Pub. No. 9

THE AMERICAN PRACTICAL NAVIGATOR

AN EPITOME OF NAVIGATION

ORIGINALLY BY

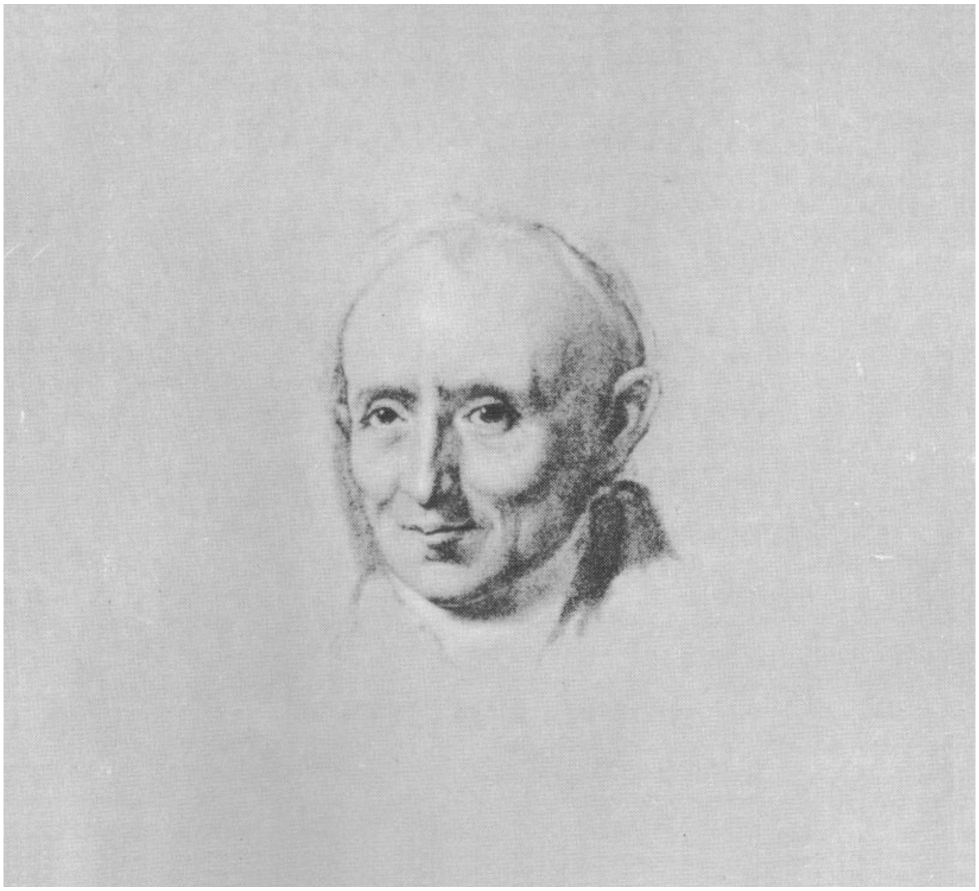
NATHANIEL BOWDITCH, LL.D.

2002 BICENTENNIAL EDITION

Prepared and published by the
NATIONAL IMAGERY AND MAPPING AGENCY
Bethesda, Maryland

NSN 7642014014652
NIMA REF. NO. NVPUB9V1

For sale by the Superintendant of Documents, U.S. Government Printing Office
Internet: bookstore.gpo.gov Phone: toll free (866) 512-1800; DC area (202) 512-1800
Fax: (202) 512-2250 Mail Stop: SSOP, Washington, DC 20402-0001



Last painting by Gilbert Stuart (1828). Considered by the family of Bowditch to be the best of various paintings made, although it was unfinished when the artist died.

NATHANIEL BOWDITCH
(1773-1838)

Nathaniel Bowditch was born on March 26, 1773, in Salem, Mass., fourth of the seven children of shipmaster Habakkuk Bowditch and his wife, Mary.

Since the migration of William Bowditch from England to the Colonies in the 17th century, the family had resided at Salem. Most of its sons, like those of other families in this New England seaport, had gone to sea, and many of them became shipmasters. Nathaniel Bowditch himself sailed as master on his last voyage, and two of his brothers met untimely deaths while pursuing careers at sea.

Nathaniel Bowditch's father is said to have lost two ships at sea, and by late Revolutionary days he returned to the trade of cooper, which he had learned in his youth. This provided insufficient income to properly supply the needs of his growing family, who were often hungry and cold. For many years the nearly destitute family received an annual grant of 15 to 20 dollars from the Salem Marine Society. By the time Nathaniel had reached the age of 10, the family's poverty forced him to leave school and join his father in the cooper's trade to help support the family.

Nathaniel was unsuccessful as a cooper, and when he was about 12 years of age, he entered the first of two ship-chandlery firms by which he was employed. It was during the nearly 10 years he was so employed that his great mind first attracted public attention. From the time he began school Bowditch had an all-consuming interest in learning, particularly mathematics. By his middle teens he was recognized in Salem as an authority on that subject. Salem being primarily a shipping town, most of the inhabitants sooner or later found their way to the ship chandler, and news of the brilliant young clerk spread until eventually it came to the attention of the learned men of his day. Impressed by his desire to educate himself, they supplied him with books that he might learn of the discoveries of other men. Since many of the best books were written by Europeans, Bowditch first taught himself their languages. French, Spanish, Latin, Greek and German were among the two dozen or more languages and dialects he studied during his life. At the age of 16 he began the study of Newton's *Principia*, translating parts of it from the Latin. He even found an error in that classic text, and though lacking the confidence to announce it at the time, he later published his findings and had them accepted by the scientific community.

During the Revolutionary War a privateer out of Beverly, a neighboring town to Salem, had taken as one of its prizes an English vessel which was carrying the philosophical library of a famed Irish scholar, Dr. Richard Kirwan. The books were brought to the Colonies and there bought by a group of educated Salem men who used them to found the Philosophical Library Company, reputed to have been the best library north of Philadelphia at the time. In 1791, when Bowditch was 18, two Harvard-educated ministers, Rev. John Prince and Rev. William Bentley, persuaded the Company to allow Bowditch the use of its library. Encouraged by these two men and a third, Nathan Read, an apothecary and also a Harvard man, Bowditch studied the works of the great men who had preceded him, especially the mathematicians and the astronomers. By the time he became of age, this knowledge, acquired when not working long hours at the chandlery, had made young Nathaniel the outstanding mathematician in the Commonwealth, and perhaps in the country.

In the seafaring town of Salem, Bowditch was drawn to navigation early, learning the subject at the age of 13 from an old British sailor. A year later he began studying surveying, and in 1794 he assisted in a survey of the town. At 15 he devised an almanac reputed to have been of great accuracy. His other youthful accomplishments included the construction of a crude barometer and a sundial.

When Bowditch went to sea at the age of 21, it was as captain's writer and nominal second mate, the officer's berth being offered him because of his reputation as a scholar. Under Captain Henry Prince, the ship *Henry* sailed from Salem in the winter of 1795 on what was to be a year-long voyage to the Ile de Bourbon (now called Reunion) in the Indian Ocean.

Bowditch began his seagoing career when accurate time was not available to the average naval or merchant ship. A reliable marine chronometer had been invented some 60 years before, but the prohibitive cost, plus the long voyages without opportunity to check the error of the timepiece, made the large investment an impractical one. A system of determining longitude by "lunar distance," a method which did not require an accurate timepiece, was known, but this product of the minds of mathematicians and astronomers was so involved as to be beyond the capabilities of the uneducated seamen of that day. Consequently, ships were navigated by a combination of dead reckoning and parallel sailing (a system of sailing north or south to the latitude of the destination and then east or west to the destination). The navigational routine of the time was "lead, log, and lookout."

To Bowditch, the mathematical genius, computation of lunar distances was no mystery, of course, but he recognized the need for an easier method of working them in order to navigate ships more safely and efficiently. Through analysis and observation, he derived a new and simplified formula during his first trip.

John Hamilton Moore's *The Practical Navigator* was the leading navigational text when Bowditch first went to sea, and had been for many years. Early in his first voyage, however, the captain's writer-second mate began turning up errors in Moore's book, and before long he found it necessary to recompute some of the tables he most often used in working his sights. Bowditch recorded the errors he found, and by the end of his second voyage, made in the higher capacity of supercargo, the news of his findings in *The New Practical Navigator* had reached Edmund Blunt, a printer at Newburyport, Mass. At Blunt's request, Bowditch agreed to participate with other learned men in the preparation of an American edition of the thirteenth (1798) edition of Moore's work. The first American edition was published at Newburyport by Blunt in 1799. This edition corrected many of the errors that Moore had included.

Although most of the errors were of little significance to practical navigation because they were errors in the fifth and sixth places of logarithm tables, some errors were significant.The most significant mistake was listing the year 1800 as a leap year in the table of the sun's declination. The consequence was that Moore gave the declination for March 1, 1800, as 7°11'. Since the actual value was 7° 33', the calculation of a meridian altitude would be in error by 22 minutes of latitude, or 22 nautical miles.

Bowditch's principal contribution to the first American edition was his chapter "The Method of Finding the Longitude at Sea," which discussed his new method for computing lunar distances. Following publication of the first American edition, Blunt obtained Bowditch's services in checking the American and English editions for further errors. Blunt then published a second American edition of Moore's thirteenth edition in 1800. When preparing a third American edition for the press, Blunt decided that Bowditch had revised Moore's work to such an extent that Bowditch should be named as author. The title was changed to *The New American Practical Navigator* and the book was published in 1802 as a first edition. Bowditch vowed while writing this edition to "put down in the book nothing I can't teach the crew," and it is said that every member of his crew including the cook could take a lunar observation and plot the ship's position.

Bowditch made a total of five trips to sea, over a period of about nine years, his last as master and part owner of the three-masted *Putnam*. Homeward bound from a 13-month voyage to Sumatra and the Ile de France (now called Mauritius) the *Putnam* approached Salem harbor on December 25, 1803, during a thick fog without having had a celestial observation since noon on the 24th. Relying upon his dead reckoning, Bowditch conned his wooden-hulled ship to the entrance of the rocky harbor, where he had the good fortune to get a momentary glimpse of Eastern Point, Cape Ann, enough to confirm his position. The *Putnam* proceeded in, past such hazards as "Bowditch's Ledge" (named after a great-grandfather who had wrecked his ship on the rock more than a century before) and anchored safely at 1900 that evening. Word of the daring feat, performed when other masters were hove-to outside the harbor, spread along the coast and added greatly to Bowditch's reputation. He was, indeed, the "practical navigator."

His standing as a mathematician and successful shipmaster earned him a well-paid position ashore within a matter of weeks after his last voyage. He was installed as president of a Salem fire and marine insurance company at the age of 30, and during the 20 years he held that position the company prospered. In 1823 he left Salem to take a similar position with a Boston insurance firm, serving that company with equal success until his death.

From the time he finished the "*Navigator*" until 1814, Bowditch's mathematical and scientific pursuits consisted of studies and papers on the orbits of comets, applications of Napier's rules, magnetic variation, eclipses, calculations on tides, and the charting of Salem harbor. In that year, however, he turned to what he considered the greatest work of his life, the translation into English of *Mecanique Celeste*, by Pierre Laplace. *Mecanique Celeste* was a summary of all the then known facts about the workings of the heavens. Bowditch translated four of the five volumes before his death, and published them at his own expense. He gave many formula derivations which Laplace had not shown, and also included further discoveries following the time of publication. His work made this information available to American astronomers and enabled them to pursue their studies on the basis of that which was already known. Continuing his style of writing for the learner, Bowditch presented his English version of *Mecanique Celeste* in such a manner that the student of mathematics could easily trace the steps involved in reaching the most complicated conclusions.

Shortly after the publication of *The New American Practical Navigator*, Harvard College honored its author with the presentation of the honorary degree of Master of Arts, and in 1816 the college made him an honorary Doctor of Laws. From the time the Harvard graduates of Salem first assisted him in his studies, Bowditch had a great interest in that college, and in 1810 he was elected one of its Overseers, a position he held until 1826, when he was elected to the Corporation. During 1826-27 he was the leader of a small group of men who saved the school from financial disaster by forcing necessary economies on the college's reluctant president. At one time Bowditch was offered a Professorship in Mathematics at Harvard but this, as well as similar offers from West Point and the University of Virginia, he declined. In all his life he was never known to have made a public speech or to have addressed any large group of people.

Many other honors came to Bowditch in recognition of his astronomical, mathematical, and marine accomplishments. He became a member of the American Academy of Arts and Sciences, the East India Marine Society, the Royal Academy of Edinburgh, the Royal Society of London, the Royal Irish Academy, the American Philosophical Society, the Connecticut Academy of Arts and Sciences, the Boston Marine Society, the Royal Astronomical Society, the Palermo Academy of Science, and the Royal Academy of Berlin.

Nathaniel Bowditch outlived all of his brothers and sisters by nearly 30 years. He died on March 16, 1838, in his sixty-fifth year. The following eulogy by the Salem Marine Society indicates the regard in which this distinguished American was held by his contemporaries:

"In his death a public, a national, a human benefactor has departed. Not this community, nor our country only, but the whole world, has reason to do honor to his memory. When the voice of Eulogy shall be still, when the tear of Sorrow shall cease to flow, no monument will be needed to keep alive his memory among men; but as long as ships shall sail, the needle point to the north, and the stars go through their wonted courses in the heavens, the name of Dr. Bowditch will be revered as of one who helped his fellow-men in a time of need, who was and is a guide to them over the pathless ocean, and of one who forwarded the great interests of mankind."

THE NEW AMERICAN

PRACTICAL NAVIGATOR;

BEING AN

EPITOME OF NAVIGATION;

CONTAINING ALL THE TABLES NECESSARY TO BE USED WITH THE

NAUTICAL ALMANAC,

IN DETERMINING THE

L A T I T U D E;

AND THE

LONGITUDE BY LUNAR OBSERVATIONS;

AND

KEEPING A COMPLETE RECKONING AT SEA:

ILLUSTRATED BY

PROPER RULES AND EXAMPLES:

THE WHOLE EXEMPLIFIED IN A

J O U R N A L,

KEPT FROM

BOSTON TO MADEIRA,

IN WHICH ALL THE RULES OF NAVIGATION ARE INTRODUCED:

A L S O

The Demonstration of the most useful Rules of Trigonometry: With many useful Problems in Mensuration, Surveying, and Gauging: And a Dictionary of Sea-Terms; with the Manner of performing the most common Evolutions at Sea.

TO WHICH ARE ADDED,

Some General Instructions and Information to Merchants, Masters of Vessels, and others concerned in Navigation, relative to Maritime Laws and Mercantile Customs.

FROM THE BEST AUTHORITIES.

ENRICHED WITH A NUMBER OF

NEW TABLES,

WITH ORIGINAL IMPROVEMENTS AND ADDITIONS, AND A LARGE VARIETY OF NEW AND IMPORTANT MATTER:

A L S O,

MANY THOUSAND ERRORS ARE CORRECTED,

WHICH HAVE APPEARED IN THE BEST SYSTEMS OF NAVIGATION YET PUBLISHED.

BY NATHANIEL BOWDITCH,

FELLOW OF THE AMERICAN ACADEMY OF ARTS AND SCIENCES.

ILLUSTRATED WITH COPPERPLATES.

First Edition.

PRINTED AT NEWBURYPORT, (Mass.) 1802,

BY

EDMUND M. BLUNT, (Proprietor)

For CUSHING & APPLETON, Salem.

SOLD BY EVERY BOOKSELLER, SHIP-CHANDLER, AND MATHEMATICAL INSTRUMENT-MAKER, IN THE UNITED STATES AND WEST-INDIES.

Original title page of *The New American Practical Navigator*, First Edition, published in 1802.

PREFACE

The Naval Observatory library in Washington, D.C., is unnaturally quiet. It is a large circular room, filled with thousands of books. Its acoustics are perfect; a mere whisper from the room's open circular balcony can be easily heard by those standing on the ground floor. A fountain in the center of the ground floor softly breaks the room's silence as its water stream gently splashes into a small pool. From this serene room, a library clerk will lead you into an antechamber, beyond which is a vault containing the Observatory's most rare books. In this vault, one can find an original 1802 first edition of the *New American Practical Navigator*.

One cannot hold this small, delicate, slipcovered book without being impressed by the nearly 200-year unbroken chain of publication that it has enjoyed. It sailed on U.S. merchantmen and Navy ships shortly after the quasi-war with France and during British impressment of merchant seamen that led to the War of 1812. It sailed on U.S. Naval vessels during operations against Mexico in the 1840's, on ships of both the Union and Confederate fleets during the Civil War, and with the U.S. Navy in Cuba in 1898. It went around the world with the Great White Fleet, across the North Atlantic to Europe during both World Wars, to Asia during the Korean and Vietnam Wars, and to the Middle East during Operation Desert Storm. It has circled the globe with countless thousands of merchant ships for 200 years.

As navigational requirements and procedures have changed throughout the years, *Bowditch* has changed with them. Originally devoted almost exclusively to celestial navigation, it now also covers a host of modern topics. It is as practical today as it was when Nathaniel Bowditch, master of the *Putnam*, gathered the crew on deck and taught them the mathematics involved in calculating lunar distances. It is that practicality that has been the publication's greatest strength, and that makes the publication as useful today as it was in the age of sail.

Seafarers have long memories. In no other profession is tradition more closely guarded. Even the oldest and most cynical acknowledge the special bond that connects those who have made their livelihood plying the sea. This bond is not comprised of a single strand; rather, it is a rich and varied tapestry that stretches from the present back to the birth of our nation and its seafaring culture. As this book is a part of that tapestry, it should not be lightly regarded; rather, it should be preserved, as much for its historical importance as for its practical utility.

Since antiquity, mariners have gathered available navigation information and put it into a text for others to follow. One of the first attempts at this involved volumes of Spanish and Portuguese navigational manuals translated into English between about 1550 to 1750. Writers and translators of the time "borrowed" freely in compiling navigational texts, a practice which continues today with works such as Sailing Directions and Pilots.

Colonial and early American navigators depended exclusively on English navigation texts because there were no American editions. The first American navigational text, *Orthodoxal Navigation*, was completed by Benjamin Hubbard in 1656. The first American navigation text published in America was Captain Thomas Truxton's *Remarks, Instructions, and Examples Relating to the Latitude and Longitude; also the Variation of the Compass, Etc., Etc.*, published in 1794.

The most popular navigational text of the late 18th century was John Hamilton Moore's *The New Practical Navigator*. Edmund M. Blunt, a Newburyport publisher, decided to issue a revised copy of this work for American navigators. Blunt convinced Nathaniel Bowditch, a locally famous mariner and mathematician, to revise and update *The New Practical Navigator*. Several other learned men assisted in this revision. Blunt's *The New Practical Navigator* was published in 1799. Blunt also published a second American edition of Moore's book in 1800.

By 1802, when Blunt was ready to publish a third edition, Nathaniel Bowditch and others had corrected so many errors in Moore's work that Blunt decided to issue the work as a first edition of the *New American Practical Navigator*. It is to that 1802 work that the current edition of the *American Practical Navigator* traces its pedigree.

The *New American Practical Navigator* stayed in the Bowditch and Blunt family until the government bought the copyright in 1867. Edmund M. Blunt published the book until 1833; upon his retirement, his sons, Edmund and George, took over publication.The elder Blunt died in 1862; his son Edmund followed in 1866. The next year, 1867, George Blunt sold the copyright to the government for $25,000. The government has published *Bowditch* ever since. George Blunt died in 1878.

Nathaniel Bowditch continued to correct and revise the book until his death in 1838. Upon his death, the editorial responsibility for the *American Practical Navigator* passed to his son, J. Ingersoll Bowditch. Ingersoll Bowditch continued editing the *Navigator* until George Blunt sold the copyright to the government. He outlived all of the principals involved in publishing and editing the *Navigator*, dying in 1889.

The U.S. government has published some 52 editions since acquiring the copyright to the book that has come to be known simply by its original author's name, "*Bowditch*." Since the government began production, the book has been known by its year of publishing, instead of by the edition number. During a revision in 1880 by Commander Phillip H. Cooper, USN, the name was changed to *American Practical Navigator*. Bowditch's original method of taking "lunars" was finally dropped from the book just after the turn of the 20th century. After several more revisions and printings through World Wars I and II, *Bowditch* was extensively revised for the 1958 edition and again in 1995.

Recognizing the limitations of the printed word, and that computers and electronic media permit us to think about the processes of both navigation and publishing in completely new ways, NIMA has, for the 2002 edition, produced the first official Compact Disk-Read Only Memory (CD-ROM) version of this work. This CD contains, in addition to the full text of the printed book, electronic enhancements and additions not possible in book form. Our goal is to put as much useful navigational information before the navigator as possible in the most understandable and readable format. We are only beginning to explore the possibilities of new technology in this area.

As much as it is a part of history, Bowditch is not a history book. As in past editions, dated material has been dropped and new methods, technologies and techniques added to keep pace with the rapidly changing world of navigation. The changes to this edition are intended to ensure that it remains the premier reference work for modern, practical marine navigation. This edition replaces but does not cancel former editions, which may be retained and consulted as to historical navigation methods not discussed herein.

PART 1, FUNDAMENTALS, includes an overview of the types and phases of marine navigation and the organizations which develop, support and regulate it. It includes chapters relating to the types, structure, use and limitations of nautical charts; a concise explanation of geodesy and chart datums; and a summary of various necessary navigational publications.

PART 2, PILOTING, emphasizes the practical aspects of navigating a vessel in restricted waters, using both traditional and electronic methods.

PART 3, ELECTRONIC NAVIGATION, explains the nature of radio waves and electronic navigation systems. Chapters deal with each of the several electronic methods of navigation--satellite, Loran C, and radar, with special emphasis on satellite navigation systems and electronic charts.

PART 4, CELESTIAL NAVIGATION, updates the former edition with more modern terminology, and discusses the use of calculators and computers for the solution of celestial navigation problems.

PART 5, NAVIGATIONAL MATHEMATICS, remains unchanged from the former edition.

PART 6, NAVIGATIONAL SAFETY, discusses recent developments in management of navigational resources, the changing role of the navigator, distress and safety communications, procedures for emergency navigation, and the increasingly complex web of navigation regulations.

PART 7, OCEANOGRAPHY, has been updated to reflect the latest science and terminology.

PART 8, MARINE WEATHER incorporates updated weather routing information and new cloud graphics.

The pronoun "he," used throughout this book as a reference to the navigator, refers to both genders.

The printed version of this volume may be corrected using the Notice to Mariners and Summary of Corrections. Suggestions and comments for changes and additions may be sent to:

NATIONAL IMAGERY AND MAPPING AGENCY
MARITIME SAFETY INFORMATION DIVISION
MAIL STOP D-44
4600 SANGAMORE RD.
BETHESDA, MARYLAND, 20816-5003
UNITED STATES OF AMERICA

This book could not have been produced without the expertise of dedicated personnel from many government organizations, among them: U.S. Coast Guard, U.S. Naval Academy, U.S. Naval Oceanographic Office, US Navy Fleet Training Center, the U.S. Naval Observatory, Office of the Navigator of the Navy, U.S. Merchant Marine Academy, U.S. Coast and Geodetic Survey, the National Ocean Service, and the National Weather Service. In addition to official government expertise, we must note the contributions of private organizations and individuals far too numerous to mention. Mariners worldwide can be grateful for the experience, dedication, and professionalism of the many people who generously gave their time in this effort. A complete list of contributors can be found in the "Contributor's Corner" of the CD-ROM version of this book.

THE EDITORS

TABLE OF CONTENTS

NATHANIEL BOWDITCH iii
PREFACE vii

PART 1 — FUNDAMENTALS

CHAPTER 1. INTRODUCTION TO MARINE NAVIGATION 1
CHAPTER 2. GEODESY AND DATUMS IN NAVIGATION 15
CHAPTER 3. NAUTICAL CHARTS 23
CHAPTER 4. NAUTICAL PUBLICATIONS 51

PART 2 — PILOTING

CHAPTER 5. SHORT RANGE AIDS TO NAVIGATION 63
CHAPTER 6. COMPASSES 81
CHAPTER 7. DEAD RECKONING 99
CHAPTER 8. PILOTING 105
CHAPTER 9. TIDES AND TIDAL CURRENTS 129

PART 3 — ELECTRONIC NAVIGATION

CHAPTER 10. RADIO WAVES 151
CHAPTER 11. SATELLITE NAVIGATION 163
CHAPTER 12. LORAN NAVIGATION 173
CHAPTER 13. RADAR NAVIGATION 187
CHAPTER 14. ELECTRONIC CHARTS 199

PART 4 — CELESTIAL NAVIGATION

CHAPTER 15. NAVIGATIONAL ASTRONOMY 217
CHAPTER 16. INSTRUMENTS FOR CELESTIAL NAVIGATION 261
CHAPTER 17. AZIMUTHS AND AMPLITUDES 271
CHAPTER 18. TIME 275
CHAPTER 19. THE ALMANACS 287
CHAPTER 20. SIGHT REDUCTION 295

PART 5 — NAVIGATIONAL MATHEMATICS

CHAPTER 21. NAVIGATIONAL MATHEMATICS 317
CHAPTER 22. CALCULATIONS AND CONVERSIONS 329
CHAPTER 23. NAVIGATIONAL ERRORS 341
CHAPTER 24. THE SAILINGS 345

PART 6 — NAVIGATIONAL SAFETY

CHAPTER 25. NAVIGATION PROCESSES 363
CHAPTER 26. EMERGENCY NAVIGATION 373
CHAPTER 27. NAVIGATION REGULATIONS 383
CHAPTER 28. MARITIME SAFETY SYSTEMS 393
CHAPTER 29. HYDROGRAPHY 409

PART 7 — OCEANOGRAPHY

CHAPTER 30. THE OCEANS 425
CHAPTER 31. OCEAN CURRENTS 433
CHAPTER 32. WAVES, BREAKERS AND SURF 441
CHAPTER 33. ICE NAVIGATION 453

PART 8 — MARINE METEOROLOGY

CHAPTER 34. WEATHER ELEMENTS 481
CHAPTER 35. TROPICAL CYCLONES 503
CHAPTER 36. WEATHER OBSERVATIONS 519
CHAPTER 37. WEATHER ROUTING 545

NAVIGATION TABLES

EXPLANATION OF NAVIGATION TABLES 557

MATHEMATICAL TABLES

TABLE 1. LOGARITHMS OF NUMBERS 565
TABLE 2. NATURAL TRIGONOMETRIC FUNCTIONS 575
TABLE 3. COMMON LOGARITHMS OF TRIGONOMETRIC FUNCTIONS 598
TABLE 4. TRAVERSE TABLES 621

CARTOGRAPHIC TABLES

TABLE 5. NATURAL AND NUMERICAL CHART SCALES 666
TABLE 6. MERIDIONAL PARTS 667
TABLE 7. LENGTH OF A DEGREE OF LATITUDE AND LONGITUDE 672

PILOTING TABLES

TABLE 8. CONVERSION TABLE FOR METERS, FEET, AND FATHOMS 673
TABLE 9. CONVERSION TABLE FOR NAUTICAL AND STATUTE MILES 674
TABLE 10. SPEED TABLE FOR MEASURED MILE 675
TABLE 11. SPEED, TIME, AND DISTANCE 676
TABLE 12. DISTANCE OF THE HORIZON 679
TABLE 13. GEOGRAPHIC RANGE 680
TABLE 14. DIP OF THE SEA SHORT OF THE HORIZON 682
TABLE 15. DISTANCE BY VERTICAL ANGLE MEASURED BETWEEN SEA HORIZON AND TOP OF OBJECT BEYOND SEA HORIZON 683
TABLE 16. DISTANCE BY VERTICAL ANGLE MEASURED BETWEEN WATERLINE AT OBJECT AND TOP OF OBJECT 685
TABLE 17. DISTANCE BY VERTICAL ANGLE MEASURED BETWEEN WATERLINE AT OBJECT AND SEA HORIZON BEYOND OBJECT 687
TABLE 18. DISTANCE OF AN OBJECT BY TWO BEARINGS 688

CELESTIAL NAVIGATION TABLES

TABLE 19. TABLE OF OFFSETS 691
TABLE 20. MERIDIAN ANGLE AND ALTITUDE OF A BODY ON THE PRIME VERTICAL CIRCLE 692
TABLE 21. LATITUDE AND LONGITUDE FACTORS 694
TABLE 22. AMPLITUDES 698
TABLE 23. CORRECTION OF AMPLITUDE AS OBSERVED ON THE VISIBLE HORIZON 700
TABLE 24. ALTITUDE FACTORS 701
TABLE 25. CHANGE OF ALTITUDE IN GIVEN TIME FROM MERIDIAN TRANSIT 706
TABLE 26. TIME ZONES, ZONE DESCRIPTIONS, AND SUFFIXES 708
TABLE 27. ALTITUDE CORRECTION FOR AIR TEMPERATURE 709
TABLE 28. ALTITUDE CORRECTION FOR ATMOSPHERIC PRESSURE 709

METEOROLOGICAL TABLES

TABLE 29. CONVERSION TABLE FOR THERMOMETER SCALES 710
TABLE 30. DIRECTION AND SPEED OF TRUE WIND IN UNITS OF SHIP'S SPEED 711
TABLE 31. CORRECTION OF BAROMETER READING FOR HEIGHT ABOVE SEA LEVEL 712
TABLE 32. CORRECTION OF BAROMETER READING FOR GRAVITY 712
TABLE 33. CORRECTION OF BAROMETER READING FOR TEMPERATURE 712
TABLE 34. CONVERSION TABLE FOR MILLIBARS, INCHES, AND MILLIMETERS OF MERCURY 713
TABLE 35. RELATIVE HUMIDITY 714
TABLE 36. DEW POINT 715

GLOSSARIES

GLOSSARY OF MARINE NAVIGATION 717
GLOSSARY OF ABBREVIATIONS AND ACRONYMS 855

INDEX

863-879

CHAPTER 1

INTRODUCTION TO MARINE NAVIGATION

DEFINITIONS

100. The Art And Science Of Navigation

Marine navigation blends both science and art. A good navigator constantly thinks strategically, operationally, and tactically. He plans each voyage carefully. As it proceeds, he gathers navigational information from a variety of sources, evaluates this information, and determines his ship's position. He then compares that position with his voyage plan, his operational commitments, and his pre-determined "dead reckoning" position. A good navigator anticipates dangerous situations well before they arise, and always stays "ahead of the vessel." He is ready for navigational emergencies at any time. He is increasingly a manager of a variety of resources--electronic, mechanical, and human. Navigation methods and techniques vary with the type of vessel, the conditions, and the navigator's experience. The navigator uses the methods and techniques best suited to the vessel, its equipment, and conditions at hand.

Some important elements of successful navigation cannot be acquired from any book or instructor. The science of navigation can be taught, but the art of navigation must be developed from experience.

101. Types of Navigation

Methods of navigation have changed throughout history. New methods often enhance the mariner's ability to complete his voyage safely and expeditiously, and make his job easier. One of the most important judgments the navigator must make involves choosing the best methods to use. Each method or type has advantages and disadvantages, while none is effective in all situations. Commonly recognized types of navigation are listed below.

- **Dead reckoning (DR)** determines position by advancing a known position for courses and distances. A position so determined is called a dead reckoning (DR) position. It is generally accepted that only course and speed determine the DR position. Correcting the DR position for leeway, current effects, and steering error result in an **estimated position (EP)**.

- **Piloting** involves navigating in restricted waters with frequent or constant determination of position relative to nearby geographic and hydrographic features.

- **Celestial navigation** involves reducing celestial measurements taken with a sextant to lines of position using calculators or computer programs, or by hand with almanacs and tables or using spherical trigonometry.

- **Radio navigation** uses radio waves to determine position through a variety of electronic devices.

- **Radar navigation** uses radar to determine the distance from or bearing of objects whose position is known. This process is separate from radar's use in collision avoidance.

- **Satellite navigation** uses radio signals from satellites for determining position.

Electronic systems and integrated bridge concepts are driving navigation system planning. Integrated systems take inputs from various ship sensors, electronically and automatically chart the position, and provide control signals required to maintain a vessel on a preset course. The navigator becomes a system manager, choosing system presets, interpreting system output, and monitoring vessel response.

In practice, a navigator synthesizes different methodologies into a single integrated system. He should never feel comfortable utilizing only one method when others are also available. Each method has advantages and disadvantages. The navigator must choose methods appropriate to each situation, and never rely completely on only one system.

With the advent of automated position fixing and electronic charts, modern navigation is almost completely an electronic process. The mariner is constantly tempted to rely solely on electronic systems. But electronic navigation systems are always subject to failure, and the professional mariner must never forget that the safety of his ship and crew may depend on skills that differ little from those practiced generations ago. Proficiency in conventional piloting and celestial navigation remains essential.

102. Phases of Navigation

Four distinct phases define the navigation process. The mariner should choose the system mix that meets the accuracy requirements of each phase.

- **Inland Waterway Phase**: Piloting in narrow canals, channels, rivers, and estuaries.
- **Harbor/Harbor Approach Phase**: Navigating to a harbor entrance through bays and sounds, and negotiating harbor approach channels.
- **Coastal Phase**: Navigating within 50 miles of the coast or inshore of the 200 meter depth contour.
- **Ocean Phase**: Navigating outside the coastal area in the open sea.

The navigator's position accuracy requirements, his fix interval, and his systems requirements differ in each phase. The following table can be used as a general guide for selecting the proper system(s).

	Inland	*Harbor/ Approach*	*Coastal*	*Ocean*
DR	X	X	X	X
Piloting	X	X	X	
Celestial			X	X
Radio		X	X	X
Radar	X	X	X	
Satellite	X*	X	X	X

*Table 102. The relationship of the types and phases of navigation. * With SA off and/or using DGPS*

NAVIGATION TERMS AND CONVENTIONS

103. Important Conventions and Concepts

Throughout the history of navigation, numerous terms and conventions have been established which enjoy worldwide recognition. The professional navigator, to gain a full understanding of his field, should understand the origin of certain terms, techniques, and conventions. The following section discusses some of the important ones.

Defining a **prime meridian** is a comparatively recent development. Until the beginning of the 19th century, there was little uniformity among cartographers as to the meridian from which to measure longitude. But it mattered little because there existed no method for determining longitude accurately.

Ptolemy, in the 2nd century AD, measured longitude eastward from a reference meridian 2 degrees west of the Canary Islands. In 1493, Pope Alexander VI established a line in the Atlantic west of the Azores to divide the territories of Spain and Portugal. For many years, cartographers of these two countries used this dividing line as the prime meridian. In 1570 the Dutch cartographer Ortelius used the easternmost of the Cape Verde Islands. John Davis, in his 1594 *The Seaman's Secrets*, used the Isle of Fez in the Canaries because there the variation was zero. Most mariners paid little attention to these conventions and often reckoned their longitude from several different capes and ports during a voyage.

The meridian of London was used as early as 1676, and over the years its popularity grew as England's maritime interests increased. The system of measuring longitude both east and west through 180° may have first appeared in the middle of the 18th century. Toward the end of that century, as the Greenwich Observatory increased in prominence, English cartographers began using the meridian of that observatory as a reference. The publication by the Observatory of the first British *Nautical Almanac* in 1767 further entrenched Greenwich as the prime meridian. An unsuccessful attempt was made in 1810 to establish Washington, D.C. as the prime meridian for American navigators and cartographers. In 1884, the meridian of Greenwich was officially established as the prime meridian. Today, all maritime nations have designated the Greenwich meridian the prime meridian, except in a few cases where local references are used for certain harbor charts.

Charts are graphic representations of areas of the Earth, in digital or graphic form, for use in marine or air navigation. Nautical charts, whether in digital or paper form, depict features of particular interest to the marine navigator. Charts have probably existed since at least 600 B.C. Stereographic and orthographic projections date from the 2nd century B.C. In 1569 Gerardus Mercator published a chart using the mathematical principle which now bears his name. Some 30 years later, Edward Wright published corrected mathematical tables for this projection, enabling other cartographers to produce charts on the Mercator projection. This projection is still the most widely used.

Sailing Directions or ***pilots*** have existed since at least the 6th century B.C. Continuous accumulation of navigational data, along with increased exploration and trade, led to increased production of volumes through the Middle Ages. "Routiers" were produced in France about 1500; the English referred to them as "rutters." In 1584 Lucas Waghenaer published the *Spieghel der Zeevaerdt (The Mariner's Mirror)*, which became the model for such publications for several generations of navigators. They were known as "Waggoners" by most sailors.

The **compass** was developed about 1000 years ago. The origin of the magnetic compass is uncertain, but

Norsemen used it in the 11th century, and Chinese navigators used the magnetic compass at least that early and probably much earlier. It was not until the 1870s that Lord Kelvin developed a reliable dry card marine compass. The fluid-filled compass became standard in 1906.

Variation was not understood until the 18th century, when Edmond Halley led an expedition to map lines of variation in the South Atlantic. **Deviation** was understood at least as early as the early 1600s, but adequate correction of compass error was not possible until Matthew Flinders discovered that a vertical iron bar could reduce certain types of errors. After 1840, British Astronomer Royal Sir George Airy and later Lord Kelvin developed combinations of iron masses and small magnets to eliminate most magnetic compass error.

The **gyrocompass** was made necessary by iron and steel ships. Leon Foucault developed the basic gyroscope in 1852. An American (Elmer Sperry) and a German (Anshutz Kampfe) both developed electrical gyrocompasses in the early years of the 20th century. Ring laser gyrocompasses and digital flux gate compasses are gradually replacing traditional gyrocompasses, while the magnetic compass remains an important backup device.

The **log** is the mariner's speedometer. Mariners originally measured speed by observing a chip of wood passing down the side of the vessel. Later developments included a wooden board attached to a reel of line. Mariners measured speed by noting how many knots in the line unreeled as the ship moved a measured amount of time; hence the term **knot.** Mechanical logs using either a small paddle wheel or a rotating spinner arrived about the middle of the 17th century. The taffrail log still in limited use today was developed in 1878. Modern logs use electronic sensors or spinning devices that induce small electric fields proportional to a vessel's speed. An engine revolution counter or shaft log often measures speed aboard large ships. Doppler speed logs are used on some vessels for very accurate speed readings. Inertial and satellite systems also provide highly accurate speed readings.

The Metric Conversion Act of 1975 and the Omnibus Trade and Competitiveness Act of 1988 established the **metric system** of weights and measures in the United States. As a result, the government is converting charts to the metric format. Notwithstanding the conversion to the metric system, the common measure of distance at sea is the **nautical mile**.

The current policy of the National Imagery and Mapping Agency (NIMA) and the National Ocean Service (NOS) is to convert new compilations of nautical, special purpose charts, and publications to the metric system. All digital charts use the metric system. This conversion began on January 2, 1970. Most modern maritime nations have also adopted the meter as the standard measure of depths and heights. However, older charts still on issue and the charts of some foreign countries may not conform to this standard.

The **fathom** as a unit of length or depth is of obscure origin. Posidonius reported a sounding of more than 1,000 fathoms in the 2nd century B.C. How old the unit was then is unknown. Many modern charts are still based on the fathom, as conversion to the metric system continues.

The sailings refer to various methods of mathematically determining course, distance, and position. They have a history almost as old as mathematics itself. Thales, Hipparchus, Napier, Wright, and others contributed the formulas that permit computation of course and distance by plane, traverse, parallel, middle latitude, Mercator, and great circle sailings.

