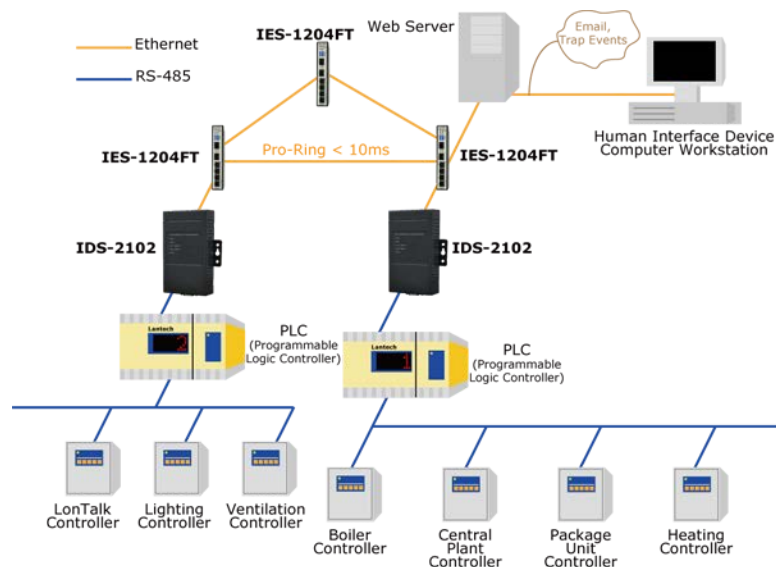
**Figure 9.7.** Frequently used RS-485 Bus Architectures. (a) Illustration of the linear bus versus the tree structure used in ethernet networking. (b) The actual realization of an RS-485 daisy chain network.

As sensors have progressed from discrete measurement devices into intelligent edge devices, the traditional star (a.k.a., “home run”) wiring architecture used in building automation and management systems has changed. The incorporation of a building automation system (or a building management system) has essentially meant that the network architecture has migrated from the aforementioned star topology into an ethernet-centric IP-addressed-based topology. The situation is illustrated in Figure 9.8 where a typical building automation network is shown.



**Figure 9.8.** Building Automation Network Riser Diagram

From a data rate requirement perspective, signaling protocols such as ModBus/RS-485 operate in the 10- to 100-s kbps realm <sup>9</sup>. Traditional wiring is used to satisfy the bandwidth/data-rate requirements for such a network. As the complexity of the building automation system components increases, there has been a major shift towards using ethernet to transport data (a vendor proprietary protocol may still be in use). In either instance, the communication network needs to support ethernet data rates. The network architecture becomes an IT-centric design such as that shown in Figure 9.9 where RS-485 and ethernet data transport is integrated.



**Figure 9.9.** A Classic IT-Centric Network Design for Building Automation Systems

<sup>9</sup> In the case of a programmable logic controller (PLC) based system, the sensors are (traditionally) wired directly to the PLC input/output. The entire system may be analog based, resulting in the data rate concept not being applicable. The physical plant choice of wire is dictated by the cable lengths (line loss) and the sensor sampling rate, again, typically in the 1 to 10 to 100 times per second category.

From a “sensors for buildings” wiring perspective, a combination of PLC and dedicated wiring may be used to achieve the architecture shown in Figure 9.9. In the case where the sensors themselves may still operate at a relatively low sampling rate with their readings “concentrated” into an ethernet-capable device (frequently referred to as a gateway) or the sensor itself may support ethernet directly, a total IT-centric ethernet network may be used.

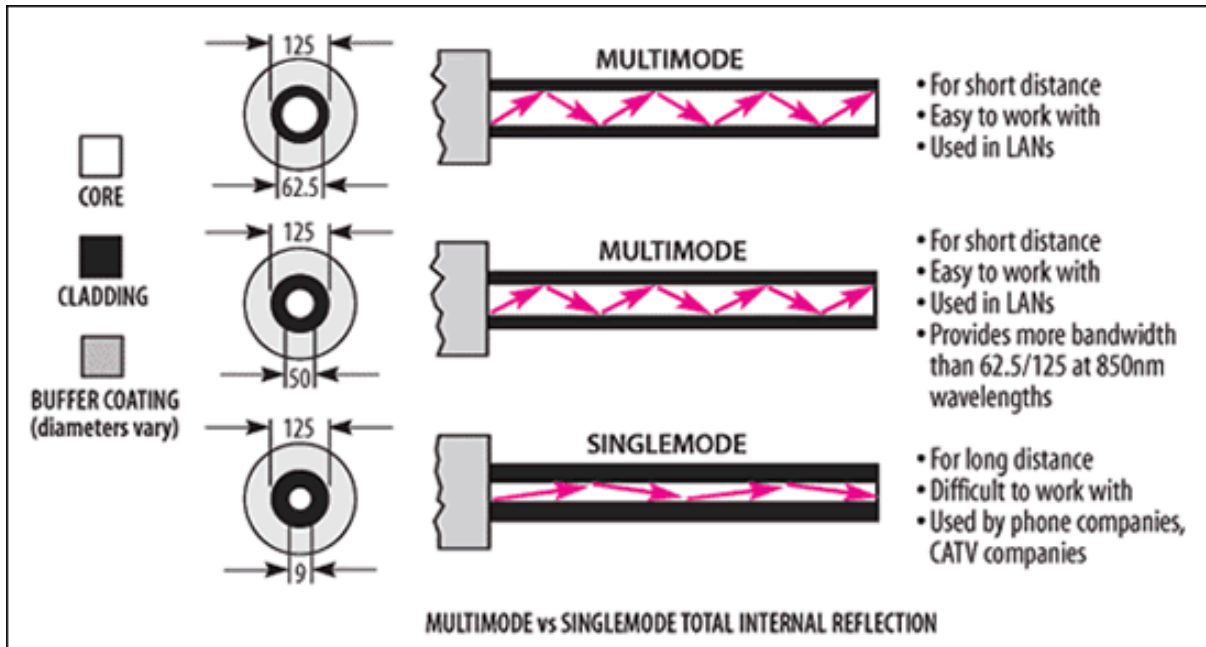
## 9.5 Fiber Optic Communications for Wired Building Sensors

As was seen in the prior sections, (traditional) electrical wire (cable) has a relatively high attenuation factor and is not well suited for long-distance transport of high-data-rate signals. The maximum usable length of a UTP cable is 100 meters. Fiber optic cable is used for runs exceeding this length. The main benefits of fiber are its exceptionally low loss (allowing long distances between amplifiers/repeaters), its absence of ground currents and other parasitic signal and power issues common to long parallel electric conductor runs (due to its reliance on light rather than electricity for transmission, and the dielectric nature of fiber optic), and its inherently high data-carrying capacity. Thousands of electrical links would be required to replace a single high bandwidth fiber cable. In comparison with electrical cable, optical fibers may be run alongside each other for long distances in regions of high electromagnetic interference with the data signals effectively experiencing no crosstalk.

Optical fiber may be categorized as either multi-mode or single-mode, with the differences being the number of spatial modes that the fiber supports<sup>10</sup>. Figure 9.10 shows the most salient details in the differentiation of single-mode (single spatial mode, only one path through the light pipe) and multi-mode (multiple paths through the light pipe with the differing paths having slightly different optical path lengths, thereby taking slightly different times resulting in the pulse being dispersed [spread out in time]).

---

<sup>10</sup> Individuals interested in more technical details associated with optical fiber communications may wish to read references [25-29].



**Figure 9.10.** Single-Mode and Multi-Mode Optical Fibers

Figure 9.10 also illustrates that there are different types of multi-mode optical fiber. In this case, the top multi-mode fiber has a core diameter of 62 micrometers, while the middle multi-mode fiber has a core diameter of 50 micrometers. From a signal dispersion perspective, the larger the core, the more spatial modes that are resident within the fiber core resulting in larger dispersion and lower effective data rate. The tradeoff is that a larger core presents a physically larger target for the photons to enter into the core with a substantially reduced tolerance on all components (i.e., inexpensive).

While most optical fiber is made of glass, there is a subset of optical fiber that is made of polymer (plastic). Such polymer optical fiber (POF) is very inexpensive to manufacture and relies on the POF-air interface for total internal reflection (and signal propagation down the core)<sup>11</sup>. POF fibers are inherently “lossy,” physically large, very durable, and suitable for use in situations where a moderate data rate is required over not too long of a distance. Table 9.2 presents data rate versus usable link distance for a number of multi-mode fibers (POF = polymer optical fiber, HCS = hard clad silica [glass], MBd = million symbols per second). This is different than the bit rate for one symbol may represent multiple bits].

**Table 9.2.** Symbol Rate Versus Length of Fiber (distance) for Various Optical Fibers

Data Rate	Link Distance
DC – 1MBd	45m (POF)
DC – 5MBd	20m (POF)
DC – 10MBd	40m (POF)
	200m (HCS)
20MBd	2700m (MM)
125MBd	50m (POF)
	100m (HCS)
125 MBd	2000m (MM)

<sup>11</sup> There is no core/cladding/jacket interface as shown in Figure 9.10, simply polymer and air.

A number of building codes provide details for the use of optical fiber in building communication systems. Table 9.3 presents distance versus type of optical fiber information for two classes of Gigabit Ethernet: 1) 1000Base-SX (1000 Mbps using 770 to 860 nm, near infrared light) and 2) 1000Base-LX (1000 Mbps using 1270 to 1355nm light)<sup>12</sup>. The TIA568 building wiring standard specifies 160/500 Mhz\*km multimode fiber (first entry in 1000Base-SX), while the ISO/IEC 11801 building wiring standard specifies 200\*500 MHz\*km multimode fiber (second entry for 62.5 micrometer fiber). ISO/IEC 11801 has proposed adoption of 500/500 MHz\*km 50 micrometer multimode fiber within its specifications.

**Table 9.3.** Gigabit Ethernet Distances for Optical Fiber

Standard	Fiber Type	Diameter (microns)	Model Bandwidth (MHz* km)	Minimum Range Meters
1000BASE-SX	MM	62.5	160	2 to 220*
	MM	62.5	200	2 to 275**
	MM	50	400	2 to 500
	MM	50	500	2 to 550***
1000BASE-LX	MM	62.5	500	2 to 550
	MM	50	400	2 to 550
	MM	50	500	2 to 550
	SM	9	N/A	2 to 5000

The choice between optical fiber and electrical (or copper) transmission for a particular system is made based on a number of tradeoffs. Fiber has a higher bandwidth than copper. For example, Cat-6A cable is classified by the Telecommunications Industry Association (TIA) to handle a bandwidth up to 600 MHz over 100 meters, which theoretically, could carry around 18,000 calls at the same time. Multi-mode fiber, on the other hand, would have a bandwidth of over 1000 MHz, which could carry almost 31,000 simultaneous calls. Fiber is a dielectric, which means that no electrical current flows through it. Copper, on the other hand does carry a current and could cause a fire concern if it is old or worn.

Regardless of the choice of transport medium, there will be attenuation. The maximum allowed industry standard of attenuation for multi-mode fiber over 100 meters is 0.15 dB. This means that the fiber loses 3 percent of its original signal strength over 100 meters. Also, the attenuation of fiber does not change as bandwidth increases or decreases. The maximum allowed industry standard of attenuation for Cat-6A cable over 100 meters at 100 MHz is 20.9 dB, which is a 94 percent loss in signal strength.

On a practical installation perspective, copper cable is relatively delicate. It has a 25 pound tension limit. Basic fiber has a 100 to 200 pound tension limit. In addition, optical fibers can be installed in buildings with the same equipment used to install copper and coaxial cables, with some modifications due to the small size and limited pull tension and bend radius of optical cables. Optical cables can typically be installed in duct systems in spans of 6000 meters or more depending on the condition, layout, and installation technique. Longer cables can be coiled at an intermediate point and pulled farther into the duct system as necessary.

<sup>12</sup> In each case, these are glass optical fibers, not POF. Such silica optical fibers have a significantly lower attenuation coefficient at the longer (nominal) 1500 nm wavelength than the (nominal) 800 nm wavelength.

Copper cable, Cat-5(etc) provides Power Over Ethernet (POE)<sup>13</sup> capabilities, thereby providing the ability to power a range of devices, including telephones, surveillance cameras, Wireless Access Points (WAPs), and many other devices through the networking cable itself. The savings in not requiring a separate power line for such devices may prove substantial.

Cat-5 cabling is less expensive than optical fiber cabling (and the associated costs for the optical versus electrical network transceivers used in all connected equipment).

## 9.6 Summary/Conclusions

A wired backbone for sensors in a building may use existing power distribution wiring (PLC) or dedicated wiring (electrical, optical). The data-rate requirements are the dominant factor in deciding which to use. It is prudent to identify the appropriate network architecture in determining what transport media is best for which need with an understanding that it may be optimal to not use uniform wiring for all parts of the network (e.g., use PLC to support one network element, Cat-5 for another, and optical fiber for yet another network element). With the primary elements within a building automation network being IP addressable, it is highly recommended that the individual(s) responsible for maintaining the building IT network be involved in the architecture and deployment discussions.



**Figure 9.11.** A Wiring Closet in a Commercial Building that Includes the Telephone Exchange Wiring (left hand side), Coaxial-Cable Based Broadband Satellite TV Distribution Wiring (further along left hand side), a Fiber Optic Video Distribution Network (far wall), and a 96 Port Ethernet Distribution Network (right hand side). Each communication network has independent wiring and shares the cable trays and building risers.

---

<sup>13</sup> Please see references [30-32] for a description of Power Over Ethernet (POE).

## 9.7 References

1. H.C. Ferreira and O. Hooijen, "Power Line Communications: An Overview", The Transactions of The S.A. Electrical Engineers, pp145-161, September, 1995
2. J.R. Nicholson and J.A. Malack, "RF Impedance of Power Lines and Line Impedance Stabilization Networks in Conducted Interference Measurements", IEEE Transactions on Electromagnetic Compatibility, Vol. EMC-15, no. 2, pp84-86, May 1973.
3. J.E. Vander Mey and T.J. Vander Mey, "United States Patent No. 5,090,024", Feb. 18,1992
4. J.E. Vander Mey and T.J. Vander Mey, "United States Patent No. 5,278,862", Jan 11,1994
5. R.M.Vines, H.J. Trussel, L.J. Gale and J.B. O'Neal, "Noise on Residential Power Distribution Circuits", IEEE Transactions on Electromagnetic Compatibility, Vol. EMC-26, No. 4, pp161-168, November, 1984.
6. R.M. Vines, H.J. Trussel, K.C. Shuey and J.B. O'Neal, "Impedance of the Residential Power-Distribution Circuit", IEEE Transactions on Electromagnetic Compatibility, Vol. EMC-27, No. 1, pp6-12, February 1985.
7. [http://en.wikipedia.org/wiki/IEC\\_60364](http://en.wikipedia.org/wiki/IEC_60364)
8. Boddie, C. A. and Curtis, R. C. "The Transmission of High-Frequency Currents for Communication over Existing Power Networks", *AIEE Transactions*, Volume 48, pp. 227-235, Disc. 236, 1929
9. Boddie, C. A. "Largest System of Power Line Telephony", *Electrical World*, Volume 83, March 13, 1926
10. Boddie, C. A. and Coke, M. W. "Limitations of Carrier-Current Telephony", *Electrical World*, Volume 81, pp. 909-913, April 21, 1923
11. Fiedler, H.J. "PLC Application Seminar" GE Protection & Control. GEZ-8199.
12. <http://www.ewh.ieee.org/soc/pes/psec/index.html>
13. Phil Sutterlin and Walter Downey, "A Power Line Communication Tutorial - Challenges and Technologies", [http://www.isplc.org/docsearch/Proceedings/1998/pdf/0450\\_001.pdf](http://www.isplc.org/docsearch/Proceedings/1998/pdf/0450_001.pdf)
14. M. Tanaka, "Transmission Characteristics of a Power Line Used for Data Communications at High Frequencies", IEEE Transactions on Consumer Electronics, Vol. 35, No. 1, pp37-42, February 1989.
15. K.M. Dostert, "Frequency Hopping Spread-Spectrum Modulation for Digital Communications Over Electrical Power Lines", IEEE Journal on Selected Areas in Communications, Vol 8, No. 4, pp700-710, May 1990.
16. J.B. O'Neal, "The Residential Power Circuit as Communication Medium", IEEE Transactions on Consumer Electronics, Vol. CE-32, No. 3, pp567-577, August 1986.
17. J. Drnovšek, J. Bojkovski, I. Pušnik and T. Repar, "A Measurement Communication System", IEEE Instrumentation and Measurement Technology Conference Proceedings, pp200-202, June 4, 1996.
18. M.H. Shwehdi and A.Z. Khan, "A Power Line Data Communication Interface Using Spread Spectrum Technology in Home Automation", IEEE Transactions on Power Delivery, Vol 11, No. 3, pp1232-1237, July, 1996

19. Chano Gomez (19 May 2008). "How To Kill The Home Networking Industry". *Electronic Design News*.
20. Mike Wilson (30 October 2008). "Setting The Standard In Wireline Home Networking". *Electronic Design News*.
21. Stefano Galli, O. Logvinov (July 2008). "Recent Developments in the Standardization of Power Line Communications within the IEEE". *IEEE Communications Magazine* **46** (7): 64–71.
22. S. Goldfisher, S. Tanabe, "IEEE 1901 access system: An overview of its uniqueness and motivation", *IEEE Commun. Mag.*, vol. 48, no. 10, October 2010, pp. 150–157.
23. <http://grouper.ieee.org/groups/1901>
24. Yashkardin Vladimir, "RS-485. Recommended Standard 485. Electrical Characteristics of Generators and Receivers for Use in Balanced Multipoint Systems", [http://www.softelectro.ru/rs485\\_en.html](http://www.softelectro.ru/rs485_en.html)
25. Mary Bellis, "The Birth of Fiber Optics", <http://inventors.about.com/library/weekly/aa980407.htm>
26. John Tyndall, "Notes of a Course of Nine Lectures on Light", 1870. <http://archive.org/stream/notesofcourseofn00tyndrich#page/n5/mode/2up>
27. Peter Fuhr, "Free Space Optical Communications Using Laser Diode Arrays", Ph.D. Dissertation, The Johns Hopkins University, 1985.
28. Encyclopedia of Laser Physics and Technology, [http://www.rp-photonics.com/optical\\_fiber\\_communications.html](http://www.rp-photonics.com/optical_fiber_communications.html)
29. H. J. R. Dutton, "Understanding Optical Communications", <http://www.redbooks.ibm.com/pubs/pdfs/redbooks/sg245230.pdf>, IBM Redbooks
30. Understanding Power Over Ethernet, <http://www.connect802.com/ethernet.htm>
31. <http://www.ieee802.org/3/af/>
32. Power Over Ethernet, POE, 802.3af /802.3at, [http://www.radio-electronics.com/info/telecommunications\\_networks/ethernet/power-over-ethernet-poe-ieee-802-3af-3at.php](http://www.radio-electronics.com/info/telecommunications_networks/ethernet/power-over-ethernet-poe-ieee-802-3af-3at.php)



# 10.0 Communications and Networking: Wireless

## 10.1 Introduction

In this chapter, we identify and discuss various wireless technologies and determine the optimal topology (if possible) and associated performance of the network and its subsystems. There are a number of wireless devices and systems that are being considered or are being deployed in building automation settings. This chapter attempts to provide guidance on how individual networks operate and, perhaps even more importantly, how various networks interoperate.

Our review begins with an examination of various radio technologies, the frequencies that are “available” for the license-free band, and dense RF deployments. Next, we present networking fundamentals and topologies associated with specific wireless systems. Finally, we conclude the chapter with an examination of facility networks from an IT perspective.

## 10.2 Wireless for Building Automation

It is easy to understand the marketplace confusion with all the various radios being deployed from personal two-way communications (e.g., walkie-talkies) to supervisory control and data acquisition (SCADA) systems to wireless sensor networks to backhaul communications systems to mention just a few. Organizing a list of wireless technologies available for use in the building automation sector is complicated by the multitude of applications that end users request. An evaluation of the deployment and value of wireless building automation systems is based on two broad application classes—those requiring mobility and those derived from the reduced cost resulting from not having to install a wired network.

Reducing the cost of installing a wired network allows the user to increase the points of measurement within a building sensor system. The process engineer can get more measurement points for every dollar invested. Incremental process measurements improve process efficiency and optimization, saving resources, energy and increasing throughput. Added condition-monitoring measurements can dramatically increase maintenance efficiency, reducing cost and preventing lost time due to asset failures. The building automation facility of the future will be based on having a complete understanding of what is happening within the facility. Incremental sensors are the foundation for collaborative applications and advanced process/business management.

Mobility versus fixed measurements has very different requirements with power management being the most obvious. Local power is often available at the location of a measurement point, while battery power is required for mobility. The following discussion will help establish context for the various technologies and approaches available for wireless building automation.

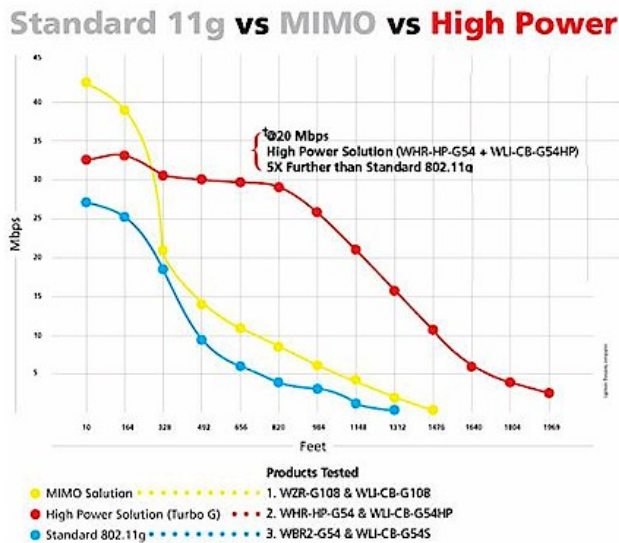
The discussions may be distilled in other ways, with the most logical being based on the RF footprint of the radio transmission and the frequency of operation. Even the definition of such an RF footprint is problematic because, in reality, the extent of the footprint is the distance from the radio transceiver where the companion transceiver operates in an “acceptable” manner. Consider the case of 802.11, which commonly is referred to as “Wi-Fi.” In this case, the transceiver operates with an output power that is compliant with 802.11 standards and local regulations (e.g., power limits) and a data-rate dependent circuitry defined receiver sensitivity. These two factors combine in a manner that, given a specific bit

error rate, defines the maximum separation distance between the two transceivers, thus defining the RF footprint. In an ideal case, the RF footprint decreases as  $1/r^2$ , where  $r$  is the distance from the RF emitter. A gradient representation of this signal variation is shown in Figure 10.1.



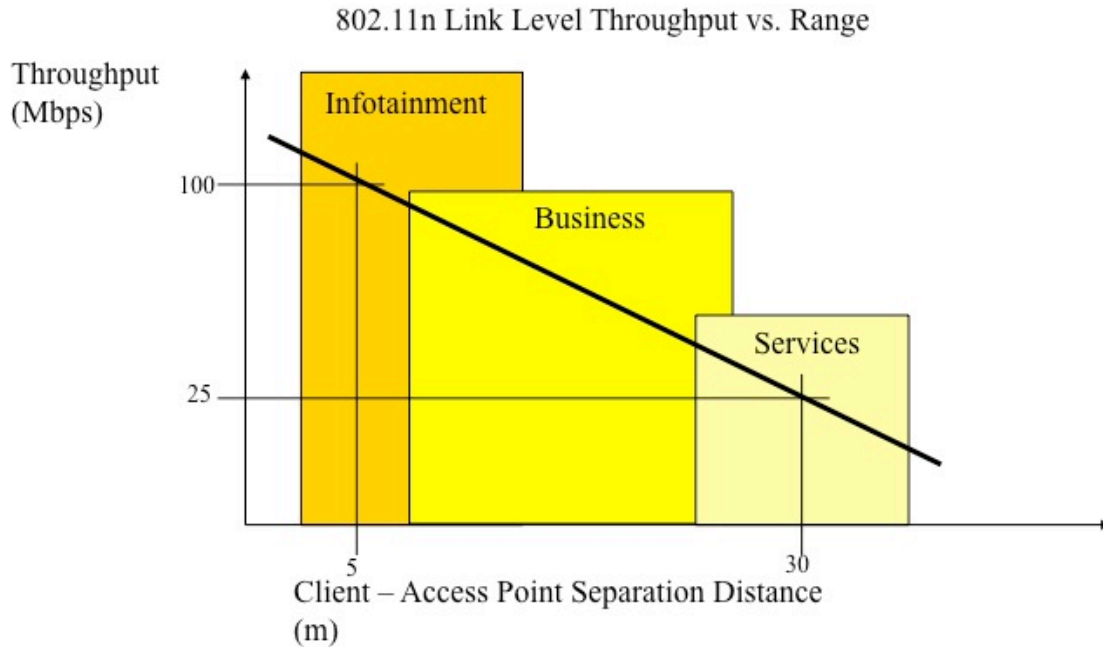
**Figure 10.1.** The Signal Level from a Transmitter Decreases in a Nonlinear Manner with Distance.  $E_1$  represents the transmitter’s output field.

The data rate (Mbps), bit error rate (BER), SNR, and distance served by a communication channel are intertwined and, in practice, the perfect situation depicted in Figure 10.1 does not exist. A more realistic case in which the data rate decreases as the transceiver separation distance increases is shown in Figure 10.2.



**Figure 10.2.** Data Rate as a Function of Separation Distance for an 802.11g Access Point and Client

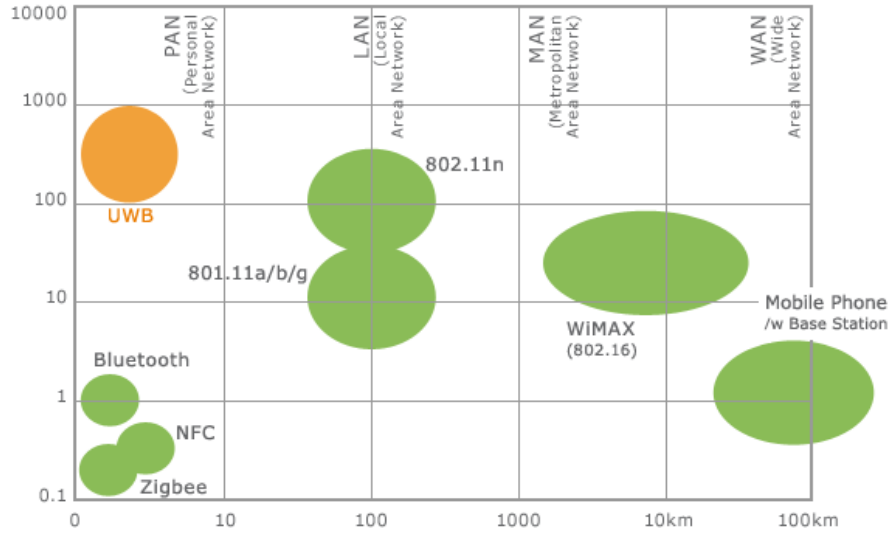
Figure 10.2 shows that, for fixed error rate, the data rate decreases as the separation distance increases. In certain high bandwidth application areas, such as wireless display communications, the separation distance should be as small as possible. Therefore, if the application requires 100 Mbps of throughput, the distance between the transceivers should be less than approximately 5 m. As the throughput required by the application decreases, the separation distance can be increased. The situation is shown Figure 10.3 for an 802.11n channel.



**Figure 10.3.** Throughput of Wireless Systems Changes as a Function of Distance (for fixed bit error rate)

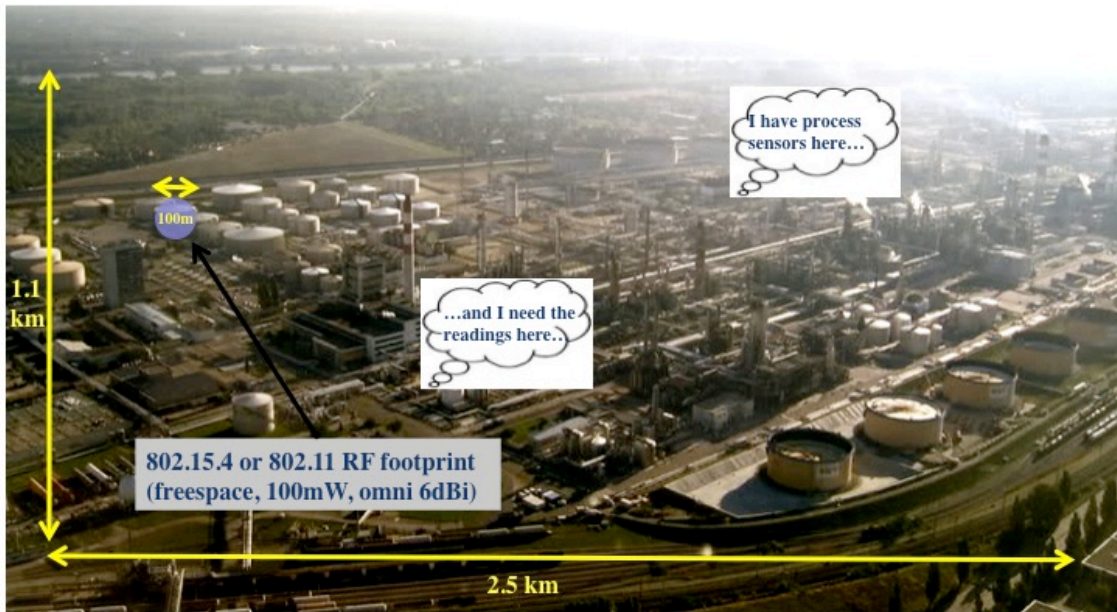
Note that Figure 10.3 is for two transceivers communicating in a point-to-point link manner. Even in this situation, at any moment, the system behaves as a point-to-point link (even in a point-multi-point topology, once the base station/access point has “given” the channel to the client, the communications channel is at that moment a point-to-point system). While the data throughput/distance relationship dictates the client-access point performance, the network design may provide slightly overlapping RF footprints, such as in cellular network design, leading to a network “footprint” that may be considerably larger than a situation in which a single radio is employed.

Given mobile client handoffs and networks of base stations/access points, the separation distance idea gets blurred (e.g., a mobile phone may be thought to work throughout the world). The operating range for a number of wireless technologies, some relying on point-to-point operation and some relying on network operation, is shown in Figure 10.4. The vertical axis depicts data throughput.



**Figure 10.4.** Separation Distance, or “Footprint,” Versus Data Throughput for a Wireless Technologies

The building automation application areas involving wireless technologies that are most frequently discussed are 1) wireless sensor networks and 2) backhaul networking (data transport across the facility). A logarithmic figure such as depicted in Figure 10.4 is augmented by the actual situation of the RF footprint and the plant size. The situation is shown in Figure 10.5.



**Figure 10.5.** The 100 m RF Footprint Associated with 802.15.4 and 802.11 Transceivers is Overlaid onto a Photograph of an Industrial Site

While it is possible to deploy a mesh array of 802.15.4-based wireless sensors to cover the facility and distribute the readings throughout the plant, user reports have shown that a more cost effective situation may be to use a backhaul transport network for longer distance communications.

## 10.3 Wireless Technologies

A number of wireless technologies are listed in Figure 10.4. Wireless technologies most commonly deployed in building automation settings are presented as Table 10.1. The table lists, where applicable, the standard, its common name, the frequency range, if the system/elements use unlicensed (ISM) frequencies and the typical application of the technology.

**Table 10.1** List of Building Automation Radio Technology Standards and Their Candidate Applications

Number	“Common” Name	Operational Frequency	Unlicensed (Yes/No)	Typical Application
802.11 a-z	Wi-Fi	2.4 GHz, 5.7 GHz	Yes	Wireless LAN
802.15.1	Bluetooth	2.4 GHz	Yes	Wireless PAN
802.15.3	WiMedia	~5 GHz	*	High data rate, short distance
802.15.4	ZigBee™/ISA100.11 a/WiHART	2.4 GHz	Yes	Low rate building automation sensors
802.15.4a	“chirped”	2.4 GHz	Yes	Low rate sensors and position
Sat Comm	Satellite Communications	Ku, K, Ka bands (12-40 GHz)	No	Broadband, data transport
802.16	WiMAX (WiBro)	2-11 GHz, 10-60 GHz	No	Broadband wireless
802.20	MBWA	<3.5 GHz	No	IP-based data transport
1451	Sensors	900 MHz, 2.4 GHz	Yes	Sensor transport using 802.15.4 and 802.11
1901	RuBee	135 kHz	Yes	location
Wi-Di	Wireless Display	5.7 GHz	Yes	HD displays using 802.11n
RF SCADA	Wireless SCADA	<1 GHz	No/Yes	SCADA transport
FRS/GMRS PMR446(Europe)	Walkie-Talkies	27, 49, 462-467 MHz, 446 MHz (E)	Yes/No	Personal communication
IS95/IS136/others (cellular)	CDMA/TDMA/etc	See Chart 1	No	Telephony
3GPP TS 45.005	GSM	See Chart 2	No	Telephony
ISO 18000-7	DASH7	433 MHz	Yes	Wireless sensors, RFID
ISO/IEC 14443	Near Field Communications	13.56 MHz	Yes	Short distance (10cm) data transfer
UWB	Wireless USB	3.1 – 10.6 GHz	Yes*	High data rate, <10m
Wireless HD		60 GHz	Yes	High def transmission
WHDI	Wireless Home Display Interface	5.7 GHz	Yes	Up to 3 Gbps, short distance

## 10.4 Distance between Transceivers

The question—*How far apart can the wireless devices be?*—is a typical and logical question. Many technical details must be considered in the answer, including bit error rate, data rate, attenuation, receiver sensitivity, transmitter output power, and receive/transmit antenna gains.