104. The Earth

The Earth is an irregular oblate spheroid (a sphere flattened at the poles). Measurements of its dimensions and the amount of its flattening are subjects of geodesy. However, for most navigational purposes, assuming a spherical Earth introduces insignificant error. The Earth's axis of rotation is the line connecting the north and south geographic poles.

A great circle is the line of intersection of a sphere and a plane through its center. This is the largest circle that can be drawn on a sphere. The shortest line on the surface of a sphere between two points on the surface is part of a great circle. On the spheroidal Earth the shortest line is called a geodesic. A great circle is a near enough approximation to

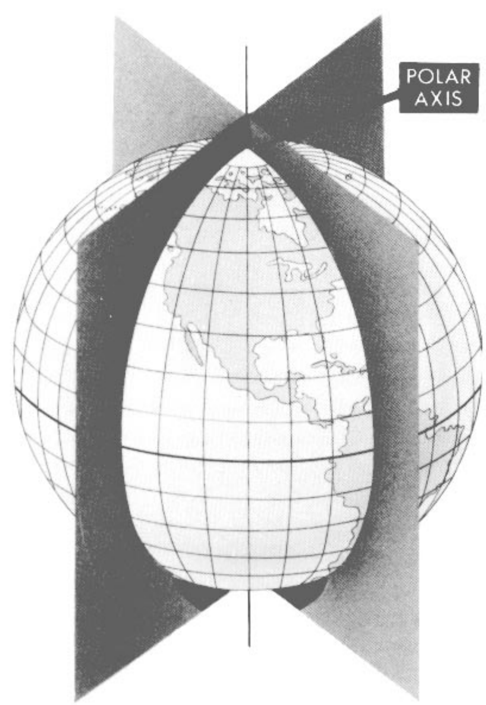


Figure 104a. The planes of the meridians at the polar axis.

a geodesic for most problems of navigation. A small circle is the line of intersection of a sphere and a plane which does not pass through the center. See Figure 104a.

The term **meridian** is usually applied to the **upper branch** of the half-circle from pole to pole which passes through a given point. The opposite half is called the **lower branch.**

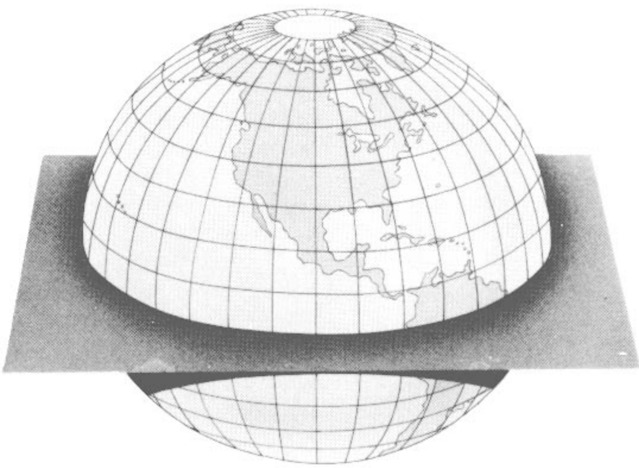


Figure 104b. The equator is a great circle midway between the poles.

A **parallel** or parallel of latitude is a circle on the surface of the Earth parallel to the plane of the equator. It connects all points of equal latitude. The equator is a great circle at latitude 0°. See Figure 104b. The poles are single points at latitude 90°. All other parallels are small circles.

105. Coordinates

Coordinates of latitude and longitude can define any position on Earth. **Latitude (L, lat.)** is the angular distance from the equator, measured northward or southward along a meridian from 0° at the equator to 90° at the poles. It is designated north (N) or south (S) to indicate the direction of measurement.

The **difference of latitude (*l*, DLat.)** between two places is the angular length of arc of any meridian between their parallels. It is the numerical difference of the latitudes if the places are on the same side of the equator; it is the sum of the latitudes if the places are on opposite sides of the equator. It may be designated north (N) or south (S) when appropriate. The middle or **mid-latitude (Lm)** between two places on the same side of the equator is half the sum of their latitudes. Mid-latitude is labeled N or S to indicate whether it is north or south of the equator.

The expression may refer to the mid-latitude of two places on opposite sides of the equator. In this case, it is equal to half the difference between the two latitudes and takes the name of the place farthest from the equator.

Longitude (l, long.) is the angular distance between the prime meridian and the meridian of a point on the Earth, measured eastward or westward from the prime meridian through 180°. It is designated east (E) or west (W) to indicate the direction of measurement.

The **difference of longitude (DLo)** between two places is the shorter arc of the parallel or the smaller angle at the pole between the meridians of the two places. If both places are on the same side (east or west) of Greenwich, DLo is the numerical difference of the longitudes of the two places; if on opposite sides, DLo is the numerical sum unless this exceeds 180°, when it is 360° minus the sum.

The distance between two meridians at any parallel of latitude, expressed in distance units, usually nautical miles, is called **departure (p, Dep.)**. It represents distance made good east or west as a craft proceeds from one point to another. Its numerical value between any two meridians decreases with increased latitude, while DLo is numerically the same at any latitude. Either DLo or p may be designated east (E) or west (W) when appropriate.

106. Distance on the Earth

Distance, as used by the navigator, is the length of the **rhumb line** connecting two places. This is a line making the same angle with all meridians. Meridians and parallels which also maintain constant true directions may be considered special cases of the rhumb line. Any other rhumb line spirals toward the pole, forming a **loxodromic curve** or **loxodrome**. See Figure 106. Distance along the great

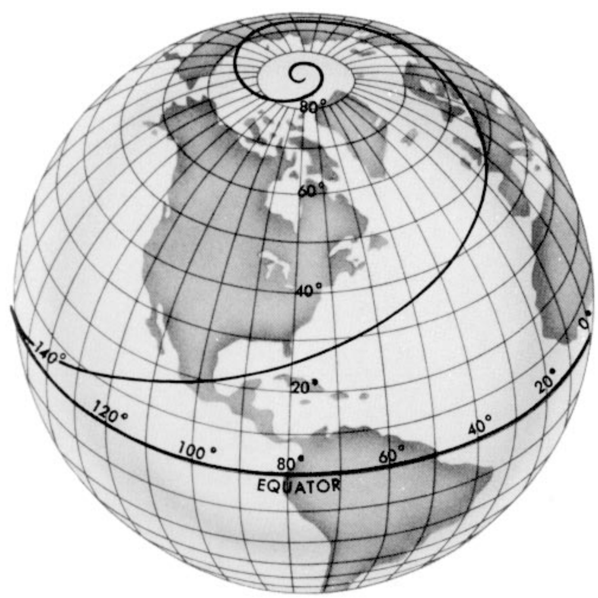


Figure 106. A loxodrome.

circle connecting two points is customarily designated **great-circle distance**. For most purposes, considering the nautical mile the length of one minute of latitude introduces no significant error

Speed (S) is rate of motion, or distance per unit of time. A **knot (kn.)**, the unit of speed commonly used in navigation, is a rate of 1 nautical mile per hour. The expression **speed of advance (SOA)** is used to indicate the speed to be made along the intended track. **Speed over the ground (SOG)** is the actual speed of the vessel over the surface of the Earth at any given time. To calculate **speed made good (SMG)** between two positions, divide the distance between the two positions by the time elapsed between the two positions.

107. Direction on the Earth

Direction is the position of one point relative to another. Navigators express direction as the angular difference in degrees from a reference direction, usually north or the ship's head. **Course (C, Cn)** is the horizontal direction in which a vessel is intended to be steered, expressed as angular distance from north clockwise through 360°. Strictly used, the term applies to direction through the water, not the direction intended to be made good over the ground.The course is often designated as true, magnetic, compass, or grid according to the reference direction.

Track made good (TMG) is the single resultant direction from the point of departure to point of arrival at any given time. **Course of advance (COA)** is the direction intended to be made good over the ground, and **course over ground (COG)** is the direction between a vessel's last fix and an EP. A **course line** is a line drawn on a chart extending in the direction of a course. It is sometimes convenient to express a course as an angle from either north or south, through 90° or 180°. In this case it is designated course angle (C) and should be properly labeled to indicate the origin (prefix) and direction of measurement (suffix). Thus, C N35°E = Cn 035° (000° + 35°), C N155°W = Cn 205° (360° - 155°), C S47°E = Cn 133° (180° - 47°). But Cn 260° may be either C N100°W or C S80°W, depending upon the conditions of the problem.

Track (TR) is the intended horizontal direction of travel with respect to the Earth. The terms intended track and trackline are used to indicate the path of intended travel. See Figure 107a. The track consists of one or a series of course lines, from the point of departure to the destination, along which one intends to proceed. A great circle which a vessel intends to follow is called a **great-circle track**, though it consists of a series of straight lines approximating a great circle

Heading (Hdg., SH) is the direction in which a vessel is pointed at any given moment, expressed as angular distance from 000° clockwise through 360°. It is easy to confuse heading and course. Heading constantly changes as a vessel yaws back and forth across the course due to sea, wind, and steering error.

Bearing (B, Brg.) is the direction of one terrestrial point from another, expressed as angular distance from 000° (North) clockwise through 360°. When measured through 90° or 180° from either north or south, it is called bearing angle (B). Bearing and azimuth are sometimes used interchangeably, but the latter more accurately refers to the horizontal direction of a point on the celestial sphere from a point on the Earth. A relative bearing is measured relative to the ship's heading from 000° (dead ahead) clockwise through 360°. However, it is sometimes conveniently measured right or left from 000° at the ship's head through 180°. This is particularly true when using the table for Distance of an Object by Two Bearings.

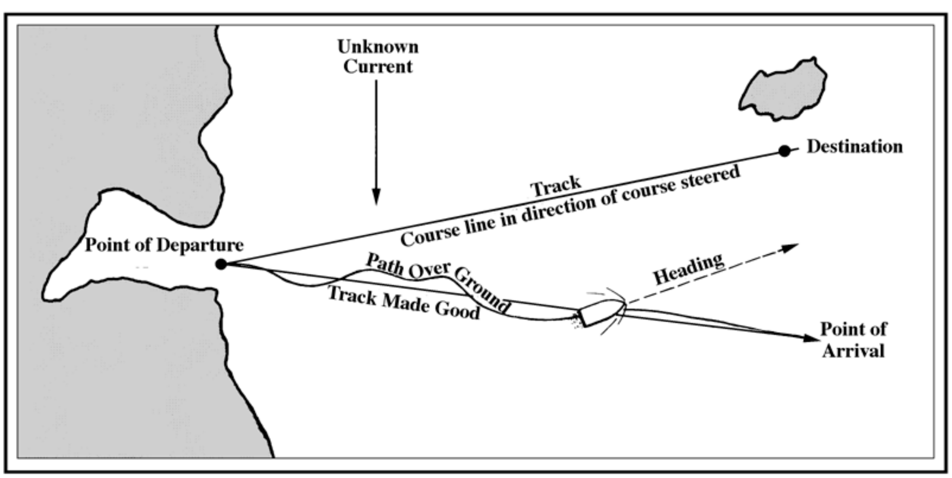


Figure 107a. Course line, track, track made good, and heading.

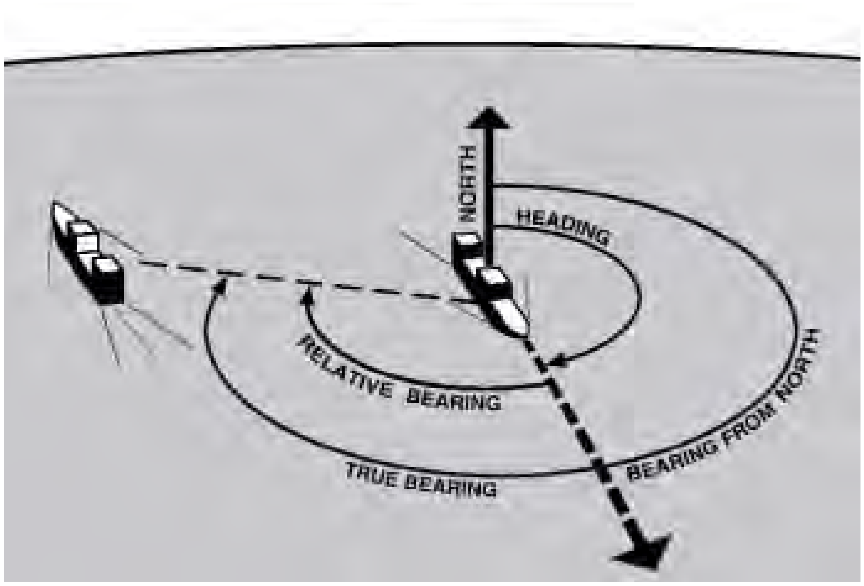


Figure 107b. Relative Bearing

To convert a relative bearing to a true bearing, add the true heading. See Figure 107b

True Bearing = Relative Bearing + True Heading.
Relative Bearing = True Bearing - True Heading.

108. Finding Latitude and Longitude

Navigators have made latitude observations for thousands of years. Accurate declination tables for the Sun have been published for centuries, enabling ancient seamen to compute latitude to within 1 or 2 degrees. Those who today determine their latitude by measuring the Sun at their meridian and the altitude of Polaris are using methods well known to 15th century navigators.

A method of finding longitude eluded mariners for centuries. Several solutions independent of time proved too cumbersome. Finding longitude by magnetic variation was tried, but found too inaccurate. The lunar distance method, which determines GMT by observing the Moon's position among the stars, became popular in the 1800s. However, the mathematics required by most of these processes were far above the abilities of the average seaman. It was apparent that the solution lay in keeping accurate time at sea.

In 1714, the British Board of Longitude was formed, offering a small fortune in reward to anyone who could provide a solution to the problem.

An Englishman, John Harrison, responded to the challenge, developing four chronometers between 1735 and 1760. The most accurate of these timepieces lost only 15 seconds on a 156 day round trip between London and Barbados. The Board, however, paid him only half the promised reward. The King finally intervened on Harrison's behalf, and at the age of 80 years Harrison received his full reward of £20,000.

Rapid chronometer development led to the problem of determining **chronometer error** aboard ship. **Time balls**, large black spheres mounted in port in prominent locations, were dropped at the stroke of noon, enabling any ship in harbor which could see the ball to determine chronometer error. By the end of the U.S. Civil War, telegraph signals were being used to key time balls. Use of radio signals to send time ticks to ships well offshore began in 1904, and soon worldwide signals were available.

109. The Navigational Triangle

Modern celestial navigators reduce their celestial observations by solving a **navigational triangle** whose points are the elevated pole, the celestial body, and the zenith of the observer. The sides of this triangle are the polar distance of the body (**codeclination**), its zenith distance (**coaltitude**), and the polar distance of the zenith (**colatitude** of the observer).

A spherical triangle was first used at sea in solving **lunar distance** problems. Simultaneous observations were made of the altitudes of the Moon and the Sun or a star near the ecliptic and the angular distance between the Moon and the other body. The zenith of the observer and the two celestial bodies formed the vertices of a triangle whose sides were the two coaltitudes and the angular distance between the bodies. Using a mathematical calculation the navigator "cleared" this distance of the effects of refraction and parallax applicable to each altitude. This corrected value was then used as an argument for entering the almanac. The almanac gave the true lunar distance from the Sun and several stars at 3 hour intervals. Previously, the navigator had set his watch or checked its error and rate with the local mean time determined by celestial observations. The local mean time of the watch, properly corrected, applied to the Greenwich mean time obtained from the lunar distance observation, gave the longitude.

The calculations involved were tedious. Few mariners could solve the triangle until Nathaniel Bowditch published his simplified method in 1802 in *The New American Practical Navigator*.

Reliable chronometers were available by1800, but their high cost precluded their general use aboard most ships. However, most navigators could determine their longitude using Bowditch's method. This eliminated the need for parallel sailing and the lost time associated with it. Tables for the lunar distance solution were carried in the American nautical almanac into the 20th century.

110. The Time Sight

The theory of the **time sight** had been known to mathematicians since the development of spherical trigonometry, but not until the chronometer was developed could it be used by mariners.

The time sight used the modern navigational triangle. The codeclination, or polar distance, of the body could be determined from the almanac. The zenith distance (coaltitude) was determined by observation. If the colatitude were known, three sides of the triangle were available. From these the meridian angle was computed. The comparison of this with the Greenwich hour angle from the almanac yielded the longitude.

The time sight was mathematically sound, but the navigator was not always aware that the longitude determined was only as accurate as the latitude, and together they merely formed a point on what is known today as a **line of position**. If the observed body was on the prime vertical, the line of position ran north and south and a small error in latitude generally had little effect on the longitude. But when the body was close to the meridian, a small error in latitude produced a large error in longitude.

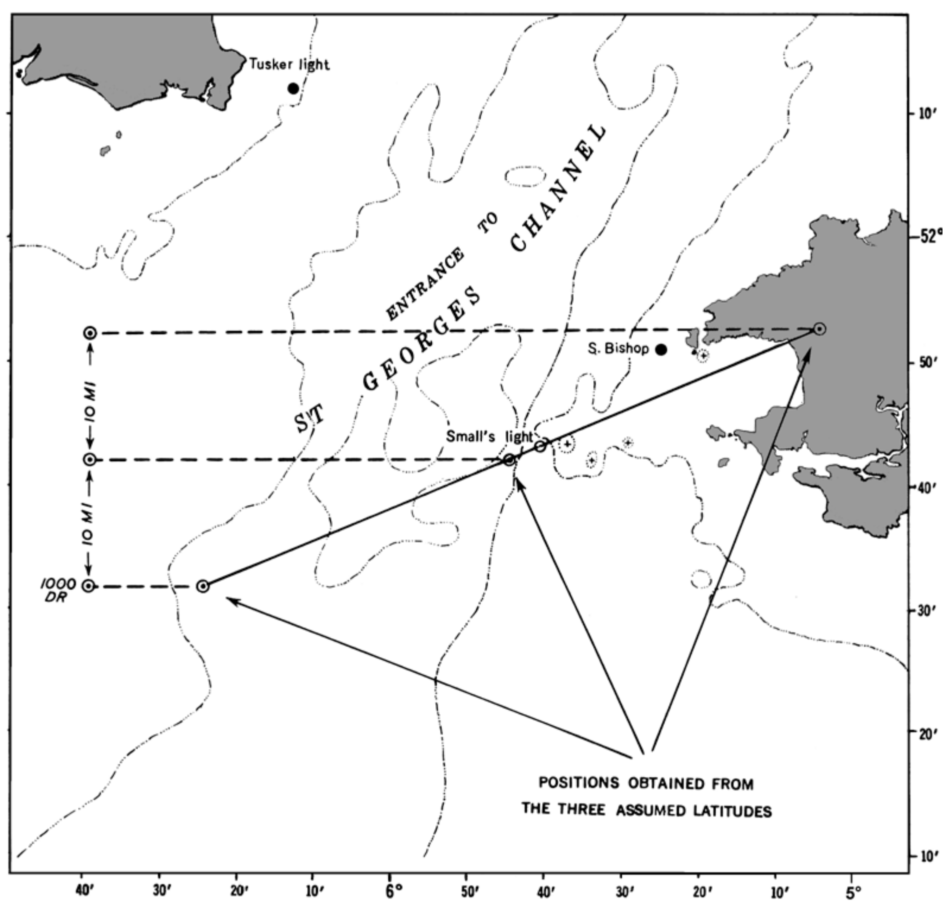


Figure 110. The first celestial line of position, obtained by Captain Thomas Sumner in 1837.

The line of position by celestial observation was unknown until discovered in 1837 by 30-year-old Captain Thomas H. Sumner, a Harvard graduate and son of a United States congressman from Massachusetts. The discovery of the "**Sumner line**," as it is sometimes called, was considered by Maury "the commencement of a new era in practical navigation." This was the turning point in the development of modern celestial navigation technique. In Sumner's own words, the discovery took place in this manner:

> *Having sailed from Charleston, S. C., 25th November, 1837, bound to Greenock, a series of heavy gales from the Westward promised a quick passage; after passing the Azores, the wind prevailed from the Southward, with thick weather; after passing Longitude 21° W, no observation was had until near the land; but soundings were had not far, as was supposed, from the edge of the Bank. The weather was now more boisterous, and very thick; and the wind still Southerly; arriving about midnight, 17th December, within 40 miles, by dead reckoning, of Tusker light; the wind hauled SE, true, making the Irish coast a lee shore; the ship was then kept close to the wind, and several tacks made to preserve her position as nearly as possible until daylight; when nothing being in sight, she was kept on ENE under short sail, with heavy gales; at about 10 AM an altitude of the Sun was observed, and the Chronometer time noted; but, having run so far without any observation, it was plain the Latitude by dead reckoning was liable to error, and could not be entirely relied on. Using, however, this Latitude, in finding the Longitude by Chronometer, it was found to put the ship 15' of Longitude E from her position by dead reckoning; which in Latitude 52° N is 9 nautical miles; this seemed to agree tolerably well with the dead reckoning; but feeling doubtful of the Latitude, the observation was tried with a Latitude 10' further N, finding this placed the ship ENE 27 nautical miles, of the former position, it was tried again with a Latitude 20' N of the dead reckoning; this also placed the ship still further ENE, and still 27 nautical miles further; these three positions were then seen to lie in the direction of Small's light. It then at once appeared that the observed altitude must have happened at all the three points, and at Small's light, and at the ship, at the same instant of time; and it followed, that Small's light must bear ENE, if the Chronometer was right. Having been convinced of this truth, the ship was kept on her course, ENE, the wind being still SE., and in less than an hour, Small's light was made bearing ENE 1/2 E, and close aboard.*

In 1843 Sumner published a book, A *New and Accurate Method of Finding a Ship's Position at Sea by Projection on Mercator's Chart*. He proposed solving a single time sight twice, using latitudes somewhat greater and somewhat less than that arrived at by dead reckoning, and joining the two positions obtained to form the line of position.

The Sumner method required the solution of two time sights to obtain each line of position. Many older navigators preferred not to draw the lines on their charts, but to fix their position mathematically by a method which Sumner had also devised and included in his book. This was a tedious but popular procedure.

111. Navigational Tables

Spherical trigonometry is the basis for solving every navigational triangle, and until about 80 years ago the navigator had no choice but to solve each triangle by tedious, manual computations.

Lord Kelvin, generally considered the father of modern navigational methods, expressed interest in a book of tables with which a navigator could avoid tedious trigonometric solutions. However, solving the many thousands of triangles involved would have made the project too costly. Computers finally provided a practical means of preparing tables. In 1936 the first volume of *Pub. No. 214* was made available; later, Pub. No. 249 was provided for air navigators. *Pub. No. 229, Sight Reduction Tables for Marine Navigation*, has replaced *Pub. No. 214*.

Electronic calculators are gradually replacing the tables. Scientific calculators with trigonometric functions can easily solve the navigational triangle. Navigational calculators readily solve celestial sights and perform a variety of voyage planning functions. Using a calculator generally gives more accurate lines of position because it eliminates the rounding errors inherent in tabular inspection and interpolation.

112. Development of Electronic Navigation

Perhaps the first application of electronics to navigation involved sending telegraphic time signals in 1865 to check chronometer error. Transmitting radio time signals for chronometer checks dates to 1904. Radio broadcasts providing navigational warnings, begun in 1907 by the U.S. Navy Hydrographic Office, helped increase the safety of navigation at sea.

By the latter part of World War I the directional properties of a loop antenna were successfully used in the radio direction finder. The first radiobeacon was installed in 1921. Early 20th century experiments by Behm and Langevin led to the U.S. Navy's development of the first practical echo sounder in 1922. Radar and hyperbolic systems grew out of WWII.

Today, electronics touches almost every aspect of navigation. Hyperbolic systems, satellite systems, and electronic charts all require an increasingly sophisticated electronics suite and the expertise to manage them. These systems' accuracy and ease of use make them invaluable assets to the navigator, but there is far more to using them than knowing which buttons to push.

113. Development of Radar

As early as 1904, German engineers were experimenting with reflected radio waves. In 1922 two American scientists, Dr. A. Hoyt Taylor and Leo C. Young, testing a communication system at the Naval Aircraft Radio Laboratory, noted fluctuations in the signals when ships passed between stations on opposite sides of the Potomac River. In 1935 the British began work on radar. In 1937 the USS Leary tested the first sea-going radar, and in 1940 United States and British scientists combined their efforts. When the British revealed the principle of the multicavity magnetron developed by J. T. Randall and H. A. H. Boot at the University of Birmingham in 1939, microwave radar became practical. In 1945, at the close of World War II, radar became available for commercial use.

114. Development of Hyperbolic Radio Aids

Various hyperbolic systems were developed beginning in World War II. These were outgrowths of the British GEE system, developed to help bombers navigate to and from their missions over Europe. Loran A was developed as a long-range marine navigation system. This was replaced by the more accurate Loran C system, deployed throughout much of the world. Various short range and regional hyperbolic systems have been developed by private industry for hydrographic surveying, offshore facilities positioning, and general navigation.

115. Other Electronic Systems

The underlying concept that led to development of satellite navigation dates to 1957 and the first launch of an artificial satellite into orbit. The first system, NAVSAT, has been replaced by the far more accurate and widely available **Global Positioning System (GPS)**, which has revolutionized all aspects of navigation

The first **inertial navigation system** was developed in 1942 for use in the V2 missile by the Peenemunde group under the leadership of Dr. Wernher von Braun. This system used two 2-degree-of-freedom gyroscopes and an integrating accelerometer to determine the missile velocity. By the end of World War II, the Peenemunde group had developed a stable platform with three single-degree-of-freedom gyroscopes and an integrating accelerometer. In 1958 an inertial navigation system was used to navigate the USS *Nautilus* under the ice to the North Pole.

NAVIGATION ORGANIZATIONS

116. Governmental Role

Navigation only a generation ago was an independent process, carried out by the mariner without outside assistance. With compass and charts, sextant and chronometer, he could independently travel anywhere in the world. The increasing use of electronic navigation systems has made the navigator dependent on many factors outside his control. Government organizations fund, operate, and regulate satellites, Loran, and other electronic systems. Governments are increasingly involved in regulation of vessel movements through traffic control systems and regulated areas. Understanding the governmental role in supporting and regulating navigation is vitally important to the mariner. In the United States, there are a number of official organizations which support the interests of navigators. Some have a policy-making role; others build and operate navigation systems. Many maritime nations have similar organizations performing similar functions. International organizations also play a significant role.

117. The Coast and Geodetic Survey

The **U.S. Coast and Geodetic Survey** was founded in 1807 when Congress passed a resolution authorizing a survey of the coast, harbors, outlying islands, and fishing banks of the United States. President Thomas Jefferson appointed Ferdinand Hassler, a Swiss immigrant and professor of mathematics at West Point, the first Director of the “Survey of the Coast.” The survey became the “Coast Survey” in 1836.

The approaches to New York were the first sections of the coast charted, and from there the work spread northward and southward along the eastern seaboard. In 1844 the work was expanded and arrangements made to simultaneously chart the gulf and east coasts. Investigation of tidal conditions began, and in 1855 the first tables of tide predictions were published. The California gold rush necessitated a survey of the west coast, which began in 1850, the year California became a state. *Coast Pilots*, or *Sailing Directions*, for the Atlantic coast of the United States were privately published in the first half of the 19th century. In 1850 the Survey began accumulating data that led to federally produced *Coast Pilots*. The 1889 *Pacific Coast Pilot* was an outstanding contribution to the safety of west coast shipping.

In 1878 the survey was renamed “Coast and Geodetic Survey.” In 1970 the survey became the “National Ocean Survey,” and in 1983 it became the “National Ocean Service.” The Office of Charting and Geodetic Services accomplished all charting and geodetic functions. In 1991 the name was changed back to the original “Coast and Geodetic Survey,” organized under the National Ocean Service along with several other environmental offices. Today it provides the mariner with the charts and coast pilots of all waters of the United States and its possessions, and tide and tidal current tables for much of the world. Its administrative order requires the Coast and Geodetic Survey to plan and direct programs to produce charts and related information for safe navigation of U.S. waterways, territorial seas, and airspace. This work includes all activities related to the National Geodetic Reference System; surveying, charting, and data collection; production and distribution of charts; and research and development of new technologies to enhance these missions.

118. The National Imagery and Mapping Agency

In the first years of the newly formed United States of America, charts and instruments used by the Navy and merchant mariners were left over from colonial days or were obtained from European sources. In 1830 the U.S. Navy established a "Depot of Charts and Instruments" in Washington, D. C., as a storehouse from which available charts, pilots and sailing directions, and navigational instruments were issued to Naval ships. Lieutenant L. M. Goldsborough and one assistant, Passed Midshipman R. B. Hitchcock, constituted the entire staff.

The first chart published by the Depot was produced from data obtained in a survey made by Lieutenant Charles Wilkes, who had succeeded Goldsborough in 1834. Wilkes later earned fame as the leader of a United States expedition to Antarctica. From 1842 until 1861 Lieutenant Matthew Fontaine Maury served as Officer in Charge. Under his command the Depot rose to international prominence.

Maury decided upon an ambitious plan to increase the mariner's knowledge of existing winds, weather, and currents. He began by making a detailed record of pertinent matter included in old log books stored at the Depot. He then inaugurated a hydrographic reporting program among ship masters, and the thousands of reports received, along with the log book data, were compiled into the "*Wind and Current Chart of the North Atlantic*" in 1847. This is the ancestor of today's *Pilot Chart*.

The United States instigated an international conference in 1853 to interest other nations in a system of exchanging nautical information. The plan, which was Maury's, was enthusiastically adopted by other maritime nations. In 1854 the Depot was redesignated the "U.S. Naval Observatory and Hydrographical Office." At the outbreak of the American Civil War in 1861, Maury, a native of Virginia, resigned from the U.S. Navy and accepted a commission in the Confederate Navy. This effectively ended his career as a navigator, author, and oceanographer. At war's end, he fled the country, his reputation suffering from his embrace of the Confederate cause.

After Maury's return to the United States in 1868, he served as an instructor at the Virginia Military Institute. He continued at this position until his death in 1873. Since his death, his reputation as one of America's greatest hydrographers has been restored.

In 1866 Congress separated the Observatory and the Hydrographic Office, broadly increasing the functions of the latter. The Hydrographic Office was authorized to carry out surveys, collect information, and print every kind of nautical chart and publication "for the benefit and use of navigators generally."

The Hydrographic Office purchased the copyright of ***The New American Practical Navigator*** in 1867. The first *Notice to Mariners* appeared in 1869. Daily broadcast of navigational warnings was inaugurated in 1907. In 1912, following the sinking of the *Titanic*, the International Ice Patrol was established.

In 1962 the U.S. Navy Hydrographic Office was redesignated the U.S. Naval Oceanographic Office. In 1972 certain hydrographic functions of the latter office were transferred to the **Defense Mapping Agency Hydrographic Center**. In 1978 the **Defense Mapping Agency Hydrographic/Topographic Center (DMAHTC)** assumed hydrographic and topographic chart production functions. In 1996 the **National Imagery and Mapping Agency (NIMA)** was formed from DMA and certain other elements of the Department of Defense. NIMA continues to produce charts and publications and to disseminate maritime safety information in support of the U.S. military and navigators generally.

119. The United States Coast Guard

Alexander Hamilton established the **U.S. Coast Guard** as the Revenue Marine, later the Revenue Cutter Service, on August 4, 1790. It was charged with enforcing the customs laws of the new nation. A revenue cutter, the *Harriet Lane*, fired the first shot from a naval unit in the Civil War at Fort Sumter. The Revenue Cutter Service became the U.S. Coast Guard when combined with the Lifesaving Service in 1915. The Lighthouse Service was added in 1939, and the Bureau of Marine Inspection and Navigation was added in 1942. The Coast Guard was transferred from the Treasury Department to the Department of Transportation in 1967.

The primary functions of the Coast Guard include maritime search and rescue, law enforcement, and operation of the nation's aids to navigation system. In addition, the Coast Guard is responsible for port safety and security, merchant marine inspection, and marine pollution control. The Coast Guard operates a large and varied fleet of ships, boats, and aircraft in performing its widely ranging duties

Navigation systems operated by the Coast Guard include the system of some 40,000 lighted and unlighted beacons, buoys, and ranges in U.S. and territorial waters; the U.S. stations of the Loran C system; differential GPS (DGPS) services in the U.S.; and Vessel Traffic Services (VTS) in major ports and harbors of the U.S.

120. The United States Navy

The **U.S. Navy** was officially established in 1798. Its role in the development of navigational technology has been singular. From the founding of the Naval Observatory to the development of the most advanced electronics, the U.S. Navy has been a leader in developing devices and techniques designed to make the navigator's job safer and easier.

The development of almost every device known to navigation science has been deeply influenced by Naval policy. Some systems are direct outgrowths of specific Naval needs; some are the result of technological improvements shared with other services and with commercial maritime industry.

121. The United States Naval Observatory

One of the first observatories in the United States was built in 1831-1832 at Chapel Hill, N.C. The Depot of Charts and Instruments, established in 1830, was the agency from which the U.S. Navy Hydrographic Office and the **U.S. Naval Observatory** evolved 36 years later. In about 1835, under Lieutenant Charles Wilkes, the second Officer in Charge, the Depot installed a small transit instrument for rating chronometers.

The Mallory Act of 1842 provided for the establishment of a permanent observatory. The director was authorized to purchase everything necessary to continue astronomical study. The observatory was completed in 1844 and the results of its first observations were published two years later. Congress established the Naval Observatory as a separate agency in 1866. In 1873 a refracting telescope with a 26 inch aperture, then the world's largest, was installed. The observatory, located in Washington, D.C., has occupied its present site since 1893.

122. The Royal Greenwich Observatory

England had no early privately supported observatories such as those on the continent. The need for navigational advancement was ignored by Henry VIII and Elizabeth I, but in 1675 Charles II, at the urging of John Flamsteed, Jonas Moore, Le Sieur de Saint Pierre, and Christopher Wren, established the **Greenwich Royal Observatory**. Charles limited construction costs to £500, and appointed Flamsteed the first Astronomer Royal, at an annual salary of £100. The equipment available in the early years of the observatory consisted of two clocks, a "sextant" of 7 foot radius, a quadrant of 3 foot radius, two telescopes, and the star catalog published almost a century before by Tycho Brahe. Thirteen years passed before Flamsteed had an instrument with which he could determine his latitude accurately.

In 1690 a transit instrument equipped with a telescope and vernier was invented by Romer; he later added a vertical circle to the device. This enabled the astronomer to determine declination and right ascension at the same time. One of these instruments was added to the equipment at Greenwich in 1721, replacing the huge quadrant previously used. The development and perfection of the chronometer in the next hundred years added to the accuracy of observations.

Other national observatories were constructed in the years that followed: at Berlin in 1705, St. Petersburg in 1725, Palermo in 1790, Cape of Good Hope in 1820, Parramatta in New South Wales in 1822, and Sydney in 1855.

123. The International Hydrographic Organization

The **International Hydrographic Organization (IHO)** was originally established in 1921 as the International Hydrographic Bureau (IHB). The present name was adopted in 1970 as a result of a revised international agreement among member nations. However, the former name, International Hydrographic Bureau, was retained for the IHO's administrative body of three Directors and their staff at the organization's headquarters in Monaco.

The IHO sets forth hydrographic standards to be agreed upon by the member nations. All member states are urged and encouraged to follow these standards in their surveys, nautical charts, and publications. As these standards are uniformly adopted, the products of the world's hydrographic and oceanographic offices become more uniform. Much has been done in the field of standardization since the Bureau was founded.

The principal work undertaken by the IHO is:

- To bring about a close and permanent association between national hydrographic offices.
- To study matters relating to hydrography and allied sciences and techniques.
- To further the exchange of nautical charts and documents between hydrographic offices of member governments.
- To circulate the appropriate documents.
- To tender guidance and advice upon request, in particular to countries engaged in setting up or expanding their hydrographic service.
- To encourage coordination of hydrographic surveys with relevant oceanographic activities.
- To extend and facilitate the application of oceanographic knowledge for the benefit of navigators.
- To cooperate with international organizations and scientific institutions which have related objectives.

During the 19th century, many maritime nations established hydrographic offices to provide means for improving the navigation of naval and merchant vessels by providing nautical publications, nautical charts, and other navigational services. There were substantial differences in hydrographic procedures, charts, and publications. In 1889, an International Marine Conference was held at

Washington, D. C., and it was proposed to establish a "permanent international commission." Similar proposals were made at the sessions of the International Congress of Navigation held at St. Petersburg in 1908 and again in 1912.

In 1919 the hydrographers of Great Britain and France cooperated in taking the necessary steps to convene an international conference of hydrographers. London was selected as the most suitable place for this conference, and on July 24, 1919, the First International Conference opened, attended by the hydrographers of 24 nations. The object of the conference was "To consider the advisability of all maritime nations adopting similar methods in the preparation, construction, and production of their charts and all hydrographic publications; of rendering the results in the most convenient form to enable them to be readily used; of instituting a prompt system of mutual exchange of hydrographic information between all countries; and of providing an opportunity to consultations and discussions to be carried out on hydrographic subjects generally by the hydrographic experts of the world." This is still the major purpose of the International Hydrographic Organization.

As a result of the conference, a permanent organization was formed and statutes for its operations were prepared. The International Hydrographic Bureau, now the International Hydrographic Organization, began its activities in 1921 with 18 nations as members. The Principality of Monaco was selected because of its easy communication with the rest of the world and also because of the generous offer of Prince Albert I of Monaco to provide suitable accommodations for the Bureau in the Principality. There are currently 59 member governments. Technical assistance with hydrographic matters is available through the IHO to member states requiring it.

Many IHO publications are available to the general public, such as the International Hydrographic Review, International Hydrographic Bulletin, Chart Specifications of the IHO, Hydrographic Dictionary, and others. Inquiries should be made to the International Hydrographic Bureau, 7 Avenue President J. F. Kennedy, B.P. 445, MC98011, Monaco, CEDEX.

124. The International Maritime Organization

The **International Maritime Organization (IMO)** was established by United Nations Convention in 1948. The Convention actually entered into force in 1959, although an international convention on marine pollution was adopted in 1954. (Until 1982 the official name of the organization was the Inter-Governmental Maritime Consultative Organization.) It is the only permanent body of the U. N. devoted to maritime matters, and the only special U. N. agency to have its headquarters in the UK.

The governing body of the IMO is the **Assembly** of 137 member states, which meets every two years. Between Assembly sessions a Council, consisting of 32 member governments elected by the Assembly, governs the organization. Its work is carried out by the Maritime Safety Committee, with subcommittees for:

- Safety of Navigation
- Radiocommunications
- Life-saving
- Search and Rescue
- Training and Watchkeeping
- Carriage of Dangerous Goods
- Ship Design and Equipment
- Fire Protection
- Stability and Load Lines/Fishing Vessel Safety
- Containers and Cargoes
- Bulk Chemicals
- Marine Environment Protection Committee
- Legal Committee
- Technical Cooperation Committee
- Facilitation Committee

IMO is headed by the Secretary General, appointed by the council and approved by the Assembly. He is assisted by some 300 civil servants.

To achieve its objectives of coordinating international policy on marine matters, the IMO has adopted some 30 conventions and protocols, and adopted over 700 codes and recommendations. An issue to be adopted first is brought before a committee or subcommittee, which submits a draft to a conference. When the conference adopts the final text, it is submitted to member governments for ratification. Ratification by a specified number of countries is necessary for adoption; the more important the issue, the more countries must ratify. Adopted conventions are binding on member governments.

Codes and recommendations are not binding, but in most cases are supported by domestic legislation by the governments involved.