*Step 1:* An example of how much path loss (attenuation) is associated with distance is provided in Table 10.2. Consider the case of a 2.4 GHz 100 mW (+20 dBm) transmitter with a 9 dBi omnidirectional antenna yielding an Effective Isotropic Radiated Power (EIRP) of +29 dBm. With the receiver antenna gain of 2 dBi, the incident received power at the receiver can be calculated as follows (given a path loss exponent of 2 [i.e.,  $1/r^2$ ]):

**Table 10.2.** Attenuation and Received Power Versus Range

Range	1 m	10 m	100 m	1000 m	4000 m
Attenuation	40 dB	60 dB	80 dB	100 dB	112 dB
Received power	-9 dBm	-29 dBm	-49 dBm	-69 dBm	-81 dBm

*Step 2:* Receiver sensitivity is data-rate dependent. Consider the situation where an 802.11b access point is used. Assuming that our transmitter was a typical 802.11b access point, the receiver sensitivities and associated data rates are shown in Table 10.3. Coupled with the information from Table 10.2 and incorporating a link margin of +20 dB, the transmitter-to-receiver separation distance is shown in the third column of Table 10.3.

**Table 10.3.** Transmitter-to-Receiver Separation Distance as a Function of Data Rate (for an 802.11b access point)

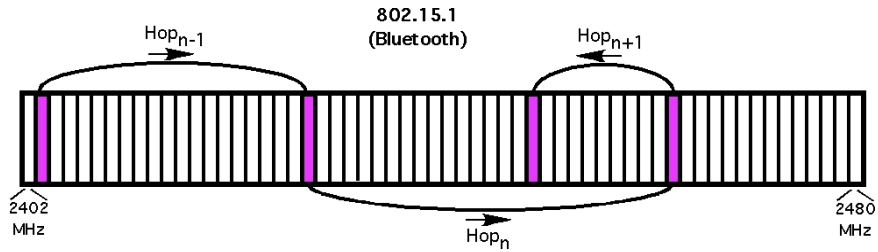
	Receiver sensitivity	Nominal separation distance (w/20 dB link margin)
11 Mbit/s mode (CCK)	-80 dBm	350 m
5.5 Mbit/s mode (CCK)	-83 dBm	500 m
2 Mbit/s mode (QPSK)	-92 dBm	1,420 m
1 Mbit/s mode (BPSK)	-95 dBm	2,000 m

## 10.5 Building Automation-Strength Wireless Sensor Networks

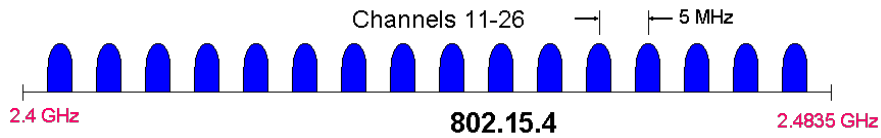
In standards-compliant wireless operation, most devices have gravitated to using either an IEEE 802.15.4-compliant wireless channel or an IEEE 802.11b/g-compliant channel. Please note that not all of the exhibited devices operate under IEEE-compliance channels, rather they could be running their own protocol and be broadcasting in the ISM bands. The result is easy to predict, namely numerous sensors/instruments/transmitters all attempting to operate in the same 2400 MHz channels resulting in considerable congestion and coexistence issues. The principal ISM radio transceivers encountered in building automation settings are based on the 802.15.1, 802.15.4, and 802.11 standards.

The 2400 MHz channel assignments for 802.15.1 are shown in Figure 10.6. Figure 10.7 shows the channel assignments for 802.15.4 while the 2400 MHz frequencies associated with 802.11 are shown in Figure 10.8.<sup>1</sup>

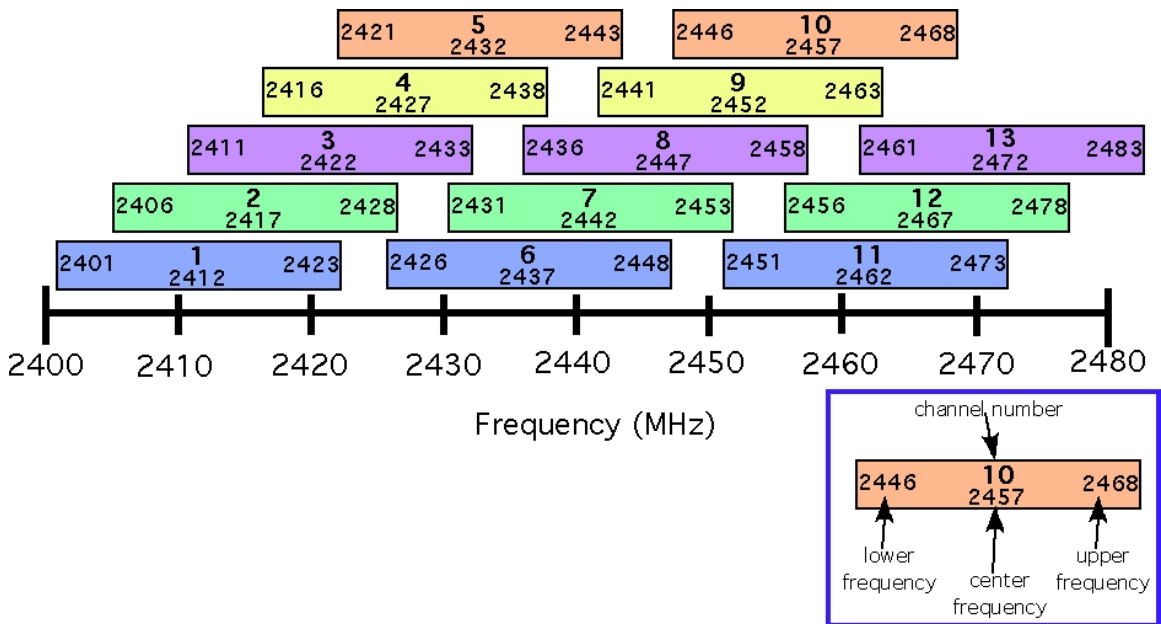
<sup>1</sup> There is seemingly a never-ending change to radio transceiver technologies and the frequencies at which they operate. While the current trend in building automation wireless field transmitters is to use the 802.15.4-2006 version of the radio specification (i.e., operating at 2400 MHz), the IEEE802.15.4c study group is considering the newly opened 314-316 MHz, 430-434 MHz, and 779-787 MHz bands in China, while the IEEE 802.15 Task Group 4d is defining an amendment to the existing standard 802.15.4-2006 to support the new 950 MHz to 956 MHz band in Japan. IEEE 802.15.4a was released expanding the four PHYs available in the earlier 2006 version to six, including one PHY using Direct Sequence Ultra-wideband (UWB) and another using Chirp Spread Spectrum (CSS). The UWB PHY is allocated frequencies in three ranges: below 1 GHz, between 3 and 5 GHz, and between 6 and 10 GHz. The CSS PHY is allocated spectrum in the 2450 MHz ISM band. In April 2009, IEEE 802.15.4c and IEEE 802.15.4d were released expanding the available PHYs with several additional PHYs: one for 780 MHz band and another for 950 MHz using various modulation formats.



**Figure 10.6.** The 802.15.1 (Bluetooth) Frequency Channels in the 2450 MHz Range



**Figure 10.7.** The 802.15.4 Frequency Channels in the 2450 MHz Range

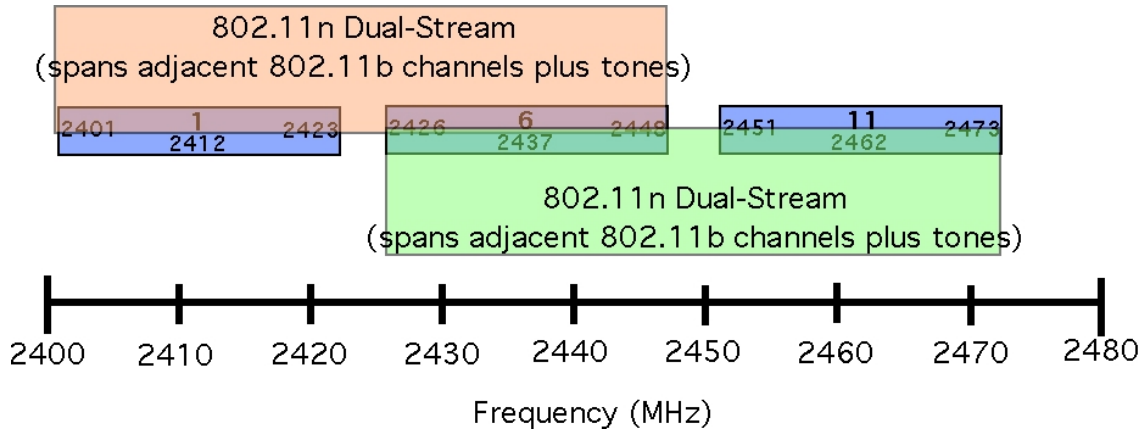


**Figure 10.8.** The Overlapping Channel Assignments for 802.11 Operation in the 2400 MHz Range

The situation for 802.11 in building automation settings warrants further examination. The IT departments at many, if not all, organizations have or are contemplating deploying 802.11 (Wi-Fi) networks in support of a wide range of applications that cross multiple business units (e.g., video surveillance, mobile operator support, etc.). From a frequency perspective, 802.11b/g utilizes the 22 MHz channels that are listed in Figure 10.9. A higher data/throughput rate is achieved in 802.11n. 802.11n provides an option to double the bandwidth per channel to 40 MHz. (802.11n operating in the 20 MHz bandwidth is frequently referred to as single stream. The 40 MHz situation is referred to as dual stream and provides approximately twice the data/throughput rate of single stream 802.11n.) 802.11n defines operation in the 2.4 and 5.7 GHz bands. However, when in 2.4 GHz, enabling the dual stream option takes up to 82 percent of the unlicensed band, which in many areas may prove to be unfeasible.

The 802.11n specification calls for requiring one main 20 MHz channel and an adjacent channel spaced  $\pm 20$  MHz away. The main channel is used for communications with clients incapable of operating

in the 40 MHz mode. When in 40 MHz mode, the center frequency is actually the mean of the main and auxiliary channel. 802.11n may operate in a “single stream” (20 MHz plus tones) bandwidth channel or “dual stream” (40 MHz+ tones) bandwidth channel. The 802.11n “dual stream” situation is shown in Figure 10.9.



**Figure 10.9.** 802.11n Dual Stream Occupies 44 MHz of Bandwidth. Dual stream 802.11n in the 2.4 GHz band is shown.

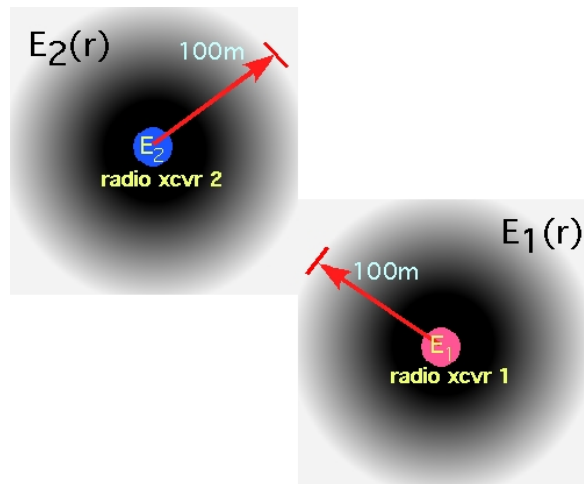
## 10.6 Coexistence of 802.11 and 802.15.4

The IEEE 802.15.2 Coexistence Working Group (WG) as well as the 802.19 RF Coexistence WG are tasked with addressing the coexistence issues associated with 802.11 and 802.15.1/802.15.4 transceivers. Coexistence between different wireless short-range devices, such as wireless sensors, using the 2.4 GHz and 5.8 GHz ISM bands is becoming increasingly more difficult and more important. Interference is increasingly an issue as wireless consumer devices proliferate. The IEEE specifications have stated that such 802.15-based devices are “secondary,” which means that they may not interfere with 802.11, and must tolerate any interference received.

An 802.15.4 transceiver employs a Clear Channel Assessment (CCA) mechanism to “determine” if there is interference on the frequency channel that it is attempting to broadcast on. Restated, CCA is used to determine if the channel is busy.

The topic of coexistence is relevant only when the RF footprints of two transceivers overlap. The situation can be as simple as that shown in Figure 10.10,





**Figure 10.10.** Non-Overlapping RF Footprints

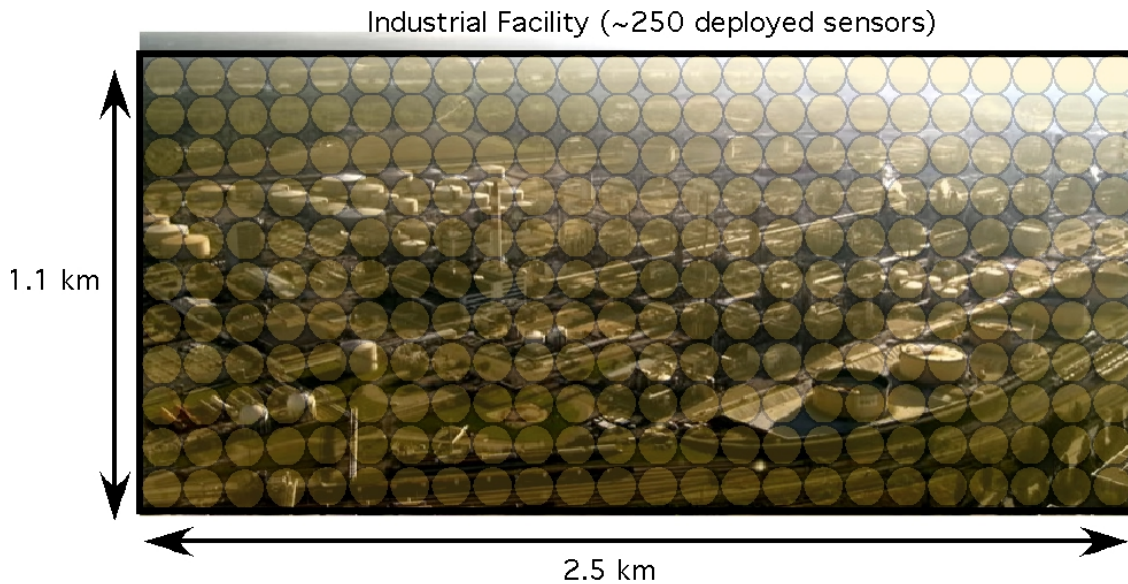
As shown in Table 10.3, the data rate will decrease as the separation distance between two wireless transceivers increases. The general rule of thumb for maximum separation between an 802.11 access point and an 802.11 client device is 100 m. A similar, 100 m, separation distance for 100 mW output omnidirectional 802.15.4 devices is routinely stated.<sup>2</sup>

Certain building automation wireless applications, particularly those related to mobile devices, greatly benefit from a seamless coverage of the wireless signal. The physical situation (e.g., reflectivity, variations in attenuation, etc.) presented by the building automation site make it problematic for a single RF transmitter (i.e., base station, access point, gateway) to provide acceptable coverage throughout the entire site. In such a situation, it is necessary to put in a network of base stations/access points/gateways.<sup>3</sup> As a mobile device moves from one coverage zone to another (i.e., roaming), the handoff of a device communicating with one base station to another must be coordinated.

The combination of RF frequency coexistence, overlapping RF coverage for seamless mobility, and the physical (spatial) distribution/placement of 802.11 access points in a large building automation site intersect directly in the situation voiced by a number of end users who are discussing the deployment of thousands of wireless field transmitters within a single building automation site. Using the inaccurate estimation of a uniform distribution of such devices, each operating with a 100 mW 802.15.4 transceiver, leads to the graphical situation depicted in Figure 10.11. This graphical depiction of the density of RF fields is for only approximately 250 wireless field devices with the same operational details as Figure 10.5 (the plant in Figure 10.11 is the same as that in Figure 10.5).

<sup>2</sup> There are seemingly innumerable factors that can, and will, influence this 100 m separation distance. It is recommended that such a value only be used for rough guidance especially at metal (reflectivity)-laden building automation sites or in buildings with metal frames or metal reinforcing between floors.

<sup>3</sup> The networking of base stations, access points, and/or gateways will be covered in an accompanying document.



**Figure 10.11.** Dense Deployment of Wireless Field Transmitters at a Building Automation Site

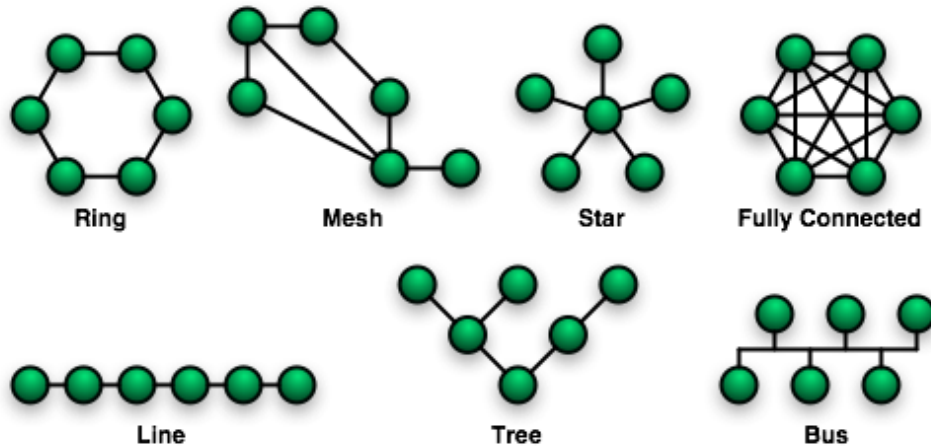
Notice that the situation in Figure 10.11 is for only 250 deployed devices. While not the focus of this report, various building wireless sensor networks may operate in a mesh network topology. This requires that each wireless sensor's RF footprint overlap at least one, and preferably more, other wireless sensors, thereby achieving the physical+RF situation required for multiple information pathways (i.e., a mesh network). Current realistic deployments of wireless sensor networks allocate approximately 20 to 50 sensors per network gateway. For most gateways, 802.11 connectivity is provided on the output side of the gateway. At the 1000 sensor device deployment density, this would imply 50 (to 20) Wi-Fi transmitters with associated 802.11 access points.<sup>4</sup> The 802.11 access points may be connected (bridge or mesh) via a wireless or a wired network. Figure 10.11 also indicates that there should be approximately 500 802.11 access points deployed for that site just to provide seamless coverage for mobility applications. The network connectivity also should be addressed. In the case of wireless connectivity between nearby access points, a typical situation is for the access points to link via the 5 GHz band while data service from each access point is conducted in the 2.4 GHz band. The result is a dense RF environment.

## 10.7 Network Topologies

The logistics associated with wiring thousands or tens of thousands of devices led to network developments in which the field transmitters could (somehow) share a common transport medium. This idea, in turn, led to a wide variety of network designs and protocols, with the vast majority being proprietary. Field transmitters (devices) would have identifiers that were transmitted within the data frame, thus allowing those network elements with some level of intelligence to sort out the readings and

<sup>4</sup> The differences arise from the 20 access point case, with 50 wireless field sensors per sensor gateway, or the 50 access point case, with 20 wireless field sensors per sensor gateway. The indeterminate nature of LAN traffic and the CSMA/CD operation of ethernet (802.3) led to Industrial Ethernet (IE used for industrial process control) operating at 1 to 5 percent of available bandwidth. The implications for building automation 802.11 is that the traffic patterns for the multi-use 802.11 (wireless Ethernet) backhaul network, perhaps carrying voice, video and data, should be considered when designing the network.

process/transport them accordingly. Improvements in network protocols and the robustness requirement of minimal, or even zero, single points of failure led to the variety of network topologies used today. Figure 10.12 provides a graphical representation of such network topologies.

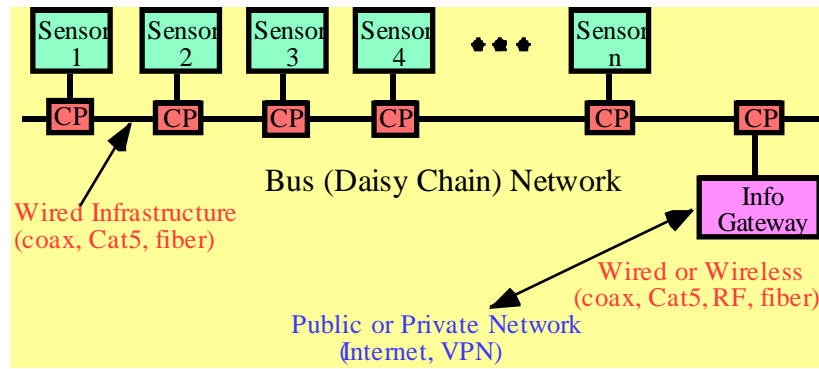


**Figure 10.12.** Networking Topologies Take Many Forms with Associated Levels of Complexity Required for Robust Fault-Tolerant Data Transport

Networking, or multiplexing, numerous sensors began almost the instant after the first sensor was developed. The issue was, and is, always the same; that is, how to optimally communicate the information from a number of sources across a signal transmission medium. While clarification or identification issues are frequently handled by being bundled into the message being transmitted, topological methods for configuring the information sources must be addressed. Networking topologies that are used in wireless sensor networks mimic those found in traditional computer networks.

### 10.7.1 Bus Networks

The simple situation of a linear series of information sources which are connected in the classic bus, or daisy chain, networking configuration is shown in Figure 10.13. In such a situation, the information source (a.k.a., the sensor) is physically attached to the network via a connection point (labeled CP in the figure). In the depicted wired network infrastructure, the physical media could be coaxial cable, Cat-5 cable, or even optical fiber. Regardless, at some place along the network, an information gateway is attached providing a means for the information to migrate from this bus network out to another network (e.g., the Internet, a Virtual Private Network, etc.).

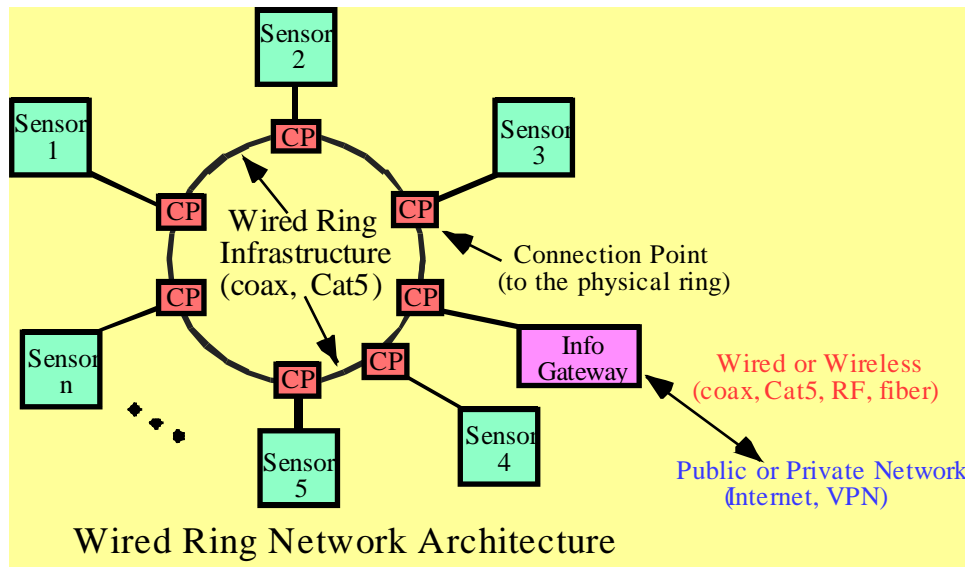


**Figure 10.13.** A Linear Chain of Sensors, with their Associated Connection Points (CP), Form an Easy-to-Scale Network

Extending this simple bus network topology to the wireless sensor network pictorially means simply replacing the wired infrastructure with some form of wireless channel. In reality, there are numerous possibilities for the wireless infrastructure's modulation and coding methods.

### 10.7.2 Ring Networks

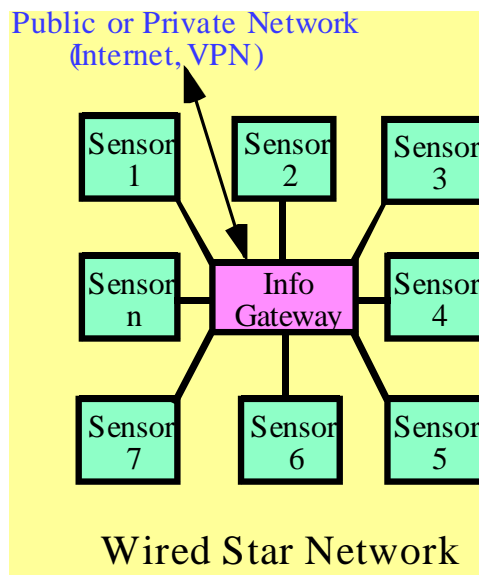
The classic ring network topology relies on the information proceeding around the ring. The function of the connection points is to provide the sensor with physical access to the ring infrastructure. The information originating from or destined to a specific sensor is somehow flagged in a manner that will cause the connection point to separate it from the other traffic and transport it to or from the sensor. In a typical configuration, some form of data concentrator, data repository, or information gateway is connected to the network. The goal of the gateway may be to provide a pathway to either a public network (e.g., the Internet) or a virtual private network (VPN). The physical connection between the gateway and the Internet/VPN may be wired (cable, Cat-5, optical fiber) or wireless (RF). The protocol used in the information transport between the gateway and the Internet/VPN, the gateway, and the network, or between connection points on the network may have many forms ranging from Ethernet to SONET (Synchronous Optical Networking) to a proprietary scheme. A traditional wired ring network system is shown in Figure 10.14.



**Figure 10.14.** Outside Access to a Ring Network Relies on an Information Gateway being Connected to the Ring

### 10.7.3 Star Networks

A wired star network, shown in Figure 10.15, is used in many server-client computer configurations. Communications between the information sources located within the star are channeled through one node. Frequently, this node also serves as the connection path to the outside world (Internet, VPN, etc.).

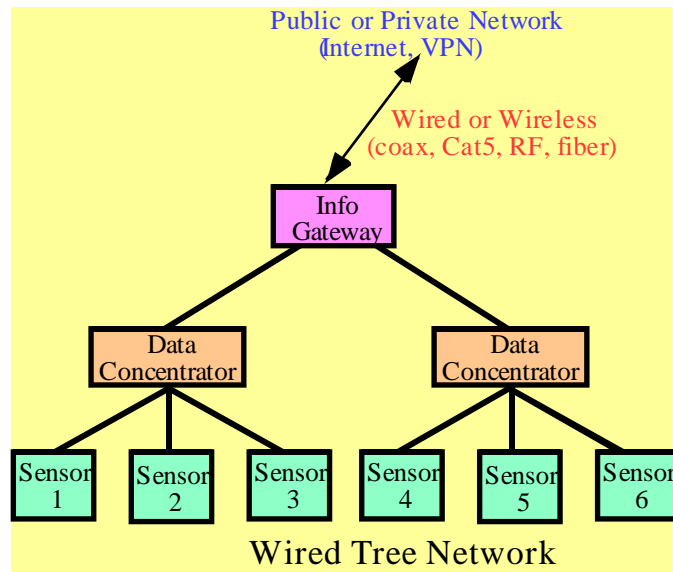


**Figure 10.15.** A Wired Star Network

### 10.7.4 Tree Networks

Tree networks have been used for many decades to provide inter-computer communications. This configuration, as shown in Figure 10.16, also provides a simple connection path for the sharing of

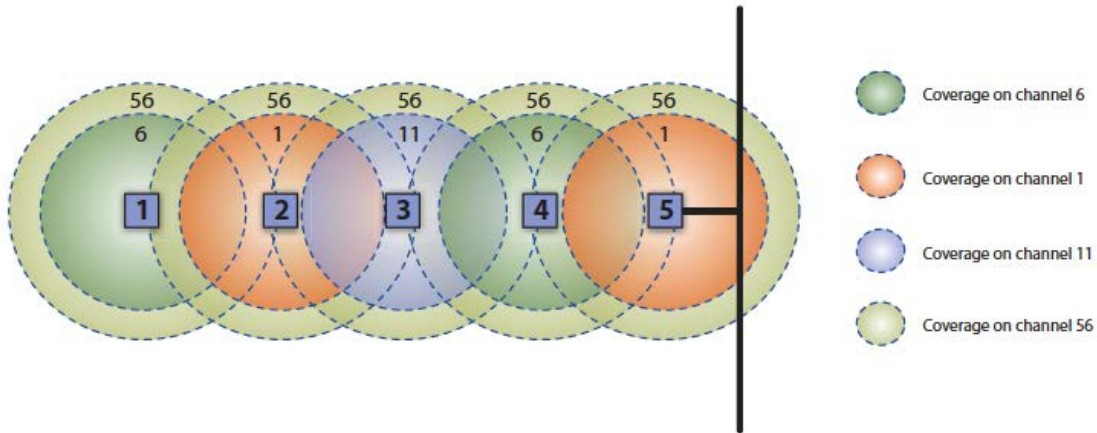
resources. A tree structure is observed in situations where a few computers are connected to a hub, and the hubs are in turn connected to a router. The router provides the access to the outside network. Such a system has proven to scale easily in computer installations.



**Figure 10.16.** Classic Tree Networking Structure

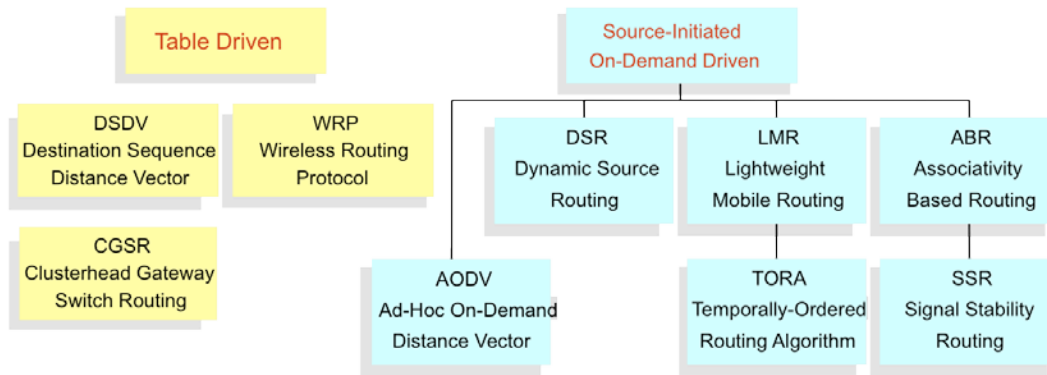
### 10.7.5 Mesh Networks

A wireless implementation that is marketed by numerous vendors is a mesh network world such as that shown in Figure 10.17. A typical ad hoc communication network may rely on multiple peer-to-peer information paths all of which reside within a given geographical region. A cornerstone of this topology is the reliance of the network on mobile wireless routers and an internetworking architecture. Each router has one or more associated host devices (e.g., computers and sensors) and implements routing protocols that tolerate rapid changes in connectivity among nodes. The routing algorithms organize the network by automatically discovering the topology of the connectivity among constituent nodes. There are many variants on mesh routing algorithms such as those illustrated in Figure 10.18. Please note that, while the sensors may be (physically) placed so they should function as a mesh, these routing algorithms are not interchangeable. In other words, a wireless sensor programmed to function using, for example, Dynamic Source Routing will not work with one that is in RF coverage range but using the Destination Sequence Distance Vector algorithm.



**Figure 10.17.** An 802.11 (Wi-Fi) Mesh Network Relying on Dual Radios (One Radio is used for “Local” Data Service, the Second Radio is used for Access Point-to-Access Point Communication). Seamless transport coverage occurs on Channel 56 while data service varies from access point to access point following, in this case, a Channel 6-Channel 11-Channel 1 repeating sequence.