The first and most far-reaching convention adopted by the IMO was the Convention of **Safety of Life at Sea (SOLAS)** in 1960. This convention actually came into force in 1965, replacing a version first adopted in 1948. Because of the difficult process of bringing amendments into force internationally, none of subsequent amendments became binding. To remedy this situation, a new convention was adopted in 1974 and became binding in 1980. Among the regulations is V-20, requiring the carriage of up-to-date charts and publications sufficient for the intended voyage.

Other conventions and amendments were also adopted, such as the International Convention on Load Lines (adopted 1966, came into force 1968), a convention on the tonnage measurement of ships (adopted 1969, came into force 1982), The International Convention on Safe Containers (adopted 1972, came into force 1977), and the convention on **International Regulations for Preventing Collisions at Sea (COLREGS)** (adopted 1972, came into force 1977).

The 1972 COLREGS convention contained, among other provisions, a section devoted to Traffic Separation

Schemes, which became binding on member states after having been adopted as recommendations in prior years.

One of the most important conventions is the **International Convention for the Prevention of Pollution from Ships (MARPOL 73/78)**, which was first adopted in 1973, amended by Protocol in 1978, and became binding in 1983. This convention built on a series of prior conventions and agreements dating from 1954, highlighted by several severe pollution disasters involving oil tankers. The MARPOL convention reduces the amount of oil discharged into the sea by ships, and bans discharges completely in certain areas. A related convention known as the London Dumping Convention regulates dumping of hazardous chemicals and other debris into the sea.

The IMO also develops minimum performance standards for a wide range of equipment relevant to safety at sea. Among such standards is one for the **Electronic Chart Display and Information System (ECDIS)**, the digital display deemed the operational and legal equivalent of the conventional paper chart.

Texts of the various conventions and recommendations, as well as a catalog and publications on other subjects, are available from the Publications Section of the IMO at 4 Albert Embankment, London SE1 7SR, United Kingdom.

125. The International Association of Marine Aids to Navigation and Lighthouse Authorities

The **International Association of Marine Aids to Navigation and Lighthouse Authorities (formerly IALA)** brings together representatives of the aids to navigation services of more than 80 member countries for technical coordination, information sharing, and coordination of improvements to visual aids to navigation throughout the world. It was established in 1957 to provide a permanent organization to support the goals of the Technical Lighthouse Conferences, which had been convening since 1929. The General Assembly of IALA meets about every 4 years. The Council of 20 members meets twice a year to oversee the ongoing programs.

Five technical committees maintain the permanent programs:

- The Marine Marking Committee
- The Radionavigation Systems Committee
- The Vessel Traffic Services (VTS) Committee
- The Reliability Committee
- The Documentation Committee

IALA committees provide important documentation to the IHO and other international organizations, while the IALA Secretariat acts as a clearing house for the exchange of technical information, and organizes seminars and technical support for developing countries.

Its principle work since 1973 has been the implementation of the IALA Maritime Buoyage System, described in Chapter 5, Visual Aids to Navigation. This system replaced some 30 dissimilar buoyage systems in use throughout the world with 2 major systems.

IALA is based near Paris, France in Saint-Germaine-en-Laye.

126. The Radio Technical Commission for Maritime Services

The **Radio Technical Commission for Maritime Services** is a non-profit organization which serves as a focal point for the exchange of information and the development of recommendations and standards related to all aspects of maritime radiocommunications and radionavigation.

Specifically, RTCM:

- Promotes ideas and exchanges information on maritime radiocommunications and radionavigation.
- Facilitates the development and exchange of views among and between government and non-government interests both nationally and internationally.
- Conducts studies and prepares reports on maritime radiocommunications and radionavigation issues to improve efficiency and capabilities.

Both government and non-government organizations are members, coming from the U.S. and many other nations. The RTCM organization consists of a Board of Directors, and the Assembly consisting of all members, officers, staff, technical advisors, and working committees.

Working committees are formed as needed to develop official RTCM recommendations regarding technical standards and regulatory policies in the maritime field. Currently committees address such issues as maritime safety information, electronic charts, emergency position-indicating radiobeacons (EPIRB's), personal locator beacons, ship radars, differential GPS, GLONASS, and maritime survivor locator devices.

The RTCM headquarters office is in Alexandria, VA.

127. The National Marine Electronic Association

The **National Marine Electronic Association (NMEA)** is a professional trade association founded in 1957 whose purpose is to coordinate the efforts of marine electronics manufacturers, technicians, government agencies, ship and boat builders, and other interested groups. In addition to certifying marine electronics technicians and professionally recognizing outstanding achievements by corporate and individual members, the NMEA sets standards for the exchange of digital data by all manufacturers of marine electronic equipment. This allows the configuration of integrated navigation system using equipment from different manufacturers.

NMEA works closely with RTCM and other private organizations and with government agencies to monitor the status of laws and regulations affecting the marine electronics industry.

It also sponsors conferences and seminars, and publishes a number of guides and periodicals for members and the general public.

128. International Electrotechnical Commission

The **International Electrotechnical Commission (IEC)** was founded in 1906 as an outgrowth of the International Electrical Congress held at St. Louis, Missouri in 1904. Some 60 countries are active members. Its mission is to develop and promote standardization among all nations in the technical specifications of electrical and electronic equipment. These technologies include electronics, magnetics, electromagnetics, electroacoustics, multimedia, telecommunications, electrical energy production and distribution, and associated fields such as terminology and symbology, compatibility, performance standards, safety, and environmental factors.

By standardizing in these areas, the IEC seeks to promote more efficient markets, improve the quality of products and standards of performance, promote interoperability, increase production efficiency, and contribute to human health and safety and environmental protection.

Standards are published by the IEC in the form of official IEC documents after debate and input from the national committees. Standards thus represent a consensus of the views of many different interests. Adoption of a standard by any country is entirely voluntary. However, failure to adopt a standard may result in a technical barrier to trade, as goods manufactured to a proprietary standard in one country may be incompatible with the systems of others.

IEC standards are vital to the success of ECDIS and other integrated navigation systems because they help to ensure that systems from various manufacturers in different countries will be compatible and meet required specifications.

CHAPTER 2

GEODESY AND DATUMS IN NAVIGATION

GEODESY, THE BASIS OF CARTOGRAPHY

200. Definition

Geodesy is the science concerned with the exact positioning of points on the surface of the Earth. It also involves the study of variations of the Earth's gravity, the application of these variations to exact measurements on the Earth, and the study of the exact size and shape of the Earth. These factors were unimportant to early navigators because of the relative inaccuracy of their methods. The precision of today's navigation systems and the global nature of satellite and other long-range positioning methods demand a more complete understanding of geodesy by the navigator than has ever before been required.

201. The Shape of the Earth

The **topographic surface** is the actual surface of the earth, upon which geodetic measurements are made. These measurements are then reduced to the **geoid**. Marine navigation measurements are made on the ocean surface which approximates the geoid.

The **geoid** is a surface along which gravity is always equal and to which the direction of gravity is always perpendicular. The latter point is particularly significant because optical instruments containing leveling devices are commonly used to make geodetic measurements. When properly adjusted, the vertical axis of the instrument coincides exactly with the direction of gravity and is by definition perpendicular to the geoid. See Figure 201.

The geoid is that surface to which the oceans would conform over the entire Earth if free to adjust to the combined effect of the Earth's mass attraction and the centrifugal force of the Earth's rotation. Uneven distribution of the Earth's mass makes the geoidal surface irregular.

The geoid refers to the actual size and shape of the Earth, but such an irregular surface has serious limitations as a mathematical Earth model because:

- It has no complete mathematical expression.
- Small variations in surface shape over time introduce small errors in measurement.
- The irregularity of the surface would necessitate a prohibitive amount of computations.

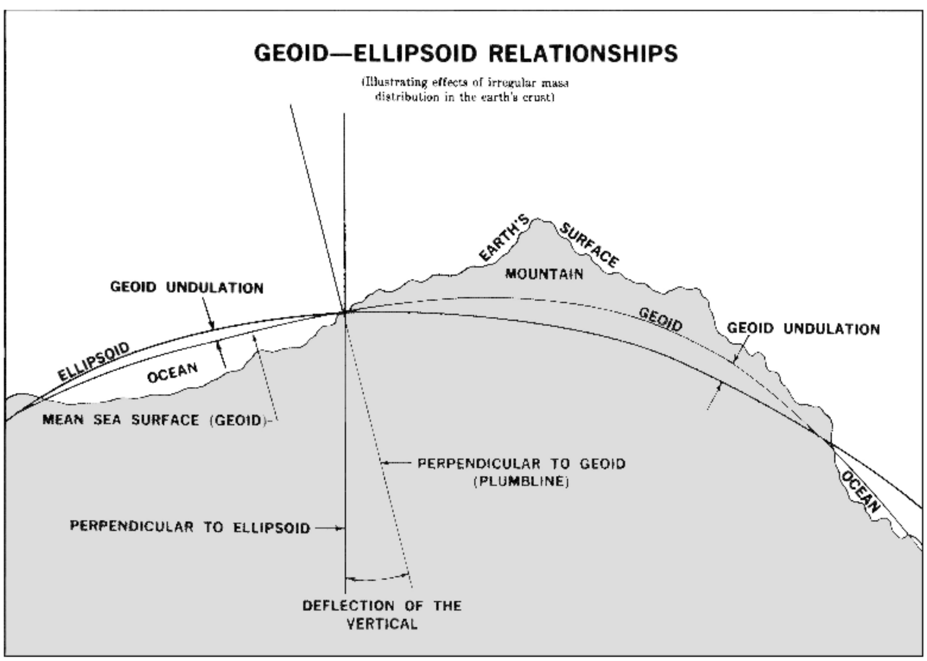


Figure 201. Geoid, ellipsoid, and topographic surface of the Earth, and deflection of the vertical due to differences in mass.

The surface of the geoid, with some exceptions, tends to rise under mountains and to dip above ocean basins.

For geodetic, mapping, and charting purposes, it is necessary to use a regular or geometric shape which closely approximates the shape of the geoid either on a local or global scale and which has a specific mathematical expression. This shape is called the **ellipsoid**.

The separations of the geoid and ellipsoid are called **geoidal heights**, **geoidal undulations**, or **geoidal separations**.

Natural irregularities in density and depths of the material making up the upper crust of the Earth also result in slight alterations of the direction of gravity. These alterations are reflected in the irregular shape of the geoid, the surface that is perpendicular to a plumb line.

Since the Earth is in fact flattened slightly at the poles and bulges somewhat at the equator, the geometric figure used in geodesy to most nearly approximate the shape of the Earth is the **oblate spheroid** or **ellipsoid of revolution**. This is the three dimensional shape obtained by rotating an ellipse about its minor axis.

202. Defining the Ellipsoid

An ellipsoid of revolution is uniquely defined by specifying two parameters. Geodesists, by convention, use the **semimajor axis** and **flattening**. The size is represented by the radius at the equator, the semimajor axis. The shape of the ellipsoid is given by the flattening, which indicates how closely an ellipsoid approaches a spherical shape. The flattening is the ratio of the difference between the semimajor and semiminor axes of the ellipsoid and the semimajor axis. See Figure 202. If a and b represent the semimajor and semiminor axes, respectively, of the ellipsoid, and f is the flattening,

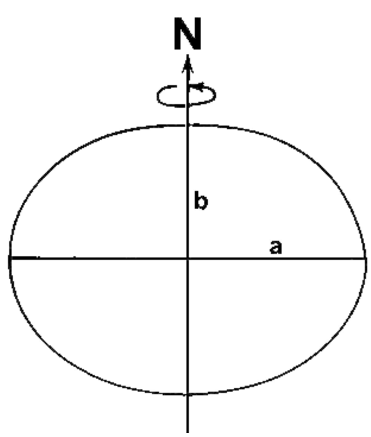


Figure 202. An ellipsoid of revolution, with semimajor axis (a), and semiminor axis (b).

$$f = \frac{a - b}{a}.$$

This ratio is about 1/300 for the Earth. The ellipsoidal Earth model has its minor axis parallel to the Earth's polar axis.

203. Ellipsoids and the Geoid as Reference Surfaces

Since the surface of the geoid is irregular and the surface of an ellipsoid is regular, no ellipsoid can provide more than an approximation of part of the geoidal surface. Figure 203 illustrates an example. A variety of ellipsoids are necessary to cover the entire earth.

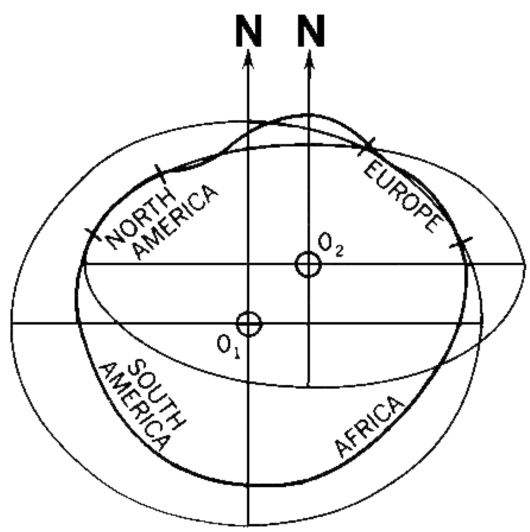


Figure 203. An ellipsoid which fits well in North America may not fit well in Europe, whose ellipsoid must have a different size, shape, and origin. Other ellipsoids are necessary for other areas

204. Coordinates

The **astronomic latitude** is the angle between a plumb line and the plane of the celestial equator. It is the latitude which results directly from observations of celestial bodies, uncorrected for deflection of the vertical component in the meridian (north-south) direction. Astronomic latitude applies only to positions on the Earth. It is reckoned from the astronomic equator (0°), north and south through 90°.

The **astronomic longitude** is the angle between the plane of the celestial meridian at a station and the plane of the celestial meridian at Greenwich. It is the longitude which results directly from observations of celestial bodies, uncorrected for deflection of the vertical component in the prime vertical (east-west) direction. These are the coordinates observed by the celestial navigator using a sextant and a very accurate clock based on the Earth's rotation.

Celestial observations by geodesists are made with optical instruments (theodolite, zenith camera, prismatic astrolabe) which all contain leveling devices. When properly adjusted, the vertical axis of the instrument coincides with the direction of gravity, which may not coincides with the plane of the meridian. Thus, geodetically derived astronomic positions are referenced to the geoid. The difference, from a navigational standpoint, is too small to be of concern.

The **geodetic latitude** is the angle which the normal to the ellipsoid at a station makes with the plane of the geodetic equator. In recording a geodetic position, it is essential that the geodetic datum on which it is based also be stated. A geodetic latitude differs from the corresponding astronomic latitude by the amount of the meridian component of the local deflection of the vertical.

The **geodetic longitude** is the angle between the plane of the geodetic meridian at a station and the plane of the geodetic meridian at Greenwich. A geodetic longitude differs from the corresponding astronomic longitude by the prime vertical component of the local deflection of the vertical divided by the cosine of the latitude. The geodetic coordinates are used for mapping.

The **geocentric latitude** is the angle at the center of the ellipsoid (used to represent the Earth) between the plane of the equator, and a straight line (or radius vector) to a point on the surface of the ellipsoid. This differs from geodetic latitude because the Earth is approximated more closely by a spheroid than a sphere and the meridians are ellipses, not perfect circles.

Both geocentric and geodetic latitudes refer to the reference ellipsoid and not the Earth. Since the parallels of latitude are considered to be circles, geodetic longitude is geocentric, and a separate expression is not used.

Because of the oblate shape of the ellipsoid, the length of a degree of geodetic latitude is not everywhere the same, increasing from about 59.7 nautical miles at the equator to about 60.3 nautical miles at the poles.

A **horizontal geodetic datum** usually consists of the astronomic and geodetic latitude, and astronomic and geodetic longitude of an initial point (origin); an azimuth of a line (direction); the parameters (radius and flattening) of the ellipsoid selected for the computations; and the geoidal separation at the origin. A change in any of these quantities affects every point on the datum.

For this reason, while positions within a given datum are directly and accurately relatable, those from different datums must be transformed to a common datum for consistency.

TYPES OF GEODETIC SURVEY

205. Triangulation

The most common type of geodetic survey is known as **triangulation**. Triangulation consists of the measurement of the angles of a series of triangles. The principle of triangulation is based on plane trigonometry. If the distance along one side of the triangle and the angles at each end are accurately measured, the other two sides and the remaining angle can be computed. In practice, all of the angles of every triangle are measured to provide precise measurements. Also, the latitude and longitude of one end of the measured side along with the length and direction (azimuth) of the side provide sufficient data to compute the latitude and longitude of the other end of the side.

The measured side of the base triangle is called a **baseline**. Measurements are made as carefully and accurately as possible with specially calibrated tapes or wires of Invar, an alloy with a very low coefficient of expansion. The tape or wires are checked periodically against standard measures of length.

To establish an arc of triangulation between two widely separated locations, the baseline may be measured and longitude and latitude determined for the initial points at each location. The lines are then connected by a series of adjoining triangles forming quadrilaterals extending from each end. All angles of the triangles are measured repeatedly to reduce errors. With the longitude, latitude, and azimuth of the initial points, similar data is computed for each vertex of the triangles, thereby establishing triangulation stations, or geodetic control stations. The coordinates of each of the stations are defined as geodetic coordinates.

Triangulation is extended over large areas by connecting and extending series of arcs to form a network or triangulation system. The network is adjusted so as to reduce observational errors to a minimum. A denser distribution of geodetic control is achieved by subdividing or filling in with other surveys.

There are four general classes or orders of triangulation. **First-order** (primary) triangulation is the most precise and exact type. The most accurate instruments and rigorous computation methods are used. It is costly and time-consuming, and is usually used to provide the basic framework of control data for an area, and the determination of the figure of the Earth. The most accurate first-order surveys furnish control points which can be interrelated with an accuracy ranging from 1 part in 25,000 over short distances to approximately 1 part in 100,000 for long distances.

Second-order triangulation furnishes points closer together than in the primary network. While second-order surveys may cover quite extensive areas, they are usually tied to a primary system where possible. The procedures are less exacting and the proportional error is 1 part in 10,000.

Third-order triangulation is run between points in a secondary survey. It is used to densify local control nets and position the topographic and hydrographic detail of the area. Error can amount to 1 part in 5,000.

The sole accuracy requirement for **fourth-order** triangulation is that the positions be located without any appreciable error on maps compiled on the basis of the control. Fourth-order control is done primarily as mapping control.

206. Trilateration, Traverse, And Vertical Surveying

Trilateration involves measuring the sides of a chain of triangles or other polygons. From them, the distance and direction from A to B can be computed. Figure 206 shows this process.

Traverse involves measuring distances and the angles between them without triangles for the purpose of computing the distance and direction from A to B. See Figure 206.

Vertical surveying is the process of determining elevations above mean sea-level. In geodetic surveys executed primarily for mapping, geodetic positions are referred to an ellipsoid, and the elevations of the positions are referred to the geoid. However, for satellite geodesy the geoidal heights must be considered to establish the correct height above the geoid.

Precise geodetic **leveling** is used to establish a basic network of vertical control points. From these, the height of other positions in the survey can be determined by supplementary methods. The mean sea-level surface used as a reference (vertical datum) is determined by averaging the hourly water heights for a specified period of time at specified tide gauges.

There are three leveling techniques: **differential**, **trigonometric**, and **barometric**. Differential leveling is the most accurate of the three methods. With the instrument locked in position, readings are made on two calibrated staffs held in an upright position ahead of and behind the instrument. The difference between readings is the difference in elevation between the points.

Trigonometric leveling involves measuring a vertical angle from a known distance with a theodolite and computing the elevation of the point. With this method, vertical measurement can be made at the same time horizontal angles are measured for triangulation. It is, therefore, a somewhat more economical method but less accurate than differential leveling. It is often the only mechanical method of establishing accurate elevation control in mountainous areas.

In barometric leveling, differences in height are determined by measuring the differences in atmospheric pressure at various elevations. Air pressure is measured by mercurial or aneroid barometer, or a boiling point thermometer. Although the accuracy of this method is not as great as either of the other two, it obtains relative heights very rapidly at points which are fairly far apart. It is used in reconnaissance and exploratory surveys where more accurate measurements will be made later or where a high degree of accuracy is not required.

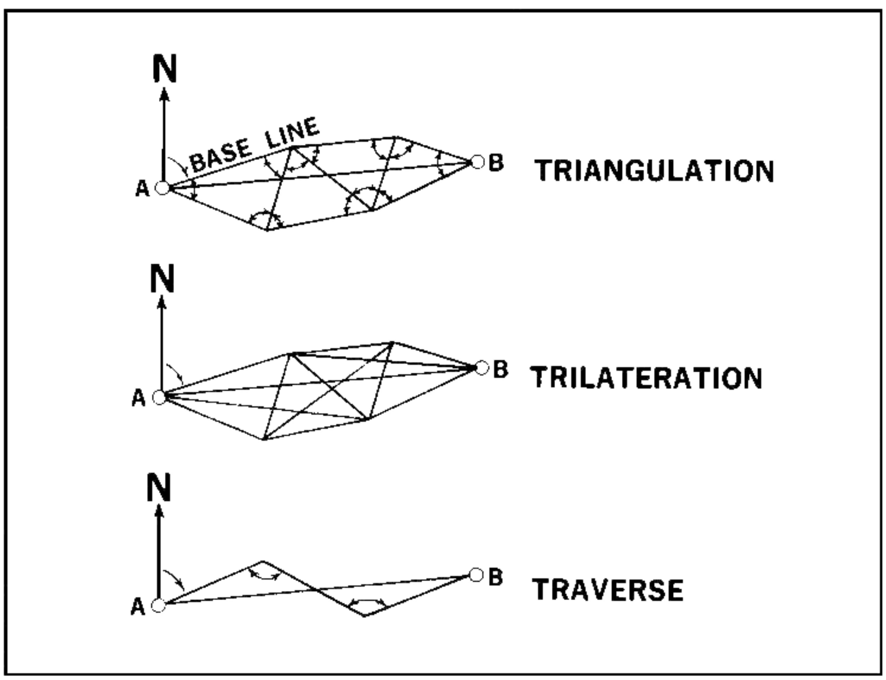


Figure 206. Triangulation, trilateration, and traverse.

DATUM CONNECTIONS

207. Definitions

A **datum** is defined as any numerical or geometrical quantity or set of such quantities which serves as a reference point from which to measure other quantities.

In geodesy, cartography, and navigation, two general types of datums must be considered: **horizontal datum** and **vertical datum**. The horizontal datum forms the basis for computations of horizontal position. The vertical datum provides the reference to measure heights or depths, and may be one of two types: **Vertical geodetic datum** is the reference used by surveyors to measure heights of topographic features, and by cartographers to portray them. This should not be confused with the various types of **tidal datums**, which are by definition vertical datums (and having no horizontal component), used to define the heights and depths of hydrographic features, such as water depths or bridge clearances. The vertical geodetic datum is derived from its mathematical expression, while the tidal datum is derived from actual tidal data. For a complete discussion of tidal datums, see Chapter 9.

This chapter will discuss only geodetic datums. For navigational purposes, vertical geodetic datums are quite unimportant, while horizontal geodetic datums and tidal datums are vital.

A horizontal datum may be defined at an origin point on the ellipsoid (local datum) such that the center of the ellipsoid coincides with the Earth's center of mass (geocentric datum). The coordinates for points in specific geodetic surveys and triangulation networks are computed from certain initial quantities, or datums.

208. Preferred Datums

In areas of overlapping geodetic triangulation networks, each computed on a different datum, the coordinates of the points given with respect to one datum will differ from those given with respect to the other. The differences can be used to derive transformation formulas. Datums are connected by developing transformation formulas at common points, either between overlapping control networks or by satellite connections.

Many countries have developed national datums which differ from those of their neighbors. Accordingly, national maps and charts often do not agree along national borders.

The **North American Datum, 1927** (NAD 27) has been used in the United States for about 60 years, but it is being replaced by datums based on the **World Geodetic System**. NAD 27 coordinates are based on the latitude and longitude of a triangulation station (the reference point) at Mead's Ranch in Kansas, the azimuth to a nearby triangulation station called Waldo, and the mathematical parameters of the Clarke Ellipsoid of 1866. Other datums throughout the world use different assumptions as to origin points and ellipsoids.

The origin of the **European Datum** is at Potsdam, Germany. Numerous national systems have been joined into a large datum based upon the International Ellipsoid of 1924 which was oriented by a modified astrogeodetic method. European, African, and Asian triangulation chains were connected, and African measurements from Cairo to Cape Town were completed. Thus, all of Europe, Africa, and Asia are molded into one great system. Through common survey stations, it was also possible to convert data from the Russian Pulkova, 1932 system to the European Datum, and as a result, the European Datum includes triangulation as far east as the 84th meridian. Additional ties across the Middle East have permitted connection of the Indian and European Datums.

The **Ordnance Survey of Great Britain 1936 Datum** has no point of origin. The data was derived as a best fit between retriangulation and original values of 11 points of the earlier Principal Triangulation of Great Britain (1783-1853).

Tokyo Datum has its origin in Tokyo. It is defined in terms of the Bessel Ellipsoid and oriented by a single astronomic station. Triangulation ties through Korea connect the Japanese datum with the Manchurian datum. Unfortunately, Tokyo is situated on a steep slope on the geoid, and the single-station orientation has resulted in large systematic geoidal separations as the system is extended from its initial point.

The **Indian Datum** is the preferred datum for India and several adjacent countries in Southeast Asia. It is computed on the Everest Ellipsoid with its origin at Kalianpur, in central India. It is largely the result of the untiring work of Sir George Everest (1790-1866), Surveyor General in India from 1830 to 1843. He is best known by the mountain named after him, but by far his most important legacy was the survey of the Indian subcontinent.

MODERN GEODETIC SYSTEMS

209. Development of the World Geodetic System

By the late 1950's the increasing range and sophistication of weapons systems had rendered local or national datums inadequate for military purposes; these new weapons required datums at least continental, if not global, in scope. In response to these requirements, the U.S. Department of Defense generated a geocentric (earth-centered) reference system to which different geodetic networks could be referred, and established compatibility

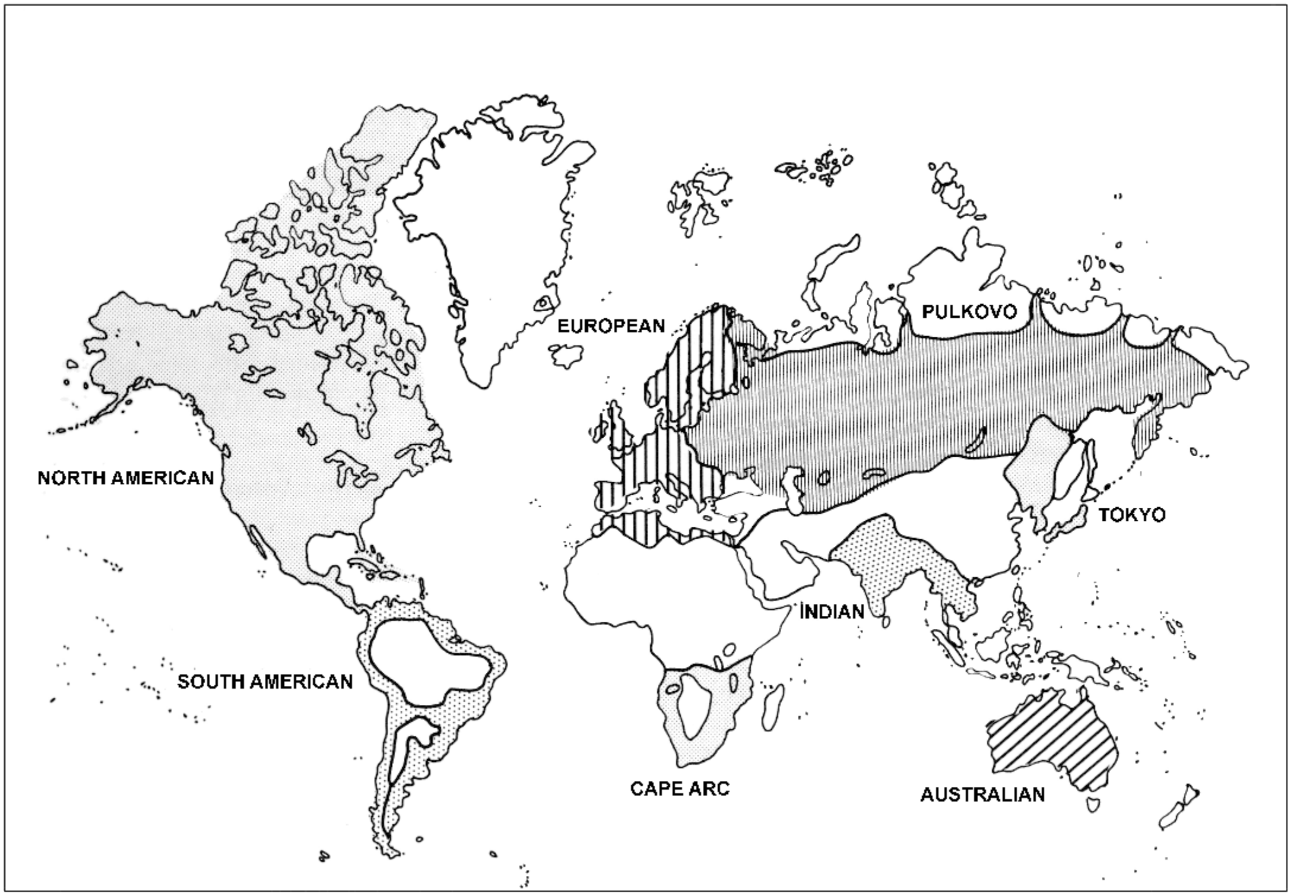


Figure 208. Major geodetic datum blocks.

between the coordinate systems. Efforts of the Army, Navy, and Air Force were combined, leading to the development of the DoD **World Geodetic System of 1960 (WGS 60)**.

In January 1966, a World Geodetic System Committee was charged with the responsibility for developing an improved WGS needed to satisfy mapping, charting, and geodetic requirements. Additional surface gravity observations, results from the extension of triangulation and trilateration networks, and large amounts of Doppler and optical satellite data had become available since the development of WGS 60. Using the additional data and improved techniques, the Committee produced **WGS 66** which served DoD needs following its implementation in 1967.

The same World Geodetic System Committee began work in 1970 to develop a replacement for WGS 66. Since the development of WGS 66, large quantities of additional data had become available from both Doppler and optical satellites, surface gravity surveys, triangulation and trilateration surveys, high precision traverses, and astronomic surveys.

In addition, improved capabilities had been developed in both computers and computer software. Continued research in computational procedures and error analyses had produced better methods and an improved facility for handling and combining data. After an extensive effort extending over a period of approximately three years, the Committee completed the development of the Department of Defense **World Geodetic System 1972 (WGS 72)**.

Further refinement of WGS 72 resulted in the new **World Geodetic System of 1984 (WGS 84),** now referred to as simply WGS. For surface navigation, WGS 60, 66, 72 and the new WGS 84 are essentially the same, so that positions computed on any WGS coordinates can be plotted directly on the others without correction.

The WGS system is not based on a single point, but many points, fixed with extreme precision by satellite fixes and statistical methods. The result is an ellipsoid which fits the real surface of the Earth, or geoid, far more accurately than any other. The WGS system is applicable worldwide. All regional datums can be referenced to WGS once a survey tie has been made.

210. The New North American Datum Of 1983

The Coast And Geodetic Survey of the National Ocean Service (NOS), NOAA, is responsible for charting United States waters. From 1927 to 1987, U.S. charts were based on NAD 27, using the Clarke 1866 ellipsoid. In 1989, the U.S. officially switched to **NAD 83** (navigationally equivalent to WGS) for all mapping and charting purposes, and all new NOS chart production is based on this new standard.

The grid of interconnected surveys which criss-crosses the United States consists of some 250,000 control points, each consisting of the latitude and longitude of the point, plus additional data such as elevation. Converting the NAD 27 coordinates to NAD 83 involved recomputing the position of each point based on the new NAD 83 datum. In addition to the 250,000 U.S. control points, several thousand more were added to tie in surveys from Canada, Mexico, and Central America.

Conversion of new edition charts to the new datums, either WGS 84 or NAD 83, involves converting reference points on each chart from the old datum to the new, and adjusting the latitude and longitude grid (known as the graticule) so that it reflects the newly plotted positions. This adjustment of the graticule is the only difference between charts which differ only in datum. All charted features remain in exactly the same relative positions.

The Global Positioning System (GPS) has transformed the science of surveying, enabling the establishment of precise ties to WGS in areas previously found to be too remote to survey to modern standards. As a result, new charts are increasingly precise as to position of features. The more recent a chart's date of publishing, the more likely it is that it will be accurate as to positions. Navigators should always refer to the title block of a chart to determine the date of the chart, the date of the surveys and sources used to compile it, and the datum on which it is based.

DATUMS AND NAVIGATION

211. Datum Shift

One of the most serious impacts of different datums on navigation occurs when a navigation system provides a fix based on a datum different from that used for the nautical chart. The resulting plotted position may be different from the actual location on that chart. This difference is known as a **datum shift**.

Modern electronic navigation systems have software installed that can output positions in a variety of datums, eliminating the necessity for applying corrections. All electronic charts produced by NIMA are compiled on WGS and are not subject to datum shift problems as long as the GPS receiver is outputting WGS position data to the display system. The same is true for NOAA charts of the U.S., which are compiled on NAD 83 datum, very closely related to WGS. GPS receivers, including the WRN-6, default to WGS, so that no action is necessary to use any U.S.-produced electronic charts.

To automate datum conversions, a number of datum transformation software programs have been written that will convert from any known datum to any other, in any location. MADTRAN and GEOTRANS-2 are two such programs. The amount of datum shift between two different datums is not linear. That is, the amount of shift is a function of the position of the observer, which must be specified for the shift to be computed. Varying differences of latitude and longitude between two different datums will be noted as one's location changes.

There are still a few NIMA-produced paper charts, and a number of charts from other countries, based on datums other than WGS. If the datum of these charts is noted in the title block of the chart, the WRN-6 and most other GPS receivers can be set to output position data in that datum, eliminating the datum shift problem. If the datum is not listed, extreme caution is necessary. An offset can sometimes be established if the ship's actual position can be determined with sufficient accuracy, and this offset applied to GPS positions in the local area. But remember that since a datum shift is not linear, this offset is only applicable locally.

Another effect on navigation occurs when shifting between charts that have been compiled using different datums. If a position is replotted on a chart of another datum using latitude and longitude, the newly plotted position will not match with respect to other charted features. The datum shift may be avoided by transferring positions using bearings and ranges to common points. If datum shift conversion notes for the applicable datums are given on the charts, positions defined by latitude and longitude may be replotted after applying the noted correction.

The positions given for chart corrections in the *Notice to Mariners* reflect the proper datum for each specific chart and edition number. Due to conversion of charts based on old datums to more modern ones, and the use of many different datums throughout the world, chart corrections intended for one edition of a chart may not be safely plotted on any other.

As noted, datum shifts are not constant throughout a given region, but vary according to how the differing datums fit together. For example, the NAD 27 to NAD 83 conversion resulted in changes in latitude of 40 meters in Miami, 11 meters in New York, and 20 meters in Seattle. Longitude changes for this conversion amounted to 22 meters in Miami, 35 meters in New York, and 93 meters in Seattle.

Most charts produced by NIMA and NOS show a "datum note." This note is usually found in the title block or in the upper left margin of the chart. According to the year of the chart edition, the scale, and policy at the time of production, the note may say "World Geodetic System 1972 (WGS-72)", "World Geodetic System 1984 (WGS-84)", or "World Geodetic System (WGS)." A datum note for a chart for which satellite positions can be plotted without correction will read: "Positions obtained from satellite navigation systems referred to (Reference Datum) can be plotted directly on this chart."

NIMA reproductions of foreign charts will usually be in the datum or reference system of the producing country. In these cases a conversion factor is given in the following format: "Positions obtained from satellite navigation systems referred to the (Reference Datum) must be moved X.XX minutes (Northward/Southward) and X.XX minutes (Eastward/ Westward) to agree with this chart."

Some charts cannot be tied in to WGS because of lack of recent surveys. Currently issued charts of some areas are based on surveys or use data obtained in the age of sailing ships. The lack of surveyed control points means that they cannot be properly referenced to modern geodetic systems. In this case there may be a note that says: "Adjustments to WGS cannot be determined for this chart."

A few charts may have no datum note at all, but may carry a note which says: "From various sources to (year)." In these cases there is no way for the navigator to determine the mathematical difference between the local datum and WGS positions. However, if a radar or visual fix can be accurately determined, and an offset established as noted above. This offset can then be programmed into the GPS receiver.

To minimize problems caused by differing datums:

- Plot chart corrections only on the specific charts and editions for which they are intended. Each chart correction is specific to only one edition of a chart. When the same correction is made on two charts based on different datums, the positions for the same feature may differ slightly. This difference is equal to the datum shift between the two datums for that area.
- Try to determine the source and datum of positions of temporary features, such as drill rigs. In general they are given in the datum used in the area in question. Since these are precisely positioned using satellites, WGS is the normal datum. A datum correction, if needed, might be found on a chart of the area.
- Remember that if the datum of a plotted feature is not known, position inaccuracies may result. It is wise to allow a margin of error if there is any doubt about the datum.
- Know how the datum of the positioning system you are using (Loran, GPS, etc.) relates to your chart. GPS and other modern positioning systems use WGS datum. If your chart is on any other datum, you must program the system to use the chart's datum, or apply a datum correction when plotting GPS positions on the chart.

CHAPTER 3

NAUTICAL CHARTS

CHART FUNDAMENTALS

300. Definitions

A **nautical chart** represents part of the spherical earth on a plane surface. It shows water depth, the shoreline of adjacent land, prominent topographic features, aids to navigation, and other navigational information. It is a work area on which the navigator plots courses, ascertains positions, and views the relationship of the ship to the surrounding area. It assists the navigator in avoiding dangers and arriving safely at his destination.

Originally hand-drawn on sheepskin, traditional nautical charts have for generations been printed on paper. **Electronic charts** consisting of a digital data base and a display system are in use and are replacing paper charts aboard many vessels. An electronic chart is not simply a digital version of a paper chart; it introduces a new navigation methodology with capabilities and limitations very different from paper charts. The electronic chart is the legal equivalent of the paper chart if it meets certain International Maritime Organization specifications. See Chapter 14 for a complete discussion of electronic charts.

Should a marine accident occur, the nautical chart in use at the time takes on legal significance. In cases of grounding, collision, and other accidents, charts become critical records for reconstructing the event and assigning liability. Charts used in reconstructing the incident can also have tremendous training value.

301. Projections

Because a cartographer cannot transfer a sphere to a flat surface without distortion, he must project the surface of a sphere onto a **developable surface**. A developable surface is one that can be flattened to form a plane. This process is known as **chart projection**. If points on the surface of the sphere are projected from a single point, the projection is said to be **perspective** or **geometric**.

As the use of electronic charts becomes increasingly widespread, it is important to remember that the same cartographic principles that apply to paper charts apply to their depiction on video screens.

302. Selecting a Projection

Each projection has certain preferable features. However, as the area covered by the chart becomes smaller, the differences between various projections become less noticeable. On the largest scale chart, such as of a harbor, all projections are practically identical. Some desirable properties of a projection are:

1. True shape of physical features
2. Correct angular relationships
3. Equal area (Represents areas in proper proportions)
4. Constant scale values
5. Great circles represented as straight lines
6. Rhumb lines represented as straight lines

Some of these properties are mutually exclusive. For example, a single projection cannot be both conformal and equal area. Similarly, both great circles and rhumb lines cannot be represented on a single projection as straight lines.

303. Types of Projections

The type of developable surface to which the spherical surface is transferred determines the projection's classification. Further classification depends on whether the projection is centered on the equator (equatorial), a pole (polar), or some point or line between (oblique). The name of a projection indicates its type and its principal features.

Mariners most frequently use a **Mercator projection**, classified as a **cylindrical projection** upon a plane, the cylinder tangent along the equator. Similarly, a projection based upon a cylinder tangent along a meridian is called **transverse** (or inverse) **Mercator** or **transverse** (or inverse) **orthomorphic**. The Mercator is the most common projection used in maritime navigation, primarily because rhumb lines plot as straight lines.

In a **simple conic projection**, points on the surface of the earth are transferred to a tangent cone. In the **Lambert conformal projection**, the cone intersects the earth (a secant cone) at two small circles. In a **polyconic projection**, a series of tangent cones is used.

In an **azimuthal** or **zenithal projection**, points on the earth are transferred directly to a plane. If the origin of the projecting rays is the center of the earth, a **gnomonic projection** results; if it is the point opposite the plane's point of tangency, a **stereographic projection**; and if at infinity (the projecting lines being parallel to each other), an **orthographic projection**. The gnomonic, stereographic, and orthographic are **perspective projections**. In an **azimuthal equidistant projection**, which is not perspective, the scale of distances is constant along any radial line from the point of tangency. See Figure 303.

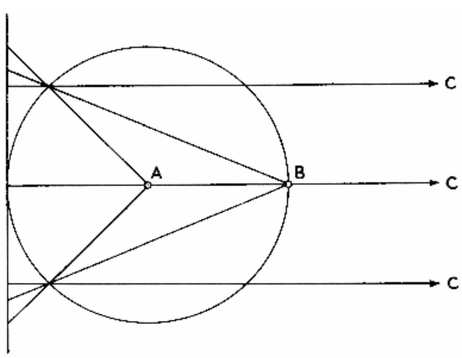


Figure 303. Azimuthal projections: A, gnomonic; B, stereographic; C, (at infinity) orthographic.

Cylindrical and **plane projections** are special conical projections, using heights infinity and zero, respectively.

A **graticule** is the network of latitude and longitude lines laid out in accordance with the principles of any projection.

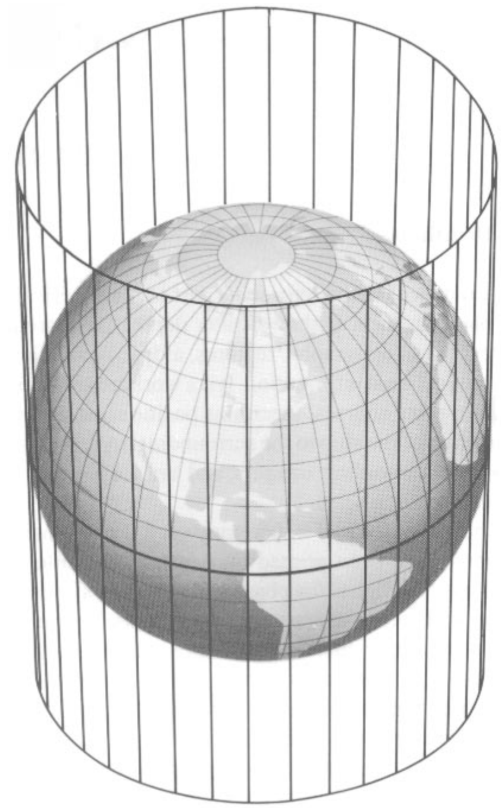


Figure 304. A cylindrical projection.

304. Cylindrical Projections

If a cylinder is placed around the earth, tangent along the equator, and the planes of the meridians are extended, they intersect the cylinder in a number of vertical lines. See Figure 304. These parallel lines of projection are equidistant from each other, unlike the terrestrial meridians from which they are derived which converge as the latitude increases. On the earth, parallels of latitude are perpendicular to the meridians, forming circles of progressively smaller diameter as the latitude increases. On the cylinder they are shown perpendicular to the projected meridians, but because a cylinder is everywhere of the same diameter, the projected parallels are all the same size.

If the cylinder is cut along a vertical line (a meridian) and spread out flat, the meridians appear as equally spaced vertical lines; and the parallels appear as horizontal lines. The parallels' relative spacing differs in the various types of cylindrical projections.

If the cylinder is tangent along some great circle other than the equator, the projected pattern of latitude and longitude lines appears quite different from that described above, since the line of tangency and the equator no longer coincide. These projections are classified as **oblique** or **transverse projections**.

305. Mercator Projection

Navigators most often use the plane conformal projection known as the **Mercator projection**. The Mercator projection is not perspective, and its parallels can be derived mathematically as well as projected geometrically. Its distinguishing feature is that both the meridians and parallels are expanded at the same ratio with increased latitude. The expansion is equal to the secant of the latitude, with a small correction for the ellipticity of the earth. Since the secant of 90° is infinity, the projection cannot include the poles. Since the projection is conformal, expansion is the same in all directions and angles are correctly shown. Rhumb lines appear as straight lines, the directions of which can be measured directly on the chart. Distances can also be measured directly if the spread of latitude is small. Great circles, except meridians and the equator, appear as curved lines concave to the equator. Small areas appear in their correct shape but of increased size unless they are near the equator.

306. Meridional Parts

At the equator a degree of longitude is approximately equal in length to a degree of latitude. As the distance from the equator increases, degrees of latitude remain approximately the same, while degrees of longitude become

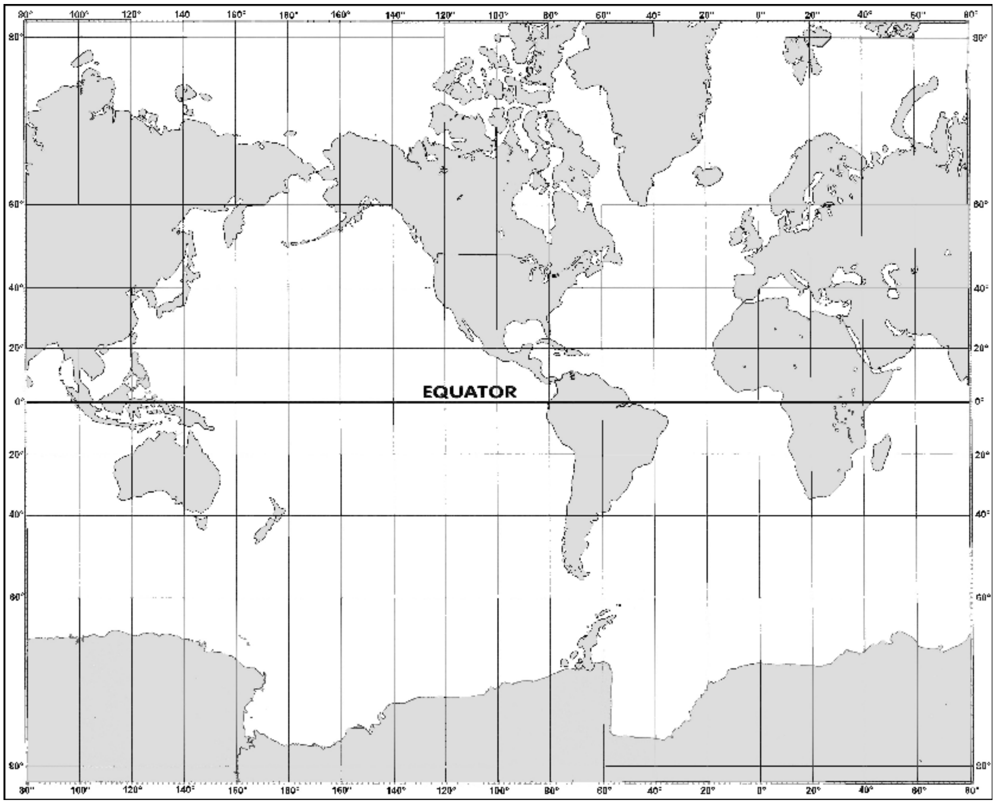


Figure 306. A Mercator map of the world.

progressively shorter. Since degrees of longitude appear everywhere the same length in the Mercator projection, it is necessary to increase the length of the meridians if the expansion is to be equal in all directions. Thus, to maintain the correct proportions between degrees of latitude and degrees of longitude, the degrees of latitude must be progressively longer as the distance from the equator increases. This is illustrated in Figure 306.

The length of a meridian, increased between the equator and any given latitude, expressed in minutes of arc at the equator as a unit, constitutes the number of meridional parts (M) corresponding to that latitude. Meridional parts, given in Table 6 for every minute of latitude from the equator to the pole, make it possible to construct a Mercator chart and to solve problems in Mercator sailing. These values are for the WGS ellipsoid of 1984.

307. Transverse Mercator Projections

Constructing a chart using Mercator principles, but with the cylinder tangent along a meridian, results in a **transverse Mercator** or **transverse orthomorphic projection**. The word "inverse" is used interchangeably with "transverse." These projections use a fictitious graticule similar to, but offset from, the familiar network of meridians and parallels. The tangent great circle is the fictitious equator. Ninety degrees from it are two fictitious poles. A group of great circles through these poles and perpendicular to the tangent great circle are the fictitious meridians, while a series of circles parallel to the plane of the tangent great circle form the fictitious parallels. The actual meridians and parallels appear as curved lines.

A straight line on the transverse or oblique Mercator projection makes the same angle with all fictitious meridians, but not with the terrestrial meridians. It is therefore a fictitious rhumb line. Near the tangent great circle, a straight line closely approximates a great circle. The projection is most useful in this area. Since the area of minimum distortion is near a meridian, this projection is useful for charts covering a large band of latitude and extending a relatively short distance on each side of the tangent meridian. It is sometimes used for star charts showing the evening sky at various seasons of the year. See Figure 307.

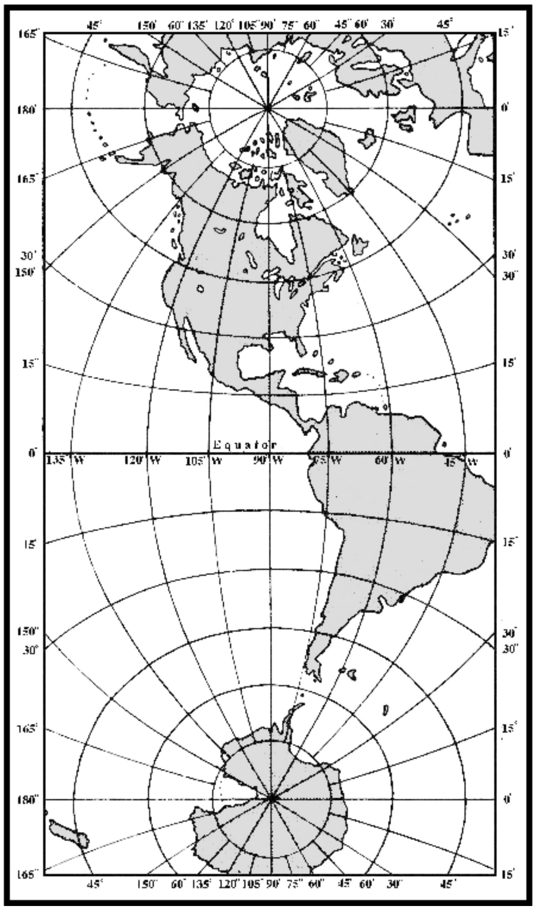


Figure 307. A transverse Mercator map of the Western Hemisphere.

308. Universal Transverse Mercator (UTM) Grid

The **Universal Transverse Mercator (UTM)** grid is a military grid superimposed upon a transverse Mercator graticule, or the representation of these grid lines upon any graticule. This grid system and these projections are often used for large-scale (harbor) nautical charts and military charts.

309. Oblique Mercator Projections

A Mercator projection in which the cylinder is tangent along a great circle other than the equator or a meridian is called an **oblique Mercator** or **oblique orthomorphic projection**. See Figure 309a and Figure 309b. This projection is used principally to depict an area in the near vicinity of an oblique great circle. Figure 309c, for example, shows the great circle joining Washington and Moscow. Figure 309d shows an oblique Mercator map with the great circle between these two centers as the tangent great circle or fictitious equator. The limits of the chart of Figure 309c are indicated in Figure 309d. Note the large variation in scale as the latitude changes.

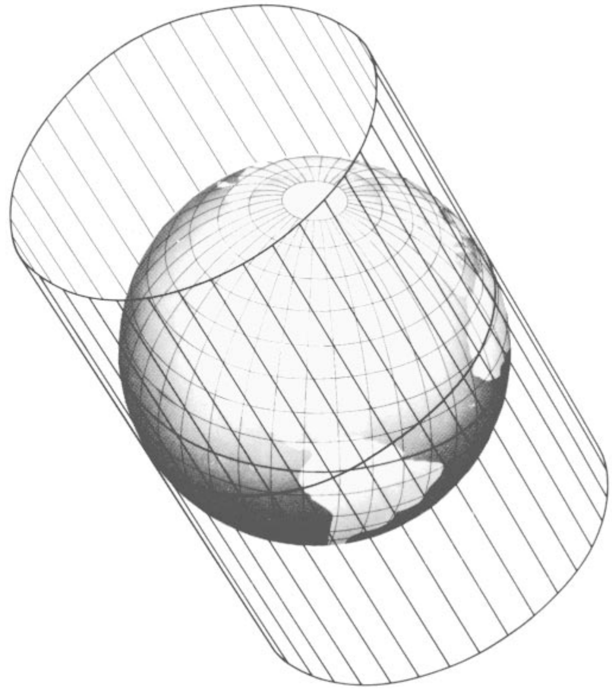


Figure 309a. An oblique Mercator projection.

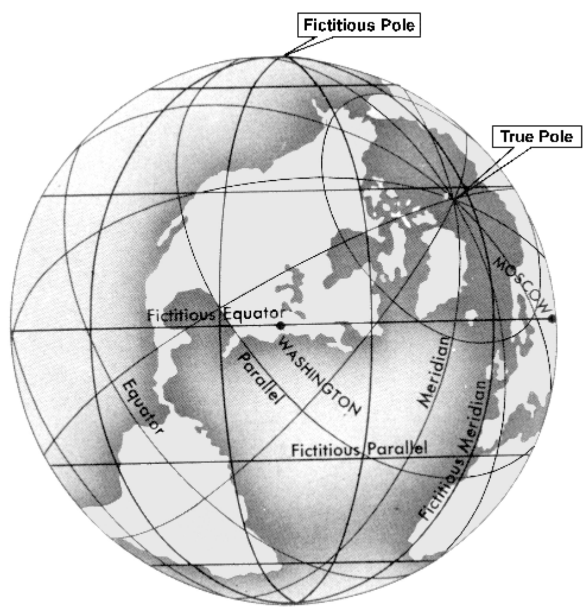


Figure 309b. The fictitious graticule of an oblique Mercator projection.

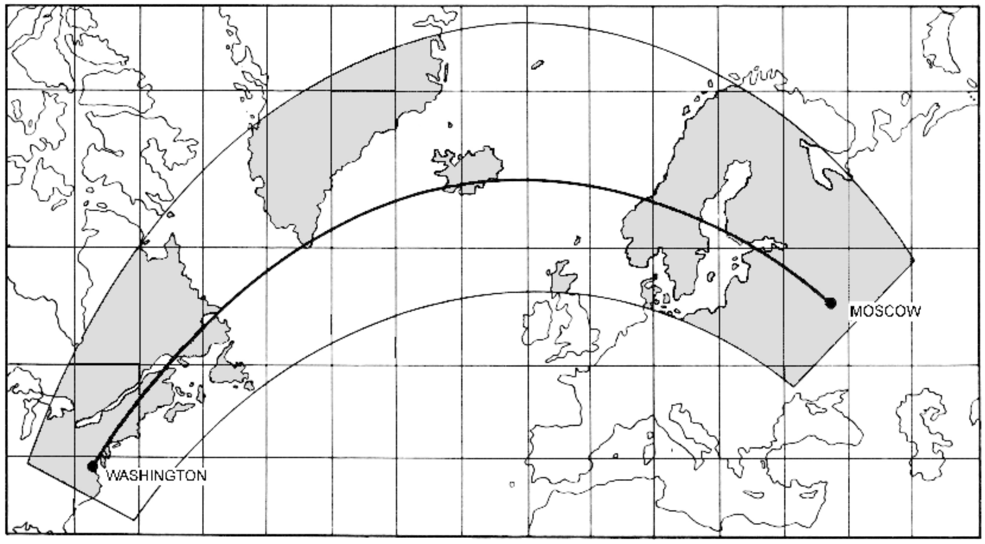


Figure 309c. The great circle between Washington and Moscow as it appears on a Mercator map.

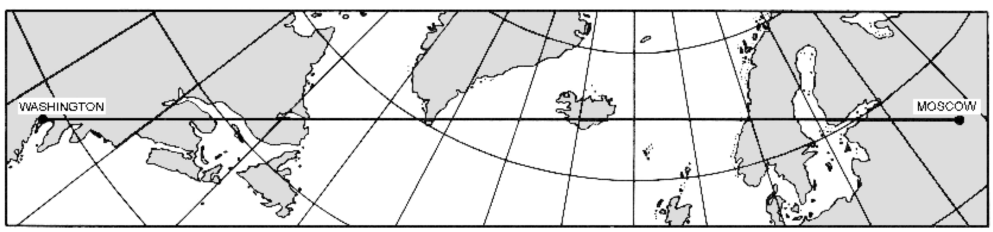


Figure 309d. An oblique Mercator map based upon a cylinder tangent along the great circle through Washington and Moscow. The map includes an area 500 miles on each side of the great circle. The limits of this map are indicated on the Mercator map of Figure 309c.

310. Rectangular Projection

A cylindrical projection similar to the Mercator, but with uniform spacing of the parallels, is called a **rectangular projection**. It is convenient for graphically depicting information where distortion is not important. The principal navigational use of this projection is for the star chart of the Air Almanac, where positions of stars are plotted by rectangular coordinates representing declination (ordinate) and sidereal hour angle (abscissa). Since the meridians are parallel, the parallels of latitude (including the equator and the poles) are all represented by lines of equal length.

311. Conic Projections

A **conic projection** is produced by transferring points from the surface of the earth to a cone or series of cones. This cone is then cut along an element and spread out flat to form the chart. When the axis of the cone coincides with the axis of the earth, then the parallels appear as arcs of circles, and the meridians appear as either straight or curved lines converging toward the nearer pole. Limiting the area covered to that part of the cone near the surface of the earth limits distortion. A parallel along which there is no distortion is called a **standard parallel**. Neither the transverse conic projection, in which the axis of the cone is in the equatorial plane, nor the oblique conic projection, in which the axis of the cone is oblique to the plane of the equator, is ordinarily used for navigation. They are typically used for illustrative maps.

Using cones tangent at various parallels, a secant (intersecting) cone, or a series of cones varies the appearance and features of a conic projection.

312. Simple Conic Projection

A conic projection using a single tangent cone is a **simple conic projection** (Figure 312a). The height of the cone increases as the latitude of the tangent parallel decreases. At the equator, the height reaches infinity and the cone becomes a cylinder. At the pole, its height is zero, and the cone becomes a plane. Similar to the Mercator projection, the simple conic projection is not perspective since only the meridians are projected geometrically, each becoming an element of the cone. When this projection is spread out flat to form a map, the meridians appear as straight lines converging at the apex of the cone. The standard parallel, where the cone is tangent to the earth, appears as the arc of a circle with its center at the apex of the cone. The other parallels are concentric circles. The distance along any meridian between consecutive parallels is in correct relation to the distance on the earth, and, therefore, can be derived mathematically. The pole is represented by a circle (Figure 312b). The scale is correct along any meridian and along the standard parallel. All other parallels are too great in length, with the error increasing with increased distance from the standard parallel. Since the scale is not the same in all directions about every point, the projection is neither a conformal nor equal-area projection. Its non-conformal nature is its principal disadvantage for navigation.

Since the scale is correct along the standard parallel and varies uniformly on each side, with comparatively little distortion near the standard parallel, this projection is useful for mapping an area covering a large spread of longitude and a comparatively narrow band of latitude. It was developed by Claudius Ptolemy in the second century A.D. to map just such an area: the Mediterranean Sea.

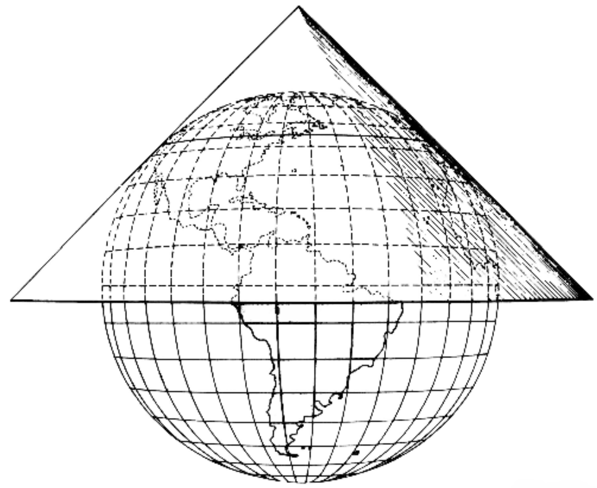


Figure 312a. A simple conic projection.

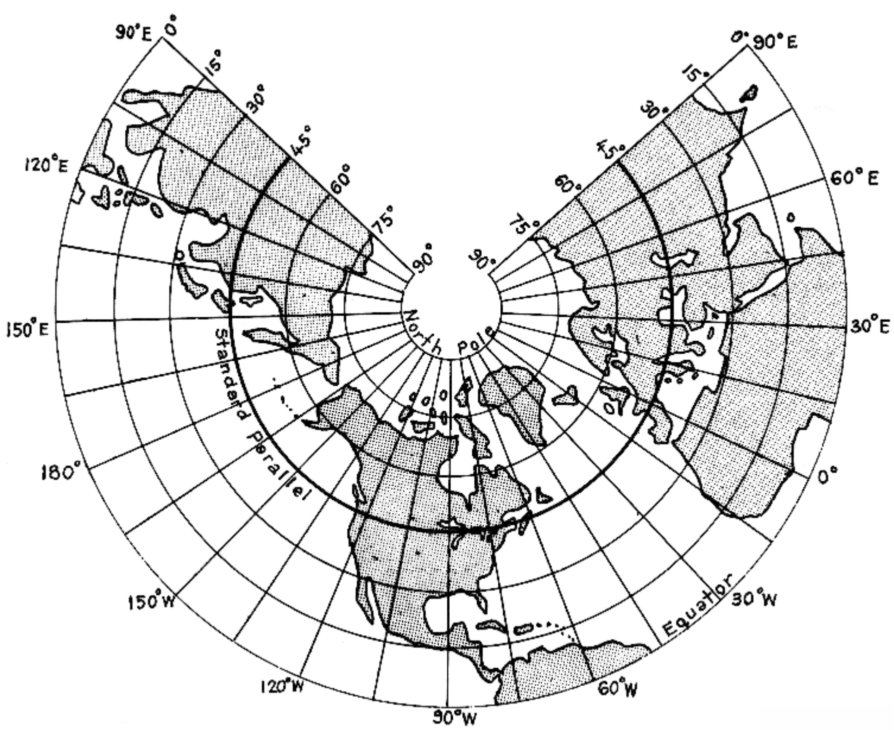


Figure 312b. A simple conic map of the Northern Hemisphere.

313. Lambert Conformal Projection

The useful latitude range of the simple conic projection can be increased by using a secant cone intersecting the earth at two standard parallels. See Figure 313. The area between the two standard parallels is compressed, and that beyond is expanded. Such a projection is called either a **secant conic** or **conic projection with two standard parallels**.

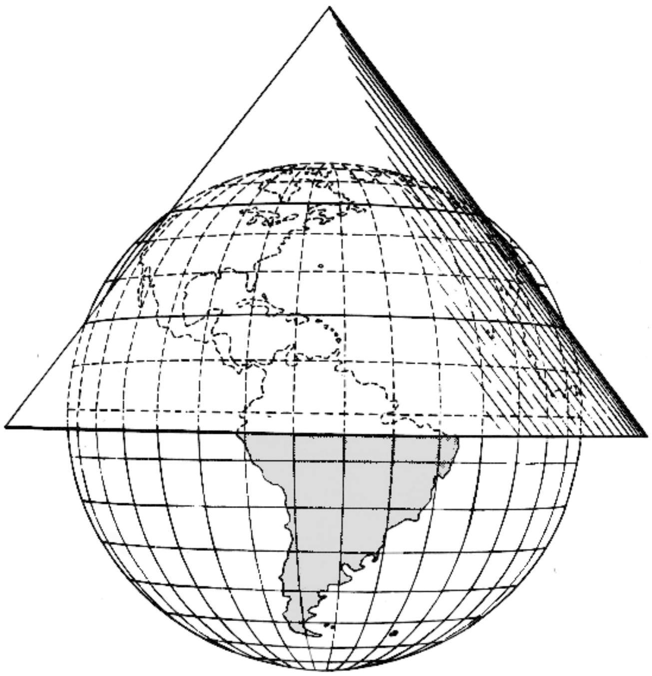


Figure 313. A secant cone for a conic projection with two standard parallels.

If in such a projection the spacing of the parallels is altered, such that the distortion is the same along them as along the meridians, the projection becomes conformal. This modification produces the **Lambert conformal projection**. If the chart is not carried far beyond the standard parallels, and if these are not a great distance apart, the distortion over the entire chart is small.

A straight line on this projection so nearly approximates a great circle that the two are nearly identical. Radio beacon signals travel great circles; thus, they can be plotted on this projection without correction. This feature, gained without sacrificing conformality, has made this projection popular for aeronautical charts because aircraft make wide use of radio aids to navigation. Except in high latitudes, where a slightly modified form of this projection has been used for polar charts, it has not replaced the Mercator projection for marine navigation.

314. Polyconic Projection

The latitude limitations of the secant conic projection can be minimized by using a series of cones. This results in a **polyconic projection**. In this projection, each parallel is the base of a tangent cone. At the edges of the chart, the area between parallels is expanded to eliminate gaps. The scale is correct along any parallel and along the central meridian of the projection. Along other meridians the scale increases with increased difference of longitude from the central meridian. Parallels appear as nonconcentric circles; meridians appear as curved lines converging toward the pole and concave to the central meridian.

The polyconic projection is widely used in atlases, particularly for areas of large range in latitude and reasonably large range in longitude, such as continents. However, since it is not conformal, this projection is not customarily used in navigation.

315. Azimuthal Projections

If points on the earth are projected directly to a plane surface, a map is formed at once, without cutting and flattening, or "developing." This can be considered a special case of a conic projection in which the cone has zero height.

The simplest case of the **azimuthal projection** is one in which the plane is tangent at one of the poles. The meridians are straight lines intersecting at the pole, and the parallels are concentric circles with their common center at the pole. Their spacing depends upon the method used to transfer points from the earth to the plane.

If the plane is tangent at some point other than a pole, straight lines through the point of tangency are great circles, and concentric circles with their common center at the point of tangency connect points of equal distance from that point. Distortion, which is zero at the point of tangency, increases along any great circle through this point. Along any circle whose center is the point of tangency, the distortion is constant. The bearing of any point from the point of tangency is correctly represented. It is for this reason that these projections are called **azimuthal**. They are also called **zenithal**. Several of the common azimuthal projections are perspective.

316. Gnomonic Projection

If a plane is tangent to the earth, and points are projected geometrically from the center of the earth, the result is a **gnomonic projection**. See Figure 316a. Since the projection is perspective, it can be demonstrated by placing a light at the center of a transparent terrestrial globe and holding a flat surface tangent to the sphere.

In an **oblique gnomonic projection** the meridians appear as straight lines converging toward the nearer pole. The parallels, except the equator, appear as curves (Figure 316b). As in all azimuthal projections, bearings from the point of tangency are correctly represented. The distance scale, however, changes rapidly. The projection is neither conformal nor equal area. Distortion is so great that shapes, as well as distances and areas, are very poorly represented, except near the point of tangency.

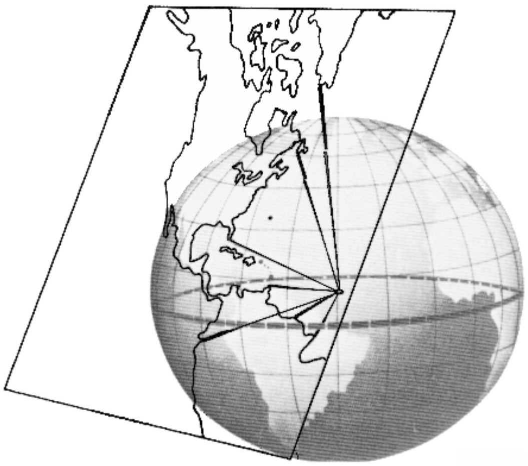


Figure 316a. An oblique gnomonic projection.

The usefulness of this projection rests upon the fact that any great circle appears on the map as a straight line, giving charts made on this projection the common name **great-circle charts**.

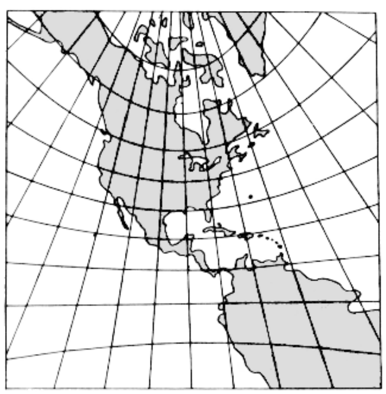


Figure 316b. An oblique gnomonic map with point of tangency at latitude 30°N, longitude 90°W.

Gnomonic charts are most often used for planning the great-circle track between points. Points along the determined track are then transferred to a Mercator projection. The great circle is then followed by following the rhumb lines from one point to the next. Computer programs which automatically calculate great circle routes between points and provide latitude and longitude of corresponding rhumb line endpoints are quickly making this use of the gnomonic chart obsolete.

317. Stereographic Projection

A **stereographic projection** results from projecting points on the surface of the earth onto a tangent plane, from a point on the surface of the earth opposite the point of tangency (Figure 317a). This projection is also called an **azimuthal orthomorphic projection**.

The scale of the stereographic projection increases with distance from the point of tangency, but it increases more slowly than in the gnomonic projection. The stereographic projection can show an entire hemisphere without excessive distortion (Figure 317b). As in other azimuthal

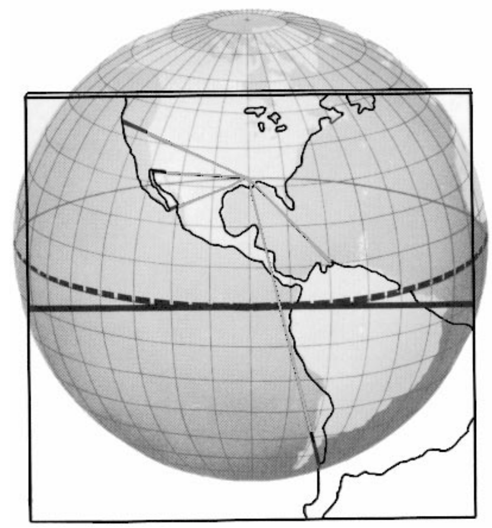


Figure 317a. An equatorial stereographic projection.

Figure 317b. A stereographic map of the Western Hemisphere.

projections, great circles through the point of tangency appear as straight lines. Other circles such as meridians and parallels appear as either circles or arcs of circles.

The principal navigational use of the stereographic projection is for charts of the polar regions and devices for mechanical or graphical solution of the navigational triangle. A **Universal Polar Stereographic (UPS)** grid, mathematically adjusted to the graticule, is used as a reference system.

318. Orthographic Projection

If terrestrial points are projected geometrically from infinity to a tangent plane, an **orthographic projection** results (Figure 318a). This projection is not conformal; nor does it result in an equal area representation. Its principal use is in navigational astronomy because it is useful for illustrating and solving the navigational triangle. It is also useful for illustrating celestial coordinates. If the plane is tangent at a point on the equator, the parallels (including the equator) appear as straight lines. The meridians would appear as ellipses, except that the meridian through the point of tangency would appear as a straight line and the one 90° away would appear as a circle (Figure 318b).

319. Azimuthal Equidistant Projection

An **azimuthal equidistant projection** is an azimuthal projection in which the distance scale along any great circle through the point of tangency is constant. If a pole is the point of tangency, the meridians appear as straight radial lines and the parallels as equally spaced concentric circles. If the plane is tangent at some point other than a pole, the concentric circles represent distances from the point of tangency. In this case, meridians and parallels appear as curves.

The projection can be used to portray the entire earth, the point 180° from the point of tangency appearing as the largest of the concentric circles. The projection is not conformal, equal area, or perspective. Near the point of tangency distortion is small, increasing with distance until shapes near the opposite side of the earth are unrecognizable (Figure 319).

The projection is useful because it combines the three features of being azimuthal, having a constant distance scale from the point of tangency, and permitting the entire earth to be shown on one map. Thus, if an important harbor or airport is selected as the point of tangency, the great-circle course, distance, and track from that point to any other point on the earth are quickly and accurately determined. For communication work with the station at the point of tangency, the path of an incoming signal is at once apparent if the direction of arrival has been determined and the direction to train a directional antenna can be determined easily. The projection is also used for polar charts and for the star finder, No. 2102D.

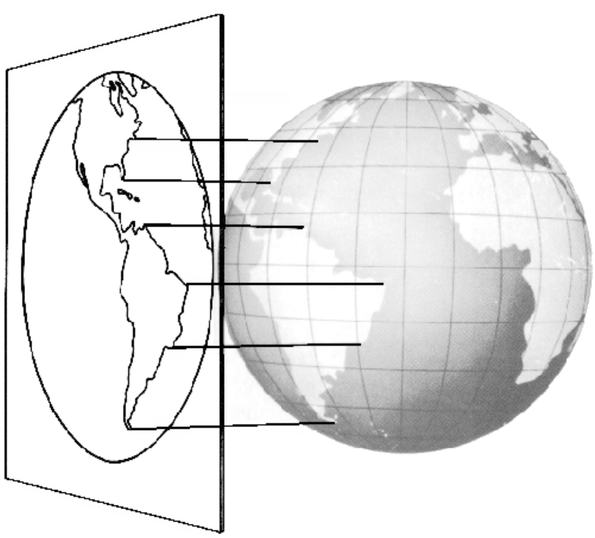


Figure 318a. An equatorial orthographic projection.

Figure 318b. An orthographic map of the Western Hemisphere.

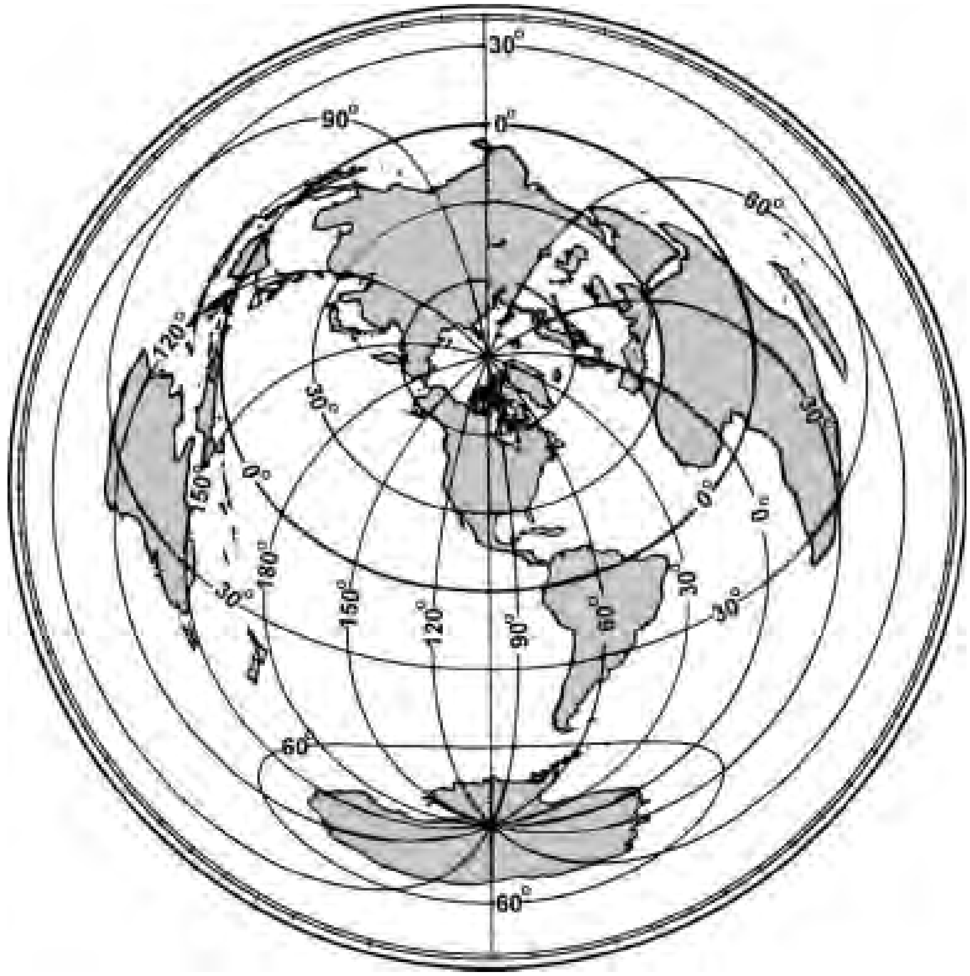


Figure 319. An azimuthal equidistant map of the world with the point of tangency latitude 40°N, longitude 100°W.

POLAR CHARTS

320. Polar Projections

Special consideration is given to the selection of projections for polar charts because the familiar projections become special cases with unique features.

In the case of cylindrical projections in which the axis of the cylinder is parallel to the polar axis of the earth, distortion becomes excessive and the scale changes rapidly. Such projections cannot be carried to the poles. However, both the transverse and oblique Mercator projections are used.

Conic projections with their axes parallel to the earth's polar axis are limited in their usefulness for polar charts because parallels of latitude extending through a full 360° of longitude appear as arcs of circles rather than full circles. This is because a cone, when cut along an element and flattened, does not extend through a full 360° without stretching or resuming its former conical shape. The usefulness of such projections is also limited by the fact that the pole appears as an arc of a circle instead of a point. However, by using a parallel very near the pole as the higher standard parallel, a conic projection with two standard parallels can be made. This requires little stretching to complete the circles of the parallels and eliminate that of the pole. Such a projection, called a **modified Lambert conformal** or **Ney's projection**, is useful for polar charts. It is particularly familiar to those accustomed to using the ordinary Lambert conformal charts in lower latitudes.

Azimuthal projections are in their simplest form when tangent at a pole. This is because the meridians are straight lines intersecting at the pole, and parallels are concentric circles with their common center at the pole. Within a few degrees of latitude of the pole they all look similar; however, as the distance becomes greater, the spacing of the parallels becomes distinctive in each projection. In the polar azimuthal equidistant it is uniform; in the polar stereographic it increases with distance from the pole until the equator is shown at a distance from the pole equal to twice the length of the radius of the earth; in the polar gnomonic the increase is considerably greater, becoming infinity at the equator; in the polar orthographic it decreases with distance from the pole (Figure 320). All of these but the last are used for polar charts.