- Taxonomy:



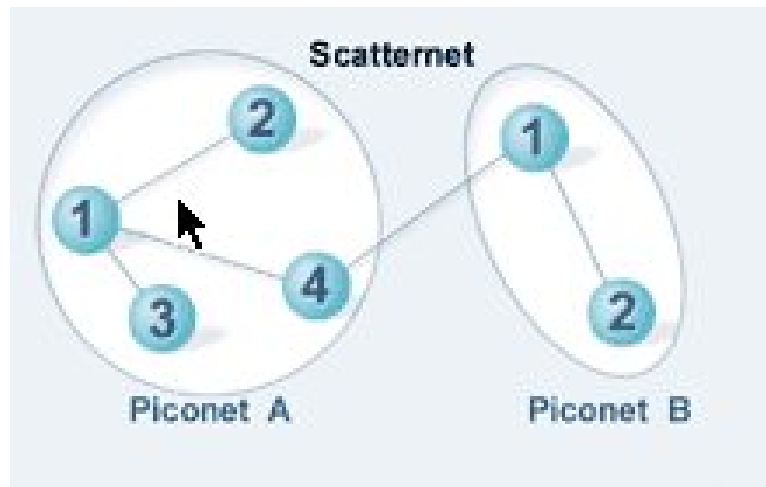
- Others:  
ZRP, FSR, LAR, CEDAR, CBRP, OLSR, and many more...

**Figure 10.18.** List of a Few of the Mesh Networking Algorithms used in Wireless Sensors

### 10.7.6 Mesh Networks Variants

A variation of the ad hoc network, the piconet structure (see Figure 10.19) advocated by the Bluetooth consortium, which in itself is a variant on the tree networking structure, may be suitable for sensor interrogation data transmission. First and foremost, the goal of this wireless radio scheme is low power transmission, low power consumption, limited number of devices within the network, and short range (<10m, sometimes referred to as a personal area network [PAN]). Within the Bluetooth implementation, the devices are to have self-awareness regarding their form and function, which is information that is transmitted into the piconet as the Bluetooth device wakes up. Trials to date have

shown that, in certain configurations, this ad hoc network, coupled with the need for one Bluetooth device to serve as the network master, consumes significant battery power as the network attempts to ascertain who the master is.

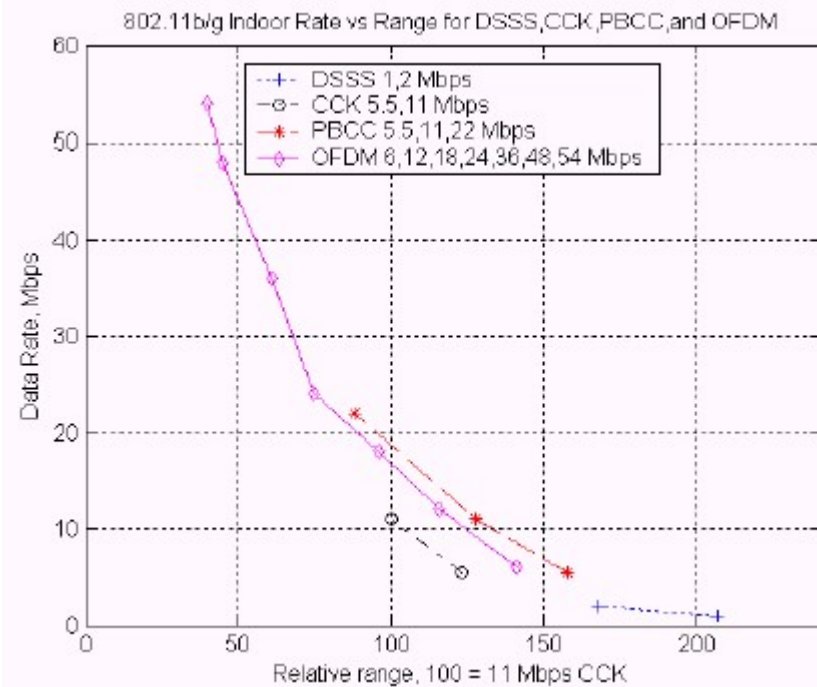


**Figure 10.19.** Bluetooth Devices rely on a Localized Network-to-Network Topology that is Meant to Allow Message Bouncing

The standard tradeoffs between transmitted power (a.k.a., range) and power consumption dominate the communication network comparison. Figure 10.20 simply states the generic names of various types of wireless communication networks (PANs, LANs, and wide area networks [WAN]) as a function of usable range. Notice that PANs tend to operate at low data rates and are typified by battery powered handheld personal digital assistants (PDAs) and digital camera transmission of information. Data rates associated with a slightly larger network “footprint” are loosely categorized as LANs. Such systems abound and are typified by the 802.11 class of wireless computer networks. The much larger wireless networks are WANs typified by cellular second- and third-generation systems.



Mode	Outdoor Range		Indoor Range	
	Meters	Feet	Meters	Feet
1 Mbps DSSS	550	1804	50	164
2 Mbps DSSS	388	1275	40	133
5.5 Mbps CCK	235	769	30	98
11 Mbps CCK	166	544	24	79
5.5 Mbps PBCC	351	1151	38	125
11 Mbps PBCC	248	814	31	101
6 Mbps OFDM	300	984	35	114
12 Mbps OFDM	211	693	28	92
18 Mbps OFDM	155	508	23	76
24 Mbps OFDM	103	339	18	60
36 Mbps OFDM	72	237	15	48
48 Mbps OFDM	45	146	11	36

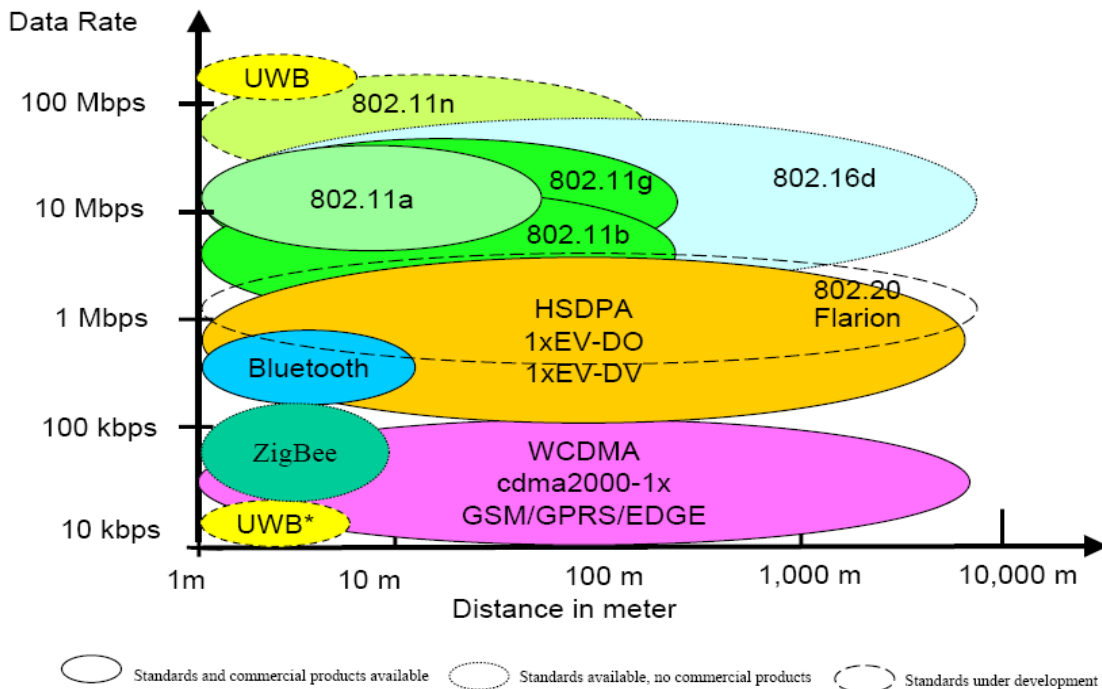


**Figure 10.20.** As the Range Increases, the SNR Decreases, which for Constant Error Rate Means that the Data Rate must Decrease

The power consumption of individual nodes within the piconet is exacerbated in the Scatternet configuration in which clouds of piconets are to communicate with each other. Given the nominal Bluetooth device separation distance of 10m (at 0 dBm radiated power), this configuration exhibits additional limitations in sensor data transmission needs to propagate from the device to a central data collection/interrogation facility. Anecdotal evidence of a Scatternet configuration of 80 Bluetooth devices attempting to self-arrange into various piconets highlighted this flaw in that the piconets, let alone the Scatternet, could not be formed before the batteries onboard the device began to fail. The net result is that, while viable, the Bluetooth implementation of wireless sensor networking is sub-optimal.

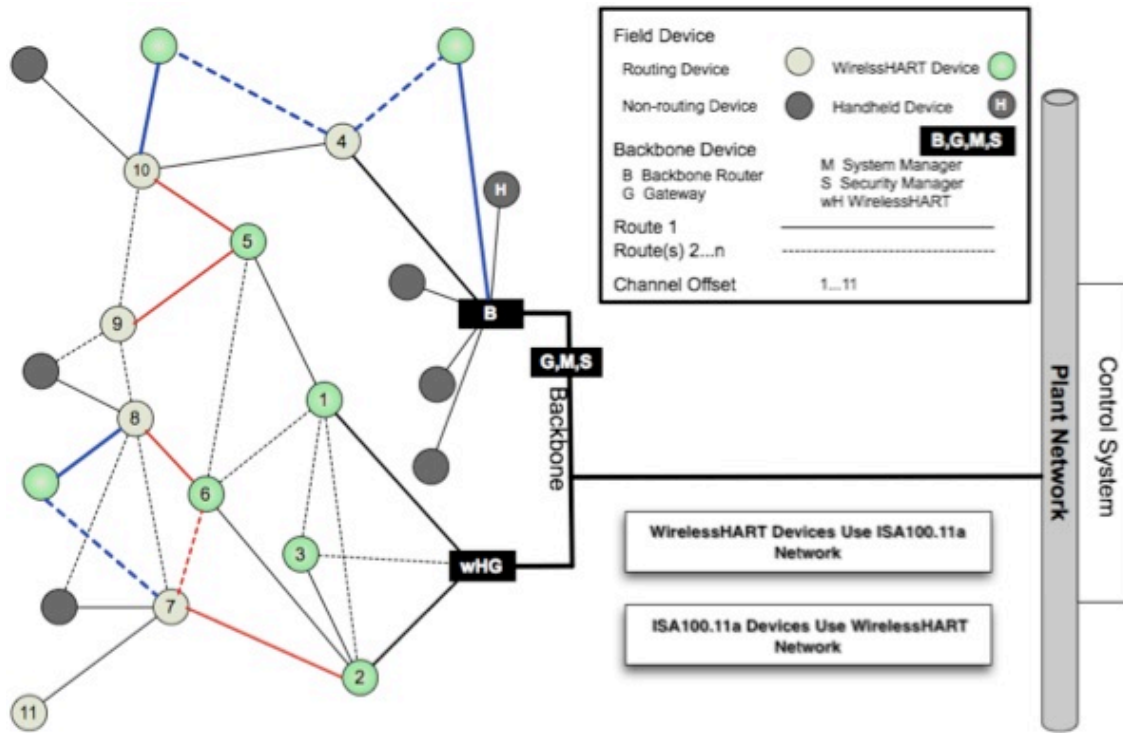
## 10.7.7 Data Rates for Backhaul Network Technologies

There are a variety of RF technologies that can be used in a building automation setting. Data rate and RF “footprints” are two parameters that are key to selecting the appropriate technology for such a setting. Consider the RF technologies shown in Figure 10.21: there are a number of operational considerations that are not listed in this figure, particularly if the system and devices are line-powered or battery-powered, which leads to their operational lifetime. While such functional matters are of considerable importance, the focus of this chapter is networking.



**Figure 10.21.** List of RF Technologies by Data Rate (vertical access) and RF Coverage (distance)

Integration of multiple technologies at a single facility requires coordination of the frequencies of the various technologies (to minimize co-channel interference). A candidate architecture for an integrated deployment in which wireless sensors and backhaul network elements are present is presented as Figure 10.22.

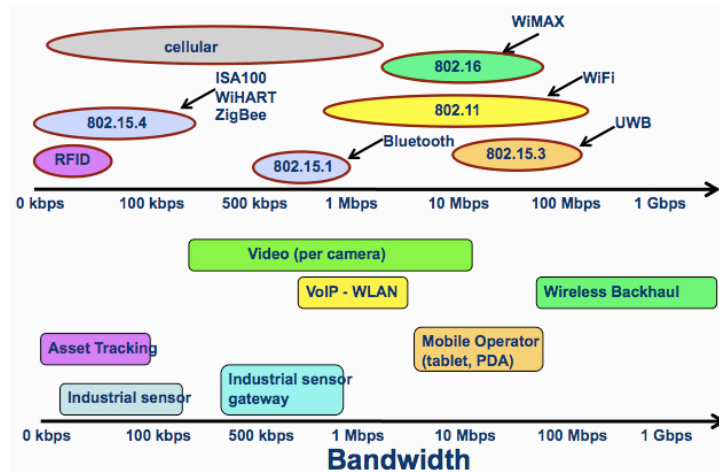


**Figure 10.22.** Integrated Wireless Environment for Building Automation Settings

## 10.8 Summary

The deployment and value of a wireless building automation system is based on two broad application classes: 1) those requiring mobility and 2) those derived from the reduced cost of attachment resulting from not having to run the wire. Such applications are best served by different wireless technologies typically based on response time and bandwidth requirements. An unscientific<sup>5</sup> mapping of applications-bandwidth-wireless technology is presented as Figure 10.23. The diagram is meant to depict the (approximate) upper boundary of the delivered bandwidth for the shown technologies.

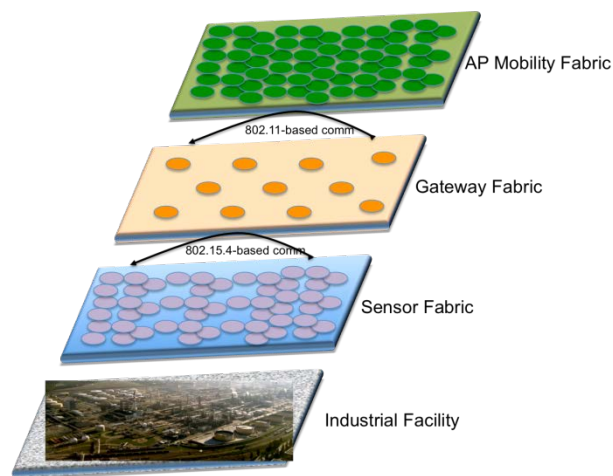
<sup>5</sup> “Unscientific” in the sense that this is not an all-inclusive list of candidate RF technologies. Also note that, for example, the bandwidth for 802.11 is depicted as roughly 1 Mbps to 200+ Mbps. The actual bandwidth may be as low as essentially 0 Mbps. Similar variations in the depiction of applicable bandwidths for the other technologies exist.



**Figure 10.23.** Bandwidth Requirements for a Variety of Applications and the Associated Wireless Technology that can Support such Requirements

Figure 10.23 illustrates that, in the case of wireless sensor networks using 802.15.4-RF underpinnings, the bandwidth for the transmission is on the order of 256 kbps, which is less than optimal for video transmission, and typically operate with a low duty rate resulting in relatively infrequent “bursty” communications. Similarly, the sensor networks are configured with typically up to 50 wireless field transmitters per gateway. The aggregate output bandwidth from the gateway is beyond the limits for efficient 802.15.4 transport and is more applicable for 802.11, 802.16 or similar backhaul technologies.

At the plant, this results in a structured fabric design as depicted in Figure 10.24. The small circles on the fabric layers represent an RF footprint originating from, for example, a 100 mW output, omnidirectional antenna transceiver 802.15.4 or 802.11 device. The primary purpose of the diagram is to illustrate how a layer of wireless sensor devices are intertwined with a layer gateway device that, in turn, may communicate with an 802.11-based dense RF footprint, which comprises the network fabric that mobility applications require.



**Figure 10.24.** Structured Fabric Design of Layered Wireless for a Building Automation Facility

Using a designed solution versus haphazard deployment results in a building automation site that may have a wide assortment of wireless technologies operating side-by-side at the plant with minimal (if any) RF coexistence, a logical integration into the operational networks.

## 10.9 References

1. O'Hara, Bob, *The IEEE 802.11 Handbook: A Designer's Companion*.
2. Gast, Matthew, *802.11 Wireless Networks: The Definitive Guide* (O'Reilly Networking).
3. Report to NIST on the Smart Grid Interoperability Standards Roadmap, Prepared for NIST by EPRI under Contract No. SB1341-09-CN-0031—Deliverable 7, June 2009.
4. Agar, Jon, *Constant Touch: A Global History of the Mobile Phone*, 2004 ISBN 1840465417
5. Ahonen, Tomi, *m-Profits: Making Money with 3G Services*, 2002, ISBN 0-470-84775-1
6. Ahonen, Kasper and Melkko, *3G Marketing 2004*, ISBN 0-470-85100-7
7. Fessenden, R. A. (1908). "Wireless Telephony". *Annual Report of The Board Of Regents Of The Smithsonian Institution*: 161–196. Retrieved 2009-08-07.
8. C. A. Balanis, *Antenna Theory Analysis and Design*, Second Edition, John Wiley & Sons, Inc., New York, 1997.
9. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, Second Edition, John Wiley & Sons, Inc., New York, 1997
10. H. Mott, *Antennas for Radar and Communications*, John Wiley & Sons, Inc., New York, 1992, pp. 115-180.
11. D. K. Cheng, *Field and Wave Electromagnetics*, Addison Wesley, Reading, Massachusetts, 1989, p. 84.
12. <http://www.ce-mag.com/arc>

## 11.0 Signal Data Aggregator: Legacy Building Interfaces

### 11.1 Introduction

Sensor systems used in buildings may have many sensing and actuation elements (i.e., transducers <sup>[1]</sup>) that are tied into a single system. The device that provides connections to these sensors and actuators and combines the resulting information is called a *data aggregator*. In the most basic sense, data aggregation simply means combining multiple measurements into a summary form. This can be as simple as tabulating data and storing it in a file or database, or as complex as taking many measurements from different sensing domains, data formats, and locations, and then calculating summary statistics. A typical data aggregator installed at a site is a box containing hardware and software components. Some of the software may also reside on a server on site or in the cloud.

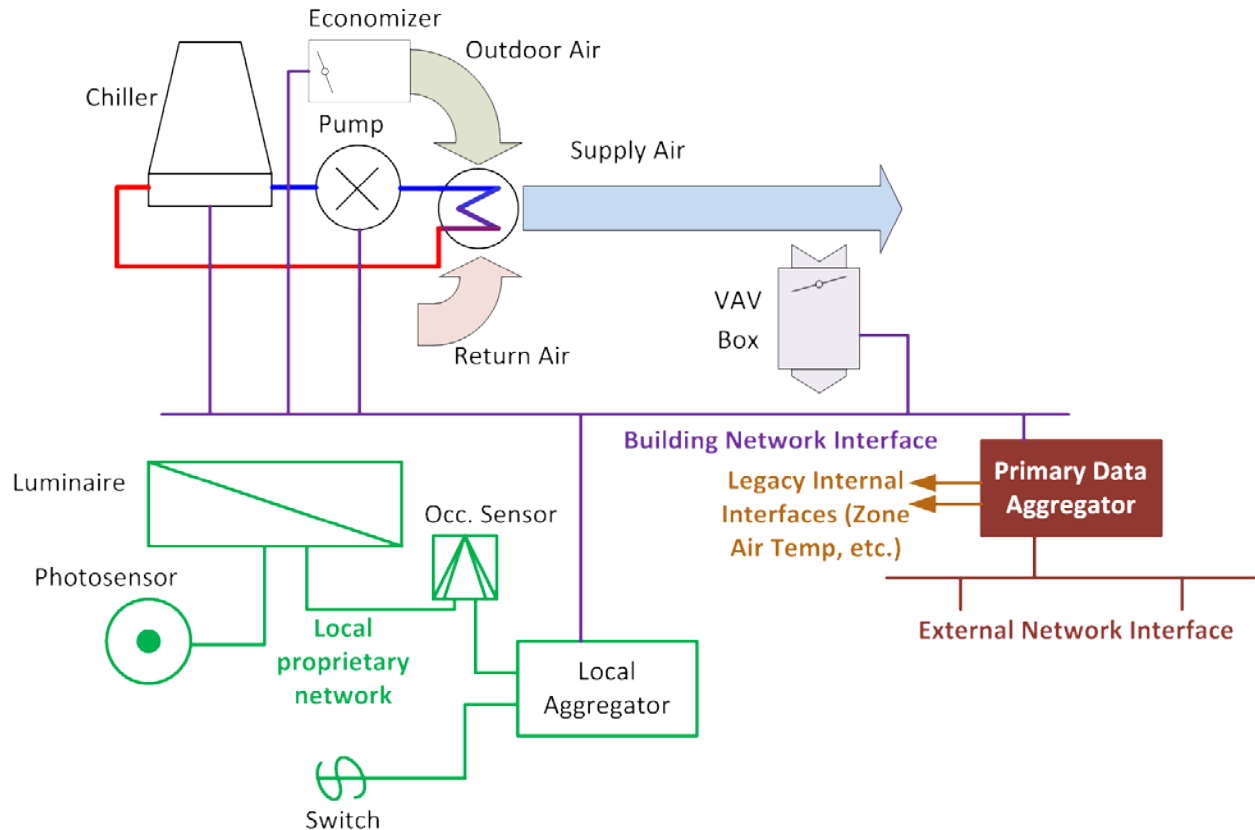
Figure 11.1 shows an example situation illustrating two different data aggregator types:

1. Multiple sensors and actuators are connected to a single local aggregator to provide local control.
2. Many data streams (in any format—analogue, digital, etc.) are collected, stored, and summarized on a primary data aggregator. This aggregator also is connected to an external network to provide external data access and display to users.

Elements of data aggregation overlap with other sensor system characteristics. The overlap generally falls into two case areas: 1) signal conversion to data and 2) communications and networking.

In the first case, an aggregator unit has inputs for low-voltage, occupant controls, light levels, and occupancy sensors and also a lighting control unit. This system receives signals from the three sensors and locally makes decisions regarding the appropriate light state and level. It can run autonomously or provide summary information over a network interface (e.g., BACnet, Ethernet) as shown in Figure 11.1. Although this type of aggregation is present in some digital control systems, it is simple and relatively uncommon. It is not the focus of this guide. While all the necessary technology aspects of this case will be covered, the guide will focus on the second data aggregation functional case.

The second case, which hereafter will be generically called data aggregation, combines many data streams from different sensors or sensor and control systems into a single, common data format. Some of these streams may be carried over existing and new network connections (i.e., building network connections), and other streams are analog or binary valued variables measured at the aggregator. This second set of connection types are considered “legacy” interfaces because they do not use advanced digital signaling. Instead, they rely on analog values, dry contacts, or other simple signaling methods. The data aggregator may include data storage and generate summary statistics, or it may simply pass the combined, reformatted data on to another system for storage and analysis. Data analysis is not covered in this guide. Instead the focus is on data conversion, aggregation, storage, and sharing.



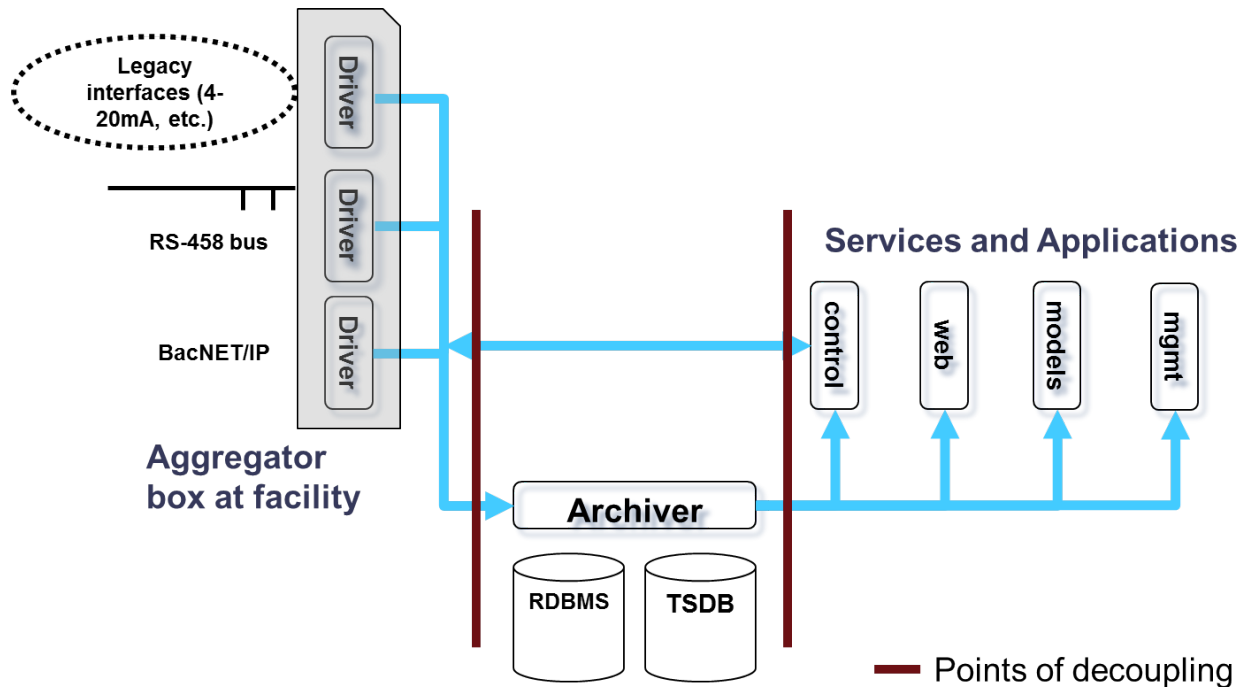
**Figure 11.1.** Common Data Aggregation Scenarios Showing Network-Based (purple) and Legacy Aggregation (orange) to a Primary Data Aggregator as well as a case where a Local Aggregator makes Control Decisions for a Lighting System and Provides Summary Data to the Primary Aggregator. The external network interface connects the data aggregator with applications and systems accessible over the traditional IT network.

A common example of data aggregation is one that connects different building subsystems that use different standard and proprietary protocols. The design of such an aggregator is such that all of these subsystems can be managed as a single system. This type of aggregator consists of many components, including the boxes with the physical interfaces, the software and network architecture for managing these boxes, and a centralized system with data storage, analysis, visualization, and control features.

Figure 11.2 shows the full architecture for a data aggregator system. Many types of data (e.g., power, temperature, humidity, occupancy, etc.) from many different locations are collected at relatively high rates and stored. The storage can be in the aggregator box, on a server at the facility site, or located offsite or in the cloud. The aggregator also may do some simple data analysis. For example, it may pre-compute hourly summaries of power data or zone average air temperature to ensure that the user or Energy Information System (EIS) can access the desired summary data rapidly.

All of the capabilities discussed above require computation, and most of the devices providing these functions use an embedded processor for this function. Embedded processors, sometimes called microcontrollers, are general-purpose processors with peripherals that provide specific functionality useful in embedded systems<sup>[3]</sup>. Examples of peripherals include analog-to-digital converters, various wire line transceivers (e.g., RS-485), encryption accelerators, digital-to-analog converters, comparators,

and amplifiers. Unlike other processor types, embedded processors often contain all of their memory on-chip or provide easy access to simple external memory peripherals. The processors are also optimized for performance under strict energy constraints and typically operate with limited memory compared to their standard counterparts. This guide does not explicitly cover this area of embedded processors, but relevant information will be included as needed.



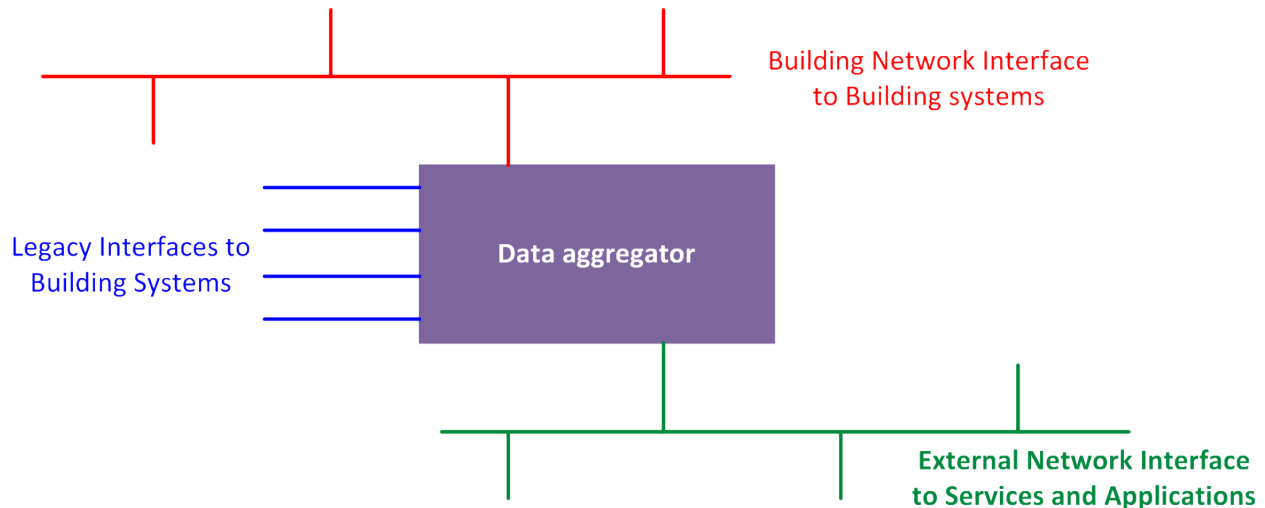
**Figure 11.2.** Full Data Aggregation System. The interfaces to the physical systems need to be located at the site, but the archiver (storage) can be located at the site or in the cloud. The services and applications shown on the right are not strictly part of a data aggregator and are not discussed in depth in this chapter. However, the link to these services is discussed. Adapted from Dawson-Haggerty <sup>[2]</sup>.

The term data aggregator is not a standard industry term, and the functionality encompassed by the term actually corresponds to functionality beyond that of a typical physical box. In this document, the systems performing the functions described above are generally called data aggregators, and the specific functionality being discussed will be identified. In conversation with other practitioners in research or industry, the particular functionality and form factor (software on a general server versus a dedicated box) should be discerned from the context or specifically defined, to ensure all parties understand the terms in use.

A critical function of the data aggregator is to work with legacy devices through simple interfaces (e.g., analog values, dry contacts) and convert them into a common format for communication or storage. Figure 11.3 shows the types of interfaces commonly found on data aggregators. This chapter will focus on the legacy interfaces on data aggregators. First, we identify the technical details associated with working with legacy interfaces, data storage, and the types of data processing used in aggregation. Then, we detail how aggregation of legacy interfaces is interdependent on other aspects of the sensor system and, in particular, the characteristics identified in this guide. Common applications in the buildings space are presented to highlight the application of the signal form conversion in data aggregation. Depending



on the type of legacy interfaces used, system cost can vary significantly, and the costs of current systems are quite high. We also review the issues that impact cost today and how changes in system design and implementation can affect cost.



**Figure 11.3.** Three Types of Interfaces in a Data Aggregator: Legacy, Building Network, and External Network. The legacy links, discussed in this chapter, are simple analog or single-bit digital links. Building network links provide higher-rate serial data with more advanced features (e.g., BACnet), and are discussed in the Chapter 12. The external network interface provides connectivity across traditional IT networks to software and services in the building or on the web (discussed in Chapter 13).

## 11.2 Technical Details

The primary technical issues with data aggregators and signal form conversion are related to the large number of physical interfaces required. In most cases, the physical interfaces are (digital) network connections that may include RS-485<sup>[4]</sup> or similar on twisted pair, various wireless connections, data on optical fiber, or any other communication media.

Some data aggregators include legacy interfaces to individual low-voltage analog or digital sensors. These are discussed in this chapter, although the specifics of how the conversions take place are covered in the Chapter 14. This chapter considers common available interfaces on low-cost hardware and the limitations of the interfaces. This section covers specific signal form types, issues associated with sensor cabling, analog versus digital transmission, and data tagging.

### 11.2.1 Legacy Interfaces

The most common legacy, non-network interface types on a data aggregator are for thermistors, thermocouples, “dry contacts,” 0 to 10 V, and 4 to 20 mA loops. Each of these interfaces requires interface electronics that enable the signal to be interpreted correctly. A discussion of each input type is included below.

### 11.2.1.1 Thermistor Interface

A thermistor is a resistor in which resistance varies with a variation in temperature. A thermistor has two common interfaces: 1) a constant current readout method, and 2) a bridge readout method<sup>[1]</sup>. In the constant current method, the voltage across the thermistor is measured when a small, known current is applied. The advantage of this method is that, when using a four-wire interface (i.e., the current is applied with one set of wires and a parallel set of wires is used for voltage measurement), the distance (i.e., length of cable) between the thermistor and the sensor interface does not affect accuracy. The disadvantage of this method is that the electronics required are relatively complex. With the bridge readout method, a temperature-stable reference resistor is placed in series with the thermistor, and a constant voltage is applied across this series pair of resistors. The relatively large change in voltage across the entire range of interest makes it easy to use this interface. The disadvantages are that the signal is non-linear with resistance and temperature, it is dependent on the reference resistor, and the cable resistance affects accuracy.