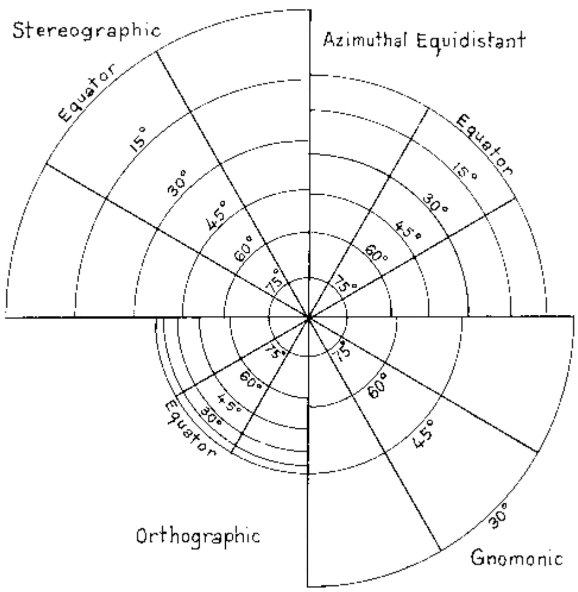


Figure 320. Expansion of polar azimuthal projections.

321. Selection of a Polar Projection

The principal considerations in the choice of a suitable projection for polar navigation are:

1. Conformality: When the projection represents angles correctly, the navigator can plot directly on the chart.
2. Great circle representation: Because great circles are more useful than rhumb lines at high altitudes, the projection should represent great circles as straight lines.
3. Scale variation: The projection should have a constant scale over the entire chart.
4. Meridian representation: The projection should show straight meridians to facilitate plotting and grid navigation
5. Limits: Wide limits reduce the number of projections needed to a minimum.

The projections commonly used for polar charts are the modified Lambert conformal, gnomonic, stereographic, and azimuthal equidistant. All of these projections are similar near the pole. All are essentially conformal, and a great circle on each is nearly a straight line.

As the distance from the pole increases, however, the distinctive features of each projection become important. The modified Lambert conformal projection is virtually conformal over its entire extent. The amount of its scale distortion is comparatively little if it is carried only to about 25° or 30° from the pole. Beyond this, the distortion increases rapidly. A great circle is very nearly a straight line anywhere on the chart. Distances and directions can be measured directly on the chart in the same manner as on a Lambert conformal chart. However, because this projection is not strictly conformal, and on it great circles are not exactly represented by straight lines, it is not suited for highly accurate work.

The polar gnomonic projection is the one polar projection on which great circles are exactly straight lines. However, a complete hemisphere cannot be represented upon a plane because the radius of 90° from the center would become infinity.

The polar stereographic projection is conformal over its entire extent, and a straight line closely approximates a great circle. See Figure 321. The scale distortion is not excessive for a considerable distance from the pole, but it is greater than that of the modified Lambert conformal projection.

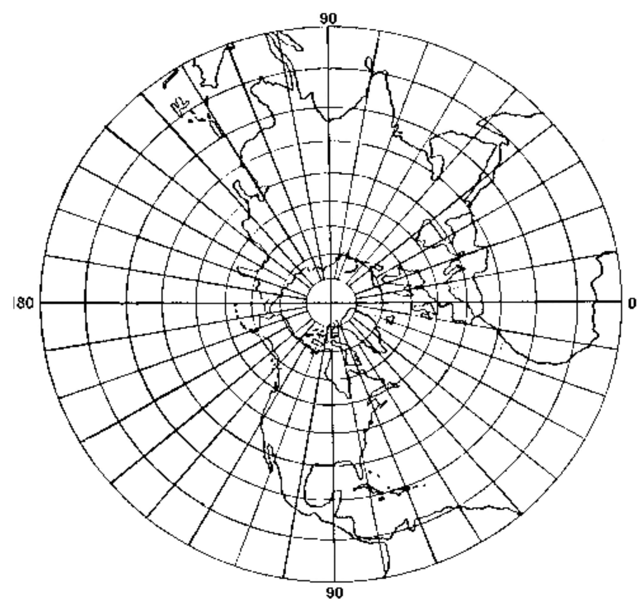


Figure 321. Polar stereographic projection.

The polar azimuthal equidistant projection is useful for showing a large area such as a hemisphere because there is no expansion along the meridians. However, the projection is not conformal and distances cannot be measured accurately in any but a north-south direction. Great circles other than the meridians differ somewhat from straight lines. The equator is a circle centered at the pole.

The two projections most commonly used for polar charts are the modified Lambert conformal and the polar stereographic. When a directional gyro is used as a directional reference, the track of the craft is approximately a great circle. A desirable chart is one on which a great circle is represented as a straight line with a constant scale and with angles correctly represented. These requirements are not met entirely by any single projection, but they are approximated by both the modified Lambert conformal and the polar stereographic. The scale is more nearly constant on the former, but the projection is not strictly conformal. The polar stereographic is conformal, and its maximum scale variation can be reduced by using a plane which intersects the earth at some parallel intermediate between the pole and the lowest parallel. The portion within this standard parallel is compressed, and that portion outside is expanded.

The selection of a suitable projection for use in polar regions depends upon mission requirements. These requirements establish the relative importance of various features. For a relatively small area, any of several projections is suitable. For a large area, however, the choice is more difficult. If grid directions are to be used, it is important that all units in related operations use charts on the same projection, with the same standard parallels, so that a single grid direction exists between any two points.

SPECIAL CHARTS

322. Plotting Sheets

Position plotting sheets are "charts" designed primarily for open ocean navigation, where land, visual aids to navigation, and depth of water are not factors in navigation. They have a latitude and longitude graticule, and they may have one or more compass roses. The meridians are usually unlabeled, so a plotting sheet can be used for any longitude. Plotting sheets on Mercator projection are specific to latitude, and the navigator should have enough aboard for all latitudes for his voyage. Plotting sheets are less expensive than charts.

A plotting sheet may be used in an emergency when charts have been lost or destroyed. Directions on how to construct plotting sheets suitable for emergency purposes are given in Chapter 26, Emergency Navigation.

323. Grids

No system exists for showing the surface of the earth on a plane without distortion. Moreover, the appearance of the surface varies with the projection and with the relation of that surface area to the point of tangency. One may want to identify a location or area simply by alpha-numeric rectangular coordinates. This is accomplished with a **grid**. In its usual form this consists of two series of lines drawn perpendicularly on the chart, marked by suitable alpha-numeric designations.

A grid may use the rectangular graticule of the Mercator projection or a set of arbitrary lines on a particular projection. **The World Geodetic Reference System (GEOREF)** is a method of designating latitude and longitude by a system of letters and numbers instead of by angular measure. It is not, therefore, strictly a grid. It is useful for operations extending over a wide area. Examples of the second type of grid are the **Universal Transverse Mercator (UTM)** grid, the **Universal Polar Stereographic (UPS)** grid, and the **Temporary Geographic Grid (TGG)**. Since these systems are used primarily by military forces, they are sometimes called military grids.

CHART SCALES

324. Types Of Scales

The **scale** of a chart is the ratio of a given distance on the chart to the actual distance which it represents on the earth. It may be expressed in various ways. The most common are:

1. A simple ratio or fraction, known as the **representative fraction**. For example, 1:80,000 or 1/80,000 means that one unit (such as a meter) on the chart represents 80,000 of the same unit on the surface of the earth. This scale is sometimes called the **natural** or **fractional** scale.
2. A **statement** that a given distance on the earth equals a given measure on the chart, or vice versa. For example, "30 miles to the inch" means that 1 inch on the chart represents 30 miles of the earth's surface. Similarly, "2 inches to a mile" indicates that 2 inches on the chart represent 1 mile on the earth. This is sometimes called the **numerical scale**.
3. A line or bar called a **graphic scale** may be drawn at a convenient place on the chart and subdivided into nautical miles, meters, etc. All charts vary somewhat in scale from point to point, and in some projections the scale is not the same in all directions about a single

point. A single subdivided line or bar for use over an entire chart is shown only when the chart is of such scale and projection that the scale varies a negligible amount over the chart, usually one of about 1:75,000 or larger. Since 1 minute of latitude is very nearly equal to 1 nautical mile, the latitude scale serves as an approximate graphic scale. On most nautical charts the east and west borders are subdivided to facilitate distance measurements.

On a Mercator chart the scale varies with the latitude. This is noticeable on a chart covering a relatively large distance in a north-south direction. On such a chart the border scale near the latitude in question should be used for measuring distances.

Of the various methods of indicating scale, the graphical method is normally available in some form on the chart. In addition, the scale is customarily stated on charts on which the scale does not change appreciably over the chart.

The ways of expressing the scale of a chart are readily interchangeable. For instance, in a nautical mile there are about 72,913.39 inches. If the natural scale of a chart is 1:80,000, one inch of the chart represents 80,000 inches of the earth, or a little more than a mile. To find the exact amount, divide the scale by the number of inches in a mile, or 80,000/72,913.39 = 1.097. Thus, a scale of 1:80,000 is the same as a scale of 1.097 (or approximately 1.1) miles to an inch. Stated another way, there are: 72,913.39/80,000 = 0.911 (approximately 0.9) inch to a mile. Similarly, if the scale is 60 nautical miles to an inch, the representative fraction is 1:(60 x 72,913.39) = 1:4,374,803.

A chart covering a relatively large area is called a **small-scale chart** and one covering a relatively small area is called a **large-scale chart**. Since the terms are relative, there is no sharp division between the two. Thus, a chart of scale 1:100,000 is large scale when compared with a chart of 1:1,000,000 but small scale when compared with one of 1:25,000.

As scale decreases, the amount of detail which can be shown decreases also. Cartographers selectively decrease the detail in a process called **generalization** when producing small scale charts using large scale charts as sources. The amount of detail shown depends on several factors, among them the coverage of the area at larger scales and the intended use of the chart.

325. Chart Classification by Scale

Charts are constructed on many different scales, ranging from about 1:2,500 to 1:14,000,000. Small-scale charts covering large areas are used for route planning and for offshore navigation. Charts of larger scale, covering smaller areas, are used as the vessel approaches land. Several methods of classifying charts according to scale are used in various nations. The following classifications of nautical charts are used by the National Ocean Service.

Sailing charts are the smallest scale charts used for planning, fixing position at sea, and for plotting the dead reckoning while proceeding on a long voyage. The scale is generally smaller than 1:600,000. The shoreline and topography are generalized and only offshore soundings, the principal navigational lights, outer buoys, and landmarks visible at considerable distances are shown.

General charts are intended for coastwise navigation outside of outlying reefs and shoals. The scales range from about 1:150,000 to 1:600,000.

Coastal charts are intended for inshore coastwise navigation, for entering or leaving bays and harbors of considerable width, and for navigating large inland waterways. The scales range from about 1:50,000 to 1:150,000.

Harbor charts are intended for navigation and anchorage in harbors and small waterways. The scale is generally larger than 1:50,000.

In the classification system used by NIMA, the sailing charts are incorporated in the general charts classification (smaller than about 1:150,000); those coast charts especially useful for approaching more confined waters (bays, harbors) are classified as approach charts. There is considerable overlap in these designations, and the classification of a chart is best determined by its use and by its relationship to other charts of the area. The use of insets complicates the placement of charts into rigid classifications.

CHART ACCURACY

326. Factors Relating to Accuracy

The accuracy of a chart depends upon the accuracy of the hydrographic surveys and other data sources used to compile it and the suitability of its scale for its intended use.

One can sometimes estimate the accuracy of a chart's surveys from the source notes given in the title of the chart. If the chart is based upon very old surveys, use it with caution. Many early surveys were inaccurate because of the technological limitations of the surveyor.

The number of soundings and their spacing indicates the completeness of the survey. Only a small fraction of the soundings taken in a thorough survey are shown on the chart, but sparse or unevenly distributed soundings indicate that the survey was probably not made in detail. See Figure 326a and Figure 326b. Large blank areas or absence of depth contours generally indicate lack of soundings in the area. Operate in an area with sparse sounding data only if required and then only with extreme caution. Run the echo sounder continuously and operate at a reduced speed.

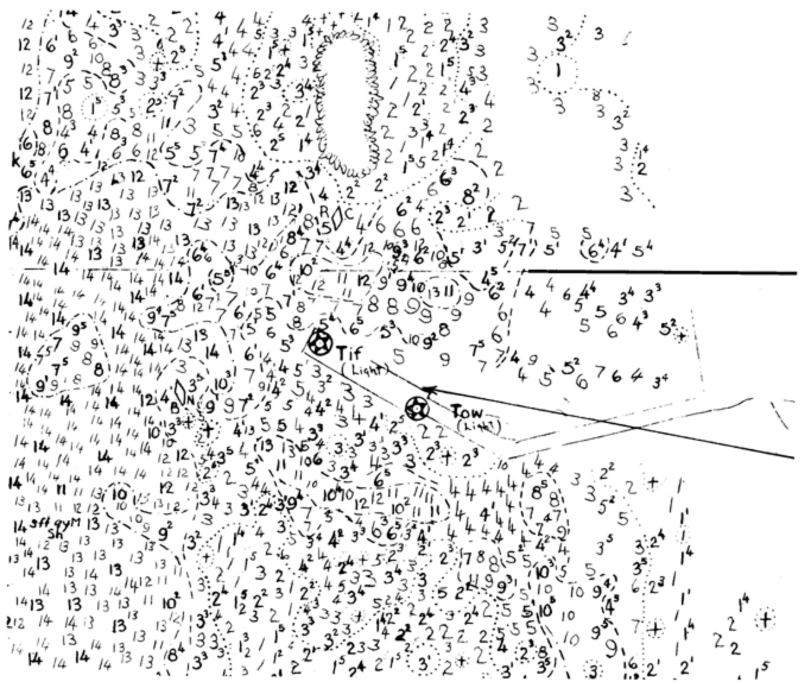


Figure 326a. Part of a "boat sheet," showing the soundings obtained in a survey.

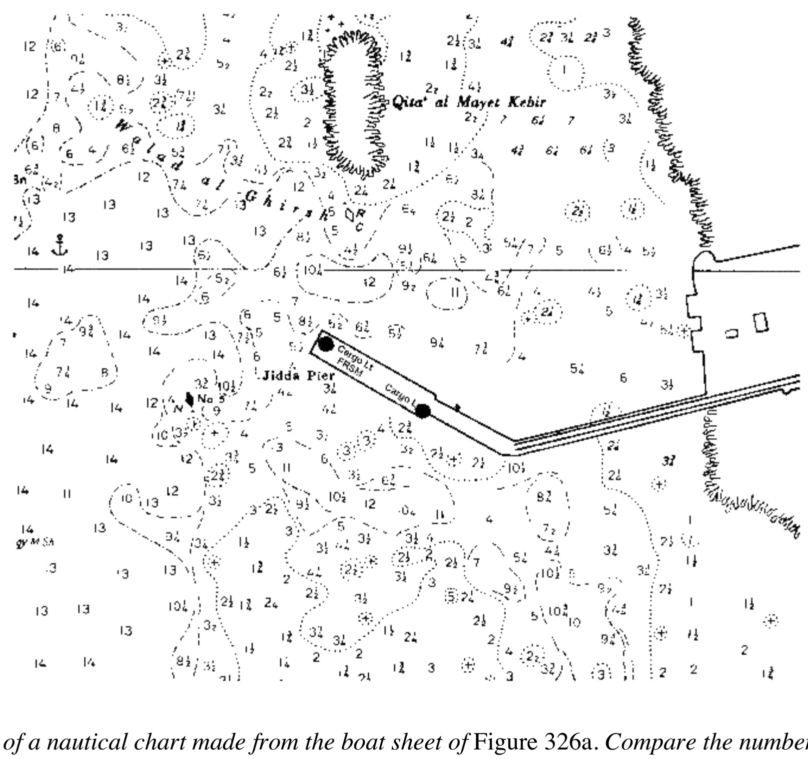


Figure 326b. Part of a nautical chart made from the boat sheet of Figure 326a. *Compare the number of soundings in the two figures.*

Sparse sounding information does not necessarily indicate an incomplete survey. Relatively few soundings are shown when there is a large number of depth contours, or where the bottom is flat, or gently and evenly sloping. Additional soundings are shown when they are helpful in indicating the uneven character of a rough bottom.

Even a detailed survey may fail to locate every rock or pinnacle. In waters where they might be located, the best method for finding them is a wire drag survey. Areas that have been dragged may be indicated on the chart by limiting lines and green or purple tint and a note added to show the effective depth at which the drag was operated.

Changes in bottom contours are relatively rapid in areas such as entrances to harbors where there are strong currents or heavy surf. Similarly, there is sometimes a tendency for dredged channels to shoal, especially if they are surrounded by sand or mud, and cross currents exist. Charts often contain notes indicating the bottom contours are known to change rapidly.

The same detail cannot be shown on a small-scale chart as on a large scale chart. On small-scale charts, detailed information is omitted or "generalized" in the areas covered by larger scale charts. The navigator should use the largest scale chart available for the area in which he is operating, especially when operating in the vicinity of hazards.

Charting agencies continually evaluate both the detail and the presentation of data appearing on a chart. Development of a new navigational aid may render previous charts inadequate. The development of radar, for example, required upgrading charts which lacked the detail required for reliable identification of radar targets.

After receiving a chart, the user is responsible for keeping it updated. Mariner's reports of errors, changes, and suggestions are useful to charting agencies. Even with modern automated data collection techniques, there is no substitute for on-sight observation of hydrographic conditions by experienced mariners. This holds true especially in less frequently traveled areas of the world.

CHART READING

327. Chart Dates

NOS charts have two dates. At the top center of the chart is the date of the first edition of the chart. In the lower left corner of the chart is the current edition number and date. This date shows the latest date through which *Notice to Mariners* were applied to the chart. Any subsequent change will be printed in the *Notice to Mariners*. Any notices which accumulate between the chart date and the announcement date in the *Notice to Mariners* will be given with the announcement. Comparing the dates of the first and current editions gives an indication of how often the chart is updated. Charts of busy areas are updated more frequently than those of less traveled areas. This interval may vary from 6 months to more than ten years for NOS charts. This update interval may be much longer for certain NIMA charts in remote areas.

New editions of charts are both demand and source driven. Receiving significant new information may or may not initiate a new edition of a chart, depending on the demand for that chart. If it is in a sparsely-traveled area, other priorities may delay a new edition for several years. Conversely, a new edition may be printed without the receipt of significant new data if demand for the chart is high and stock levels are low. *Notice to Mariners* corrections are always included on new editions.

NIMA charts have the same two dates as the NOS charts; the current chart edition number and date is given in the lower left corner. Certain NIMA charts are reproductions of foreign charts produced under joint agreements with a number of other countries. These charts, even though of recent date, may be based on foreign charts of considerably earlier date. Further, new editions of the foreign chart will not necessarily result in a new edition of the NIMA reproduction. In these cases, the foreign chart is the better chart to use.

328. Title Block

The chart title block should be the first thing a navigator looks at when receiving a new edition chart. Refer to Figure 328. The title itself tells what area the chart covers. The chart's scale and projection appear below the title. The chart will give both vertical and horizontal datums and, if necessary, a datum conversion note. Source notes or diagrams will list the date of surveys and other charts used in compilation.

329. Shoreline

The shoreline shown on nautical charts represents the line of contact between the land and water at a selected vertical datum. In areas affected by tidal fluctuations, this is usually the mean high-water line. In confined coastal waters of diminished tidal influence, a mean water level line may be used. The shoreline of interior waters (rivers, lakes) is usually a line representing a specified elevation above a

BALTIC SEA

GERMANY—NORTH COAST

DAHMESHÖVED TO WISMAR

From German Surveys

SOUNDINGS IN METERS

reduced to the approximate level of Mean Sea Level

HEIGHTS IN METERS ABOVE MEAN SEA LEVEL

MERCATOR PROJECTION

EUROPEAN DATUM

SCALE 1:50,000

Figure 328. A chart title block.

selected datum. A shoreline is symbolized by a heavy line. A broken line indicates that the charted position is approximate only. The nature of the shore may be indicated.

If the low water line differs considerably from the high water line, then a dotted line represents the low water line. If the bottom in this area is composed of mud, sand, gravel or stones, the type of material will be indicated. If the bottom is composed of coral or rock, then the appropriate symbol will be used. The area alternately covered and uncovered may be shown by a tint which is usually a combination of the land and water tint.

The apparent shoreline shows the outer edge of marine vegetation where that limit would appear as shoreline to the mariner. It is also used to indicate where marine vegetation prevents the mariner from defining the shoreline. A light line symbolizes this shoreline. A broken line marks the inner edge when no other symbol (such as a cliff or levee) furnishes such a limit. The combined land-water tint or the land tint marks the area between inner and outer limits.

330. Chart Symbols

Much of the information contained on charts is shown by symbols. These symbols are not shown to scale, but they indicate the correct position of the feature to which they refer. The standard symbols and abbreviations used on charts published by the United States of America are shown in ***Chart No. 1, Nautical Chart Symbols and Abbreviations***. See Figure 330.

Electronic chart symbols are, within programming and display limits, much the same as printed ones. The less expensive electronic charts have less extensive symbol libraries, and the screen's resolution may affect the presentation detail.

Most of the symbols and abbreviations shown in U.S. *Chart No. 1* agree with recommendations of the International Hydrographic Organization (IHO). The layout is explained in the general remarks section of *Chart No. 1*.

The symbols and abbreviations on any given chart may differ somewhat from those shown in *Chart No. 1*. In addition, foreign charts may use different symbology. When using a foreign chart, the navigator should have available the *Chart No. 1* from the country which produced the chart.

Chart No. 1 is organized according to subject matter, with each specific subject given a letter designator. The general subject areas are General, Topography, Hydrography, Aids and Services, and Indexes. Under each heading, letter designators further define subject areas, and individual numbers refer to specific symbols.

Information in *Chart No. 1* is arranged in columns. The first column contains the IHO number code for the symbol in question. The next two columns show the symbol itself, in NOS and NIMA formats. If the formats are the same, the two columns are combined into one. The next column is a text description of the symbol, term, or abbreviation. The next column contains the IHO standard symbol. The last column shows certain symbols used on foreign reproduction charts produced by NIMA.

331. Lettering

Except on some modified reproductions of foreign charts, cartographers have adopted certain lettering stan-

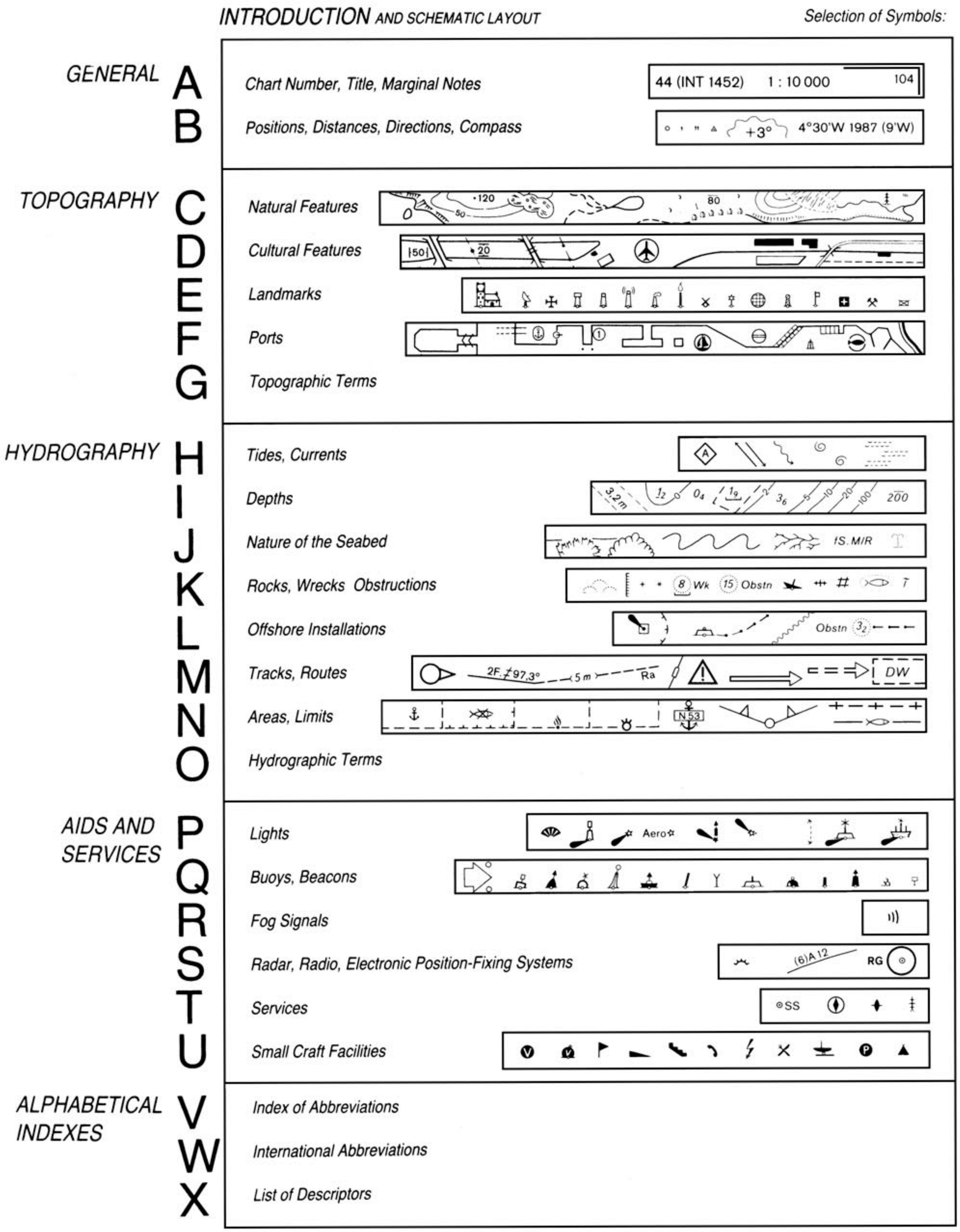


Figure 330. Contents of U.S. Chart No. 1.

dards. Vertical type is used for features which are dry at high water and not affected by movement of the water; slanting type is used for underwater and floating features.

There are two important exceptions to the two general rules listed above. Vertical type is not used to represent heights above the waterline, and slanting type is not used to indicate soundings, except on metric charts. Section 332 below discusses the conventions for indicating soundings.

Evaluating the type of lettering used to denote a feature, one can determine whether a feature is visible at high tide. For instance, a rock might bear the title "Rock" whether or not it extends above the surface. If the name is given in vertical letters, the rock constitutes a small islet; if in slanting type, the rock constitutes a reef, covered at high water.

332. Soundings

Charts show soundings in several ways. Numbers denote individual soundings. These numbers may be either vertical or slanting; both may be used on the same chart, distinguishing between data based upon different U.S. and foreign surveys, different datums, or smaller scale charts.

Large block letters at the top and bottom of the chart indicate the unit of measurement used for soundings. SOUNDINGS IN FATHOMS indicates soundings are in fathoms or fathoms and fractions. SOUNDINGS IN FATHOMS AND FEET indicates the soundings are in fathoms and feet. A similar convention is followed when the soundings are in meters or meters and tenths.

A **depth conversion scale** is placed outside the neatline on the chart for use in converting charted depths to feet, meters, or fathoms. "No bottom" soundings are indicated by a number with a line over the top and a dot over the line. This indicates that the spot was sounded to the depth indicated without reaching the bottom. Areas which have been wire dragged are shown by a broken limiting line, and the clear effective depth is indicated, with a characteristic symbol under the numbers. On NIMA charts a purple or green tint is shown within the swept area.

Soundings are supplemented by **depth contours**, lines connecting points of equal depth. These lines present a picture of the bottom. The types of lines used for various depths are shown in Section I of Chart No. 1. On some charts depth contours are shown in solid lines; the depth represented by each line is shown by numbers placed in breaks in the lines, as with land contours. Solid line depth contours are derived from intensively developed hydrographic surveys. A broken or indefinite contour is substituted for a solid depth contour whenever the reliability of the contour is questionable.

Depth contours are labeled with numerals in the unit of measurement of the soundings. A chart presenting a more detailed indication of the bottom configuration with fewer numerical soundings is useful when bottom contour navigating. Such a chart can be made only for areas which have undergone a detailed survey

Shoal areas often are given a blue tint. Charts designed to give maximum emphasis to the configuration of the bottom show depths beyond the 100-fathom curve over the entire chart by depth contours similar to the contours shown on land areas to indicate graduations in height. These are called **bottom contour** or **bathymetric charts**.

On electronic charts, a variety of other color schemes may be used, according to the manufacturer of the system. Color perception studies are being used to determine the best presentation.

The side limits of dredged channels are indicated by broken lines. The project depth and the date of dredging, if known, are shown by a statement in or along the channel. The possibility of silting is always present. Local authorities should be consulted for the controlling depth. NOS Charts frequently show controlling depths in a table, which is kept current by the *Notice to Mariners*.

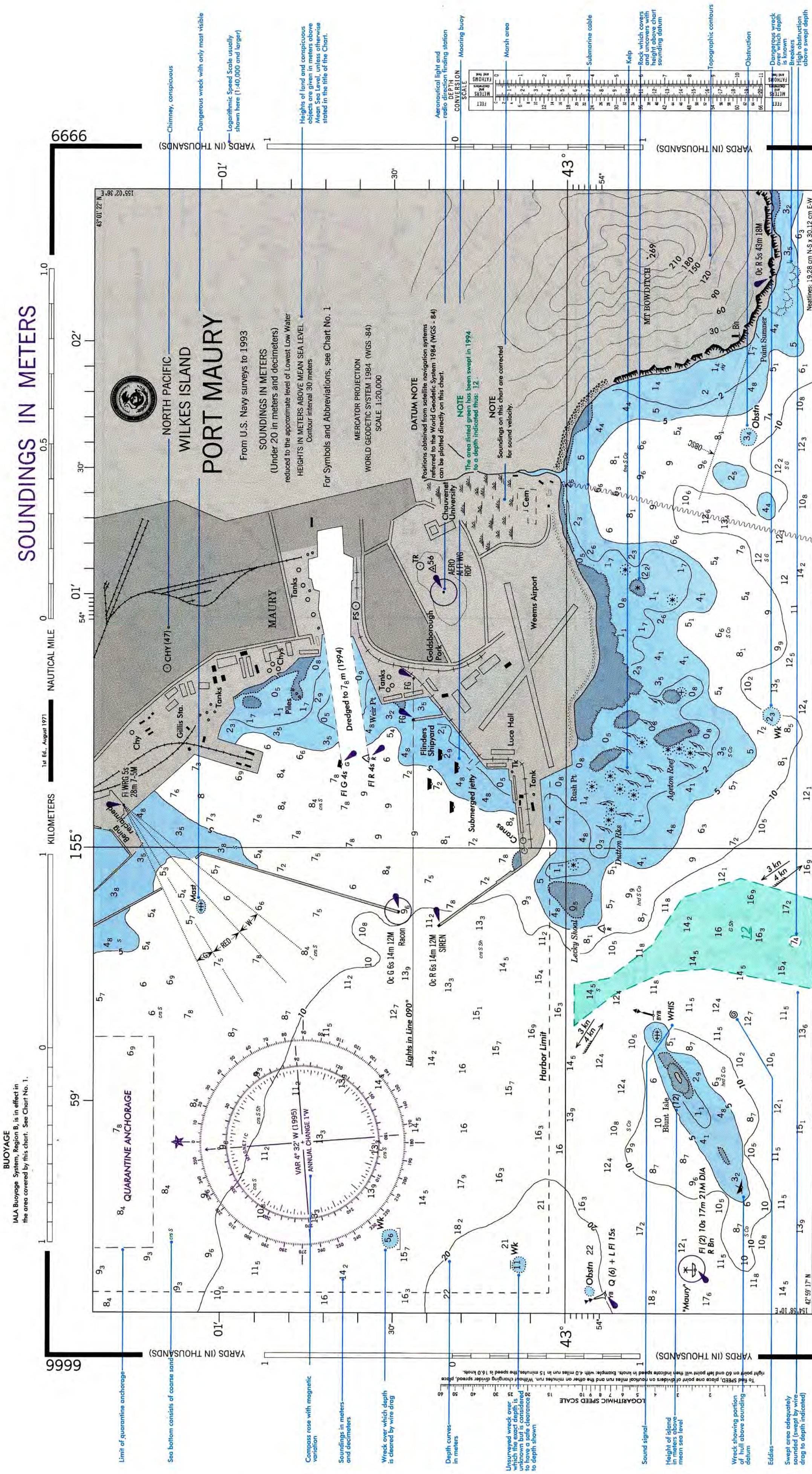
The chart scale is generally too small to permit all soundings to be shown. In the selection of soundings, least depths are shown first. This conservative sounding pattern provides safety and ensures an uncluttered chart appearance. Steep changes in depth may be indicated by more dense soundings in the area. The limits of shoal water indicated on the chart may be in error, and nearby areas of undetected shallow water may not be included on the chart. Given this possibility, areas where shoal water is known to exist should be avoided. If the navigator must enter an area containing shoals, he must exercise extreme caution in avoiding shallow areas which may have escaped detection. By constructing a "safety range" around known shoals and ensuring his vessel does not approach the shoal any closer than the safety range, the navigator can increase his chances of successfully navigating through shoal water. Constant use of the echo sounder is also important.

Abbreviations listed in Section J of Chart No. 1 are used to indicate what substance forms the bottom. The meaning of these terms can be found in the Glossary of this volume. While in ages past navigators might actually navigate by knowing the bottom characteristics of certain local areas, today knowing the characteristic of the bottom is most important when anchoring.

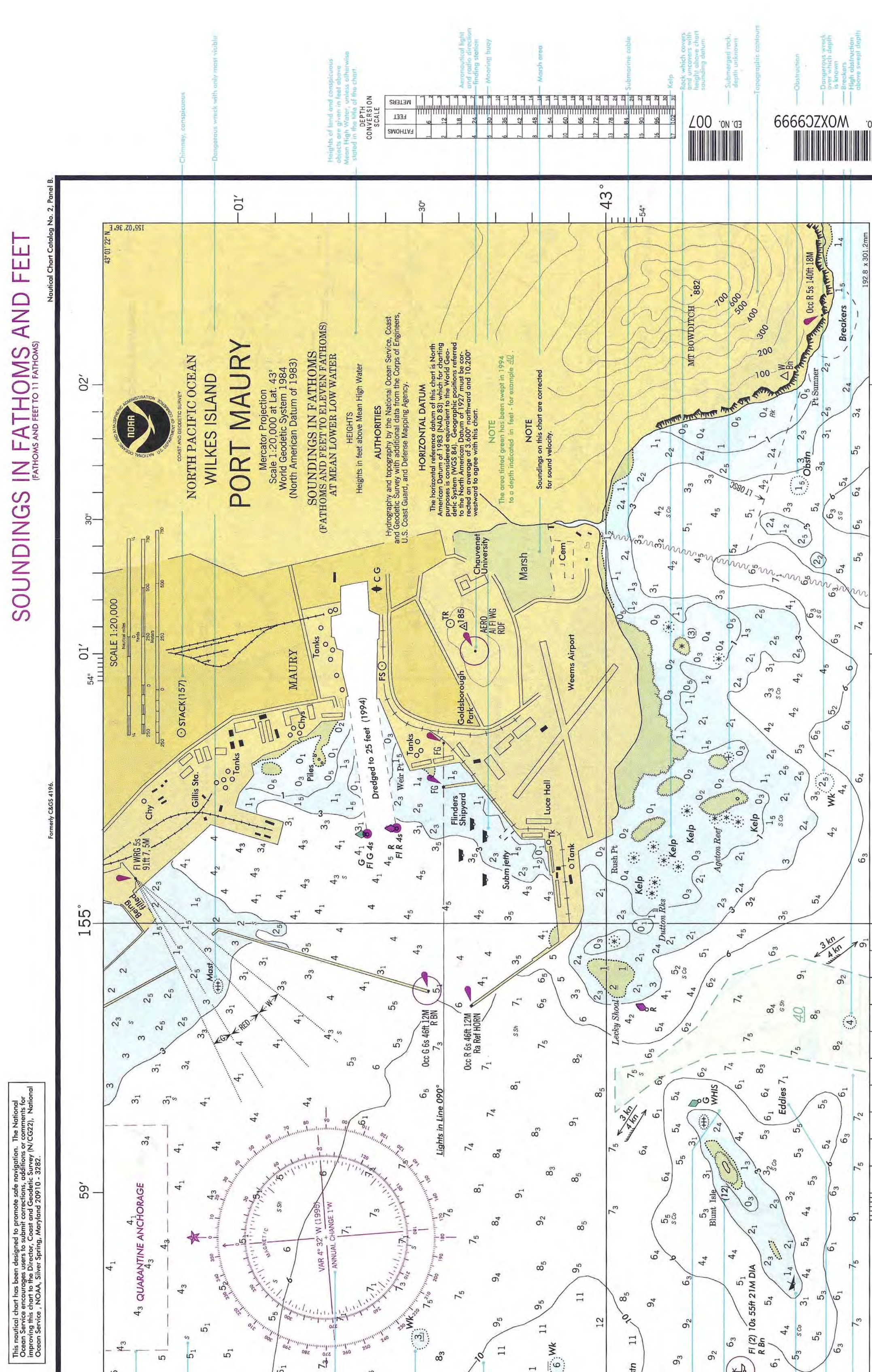
333. Depths and Datums

Depths are indicated by soundings or explanatory notes. Only a small percentage of the soundings obtained in a hydrographic survey can be shown on a nautical chart. The least depths are generally selected first, and a pattern built around them to provide a representative indication of bottom relief. In shallow water, soundings may be spaced 0.2 to 0.4 inch apart. The spacing is gradually increased as water deepens, until a spacing of 0.8 to 1.0 inch is reached in deeper waters offshore. Where a sufficient number of soundings are available to permit adequate interpretation, depth curves are drawn in at selected intervals.

All depths indicated on charts are reckoned from a selected level of the water, called the **sounding datum**, (sometimes referred to as the **reference plane** to distinguish this term from the geodetic datum). The various



SOUNDINGS IN METERS
NORTH PACIFIC
WILKES ISLAND
PORT MAURY
From U.S. Navy surveys to 1993
SOUNDINGS IN METERS
(Under 20 in meters and decimeters)
reduced to the approximate level of Lowest Low Water
HEIGHTS IN METERS ABOVE MEAN SEA LEVEL
Contour interval 30 meters
For Symbols and Abbreviations, see Chart No. 1
MERCATOR PROJECTION
WORLD GEODETIC SYSTEM 1984 (WGS-84)
SCALE 1:20,000
DATUM NOTE
NOTE
QUARANTINE ANCHORAGE
Harbor Limit
WARNING
Port Maury
SOUNDINGS IN METERS - SCALE 1:20,000
DMA STOCK NO. WOXZC9999



SOUNDINGS IN FATHOMS AND FEET
(FATHOMS AND FEET TO 11 FATHOMS)
NORTH PACIFIC OCEAN
WILKES ISLAND
PORT MAURY
Mercator Projection
Scale 1:20,000 at Lat. 43°
World Geodetic System 1984
(North American Datum of 1983)
SOUNDINGS IN FATHOMS
(FATHOMS AND FEET TO ELEVEN FATHOMS)
AT MEAN LOWER LOW WATER
HEIGHTS
Heights in feet above Mean High Water
AUTHORITIES
HORIZONTAL DATUM
NOTE
QUARANTINE ANCHORAGE
MT BOWDITCH
Weems Airport
Chauvenet University
Goldsborough Park
Flinders Shipyard
Luce Hall
Marsh
Eddies
Breakers
Blunt Isle
Port Maury
SOUNDINGS IN FATHOMS - SCALE 1:20,000
WARNING
CAUTION
LOGARITHMIC SPEED SCALE
DEPTH CONVERSION SCALE
ED. NO. 007
DMA STOCK NO. WOXZC9999
9999

sounding datums are explained in Chapter 9, Tides and Tidal Currents. On charts produced from U.S. surveys, the sounding datum is selected with regard to the tides of the region. Depths shown are the least depths to be expected under average conditions. On charts compiled from foreign charts and surveys the sounding datum is that of the original authority. When it is known, the sounding datum used is stated on the chart. In some cases where the chart is based upon old surveys, particularly in areas where the range of tide is not great, the sounding datum may not be known.