### 11.2.1.2 Thermocouple Interface

A thermocouple is a device that generates a voltage when a temperature gradient is applied across the device. One side of the device is held at a known temperature and the other side is put in the environment to be measured. The resulting voltage is very small and must be amplified to be interpreted. Thermocouples can measure extremely high or low temperatures; however, the reference temperature must be known, and the interface requires relatively complex electronics.

### 11.2.1.3 Dry Contact Interface

A dry contact is an electrical connection that does not carry the primary current of an application. An example is a switch for controlling a motor where the current to drive the motor does not pass through the switch. Instead, the switch state is an input to another circuit that controls the motor current<sup>[5]</sup>. This interface is used so the unit can be controlled from a distance without issues related to a long current loop. It also alleviates issues with contact design and wear. Dry contacts are sensed with a resistor in series with the dry contact. A voltage is applied across this series pair, and the voltage between the resistor and contact point is measured. If the voltage is near zero, the contact is closed. If the voltage is near the applied voltage, the contact is open. A simple digital input is all that is needed to sense this type of interface, making it popular in many applications. Dry contacts often are used in controls, so aggregator systems need to generate this type of interface and be able to measure it. A relay is the best method for generating a dry contact (because it is voltage independent), but it is possible to achieve a similar effect using a transistor under certain conditions.

### 11.2.1.4 0 to 10 V Interface

A 0 to 10 V interface is a low-voltage analog interface. The voltage applied or measured is proportional to the sensor or actuator output. In some cases, a non-zero voltage is selected (1 V is the most common selection) for the minimum signal level to allow the system to detect a disconnected sensor, which will read as a zero voltage. This is a simple interface, but it is easily affected by voltage drops across cabling. Most low-cost electronic devices do not natively support voltages this high, and these systems need external interface electronics to reduce the voltage.

### **11.2.1.5 4-20mA Interface**

Traditionally, the 4-20mA interface is an analog interface, with 4 mA representing the minimum condition and 20 mA representing the maximum condition. A disconnected sensor can be detected if the current is less than 4 mA <sup>[6]</sup>. The advantage of this approach is that the length of the cable is not important because the current is independent of resistance (as long as the source can provide a high enough voltage to maintain the current). This is more complex than the 0 to 10 V method because a current source must be used, but it is popular when accuracy over long distances is needed. As a result of its robustness, many digital systems use the 4-20mA interface as well. These applications are discussed in Section 12.

## **11.2.2 Sensor Cabling Advantages and Disadvantages**

Sensors are connected to data aggregators almost exclusively using electrical cabling (i.e., a set of wires). There are wireless sensors and optically wired sensors, but these sensor systems are not commonly used. Cabling enables the sensor or actuator to be placed away from the aggregator or controller. In addition, cabling brings several other advantages and limitations to sensor systems, including accuracy issues, fault detection, installation cost, and exposure to single-point failures <sup>[1]</sup>.

### **11.2.2.1 Cabling and Sensor Accuracy**

Using a cable to transfer analog signals is a solution that provides high or poor accuracy depending on the implementation. Connecting a highly precise and accurate sensor via a well-engineered cable to a highly precise and accurate aggregator can work very well. However, there are significant costs associated with the engineering, the shielded cables, and other factors. Typical installations use systems in which a voltage is applied to a wire and some current flows. The resistance in the wire results in a voltage drop and a reduction in accuracy. The resistance in the wire is dependent on many things, including the diameter of the conductors, the materials used, the condition (e.g., bends or other stress) of the wires, and the length of the wires. Given all of these factors, and that some change over time; they contribute error to the measurement.

### **11.2.2.2 Digital Transmission**

The primary alternative to analog signals over wires is to digitize the signal close to the source and send it over a cable or wireless connection. Digital data in well-designed systems are either received error free (i.e., with no loss in accuracy) or are received in a way where they are detectably corrupt (i.e., due to features conceptually related to checksums, which are a fixed-size datum computed from an arbitrary block of digital data for the purpose of detecting accidental errors that may have been introduced during its transmission or storage). Digital transmission provides a significant advantage compared to analog transmission in that the signal can then be sent over arbitrary distances and is uncorrupted by noise, interference, or time. The primary disadvantage is that users must provide reasonable quality interface circuitry (including analog-to-digital converters) at many places close to the sensor instead of potentially just a single location at the aggregator. Depending on the precision needed, the circuitry may be expensive or power hungry. With advances in electronics over the last several decades, high precision can often be achieved for low cost and low power, mitigating this concern to some extent. Another disadvantage is that, although digital transmissions are received error-free in most cases, digital signals do

not gracefully decrease in quality with increasing noise or attenuation. The signal is either received successfully or not, and no conclusion can be made from unsuccessful data. In an analog system, some decrease in accuracy may still be acceptable in some cases.

### 11.2.3 Data Tagging with Metadata

Metadata is information about data. In the case of time-series data of interest in sensor systems, data consist of a series of time-value pairs. The metadata can describe the units of the values, the location of the sensor both physically and in the system, manufacturer data, model number, and many other pieces of information useful for understanding the data. With traditional analog interfaces, this tagging must be done in the aggregator, and any changes outside the aggregator, such as the replacement of one sensor with another, must be recorded and manually entered into the aggregator. This prevents “plug-and-play” systems because the sensors cannot provide information about themselves. Generation of the data at the conversion step is critical, but the metadata are just as critical for a system that can be used, modified, and repaired. Digital systems often include some metadata, either with the measurements or in reply to requests. These metadata can be highly valuable for understanding, expanding, and improving an existing sensor system.

## 11.3 Characteristic Interdependencies

This section discusses the legacy interfaces and related functions of a data aggregator and how those interfaces either provide or limit the functionality of the aggregator. The physical processes for taking incoming signals (e.g., pressure, voltage, current, temperature, etc.) and extracting digital bits overlaps Chapter 14 of this guide, and the details are left to that chapter. The way signals and data are transferred through and out of a building overlaps with Chapter 10, and this chapter and Chapter 12 and 13 do not discuss those issues.

As described in the introduction, the aggregator may include multiple input and output types, including analog and binary valued signals along with different devices connected over a variety of network interfaces. All of these must be converted to a single format and stored and/or transmitted along another network link. This identifies a significant need that is not included in the reference guide: embedded processing capability. All of the functions related to the aggregator require some sort of processing capability, and the solution used in virtually all applications today is an embedded general purpose processor. Two classes of processors are common for these systems: microcontrollers and embedded versions of standard processors. Microcontrollers tend to be simpler, cheaper, and use less power than other processors while also integrating analog and other peripherals on-chip. Embedded versions of standard processors are more powerful, can run more standard operating systems (such as embedded Linux), but still integrate memory and other peripherals on chips to reduce size and cost <sup>[3]</sup>.

The processor is used to control the analog and digital interfaces, to process the data, and to manage network connections. Controlling analog and digital interfaces includes control of external interface circuits (e.g., amplifiers or level shifters) and on-chip subsystems, which include analog-to-digital and digital-to-analog converters. This control typically uses time information to determine when to measure a signal, but it can use measurements of other sensors or other information to trigger a sample. Any of

these options requires a processor to make decisions, control peripherals, and manage the resulting data. Managing a network connection also requires a processor because of the networking stack (even if it is a simple, peer-to-peer serial protocol).

Processors do not operate “out of the box” because they require specialized software to operate. On the lowest power and cost systems, this software is often called firmware because the software is difficult to change after the system is deployed in the field. Larger systems often use an embedded version of Linux, and some even run a version of Microsoft Windows. Developing software on the larger systems is far easier because only standard software development skills are needed. The smaller systems have higher development costs, generally for a similar amount of functionality. The hardware costs for the smaller systems, however, can be significantly lower.

Fundamentally, all hardware technology subsystems of a data aggregator are available at low cost today, but the software that is needed is only available at relatively high cost. Therefore, no particular hardware advance is needed to enable data aggregators. Software advances will provide value, but there are specific changes needed to enable truly low-cost aggregators.

## 11.4 Typical Applications

There are three common applications of data aggregators today: 1) to tie together multiple systems in a building when some systems are upgraded and others are left with legacy components; 2) for an operator who simultaneously manages several facilities that have disparate control systems that may not be connected, and 3) to use sensors to make local decisions (this third application is largely outside the topic area of this guide) <sup>[8]</sup>.

In the first case (i.e., upgrading a building controls system), the data aggregator serves to join new and legacy systems to provide better whole building visibility and control. These systems normally could not be connected into a single system (due to different vendors, incompatible versions, and other factors) but a data aggregator can provide the interfaces needed to achieve integration. Often, non-network connections (i.e., the connections discussed in this chapter) are used to connect to legacy systems because network connectivity, if it exists, may use a proprietary bus or protocol that is not supported by other systems or documented sufficiently for an interface to be created. The simple legacy input and output options are critical for this functionality because older systems often support these options and can be easily configured to “listen” to such signals or provide data using them. This type of deployment requires an aggregator box to be deployed onsite to physically interface to the legacy building systems.

The major issue with this use case is that these input/output options must be configured to operate on the various control systems. Although this often is possible and even relatively easy, it can be challenging, requires some site-specific engineering, and the labor can lead to high costs.

The second case (i.e., operation of multiple facilities) often occurs on campuses of multi-location corporate sites, where buildings, built over many years, have many different vintages or vendors of controls. Some companies provide solutions entirely based on getting data out of these buildings into a common web database for archiving and viewing by operators. This functionality goes beyond that of a simple data aggregator, but the data aggregator function is included in the service provided. An

aggregator box is deployed, sometimes at each building. In other cases, existing legacy links can be translated in the cloud to a new format and stored in the aggregation system.

Every new building (or control system) requires some amount of engineering to connect with an aggregator. An engineer must identify the existing systems, the software versions used, and available input and output options. They must then configure the aggregator to interface to those systems. Often, metadata must be manually tagged as well.

A third case in which a box receives sensing information to make local decisions is common, but it falls outside the main scope of this section of the guide. A lighting system is a common example where switch inputs, occupancy, and light-level sensors are used to make lighting-control decisions. This kind of aggregator is relatively small and does not provide the multi-system aggregation of primary interest in this guide.

## 11.5 Relative Cost Impacts

The cost of data aggregation systems is dictated by three items: 1) hardware cost, 2) non-recurring engineering (NRE) costs for setup, and 3) license or service fees. The NRE costs and license fees tend to dominate the overall costs because data aggregation is primarily a software issue rather than a hardware issue. Each of these cost areas will be addressed as it pertains to signal form conversion, and a speculative discussion about lower cost future is included.

### 11.5.1 Current Cost Impacts

Hardware costs are relatively low, but these costs could be reduced even more. The lowest-cost data aggregator boxes range in cost from a few hundred dollars to a few thousand dollars, depending on capability and vendor. Server costs are comparable to any database server (typically thousands of dollars but less than \$10,000). Servers are commoditized at this level, and costs are likely to follow market trends. The aggregator boxes, however, are highly specialized hardware that is priced higher than it would be in a high-volume, competitive market. Innovation is not required to reduce cost in any particular way, as volume manufacturing will drive design optimization and cost reduction.

Breaking the hardware costs down into components shows that no single component is a cost issue. An embedded processor that is easy to use, widely supported in embedded Linux, and reasonably powerful (capable of running an embedded web server and interfacing with multiple network links and many simple I/O options) is less than \$10 in small volume<sup>[8].<sup>1</sup></sup> Storage is becoming increasingly inexpensive. Hard disk drives have become very inexpensive on a per-gigabyte (GB) level (less than \$0.10/GB retail), and solid-state storage has also become very low cost (less than \$1/GB retail). For reference, 1GB of storage is roughly enough for 1 million readings. Depending on many variables, 1GB of storage can provide several days to many months of data buffering. All other components are typically less expensive than the processor and storage components. The NRE costs for hardware design can be large. Small volumes are common, resulting in high per-unit costs as the NRE is amortized over the small volume.

---

<sup>1</sup> The processor considered here is an ARM11, a popular embedded processor capable of running many embedded Linux distributions. The ARM11 processor has been widely used in applications ranging from the Apple iPad to the automation of large industrial systems.

NRE costs for installing and commissioning these systems tend to be large. In many cases, there is little to no documentation on the meaning of various signals. This information must be extracted from paper documents, current system setups, technician knowledge, and, if needed, through trial-and-error efforts. When the systems are well documented, this process is relatively quick. Standard efforts to solve this problem take a week or more of engineering effort. Although the cost to integrate legacy connections with a data aggregator can be large, they are typically less than integrating building network interfaces (e.g., BACnet) that connect to building systems.

Because of the closed nature of building control systems and the high NRE costs of installing an aggregator at a site, vendors tend to charge high one-time or moderate ongoing fees for their services. Some products have a per-instance fee (typically thousands of dollars but can be as low as a few hundred dollars) as well as ongoing fees (tens to hundreds of dollars a month). This cost, sometimes partially buried in hardware costs, can dominate the overall cost. There are limited examples of open source solutions to the data aggregator problem<sup>[9,10]</sup>, and these efforts are not well supported in industry.

System management is the final cost area. Hosting a local server appears to be relatively inexpensive, but there are ongoing costs associated with keeping a system patched and functioning. An employee or contractor must either be responsible for the system or the organization may experience downtime, service interruptions, and data loss, and this responsibility or reliability issue translates to a soft cost. Using a cloud services for storage minimizes management and improves some aspects of reliability, but there are ongoing, hard costs associated with using this system as well. In addition, loss of Internet connectivity can result in service interruptions, depending on how the system is setup. All of these factors influence system management costs.

### **11.5.2 Future Low-Cost Possibilities**

A lack of documentation and poor standardization of controls limits cost reduction for integration into a new site. The integration problem is primarily a software problem rather than one of designing new chips and complex hardware. Reducing the labor required to bring a system up in a new site is the critical issue for making data aggregators cost-effective in the market.

Standardized control interfaces, improved documentation, and, eventually, plug-and-play building systems are required to achieve low-cost options. To enable plug-and-play systems, two key things must happen related to data aggregators. First, devices need to be able to identify themselves and provide significant metadata on customizations to the system. Second, a flexible, portable driver library needs to be widely available so a system can use the identification information to select and configure an appropriate driver from an existing library.

## **11.6 Summary/Conclusions**

Legacy building interfaces on data aggregators provide a critical function: they enable the relatively simple connection of new, IT-connected systems to legacy systems with limited capability. There are many types of legacy interfaces that are analog or digital, and each has a place where it performs well. A key limiting factor to the use of these legacy interfaces is that the metadata, information about how to

interpret or adjust the information on the legacy interface, is frequently poorly documented or not known. As a result, drivers must be created on an individual interface basis to enable integration with data aggregation systems.

## 11.7 References

1. J. Fraden, Handbook of Modern Sensors: Physics, Designs, and Applications, 4th Edition, Springer 2010.
2. S. Dawson-Haggerty, "sMAP 2.0: Integrating and Managing Physical Data," LoCal Retreat, Jan 2012. <http://local.cs.berkeley.edu/wiki2/index.php/File:Local-winter-smap-2012.pptx>
3. J. Ganssle, Embedded Systems, Elsever, 2008
4. Guidelines for Proper Wiring of an RS-485 (TIA/EIA-485-A) Network, TUTORIAL 763, Nov 19, 2001, <http://www.maxim-ic.com/app-notes/index.mvp/id/763>
5. "LG System Controls Dry Contact," <http://www.lg-vrf.com/dry-contact.aspx>
6. "Understanding 4-20 mA Current Loops," [http://www.bapihvac.com/content/uploads/2010/07/Understanding-4-20-mA-Current\\_Loops.pdf](http://www.bapihvac.com/content/uploads/2010/07/Understanding-4-20-mA-Current_Loops.pdf)
7. S. Dawson-Haggerty, et al., "sMAP — a Simple Measurement and Actuation Profile for Physical Information," in Proceedings of the Eighth ACM Conference on Embedded Networked Sensor Systems (SenSys '10), November 2010
8. [www.digikey.com](http://www.digikey.com)
9. Tridium Sedona Framework, <http://www.sedonadev.org/>
10. sMAP 2.0, <http://code.google.com/p/smmap-data/>

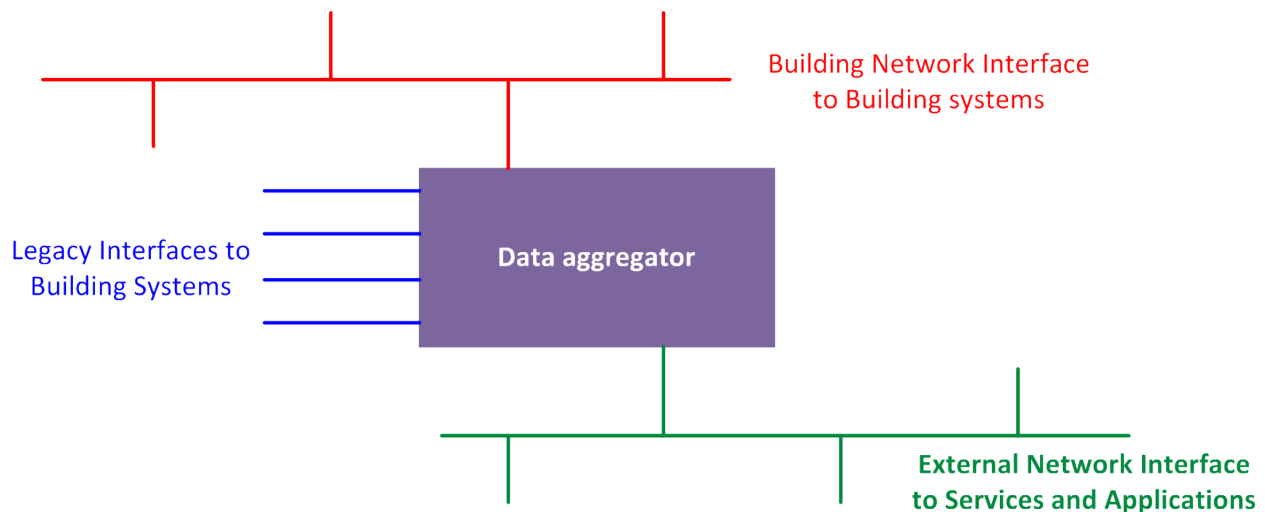


## 12.0 Signal Data Aggregator: Building Network Interfaces

### 12.1 Introduction

A full introduction to the general topic of data aggregators is included in Chapter 11, which provides the necessary background and context on data aggregators needed to understand many data aggregator issues not covered explicitly or in depth in this chapter.

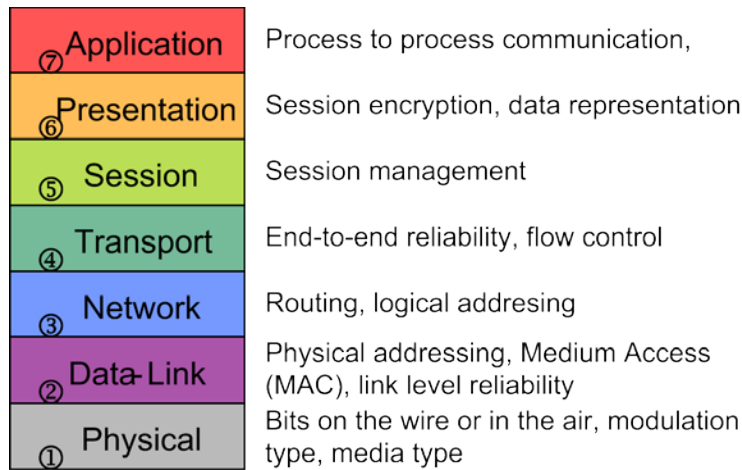
A critical function of the data aggregator is to connect to one or more building systems over relatively complex single or multi-system network links. “Network interfaces” are differentiated from “legacy interfaces” in that more than a single variable can be communicated over the link, and communication is often bidirectional. Figure 12.1 shows the different interface types.



**Figure 12.1.** Data Aggregator with Three Types of Interfaces: 1) Legacy, 2) Building Network, and 3) External Network. The legacy links include simple analog or single-bit digital links. Building network links provide higher-rate serial data with more advanced features (e.g., BACnet) and are discussed in this chapter. The external network interface provides connectivity across traditional IT networks to software and services in the building or on the web.

Networks as discussed in this document refer specifically to computer networks<sup>[1]</sup>. Although the computers on the network may be quite simple (i.e., just an inexpensive processor with limited software and capability), the network provides data connectivity and the ability to share resources (e.g., storage, computation). Conceptually, networks are simple interconnections between devices to share information. In practice, networks are complex software and physical systems with many opportunities for both confusion and failure.

The software and hardware architecture of a network is called the network stack. Ideally, the stack consists of segmented layers with well-defined interfaces at the top and bottom to ensure interoperability and the option to replace any single layer with another compatible layer without the user noticing any particular changes in behavior. Figure 12.2 shows the Open Systems Interconnection (OSI) model for a network stack<sup>[1,2]</sup>.



**Figure 12.2.** Open Systems Interconnection (OSI) Model for a Network Stack, Showing the Name and Number of Each Layer Along with Information about What Happens at each Layer <sup>[1,2]</sup>

Layer 1 (the bottom layer) is called the physical layer (PHY) and includes the definition of how bits are encoded on the wire, over the air or encoded on other media. The most basic data processing aspects of the system, which include how the data is verified as valid in the most basic way, are included as well. Layer 2 is the data link layer, which more commonly is called the MAC layer<sup>1</sup> because this layer controls how and when the device decides to listen for a signal to receive or transmit a signal for others to receive. This layer also includes physical addressing (e.g., the MAC address to your computer), and link layer (single hop) reliability. Layer 3 is the network layer, which plans a path for data to get from one location to another. Layer 4 is the transport layer, which provides end-to-end reliability as well as flow control (how much data can be packed onto the link before it becomes unreliable or unusable to other users). Layers 5 and 6 are the session and presentation layers, respectively. These layers include management of ongoing communication sessions and end-to-end encryption among other things. Layer 7 is the application layer, and, as the name suggests, it handles the communication between applications on the devices at each end of the network.

An example application layer protocol is HTTP<sup>[3]</sup>. HTTP is the application layer protocol used by the web server on one end and the web browser on the other. Note, this is different than HTML<sup>[4]</sup>, which is the language used to encode the web page itself. This distinction is important because HTTP can be used to transfer any content, whether it is one of many versions of HTML or just some proprietary binary data. Just because two devices both have compatible stacks and are using the same application layer protocol, it does not mean that they can actually communicate. If the language of communication is different (e.g., HTMLv4 on the browser and HTMLv5 on the server), the browser will not be able to display the webpage correctly or respond to the server appropriately.

In the networks of interest for this work, not all of these layers have equal relevance. Application, transport, network, link/MAC, and physical layers all play a role and will be used in the following discussion. This is not a chapter about networking in the direct sense, but it is important to understand some networking basics in a discussion on network interfaces being used with data aggregators.

<sup>1</sup> Technically, the MAC layer is a subset of Layer 2, with the remainder of Layer 2 being made up of the logical link control sub-layer. This distinction is rarely made in most discussions.

In this chapter, we discuss the importance of building network interfaces and the issues they present for data aggregators. We first identify the technical details associated with the network interfaces, including options at the various layers and how they present interoperability challenges. With this technical basis covered, we provide further detail on how this chapter is interdependent on other aspects of the sensor system and, in particular, the characteristics identified in this reference guide. We present common applications in the buildings space to highlight the application of network interfaces in data aggregation, and we discuss the cost impacts of parts of the system.

## 12.2 Technical Details

The primary technical issues with data aggregators and building network interfaces are related to the large number of interfaces used and the wide variety of languages carried over these communication links. In most cases, the physical interfaces are (digital) network connections that may include RS-485 or similar on twisted pairs of wires, various wireless connections, digital communication over legacy analog links, among others and these are the interfaces of particular interest in this chapter.

This section covers specific interface types, issues associated with choices at each layer in the network stack, bidirectional versus unidirectional communication, and understanding received data.

### 12.2.1 Common Network Interfaces

This section will introduce many of the common network interfaces used but will not provide in-depth information about the specifics of the protocols themselves. The common network interfaces will be used primarily to introduce the variety of connection types across the network stack and set up discussion on the advantages, disadvantages, and challenges of particular network interface decisions.

The most common network interfaces are over a single twisted pair of wires and include the most common BACnet<sup>[5]</sup>, Modbus<sup>[6]</sup>, and LonTalk<sup>[7]</sup> implementations, as well as countless proprietary network links. The primary advantage of this type of wiring is that the cabling is inexpensive (single-line telephone wire is often used) but sufficient for the low-bandwidth communication required in this application. The primary disadvantage is that running cable of any kind is expensive, and most network installers are unfamiliar with its use and awkward termination strategy. The limited bandwidth makes aggregation challenging because less bandwidth means fewer devices on the single network for a given communication latency and throughput per device. The physical interface in this case is often a set of screw terminals where stripped wire must be screwed into a connection point. RS-485 links<sup>[8]</sup>, the most common Layer 1 used on this media, also needs “termination” resistors added to the end of the link to enhance reliability, and the addition of termination resistors is widely seen as a non-issue despite the obvious (and common) errors and faults that occur with their use. Many of these links use a token passing- or polling-based MAC method rather than a more modern MAC architecture.

Legacy analog links can provide digital data connectivity over a single twisted pair as well. The most common example is using a 4 to 20 mA analog loop to carry network traffic. The Highway Addressable Remote Transducer (HART) protocol is the most common standardized version of this system<sup>[9]</sup>. In HART, both the legacy 4 to 20 mA analog signal and a new digital signal riding on top of the analog signal can be used simultaneously for communication between a single device and a single endpoint. Alternatively, multiple devices can communicate with a single endpoint using the digital-only signal.

This twisted-pair link typically connects using screw terminals but does not require a load resistor, thus simplifying installation. However, a link of 4 to 20 mA loop also tends to have lower throughput than an RS-485-based link. HART links use a polling-based MAC.

Ethernet- and TCP/IP-based communications provide the other most common methods of communication<sup>[10]</sup>. Ethernet consists of a specific set of physical layers, the most common of which are over four twisted pairs with a standard RJ-45 connector. So-called Ethernet cable is extremely cheap (comparable with telephone cable), includes standard and robust connectors, and is widely understood by installers. Ethernet also includes the ability to provide power over Ethernet using a standard mechanism widely used in the computer network industry. Using Ethernet and TCP/IP provides proven reliability, lower-cost software development, and, in theory, communication on existing IT networks. Modbus, BACnet, and LonTalk all provide a way to communicate over Ethernet more or less gracefully. More important, the relatively high data rate (commonly 100 Mb/s) means that many devices can be included on a bus, and these devices can communicate a great deal of data without issue. However, the infrastructure required is far greater than for an RS-485 link because network switches are commonly used to connect devices. Ethernet uses carrier sense multiple access (CSMA)<sup>1</sup> as its MAC method, and this provides advantages in both simplicity and performance compared to token passing or polling-based MACs.

The last increasingly common network connection is a wireless connection. These span the use of IEEE 802.11 based Wi-Fi networks<sup>[11]</sup> using TCP/IP to IEEE 802.15.4<sup>[12]</sup> based ZigBee™<sup>[13]</sup> or 6LoWPAN-based networks<sup>[14]</sup>. Cellular data connections using SMS, IP-based, or other connections are also possible and popular in some situations. The MAC choices here are almost universally CSMA for local area networks and a combination of time, frequency, and code division multiple access (TDMA, FDMA, and CDMA, respectively) for cellular networks. Some local area networks like IEEE 802.15.4e (based on technology from Dust Networks)<sup>[15]</sup> do use TDMA and FDMA to provide higher reliability and network performance.

### 12.2.2 Constraints Set by the Network Stack

The network selected by the vendor of the systems installed in a building can limit the performance of the aggregator. In many cases, these links were included as either an afterthought or never intended to provide real-time access to the full set of variables available in the system. Because many systems share the same link, it can quickly become the case that, although information can be extracted from the entire network, it comes at the price of significant restrictions on the sample rate and latency of the information. This section will particularly focus on decisions at Layers 1 and 2, as well as the top layer (i.e., the application layer), rather than focusing on any specific network issues.

In Layer 1, the data rate and resilience to interference or noise are directly related to the throughput and latency available for communication. These parameters are critical when considering how many devices can be supported on the network, and how much data can be reliably passed over the link. The encoding can affect reliability as well if error correction and/or error detection are included in the PHY specification. Timing of bits and synchronization signals must be precisely controlled, or no

---

<sup>1</sup> CSMA means that the device listens to see if someone is already transmitting on the medium before it starts to transmit its message. Typically, if the channel is in use, a “backoff” strategy is implemented in which the device checks the channel periodically to see if it can send its information.

communication can take place. Data aggregators must implement multiple PHY protocols with the flexibility to adjust to slight variations in implementation from vendor to vendor. Even standards like RS-485 and RS-232 can take on a variety of bit rates, data set ups, and checksum formats. The data aggregator not only has to deal with a network provisioned for a purpose other than providing aggregation services, it must also provide great flexibility.

Layer 2 decisions can have a significant impact on throughput and latency as well, and the number of variations available for a particular MAC is very large. First, the MAC method impact is primarily on latency and reliability. In a token passing network, a device must receive permission from the network to transmit anything. The token (i.e., the permission) is passed serially around the network so each device has a chance to transmit. If the network is large, a device with much to say must wait to get the key and then may only have a specified amount it can transmit before it must pass the token on. Although this provides a degree of fairness, it makes networking less deterministic than methods where devices are allocated specific intervals for transmission.<sup>1</sup> A CSMA MAC on wired links tends to be very high performance as long as only a few devices are trying to use the link at any one time. TDMA links are very similar to token passing links in that a schedule determines when a device can transmit, but the amount of bandwidth allocated to any particular device can be increased or decreased according to need.

Layer 2 implementation decisions regarding back-off, methods of access control (e.g., CSMA, TDMA) can make communication impossible between two devices with similar but slightly different MAC structures. Some MACs are implemented in hardware, and others are completely implemented in software. Hardware MACs tend to be highly deterministic (consistent transmission to transmission) but cannot be changed to match other systems. Software MACs often do not provide the same level of timing accuracy sometimes critical for operation but are flexible enough to accommodate a variety of systems. As a result, as much of the MAC as possible is implemented in software in most data aggregators. The low cost of high-speed electronics and efficient software design strategies have enabled software MAC layers to be highly effective. It is not uncommon, however, for systems with different MAC implementations to be sold under the same nominal interface type, with the most common example of this being ZigBee™. These systems can be completely incompatible at Layer 2 because of small variations in implementation.

The interior layers of the stack also must be implemented in a compatible way or communication is not possible. This information is covered in the other networking chapters, but the general issues are similar to those discussed for Layers 1 and 2. For most of the non-wireless links discussed here, the interior layers are very simple and implemented in software, resulting in a high degree of flexibility being possible. The data aggregator must be aware of the selections in these interior layers to ensure interoperability.

The application layers supported by the data aggregator must be broad enough to enable integration with a wide variety of systems. There are a very large set of application layers, and this layer is the most common to be implemented as fully proprietary or as some proprietary variant of common application layer. Application layers include the top layers of Modbus, BACnet, LonTalk, the Smart Energy Profile (and other ZigBee™ profiles), as well as Internet standards like HTTP, SNMP, etc. An important point is

---

<sup>1</sup> No method of communication is deterministic, in that errors are always possible because of a variety of factors. Some networks are more predictable than others.

that implementing the application layer correctly is critical, but the determination of what information is communicated using the application layer is still left to the developer, with a great degree of flexibility.

### **12.2.3 Bidirectional Versus Unidirectional Data Transfer**

Data aggregators and commissioning agents often must be able to interrogate the devices of interest in order to determine variables, gather metadata, participate in control, change set points, etc. Some links are only unidirectional (the device can talk but cannot listen). The key limitations of unidirectional data transfer are:

- A link can only support a single transmitting device because there is no MAC.
- There is no way to request information about the contents of its data packets from the transmitting device.
- Data often are asynchronous (i.e., not synchronized in time with other data sources), making analysis more challenging.

The addition of digital data on 4-20mA loops, in part, was done to alleviate these problems, and other methods have appeared over time as well. The end device must support these methods, however, for the data aggregator to be able to take advantage of them. The data aggregator must be robust to the issues with unidirectional data.

### **12.2.4 Understanding Received Data**

The largest challenge for data aggregators is in understanding what to send and how to interpret what it receives. Despite all of the challenges identified thus far, the greatest complexity comes from the place where standards are least prevalent and system implementers take the greatest liberties. We will use BACnet as a key example because it has a standard set of objects and the ability to extract metadata (if it was programmed into the device) over the available network link.

BACnet defines objects such as analog output, binary value, etc., and each of these objects comes with a variety of properties. Examples include a name, identifier, type, description, flags, units, current value, and many others. Only the first three properties are required, however, and there are no standards as to how the other properties are filled in (or even if they are provided). Therefore, to commission a data aggregator, it is common to come across BACnet objects where the properties useful for understanding the system are either cryptic or missing altogether. Unfortunately, BACnet is considered the easiest of existing protocols to integrate because vendors usually put some information into the objects and properties.

In many other cases, there is no mechanism defined for extracting information about the data format from the system, and the data aggregator installation requires significant effort just to understand the data that is available over a link.

### **12.2.5 Device Drivers for Data Aggregators**

A data aggregator system must have a driver for each connected device over a particular network link. That driver will specify not only how the network link itself is configured but also how to interpret the

data that is received and how to format data to transmit it to the device of interest. Today drivers tend to be developed as needed by system integrators around the country and placed into that integrator's local library of drivers. Customers pay for NRE needed to develop a driver from their system from scratch in many cases. Some systems come with a library of drivers, but it is common for systems in the field to vary somewhat from the design specifications because of slight version changes in software, customizations to the facility in the field, or any number of other reasons. Therefore, even existing drivers often are edited or tuned to meet the needs of a particular application. These drivers form the backbone of the data aggregator and provide the largest value to the overall system.

The commercial system with what is believed to be the largest semi-portable driver set is the Tridium Niagara Framework<sup>[16]</sup>. Integrators around the country have developed drivers to a large number of systems, and there are some reports of sharing between integrators and Honeywell/Tridium and between Tridium and the integrators. This driver-based model is critical to the cost reduction and increased use of data aggregators in sensor systems.

Open-source alternatives to the relatively high cost Tridium Niagara Framework include the Sedona Framework<sup>[17]</sup> and sMAP<sup>[18]</sup>. Sedona, owned by Tridium, is an open source implementation of Tridium Niagara without the licensing fees, but it has greatly reduced capability and driver support at this time. The sMAP system, developed by the University of California, Berkeley, is another open-source project looking at developing drivers to interface to devices in buildings. Both Sedona and sMAP are a combination of software tools and hardware that interface to the physical network links. As such, they are outside the scope of this guide.

## 12.3 Characteristic Interdependencies

Network interfaces in data aggregators have significant overlap with other areas of this guide. As described in the introduction to Chapter 11, there are many types of connections on data aggregators. The data aggregator's network capabilities are the key reason that it provides so much value.

Data aggregators typically provide two or more types of network connections:

- The connections to the devices of interest in the building. These connections are discussed in this chapter.
- The connection to a building automation system, data archiver, or other tool. This connection is discussed in Chapter 14.

Detailed aspects of understanding which network connection to use and how performance is determined are covered in the chapters on communications and networking in this guide. This chapter only includes the basic information that is needed to understand the aggregator itself.

The missing components of highly successful and low-cost data aggregators are all associated with software libraries, standardization, and interoperability. The hardware, as discussed in the legacy interfaces chapter, is available and inexpensive. There are too many poorly specified or followed networking standards used by many poorly documented devices to enable a universal and inexpensive aggregator solution. There are two potential solutions:

- A dramatic change in the way building systems are built and documented, to ensure that all systems are easy to interface with one another using a minimal amount of labor.
- A large and growing open library of device drivers that is easy to modify to new situations and apply to an open aggregator platform.

The first option is an important step toward an improved future because it enables low-cost and efficient buildings in a way that other solutions do not. This is a plug-and-play vision for building controls, where devices can provide enough information to the aggregator so that the aggregator can configure a skeleton driver to achieve full operation. However, this option is not particularly relevant in most retrofit situations, where only the monitoring (and perhaps controls) systems are to be replaced. If the legacy devices remain, then they will not support the high levels of interoperability, automation, and documentation.

This first option requires only limited technology advancement but significant changes in policy and/or implementer and customer decisions. Control systems currently are built with a reasonable model of the building physical systems either explicitly in the control system or implicitly through the design of the sequence of operations (e.g., the programmer knew there was a chilled water loop powered by a certain pump and cooling tower and built controls accordingly). It is very important to include a flexible controls description language that enables abstract control functionality to be applied to varying hardware configurations, to increase data aggregator usefulness and reduce costs.

The second option is the intermediate solution that, if pursued, could reduce costs and increase market uptake in the relative near term. The library of drivers would need to be supported by a lively community of both developers and integrators to be successful. The key to the success of this model is that the aggregator and drivers are just part of the overall service. Tied in with the open aggregator would have to be features that create real value for the customer, such as fault detection and diagnosis and continuous commissioning, among others.

## 12.4 Typical Applications

There are three common applications of data aggregators today: 1) in an existing building that is receiving a control system upgrade that combines an upgrade of some components with the use of some legacy components; 2) for an operator who simultaneously manages several facilities that have disparate control systems and may not be Internet connected, and 3) to use sensors to make local decisions (this application is largely outside the topic area of this guide).

In the first case, upgrading a building controls system, the data aggregator serves to join new and legacy systems to provide better whole-building visibility and control. These systems could not be integrated into a single system normally, but a data aggregator can provide the interfaces needed to achieve integration. The network connections used between existing devices and a legacy automation system may be used to connect to the aggregator. The legacy automation system may be disconnected from the building with controls provided through the aggregator, or it may remain in the loop with the aggregator, serving primarily to monitor systems. Using these network interfaces requires the development of software drivers to connect to the devices over that network link. This deployment requires an aggregator box to be deployed onsite to physically interface to the legacy (and often the new) building systems.



The second case, operation of multiple facilities, often occurs on campuses of multi-location corporate sites, where buildings built over many years have many different vintages of controls. Some companies provide solutions entirely based on getting data out of these buildings into a common web database for archiving and viewing by operators. This functionality goes beyond that of a simple data aggregator, but the data aggregator function is included in the service provided. An aggregator box is deployed sometimes at each building. In other cases, existing legacy links can be translated in the cloud to a new format and stored in the aggregation system.

Every new building (or control system) requires some amount of engineering today to connect with an aggregator. An engineer must identify the existing systems, their software version, available input and output options, and configure the aggregator to interface to those systems. Often, metadata must be manually tagged as well.

A third case, a box that receives sensing information to make local decisions is common, but it falls outside the main scope of this section of the guide. A lighting system is a common example in which switch inputs, occupancy, and light-level sensors are all used to make lighting control decisions. This kind of aggregator is relatively small and does not provide the multi-system aggregation of primary interest in this guide.

## 12.5 Relative Cost Impacts

The cost of a data aggregation system is dictated by three items: hardware costs, NRE cost for setup, and license or service fees. The NRE costs and license fees tend to dominate the overall costs because data aggregation is primarily a software issue rather than a hardware issue. Each of these cost areas will be addressed as it pertains to signal form conversion, and a speculative discussion about lower cost future is included.

### 12.5.1 Current Cost Impacts

Hardware costs are relatively low and are discussed in more depth in Chapter 11. The same issues apply to network interfaces.

The NRE costs for installing and commissioning these systems tend to be large. In many cases, there is little to no documentation, as described in interdependencies section of this chapter. This information must be extracted from device-reported metadata (e.g., BACnet object properties), paper or electronic documents, the current system setup, technician knowledge, and, if needed, trial-and-error efforts. When the systems are well documented, this process is relatively quick. Standard efforts to solve this problem can take weeks of engineering effort per device.

Because of the closed nature of building controls systems and the high NRE costs of installing an aggregator at a site, vendors tend to charge a high one-time fee or moderate ongoing fees for their services. Some products have a per-instance fee as well as ongoing fees. This cost, sometimes buried in hardware costs, can dominate the overall cost. There are limited examples of open-source solutions to the data aggregator problem<sup>[17, 18]</sup>, and these efforts are not well supported in industry.

System management is the final cost area. Hosting a local server appears to be relatively inexpensive, but there are ongoing costs associated with keeping a system patched and functioning. These costs result from the need for some employee or contractor to either be responsible for the system or the organization may experience downtime, service interruptions, and data loss. Using a cloud services for storage minimizes management and improves some aspects of reliability, but there are ongoing costs associated with using this system as well. In addition, loss of Internet connectivity can result in service interruptions, depending on how the system is set up. All of these factors influence system management costs.

## 12.5.2 Future Low-cost Possibilities

The limiting cost is the lack of documentation and poor standardization of controls that result in high NRE costs for integration into a site. This is primarily a software problem rather than a problem of designing new chips and complex hardware. Reducing the labor required to bring a system up in a new site is the critical issue for making data aggregators cost effective in the market.

Standardized control interfaces, improved documentation, and, eventually, plug-and-play building systems are required to achieve lower costs. To enable plug-and-play systems, two key things must happen related to data aggregators. First, devices need to be able to identify themselves and provide significant metadata on customizations to the system. Second, a flexible, portable driver library needs to be widely available so that a system can use the identification information and select an appropriate driver from an existing library.

## 12.6 Summary/Conclusions

Building network interfaces on data aggregators provide a critical function: they connect building automation networks inside the building to new, IT-connected systems. There are many types of building network interfaces, and all of these networks strive to connect more devices together to pass more data between those devices. A key limiting factor to the use of a data aggregator with building network interfaces is that the links and the data carried on the links are often not well documented. It is time consuming and expensive to build drivers to interface to the many unique systems in existence, and this lack of documentation and drivers limits the use of data aggregators more than any single other factor.

## 12.7 References

1. Cisco, Internetworking Basics, [http://docwiki.cisco.com/wiki/Internetworking\\_Basics](http://docwiki.cisco.com/wiki/Internetworking_Basics)
2. "Information technology -- Open Systems Interconnection -- Basic Reference Model: The Basic Model," ISO/IEC 7498-1:1994
3. R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. Rfc 2616 – hypertext transfer protocol – http/1.1. Technical report.
4. T. Berners-Lee, D. Connolly, Rfc 1866, Hypertext Markup Language - 2.0
5. ASHRAE SSPC 135, [www.bacnet.org](http://www.bacnet.org)
6. <http://www.modbus.org/>

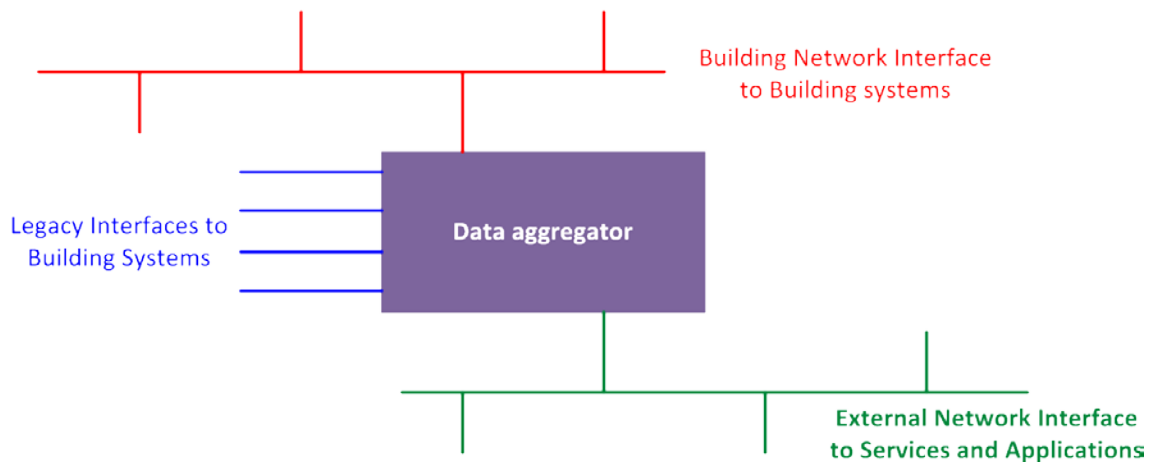
7. Echelon Corp. "LonTalk Protocol Specification,"  
<http://www.enerlon.com/JobAids/Lontalk%20Protocol%20Spec.pdf>
8. Guidelines for Proper Wiring of an RS-485 (TIA/EIA-485-A) Network, TUTORIAL 763, Nov 19, 2001m <http://www.maxim-ic.com/app-notes/index.mvp/id/763>
9. <http://www.hartcomm.org/>
10. IEEE 802.3, <http://www.ieee802.org/3/>
11. IEEE 802.11 WIRELESS LOCAL AREA NETWORKS, <http://www.ieee802.org/11/>
12. IEEE 802.15 WPAN Task Group 4, <http://www.ieee802.org/15/pub/TG4.html>
13. ZigBee™ Alliance, <http://www.zigbee.org/>
14. G. Montenegro, N. Kushalnagar, J. Hui, , and D. Culler. Rfc 4944 – transmission of ipv6 packets over ieee 802.15.4 networks. Technical report.
15. IEEE 802.15 WPAN Low Rate Alternative PHY Task Group 4a,  
<http://www.ieee802.org/15/pub/TG4a.html>
16. <http://www.tridium.com/>
17. Tridium Sedona Framework, <http://www.sedonadev.org/>
18. sMAP 2.0, <http://code.google.com/p/smap-data/>

## 13.0 Signal Data Aggregator: External Network Interface

### 13.1 Introduction

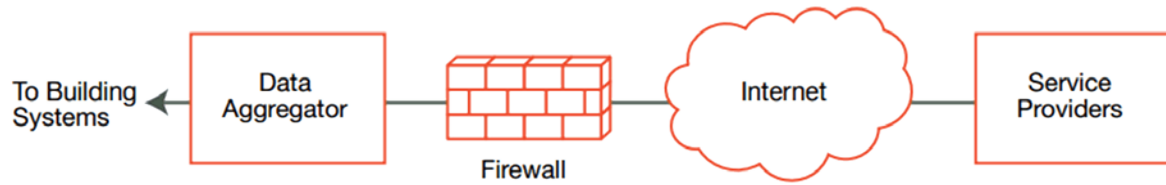
A full introduction to the general topic of data aggregators is included in the introduction to Chapter 11, which provides the necessary background and context on data aggregators needed to understand many data aggregator issues not covered explicitly or in depth in this chapter. Chapter 12 provides additional context on aggregators including some background on networking and communication that can be helpful in understanding this chapter.

A critical function of the data aggregator is to connect to one or more automation, information, or other “external” systems. The external network interface is differentiated from other links the aggregator supports in that it does not connect to traditional building systems. This link provides connectivity to data storage, energy information systems, new building automation systems, or similar tools. Figure 13.1 shows the different interface types. This link is critical to extracting the most value from a data aggregation system, as it provides the link into the modern IT world.



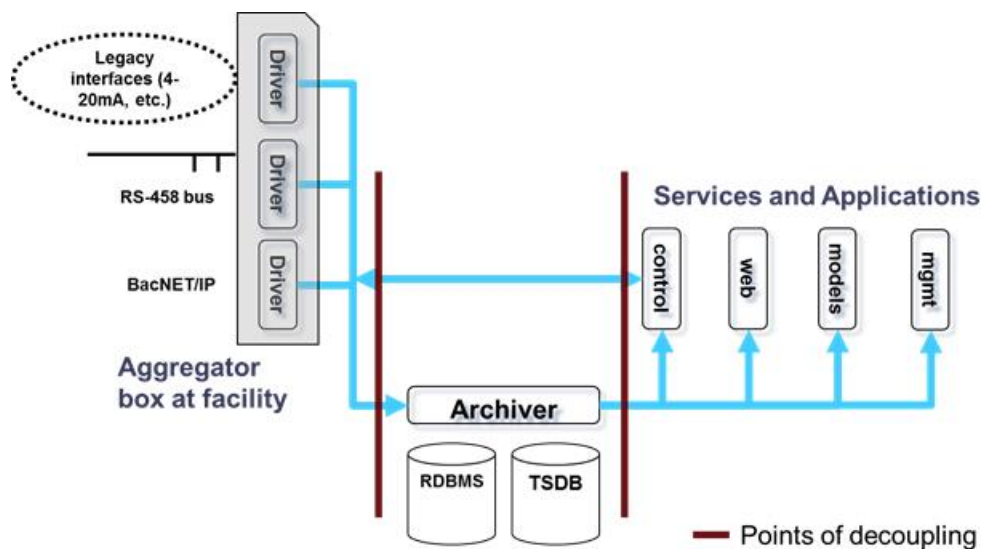
**Figure 13.1.** A Data Aggregator has Three Types of Interfaces: 1) Legacy, 2) Building Network, and 3) External Network. The legacy links include simple analog or single-bit digital links. Building network links provide higher-rate serial data with more advanced features (e.g., BACnet). The external network interface provides connectivity across traditional IT networks to software and services in the building or on the web. The external links are the topic of this chapter.

For the purposes of this chapter, the external interface can be considered at a very abstract level, with the functionality provided being critical. This interface uses standard IP-based protocols over standard media to connect between computers, just like any other IT network. Some features, like crossing firewalls, differ slightly from other IT network connections and will be discussed in this chapter. The term external interface is used in this chapter to avoid confusion between this interface and the building network interfaces to building systems. The external interface is shown in Figure 13.2, extending from the right of the data aggregator through a firewall and the Internet to external services. The external network link does not need to include the Internet, but web services running in the cloud are becoming increasingly popular with service providers.



**Figure 13.2.** A Data Aggregator Located in a Facility is Connected via the Local IT Network Through a Firewall to External Services and Applications. The external network link extends from the right of the data aggregator, and this is the link covered in this chapter.

Figure 13.3 shows a complete aggregator system that includes the interface to building systems, as well as the data archiver and interface to external applications and services. In this chapter, we identify and discuss issues associated with the aggregator external interface, including the archiver.



**Figure 13.3.** A Full Data Aggregation System. The interfaces to the physical systems need to be located at the site, but the archiver (storage) can be located at the site or in the cloud. The services and applications shown on the right are not strictly part of a data aggregator and are not discussed in depth in this chapter. However, the link to these services is discussed.

We, also, discuss the importance of the external interface and the issues it presents for data aggregators. We first identify the technical details associated with the external interface and interoperability challenges. With this technical basis covered, we further detail how this chapter is interdependent on other aspects of the sensor system and, in particular, those characteristics identified in this reference guide. We present common applications and cost impacts in the buildings space to highlight the application of the external interface in data aggregation.

## 13.2 Technical Details

The primary technical issues with data aggregators and the external interfaces are related to connecting across existing IT environments and providing the required data in a timely manner. In most cases, the external interface is a network connection over a LAN or the Internet, but there are examples of

other digital connections in use. The archiving service is often provided by a simple trending application running, but it can be a production database on a local or cloud-based server.

This section covers the external interface including the application programming interface (API) to the aggregator and the structure of the archiver.

### 13.2.1 Application Programming Interfaces

This section will introduce many of the common features required in an API<sup>[1]</sup> for both the aggregator and the archiver. The aggregator typically resides at the facility, while the archiver can reside in the facility (even in the aggregator) or on the Internet. This section introduces a general set of common API paradigms used in data aggregators. This discussion will identify which API structures are useful under different situations.

An API is a specification for how two pieces of software can interact to achieve certain outcomes<sup>[1]</sup>. Often APIs are for software running on the same local system, but network-based APIs are common as well. This second class, APIs for software running on different systems connected by a network, is the one of particular interest for data aggregators. Common example APIs used in the buildings application are Simple Object Access Protocol (SOAP), Open Building Information Exchange (oBIX), and OPC. An emerging open-source API is sMAP. Many custom or proprietary APIs also exist that are based most commonly on HTTP/TCP/IP stacks similar to SOAP.

The SOAP API is a general-purpose interface that has been a popular choice for building systems for a number of reasons<sup>[2]</sup>. It is lightweight and highly flexible while supporting secure transactions. A SOAP system has a device that contains information (server) and another device that requests this information (client). The most common case is SOAP implemented over HTTP, and the client makes an HTTP “get” request containing information about what data are needed. The server replies with an XML file containing the data requested. In a SOAP interface, both ends of the link can serve as the client and the server enabling bi-directional data transfer. Unfortunately, many firewalls do not permit incoming HTTP requests, making this technique difficult across firewalls. It is possible to have the device on the protected side of the firewall send periodic requests to the other device so that all incoming information is in the form of a reply (firewalls allow this type of traffic). Modifications to basic SOAP to allow data to pass across the firewall, such as this one, move away from the basic SOAP model, however, and start to approach other API models. Using a SOAP API does not ensure interoperability even if an aggregator has SOAP support because the high degree of flexibility means that much more than the simple SOAP API must be known to ensure that two systems can interact.

The oBIX<sup>[3]</sup> API is a REpresentational State Transfer (REST) interface that is specifically intended to enable sharing of building monitoring and control information. A RESTful interface is one in which the server web service (different from other server software) does not keep track of individual clients or their state<sup>[4]</sup>. The clients are responsible for updating the server on their state and needs through requests, and the server is responsible only for providing for those needs in the given state of the client. This interface is simpler to implement than a SOAP interface and automatically assumes the model where one side may not be able to receive information from the server without first making a request. RESTful interfaces are built on HTTP just like SOAP, making these two interface types very similar. The oBIX interface provides a normalized representation of both time-stamped data (real-time and historical) and alarm

conditions. This framework also allows the individual control or sense points to be addressed if they are known to ensure no loss of functionality in the normalization process.

The Simple Measurement and Actuation Profile (sMAP) is a RESTful interface similar to oBIX in some ways, but it also provides explicit support for drivers to connect to building systems on building network and legacy interfaces. It also provides a data archiver for both local buffering and long-term remote storage<sup>[5]</sup>. An entire data aggregation software solution is provided by sMAP, making it unique among the API models. This API does provide data normalization, as well as raw data support, and it includes an advanced, comprehensive method of tagging data with metadata, searching based on a variety of tags, and identifying unusual conditions.

The Object Linking and Embedding (OLE) for Process Control (OPC) framework was initially popularized in industrial controls but has been used in buildings as well<sup>[6]</sup>. OPC is intended to exchange real-time data between systems from different manufacturers, and this intention has made it a potential option for data aggregation. One reason OPC has seen some use is that it integrates well with the Microsoft Windows platform and now with the .NET framework. The OPC Data Access specification focuses on streaming real-time data from multiple sources to archiving and human display interfaces, and the Historical Data Access (HDA) specification focuses on archived data. HDA includes interpolation, resampling support to simplify the use of data for display or analytics. The architecture for OPC is based on Distributed Component Object Models (DCOMs) which are a Microsoft proprietary API for data sharing across networks. This method does not operate through firewalls cleanly or even in an awkward way (like SOAP). Therefore, it is only useful if the aggregation occurs all within a particular facility. A different API would be needed for communication outside the facility.

Other APIs are common and are largely customized by vendors for their particular tools. Most data aggregators are provided as part of a larger server in which the aggregator talks with web services developed by the vendor. Therefore, the need for open or standards-based APIs has been limited. General SOAP and REST APIs can be considered custom as well because the specifications are general rather than specific.

### 13.2.2 Archiver Structure

Storing data for later analysis or viewing, often called trending, is an important feature of many data aggregators. Although some aggregators only pass data in real time, it is more common for systems to contain a method for storing some data either locally or in a database elsewhere on the local network or in the cloud. These three storage situations will be considered in this section.

Local storage of data on the aggregator itself is primarily limited by the amount of local storage and the processing power of the aggregator. To optimize the storage space and minimize computation, local data logging usually stores data in a binary file with time-stamp, value pairs. In some cases, more than one value can be associated with a single time stamp. This storage method does not allow for efficient searching (i.e., querying), but rudimentary range queries are possible by searching the archive as a linked list based on time stamp. The data can be retrieved point-by-point or in a stream across a specified time range, and data are often output in a text (e.g., CSV), XML<sup>[7]</sup>, or JSON<sup>[8]</sup> format, with CSV being the most common format for human use and XML and JSON being common for exchange to another system (possibly a larger archiver or an automation system). Local storage typically serves only as a buffer before moving the data to another system. Local storage also can be used to store data that is not

routinely used for any further purpose. In this second case, a circular buffer (where the oldest data are overwritten by newer data) is common, so that a user can retrieve the newest data in the event of a fault or other event.

Storage local to the facility but outside the physical aggregator box is relatively common and enables greater storage while potentially limiting ongoing costs, increasing data security, and providing fast, local access to data. Typically, a local desktop-based computer is used as a server to run vendor-provided software in a standard operating system (usually Windows). This software can connect to data aggregators in the facility and either retrieve data in real time or periodically download data into a local archive. The local archive is sometimes just a simple buffer with limited querying capability, and this is usually the case when only plotting, basic analytics (e.g., percent time on, averages, etc.), and data export to CSV are supported.

In other cases, a database is implemented to enable access to data using queries for different values, fault detection, and other services. The use of a reasonably designed database can provide fast access to local data, and this can be valuable for building data-driven models of building operation to make control decisions or other optimizations. Poor design of the database or exceeding the data management capacity of the system can lead to slow data access, however <sup>[9]</sup>.

The scalability of a single, local server to very large data sets (billions of readings) is poor. A local server also does not require explicit ongoing charges to a cloud provider, and data security is not limited by the policies of a third party. However, the use of a local server adds management costs because someone must maintain or fix this system if it fails. If there is no maintenance and the system is on the local IT network, it may be susceptible to network-based attacks on security holes in the system. The local database can be communicated with using a variety of protocols that are not appropriate for Internet communication, further making it an attractive choice in some cases.

Using a cloud-based data archiver reduces local management costs; shifts security concerns to professional staff, and can provide scalability with data unavailable in local systems. Further, cloud services provide data access anywhere with an Internet connection, and this is not common for local systems.

### **13.2.3 Additional IT Network Considerations**

Connecting to an existing IT network has its own advantages and disadvantages. The goal of the external interface is to get system data and control information between the aggregator and other systems or web applications. Connecting to an existing, Internet-connected IT network is the fastest conceptual way to make this connection. Although conceptually easy, connecting to an existing network can be very challenging because of restrictions set by the facility IT provider. For example, corporate IT rules may prevent a contractor from connecting a device to the IT network, making it difficult for a data aggregator to interface with the corporate network. Buy-in and support from high-level management and/or IT officials can alleviate this issue. In many cases, however, connecting to the IT network may take so much logistical time and effort that it is not cost effective. Providing a dedicated Internet link using either a wired or wireless (e.g., cellular) Internet connection can solve this problem. When using a dedicated connection in this way, however, the potential benefits must be greater than the ongoing cost of the Internet connection.



## 13.3 Characteristic Interdependencies

External interfaces in data aggregators have significant overlap with other areas of this guide. As described in the introduction to Chapter 11, there are many types of connections on data aggregators. The data aggregator's network capabilities are the key reason that it provides so much value.

Data aggregators typically provide two or more types of network connections:

- The connections to the devices of interest in the building. These connections are discussed in Chapter 12.
- The connection to a building automation system, data archiver, or other tool. This connection is discussed in this chapter.

Detailed aspects of understanding which network connection to use and how performance is determined are covered in the chapters on communications and networking in this guide. This chapter only includes the basic information that is needed to understand the aggregator itself.

The missing components of highly successful and low-cost data aggregators with respect to the external interface are associated with standardization and interoperability. The hardware, as discussed in Chapter 11, is available and low-cost. Because of a large number of options and proprietary solutions, it is difficult to have an aggregator unit that connects to third-party systems. As a result, users are forced to buy entire systems from a single vendor. Providing an open interface standard such as oBIX or sMAP enables systems from multiple vendors to be used together, potentially improving performance and reducing costs.

## 13.4 Typical Applications

The external interface is typically used to connect between physical aggregator boxes at the facility and data archiving, analysis, and human interface tools. Using modern IT tools and software, it is possible to create applications quickly and easily to provide a variety of services to facility managers and technicians. The external interface is specifically intended to connect the building systems to these tools so that greater levels of automation, fault detection, data visualization, etc., are all possible, while also using fewer staff hours for site visits.

A common example is EIS. An EIS gathers operational data from a building (often using an aggregator provided by the EIS vendor) and displays trends, baselines, energy analysis, and fault detection services<sup>[10]</sup>. The aggregator provides the physical connection to the building resources while also connecting back to (typically) a cloud service that provides services to users in a web application.

Other examples include controls companies that provide web-based services. These services may provide methods for changing the sequence of operations in controllers that are connected to onsite data aggregators. Displaying trends related to operation, detecting alarms, or anomalous situations also are included sometimes.

## 13.5 Relative Cost Impacts

The cost of a data aggregation system is dictated by three items: hardware costs, NRE costs for setup, and license or service fees. In the case of the external interface, license or ongoing service fees tend to dominate because the other costs are associated with the other interfaces of the system.

The web-based services provided are often the ongoing revenue and profit generation portion of the aggregator service provider's business model. This is appropriate in that it is the services that provide the value rather than the hardware installed itself. Given the fast growth of web applications for every aspect of life and the relative low-cost of development, it is likely that energy-based web applications on building data are going to continue to develop and compete with one another.

In an ideal future, these services will compete with one another based on service quality rather than on the cost of installing the aggregation solution in the facility. This will only occur when the aggregation solutions are commoditized and provide data over well-defined and common interfaces so that any third party can access the data. When this occurs, it will be possible for a service provider to offer incentives to building operators to try out their services, and building operators will be free to move periodically to a new provider of better or cheaper services. This will revolutionize the building controls and information industry, leading to efficiency improvements and cost reductions.

## 13.6 Summary/Conclusions

The external network interface on data aggregators provides a critical function: it connects the aggregator to new, external applications and services like energy information systems, new building automation systems, or other web applications. Combined with the legacy and network interfaces, this allows the aggregator to connect the building physical systems not only together but to applications that reside in the standard IT domain. The data aggregator should include an open API to allow systems to access the aggregated data, and this API should be selected and designed with IT network issues like firewalls in mind.

## 13.7 References

1. API versus Protocol, <http://c2.com/cgi/wiki?ApiVsProtocol>
2. W3C, SOAP Specification, <http://www.w3.org/TR/soap/>
3. Open Building Information Xchange, <http://www.obix.org/>
4. R. T. Fielding. REST: Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation, University of California, Irvine, 2000.
5. sMAP 2.0, <http://code.google.com/p/smap-data/>
6. OPC Foundation, <http://www.opcfoundation.org/>
7. XML Protocol Working Group, <http://www.w3.org/2000/xp/Group/>
8. Introducing JSON, <http://www.json.org/>

9. S. Dawson-Haggerty, "Physical Data Storage", LoCal Retreat June 2011, <http://local.cs.berkeley.edu/files/local-summer-2011-sdh.pptx>
10. J. Granderson, M.A. Piette, G. Ghatikar, P. Price, "Building Energy Information Systems: State of the Technology and User Case Studies," LBNL Report 2899E, <http://eis.lbl.gov/pubs/lbnl-2899e.pdf>

## 14.0 Signal Conversion to Data: Signal Form Conversion

### 14.1 Introduction

The signals provided by sensors must be converted to a usable form of data. In the most simple analog control systems, data conversion may be as simple as amplification of a voltage signal, conversion of voltage to current, or conversion of current to voltage. The voltage or current signal could then be directly fed to an analog control circuit. However, in any advanced control system, signals must be converted to digital form to be processed by a microprocessor-based control system. The bulk of this chapter will discuss the conversion of inherently analog signals, voltages, and currents to digital form (i.e., analog-to-digital conversion), but we will also discuss the conversion of optical, IR, and RF signals to data.

The signals produced by sensors are usually one of three forms: 1) analog voltage, 2) analog current, or 3) digital data. Sensors that output direct digital data, either because they are analog sensors with built in analog-to-digital conversion or they are inherently digital sensors (e.g., on/off, occupied/empty, flow/no-flow, etc.), are already digital in nature and, hence, do not need signal conversion per-se. Instead they need interfaces to manage noise and to convert the output digital signal into a form that can be handled by the digital communication systems.

The method of analog-to-digital conversion depends greatly on the temporal nature of the signal. Nearly all sensors encode information in the amplitude of the analog voltage or current output or can be incorporated into a circuit that creates a changing voltage or current amplitude that is converted to data using a voltage or current analog-to-digital converter. However, some sensors, such as capacitive sensors, can be designed to function as part of a resonance circuit, and the output is instead a changing frequency in which the amplitude itself does not carry information. Other sensors can encode information as pulses in which the timing of voltage or current changes is important, but the magnitude of the changes is not. These frequency and pulse signals are converted to data through frequency counters or timing circuits. Because most sensor output signals are voltage amplitude variations, our focus in this chapter is primarily on voltage analog-to-digital converters.

### 14.2 Analog-to-Digital Converters

The bulk of analog signal to digital conversion is the conversion of analog voltages to digital data, which is accomplished with analog-to-digital converters (ADCs). With the exception of the integrating ADC, ADCs are sampling devices that take a series of snapshots of the signal and convert the snapshots to data in various ways. The amplitude of the signal is sampled at a series of regular time steps, and the sampled signal amplitude is converted to a digital number.