For most National Ocean Service charts of the United States and Puerto Rico, the sounding datum is mean lower low water. Most NIMA charts are based upon mean low water, mean lower low water, or mean low water springs. The sounding datum for charts published by other countries varies greatly, but is usually lower than mean low water. On charts of the Baltic Sea, Black Sea, the Great Lakes, and other areas where tidal effects are small or without significance, the sounding datum adopted is an arbitrary height approximating the mean water level.

The sounding datum of the largest scale chart of an area is generally the same as the reference level from which height of tide is tabulated in the tide tables.

The chart datum is usually only an approximation of the actual mean value, because determination of the actual mean height usually requires a longer series of tidal observations than is usually available to the cartographer. In addition, the heights of the tide vary over time.

Since the chart datum is generally a computed mean or average height at some state of the tide, the depth of water at any particular moment may be less than shown on the chart. For example, if the chart datum is mean lower low water, the depth of water at lower low water will be less than the charted depth about as often as it is greater. A lower depth is indicated in the tide tables by a minus sign (–).

334. Heights

The shoreline shown on charts is generally mean high water. A light's height is usually reckoned from mean sea level. The heights of overhanging obstructions (bridges, power cables, etc.) are usually reckoned from mean high water. A high water reference gives the mariner the minimum clearance expected.

Since heights are usually reckoned from high water and depths from some form of low water, the reference levels are seldom the same. Except where the range of tide is very large, this is of little practical significance.

335. Dangers

Dangers are shown by appropriate symbols, as indicated in Section K of *Chart No. 1*.

A rock uncovered at mean high water may be shown as an islet. If an isolated, offlying rock is known to uncover at the sounding datum but to be covered at high water, the chart shows the appropriate symbol for a rock and gives the height above the sounding datum. The chart can give this height one of two ways. It can use a statement such as "Uncov 2 ft.," or it can indicate the number of feet the rock protrudes above the sounding datum, underline this value, and enclose it in parentheses (i.e. (2)). A rock which does not uncover is shown by an enclosed figure approximating its dimensions and filled with land tint. It may be enclosed by a dotted depth curve for emphasis.

A tinted, irregular-line figure of approximately true dimensions is used to show a detached coral reef which uncovers at the chart datum. For a coral or rocky reef which is submerged at chart datum, the sunken rock symbol or an appropriate statement is used, enclosed by a dotted or broken line if the limits have been determined.

Several different symbols mark wrecks. The nature of the wreck or scale of the chart determines the correct symbol. A sunken wreck with less than 11 fathoms of water over it is considered dangerous and its symbol is surrounded by a dotted curve. The curve is omitted if the wreck is deeper than 11 fathoms. The safe clearance over a wreck, if known, is indicated by a standard sounding number placed at the wreck. If this depth was determined by a wire drag, the sounding is underscored by the wire drag symbol. An unsurveyed wreck over which the exact depth is unknown but a safe clearance depth is known is depicted with a solid line above the symbol.

Tide rips, eddies, and kelp are shown by symbol or legend. Piles, dolphins (clusters of piles), snags, and stumps are shown by small circles and a label identifying the type of obstruction. If such dangers are submerged, the letters "Subm" precede the label. Fish stakes and traps are shown when known to be permanent or hazardous to navigation.

336. Aids to Navigation

Aids to navigation are shown by symbols listed in Sections P through S of Chart No. 1. Abbreviations and additional descriptive text supplement these symbols. In order to make the symbols conspicuous, the chart shows them in size greatly exaggerated relative to the scale of the chart. "Position approximate" circles are used on floating aids to indicate that they have no exact position because they move around their moorings. For most floating aids, the position circle in the symbol marks the approximate location of the anchor or sinker. The actual aid may be displaced from this location by the scope of its mooring.

The type and number of aids to navigation shown on a chart and the amount of information given in their legends varies with the scale of the chart. Smaller scale charts may have fewer aids indicated and less information than larger scale charts of the same area.

Lighthouses and other navigation lights are shown as black dots with purple disks or as black dots with purple flare symbols. The center of the dot is the position of the light. Some modified facsimile foreign charts use a small star instead of a dot.

On large-scale charts the legend elements of lights are shown in the following order:

Legend	*Example*	*Meaning*
Characteristic	F1(2)	group flashing; 2 flashes
Color	R	red
Period	10s	2 flashes in 10 seconds
Height	80m	80 meters
Range	19M	19 nautical miles
Designation	"6"	light number 6

The legend for this light would appear on the chart:

Fl(2) R 10s 80m 19M "6"

As chart scale decreases, information in the legend is selectively deleted to avoid clutter. The order of deletion is usually height first, followed by period, group repetition interval (e.g. (2)), designation, and range. Characteristic and color will almost always be shown.

Small triangles mark red daybeacons; small squares mark all others. On NIMA charts, pictorial beacons are used when the IALA buoyage system has been implemented. The center of the triangle marks the position of the aid. Except on Intracoastal Waterway charts and charts of state waterways, the abbreviation "Bn" is shown beside the symbol, along with the appropriate abbreviation for color if known. For black beacons the triangle is solid black and there is no color abbreviation. All beacon abbreviations are in vertical lettering.

Radiobeacons are indicated on the chart by a purple circle accompanied by the appropriate abbreviation indicating an ordinary radiobeacon (R Bn) or a radar beacon (Ramark or Racon, for example).

A variety of symbols, determined by both the charting agency and the types of buoys, indicate navigation buoys. IALA buoys (see Chapter 5, Short Range Aids to Navigation) in foreign areas are depicted by various styles of symbols with proper topmarks and colors; the position circle which shows the approximate location of the sinker is at the base of the symbol.

A mooring buoy is shown by one of several symbols as indicated in Chart No. 1. It may be labeled with a berth number or other information.

A buoy symbol with a horizontal line indicates the buoy has horizontal bands. A vertical line indicates vertical stripes; crossed lines indicate a checked pattern. There is no significance to the angle at which the buoy symbol appears on the chart. The symbol is placed so as to avoid interference with other features.

Lighted buoys are indicated by a purple flare from the buoy symbol or by a small purple disk centered on the position circle.

Abbreviations for light legends, type and color of buoy, designation, and any other pertinent information given near the symbol are in slanted type. The letter C, N, or S indicates a can, nun, or spar, respectively. Other buoys are assumed to be pillar buoys, except for special buoys such as spherical, barrel, etc. The number or letter designation of the buoy is given in quotation marks on NOS charts. On other charts they may be given without quotation marks or other punctuation.

Aeronautical lights included in the light lists are shown by the lighthouse symbol, accompanied by the abbreviation "AERO." The characteristics shown depend principally upon the effective range of other navigational lights in the vicinity and the usefulness of the light for marine navigation.

Directional ranges are indicated by a broken or solid line. The solid line, indicating that part of the range intended for navigation, may be broken at irregular intervals to avoid being drawn through soundings. That part of the range line drawn only to guide the eye to the objects to be kept in range is broken at regular intervals. The direction, if given, is expressed in degrees, clockwise from true north.

Sound signals are indicated by the appropriate word in capital letters (HORN, BELL, GONG, or WHIS) or an abbreviation indicating the type of sound. Sound signals of any type except submarine sound signals may be represented by three purple 45° arcs of concentric circles near the top of the aid. These are not shown if the type of signal is listed. The location of a sound signal which does not accompany a visual aid, either lighted or unlighted, is shown by a small circle and the appropriate word in vertical block letters.

Private aids, when shown, are marked "Priv" on NOS charts. Some privately maintained unlighted fixed aids are indicated by a small circle accompanied by the word "Marker," or a larger circle with a dot in the center and the word "MARKER." A privately maintained lighted aid has a light symbol and is accompanied by the characteristics and the usual indication of its private nature. Private aids should be used with caution.

A light sector is the sector or area bounded by two radii and the arc of a circle in which a light is visible or in which it has a distinctive color different from that of adjoining sectors. The limiting radii are indicated on the chart by dotted or dashed lines. Sector colors are indicated by words spelled out if space permits, or by abbreviations (W, R, etc.) if it does not. Limits of light sectors and arcs of visibility as observed from a vessel are given in the light lists, in clockwise order.

337. Land Areas

The amount of detail shown on the land areas of nautical charts depends upon the scale and the intended purpose of the chart. Contours, form lines, and shading indicate relief.

Contours are lines connecting points of equal elevation. Heights are usually expressed in feet (or in meters with means for conversion to feet). The interval between contours is uniform over any one chart, except that certain intermediate contours are sometimes shown by broken line. When contours are broken, their locations are approximate.

Form lines are approximations of contours used for the purpose of indicating relative elevations. They are used in areas where accurate information is not available in sufficient detail to permit exact location of contours. Elevations of individual form lines are not indicated on the chart.

Spot elevations are generally given only for summits or for tops of conspicuous landmarks. The heights of spot elevations and contours are given with reference to mean high water when this information is available.

When there is insufficient space to show the heights of islets or rocks, they are indicated by slanting figures enclosed in parentheses in the water area nearby.

338. Cities and Roads

Cities are shown in a generalized pattern that approximates their extent and shape. Street names are generally not charted except those along the waterfront on the largest scale charts. In general, only the main arteries and thoroughfares or major coastal highways are shown on smaller scale charts. Occasionally, highway numbers are given. When shown, trails are indicated by a light broken line. Buildings along the waterfront or individual ones back from the waterfront but of special interest to the mariner are shown on large-scale charts. Special symbols from Chart No. 1 are used for certain kinds of buildings. A single line with cross marks indicates both single and double track railroads. City electric railways are usually not charted. Airports are shown on small-scale charts by symbol and on large-scale charts by the shape of runways. The scale of the chart determines if single or double lines show breakwaters and jetties; broken lines show the submerged portion of these features.

339. Landmarks

Landmarks are shown by symbols in Chart No. 1.

A large circle with a dot at its center is used to indicate that the position is precise and may be used without reservation for plotting bearings. A small circle without a dot is used for landmarks not accurately located. Capital and lower case letters are used to identify an approximate landmark: “Mon,” “Cup,” or “Dome.” The abbreviation “PA” (position approximate) may also appear. An accurate landmark is identified by all capital type (“MON,” “CUP,” “DOME”).

When only one object of a group is charted, its name is followed by a descriptive legend in parenthesis, including the number of objects in the group, for example “(TALLEST OF FOUR)” or “(NORTHEAST OF THREE).”

340. Miscellaneous Chart Features

A measured nautical mile indicated on a chart is accurate to within 6 feet of the correct length. Most measured miles in the United States were made before 1959, when the United States adopted the International Nautical Mile. The new value is within 6 feet of the previous standard length of 6,080.20 feet. If the measured distance differs from the standard value by more than 6 feet, the actual measured distance is stated and the words “measured mile” are omitted.

Periods after abbreviations in water areas are omitted because these might be mistaken for rocks. However, a lower case i or j is dotted.

Commercial radio broadcasting stations are shown on charts when they are of value to the mariner either as landmarks or sources of direction-finding bearings.

Lines of demarcation between the areas in which international and inland navigation rules apply are shown only when they cannot be adequately described in notes on the chart.

Compass roses are placed at convenient locations on Mercator charts to facilitate the plotting of bearings and courses. The outer circle is graduated in degrees with zero at true north. The inner circle indicates magnetic north.

On many NIMA charts magnetic variation is given to the nearest 1' by notes in the centers of compass roses. the annual change is given to the nearest 1' to permit correction of the given value at a later date. On NOS charts, variation is to the nearest 15', updated at each new edition if over three years old. The current practice of NIMA is to give the magnetic variation to the nearest 1', but the magnetic information on new editions is only updated to conform with the latest five year epoch. Whenever a chart is reprinted, the magnetic information is updated to the latest epoch. On some smaller scale charts, the variation is given by isogonic lines connecting points of equal variation; usually a separate line represents each degree of variation. The line of zero variation is called the agonic line. Many plans and insets show neither compass roses nor isogonic lines, but indicate magnetic information by note. A local magnetic disturbance of sufficient force to cause noticeable deflection of the magnetic compass, called local attraction, is indicated by a note on the chart.

Currents are sometimes shown on charts with arrows giving the directions and figures showing speeds. The information refers to the usual or average conditions. According to tides and weather, conditions at any given time may differ considerably from those shown.

Review chart notes carefully because they provide important information. Several types of notes are used. Those in the margin give such information as chart number, publication notes, and identification of adjoining charts. Notes in connection with the chart title include information on scale, sources of data, tidal information, soundings, and cautions. Another class of notes covers such topics as local magnetic disturbance, controlling depths of channels, hazards to navigation, and anchorages.

A datum note will show the geodetic datum of the chart (Do not confuse with the sounding datum. See Chapter 2, Geodesy and Datums in Navigation.) It may also contain instructions on plotting positions from the WGS 84 or NAD 83 datums on the chart if such a conversion is needed.

Anchorage areas are labeled with a variety of magenta, black, or green lines depending on the status of the area. Anchorage berths are shown as purple circles, with the number or letter assigned to the berth inscribed within the circle. Caution notes are sometimes shown when there are specific anchoring regulations.

Spoil areas are shown within short broken black lines. Spoil areas are tinted blue on NOS charts and labeled. These areas contain no soundings and should be avoided.

Firing and bombing practice areas in the United States territorial and adjacent waters are shown on NOS and NIMA charts of the same area and comparable scale.

Danger areas established for short periods of time are not charted but are announced locally. Most military commands charged with supervision of gunnery and missile firing areas promulgate a weekly schedule listing activated danger areas. This schedule is subjected to frequent change; the mariner should always ensure he has the latest schedule prior to proceeding into a gunnery or missile firing area. Danger areas in effect for longer periods are published in the *Notice to Mariners*. Any aid to navigation established to mark a danger area or a fixed or floating target is shown on charts.

Traffic separation schemes are shown on standard nautical charts of scale 1:600,000 and larger and are printed in magenta.

A logarithmic time-speed-distance nomogram with an explanation of its application is shown on harbor charts.

Tidal information boxes are shown on charts of scales 1:200,000 and larger for NOS charts, and various scales on DMA charts, according to the source. See Figure 340a.

Tabulations of controlling depths are shown on some National Ocean Service harbor and coastal charts. See Figure 340b.

Study Chart No. 1 thoroughly to become familiar with all the symbols used to depict the wide variety of features on nautical charts.

TIDAL INFORMATION						
Place	Position		Height above datum of soundings			
			Mean High Water		Mean Low Water	
	N. Lat.	E. Long.	Higher	Lower	Lower	Higher
			meters	meters	meters	meters
Olongapo	14°49'	120°17'	. . . 0.9 . . .	. . . 0.4 . . .	. . . 0.0 . . .	. . . 0.3 . . .

Figure 340a. Tidal box.

NANTUCKET HARBOR Tabulated from surveys by the Corps of Engineers - report of June 1972 and surveys of Nov. 1971							
Controlling depths in channels entering from seaward in feet at Mean Low Water					Project Dimensions		
Name of Channel	Left outside quarter	Middle half of channel	Right outside quarter	Date of Survey	Width (feet)	Length (naut. miles)	Depth M. L. W. (feet)
Entrance Channel	11.1	15.0	15.0	11 - 71	300	1.2	15
Note.-The Corps of Engineers should be consulted for changing conditions subsequent to the above.							

Figure 340b. Tabulations of controlling depths.

REPRODUCTIONS OF FOREIGN CHARTS

341. Modified Facsimiles

Modified facsimile charts are modified reproductions of foreign charts produced in accordance with bilateral international agreements. These reproductions provide the mariner with up-to-date charts of foreign waters. Modified facsimile charts published by NIMA are, in general, reproduced with minimal changes, as listed below:

1. The original name of the chart may be removed and replaced by an anglicized version.
2. English language equivalents of names and terms on the original chart are printed in a suitable glossary on the reproduction, as appropriate.
3. All hydrographic information, except bottom characteristics, is shown as depicted on the original chart.
4. Bottom characteristics are as depicted in Chart No. 1, or as on the original with a glossary.
5. The unit of measurement used for soundings is shown in block letters outside the upper and lower neatlines.
6. A scale for converting charted depth to feet, meters, or fathoms is added.
7. Blue tint is shown from a significant depth curve to the shoreline.
8. Blue tint is added to all dangers enclosed by a dotted danger curve, dangerous wrecks, foul areas, obstructions, rocks awash, sunken rocks, and swept wrecks.
9. Caution notes are shown in purple and enclosed in a box.
10. Restricted, danger, and prohibited areas are usually outlined in purple and labeled appropriately.
11. Traffic separation schemes are shown in purple.
12. A note on traffic separation schemes, printed in black, is added to the chart.
13. Wire dragged (swept) areas are shown in purple or green.
14. Corrections are provided to shift the horizontal datum to the World Geodetic System (1984).

INTERNATIONAL CHARTS

342. International Chart Standards

The need for mariners and chart makers to understand and use nautical charts of different nations became increasingly apparent as the maritime nations of the world developed their own establishments for the compilation and publication of nautical charts from hydrographic surveys. Representatives of twenty-two nations formed a Hydrographic Conference in London in 1919. That conference resulted in the establishment of the **International Hydrographic Bureau (IHB)** in Monaco in 1921. Today, the IHB's successor, the **International Hydrographic Organization (IHO)** continues to provide international standards for the cartographers of its member nations. (See Chapter 1, Introduction to Marine Navigation, for a description of the IHO.)

Recognizing the considerable duplication of effort by member states, the IHO in 1967 moved to introduce the first **international chart**. It formed a committee of six member states to formulate specifications for two series of international charts. Eighty-three small-scale charts were approved; responsibility for compiling these charts has subsequently been accepted by the member states' Hydrographic Offices.

Once a Member State publishes an international chart, reproduction material is made available to any other Member State which may wish to print the chart for its own purposes.

International charts can be identified by the letters INT before the chart number and the International Hydrographic Organization seal in addition to other national seals which may appear.

CHART NUMBERING

343. The Chart Numbering System

NIMA and NOS use a system in which numbers are assigned in accordance with both the scale and geographical area of coverage of a chart. With the exception of certain charts produced for military use only, one- to five-digit numbers are used. With the exception of one-digit numbers, the first digit identifies the area; the number of digits establishes the scale range. The one-digit numbers are used for certain products in the chart system which are not actually charts.

Number of Digits	*Scale*
1	No Scale
2	1:9 million and smaller
3	1:2 million to 1:9 million
4	Special Purpose
5	1:2 million and larger

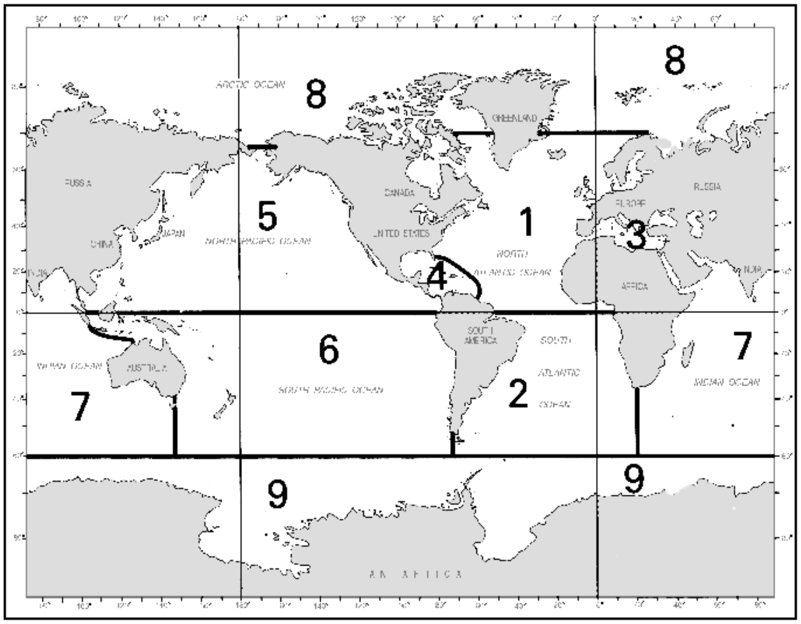


Figure 343a. Ocean basins with region numbers.

Two- and three-digit numbers are assigned to those small-scale charts which depict a major portion of an ocean basin or a large area. The first digit identifies the applicable ocean basin. See Figure 343a. Two-digit numbers are used for charts of scale 1:9,000,000 and smaller. Three-digit numbers are used for charts of scale 1:2,000,000 to 1:9,000,000.

Due to the limited sizes of certain ocean basins, no charts for navigational use at scales of 1:9,000,000 and smaller are published to cover these basins. The otherwise unused two-digit numbers (30 to 49 and 70 to 79) are assigned to special world charts.

One exception to the scale range criteria for three-digit numbers is the use of three-digit numbers for a series of position plotting sheets. They are of larger scale than 1:2,000,000 because they have application in ocean basins and can be used in all longitudes.

Four-digit numbers are used for non-navigational and special purpose charts, such as chart 5090, *Maneuvering Board*.

Five-digit numbers are assigned to those charts of scale 1:2,000,000 and larger that cover portions of the coastline rather than significant portions of ocean basins. These charts are based on the regions of the nautical chart index. See Figure 343b.

The first of the five digits indicates the region; the second digit indicates the subregion; the last three digits indicate the geographical sequence of the chart within the subregion. Many numbers have been left unused so that any future charts may be placed in their proper geographical sequence.

In order to establish a logical numbering system within the geographical subregions (for the 1:2,000,000 and larger-scale charts), a worldwide skeleton framework of coastal charts was laid out at a scale 1:250,000. This series was used as basic coverage except in areas where a coordinated series at about this scale already existed (such as the coast of Norway where a coordinated series of 1:200,000 charts was available).

Within each region, the geographical subregions are numbered counterclockwise around the continents, and within each subregion the basic series also is numbered counterclockwise around the continents. The basic coverage is assigned generally every 20th digit, except that the first 40 numbers in each subregion are reserved for smaller-scale coverage. Charts with scales larger than the basic coverage are assigned one of the 19 numbers following the number assigned to the sheet within which it falls. Figure 343c shows the numbering sequence in Iceland. Note the sequence of numbers around the coast, the direction of numbering, and the numbering of larger scale charts within the limits of smaller scales.

Five-digit numbers are also assigned to the charts produced by other hydrographic offices. This numbering system is applied to foreign charts so that they can be filed in logical sequence with the charts produced by the National Imagery and Mapping Agency and the National Ocean Service.

Certain exceptions to the standard numbering system have been made for charts intended for the military. Bottom contour charts depict parts of ocean basins. They are identified with a letter plus four digits according to a scheme best shown in the catalog, and are not available to civilian navigators.

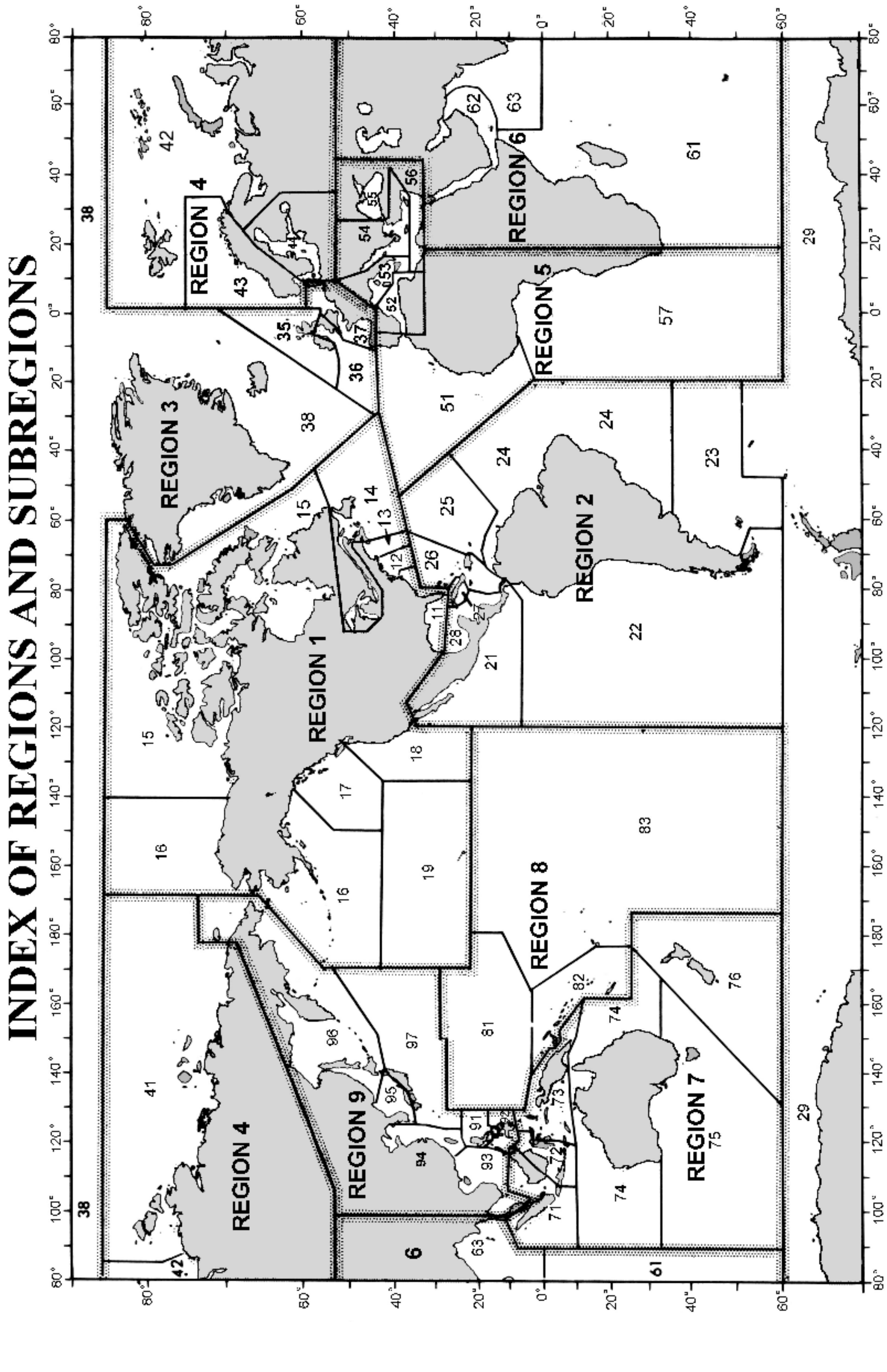


Figure 343b. Regions and subregions of the nautical chart index.

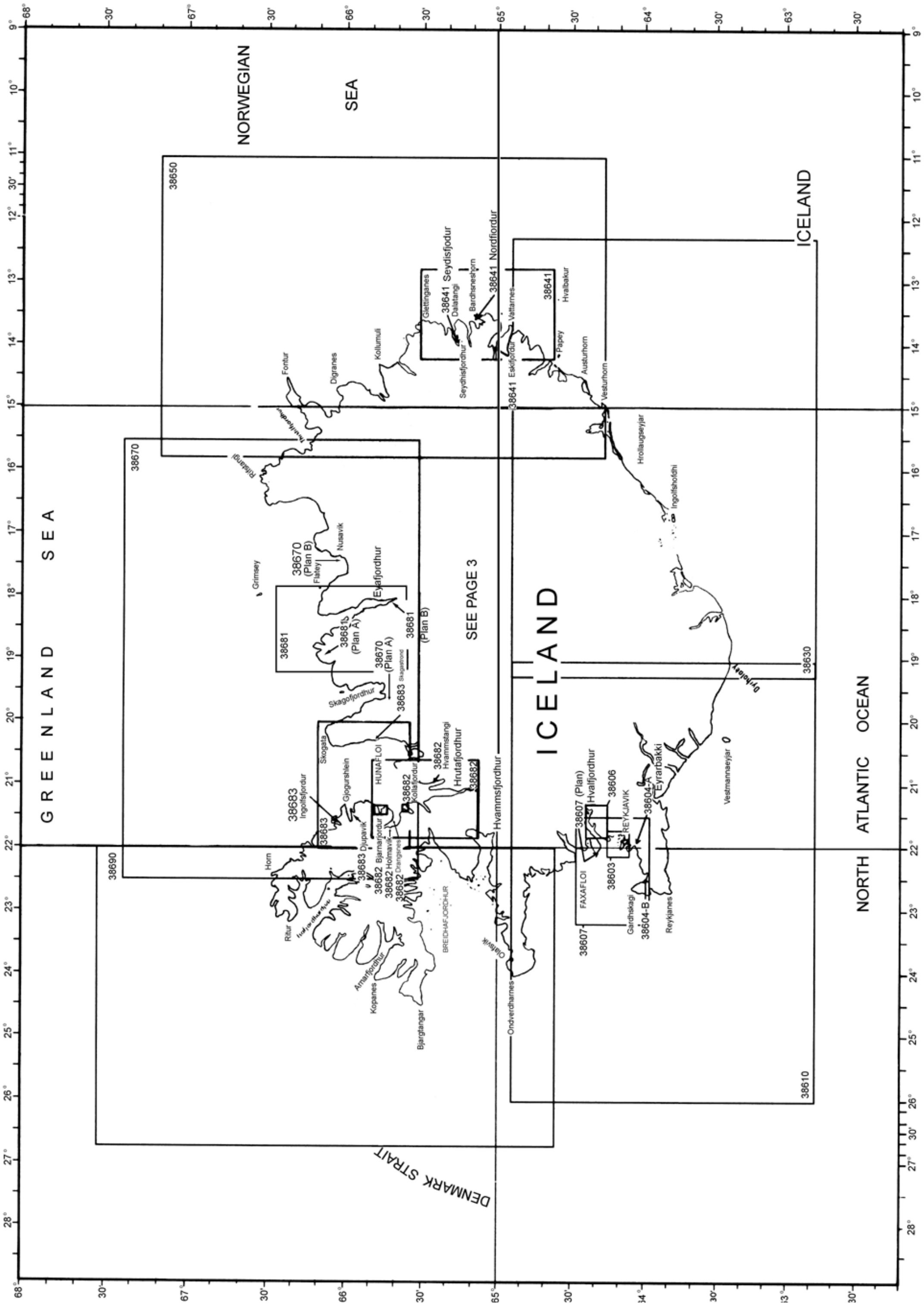


Figure 343c. Chart coverage of Iceland, illustrating the sequence and direction of the U.S. chart numbering system.

Combat charts have 6-digit numbers beginning with an "8." Neither is available to civilian navigators.

344. Catalogs and Stock Numbers

Chart catalogs provide information regarding not only chart coverage, but also a variety of special purpose charts and publications of interest. Keep a corrected chart catalog aboard ship for review by the navigator. The NIMA catalog contains operating area charts and other special products not available for civilian use, but does not contain any classified listings. The NOS catalogs contain all unclassified civilian-use NOS and NIMA charts. Military navigators receive their nautical charts and publications automatically; civilian navigators purchase them from chart sales agents.

The stock number and bar code are generally found in the lower left corner of a NIMA chart, and in the lower right corner of an NOS chart. The first two digits of the stock number refer to the region and subregion. These are followed by three letters, the first of which refers to the portfolio to which the chart belongs; the second two denote the type of chart: CO for coastal, HA for harbor and approach, and OA for military operating area charts. The last five digits are the actual chart number.

USING CHARTS

345. Preliminary Steps

Before using a new edition of a chart, verify its announcement in the *Notice to Mariners* and correct it with all applicable corrections. Read all the chart's notes; there should be no question about the meanings of symbols or the units in which depths are given. Since the latitude and longitude scales differ considerably on various charts, carefully note those on the chart to be used.

Place additional information on the chart as required. Arcs of circles might be drawn around navigational lights to indicate the limit of visibility at the height of eye of an observer on the bridge. Notes regarding other information from the light lists, tide tables, tidal current tables, and sailing directions might prove helpful.

346. Maintaining Charts

A mariner navigating on an uncorrected chart is courting disaster. The chart's print date reflects the latest *Notice to Mariners* used to update the chart; responsibility for maintaining it after this date lies with the user. The weekly *Notice to Mariners* contains information needed for maintaining charts. Radio broadcasts give advance notice of urgent corrections. Local *Notice to Mariners* should be consulted for inshore areas. The navigator must develop a system to keep track of chart corrections and to ensure that the chart he is using is updated with the latest correction. A convenient way of keeping this record is with a *Chart/Publication Correction Record Card* system. Using this system, the navigator does not immediately update every chart in his portfolio when he receives the *Notice to Mariners*. Instead, he constructs a card for every chart in his portfolio and notes the correction on this card. When the time comes to use the chart, he pulls the chart and chart's card, and he makes the indicated corrections on the chart. This system ensures that every chart is properly corrected prior to use.

A *Summary of Corrections*, containing a cumulative listing of previously published *Notice to Mariners* corrections, is published annually in 5 volumes by NIMA. Thus, to fully correct a chart whose edition date is several years old, the navigator needs only the Summary of Corrections for that region and the notices from that Summary forward; he does not need to obtain notices all the way back to the edition date. See Chapter 4, Nautical Publications, for a description of the *Summaries* and *Notice to Mariners*.

When a new edition of a chart is published, it is normally furnished automatically to U.S. Government vessels. It should not be used until it is announced as ready for use in the *Notice to Mariners*. Until that time, corrections in the Notice apply to the old edition and should not be applied to the new one. When it is announced, a new edition of a chart replaces an older one.

Commercial users and others who don't automatically receive new editions should obtain new editions from their sales agent. Occasionally, charts may be received or purchased several weeks in advance of their announcement in the *Notice to Mariners*. This is usually due to extensive rescheming of a chart region and the need to announce groups of charts together to avoid lapses in coverage. The mariner bears the responsibility for ensuring that his charts are the current edition. The fact that a new edition has been compiled and published often indicates that there have been extensive changes that cannot be made by hand corrections.

347. Using and Stowing Charts

Use and stow charts carefully. This is especially true with digital charts contained on electronic media. Keep optical and magnetic media containing chart data out of the sun, inside dust covers, and away from magnetic influences. Placing a disk in an inhospitable environment may destroy the data.

Make permanent corrections to paper charts in ink so that they will not be inadvertently erased. Pencil in all other markings so that they can be easily erased without damaging the chart. Lay out and label tracks on charts of frequently-traveled ports in ink. Draw lines and labels no larger than necessary. Do not obscure sounding data or other information when labeling a chart. When a voyage is completed, carefully erase the charts unless there has been a grounding or collision. In this case, preserve the charts without change because they will play a critical role in the investigation.

When not in use, stow charts flat in their proper portfolio. Minimize their folding and properly index them for easy retrieval.

348. Chart Lighting

Mariners often work in a red light environment because red light is least disturbing to night adapted vision. Such lighting seriously affects the appearance of a chart. Before using a chart in red light, test the effect red light has on its markings. Do not outline or otherwise indicate navigational hazards in red pencil because red markings disappear under red light.

349. Small-Craft Charts

NOS publishes a series of small craft charts sometimes called "strip charts." These charts depict segments of the Atlantic Intracoastal Waterway, the Gulf Intracoastal Waterway, and other inland routes used by yachtsmen, fishermen, and small commercial vessels for coastal travel. They are not "north-up" in presentation, but are aligned with the waterway they depict, whatever its orientation is. Most often they are used as a piloting aid for "eyeball" navigation and placed "course-up" in front of the helmsman, because the routes they show are too confined for taking and plotting fixes.

Although NOS small-craft charts are designed primarily for use aboard yachts, fishing vessels and other small craft, these charts, at scales of 1:80,000 and larger, are in some cases the only charts available depicting inland waters transited by large vessels. In other cases the small-craft charts may provide a better presentation of navigational hazards than the standard nautical chart because of better scale and more detail. Therefore, navigators should use these charts in areas where they provide the best coverage.

CHAPTER 4

NAUTICAL PUBLICATIONS

INTRODUCTION

400. Hardcopy vs. Softcopy Publications

The navigator uses many textual information sources when planning and conducting a voyage. These sources include notices to mariners, summary of corrections, sailing directions, light lists, tide tables, sight reduction tables, and almanacs. Historically, this information has been contained in paper or so-called "hardcopy" publications. But electronic methods of production and distribution of textual material are now commonplace, and will soon replace many of the navigator's familiar books. This volume's CD-ROM version is only one of many. Regardless of how technologically advanced we become, the printed word will always be an important method of communication. Only the means of access will change.

While it is still possible to obtain hard-copy printed publications, increasingly these texts are found on-line or in the form of Compact Disc-Read Only Memory (CD-ROM's). CD-ROM's are much less expensive than printed publications to reproduce and distribute, and on-line publications have no reproduction costs at all for the producer, and only minor costs to the user, if he chooses to print them at all. Also, a few CD-ROM's can hold entire libraries of information, making both distribution and on-board storage much easier.

The advantages of electronic publications go beyond their cost savings. They can be updated easier and more often, making it possible for mariners to have frequent or even continuous access to a maintained publications database instead of receiving new editions at infrequent intervals and entering hand corrections periodically. Generally, digital publications also provide links and search engines to quickly access related information.

Navigational publications are available from many sources. Military customers automatically receive or requisition most publications. The civilian navigator obtains his publications from a publisher's agent. Larger agents representing many publishers can completely supply a ship's chart and publication library. On-line publications produced by the U.S. government are available on the Web.

This chapter will refer generally to printed publications. If the navigator has access to this data electronically, his methods of access and use will differ somewhat, but the discussion herein applies equally to both electronic and hard-copy documents.

NAUTICAL TEXTS

401. *Sailing Directions*

National Imagery and Mapping Agency *Sailing Directions* consist of 37 ***Enroutes*** and 5 ***Planning Guides***. *Planning Guides* describe general features of ocean basins; *Enroutes* describe features of coastlines, ports, and harbors.

Sailing Directions are updated when new data requires extensive revision of an existing volume. These data are obtained from several sources, including pilots and foreign Sailing Directions.

One book comprises the *Planning Guide* and *Enroute* for Antarctica. This consolidation allows for a more effective presentation of material on this unique area.

The *Planning Guides* are relatively permanent; by contrast, *Sailing Directions (Enroute)* are frequently updated. Between updates, both are corrected by the *Notice to Mariners*.

402. *Sailing Directions (Planning Guide)*

Planning Guides assist the navigator in planning an extensive oceanic voyage. Each of the Guides provides useful information about all the countries adjacent to a particular ocean basin. The limits of the *Sailing Directions* in relation to the major ocean basins are shown in Figure 402.

Planning Guides are structured in the alphabetical order of countries contained within the region. Information pertaining to each country includes Buoyage Systems, Currency, Government, Industries, Holidays, Languages, Regulations, Firing Danger Areas, Mined Areas, Pilotage, Search and Rescue, Reporting Systems, Submarine Operating Areas, Time Zone, and the location of the U.S. Embassy.