In the case of most building sensors, the frequency of output signals is quite low, measured in Hz, tens of Hz, or at most in the kHz range, with the exception being ultrasonic signals produced by some occupancy sensors and video signals produced by video cameras that may be used within building controls. The raw ultrasonic or video voltage signals are rarely fed to any building control system. Ultrasonic signals are usually converted to digital occupancy information as part of the occupancy sensor, and video signals usually are converted to digital form within the camera system. Optical, IR, and RF

signals are inherently extremely high frequency (MHz – THz), but in most cases, it is a low frequency peak amplitude variation of the signal rather than the actual electromagnetic amplitude that is desired and converted to data.

The converter SNR is the ratio of the highest amplitude signal that can be digitized to the lowest signal level that can be digitized by the ADC (i.e., to the maximum quantization error). In practice, the SNR is usually several dB less than the theoretical maximum. For high SNR devices, the SNR usually is limited by the noise floor of the interface electronics and not the quantization error of the ADC.

When ADCs were first developed and high resolution sampling ADCs were expensive, ensuring that the converter system worked at close to the theoretical SNR was extremely important. To randomize the quantization error, which allows performance to be enhanced through signal averaging, low-level random noise was often added to the analog signal before digitization<sup>[34]</sup>. This process, known as dithering, has been largely eliminated in low-speed data conversion by the availability of extremely high-resolution, low-cost ADCs.

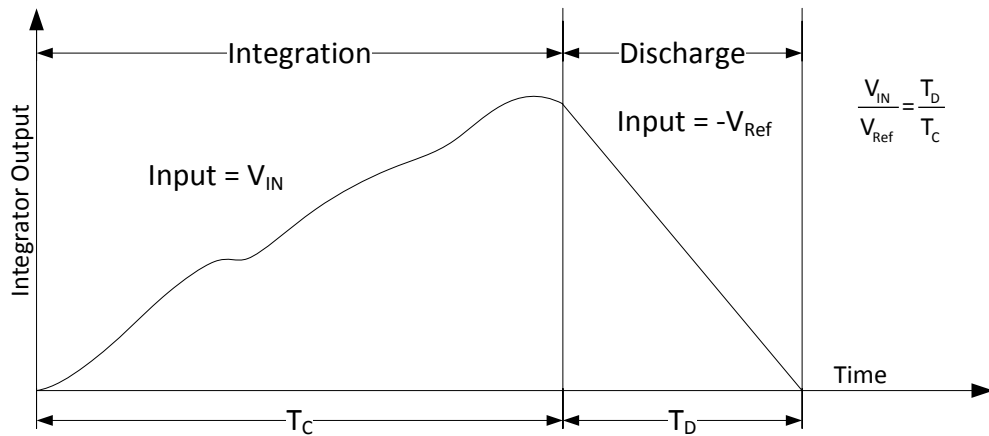
The following sections review the main types of ADCs (i.e., integrating, sigma-delta modulation [SDM], flash, successive approximation) available on the market today.

## 14.2.1 Integrating ADC

### 14.2.1.1 Technical Details

The oldest ADC technology is the integrating ADC type, which is useful for medium-resolution digitization of low-bandwidth signals<sup>[18]</sup>. The component count, microchip die size, and power usage of the integrating ADC is fairly low, and because of the small size and low power usage, integrating ADCs are quite often found in building sensor instrumentation. Integrating ADCs fall into one of three categories: 1) single-slope, 2) dual-slope, or 3) multi-slope. The most popular type is the dual-slope ADC.

In basic dual-slope integrating ADCs the signal input is integrated for a fixed time,  $T_C$ , at which point the integrator input is switched to a fixed reference voltage ( $V_{Ref}$  or  $-V_{Ref}$ ) until the integrating capacitor voltage returns to zero, which takes a variable amount of time,  $T_D$ . The input voltage is then found as  $V_{IN}/V_{Ref} = T_D/T_C$ . The actual clock rate or integrator capacitor values need not be known to calculate  $V_{IN}$ . A plot of the integrator output is shown for two different input values  $V_{IN}$  and  $-V_{Ref}$  in Figure 14.1.



**Figure 14.1.** Output of the Integrator in a Dual-Slope ADC

In an integrating ADC, the sampling resolution is related to the integration time, so medium-to-high resolution digitization is slow. However, the integration of the input naturally provides a low pass filtering of the input to reject noise and, in particular, frequencies that are multiples of  $1/T_C$ , are integrated out. Thus, if the integration period is a multiple of  $1/(60 \text{ Hz})$ , 60, and 120 Hz, line noise will be naturally filtered by the system <sup>[17]</sup>.

#### 14.2.1.2 Characteristic Interdependencies

Integrating ADCs are available in 10 bit to 15 bit resolution with conversion rates of 1 to 100 SPS. Integrating ADCs have inherently very few components, and those components are fairly low power, but the integrating ADC typically requires a precision resistor and capacitor for integration and another precision capacitor for auto-zeroing. The capacitors must have extremely low dielectric absorption and circuit board layout details are critical. The performance of the capacitors is one of the primary limitations to accuracy and resolution of the integrating ADC. While these external capacitors are usually fairly large and expensive, no anti-aliasing filters are required by the system and the integrating nature of the device signal naturally filters out signals with frequencies  $f=N/T_C$  where  $N$  is an integer and  $T_C$  is the integration time of the input signal <sup>[17]</sup>.

#### 14.2.1.3 Typical Applications

Integrating ADCs have been used for decades as the front end to digital panel meters, but are quickly being supplanted by sigma-delta devices. Because resolution is dependent on the integration time and not on fabricated resistor networks or decimating filters, the resolution of the device and, hence, the power used can be adjusted as needed. If rejection of a harmonic external noise is required, integrating ADCs offer natural notch filtering, which eliminates the need for additional components.

Recently, modified integrating ADC architectures have been proposed. This design allows for single-chip designs with low power usage, but the performance characteristics are no better than high performance sigma-delta modulation converters <sup>[32]</sup>.

#### **14.2.1.4 Relative Costs**

Typical integrating ADCs require precision, low dielectric capacitors that are fairly expensive, and as a result, integrating ADCs cannot be made extremely inexpensive. The typical integrating ADC, as used in a panel meter, has a sample rate of 10 to 100 SPS, a resolution of 10 to 15 bits, and costs ranging from \$2 to \$10. The required low dielectric absorption capacitors are relatively expensive, costing upwards of \$0.50 each.

### **14.2.2 Sigma-Delta Modulation ADC**

#### **14.2.2.1 Technical Details**

The SDM ADC is one of the oldest ADC technologies. It began as an offshoot of delta-modulation transmission methods where only changes to the signal were transmitted to reduce the required transmission bandwidth<sup>[15]</sup>. For a long time, implementation of SDM was limited as ADC development was focused on successive approximation register and flash converters. However, with improvements in semiconductor fabrication technology, SDM has found resurgence and now most advances in the ADC industry (namely, increased resolution and decreased power) start with SDM converters. The SDM converter consists of two main components: a modulator and a digital filter.

SDM converters offer the ability to provide high resolution, with low noise, at low cost, and with low power usage.

#### **14.2.2.2 Characteristic Interdependencies**

SDM converters have no precision components and are inherently linear and monotonic. Because of the oversampling nature of the ADC, simple resistor-capacitor low-pass filters without precision components can be used as anti-alias filters. Alternatively, if there is no high-frequency noise or component in the signal to be digitized, anti-alias filters often can be omitted altogether. Additionally, simple resistive sensors and thermocouples can often be interfaced directly to the SDM converter without need for additional interface or conditioning circuitry. Many modern SDM packages include circuitry for self-calibration and are easily interfaced to microcontrollers through two-wire serial interfaces. SDM converters are available in resolutions of 16 bits with single input sampling rates as high as 10 MSPS, 24 bits at 4 MSPS and resolutions as high as 31 bits at a rate of 250 SPS.

However, SDM converters do have some drawbacks. In particular, multiplexing of several signals to a single SDM converter can be problematic. Because of the “memory” of the previous signal in the feedback loop and the decimation of the digital filter, several input samples are used for each output sample generated. As a result, all the information about the old input must be flushed from the system before the output from the ADC truly represents the value from the sampled input. This delay reduces the effective sample rate of a multiplex system<sup>[15]</sup>. Most manufacturers have developed multichannel sigma-delta ADC chipsets that automatically handle the timing and only output valid digital outputs. In addition, because of the low cost, small size, and low power usage of SDM, multiple SDM converters can often be used in place of a single multiplexed successive approximation register (SAR) or integrating ADC.

Because the components for SDM can be completely implemented using complementary metal-oxide semiconductor CMOS technology, SDM can be made physically small with extremely low power requirements. Single-channel converters are available with current power draws of less than 300  $\mu\text{W}$  and standby power draws of less than 3  $\mu\text{W}$ . Six-channel converters are available with current draws of less than 1200  $\mu\text{W}$  and standby power draws of less than 3  $\mu\text{W}$ , including all interface and clock circuitry.

SDM ADCs are becoming the dominant technology used in the industry, especially when multiplexing of many channels is not necessary. SDM should probably be considered before all other forms of ADC, especially when looking to develop low power, high resolution solutions for low sample rates.

### 14.2.2.3 Relative Cost Impacts

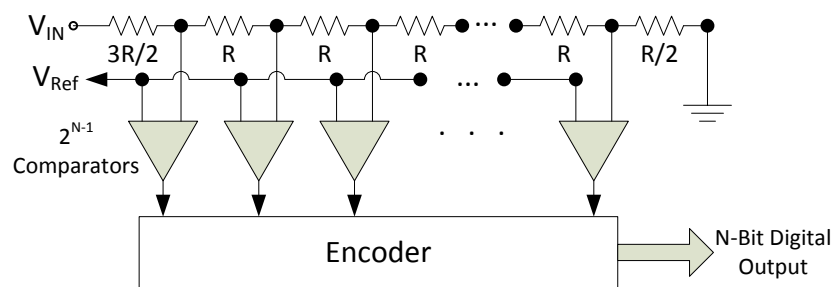
SDM ADC costs generally are the lowest of all the technologies, partially because new designs are being actively developed that utilize the newest low cost manufacturing techniques.

SDM converters with 16-bit resolution, a sample rate of 120 SPS, and all required interface components on chip are available with power draws of less than 500  $\mu\text{W}$ , in a 3 mm square package for \$2.00 each. SDM converters with 6 channels, 16-bit resolution, and 470 SPS sample rates, all required interface circuitry, and a 8 mm x 5mm chip size are available for \$4.00 each.

## 14.2.3 Direct or Flash ADC

### 14.2.3.1 Technical Details

The flash ADC has perhaps the simplest conceptual architecture of all ADCs. As shown in Figure 14.2, the input voltage to a flash ADC is compared to a number of different voltages all at the same time. The output of the comparators is the digital data itself<sup>[18]</sup>. This architecture allows for extremely fast data-conversion rates.



**Figure 14.2.** Diagram of a Flash ADC System

An N-bit flash ADC requires  $2^N$  comparators and precision resistors and, thus, is only used in situations when high speed and limited resolution ( $N \leq 12$ ) are necessary. Within building systems, the high speed of flash ADCs is rarely needed. A variation of the basic flash ADC is the pipeline ADC which uses several stages of low-resolution, but highly accurate flash ADCs<sup>[27]</sup>.



### 14.2.3.2 Characteristic Interdependencies

Because of the required number of comparators and precision resistors, flash ADC chips are fairly large and use a fairly high amount of power (tens of mW to several W). Flash ADCs also require anti-aliasing filters, which typically are not included in the flash ADC chipset. Because flash ADCs are usually used at high sample rates, the low pass anti-alias filters have fairly high corner frequencies and can thus be frequently implemented entirely with passive components<sup>[38]</sup>.

Flash ADCs are available with sample rates beyond 1 GSPS, and specialized pipelined flash ADC are available with sample rates up to 40 GSPS and with resolutions of 10 to 12 bits.

### 14.2.3.3 Typical Applications

Flash ADCs are typically used for digitizing RF and video signals. Their use in building sensor systems is limited to communication systems or video sensors. Most modern camera chipsets that might be used within buildings will utilize CMOS or charge-coupled device (CCD) imaging sensors that have built in analog-to-digital converters so there is less and less need to digitize video signals.

### 14.2.3.4 Relative Costs

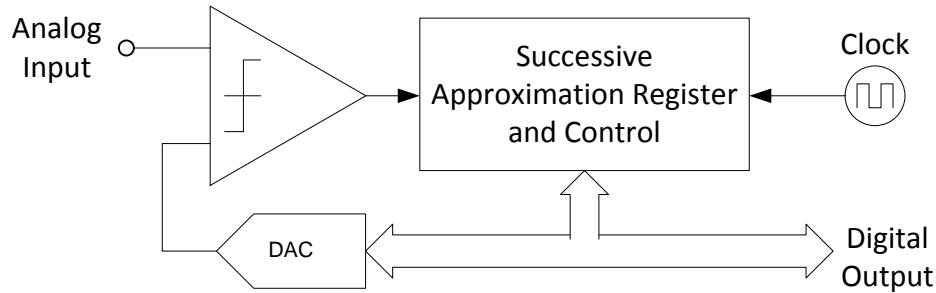
Flash ADCs have many on-chip precision comparators and, as a result, are expensive, especially for high-resolution converters. While 8-bit flash converters at several MSPS are available for \$5.00 to \$10.00, high-speed 12-bit converters cost considerably more, with prices upwards of \$100.00 converters for very high sample rates

## 14.2.4 Successive Approximation Register ADC

### 14.2.4.1 Technical Details

SAR ADCs are used where medium-to-high resolution is needed but extremely high speed is not. At one time, SAR ADCs were dominant in the ADC industry; however, they are being supplanted by SDM ADCs because SDM can provide increased accuracy at reduced cost. Latency problems make using SDM ADCs more difficult in multichannel systems so SAR ADCs often are still used for multichannel digitization.

SARs work by comparing a sample of the input voltage to a precision voltage generated by a high-performance digital-to-analog convertor (DAC) as shown in Figure 14.3. The value of the DAC is increased until the comparator is tripped, indicating that the DAC voltage matches the input voltage. A SAR often employs a sample-and-hold amplifier to maintain the voltage of the input while the comparison process is taking place.



**Figure 14.3.** Basic SAR ADC Architecture

#### 14.2.4.2 Characteristic Interdependencies

SAR ADCs have been used as the main general purpose converter technology for the last 30 to 40 years. They can be designed to be physically small with low power usage, especially for low sample-rate applications. A precision digital-to-analog converter is required by the ADC, but this can be manufactured and placed on chip and mass produced at low cost.

SAR ADCs require anti-aliasing filters and many are unipolar, so level shifting circuitry is also required. These required chips are sometimes available as part of an interface chipset or are incorporated directly into the ADC package.

SAR converters are available in resolutions up to 24 bits and sampling rates to 40 MSPS at lower resolutions. SARs generate a digital output for each input sample and are thus easily multiplexed. For applications in which many sensors are multiplexed to a single ADC, especially where medium to high sample rates are required, SAR converters should be considered, but for single-channel conversion, SDM converters are a good alternative.

#### 14.2.4.3 Typical Applications

Because of their low cost, modest power requirements and flexibility, SARs are still popular for use in low-to-medium speed, medium-resolution and some low-speed, high-resolution situations, especially when input signals are multiplexed.

Complete SAR chipsets that incorporate amplifiers, multiplexers, and anti-aliasing filters are available from many manufacturers. SARs should be considered for building sensors, especially for situations where the counts are high and converter counts are low.

#### 14.2.4.4 Relative Costs

Basic SAR converters are available for less than \$1.00 per unit, and low-power, multichannel chipsets with integrated signal conditioning are available for \$10.00 to \$20.00. SAR converters with 16-bit resolution, 100 kSPS, and power draws of less than 1 mW are available for under \$10.00.

## 14.2.5 ADC Summary

A summary of the tradeoffs of current ADC technologies is shown below in Table 14.1. Except for applications needing very high frequency sampling rates, SAR or SDM technology are the best performing, lowest cost, and lowest power. They should be considered first for the conversion of building sensor voltage signals to digital data.

For a more complete review of ADC technology and more details related to actual ADC implementation, the reader is referred to the excellent textbooks on data acquisition including those authored by Kester (2005), and Phillips (2006), as well as the application notes available from the ADC chipset manufacturers such as Analog Devices, Texas Instruments, Linear Technology, Maxim, Microchip, and National Semiconductor.

**Table 14.1.** Summary of ADC Characteristics

Characteristic	Integrating	Flash	SAR	SDM
Resolution (bits)	8-16	4-12	10-24	16-30
Max Sample Rate	100 SPS	3 GSPS	40 MSPS	10 MSPS
Requires Anti-aliasing filter	N	Y	Y	Y <sup>1</sup>
Easily Multiplexed	Y	Y	Y	N
Size	Large <sup>2</sup>	Large	Small	Small
Cost	Medium <sup>2</sup>	High	Low	Low
Power	Low	High	Low	Low

<sup>1</sup>The anti-aliasing filter of an SDM converter is quite relaxed because of oversampling and can often be eliminated altogether.

<sup>2</sup>The low dielectric precision capacitor required by the integrating ADC is both fairly large and costly and so the integrating ADC package is fairly large and of medium cost.

## 14.3 Current-to-Digital Conversion

Some sensors, such as photodiodes, inherently generate current outputs. Other sensors, such as humidity sensors, change a resistance or capacitance that is most easily encoded as a current output. In most cases, designers convert current to voltage using integrator circuits or precision resistors and then off-the-shelf ADCs for digitization. However, several manufacturers offer direct current-to-digital converter circuits. These circuits usually incorporate either precision integrators or precision resistors as part of the frontend circuitry to convert current to voltage and then convert voltage to data using an ADC<sup>[18]</sup>. Older current-to-digital converters use SARs as the ADC but newer designs use SDMs. Very low power, advanced current-mode ADC devices are being designed that eliminate the need for an on-chip DAC<sup>[1]</sup>.

Some manufacturers are producing high-channel-count, high-resolution current-to-digital conversion chipsets specifically for imaging sensors. Current-to-digital converters with 50+ channels, 20+ bit resolutions, and power consumption of less than 5 mW per channel and with chip sizes of 1 cm x 1 cm are now available at a cost of about \$100.00 per chip.

## 14.4 Capacitance-to-Digital Conversion

Another specialized form of ADC is the capacitance-to-digital converter (CDC). Capacitive sensors are found throughout buildings, and are used as pressure sensors, position sensors, humidity sensors, etc.

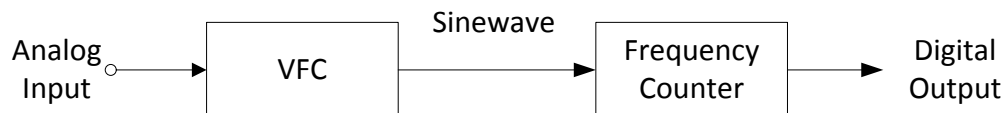
One method of converting the changing sensor capacitance to data is to connect the sensor capacitor in series with a reference capacitor across a reference voltage and then digitize the voltage across the two capacitors and use the ratio of voltages to determine the capacitance of the sensor. These CDCs typically use SDMs<sup>[2]</sup>. A second technique involves making the capacitive sensor part of an oscillator circuit, and the frequency of the oscillator output is converted to digital information through a standard frequency counter<sup>[36]</sup>. A third technique integrates the capacitive sensor into a switched capacitor network<sup>[11]</sup>.

CDCs with 12-bit resolution, 100 SPS sample rates, and less than 300  $\mu\text{W}$  power draw are available for less than \$1.50 per chip. 14 channel converters with 16-bit resolution and power draws of less than 300  $\mu\text{W}$  are available for less than \$2.00 per chip.

### 14.4.1 Voltage-to-Frequency Converters (VFCs)

Instead of being digitized directly, voltage can be converted to a frequency with a voltage-to-frequency converter (VFC). This was commonly done when the analog signal had to be transmitted a long distance, especially in a noisy environment. A VFC is essentially an oscillator with a frequency that is proportional to a control voltage. The analog signal is digitized by simply finding the frequency of the oscillator output with a counter as shown in Figure 14.4. The output of the frequency counter is the average of the frequency over the time it takes to convert, and thus, an integrating digitizer essentially is a VFC and a frequency counter.

VFCs can be made fairly small, fairly low power, and fairly inexpensively. VFCs are monotonic and can be made linear up to a resolution of about 14 bits. The time it takes to convert the signal to digital is related to the accuracy of frequency counting. Low-level voltages generate very low frequency (Hz to 10s of Hz) which can take several seconds to accurately count<sup>[18]</sup>.



**Figure 14.4.** Analog-to-Digital Conversion using a Voltage-to-Frequency Converter and a Frequency Counter

With the advent of the SDM ADC, the VFC has little use as an ADC. If the analog signal itself must be transmitted over a long distance or in a noisy environment, the VFC can still be useful, but given the size and power requirements of current SDM converters, it is better, in most cases, to convert the voltage to digital form before transmission. Because the SDMs use new manufacturing technologies, they can be made smaller, with lower power usage and higher resolution and at a lower cost than a VFC.

## 14.5 Frequency-to-Digital Conversion and Frequency Counters

A frequency-to-digital converter, or frequency counter, is a circuit that will determine the frequency or period of an analog sine wave or digital pulse train. Frequency counters are essential parts of VFC-based ADCs as well as the method of converting frequency output sensors or pulse width modulation output sensors to digital data. In buildings, equipment rotation and position sensors often will generate

pulse trains in which the signal is encoded as a pulse width or as a frequency of pulses. Note that, with a frequency-to-digital converter, it is only the signal frequency or the pulse train period that is measured, and the actual amplitude of the sine wave or pulse train is not important as long as the frequency of the sine wave or the period of the pulses can be determined. This method of encoding a signal and digitization is highly immune to electrical noise <sup>[18]</sup>.

Frequency counters usually work by one of two mechanisms: counting or timing. A counting converter will count the number of zero crossings (for sine waves) or number of transitions (pulse train),  $N$ , in a fixed amount of time  $T_C$  and then  $f = N/(2T_C)$ . For high accuracy many zero crossing or transitions should be counted and so the conversion time for low frequency or long period signals can be high. Alternatively, a timing converter determines the amount of time necessary,  $T$ , for a fixed number of zero crossings or transitions,  $M$ , to occur and then  $f = M/(2T)$ . The details of both these methods are well described in the literature <sup>[12]</sup>.

## 14.6 Optical Signal Conversion

Analog optical signals are not common within building systems. In some harsh industrial environments, analog optical signals are transmitted via fiber optic cable because it is not affected by electromagnetic noise and conversion to data is necessary.

Electrical analog-to-digital conversion of optical signals can be done by converting the optical signal to an electrical signal with a photodetector such as a photodiode, photo-resistor, or CCD and then digitizing the electrical signal <sup>[14]</sup>. In high-speed optical communication systems, the high-frequency electrical signal can be digitized using an interleaved architecture with extremely high sample rates (tens of GS/s), which requires high power and expensive components and has limited (usually 2 or 3 bit) resolution <sup>[29]</sup>.

Alternately, one can utilize photonic time stretching in which a high-frequency electrical signal is used to modulate the optical source and then a dispersive optical medium is used to stretch the signal in time, slowing and stretching the high frequency signal to a frequency at which standard converters can be used <sup>[9]</sup>. More recently, an all-optical variation has been demonstrated <sup>[19, 5]</sup>. Because they are inherently passive, non-linear optical systems could be extremely low power, but they are physically quite large and extremely expensive.

## 14.7 Infrared Signal Conversion

Infrared (IR) sensors are used in a variety of ways in buildings, from sensing of solar radiation (pyrometers), to occupancy detectors (passive IR), to temperature sensing (thermal imaging). In these applications, analog infrared signals often need to be converted to digital data. IR signals cover an extremely wide range of wavelengths of electromagnetic radiation and the technology to convert these signals to data depends upon the wavelength <sup>[33]</sup>. The definitions of near-, mid-, and far-infrared in this chapter follow the ISO 20473 standard <sup>[13]</sup>.

Near infrared (NIR) signals have wavelengths just longer than the visible spectrum (0.75 to 3  $\mu\text{m}$ ), are commonly used in telecommunication systems, and are the basis of most night vision devices. NIR signals are usually digitized using the same optoelectronic technology as other optical signals. Mid-

infrared (MIR) signals have longer wavelengths (3 to 50  $\mu\text{m}$ ) and are commonly known as the “thermal infrared.” The short end of the MIR spectrum can be detected using standard photodiodes but the longer end of the spectrum requires specialty photodiodes or thermal imaging with bolometers, thermopiles, or pyroelectric detectors. Far-infrared (FIR) signals have extremely long wavelengths (50 to 1000  $\mu\text{m}$ ). A discussion of IR signal detectors themselves is beyond the scope of this chapter.

## 14.8 RF Signal Conversion

RF signals are the basis of most wireless communication systems and are used with building sensors for communication. These signals are already digital data and do not need conversion. Analog RF signals themselves are rarely output by building sensors. RF signals are sometimes emitted by faulty electronic equipment (e.g., electronic lighting ballasts, AC inverters, switching power supplies, AC motors, etc.), and the detection of these RF signals can be used for fault detection and diagnostics. RF detectors designed to detect Wi-Fi communication devices have been used as occupancy sensors<sup>[26]</sup> and in RF identification (RFID) systems used for occupancy detectors<sup>[23]</sup>. RF signals have even been used as part of an active probe sensor to determine information about building construction<sup>[21]</sup>.

The design of RF detection and signal conditioning is beyond the scope of this chapter. Readers are referred to other chapters in this review or to one of the many excellent introductory texts on the subject including those authored by Bowick [2007] or Glover [2010].

## 14.9 Signal Conditioning

Most sensor signal outputs require conditioning before conversion to data. Among the more common components are multiplexers, sample-and-hold amplifiers (SHA), gain circuits, filters, and voltage references.

SHAs are a critical part of most ADC systems. An SHA captures an analog signal and holds that value during the conversion process. In multichannel systems, SHAs are required if sampling of the multiple channels must occur at the same time. In most cases, SHAs are part of the ADC itself or are part of an ADC interface chipset.

Multiplexers are components that control and select a transmission path for an analog signal. These devices are used in multichannel data acquisition systems so that a single ADC can be used to digitize multiple signal channels. Most multiplexers are CMOS-based devices although for the very high speed of video or RF signals, higher power analog switches utilizing junction field effect transistors can be used. Multiplexers are commonly found in ADC interface chipsets, or as part of a multichannel ADC itself.

Amplifier (or gain) circuits are used to increase or reduce the level of a sensor signal to better match the signal range to the input range of the ADC, to increase the signal level above external noise, or both. Signal levels are typically decreased using precision resistor dividers and increased using op-amp or transistor amplifiers. Gain circuits are commonly found in ADC interface circuitry or built into the ADC chipset itself. The selection of amplifiers is critical to the performance of signal conditioning and is discussed further in Chapter 16.

Level shifters are special amplifier circuits that are used to shift the DC component of a signal. These are typically used to convert a bipolar AC signal (i.e., a signal having positive and negative voltages) to a signal having only positive voltages, allowing the ADC circuitry to be simplified to deal only with a unipolar signal. Such a circuit is commonly combined with an amplifier circuit <sup>[18]</sup>.

Filters are used to separate high, medium, and low frequencies and to pass only the appropriate frequencies from the sensors to the ADCs. High pass filters are usually used to block DC voltages that are often used to power the sensor. Low pass filters are used to eliminate high frequency noise that is not part of the signal being converted to data. They are also used as anti-aliasing filters for the ADC systems themselves. Bandpass filters are used to only allow a limited range of frequency to the converter circuit. They are used primarily for noise rejection when the sensor signal is an amplitude modulated sine wave or the sensors are resonant devices that output signals in a very narrow frequency range. Bandstop filters are used to block certain frequencies, usually harmonically related, and to reject harmonic noise such as 60/120/180 Hz noise that arises from the building electrical system and electrical equipment.

Filters for ADC signal conditioning usually are implemented using op-amp or transistor-based resistor-capacitor circuits or passive inductor-capacitor circuits <sup>[38]</sup>, or increasingly with switched-capacitor circuits <sup>[25]</sup>. More details about analog and digital filters can be found in Chapter 16. Further information about signal conditioning and analog circuit design can be found in the texts authored by Pallas-Areny [2000], Jung [2005], and Kester [2005] as well as the excellent application notes available from the ADC manufacturers.

Nearly all ADCs work by comparing the input voltage to another known voltage (i.e., a voltage reference). As a result, most ADCs need a stable and accurate precision voltage references for comparison. Integrating ADCs are inherently ratiometric, and a precision voltage source is not necessary.

Voltage references often are built in to some of the more advanced ADC chipsets or ADC interface circuitry, but for some of the lower cost units, external voltage references are required. A typical voltage reference uses zener diodes, bandgap references, or XFET™ (eXtra implanted Field-Effect Transistor). Among the important properties are stability (i.e., ability to maintain a constant voltage throughout its lifetime), low thermal drift (i.e., ability to maintain a constant voltage through temperature changes), accuracy, low power draw, and high output current capability <sup>[18]</sup>. While many electronic components can be shut down during idle periods to reduce power usage, it can take a good deal of time for voltage reference outputs to stabilize, so these components are often left on so low power draw can be extremely important.

## 14.10 References

1. Agarwal, A., Y. B. Kim, and S. Sonkusale (2005), Low power current mode ADC for CMOS sensor IC, *ISCAS 2005*, 584–587 Vol. 1.
2. Bingesser, M., T. Loeliger, W. Hinn, J. Hauer, and S. Modl (2008), Low-Noise Sigma-Delta Capacitance-to-Digital Converter for Sub-pF Capacitive Sensors with Integrated Dielectric Loss Measurement, *Design, Automation and Test in Europe, 2008*. 868–872.
3. Boser, B. E., and B. A. Wooley (1988), The design of sigma-delta modulation analog-to-digital converters, *IEEE Journal of Solid State Circuits*, 23(6), 1298–1308.
4. Bowick, C., J. Blyler, and C. Ajluni (2007), *RF circuit design*, 2nd ed., IEEE.

5. Fard, a. M., B. Buckley, S. Zlatanovic, C.-S. Brès, S. Radic, and B. Jalali (2012), All-optical time-stretch digitizer, *Applied Physics Letters*, 101(5), 051113.
6. Garg, V., and N. K. Bansal (2000), Smart occupancy sensors to reduce energy consumption, *Energy and Buildings*, 32(1), 81–87.
7. Graf, A., M. Arndt, M. Sauer, and G. Gerlach (2007), Review of micromachined thermopiles for infrared detection, *Measurement Science and Technology*, 18(7), R59.
8. Glover, B., and H. Bhatt (2006), *RFID Essentials*, O'Reilly Media, Inc.
9. Han, Y., and B. Jalali (2003), Photonic Time-Stretched Analog-to-Digital Converter : Fundamental Concepts and Practical Considerations, *Journal of Lightwave Technology*, 21(12), 3085–3103.
10. Hauser, M. W. (1991), Principles of oversampling A/D conversion, *J. Audio Eng. Soc.*, 39(1/2), 3–26.
11. Heidary, A., and G. C. Meijer (2007), An integrated switched-capacitor front-end for capacitive sensors with a wide dynamic range, *Solid State Circuits 2007. ESSCIRC 2007. 33rd European*, 404–407.
12. Hewlett Packard (1997), *HP App Note 200: Fundamentals of the Electronic Counters*.
13. ISO (2007), *ISO 20473:2007 Optics and photonics -- Spectral bands*, ISO.
14. Kasap, S. O. (2001), *Optoelectronics and photonics: principles and practices*, Prentice Hall.
15. Kester, W. (2008), ADC Architectures III: Sigma-Delta ADC Basics, *Analog Devices MT-022*.
16. Kester, W. (2008a), Understand SNID, ENOB, SNR, THD, THD+N and SFDR so You Don't Get Lost in the Noise Floor, *Analog Devices MT-003*, pp. 1–9.
17. Kester, W., and J. Bryant (2009), ADC Architectures VIII: Integrating ADC, *Analog Devices MT-027*.
18. Kester, W. (Ed.) (2005), *Data conversion handbook*, Newnes.
19. Konishi, T., and H. Goto (2009), All optical analog-to-digital conversion: Principle and recent progress, in *APCC 2009*, pp. 1–4.
20. Kruse, P. W. (2001), *Uncooled thermal imaging: Arrays, systems, and applications*, SPIE.
21. Ivashov, S., V. Razevig, A. Sheyko, I. Vasilyev, A. Zhuravlev, and T. Bechtel (2008), Holographic Subsurface Radar Technique and its Applications, in *12th International Conference on Ground Penetrating Radar*.
22. Jung, W. G. (2005), *Op Amp applications handbook*, Newnes.
23. Li, N., G. Calis, and B. Becerik-Gerber (2012), Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations, *Automation in Construction*, 24, 89–99.
24. Liddiard, K. C. (2004), The active microbolometer: a new concept in infrared detection, in *Proceedings of SPIE*, vol. 5274, p. 227.
25. Liu, M., and M. Liu (2006), *Demystifying switched capacitor circuits*, Newnes.
26. Martani, C., D. Lee, P. Robinson, R. Britter, and C. Ratti (2012), ENERNET: Studying the dynamic relationship between building occupancy and energy consumption, *Energy and Buildings*, 47, 584–591.

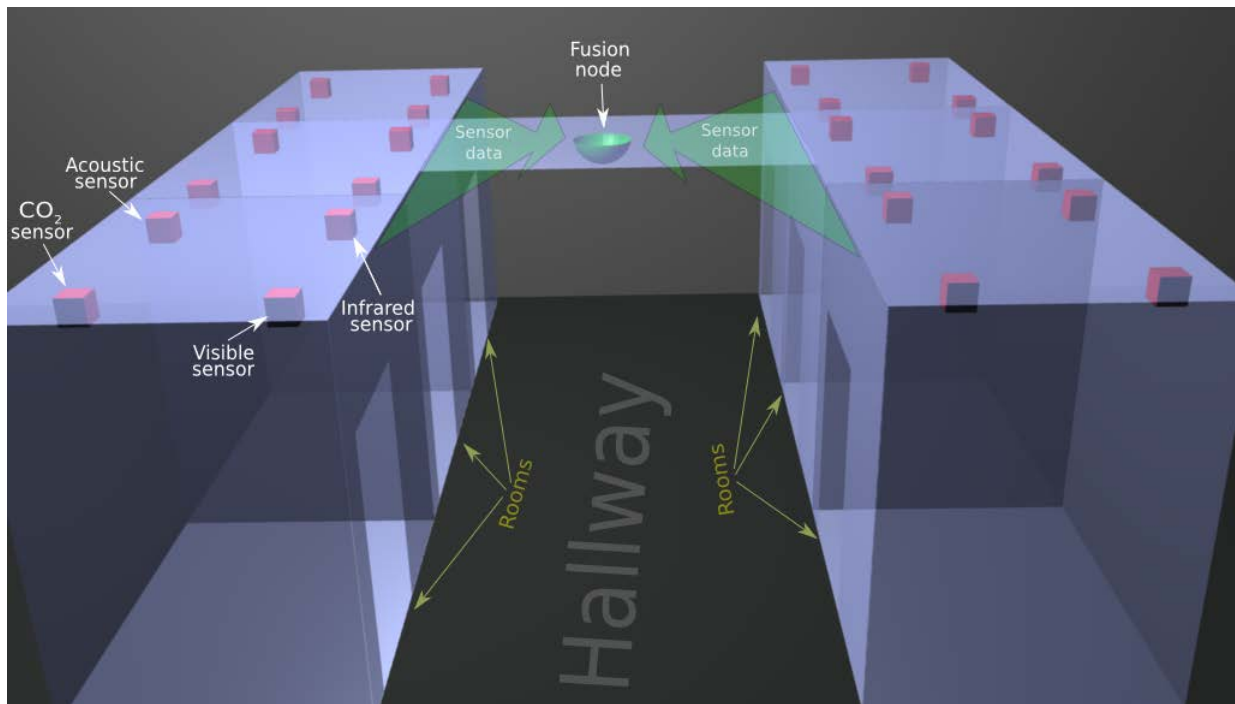


27. Maxim (2001), Maxim Application Note 1023: Understanding Pipelined ADCs.
28. Maxim (2011), Maxim Application Note 2101: Delta-Sigma ADCs replacing Integrating ADCs for Panel Meters.
29. Miyoshi, Y., and K. Ikeda (2007), All-optical analog-to-digital conversion using split-and-delay technique, *Journal of Lightwave Technology*, 25(6), 1339–1347.
30. Pallas-Areny, R., and J. G. Webster (2000), *Sensors & signal conditioning*, 2nd ed., Wiley Interscience.
31. Philips, K., and A. H. M. Van Roermund (2006), *Sigma delta A/D conversion for signal conditioning*, Springer Verlag.
32. Rahiminejad, E., and R. Lotfi (2009), A low-power architecture for integrating analog-to-digital converters, in *ICECS 2009*, pp. 411–414, IEEE.
33. Rogalski, A. (2002), Infrared detectors: an overview, *Infrared Physics & Technology*, 43(3–5), 187–210.
34. Schuchman, L. (1964), Dither Signals and Their Effect on Quantization Noise, *IEEE Transactions on Communications*, 12(4), 162–165.
35. Shannon, C. E. (1998), Communication In The Presence Of Noise, *Proceedings of the IEEE*, 86(2), 447–457.
36. Shibata, H., M. Ito, M. Asakura, and K. Watanabe (1995), A digital hygrometer using a capacitance-to-frequency converter, *Instrumentation and Measurement Technology Conference, 1995. IMTC/95. Proceedings. "Integrating Intelligent Instrumentation and Control".*, IEEE, 100.
37. Walden, R. H. (1999), Analog-to-digital converter survey and analysis, *Selected Areas in Communications, IEEE Journal on*, 17(4), 539–550.
38. Williams, A., and F. J. Taylor (2006), *Electronic Filter Design Handbook*, 4th ed., McGraw-Hill Professional.

## 15.0 Signal Conversion to Data: Data Fusion

### 15.1 Introduction

Data fusion can be loosely defined as the exploitation of information from multiple data sources (sensors) to estimate or infer a set of desired attributes about a set of target entities<sup>[8]</sup>. The sensors can be similar in nature (e.g., temperature sensors) or quite diverse (i.e., multi-modal), such as a combination of temperature sensors, pressure sensors, humidity sensors, cameras, etc. Data collected usually overlap in time and/or space, or are complementary in nature. A natural question to ask is why anyone should care about such fusion of data from multiple sources. A short yet quite general answer is that it provides the following advantages over a single data source scenario: 1) wider spatial/temporal coverage, 2) improved robustness/fault tolerance, and 3) improved estimation of relevant information from raw data. Figure 15.1 illustrates the basic idea of fusing data from multiple sensors.



**Figure 15.1.** A Sensor Fusion Scenario Inside a Building

The figure portrays a scenario of a typical office space with a hallway and multiple rooms on each side. Four different sensors are mounted on the ceiling in each room: 1) a carbon-dioxide sensor, 2) an acoustic sensor, 3) an infrared sensor, and 4) a visible camera. Data from each of these sensors in each room are combined by a data fusion node shown mounted in the ceiling over the hallway. This node combines the data from all the sensors and produces high-level knowledge about this space that can be used to automatically control the indoor environment or by a building manager to assess the operational efficiency of the lighting system and HVAC (heating, ventilation, and air conditioning) in the facility.

### 15.1.1 Wider Spatial/Temporal Coverage

Wider coverage is obtained by aggregating data from individual sensors with each sensor having a relatively limited coverage. For example, suppose a building is monitored through security cameras which are the data sources in this case. One camera may be able to only see the main entrance area but not the rest of the building. So, with this camera alone, one would not have any information about the status of the other areas in the building. However, if a second camera is placed on the opposite corner and the data from both the cameras are combined, one can now monitor the entrance area as well as the opposite end of the building. Adding more cameras will provide surveillance coverage for more and more areas by aggregating or fusing the data from each camera. Thus, fusion of data from multiple sensors can provide wider spatial coverage. The same can be true in terms of temporal coverage as well. A visible camera may only work during the day when there is adequate light available, but will likely not work at night when the building is dark. If an infrared camera with the ability to see during the night is also deployed, data from these two types of cameras can be combined to gain wider temporal coverage. Such spatial and temporal monitoring without any coverage gap is critical in many applications, and data fusion is a key enabler behind such capabilities.

### 15.1.2 Better Robustness/Fault Tolerance

Having more than one source of data (i.e., redundant sources) can provide better robustness or fault tolerance. If a small subset of the sensors happens to fail or malfunction, the data from the other sensors can fill the gap and maintain the flow of critical data. Care needs to be taken, however, to differentiate data streams obtained from malfunctioning sensors from data provided by properly functioning sensors. Otherwise, completely inappropriate decisions may be made in a given situation. Mathematical frameworks for identifying deviant data are quite mature. A simple and often effective technique is a consensus-based approach in which the majority of compatible data are assumed to be the standard and any disagreements from this standard are assumed to be problematic. The underlying assumption is that failures are rare and that most of the sensors are not faulty.

### 15.1.3 Better Estimation of Information from Data

Having a multitude of sensors measuring raw target parameters creates the possibility of inferring a much higher level of information about the target than is usually possible with a single sensor. However, it is important that raw data from multiple sensors are not highly correlated. The more diverse the data, the better the estimation will be.

For example, suppose we are trying to extract the three-dimensional shape of an object from three different cameras. If one camera each is dedicated to capturing the front view, the rear view, and the top view, then combining the images from all three cameras can provide a fairly decent idea about the shape of the object. This is because the images from the three cameras were taken from three very different viewpoints and hence they were *diverse* or fairly *uncorrelated*. However, if all three cameras are placed near each other in front of the object, there is no good way to infer the shape at the top or the back side of the object. In this case, the data from the multiple sensors (cameras) are very similar and not diverse enough to provide more inference than can be determined from a single sensor. Such data are said to be highly *correlated*. So, if a set of sensors is deployed to produce data that are significantly *uncorrelated*, then data fusion across those sensors can greatly improve estimates of the target's attributes of interest.

Various research efforts have shown how using data fusion helps improve estimating target parameters. For example,<sup>[17]</sup> fused optical, as well as inertial, measurements were used to obtain optimal motion estimates of targets. Also,<sup>[19]</sup> research has demonstrated the value of integrating optical and inertial measurement units for navigation.

## 15.2 Technical Details

Sensor fusion can be broadly classified into two categories: 1) *homogeneous* and 2) *heterogeneous*. In the first case, the sensors that produce data to be fused are identical in terms of sensing capabilities and output properties, whereas in the second case, the system consists of diverse kinds of sensors, such as acoustic sensors, imaging sensors, temperature sensors, pressure sensors, etc. In general, fusion across a set of homogeneous sensors is easier than fusion of data from heterogeneous sensors, primarily because data alignment (Section 15.2.5), one of the most fundamental steps in data fusion, is easier in the case of homogeneous sensors.

The data fusion process can also be categorized into three broad classes based on the amount of pre-processing done on the raw data before fusing them: 1) low-level fusion, 2) feature (intermediate)-level fusion, and 3) decision (high)-level fusion.

### 15.2.1 Low-Level Fusion

In low-level fusion, raw data from the sensors are fed to the fusion engine with minimal pre-processing and the engine extracts all high-level information. The volume of data transferred between a data source and the fusion node is relatively high because the sensors can collect quite a large volume of raw data to be fused. This may make this paradigm infeasible where the communication bandwidth is low. Figure 15.2 illustrates this fusion paradigm.

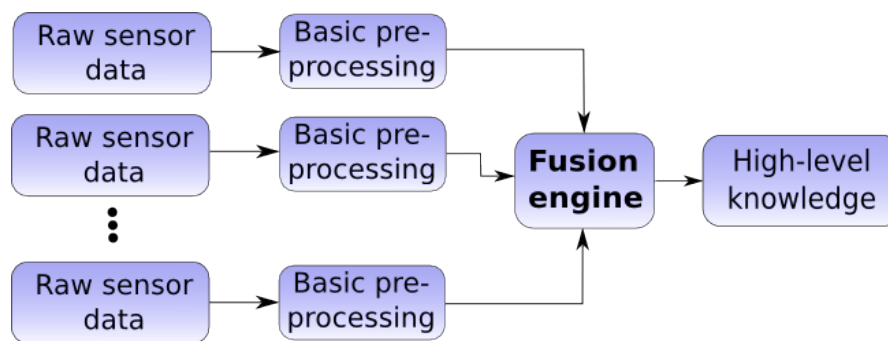
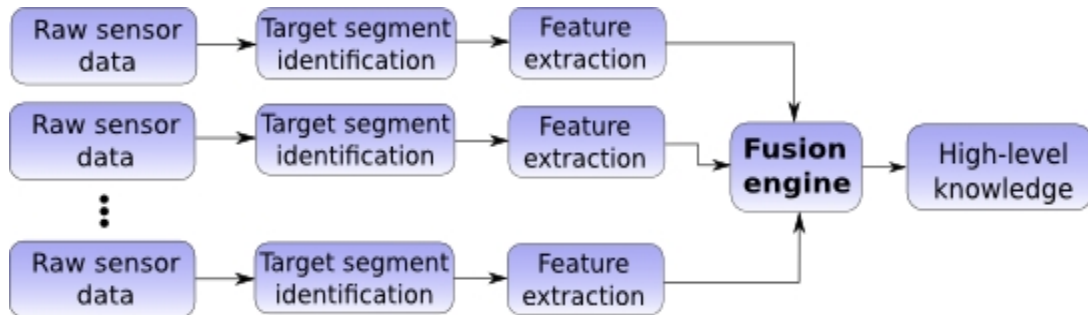


Figure 15.2. Low-Level Data Fusion Scheme

### 15.2.2 Feature-Level Fusion

This scenario can be thought of as an intermediate-level data fusion scheme. Each sensor platform has sufficient computational capability to identify interesting properties corresponding to the desired targets and extract a set of relevant features to be input to the fusion engine. It then extracts high-level information by combining these features from the individual sensors. Figure 15.3 illustrates this fusion paradigm. The paradigm is based on the assumption that it is possible to identify and segment the targets in each sensor data on its own without any input from any other sensor. The extracted features generally

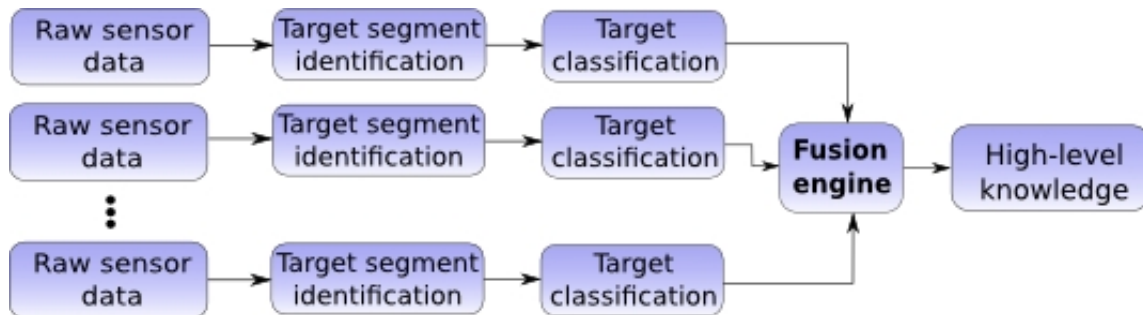
require a significantly smaller data volume than that of the raw sensor data. Because the only data transferred between a sensor node and the fusion engine are the extracted features, the communication bandwidth required under this fusion scheme is significantly less than what is required in the case of low-level fusion.



**Figure 15.3.** Feature-Level Fusion Scheme

### 15.2.3 Decision-Level Fusion

In this case, each of the data sources have sufficient onboard computing capability to process the raw sensor data to generate higher-level information with a more compact representation than the raw data. It is this higher-level information that the sensor nodes feed to the fusion engine. As a result the communication bandwidth requirements are more modest than in the case of low-level fusion. Figure 15.4 illustrates the concept of decision level data fusion.



**Figure 15.4.** Decision-Level Fusion Scheme

As an example, suppose one wanted to count the number of people in a building where each office is equipped with a camera to capture images at pre-defined time intervals. For this scenario, assume that image processing algorithms have been developed to extract the number of people that appear in a captured image. In a low-level fusion scenario, the camera in each room will send the raw image pixel data to a fusion engine which will then process each image to find the number of people in each office, and subsequently add the numbers from each office to report the total number of people in all the offices combined. Because each camera transmits the raw image, the bandwidth requirement is quite high.

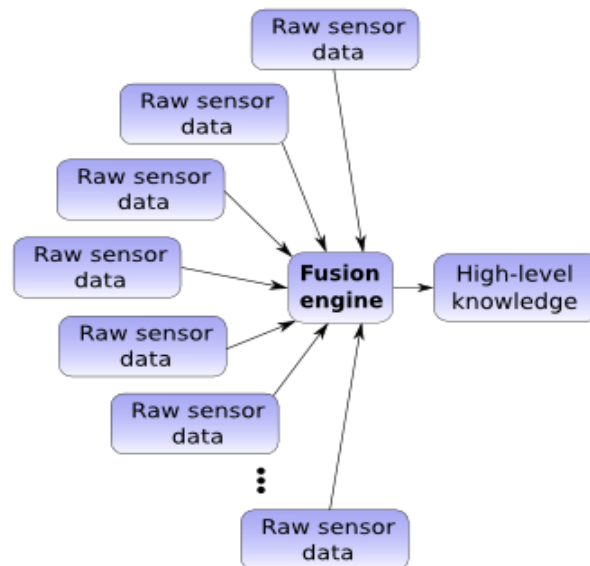
In the case of feature-level fusion, each sensor may analyze its image to determine the segments that show change from a reference image, such as an empty room. Then the changes would likely correspond to human occupants. However, a change may also correspond to an item in the room that has been moved since the reference image was captured, such as a chair. The sensor in this paradigm does not

differentiate humans from non-humans. It simply identifies the image segments corresponding to such changes and extracts relevant features, such as the area, shapes, and locations in the image of these segments. It then sends these features to the fusion node which combines them to produce the high-level knowledge about the environment.

In contrast, in a high-level fusion scenario, each camera system will locally run all the algorithms necessary to extract the accurate count of people in the respective room by fully analyzing the captured raw image and only send this number (a single integer) to the fusion engine, which will add up the numbers sent by each camera to output the final result. Obviously, the bandwidth requirement in this case for the transfer of high-level information (a single number) by each sensor to the fusion engine is significantly lower than that in the previous cases where the whole image or the features had to be transferred. But this reduction in bandwidth comes at the expense of each camera system requiring enough computational power to perform complete image analysis.

#### 15.2.4 Distributed versus Centralized Fusion

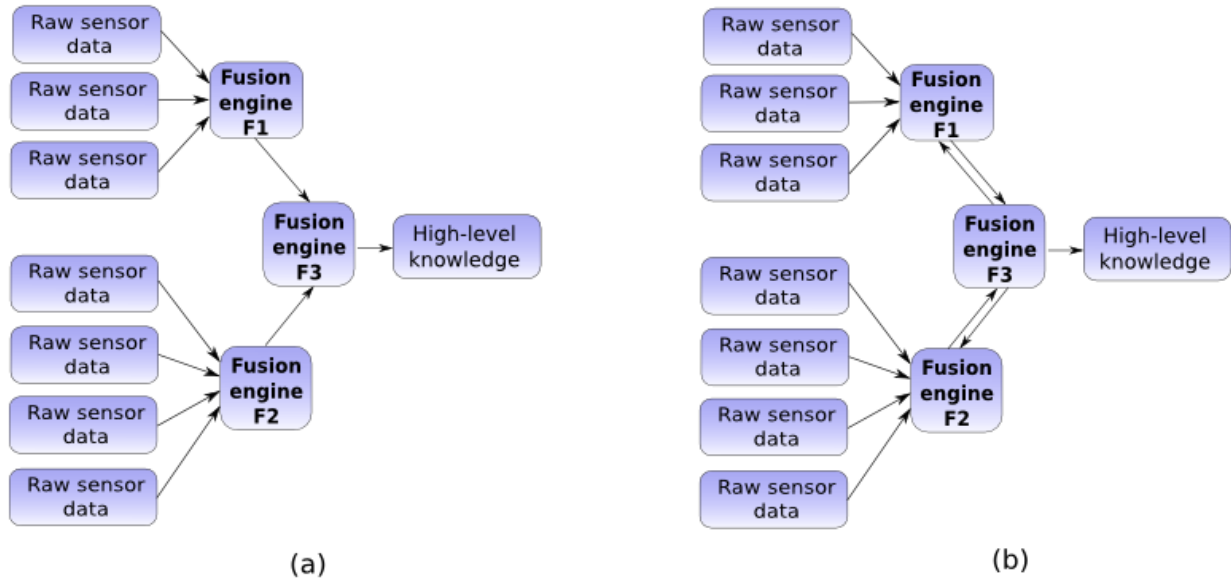
Traditionally, a centralized architecture has been common in data-fusion systems. In such a system, there is a single fusion processing node, and all the sensors send their data to this node. The fusion node is responsible for all aspects of data merging and extraction of high level information. Figure 15.5 illustrates a centralized fusion system. This approach is algorithmically simple to implement but requires significant communication bandwidth because every sensor needs to send the entire output to the fusion node. A major drawback of this approach is its susceptibility to a single point of failure. If the fusion node fails, the entire system ceases operation.



**Figure 15.5.** Centralized Fusion Architecture

However, a data-fusion system does not necessarily need to have a centralized architecture. The fusion task can be performed in a distributed manner. An incremental step toward distributed fusion from a centralized framework is an architecture in which the overall fusion task can be divided into smaller

subtasks that are performed by separate fusion sub-engines. The output of each of these sub-engines can then be merged at a master fusion node to produce the final fused output. This is a bi-level hierarchical fusion architecture. The obvious generalization is a multi-level hierarchy. Hierarchical fusion architectures can be broadly classified into two categories—one with feedback and the other without feedback among the fusion nodes. Figure 15.6 illustrates the concept of hierarchical fusion. Figure 15.6(a) shows an architecture without feedback, and Figure 15.6(b) shows one with feedback.



**Figure 15.6.** Hierarchical Data-Fusion Architectures

There are several advantages to hierarchical data-fusion schemes. It requires significantly less communication bandwidth because a low-level fusion node receives data from a smaller number of sensor nodes compared to the case of centralized fusion. And, the data volume transferred between any two fusion nodes is significantly smaller than the raw sensor data size because the fusion nodes usually transfer aggregated information at a higher level of information abstraction with a more compact representation than raw sensor output. In addition, each fusion node needs smaller computational capacity because the computational tasks are distributed across multiple fusion nodes. These also do not have any single point of catastrophic failure and, hence, are more resilient.

The extreme case of a distributed and decentralized data fusion framework is one in which there are no designated fusion engines but the network of sensors, as a whole, produces the desired fused output through cooperative computation. Such computational frameworks are known as “swarm” algorithms. Such systems are more robust and fault tolerant than centralized or semi-centralized architectures because the swarms are robust against the failure of a subset of nodes. There is no predefined communication hierarchy or topology and the network is capable of autonomously reconfiguring its communication pathways if a set of nodes happen to fail. Also, adding new nodes into the system is seamless. This makes this type of architecture highly scalable.

Despite these advantages, the distributed and decentralized fusion architectures pose a number of challenges. In the hierarchical scheme, it is not trivial to design an optimal architecture, such as defining the assignment of sensor subsets to first-level fusion nodes and the communication topology for the

fusion nodes. Also, one cannot take a centralized fusion algorithm and apply it in a hierarchical fusion system. The algorithms need to be tailored to the architecture, which is not an easy task. Designing a swarm-based algorithm for data fusion is likely the most complex of all the decentralized fusion schemes. This has led to research to address these challenges. For more information, see [3], [4], [7], [12], [13], and [15].

### 15.2.5 Data Fusion Process Model

Data fusion systems are highly domain specific, and there is no one-size-fits-all solution that can be deployed without considering the specific requirements for the application. Because of such diverse system requirements, it was important to develop a common descriptive substrate to elucidate the core concepts of data fusion as a discipline. This led to the development of the data-fusion process model which decomposes any data fusion task into a set of six subtasks, or levels. This general process mode is applicable across diverse application domains as elaborated in [16] and [20].

These six levels are:

- Level 0: Source Preprocessing/Data Alignment
- Level 1: Object Refinement
- Level 2: Situation Refinement
- Level 3: Impact Assessment (or Threat Refinement)
- Level 4: Process Refinement
- Level 5: User Refinement (or Cognitive Refinement)

#### 15.2.5.1 Level 0: Source Preprocessing/Data Alignment

Source pre-processing is one of the most fundamental steps in any data-fusion framework. Here, the raw data are subjected to a set of conditioning steps to make them ready for fusion. Examples of such conditioning are noise reduction for noisy images, color space translation (such as conversion of color images to monochrome images), orienting all geo-spatial data to north up, scaling all data to a desired range, etc. It is important to understand that no actual fusion of data from different sources takes place at this level. One particular pre-processing—*data alignment*—task deserves elaboration as it is almost ubiquitous in data fusion applications. Data alignment refers to the task of bringing raw data from all the sensors to a common representational framework so that one can make a valid comparison and assess the similarities or dissimilarities among the data. Data alignment is not a simple or easy task by any means. In fact, this can be one of the most critical steps in a data fusion pipeline. If the alignment is incorrect; all subsequent analyses will produce incorrect inferences.

The situation gets far more complex when the sensors are heterogeneous in nature. For example, if one installs microphones to detect human voices in areas where cameras may not be deployed or carbon dioxide sensors to get an estimate of the number of people, the data from all these different sensors need to be aligned before they can be fused to make a unified interpretation. Aligning such disparate data sources is a challenging task.



#### **15.2.5.2 Level 1: Object Refinement**

The core task is to fuse the Level 0 data from the various data sources to identify items of interest. Sensors can have different resolutions, angles of view, or other characteristics. The classification step essentially groups the identified items into categories. The processes used in Level 1 are highly application specific.

Level 1 processing becomes especially challenging when a larger number of items are to be tracked. Not only does the computation complexity increase, but the situation injects uncertainty and ambiguity through occlusions, overlaps, track intersections, etc. <sup>[18] [1]</sup>.

#### **15.2.5.3 Level 2: Situation Refinement**

The goal of Level 2 is to determine relationships among the detected targets/entities as well as the environment to extract a context. The targets (objects, spatio-temporal events, etc.) detected in Level 1 need to be aggregated in this stage to obtain desired contextual knowledge.

Besides aggregating target objects, it is also useful to aggregate the temporal information of events because the same event can imply a vastly different context depending on its time of occurrence.

A time sequence also can provide insight. A sequence of events in a specific time order can provide valuable contextual information about a situation.

#### **15.2.5.4 Level 3: Impact Assessment**

The essence of impact assessment is to extrapolate the situational awareness obtained in Level 2 to forecast the evolution of the scenario in the future. This requires prediction based on currently observed events and estimating the ramifications of those events.

#### **15.2.5.5 Level 4: Process Refinement**

In essence, process refinement is not part of any core fusion task, but it refers to the auxiliary task of monitoring the performance of the current fusion infrastructure and dynamically modifying aspects of the framework to achieve optimal performance. For example, the sensors may need to be re-oriented to improve coverage, or the Level 1 fusion may need to be switched from a lower level fusion mode to a higher decision level fusion.

#### **15.2.5.6 Level 5: User Refinement**

Level 5 involves the process of incorporating a human-in-the-loop facility by incorporating various human-computer interaction (HCI) mechanisms into the data-fusion pipeline. Examples of such HCI methods are information visualization, haptic feedback, and verbal interaction.

## 15.3 Characteristic Interdependencies

Any data-fusion task involves a set of subtasks. The fusion model formalizes these subtasks into six levels as elaborated in Section 15.2.5. However, each of these levels can be decomposed further depending on the application domain. Most of these subtasks represent interdependent problems.

The most basic interdependency is between Level 0 pre-processing and estimation of the various states of the target. The uncertainty in these state estimates is highly dependent on the fidelity of the data-alignment task. Any error in data alignment will likely be amplified in the subsequent stages of the fusion pipeline.

Another crucial interdependency is between the uncertainty in the final fused result and the uncertainty in the data gathered by the sensors. Uncertainty in raw sensor data can arise from various sources, such as sensor calibration drift, uncertainty in the location of the sensor itself (e.g., in a GPS-denied environment), uncertainty in the location of a target (e.g., when the target is occluded), etc. Such uncertainties, if not taken into account in the fusion process, can have significant impact on the accuracy of the fused output.

The communication bandwidth among the sensors or between a sensor and a fusion engine can affect the choice of fusion architecture to use, as alluded to in Section 15.2.

The available power source also determines the type of fusion system to be deployed. For example, in a building surveillance system, all sensors, as well as the fusion engine, can draw power from the buildings electrical infrastructure. So, power availability is not a constraint. Hence, they can operate almost indefinitely and be of a permanent nature. However, if we want to deploy a set of temporary sensors, or sensors without a hard-wired power source, this will impose limits on the architecture.

Criticality of the fused product influences the architecture of the data-fusion system as well. If it is imperative, for some reason, to collect data without interruption, it is essential to build as much redundancy into the system as possible within the constraints of the project.

## 15.4 Typical Applications

Data fusion has tremendous potential in the building management domain by reducing the cost of day-to-day operations and maintenance through efficient coordination of various active components such as HVAC, electrical systems, and mechanical systems, as well as reducing waste of valuable resources such as water and power. Operation and maintenance are usually the most expensive elements of a building's life cycle expenses, and these costs usually increase over time as a building gets older. An appropriate sensor fusion framework installed in a building can significantly reduce these costs. Bogen<sup>[2]</sup> states that, "... *approximately \$15.8 billion of annual U.S. capital facility industry efficiency losses are due to inadequate interoperability in design, engineering, facilities management...*" A sensor suite that is properly chosen and deployed in a building and the appropriate fusion of the data from these sensors can provide reliable, accurate, and actionable near real-time information that a building operations manager can leverage to make decisions about the most efficient control of the various components. This can result in substantial energy savings and reduced operational cost.

A key component in efficient building operations is real-time information about occupancy load distribution across a facility. If the dynamic occupancy profile can be determined in real-time, the information can be exploited for automatic dynamic control of lighting, temperature, air-flow, and other indoor environmental parameters to provide optimal comfort for the occupants without wasting energy on maintaining the same level of comfort in unoccupied areas. To determine a building-wide occupancy profile, it is necessary to deploy a suite of sensors of different modalities (such as temperature sensors, humidity sensors, visible cameras, infrared sensors, etc.) distributed throughout the building. Information from these sensors could then be fused to produce a unified and actionable knowledge product that can be used to optimally control the various building systems.

Another area of interest in energy efficient building management is predictive occupancy modeling. By discretely sampling the occupancy load distribution data obtained through multi-sensor fusion and archiving the data in a database over a significant period of time, machine-learning techniques can be applied to discern occupancy patterns. These patterns can then be used to predict spatio-temporal occupancy and activate the systems appropriately to bring the relevant areas to the desired environmental state at the right time.

Human activity detection in an indoor environment has been an active area of research, motivated not only by energy conservation, but also by other application domains, such as gaming. However, the fundamental technological advances resulting from these diverse research activities can be exploited in smart building management systems. Coen<sup>[5]</sup> carried out one of the earliest instances of research about tracking people and activity detection in a room. His team used multiple sensors to detect locations of people and their activities and offered automated help using artificial intelligence algorithms. Mozer<sup>[14]</sup> developed a framework where a building learns by observing occupant behaviors over time. The key differentiator of this research was that the building had the ability to adapt its functionality to the inhabitants' desires and habits. The MavHome framework created by Das<sup>[6]</sup> imparts rational agent attributes to a house with the goal of maximizing occupants' comfort and minimizing operating costs. The MavHome exploits multi-sensor information fusion to create predictive models that anticipate occupants' states and automate routine and repetitive actions.

At a finer level, data fusion can play an important role in getting reliable estimates about the state of a specific sub-system in a building. For example, Huang<sup>[11]</sup> developed a data fusion scheme for improving the measurement of the cooling load in chiller plants in buildings.

Besides smart building management, another active application domain for data fusion is the defense sector. In 1999 and 2000, the U.S. Department of Defense conducted a survey of operational as well as their currently planned data-fusion applications. Hall<sup>[9]</sup> presents an extensive list of these applications in Chapter 22 of their *Handbook of Multi-Sensor Data Fusion*. Data fusion plays an important role in obtaining timely actionable intelligence through surveillance. It is desirable to gather as much information about an adversarial target as possible to develop the most effective strategy to counter a threat.

## 15.5 Relative Cost Impacts

There is no one-size-fits-all data-fusion solution for all application domains. In most cases, fusion systems are highly tailored to specific requirements and the cost-impact analysis should be carried out for

each specific case. However, fusion systems are becoming more and more viable from a cost perspective because of the falling price of many commodity sensors that can play a fundamental role. Multi-sensor fusion systems can provide significant cost reductions in the operations-and-maintenance phases in many domains.

An application domain that can see significant cost saving through the use of multi-sensor data fusion is in lighting and interior environment control systems in buildings. By deploying appropriate sensors to obtain real-time occupancy information in buildings, it allows a building management system to optimally manage various environmental parameters, such as temperature, humidity, and air-flow in real time. This can provide significant reductions in energy consumption by dynamically adjusting the environment based on occupancy levels and patterns, as opposed to the current practice of setting the environment globally across the building to address the worst case scenario.