403. *Sailing Directions (Enroute)*

Each volume of the *Sailing Directions (Enroute)*

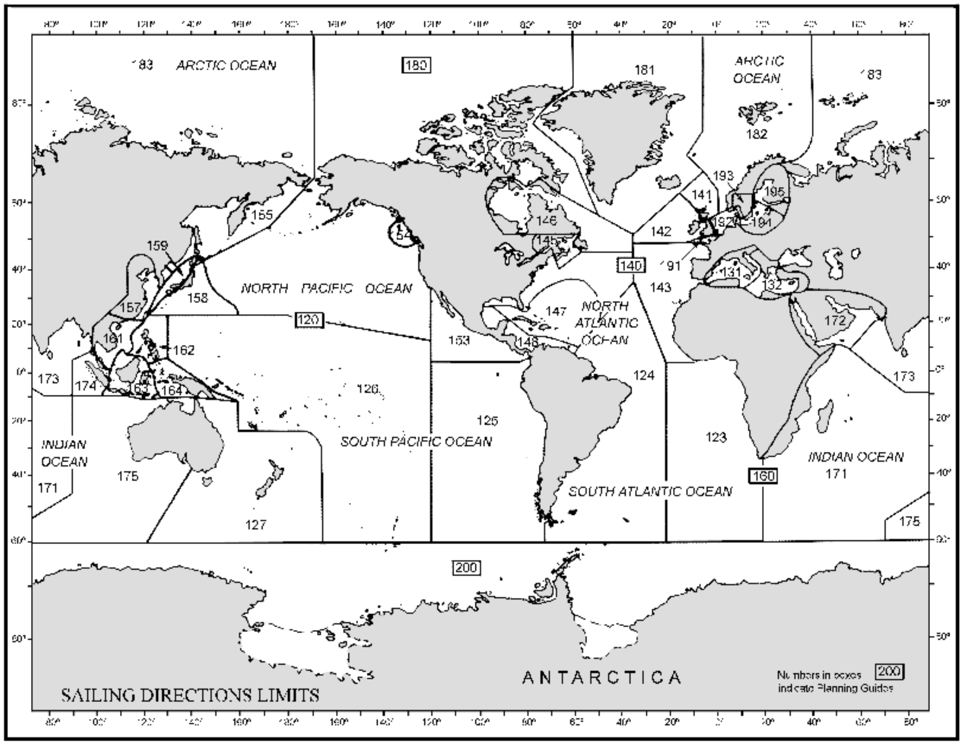


Figure 402. Sailing Directions limits in relation to the major ocean basins.

contains numbered sections along a coast or through a strait. Figure 403a illustrates this division. Each sector is sub-divided into paragraphs and discussed in turn. A preface with information about authorities, references, and conventions used in each book precedes the sector discussions. Each book also provides conversions between feet, fathoms, and meters, and an Information and Suggestion Sheet.

The Chart Information Graphic, the first item in each sector, is a graphic key for charts pertaining to that area. See Figure 403b. The graduation of the border scale of the chartlet enables navigators to identify the largest scale chart for a location and to find a feature listed in the Index-Gazetteer. These graphics are not maintained by *Notice to Mariners*; one should refer to the chart catalog for updated chart listings. Other graphics may contain special information on anchorages, significant coastal features, and navigation dangers.

A foreign terms glossary and a comprehensive Index-Gazetteer follow the sector discussions. The Index-Gazetteer is an alphabetical listing of described and charted features. The Index lists each feature by geographic coordinates and sector paragraph number.

U.S. military vessels have access to special files of data reported via official messages known as Port Visit After Action Reports. These reports, written in text form according to a standardized reporting format, give complete details of recent visits by U.S. military vessels to all foreign ports visited. Virtually every detail regarding navigation, services, supplies, official and unofficial contacts, and other matters is discussed in detail, making these reports an extremely useful adjunct to the *Sailing Directions*. These files are available to ".mil" users only, and may be accessed on the Web at: http://cnsl.spear.navy.mil, under the "Force Navigator" link. They are also available via DoD's classified Web.

404. *Coast Pilots*

The National Ocean Service publishes nine *United States Coast Pilots* to supplement nautical charts of U.S. waters. Information comes from field inspections, survey vessels, and various harbor authorities. Maritime officials and pilotage associations provide additional information. *Coast Pilots* provide more detailed information than *Sailing Directions* because *Sailing Directions* are intended exclusively for the oceangoing mariner. The *Notice to Mariners* updates *Coast Pilots*.

Each volume contains comprehensive sections on local operational considerations and navigation regulations. Following chapters contain detailed discussions of coastal navigation. An appendix provides information on obtaining additional weather information, communications services, and other data. An index and additional tables complete the volume.

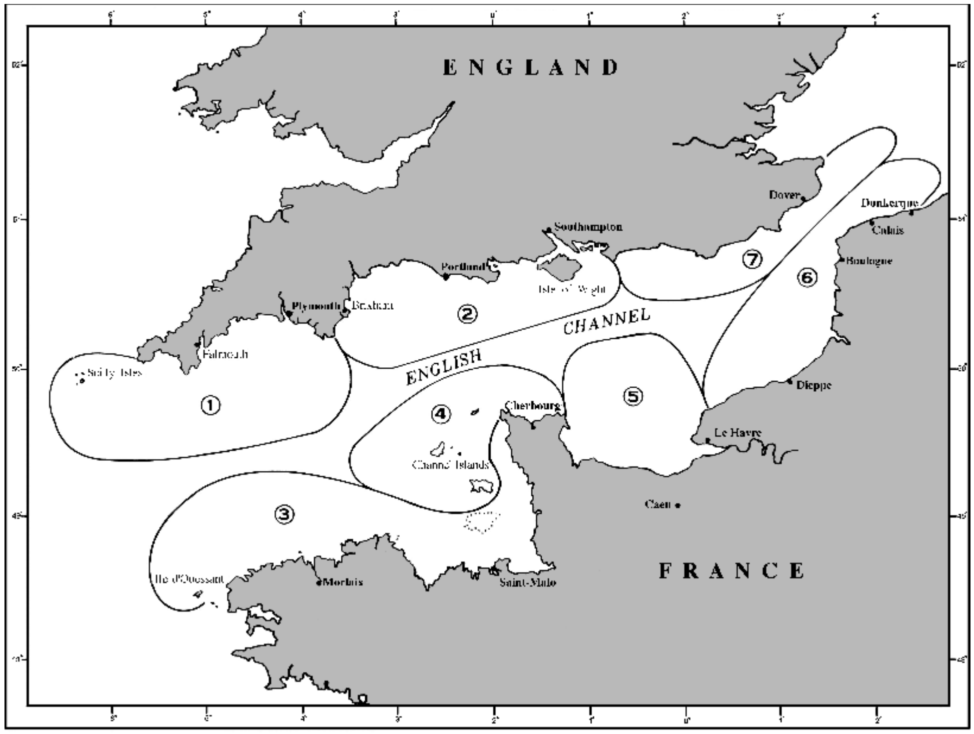
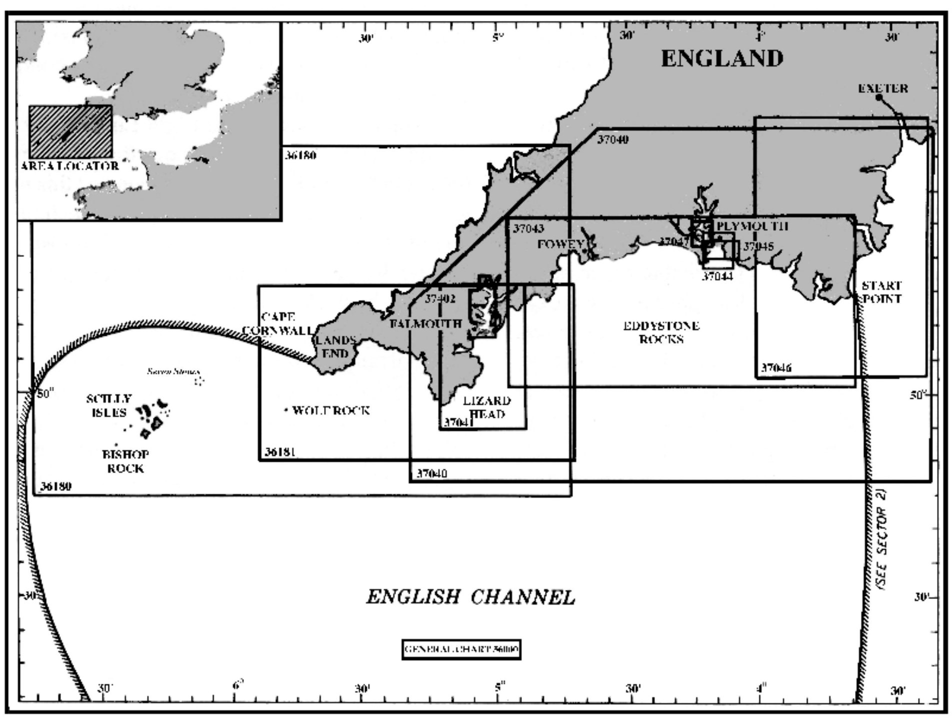


Figure 403a. Sector Limits graphic.



Additional chart coverage may be found in CATP2 Catalog of Nautical Charts.

Figure 403b. Chart Information graphic.

405. Other Nautical Texts

The government publishes several other nautical texts. NIMA, for example, publishes *Pub. 1310, Radar Navigation and Maneuvering Board Manual* and *Pub. 9, American Practical Navigator*.

The U.S. Coast Guard publishes *Navigation Rules* for international and inland waters. This publication, officially known as Commandant Instruction M16672.2d, contains the Inland Navigation Rules enacted in December 1980 and effective on all inland waters of the United States including the Great Lakes, as well as the *International Regulations for the Prevention of Collisions at Sea*, enacted in 1972 (1972 COLREGS). Mariners should ensure that they have the updated issue. The Coast Guard also publishes comprehensive user's manuals for the Loran and GPS navigation systems; *Navigation and Vessel Inspection Circulars*; and the *Chemical Data Guide for Bulk Shipment by Water*.

The Government Printing Office provides several publications on navigation, safety at sea, communications, weather, and related topics. Additionally, it publishes provisions of the Code of Federal Regulations (CFR) relating to maritime matters. A number of private publishers also provide maritime publications.

The International Maritime Organization, International Hydrographic Organization, and other governing international organizations provide information on international navigation regulations. Chapter 1 gives these organizations' addresses. Regulations for various Vessel Traffic Services (VTS), canals, lock systems, and other regulated waterways are published by the authorities which operate them. Nautical chart and publication sales agents are a good source of information about publications required for any voyage. Increasingly, many regulations, whether instituted by international or national governments, can be found online. This includes regulations for Vessel Traffic Services, Traffic Separation Schemes, special regulations for passage through major canal and lock systems, port and harbor regulations, and other information. A Web search can often find the textual information the navigator needs.

USING THE LIGHT LISTS

406. Light Lists

The United States publishes two different light lists. The U.S. Coast Guard publishes the *Light List* for lights in U.S. territorial waters; NIMA publishes the *List of Lights* for lights in foreign waters.

Light lists furnish detailed information about navigation lights and other navigation aids, supplementing the charts, *Coast Pilots*, and *Sailing Directions*. Consult the chart for the location and light characteristics of all navigation aids; consult the light lists to determine their detailed description.

The *Notice to Mariners* corrects both lists. Corrections which have accumulated since the print date are included in the *Notice to Mariners* as a *Summary of Corrections*. All of these summary corrections, and any corrections published subsequently, should be noted in the "Record of Corrections."

A navigator needs to know both the identity of a light and when he can expect to see it; he often plans the ship's track to pass within a light's range. If lights are not sighted when predicted, the vessel may be significantly off course and standing into danger.

A circle with a radius equal to the visible range of the light usually defines the area in which a light can be seen. On some bearings, however, obstructions may reduce the range. In this case, the obstructed arc might differ with height of eye and distance. Also, lights of different colors may be seen at different distances. Consider these facts both when identifying a light and predicting the range at which it can be seen.

Atmospheric conditions have a major effect on a light's range. Fog, haze, dust, smoke, or precipitation can obscure a light. Additionally, a light can be extinguished. Always report an extinguished light so maritime authorities can issue a warning and make repairs.

On a dark, clear night, the visual range is limited by either: (1) luminous intensity, or (2) curvature of the Earth. Regardless of the height of eye, one cannot see a weak light beyond a certain luminous range. Assuming light travels linearly, an observer located below the light's visible horizon cannot see it. The Distance to the Horizon table gives the distance to the horizon for various heights of eye. The light lists contain a condensed version of this table. Abnormal refraction patterns might change this range; therefore, one cannot exactly predict the range at which a light will be seen.

407. Finding Range and Bearing of a Light at Sighting

A light's **luminous range** is the maximum range at which an observer can see a light under existing visibility conditions. This luminous range ignores the elevation of the light, the observer's height of eye, the curvature of the Earth, and interference from background lighting. It is determined from the known **nominal range** and the existing visibility conditions. The nominal range is the maximum distance at which a light can be seen in weather conditions where visibility is 10 nautical miles.

The U.S. Coast Guard *Light List* usually lists a light's nominal range. Use the Luminous Range Diagram shown in the *Light List* and Figure 407a to convert this nominal range to luminous range. Remember that the luminous ranges obtained are approximate because of atmospheric or background lighting conditions. To use the Luminous Range

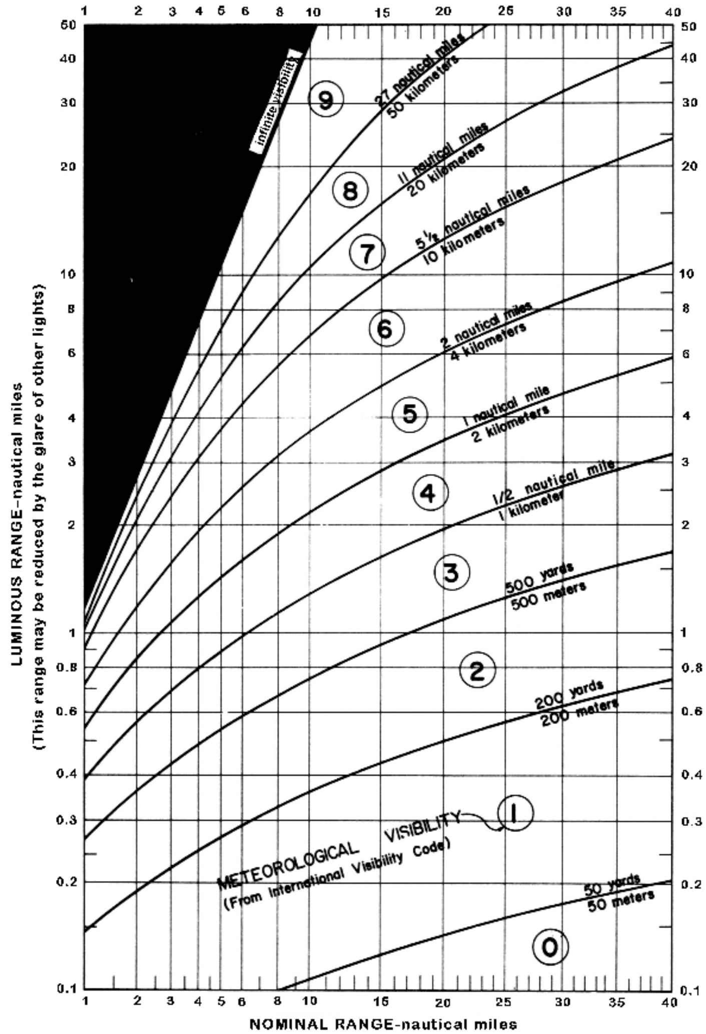


Figure 407a. Luminous Range Diagram.

Diagram, first estimate the meteorological visibility by the Meteorological Optical Range Table, Figure 407b. Next, enter the Luminous Range Diagram with the nominal range on the horizontal nominal range scale. Follow a vertical line until it intersects the curve or reaches the region on the diagram representing the meteorological visibility. Finally, follow a horizontal line from this point or region until it intersects the vertical luminous range scale.

Example 1: *The nominal range of a light as extracted from the Light List is 15 nautical miles.*

Required: *The luminous range when the meteorological visibility is (1) 11 nautical miles and (2) 1 nautical mile.*

Solution: *To find the luminous range when the meteorological visibility is 11 nautical miles, enter the Luminous Range Diagram with nominal range 15 nautical miles on the horizontal nominal range scale; follow a vertical line upward until it intersects the curve on the diagram representing a meteorological visibility of 11 nautical miles; from this point follow a horizontal line to the right until it intersects the vertical luminous range scale at 16 nautical miles. A similar procedure is followed to find the luminous range when the meteorological visibility is 1 nautical mile.*

Answers: *(1) 16 nautical miles; (2) 3 nautical miles.*

A light's **geographic range** depends upon the height of both the light and the observer. The sum of the observer's dis-

Code No.	Weather	Yards
0	Dense fog	Less than 50
1	Thick fog	50-200
2	Moderate fog	200-500
3	Light fog	500-1000
		Nautical Miles
4	Thin fog	1/2-1
5	Haze	1-2
6	Light Haze	2-5 1/2
7	Clear	5 1/2-11
8	Very Clear	11.0-27.0.
9	Exceptionally Clear	Over 27.0

From the International Visibility Code.

Figure 407b. Meteorological Optical Range Table.

tance to the visible horizon (based on his height of eye) plus the light's distance to the horizon (based on its height) is its geographic range. See Figure 407c. This illustration uses a light 150 feet above the water. Table 12, Distance of the Horizon, yields a value of 14.3 nautical miles for a height of 150 feet. Within this range, the light, if powerful enough and atmospheric conditions permit, is visible regardless of the height of eye of the observer. Beyond 14.3 nautical miles, the geographic range depends upon the observer's height of eye. Thus, by the Distance of the Horizon table mentioned above, an observer with height of eye of 5 feet can see the light on his horizon if he is 2.6 miles beyond the horizon of the light. The geographic range of the light is therefore 16.9 miles. For a height of 30 feet the distance is 14.3 + 6.4 = 20.7 miles. If the height of eye is 70 feet, the geographic range is 14.3 + 9.8 = 24.1 miles. A height of eye of 15 feet is often assumed when tabulating lights' geographic ranges.

To predict the bearing and range at which a vessel will initially sight a light first determine the light's geographic range. Compare the geographic range with the light's luminous range. The lesser of the two ranges is the range at which the light will first be sighted. Plot a visibility arc centered on the light and with a radius equal to the lesser of the geographic or luminous ranges. Extend the vessel's track until it intersects the visibility arc. The bearing from the intersection point to the light is the light's predicted bearing at first sighting.

If the extended track crosses the visibility arc at a small angle, a small lateral track error may result in large bearing and time prediction errors. This is particularly apparent if the vessel is farther from the light than predicted; the vessel may pass the light without sighting it. However, not sighting a light when predicted does not always indicate the vessel is farther from the light than expected. It could also mean that atmospheric conditions are affecting visibility.

Example 2: *The nominal range of a navigational light 120 feet above the chart datum is 20 nautical miles. The meteorological visibility is 27 nautical miles.*

Required: *The distance at which an observer at a height of eye of 50 feet can expect to see the light.*

Solution: *The maximum range at which the light may be seen is the lesser of the luminous or geographic ranges. At 120 feet the distance to the horizon, by table or formula, is 12.8 miles. Add 8.3 miles, the distance to the horizon for a height of eye of 50 feet to determine the geographic range. The geographic range, 21.1 miles, is less than the luminous range, 40 miles.*

Answer: *21 nautical miles. Because of various uncertainties, the range is rounded off to the nearest whole mile.*

When first sighting a light, an observer can determine if it is on the horizon by immediately reducing his height of eye. If the light disappears and then reappears when the observer returns to his original height, the light is on the horizon. This process is called **bobbing a light**.

If a vessel has considerable vertical motion due to rough seas, a light sighted on the horizon may alternately appear and disappear. Wave tops may also obstruct the light periodically. This may cause the characteristic to appear different than expected. The light's true characteristics can be ascertained either by closing the range to the light or by increasing the observer's height of eye.

If a light's range given in a foreign publication approximates the light's geographic range for a 15-foot observer's height of eye, one can assume that the printed range is the light's geographic range. Also assume that publication has listed the lesser of the geographic and nominal ranges. Therefore, if the light's listed range approximates the geographic range for an observer with a height of eye of 15 feet, then assume that the light's limiting range is the geographic range. Then, calculate the light's true geographic range using the actual observer's height of eye, not the assumed height of eye of 15 feet. This calculated true geographic range is the range at which the light will first be sighted.

Example 3: *The range of a light as printed on a foreign chart is 17 miles. The light is 120 feet above chart datum. The meteorological visibility is 10 nautical miles.*

Required: *The distance at which an observer at a height of eye of 50 feet can expect to see the light.*

Solution: *Calculate the geographic range of the light assuming a 15 foot observer's height of eye. At 120 feet the distance to the horizon is 12.8 miles. Add 4.5 miles (the distance to the horizon at a height of 15 feet) to 12.8 miles; this range is 17.3 miles. This approximates the range listed on the chart. Then assuming that the charted range is the*

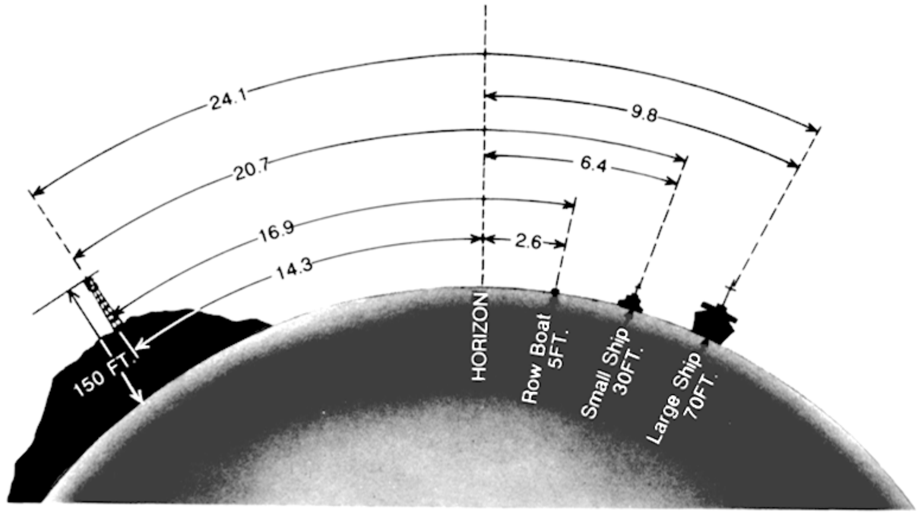


Figure 407c. Geographic Range of a light.

geographic range for a 15-foot observer height of eye and that the nominal range is the greater than this charted range, the predicted range is found by calculating the true geographic range with a 50 foot height of eye for the observer.

Answer: *The predicted range = 12.8 mi. + 8.3 mi. = 21.1 mi. The distance in excess of the charted range depends on the luminous intensity of the light and the meteorological visibility.*

408. USCG *Light Lists*

The U.S. Coast Guard *Light List* (7 volumes) gives information on lighted navigation aids, unlighted buoys, radiobeacons, radio direction finder calibration stations, daybeacons, racons, and Loran stations.

Each volume of the *Light List* contains aids to navigation in geographic order from north to south along the Atlantic coast, from east to west along the Gulf coast, and from south to north along the Pacific coast. It lists seacoast aids first, followed by entrance and harbor aids listed from seaward. Intracoastal Waterway aids are listed last in geographic order in the direction from New Jersey to Florida to the Texas/Mexico border.

The listings are preceded by a description of the aids to navigation system in the United States, luminous range diagram, geographic range tables, and other information.

409. NIMA *List of Lights, Radio Aids, and Fog Signals*

The National Imagery and Mapping Agency publishes the *List of Lights, Radio Aids, and Fog Signals* (usually referred to as the *List of Lights*, not to be confused with the Coast Guard's *Light List*). In addition to information on lighted aids to navigation and sound signals in foreign waters, the NIMA *List of Lights* provides information on storm signals, signal stations, racons, radiobeacons, radio direction finder calibration stations located at or near lights, and DGPS stations. For more details on radio navigational aids, consult *Pub. 117, Radio Navigational Aids*.

The NIMA *List of Lights* generally does not include information on buoys, although in certain instances, a large offshore buoy with a radio navigational aid may be listed. It does include certain aeronautical lights situated near the coast. However, these lights are not designed for marine navigation and are subject to unreported changes.

Foreign notices to mariners are the main correctional information source for the NIMA *Lists of Lights*; other sources, such as ship reports, are also used. Many aids to navigation in less developed countries may not be well maintained. They are subject to damage by storms and vandalism, and repairs may be delayed for long periods.

MISCELLANEOUS NAUTICAL PUBLICATIONS

410. NIMA *Radio Navigational Aids (Pub. 117)*

This publication is a selected list of worldwide radio stations which perform services to the mariner. Topics covered include radio direction finder and radar stations, radio time signals, radio navigation warnings, distress and safety communications, medical advice via radio, long-range navigation aids, the AMVER system, and interim procedures for U.S. vessels in the event of an outbreak of hostilities. *Pub. 117* is corrected via the

Notice to Mariners and is updated periodically with a new edition.

Though *Pub. 117* is essentially a list of radio stations providing vital maritime communication and navigation services, it also contains information which explains the capabilities and limitations of the various systems.

411. *Chart No. 1*

Chart No. 1 is not actually a chart but a book containing a key to chart symbols. Most countries which produce charts also produce such a list. The U.S. *Chart No. 1* contains a listing of chart symbols in four categories:

- Chart symbols used by the National Ocean Service
- Chart symbols used by NIMA
- Chart symbols recommended by the International Hydrographic Organization
- Chart symbols used on foreign charts reproduced by NIMA

Subjects covered include general features of charts, topography, hydrography, and aids to navigation. There is also a complete index of abbreviations and an explanation of the IALA buoyage system.

412. NIMA *World Port Index (Pub. 150)*

The *World Port Index* contains a tabular listing of thousands of ports throughout the world, describing their locations, characteristics, facilities, and services available. Information is arranged geographically; the index is arranged alphabetically.

Coded information is presented in columns and rows. This information supplements information in the *Sailing Directions*. The applicable volume of *Sailing Directions* and the number of the harbor chart are given in the *World Port Index*. The *Notice to Mariners* corrects this book.

413. NIMA *Distances Between Ports (Pub. 151)*

This publication lists the distances between major ports. Reciprocal distances between two ports may differ due to different routes chosen because of currents and climatic conditions. To reduce the number of listings needed, junction points along major routes are used to consolidate routes converging from different directions.

This book can be most effectively used for voyage planning in conjunction with the proper volume(s) of the *Sailing Directions (Planning Guide)*. It is corrected via the *Notice to Mariners*.

414. NIMA *International Code of Signals (Pub. 102)*

This book lists the signals to be employed by vessels at sea to communicate a variety of information relating to safety, distress, medical, and operational information. This publication became effective in 1969.

According to this code, each signal has a unique and complete meaning. The signals can be transmitted via Morse code light and sound, flag, radio telegraph and telephone, and semaphore. Since these methods of signaling are internationally recognized, differences in language between sender and receiver are immaterial; the message will be understood when decoded in the language of the receiver, regardless of the language of the sender. The *Notice to Mariners* corrects *Pub. 102*.

415. Almanacs

For celestial sight reduction, the navigator needs an **almanac** for ephemeris data. The *Nautical Almanac*, produced jointly by H.M. Nautical Almanac Office and the U.S. Naval Observatory, is the most common almanac used for celestial navigation. It also contains information on sunrise, sunset, moonrise, and moonset, as well as compact sight reduction tables. The *Nautical Almanac* is published annually.

The *Air Almanac* contains slightly less accurate ephemeris data for air navigation. It can be used for marine navigation if slightly reduced accuracy is acceptable.

Chapter 19 provides more detailed information on using the *Nautical Almanac*.

416. *Sight Reduction Tables*

Without a calculator or computer programmed for sight reduction, the navigator needs ***sight reduction tables*** to solve the celestial triangle. Two different sets of tables are commonly used at sea.

NIMA *Pub. 229, Sight Reduction Tables for Marine Navigation*, consists of six volumes of tables designed for use with the *Nautical Almanac* for solution of the celestial triangle by the **Marcq Saint Hilaire** or **intercept** method. The tabular data are the solutions of the navigational triangle of which two sides and the included angle are known and it is necessary to find the third side and adjacent angle.

Each volume of *Pub. 229* includes two 8 degree zones, comprising 15 degree bands from 0 to 90 degrees, with a 1° degree overlap between volumes. *Pub. 229* is a joint publication produced by the National Imagery and Mapping Agency, the U.S. Naval Observatory, and the Royal Greenwich Observatory.

Sight Reduction Tables for Air Navigation, Pub. 249, is also a joint production of the three organizations above. It is issued in three volumes. Volume 1 contains the values of the altitude and true azimuth of seven selected stars chosen to provide, for any given position and time, the best celestial observations. A new edition is issued every 5 years for the upcoming astronomical epoch. Volumes 2 (0° to 40°) and 3 (39° to 89°) provide for sights of the Sun, Moon, and planets.

417. Catalogs

A chart catalog is a valuable reference to the navigator for voyage planning, inventory control, and ordering. The catalog is used by military and civilian customers.

The navigator will see the NIMA nautical chart catalog as part of a larger suite of catalogs including aeronautical (Part 1), hydrographic (Part 2), and topographic (Part 3) products. Each Part consists of one or more volumes. Unclassified NIMA nautical charts are listed in Part 2, Volume 1.

This catalog contains comprehensive ordering instructions and information about the products listed. Also listed are addresses of all Map Support Offices, information on crisis support, and other special situations. The catalog is organized by geographic region corresponding to the chart regions 1 through 9. A special section of miscellaneous charts and publications is included. This section also lists products produced by NOS, the U.S. Army Corps of Engineers, U.S. Coast Guard, U.S. Naval Oceanographic Office, and some foreign publications from the United Kingdom and Canada.

The civilian navigator should also refer to catalogs produced by the National Ocean Service. For U.S. waters, NOS charts are listed in a series of large sheet "charts" showing a major region of the U.S. with individual chart graphics depicted. These catalogs also list charts showing titles and scales. They also list sales agents from whom the charts may be purchased.

NIMA products for the civilian navigator are listed by NOS in a series of regionalized catalogs similar to Part 2 Volume 1. These catalogs are also available through authorized NOS chart agents.

MARITIME SAFETY INFORMATION

418. *Notice to Mariners*

The ***Notice to Mariners*** is published weekly by the National Imagery and Mapping Agency (NIMA), prepared jointly with the National Ocean Service (NOS) and the U.S. Coast Guard. It advises mariners of important matters affecting navigational safety, including new hydrographic information, changes in channels and aids to navigation, and other important data. The information in the *Notice to Mariners* is formatted to simplify the correction of paper charts, sailing directions, light lists, and other publications produced by NIMA, NOS, and the U.S. Coast Guard.

It is the responsibility of users to decide which of their charts and publications require correction. Suitable records of *Notice to Mariners* should be maintained to facilitate the updating of charts and publications prior to use.

Information for the *Notice to Mariners* is contributed by: NIMA (Department of Defense) for waters outside the territorial limits of the United States; National Ocean Service (National Oceanic and Atmospheric Administration, Department of Commerce), which is charged with surveying and charting the coasts and harbors of the United States and its territories; the U.S. Coast Guard (Department of Transportation) which is responsible for, among other things, the safety of life at sea and the establishment and operation of aids to navigation; and the Army Corps of Engineers (Department of Defense), which is charged with the improvement of rivers and harbors of the United States. In addition, important contributions are made by foreign hydrographic offices and cooperating observers of all nationalities.

Over 60 countries which produce nautical charts also produce a notice to mariners. About one third of these are weekly, another third are bi-monthly or monthly, and the rest irregularly issued according to need. Much of the data in the U.S. *Notice to Mariners* is obtained from these foreign notices.

U.S. charts must be corrected only with a U.S. *Notice to Mariners*. Similarly, correct foreign charts using the foreign notice because chart datums often vary according to region and geographic positions are not the same for different datums.

The *Notice to Mariners* consists of a page of **Hydrograms** listing important items in the notice, a chart correction section organized by ascending chart number, a publications correction section, and a summary of broadcast navigation warnings and miscellaneous information.

Mariners are requested to cooperate in the correction of charts and publications by reporting all discrepancies between published information and conditions actually observed and by recommending appropriate improvements. A convenient reporting form is provided in the back of each *Notice to Mariners*.

Notice to Mariners No. 1 of each year contains important information on a variety of subjects which supplements information not usually found on charts and in navigational publications. This information is published as ***Special Notice to Mariners Paragraphs***. Additional items considered of interest to the mariner are also included in this *Notice*.

419. *Summary of Corrections*

A close companion to the *Notice to Mariners* is the

Summary of Corrections. The *Summary* is published in five volumes. Each volume covers a major portion of the Earth including several chart regions and their subregions. Volume 5 also includes special charts and publications corrected by the *Notice to Mariners*. Since the *Summaries* contain cumulative corrections, any chart, regardless of its print date, can be corrected with the proper volume of the *Summary* and all subsequent *Notice to Mariners*.

420. The Maritime Safety Information Website

The NIMA **Maritime Safety Information Website** provides worldwide remote query access to extensive menus of maritime safety information 24 hours a day. The Maritime Safety Information Website can be accessed via the NIMA Homepage (www.nima.mil) under the Safety of Navigation icon or directly at http://pollux.nss.nima.mil.

Databases made available for access, query and download include Chart Corrections, Publication Corrections, NIMA Hydrographic Catalog Corrections, Chart and Publication Reference Data (current edition number, dates, title, scale), NIMA *List of Lights*, U.S. Coast Guard *Light Lists*, World Wide Navigational Warning Service (WWNWS) Broadcast Warnings, Maritime Administration (MARAD) Advisories, Department of State Special Warnings, Mobile Offshore Drilling Units (MODUs), Anti-Shipping Activity Messages (ASAMs), *World Port Index*, and *Radio Navigational Aids*. Publications that are also made available as Portable Document Format (PDF) files include the U.S. *Notice to Mariners*, U.S. *Chart No. 1*, *The American Practical Navigator*, *International Code of Signals*, *Radio Navigational Aids*, *World Port Index*, *Distances Between Ports*, *Sight Reduction Tables for Marine Navigation*, *Sight Reduction Tables for Air Navigation*, and the *Radar Navigation and Maneuvering Board Manual*.

Navigators have online access to, and can download, all the information contained in the printed *Notice to Mariners* including chartlets. Information on this website is updated daily or weekly according to the *Notice to Mariners* production schedule. Broadcast Warnings, MARAD Advisories, ASAMs and MODUs are updated on a daily basis; the remaining data is updated on a weekly basis.

Certain files, for example U.S. Coast Guard *Light List* data, are entered directly into the database without editing and the accuracy of this information cannot be verified by NIMA staff. Also, drill rig locations are furnished by the companies which operate them. They are not required to provide these positions, and they cannot be verified. However, within these limitations, the Website can provide information 2 weeks sooner than the printed *Notice to Mariners*, because the paper *Notice* must be printed and mailed after the digital version is completed and posted on the Web.

Users can provide suggestions, changes, corrections or comments on any of the Maritime Safety Information Division products and services by submitting an online version of the Marine Information Report and Suggestion Sheet.

Access to the Maritime Safety Information Website is free, but the user must pay the applicable charges for internet service. Any questions concerning the Maritime Safety Information Website should be directed to the Maritime Safety Information Division, Attn.: NSS STAFF, Mail Stop D-44, NIMA, 4600 Sangamore Rd., Bethesda, MD, 20816-5003; telephone (1) 301-227-3296; fax (1) 301-227-4211; e-mail webmaster_nss@nima.mil.

421. *Local Notice to Mariners*

The ***Local Notice to Mariners*** is issued by each U.S. Coast Guard District to disseminate important information affecting navigational safety within that District. This Notice reports changes and deficiencies in aids to navigation maintained by the Coast Guard. Other marine information such as new charts, channel depths, naval operations, and regattas is included. Since temporary information of short duration is not included in the NIMA *Notice to Mariners*, the *Local Notice to Mariners* may be the only source for it. Since correcting information for U.S. charts in the NIMA *Notice* is obtained from the Coast Guard local notices, there is a lag of 1 or 2 weeks for NIMA *Notice* to publish a correction from this source.

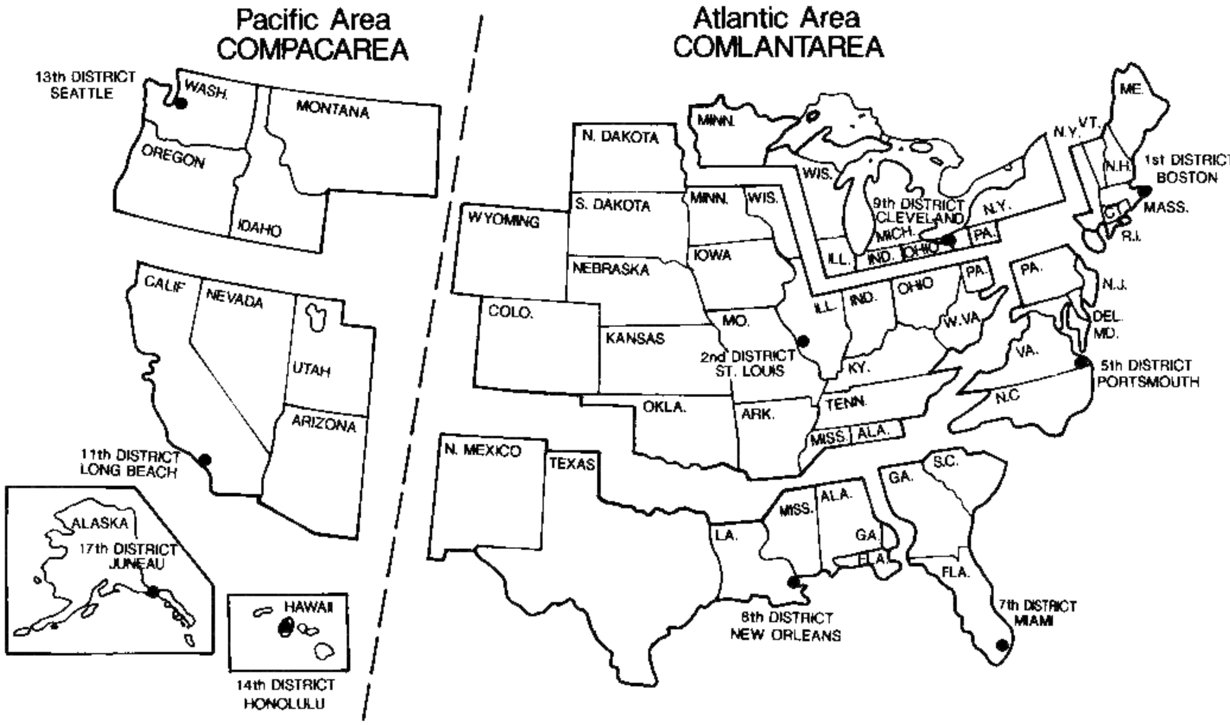
The *Local Notice to Mariners* may be obtained free of charge by contacting the appropriate Coast Guard District Commander. Vessels operating in ports and waterways in several districts must obtain the *Local Notice to Mariners* from each district. See Figure 421 for a complete list of U.S. Coast Guard Districts.

422. Electronic Notice to Mariners

One major impediment to full implementation of electronic chart systems has been the issue of how to keep them up to date. The IMO, after reviewing the range standards which might be employed in the provision of updates to ECDIS charts, decided that the correction system must be "hands off" from the mariner's point of view. That is, the correction system could not rely on the ability of the mariner to enter individual correction data himself, as he would do on a paper chart. The process must be automated to maintain the integrity of the data and prevent errors in data entry by navigators.