## 15.6 Summary/Conclusions

Data fusion is a key technology for making optimal use of the vast and ever expanding network of homogeneous and heterogeneous sensors being deployed in more and more application domains. It is a critical fundamental technology underlying most automated, adaptive, smart, real-time infrastructures. A truly smart system needs to be adaptive because the environment in which a system operates is rarely static, and a system needs to adjust its operational profile according to circumstantial changes. For example, in a smart building system geared towards automated lighting and HVAC control, the automated control logic needs to adapt because the needs of each occupant may differ in winter and summer. At a finer level, the control system may need to learn individual preferences and may need to keep one office warmer than another. Multi-modal data fusion can provide the core automated decision-making layer in the overall automation stack. In the early stages of this field, the U.S. Department of Defense was the main beneficiary of data fusion, but eventually, civilian sectors began to recognize the value of this technology in their domains. Multi-sensor fusion has a tremendous potential in reducing the energy budget of building infrastructures, thus reducing the environmental impacts of operations as well.

## 15.7 References

1. Bar-Shalom Y. and Li X. R., “*Multitarget-multisensor tracking: Principles and Techniques*”, YBS Press, 1995.
2. Bogen, C., Rashid, M. and East, E. W., “*A framework for building information fusion*”, Proceedings of the CIB W78-W102 2011: International Conference –Sophia Antipolis, France, 26-28 October, 2011.
3. Cetin M., Chen L., Fisher III J., Ihler A., Moses R., Wainwright M., and Willsky A., “*Distributed fusion in sensor networks*,” Signal Processing Magazine, IEEE, vol. 23, no. 4, pp. 42-55, 2006.
4. Chair Z. and Varshney P. K., “*Optimal data fusion in multiple sensor detection systems*,” IEEE Trans. Aerosp. Electron. Syst., vol. AES-22, pp. 98–101, Jan. 1986.
5. Coen, M., “*Design principles for Intelligent Environments*”, AAAI Spring Symposium, Stanford, pp. 36-43, March 1998.

6. Das S., Cook D., Bhattacharaya A., Heierman E., and Lin T., "*The Role of Prediction Algorithms in the MavHome Smart Home Architecture*", IEEE Wireless Communications, vol. 9, no. 6, pp. 77-84, December 2002.
7. Durrant-Whyte, H. F., Rao B. S. Y., and Hu H., "*Toward a fully decentralized architecture for multi-sensor data fusion*," in Proc. IEEE Int. Conf. Robot. Automat., 1990.
8. Hall D.L. and Llinas J., "*An introduction to multisensor data fusion*", Proceedings of the IEEE, vol 85(1), pp. 6 – 23, 1997.
9. Hall D.L. and Llinas J., "*Handbook of multi-sensor data fusion*", CRC Press, 2001.
10. Hall D. H. and McMullen A. H., "*Mathematical techniques in multisensor data fusion*", Artech House, 2004.
11. Huang G., Wang S., Xiao F., and Sun Y., "*A data fusion scheme for building automation systems of building central chilling plants*", Automation in Construction, vol. 18, pp. 302-309, 2009.
12. Julier S. and Uhlmann J., "*General decentralised data fusion with covariance intersection (CI)*," in Handbook of Data Fusion, D. Hall and J. Llinas, Eds. CRC Press, 2001, pp. 319-343.
13. Makarenko A. and Durrant-Whyte H., "*Decentralised data fusion and control in active sensor networks*", in Proc. Int. Conf. Information Fusion, 2004.
14. Mozer, M., "*An intelligent environment must be adaptive*", IEEE Intelligent Systems, vol. 14, no. 2, pp. 11-13, March/April 1999.
15. Rao B. S. Y., Durrant-Whyte H. F., and Sheen J. A., "*A fully decentralized multi-sensor system for tracking and surveillance*," Int. J. Robot. Res., vol. 12, no. 1, pp. 20–44, Feb. 1993.
16. Steinberg A. N., Bowman C. L., and White Jr. E. E., "*Revision to the JDL data fusion model*", Proceedings 3<sup>rd</sup>. NATO/IRIS Conference, Quebec City, Canada, 1998.
17. Strelow D. and Singh S., "*Optimal Motion Estimation from Visual and Inertial Measurements*", Proceedings of the Workshop on Applications of Computer Vision, December 2002.
18. Uhlmann J. K., "*Algorithms for multiple-target tracking*", American Scientist, 80(2), 1992.
19. Veth M, J. and J. F. Raquet J. F., "*Two-dimensional stochastic projections for tight integration of optical and inertial sensors for navigation*", National Technical Meeting Proceedings of the Institute of Navigation, pp. 587–596, 2006.
20. White, Jr. E. E., "*A model for data fusion*", Proceedings 1<sup>st</sup> National Symposium on Sensor Fusion, vol. 2, 1988.

## 16.0 Signal Conversion to Data: Signal Processing

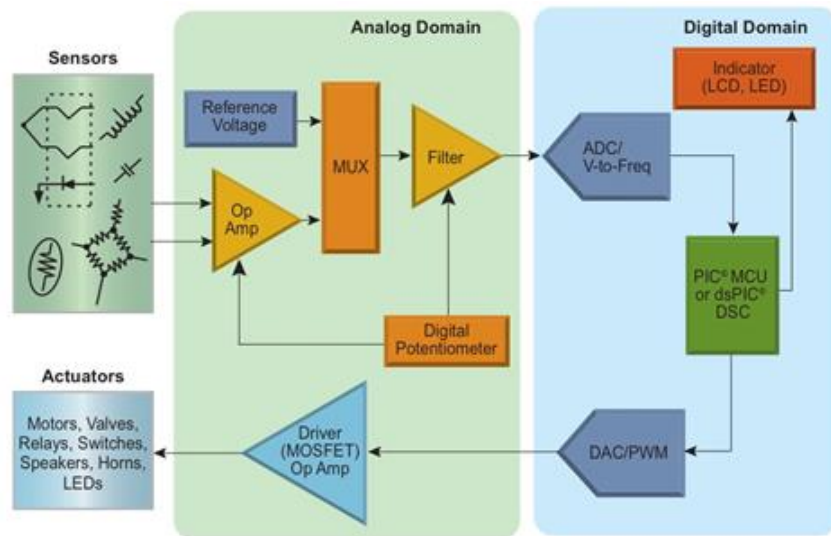
### 16.1 Introduction

Sensors are used to measure and detect a wide variety of physical parameters. These parameters are electrical (voltage, current, resistance, etc.), mechanical (force, movement, etc.), pneumatic/hydraulic (pressure, flow, etc.), optical (light levels, colors, etc.), radio waves (frequency, phase, etc.), etc. To extract useful information from raw signals generated by the transducer elements in a sensor, modern systems widely use signal conditioning and signal processing techniques that operate on the electrical signal outputs of the transducer. This is accomplished by converting non-electrical parameters into electrical parameters, if necessary. A list of widely and commonly used physical parameters and the conversion to electrical signals is presented in Table 16.1<sup>[1]</sup>.

**Table 16.1.** Physical Parameters Measured and Conversion to Electrical Signals

Phenomena	Sensor	Electrical Output
Magnetic	Hall Effect	Voltage
	Magneto-Resistive	Resistance
Temperature	Thermocouple	Voltage
	RTD	Resistance
	Thermistor	Resistance
	IC	Voltage
	Infrared	Current
	Thermopile	Voltage
Humidity	Capacitive	Capacitance
	Infrared	Current
Force, Weight, Torque, Pressure	Strain Gauge	Resistance/Voltage
	Load Cell	Resistance
	Piezo-electric	Voltage or Charge
	Mechanical Transducer	Resistance, Voltage, Capacitance
Motion and Vibration	LVDT	AC Voltage
	Piezo-electric	Voltage or Charge
	Microphone	Voltage
	Ultrasonic	Voltage, Resistive, Current
	Accelerometer	Voltage
Flow	Magnetic Flowmeter	Voltage
	Mass Flowmeter	Resistance/Voltage
	Ultrasound/Doppler	Frequency
	Hot-wire Anemometer	Resistance
	Mechanical Transducer (turbine)	Voltage
Fluid Level and Volume	Ultrasound	Time Delay
	Mechanical Transducer	Resistance, Voltage
	Capacitor	Capacitance
	Switch	On/Off
	Thermal	Voltage
Light	Photodiode	Current
Chemical	pH Electrode	Voltage
	Solution Conductivity	Resistance/Current
	CO Sensor	Voltage or Charge
	Photodiode (turbidity, colorimeter)	Current

Electrical signals then are conditioned, almost entirely implemented as electronics hardware in the analog domain which is very inexpensive. Signal conditioning entails mapping the output signal into a range that the remaining electronic circuitry can process, reducing or eliminating electronic noise, and then converting the electrical/electronic signal from analog domain to the digital domain. Modern sensors also have an increasing amount of mid-end (or preliminary) digital signal processing. These are primarily digital filters that can be implemented more efficiently on low to moderate cost microcontroller units. These steps are illustrated in Figure 16.1 <sup>[1]</sup>.



**Figure 16.1.** Signal Conditioning

In addition, the process of converting increasingly complex and large amounts of digitized and filtered sensor data to information requires complex and advanced signal-processing algorithms. These are almost always implemented as software on advanced digital signal processors, complex system-on-chip microprocessors, and even desktop computers. These back-end processing systems are also used to aggregate data from a multitude of sensors to amortize the cost of processing data as well as to extract system-level information.

## 16.2 Definitions:

### 16.2.1 Technical Details

#### 16.2.1.1 Signal Conditioning

There are three classes of operations typically composed in a sequence to accomplish signal conditioning: 1) amplification, 2) anti-alias filtering, and 3) ADC. In addition to these, we will review an important process known as *calibration*.

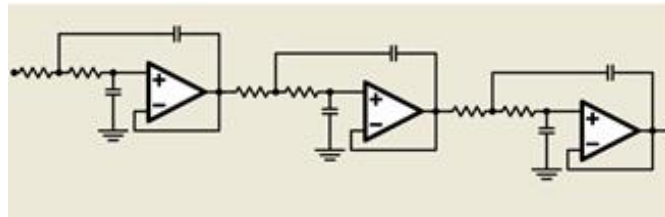
#### Amplification

Signals generated by sensors are usually weak. To process the signal accurately and reduce the effect of noise, the signal needs to be amplified <sup>[2], [24]</sup>.

The most critical element of an amplifier circuit is usually an operational amplifier (a.k.a., op-amp). The op-amp is an extremely efficient and versatile device. It is essential to perform due diligence in selecting the right op-amp for a particular application. There are some key electrical characteristics of op-amps and their effects on system performance. The first characteristic is the gain-bandwidth product, which defines how the op-amp gain factor decreases as signal frequency increases. Two other key quality metrics of an op-amp are its common mode rejection ratio and power supply rejection ratio. Common mode gain is a term used to describe how much the op-amp output would change if the same input signal were to be fed into both inputs and this common signal were to be changed by 1 volt. The power supply rejection ratio is computed in a similar manner, but this metric relates to the change in output voltage with a change in power supply voltage.

## Anti-Alias and Analog Filtering

While it is necessary that the signals generated by the sensor must be as free of noise as possible, it is also equally important that the frequency content of the signal, in other words its bandwidth, must be limited to a certain range. Anti-aliasing filters are essentially low-pass filters implemented in the analog domain (see Figure 16.2). There are numerous tools/software available that can help design the correct anti-alias filter for a given sensor signal. These require the designer to supply certain specifications: passband cut-off frequency, ADC sampling frequency, ADC resolution (number of bits), and desired SNR.



**Figure 16.2.** Example of Anti-Aliasing Filter Implementation

## Analog to Digital Conversion

To be processed by a digital microprocessor or microcontroller, the signal needs to be converted into digital data. The primary process performed in an ADC is known as *quantization*. The hypothesis behind this process is that if analog signals can be converted to a stream of digital data, the power of digital computers and software can be used to great advantage to perform many manipulations on the signal.

For most real-world applications, this is still very useful despite its limited nature. This inherent uncertainty in digitizing an analog value is referred to as the quantization error. The quantization error depends on the number of bits in the converter, along with its errors, noise, and non-linear attributes.

## Calibration

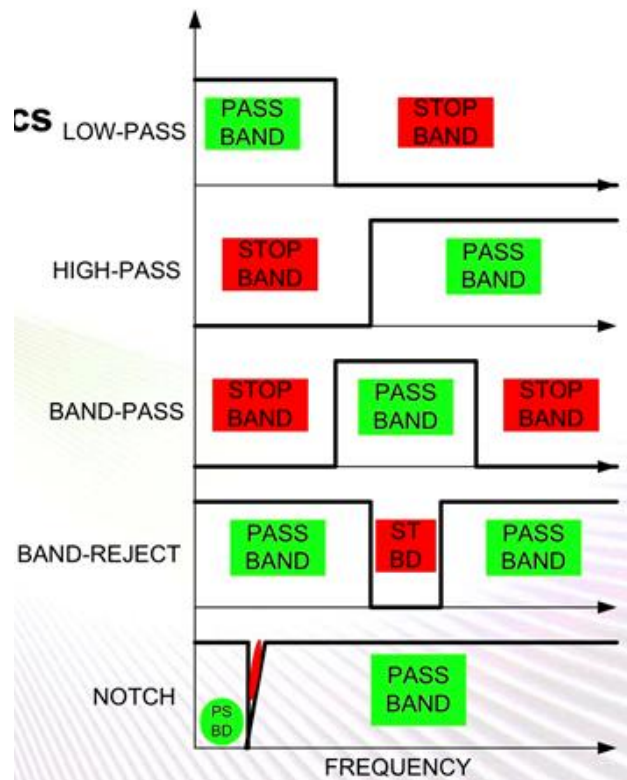
The output of many modern sensors contains errors that must be eliminated through calibration. One major source of error is variation of sensor span and offset from device to device and their respective drifts with the physical parameter being measured (such as temperature). These drifts can be non-linear. Another source of error is the non-linearity of the sensor output with applied stimulus. Signal



conditioning electronics also can introduce errors. The complexity and duration of the calibration process that removes these errors is a major cost factor in modern sensor production. Multiple measurements have to be made while applying stimulus to a sensor during the calibration process. Programmable signal conditioning amplifiers can be used to store the computed calibration data and then use it to eliminate initial errors <sup>[7]</sup>.

### 16.2.1.2 Preliminary Signal Processing

As modern control systems become increasingly more complex, so does the requirement for embedded processing within individual sensor units. Whether for localized data analysis or to support the growing need for digital communication protocols, many sensor manufacturers are now including a microcontroller unit within the sensor to enable these new features. Preliminary signal processing primarily consists of digital filtering. Based on frequency characteristics, digital filtering can be partitioned into low-pass, high-pass, band-pass, band-reject, and notch filters as shown in Figure 16.3 <sup>[9]</sup>.



**Figure 16.3.** Types of Filters used in Digital Filters

Based on the type, they can be viewed through the traditional dichotomy of finite impulse response (FIR) and infinite impulse response (IIR) filters. In addition to these, lattice wave digital filters are also important in preliminary signal processing.

### 16.2.1.3 Advanced Signal Processing (ASP)

Advanced signal processing can broadly be thought of as a process to extract meaningful information from raw and preliminary filtered digital data. It is primarily accomplished through the implementation

of digital signal processing algorithms (and their variants such as image processing etc.) as software on digital microprocessors (such as digital signal processors, SoCs with embedded graphics processing units, etc.). In this section we will list widely used advanced signal processing algorithms and explain them in simple terms. Beyond these fundamental advanced signal processing algorithms, there are an infinite number of ways to combine them to create novel solutions to extract intelligent information at a sensing node level as well as at a system level from a multitude of sensing nodes.

## **Transforms**

This is one of the broadest categories of fundamental signal processing. It essentially encompasses a set of algorithms that convert digital signals from one domain to another. Mathematical transformations are applied to signals to obtain further information from the signal that is not readily available in the raw signal. For example, the transformation between the time and frequency domains is widely known and used through the discrete Fourier transform to more easily apply digital filters.

## **Compression Algorithms**

Data compression, source coding, or bit-rate reduction involves encoding information using fewer bits than the original representation. Compression can be either “lossy” or “lossless.” Lossless compression reduces bits by identifying and eliminating statistical redundancy. No information is lost in lossless compression. Lossy compression reduces bits by identifying marginally important information and removing it. Compression is useful because it helps reduce the consumption of resources such as data space or transmission capacity. Because compressed data must be decompressed to be used, this extra processing imposes computational or other costs through decompression.

### **16.2.1.4 Characteristic Interdependencies**

The characteristics identified so far are signal conditioning (amplification, analog domain filtering, ADC, and calibration), preliminary digital signal processing (FIR and IIR filters), and advanced digital signal processing (transforms and compression). The chain of operations has a direct and sequential impact on all operations up the chain. For example, amplification is often necessary prior to filtering, ADC, or calibration. Signal conditioning is necessary before any form of digital signal processing and so on. However, there is one exception to this linear chain of interdependencies. Typically, an amplifier is required if the magnitude of the input signal is significantly lower than the full-scale input range of the ADC. However, by selecting an ADC with a higher resolution, the need for an amplifier can be eliminated<sup>[19]</sup>.

### **16.2.1.5 Application**

#### **Choice of Amplification**

The choice of the amplification method is based on noise in the environment, need for simplicity of system/reduced components, location of sensing element, desired strength of amplification, and proportionality of output to varying nature and range of parameter being sensed.

## Single-Ended Amplifiers

Local sensor systems<sup>[1], [2]</sup> are those in which the sensors are located relatively close to their respective signal conditioning circuits; in such cases, the noise environment is not severe. Single-ended, non-inverting amplifiers are a good choice for amplifying the sensor output because they require a minimal amount of discrete components. Systems such as passive infrared sensors, thermistors, Hall effect sensors, smoke detectors, and humidity sensors, etc., can be designed using single-ended amplifier circuits/systems.

## Differential Amplifier

When signals are noisier, particularly in environments where there is excessive noise that can corrupt signals from the sensor or the conditioning circuit cannot be placed in close physical proximity to the sensor/transducer element, it is necessary to transmit raw signals over differential wires from a remote sensor to a type of amplifier called differential amplifier. For example, thermocouples and pressure sensors often need the help of differential amplifiers<sup>[1], [2]</sup>.

## Instrumentation Amplifiers

When high levels of noise cleansing and purer forms of the signal from the sensor/transducer are desired, such as in instrumentation systems, a class of differential amplifiers known as instrumentation amplifiers are preferred<sup>[1], [2]</sup>. A differential amplifier based on a single op-amp does not provide the high performance and precision needed by an instrumentation amplifier. Therefore, many instrumentation amplifiers use two op-amps, and ideally two op-amps located in the same integrated circuit chip, which provides better matching.

## Programmable Gain Amplifiers

In some sensing applications, it may be desirable to amplify signals to different extents if the form of energy (physical parameter) being sensed is non-linear over a range. For example, a thermistor producing non-linear outputs across the temperature range will need a form of programmability/flexibility to adjust its gain. For this purpose it is desirable to use a class of amplifiers known as programmable gain amplifiers<sup>[1], [2]</sup>.

## Trans-impedance Amplifiers

In some applications, when it is necessary to generate a current that is proportional to the parameter being sensed, a class of amplifiers known as trans-impedance amplifiers is needed. Unlike classical amplifiers, these do not simply amplify the output of a sensing element, they consume current produced by a sensing element (that varies according to a physical parameter) and converts it into a proportionally varying voltage waveform. Examples of trans-impedance amplifiers are photodiodes or CCDs that produce a current in proportion to light intensity.

## RC Op-Amp Oscillators

In cost sensitive applications that cannot afford an ADC process and use sensing transducers that are effectively resistive (Resistance Temperature Detectors (RTDs), humidity, and thermistors) or capacitive sensors (humidity/pressure/oil), RC op-amp oscillators can be used to directly obtain digitized signals. Unlike the other classes of amplifiers, this class does not need downstream ADCs. Op-amp or state-variable oscillators can be used to accurately measure resistive and capacitive sensors. Oscillators provide a sensor measurement in which conversion to digital has an accuracy that is limited only by the reference clock signal. State-variable oscillators are often used in sensor conditioning applications because they have a reliable startup and a low sensitivity to stray capacitance. Absolute quartz pressure sensors and humidity sensors are examples of capacitive sensors that can use the state-variable oscillator.

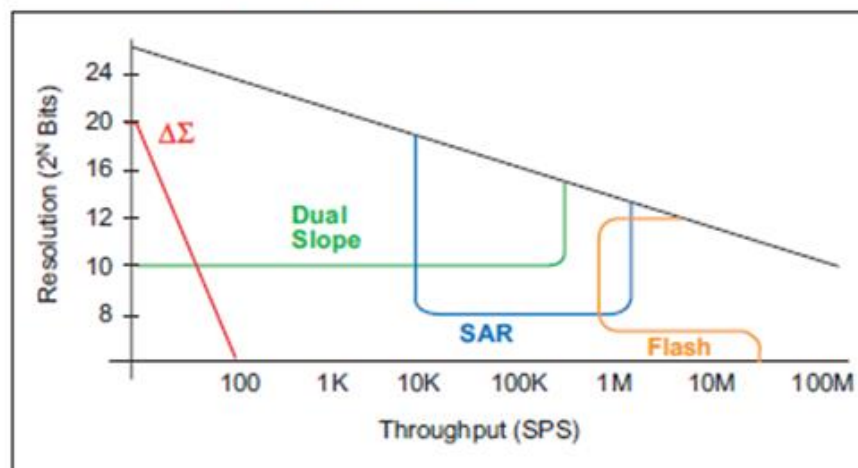
## Nature of Application that Determines Factors Related to the Design of Anti-Aliasing Filters and Analog Filters

While there are numerous tools/software available that can help design the correct anti-alias filter for a given sensor signal, the criteria that guide the design process include pass-band cutoff frequency, ADC sampling frequency, ADC resolution (number of bits), and desired SNR. Analog filtering is more suitable for higher speed systems (i.e., those above approximately 5 kHz). In these types of systems, an analog filter can reduce noise in the out-of-band frequency region.

## Factors to be Considered When Choosing an ADC

Selecting the most suitable ADC for an application is based on more than just the precision of bits. Different architectures are available, each exhibiting advantages and disadvantages in various data acquisition systems. The required accuracy or precision of the system puts the choice in a category based on the number of bits required. It is important to always design the system to allow for more bits than initially required: if an application calls for 10 bits of accuracy, the designer must choose a 12-bit converter. The achievable accuracy of a converter will always be less than the total number of bits available<sup>[19]</sup>. Figure 16.4 shows the different architectures versus bits and bandwidth.

**Architecture vs. Bits and Bandwidth**



**Figure 16.4.** Different ADC Architectures versus Bits and Bandwidth

## Dual-Slope Integrating Type of ADC

If an application requires noise immunity, the dual-slope integrating ADC is useful because input noise spikes are integrated (averaged to zero) during the integration periods. Integrating ADCs are immune to the large conversion errors that plague successive approximation converters in high-noise environments. Integrating converters provide inherent noise rejection, with at least a 20 dB/decade attenuation rate. Interference signals with frequencies at integral multiples of the integration period are, theoretically, completely removed because the average value of a sine wave of frequency  $(1/t)$  averaged over a period  $(t)$  is zero. A dual-slope integrating ADC operates by charging a capacitor from the input voltage during a fixed time and then discharging it to zero.

## Successive Approximation Register Type ADC

SAR ADC systems are useful when there is a need to connect the converter to multiplexed inputs at a high data acquisition rate. The input is sampled and held on an internal capacitor then the charge is converted to a digital output code using the successive approximation routine. Because this charge is held throughout the conversion time, only the initial sample and hold period or acquisition time is of concern to a fast-changing input. The conversion time is the same for all conversions. This makes the SAR converter ideal for many real-time applications, including motor control, touch-screen sensing, medical, and other data acquisition systems. SAR converters typically lie within the 8- to 16-bit resolution range and can have sample speeds up to 1 MSPS. Therefore, when the speed of sampling is of the utmost importance, a high-speed SAR is the best choice.

## Delta-Sigma ADC

The delta-sigma ADC is an oversampling converter with many benefits over the traditional SAR converter. High resolution, excellent line frequency rejection, limited external component requirement, minimal anti-aliasing filtering, and low power consumption are some of the prominent benefits. Therefore, a delta-sigma ADC is the best choice when speed of sampling is only a moderate constraint and other features (mentioned above) are more important, including accuracy requirements of up to 24 bits of resolution.

## Application Factors that Influence Sensor Calibration Process, Particularly the Nature and Range of the Parameter Being Sensed

Usually the process of calibrating a sensor in a narrow range is straightforward through the use of look-up tables. However, sometimes when it is necessary to precisely characterize a transducer over its entire range, this may prove difficult. An example is the case of a transducer with a highly nonlinear transfer function, such as a thermocouple. For such cases, Transducer Electronic Data Sheets (TEDS) provide an alternative means of calibration through use of a multi-segment polynomial curve describing the electrical-to-physical transfer function of a transducer.

Many RTDs, for example, exhibit linear behavior for part of their measurement range, and polynomial for the rest. To make effective use of the calibration curve template, several calibration points must be taken and then fitted to a polynomial curve. This may take significantly more time than collecting a few data points for a calibration table. However, the tradeoff is greater accuracy and a wider range of calibrated values. Some measurements may require information beyond the time domain. The

frequency response transfer function of a transducer is particularly useful for microphones and accelerometers. The frequency response of any vibration transducer, in fact, is very important when comparing readings. Sensors with a broader frequency response will see more vibration if it is present than a narrower bandwidth transducer.

## Application Factors to Help Choose Between FIR and IIR Digital Filtering

FIR filters should be used for purposes of anti-aliasing, low-pass filtering, reconstruction or baseband filtering that are less demanding to design and implement, robust, and low-cost. FIR filters are always stable, which is why it is a rule of thumb when implementing a design with FIR to first check if the eventual frequency shaping can be achieved before advancing to IIR filter, which will incur additional design effort. From a low-power-consumption design perspective, which is getting extremely important these days in portable electronics, IIR filters consume more power because they require more coefficients to be stored in memory and more processing variables are involved. FIR is preferably used in low power applications instead <sup>[25]</sup>.

IIR filters should be used for purposes of notch, band-limiting, noise-shaping or higher order filtering where sharp cutoff is required. All recursive/iterative/feedback structures have stability issues that must be dealt with effectively. All poles must be conjugated to ensure stability, and it is usually quite difficult to match the conjugate with a good coefficient in actual implementation because of the bit width of the digital filter that needs to be implemented. Another shortcoming of the IIR filter is that the structure is recursive, which implies increased complexity of the structure and additional coefficients that must be properly selected. Such problems are easily taken care of with FIR filters, which have a much simpler structure and fewer coefficients <sup>[25]</sup>.

The effects of using a limited number of bits to implement filters (on a digital microprocessor or microcontroller), such as rounding off noise and coefficient quantization errors, are much less severe in FIR filters compared to IIR filters. FIR filters require more coefficients for sharp cutoff filters compared to IIR filters. Thus, additional processing time and storage will be required for FIR implementation. For a given number of taps, the IIR filter is 5 to 10 times more efficient in controlling its gain response than the FIR filter. FIR can have an exactly linear-phase response. This is an important requirement in data transmission, biomedicine, digital audio, and image processing. The phase response of IIR filters is non-linear. In general, the effect of noise that is introduced is more marked in IIR filters because their feedback elements cause the errors to be accumulated over time. FIR filters, however, are simply feed forward circuits so that any errors appear in the output signal only once per sampling period.

## Transforms

### Discrete Fourier Transform (DFT)

There are many DSP applications where a long signal must be filtered in *segments*. For instance, high fidelity digital *audio* requires a data rate of about 5 Mbytes/min, while digital *video* requires about 500 Mbytes/min. With data rates this high, it is common for computers to have insufficient memory to simultaneously hold the entire signal to be processed. There also are systems that process segment-by-segment because they operate in *real time*. For example, telephone signals cannot be delayed by more than a few hundred milliseconds, limiting the amount of data that are available for processing at any one

instant. In other applications, the *processing* may require that the signal be segmented. These characteristics necessitate the use of a process known as finite Fourier transform (FFT) convolution using the overlap-add method <sup>[27]</sup>. This method is based on 1) decomposing the signal into simple components, 2) processing each of the components in some useful way, and 3) recombining the processed components into the final signal. However, performing convolution in the time domain (step 2 and 3) is extremely computationally intensive. Therefore, FFTs are suitable in such cases to transform short duration time domain signals into the frequency domain, enabling fast multiplication in the frequency domain, followed by an inverse FFT into the time domain.

### **Discrete Cosine Transform (DCT)**

DCTs <sup>[13]</sup> are important to numerous applications from lossy compression of audio (e.g., MP3) and images (e.g., JPEG) (where small high-frequency components can be discarded), to spectral methods for the numerical solution of partial differential equations. There are eight versions of DCT (all of which have better energy-compaction properties than the DFT), which are classified into even and odd with four types. All present digital signal and image processing applications (mainly transform coding and digital filtering of signals) involve only *even* types of the DCT <sup>[22]</sup>. It is a real transform with better computational efficiency than DFT, which by definition is a complex transform and does not introduce discontinuity while imposing periodicity in the time signal. In DFT, as the time signal is truncated and assumed periodic, discontinuity is introduced in time domain and some corresponding artifacts are introduced in frequency domain. But as even symmetry is assumed while truncating the time signal, no discontinuity and related artifacts are introduced in DCT. Therefore, choose DCT for applications that require better computational efficiency and energy compaction than is possible with a DFT, and which cannot tolerate discontinuity related artifacts.

### **Discrete Wavelet Transform (DWT)**

A key advantage DWT <sup>[14]</sup> has over Fourier transforms <sup>[10]</sup> is temporal resolution: it captures both frequency *and* location information (location in time). Additionally, in practice, the transform is applied to many scales and sizes of the signal resulting in vectors that encode both approximation and detail information. There are numerous filters available for the DWT such as Haar, Daubechies, Coiflets, Symlets, etc. <sup>[23]</sup>. For applications with importance of time, the wavelet filter with lesser slope needs to be used since it has a smaller time delay.

## **Compression**

In general, across the spectrum of lossless compression techniques, the compression time increases as file size increases. However, for run-length encoding, it is a constant value and not affected by the file size. Applications seeking a constant compression time could choose run-length encoding.

### **16.2.1.6 Cost**

#### **Signal Conditioning**

Most of the systems in the sub-topic of signal conditioning are considered relatively low cost compared to system components required for preliminary or even advanced signal processing.

## Preliminary Signal Processing

The microcontroller units needed to implement preliminary signal processing filters are considered moderate to low cost.

## Advanced Signal Processing

DSPs, system-on-chip, and other advanced digital computing chips needed to implement ASP algorithms are considered moderate to high cost.

## 16.3 References

1. "Signal Chain Design Guide- Devices For Use With Sensors", Microchip Technology Inc.
2. "Hardware Conditioning of Sensor Signals", Priyabrata Sinha. Jan 2010. Microchip Technology Inc.
3. "A Practical Technique for Minimizing the Number of Measurements in Sensor Signal Conditioning Calibration", Art Kay, Mikhail Ivanov, Viola Schäffer. *Application Report SBOA111–June 2005. Texas Instruments Inc.*
4. "Optimized Digital Filtering for the MSP430", Kripasagar Venkat. MSP430 advanced technical conference. Sonthofen , Germany, 2006.
5. J. Cooley, P. Lewis, and P. Welch (1969). "The finite Fourier transform". *IEEE Trans. Audio Electroacoustics* 17 (2): 77–85.
6. N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform", *IEEE Trans. Computers*, 90-93, Jan 1974.
7. S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. San Diego, CA: Academic, 1999.
8. "Analog-to-Digital Converter Design Guide", Microchip Technology Inc. 2012.
9. "The Transform and Data Compression Handbook", Edited by K. R. Rao and P. C. Yip. CRC Press 2000. Print ISBN: 978-0-8493-3692-8.
10. Ingrid Daubechies. 1993. Orthonormal bases of compactly supported wavelets II: variations on a theme. *SIAM J. Math. Anal.* 24, 2 (March 1993), 499-519.
11. "AN990: Analog Sensor Conditioning Circuits – An Overview", Kumen Blake. Microchip Technology Inc. DS00990A. 2005.
12. Rabiner, Lawrence R., and Gold, Bernard, 1975: *Theory and Application of Digital Signal Processing* (Englewood Cliffs, New Jersey: Prentice-Hall, Inc.) ISBN 0-13-914101-4
13. "The Scientist and Engineer's Guide to Digital Signal Processing, copyright ©1997-1998 by Steven W. Smith.



## Distribution

**No. of  
Copies**

# Name  
Organization  
Address  
City, State and ZIP Code

# Organization  
Address  
City, State and ZIP Code  
Name  
Name  
Name  
Name (#)

# Name  
Organization  
Address  
City, State and ZIP Code

**No. of  
Copies**

# **Foreign Distribution**

# Name  
Organization  
Address  
Address line 2  
COUNTRY

# **Local Distribution**

Pacific Northwest National Laboratory  
Name Mailstop  
Name Mailstop  
Name Mailstop  
Name Mailstop  
Name (PDF)





**Pacific Northwest**  
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

902 Battelle Boulevard  
P.O. Box 999  
Richland, WA 99352  
1-888-375-PNNL (7665)  
[www.pnnl.gov](http://www.pnnl.gov)



U.S. DEPARTMENT OF  
**ENERGY**