National hydrographic offices which publish electronic charts must also publish corrections for them. The manner of doing so varies among the different types of systems. The corrections are applied to the data as the chart to be displayed is created, leaving the database unchanged.

Another possibility exists, and that is to simply reload the entire chart data file with updated information. This is not as crazy as it sounds when one considers the amount of data that can be stored on a single CD-ROM and the ease



COMMANDER, FIRST COAST GUARD DISTRICT
408 ATLANTIC AVENUE
BOSTON, MA 02110-3350
PHONE: DAY 617-223-8338, NIGHT 617-223-8558

COMMANDER, SECOND COAST GUARD DISTRICT
1222 SPRUCE STREET
ST. LOUIS, MO 63103-2832
PHONE: DAY 314-539-3714, NIGHT 314-539-3709

COMMANDER, FIFTH COAST GUARD DISTRICT
FEDERAL BUILDING
431 CRAWFORD STREET
PORTSMOUTH, VA 23704-5004
PHONE: DAY 804-398-6486, NIGHT 804-398-6231

COMMANDER, SEVENTH COAST GUARD DISTRICT
BRICKELL PLAZA FEDERAL BUILDING
909 SE 1ST AVENUE, RM: 406
MIAMI, FL 33131-3050
PHONE: DAY 305-536-5621, NIGHT 305-536-5611

COMMANDER GREATER ANTILLES SECTION
U.S. COAST GUARD
P.O. BOX S-2029
SAN JUAN, PR 00903-2029
PHONE: 809-729-6870

COMMANDER, EIGHTH COAST GUARD DISTRICT
HALE BOGGS FEDERAL BUILDING
501 MAGAZINE STREET
NEW ORLEANS, LA 70130-3396
PHONE: DAY 504-589-6234, NIGHT 504-589-6225

COMMANDER, NINTH COAST GUARD DISTRICT
1240 EAST 9TH STREET
CLEVELAND, OH 44199-2060
PHONE: DAY 216-522-3991, NIGHT 216-522-3984

COMMANDER, ELEVENTH COAST GUARD DISTRICT
FEDERAL BUILDING
501 W. OCEAN BLVD.
LONG BEACH, CA 90822-5399
PHONE: DAY 310-980-4300, NIGHT 310-980-4400

COMMANDER, THIRTEENTH COAST GUARD DISTRICT
FEDERAL BUILDING
915 SECOND AVENUE
SEATTLE, WA 98174-1067
PHONE: DAY 206-220-7280, NIGHT 206-220-7004

COMMANDER, FOURTEENTH COAST GUARD DISTRICT
PRINCE KALANIANAOLE FEDERAL BLDG.
9TH FLOOR, ROOM 9139
300 ALA MOANA BLVD.
HONOLULU, HI 96850-4982
PHONE: DAY 808-541-2317, NIGHT 808-541-2500

COMMANDER, SEVENTEENTH COAST GUARD DISTRICT
P.O. BOX 25517
JUNEAU, AK 99802-5517
PHONE: DAY 907-463-2245, NIGHT 907-463-2000

Figure 421. U.S. Coast Guard Districts.

with which it can be reproduced. At present, these files are too large to be broadcast effectively, but with the proper bandwidth the concept of transferring entire chart portfolios worldwide via satellite or fiber-optic cable is entirely feasible.

Corrections to the DNC published by NIMA are being made by Vector Product Format Database Update (VDU). These are patch corrections and are available via the Web and by classified data links used by the Department of Defense.

Corrections to raster charts issued by NOAA are also available via the internet. To produce the patch, each chart is corrected and then compared, pixel by pixel, with the previous, uncorrected version. Any differences between the two must have been the result of a correction, so those files are saved and posted to a site for access by subscription users. The user accesses the site, downloads the compressed files, uncompresses them on his own terminal, and writes the patches onto his raster charts. He can then toggle between old and new versions to see exactly what has changed, and can view the patch by itself.

NOAA developed this process under an agreement with a commercial partner, which produces the CD-ROM containing chart data. The CD-ROM also contains *Coast Pilots*, *Light Lists*, *Tide Tables*, and *Tidal Current Tables*, thus comprising on one CD-ROM the entire suite of publications required by USCG regulations for certain classes of vessels. Additional information can be found at the NOAA Web site at: http://chartmaker.ncd.noaa.gov.

See Chapter 14 for a complete discussion on electronic charts and the means of correcting them.

CHAPTER 5

SHORT RANGE AIDS TO NAVIGATION

DEFINING SHORT RANGE AIDS TO NAVIGATION

500. Terms and Definitions

Short range aids to navigation are those intended to be used visually or by radar while in inland, harbor and approach, and coastal navigation. The term encompasses lighted and unlighted beacons, ranges, leading lights, buoys, and their associated sound signals. Each short range aid to navigation, commonly referred to as a NAVAID, fits within a system designed to warn the mariner of dangers and direct him toward safe water. An aid's function determines its color, shape, light characteristic, and sound. This chapter explains the U.S. Aids to Navigation System as well as the IALA Maritime Buoyage System.

The placement and maintenance of marine aids to navigation in U.S. waters is the responsibility of the United States Coast Guard. The Coast Guard maintains lighthouses, radiobeacons, racons, sound signals, buoys, and daybeacons on the navigable waters of the United States, its territories, and possessions. Additionally, the Coast Guard exercises control over privately owned navigation aid systems.

A **beacon** is a stationary, visual navigation aid. Large lighthouses and small single-pile structures are both beacons. Lighted beacons are called **lights**; unlighted beacons are **daybeacons**. All beacons exhibit a **daymark** of some sort. In the case of a lighthouse, the color and type of structure are the daymarks. On small structures, these daymarks, consisting of colored geometric shapes called **dayboards**, often have lateral significance. The markings on lighthouses and towers convey no lateral significance.

FIXED LIGHTS

501. Major and Minor Lights

Lights vary from tall, high intensity coastal lights to battery-powered lanterns on single wooden piles. Immovable, highly visible, and accurately charted, fixed lights provide navigators with an excellent source for bearings. The structures are often distinctively colored to aid in identification. See Figure 501a.

A **major light** is a high-intensity light exhibited from a fixed structure or a marine site. Major lights include primary seacoast lights and secondary lights. **Primary seacoast lights** are major lights established for making landfall from sea and coastwise passages from headland to headland. **Secondary lights** are major lights established at harbor entrances and other locations where high intensity and reliability are required.

A **minor light** usually displays a light of low to moderate intensity. Minor lights are established in harbors, along channels, rivers, and in isolated locations. They usually have numbering, coloring, and light and sound characteristics that are part of the lateral system of buoyage.

Lighthouses are placed where they will be of most use: on prominent headlands, at harbor and port entrances, on isolated dangers, or at other points where mariners can best use them to fix their position. The lighthouse's principal purpose is to support a light at a considerable height above the water, thereby increasing its geographic range. Support equipment is often housed near the tower.

With few exceptions, all major lights operate automatically. There are also many automatic lights on smaller structures maintained by the Coast Guard or other attendants. Unmanned major lights may have emergency generators and automatic monitoring equipment to increase the light's reliability.

Light structures' appearances vary. Lights in low-lying areas usually are supported by tall towers; conversely, light structures on high cliffs may be relatively short. However its support tower is constructed, almost all lights are similarly generated, focused, colored, and characterized.

Some major lights use modern rotating or flashing lights, but many older lights use **Fresnel** lenses. These lenses consist of intricately patterned pieces of glass in a heavy brass framework. Modern Fresnel-type lenses are cast from high-grade plastic; they are much smaller and lighter than their glass counterparts.

A **buoyant beacon** provides nearly the positional accuracy of a light in a place where a buoy would normally be used. See Figure 501b. The buoyant beacon consists of a heavy sinker to which a pipe structure is tightly moored. A buoyancy chamber near the surface supports the pipe. The light, radar reflector, and other devices are located atop the pipe above the surface of the water. The pipe with its buoyancy chamber tends to remain upright even in severe weather and heavy currents, providing a smaller watch cir-

Figure 501a. Typical offshore light station.

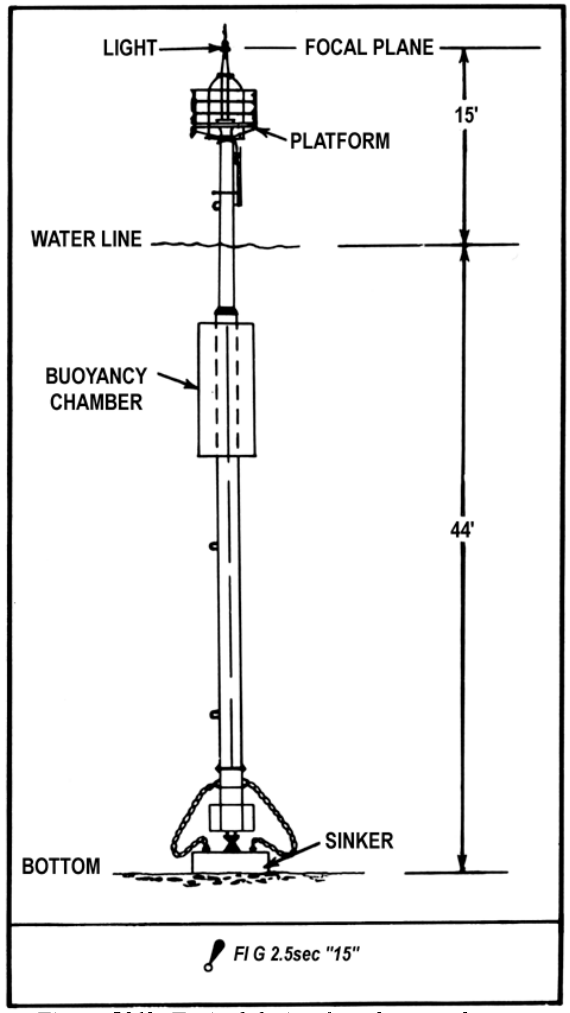


Figure 501b. Typical design for a buoyant beacon.

cle than a buoy. The buoyant beacon is most useful along narrow ship channels in relatively sheltered water.

502. Range Lights

Range lights are light pairs that indicate a specific line of position when they are in line. The higher rear light is placed behind the front light. When the mariner sees the lights vertically in line, he is on the range line. If the front light appears left of the rear light, the observer is to the right of the range line; if the front appears to the right of the rear, the observer is left of the range line. Range lights are sometimes equipped with high intensity lights for daylight use. These are effective for long channels in hazy conditions when dayboards might not be seen. The range light structures are usually also equipped with dayboards for ordinary daytime use. Some smaller ranges, primarily in the Intercoastal Waterway, rivers, and other inland waters, have just the dayboards with no lights. See Figure 502.

To enhance the visibility of range lights, the Coast Guard has developed 15-foot long lighted tubes called **light pipes**. They are mounted vertically, and the mariner sees them as vertical bars of light distinct from background lighting. Installation of light pipes is proceeding on several range markers throughout the country. The Coast Guard is also experimenting with long range sodium lights for areas requiring visibility greater than the light pipes can provide.

The output from a low pressure sodium light is almost entirely at one wavelength. This allows the use of an inexpensive band-pass filter to make the light visible even during the daytime. This arrangement eliminates the need for high intensity lights with their large power requirements.

Range lights are usually white, red, or green. They display various characteristics differentiating them from surrounding lights.

A **directional light** is a single light that projects a high intensity, special characteristic beam in a given direction. It is used in cases where a two-light range may not be practicable. A **directional sector light** is a directional light that emits two or more colored beams. The beams have a pre-

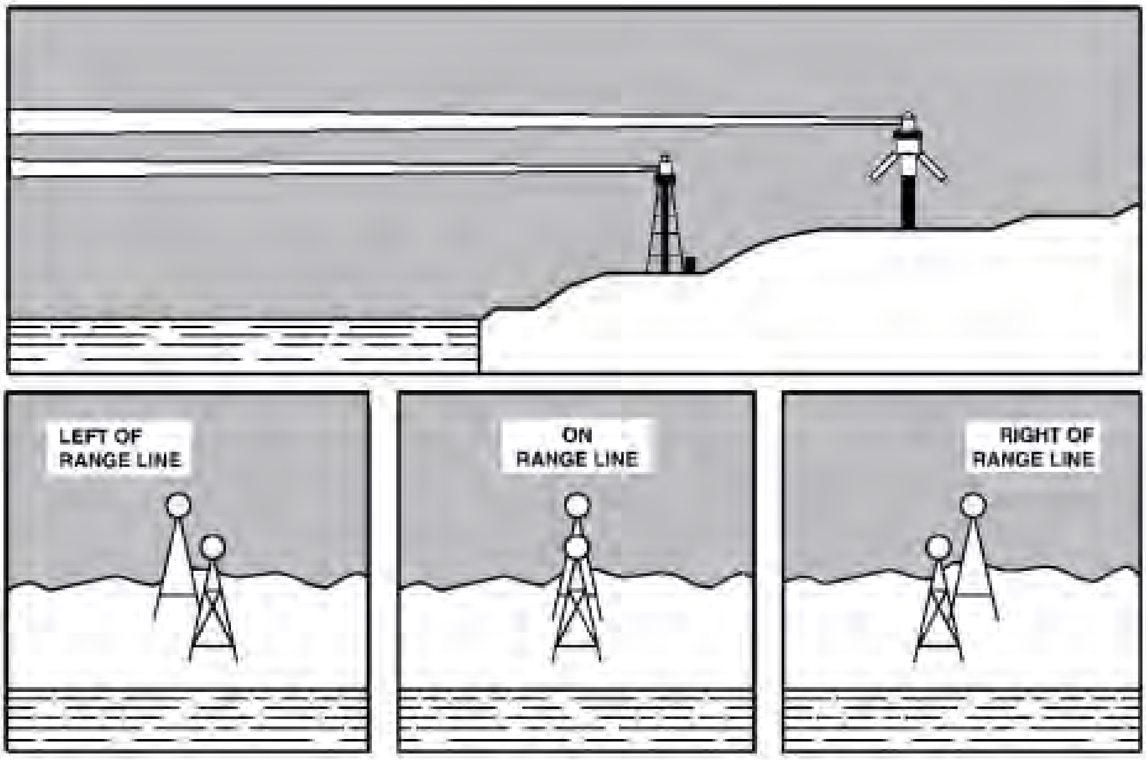


Figure 502. Range lights.

cisely oriented boundary between them. A normal application of a sector light would show three colored sections: red, white, and green. The white sector would indicate that the vessel is on the channel centerline; the green sector would indicate that the vessel is off the channel centerline in the direction of deep water; and the red sector would indicate that the vessel is off the centerline in the direction of shoal water.

503. Aeronautical Lights

Aeronautical lights may be the first lights observed at night when approaching the coast. Those situated near the coast and visible from sea are listed in the *List of Lights*. These lights are not listed in the Coast Guard *Light List*. They usually flash alternating white and green.

Aeronautical lights are sequenced geographically in the *List of Lights* along with marine navigation lights. However, since they are not maintained for marine navigation, they are subject to changes of which maritime authorities may not be informed. These changes will be published in *Notice to Airmen* but perhaps not in *Notice to Mariners*.

504. Bridge Lights

Navigational lights on bridges in the U.S. are prescribed by Coast Guard regulations. Red, green, and white lights mark bridges across navigable waters. Red lights mark piers and other parts of the bridge. Red lights are also used on drawbridges to show when they are in the closed position. Green lights mark open drawbridges and mark the centerline of navigable channels through fixed bridges. The position will vary according to the type of structure.

Infrequently-used bridges may be unlighted. In foreign waters, the type and method of lighting may be different from those normally found in the United States. Drawbridges which must be opened to allow passage operate upon sound and light signals given by the vessel and acknowledged by the bridge. These required signals are detailed in the Code of Federal Regulations and the applicable *Coast Pilot*. Certain bridges may also be equipped with sound signals and radar reflectors.

505. Shore Lights

Shore lights usually have a shore-based power supply. Lights on pilings, such as those found in the Intracoastal Waterway, are battery powered. Solar panels may be installed to enhance the light's power supply. The lights consist of a power source, a flasher to determine the characteristic, a lamp changer to replace burned-out lamps, and a focusing lens.

Various types of rotating lights are in use. They do not have flashers but remain continuously lit while a lens or reflector rotates around the horizon.

The aids to navigation system is carefully engineered to provide the maximum amount of direction to the mariner for the least expense. Specially designed filaments and special grades of materials are used in the light to withstand the harsh marine environment.

The **flasher** electronically determines the characteristic by selectively interrupting the light's power supply according to the chosen cycle.

The **lamp changer** consists of several sockets arranged around a central hub. When the circuit is broken by a burned-out filament, a new lamp is rotated into position. Almost all lights have daylight switches which turn the light off at sunrise and on at dusk.

The **lens** for small lights may be one of several types. The common ones in use are omni-directional lenses of 155mm, 250mm, and 300mm diameter. In addition, lights using parabolic mirrors or focused-beam lenses are used in leading lights and ranges. The lamp filaments must be carefully aligned with the plane of the lens or mirror to provide the maximum output of light. The lens' size is chosen according to the type of platform, power source, and lamp characteristics. Additionally, environmental characteristics of the location are considered. Various types of light-condensing panels, reflex reflectors, or colored sector panels may be installed inside the lens to provide the proper characteristic. A specially reinforced 200mm lantern is used in locations where ice and breaking water are a hazard.

LIGHT CHARACTERISTICS

506. Characteristics

A light has distinctive **characteristics** which distinguish it from other lights or convey specific information by showing a distinctive sequence of light and dark intervals. Additionally, a light may display a distinctive color or color sequence. In the *Light Lists*, the dark intervals are referred to as **eclipses**.

An **occulting** light is a light totally eclipsed at regular intervals, the duration of light always being greater than the duration of darkness. A **flashing** light flashes on and off at regular intervals, the duration of light always being less than the duration of darkness. An **isophase** light flashes at regular intervals, the duration of light being equal to the duration of darkness.

Light phase characteristics (See Table 506) are the distinctive sequences of light and dark intervals or sequences in the variations of the luminous intensity of a light. The light phase characteristics of lights which change color do not differ from those of lights which do not change color. A light showing different colors alternately is described as an **alternating** light. The alternating characteristic may be used with other light phase characteristics.

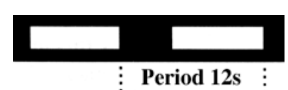
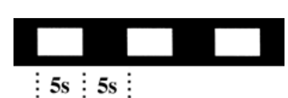
TYPE	ABBREVIATION	GENERAL DESCRIPTION	ILLUSTRATION*
Fixed	F.	A continuous and steady light.	
Occulting	Oc.	The total duration of light in a period is longer than the total duration of darkness and the intervals of darkness (eclipses) are usually of equal duration. Eclipse regularly repeated.	
Group occulting	Oc.(2)	An occulting light for which a group of eclipses, specified in number, is regularly repeated.	
Composite group occulting	Oc.(2+1)	A light similar to a group occulting light except that successive groups in a period have different numbers of eclipses.	
Isophase	Iso	A light for which all durations of light and darkness are clearly equal.	

Table 506. Light phase characteristics.

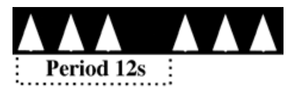
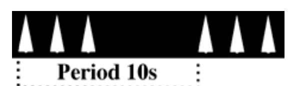
TYPE	ABBREVIATION	GENERAL DESCRIPTION	ILLUSTRATION*
Flashing	Fl.	A light for which the total duration of light in a period is shorter than the total duration of darkness and the appearances of light (flashes) are usually of equal duration (at a rate of less than 50 flashes per minute).	
Long flashing	L.Fl.	A single flashing light for which an appearance of light of not less than 2 sec. duration (long flash) is regularly repeated.	
Group flashing	Fl.(3)	A flashing light for which a group of flashes, specified in number, is regularly repeated.	
Composite group flashing	Fl.(2+1)	A light similar to a group flashing light except that successive groups in a period have different numbers of flashes.	
Quick flashing	Q.	A light for which a flash is regularly repeated at a rate of not less than 50 flashes per minute but less than 80 flashes per minute.	
Group quick flashing	Q.(3)	A light for which a specified group of flashes is regularly repeated; flashes are repeated at a rate of not less than 50 flashes per minute but less than 80 flashes per minute.	
	Q.(9)		
	Q.(6)+L.Fl.		
Interrupted quick flashing	I.Q.	A light for which the sequence of quick flashes is interrupted by regularly repeated eclipses of constant and long duration.	
Very quick flashing	V.Q.	A light for which a flash is regularly repeated at a rate of not less than 80 flashes per minute but less than 160 flashes per minute.	

Table 506. Light phase characteristics.

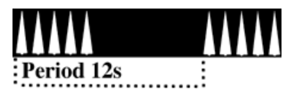
TYPE	ABBREVIATION	GENERAL DESCRIPTION	ILLUSTRATION*
Group very quick flashing	V.Q.(3)	A light for which a specified group of very quick flashes is regularly repeated.	
	V.Q.(9)		
	V.Q.(6)+L.Fl.		
Interrupted very quick flashing	I.V.Q.	A light for which the sequence of very quick flashes is interrupted by regularly repeated eclipses of constant and long duration.	
Ultra quick flashing	U.Q.	A light for which a flash is regularly repeated at a rate of not less than 160 flashes per minute.	
Interrupted ultra quick flashing	I.U.Q.	A light for which the sequence of ultra quick flashes is interrupted by regularly repeated eclipses of constant and long duration.	
Morse code	Mo.(U)	A light for which appearances of light of two clearly different durations are grouped to represent a character or characters in Morse Code.	
Fixed and flashing	F.Fl.	A light for which a fixed light is combined with a flashing light of greater luminous intensity	
Alternate light	Al.	A light showing different colors alternately	* Periods shown are examples only.
NOTE: Alternating lights may be used in combined form with most of the previous types of lights			

Table 506. Light phase characteristics.

Light-sensitive switches extinguish most lighted navigation aids during daylight hours. However, owing to the various sensitivities of the light switches, all lights do not turn on or off at the same time. Mariners should account for this when identifying aids to navigation during twilight periods when some lighted aids are on while others are not.

507. Light Sectors

Sectors of colored glass or plastic are sometimes placed in the lanterns of certain lights to indicate dangerous waters. Lights so equipped show different colors when observed from different bearings. A sector changes the color of a light, but not its characteristic, when viewed from certain directions. For example, a four second flashing white light having a red sector will appear as a four second flashing red light when viewed from within the red sector.

Sectors may be only a few degrees in width or extend in a wide arc from deep water toward shore. Bearings referring to sectors are expressed in degrees true as observed from a vessel. In most cases, areas covered by red sectors should be avoided. The nature of the danger can be determined from the chart. In some cases a narrow sector may mark the best water across a shoal, or a turning point in a channel.

The transition from one color to another is not abrupt. The colors change through an arc of uncertainty of 2° or greater, depending on the optical design of the light. Therefore determining bearings by observing the color change is less accurate than obtaining a bearing with an azimuth circle.

508. Factors Affecting Range and Characteristics

The condition of the atmosphere has a considerable effect upon a light’s range. Lights are sometimes obscured by fog, haze, dust, smoke, or precipitation. On the other hand, refraction may cause a light to be seen farther than under ordinary circumstances. A light of low intensity will be easily obscured by unfavorable conditions of the atmosphere. For this reason, the intensity of a light should always be considered when looking for it in thick weather. Haze and distance may reduce the apparent duration of a light’s flash. In some conditions of the atmosphere, white lights may have a reddish hue. In clear weather green lights may have a more whitish hue.

Lights placed at higher elevations are more frequently obscured by clouds, mist, and fog than those near sea level. In regions where ice conditions prevail, an unattended light’s lantern panes may become covered with ice or snow This may reduce the light’s luminous range and change the light’s observed color.

The distance from a light cannot be estimated by its apparent brightness. There are too many factors which can change the perceived intensity. Also, a powerful, distant light may sometimes be confused with a smaller, closer one with similar characteristics. Every light sighted should be carefully evaluated to determine if it is the one expected.

The presence of bright shore lights may make it difficult to distinguish navigational lights from background lighting. Lights may also be obscured by various shore obstructions, natural and man-made. The Coast Guard requests mariners to report these cases to the nearest Coast Guard station.

A light’s **loom** is sometimes seen through haze or the reflection from low-lying clouds when the light is beyond its geographic range. Only the most powerful lights can generate a loom. The loom may be sufficiently defined to obtain a bearing. If not, an accurate bearing on a light beyond geographic range may sometimes be obtained by ascending to a higher level where the light can be seen, and noting a star directly over the light. The bearing of the star can then be obtained from the navigating bridge and the bearing to the light plotted indirectly.

At short distances, some of the brighter flashing lights may show a faint continuous light, or faint flashes, between regular flashes. This is due to reflections of a rotating lens on panes of glass in the lighthouse.

If a light is not sighted within a reasonable time after prediction, a dangerous situation may exist. Conversely, the light may simply be obscured or extinguished. The ship’s position should immediately be fixed by other means to determine any possibility of danger.

The apparent characteristic of a complex light may change with the distance of the observer. For example, a light with a characteristic of fixed white and alternating flashing white and red may initially show as a simple flashing white light. As the vessel draws nearer, the red flash will become visible and the characteristic will apparently be alternating flashing white and red. Later, the fainter fixed white light will be seen between the flashes and the true characteristic of the light finally recognized as fixed white, alternating flashing white and red (F W Al W R). This is because for a given candlepower, white is the most visible color, green less so, and red least of the three. This fact also accounts for the different ranges given in the *Light Lists* for some multi-color sector lights. The same lamp has different ranges according to the color imparted by the sector glass.

A light may be **extinguished** due to weather, battery failure, vandalism, or other causes. In the case of unattended lights, this condition might not be immediately corrected. The mariner should report this condition to the nearest Coast Guard station. During periods of armed conflict, certain lights may be deliberately extinguished without notice. Offshore light stations should always be left well off the course whenever searoom permits.

BUOYS

509. Definitions and Types

Buoys are floating aids to navigation. They mark channels, indicate shoals and obstructions, and warn the mariner of dangers. Buoys are used where fixed aids would be uneconomical or impractical due to the depth of water. By their color, shape, topmark, number, and light characteristics, buoys indicate to the mariner how to avoid hazards and stay in safe water. The federal buoyage system in the U.S. is maintained by the Coast Guard.

There are many different sizes and types of buoys designed to meet a wide range of environmental conditions and user requirements. The size of a buoy is determined primarily by its location. In general, the smallest buoy which will stand up to local weather and current conditions is chosen.

There are five types of buoys maintained by the Coast Guard. They are:

1. Lateral marks
2. Isolated danger marks
3. Safe water marks
4. Special marks
5. Information/regulatory marks

These conform in general to the specifications of the **International Association of Lighthouse Authorities (IALA)** buoyage system.

A **lighted buoy** is a floating hull with a tower on which a light is mounted. Batteries for the light are in watertight pockets in the buoy hull or in watertight boxes mounted on the buoy hull. To keep the buoy in an upright position, a counterweight is attached to the hull below the water's surface. A radar reflector is built into the buoy tower.

The largest of the typical U.S. Coast Guard buoys can be moored in up to 190 feet of water, limited by the weight of chain the hull can support. The focal plane of the light is 15 to 20 feet high. The designed nominal visual range is 3.8 miles, and the radar range 4 miles. Actual conditions will cause these range figures to vary considerably.

The smallest buoys are designed for protected water. Some are made of plastic and weigh only 40 pounds. Specially designed buoys are used for fast current, ice, and other environmental conditions.

A variety of special purpose buoys are owned by other governmental organizations. Examples of these organizations include the St. Lawrence Seaway Development Corporation, NOAA, and the Department of Defense. These buoys are usually navigational marks or data collection buoys with traditional round, boat-shaped, or discus-shaped hulls.

A special class of buoy, the **Ocean Data Acquisition System (ODAS)** buoy, is moored or floats free in offshore

Figure 509. Buoy showing counterweight.

waters. Positions are promulgated through radio warnings. These buoys are generally not large enough to cause damage to a large vessel in a collision, but should be given a wide berth regardless, as any loss would almost certainly result in the interruption of valuable scientific experiments. They are generally bright orange or yellow in color, with vertical stripes on moored buoys and horizontal bands on free-floating ones, and have a strobe light for night visibility.

Even in clear weather, the danger of collision with a buoy exists. If struck head-on, a large buoy can inflict severe damage to a large ship; it can sink a smaller one. Reduced visibility or heavy background lighting can contribute to the problem of visibility. The Coast Guard sometimes receives reports of buoys missing from station that were actually run down and sunk. Tugboats and towboats towing or pushing barges are particularly dangerous to buoys because of poor over-the-bow visibility when pushing or yawing during towing. The professional mariner must report *any* collision with a buoy to the nearest Coast Guard unit. Failure to do so may cause the next vessel to miss the channel or hit the obstruction marked by the buoy; it can also lead to fines and legal liability.

Routine on-station buoy maintenance consists of inspecting the mooring, cleaning the hull and superstructure, replacing the batteries, flasher, and lamps, checking wiring and venting systems, and verifying the buoy's exact position. Every few years, each buoy is replaced by a similar aid and returned to a Coast Guard maintenance facility for complete refurbishment.

The placement of a buoy depends on its purpose and its position on the chart. Most buoys are placed on their charted positions as accurately as conditions allow. However, if a buoy's purpose is to mark a shoal and the shoal is found to be in a different position than the chart shows, the buoy will be placed to properly mark the shoal, and not on its charted position.

510. Lights on Buoys

Buoy light systems consist of a **battery pack**, a **flasher** which determines the characteristic, a **lamp changer** which automatically replaces burned-out bulbs, a **lens** to focus the light, and a **housing** which supports the lens and protects the electrical equipment.

The **batteries** consist of 12-volt lead/acid type batteries electrically connected to provide sufficient power to run the proper flash characteristic and lamp size. These battery packs are contained in pockets in the buoy hull, accessible through water-tight bolted hatches or externally mounted boxes. Careful calculations based on light characteristics determine how much battery power to install.

The **flasher** determines the characteristic of the lamp. It is installed in the housing supporting the lens.

The **lamp changer** consists of several sockets arranged around a central hub. A new lamp rotates into position if the active one burns out.

Under normal conditions, the **lenses** used on buoys are 155mm in diameter at the base. 200 mm lenses are used where breaking waves or swells call for the larger lens. They are colored according to the charted characteristic of the buoy. As in shore lights, the lamp must be carefully focused so that the filament is directly in line with the focal plane of the lens. This ensures that the majority of the light produced is focused in a 360° horizontal fan beam. A buoy light has a relatively narrow vertical profile. Because the buoy rocks in the sea, the focal plane may only be visible for fractions of a second at great ranges. A realistic range for sighting buoy lights is 4-6 miles in good visibility and calm weather.

511. Sound Signals on Buoys

Lighted sound buoys have the same general configuration as lighted buoys but are equipped with either a bell, gong, whistle, or horn. **Bells** and **gongs** are sounded by tappers hanging from the tower that swing as the buoy rocks in the sea. Bell buoys produce only one tone; gong buoys produce several tones. The tone-producing device is mounted between the legs of the pillar or tower.

Whistle buoys make a loud moaning sound caused by the rising and falling motions of the buoy in the sea. A sound buoy equipped with an electronic **horn** will produce a pure tone at regular intervals regardless of the sea state. Unlighted sound buoys have the same general appearance as lighted buoys, but their underwater shape is designed to make them lively in all sea states.

512. Buoy Moorings

Buoys require **moorings** to hold them in position. Typically the mooring consists of **chain** and a large concrete or cast iron **sinker**. See Figure 512. Because buoys are subjected to waves, wind, and tides, the moorings must be deployed with chain lengths much greater than the water depth. The scope of chain will normally be about 3 times the water depth. The length of the mooring chain defines a **watch circle** within which the buoy can be expected to swing. It is for this reason that the charted buoy symbol has a "position approximate" circle to indicate its charted position, whereas a light position is shown by a dot at the exact location. Actual watch circles do not necessarily coincide with the "position approximate" circles which represent them.

Figure 512. A sinker used to anchor a buoy.

Over several years, the chain gradually wears out and must be replaced. The worn chain is often cast into the concrete of new sinkers.

513. Large Navigational Buoys

Large navigational buoys are moored in open water at approaches to certain major seacoast ports and monitored from shore stations by radio signals. These 40-foot diameter buoys (Figure 513) show lights from heights of about 36 feet above the water. Emergency lights automatically energize if the main light is extinguished. These buoys may also have a radiobeacon and sound signals.

514. Wreck Buoys

A **wreck buoy** usually cannot be placed directly over the wreck it is intended to mark because the buoy tender may not want to pass over a shallow wreck or risk fouling the buoy mooring. For this reason, a wreck buoy is usually

Figure 513. Large navigational buoy.

placed as closely as possible on the seaward or channelward side of a wreck. In some situations, two buoys may be used to mark the wreck, one lying off each end. The wreck may lie directly between them or inshore of a line between them, depending on the local situation. The *Local Notice to Mariners* should be consulted concerning details of the placement of wreck buoys on individual wrecks. Often it will also give particulars of the wreck and what activities may be in progress to clear it.

The charted position of a wreck buoy will usually be offset from the actual geographic position so that the wreck and buoy symbols do not coincide. Only on the largest scale chart will the actual and charted positions of both wreck and buoy be the same. Where they might overlap, it is the wreck symbol which occupies the exact charted position and the buoy symbol which is offset.

Wreck buoys are required to be placed by the owner of the wreck, but they may be placed by the Coast Guard if the owner is unable to comply with this requirement. In general, privately placed aids are not as reliable as Coast Guard aids.

Sunken wrecks are sometimes moved away from their buoys by storms, currents, freshets, or other causes. Just as shoals may shift away from the buoys placed to mark them, wrecks may shift away from wreck buoys.

515. Fallibility of Buoys

Buoys cannot be relied on to maintain their charted positions consistently. They are subject to a variety of hazards including severe weather, collision, mooring casualties, and electrical failure. Mariners should report discrepancies to the authority responsible for maintaining the aid.

The buoy symbol shown on charts indicates the approximate position of the sinker which secures the buoy to the seabed. The approximate position is used because of practical limitations in keeping buoys in precise geographical locations. These limitations include prevailing atmospheric and sea conditions, the slope and type of material making up the seabed, the scope of the mooring chain, and the fact that the positions of the buoys and the sinkers are not under continuous surveillance. The position of the buoy shifts around the area shown by the chart symbol due to the forces of wind and current.

A buoy may not be in its charted position because of changes in the feature it marks. For example, a buoy meant to mark a shoal whose boundaries are shifting might frequently be moved to mark the shoal accurately. A *Local Notice to Mariners* will report the change, and a *Notice to Mariners* chart correction may also be written. In some small channels which change often, buoys are not charted even when considered permanent; local knowledge is advised in such areas.

For these reasons, a mariner must not rely completely upon the position or operation of buoys, but should navigate using bearings of charted features, structures, and aids to navigation on shore. Further, a vessel attempting to pass too close aboard a buoy risks a collision with the buoy or the obstruction it marks.

BUOYAGE SYSTEMS

516. Lateral and Cardinal Systems

There are two major types of buoyage systems: the **lateral system** and the **cardinal system**. The lateral system is best suited for well-defined channels. The description of each buoy indicates the direction of danger relative to the course which is normally followed. In principle, the positions of marks in the lateral system are determined by the **general direction** taken by the mariner when approaching port from seaward. These positions may also be determined with reference to the main stream of flood current. The United States Aids to Navigation System is a lateral system.

The cardinal system is best suited for coasts with numerous isolated rocks, shoals, and islands, and for dangers in the open sea. The characteristic of each buoy indicates the approximate true bearing of the danger it marks. Thus, an eastern quadrant buoy marks a danger which lies to the west of the buoy. The following pages diagram the cardinal and lateral buoyage systems as found outside the United States.

517. The IALA Maritime Buoyage System

Although most of the major maritime nations have used either the lateral or the cardinal system for many years, details such as the buoy shapes and colors have varied from country to country. With the increase in maritime commerce between countries, the need for a uniform system of buoyage became apparent.

In 1889, an International Marine Conference held in Washington, D.C., recommended that in the lateral system, starboard hand buoys be painted red and port hand buoys black. Unfortunately, when lights for buoys were introduced some years later, some European countries placed red lights on the black port hand buoys to conform with the red lights marking the port side of harbor entrances, while in North America red lights were placed on red starboard hand buoys. In 1936, a League of Nations subcommittee recommended a coloring system opposite to the 1889 proposal.

The **International Association of Lighthouse Authorities (IALA)** is a non-governmental organization which consists of representatives of the worldwide community of aids to navigation services. It promotes information exchange and recommends improvements based on new technologies. In 1980, with the assistance of IMO and the IHO, the lighthouse authorities from 50 countries and representatives of 9 international organizations concerned with aids to navigation met and adopted the **IALA Maritime Buoyage System**. They established two regions, **Region A** and **Region B**, for the entire world. Region A roughly corresponds to the 1936 League of Nations system, and Region B to the older 1889 system.

Lateral marks differ between Regions A and B. Lateral marks in Region A use red and green colors by day and night to indicate port and starboard sides of channels, respectively. In Region B, these colors are reversed with red to starboard and green to port. In both systems, the conventional direction of buoyage is considered to be returning from sea, hence the phrase “red right returning” in IALA region B.

518. Types of Marks

The **IALA Maritime Buoyage System** applies to all fixed and floating marks, other than lighthouses, sector lights, range lights, daymarks, lightships and large navigational buoys, which indicate:

1. The side and center-lines of navigable channels
2. Natural dangers, wrecks, and other obstructions
3. Regulated navigation areas
4. Other important features

Most lighted and unlighted beacons other than range marks are included in the system. In general, beacon topmarks will have the same shape and colors as those used on buoys. The system provides five types of marks which may be used in any combination:

1. Lateral marks indicate port and starboard sides of channels.
2. Cardinal marks, named according to the four points of the compass, indicate that the navigable water lies to the named side of the mark.
3. Isolated danger marks erected on, or moored directly on or over, dangers of limited extent.
4. Safe water marks, such as midchannel buoys.

5. Special marks, the purpose of which is apparent from reference to the chart or other nautical documents.

Characteristics of Marks

The significance of a mark depends on one or more features:

1. By day—color, shape, and topmark
2. By night—light color and phase characteristics

Colors of Marks

The colors red and green are reserved for lateral marks, and yellow for special marks. The other types of marks have black and yellow or black and red horizontal bands, or red and white vertical stripes.

Shapes of Marks

There are five basic buoy shapes:

1. Can
2. Cone
3. Sphere
4. Pillar
5. Spar

In the case of can, conical, and spherical, the shapes have lateral significance because the shape indicates the correct side to pass. With pillar and spar buoys, the shape has no special significance.

The term "pillar" is used to describe any buoy which is smaller than a large navigation buoy (LNB) and which has a tall, central structure on a broad base; it includes beacon buoys, high focal plane buoys, and others (except spar buoys) whose body shape does not indicate the correct side to pass.

Topmarks

The IALA System makes use of **can, conical, spherical,** and **X-shaped** topmarks only. Topmarks on pillar and spar buoys are particularly important and will be used wherever practicable, but ice or other severe conditions may occasionally prevent their use.

Colors of Lights

Where marks are lighted, red and green lights are reserved for lateral marks, and yellow for special marks. The other types of marks have a white light, distinguished one from another by phase characteristic.

Phase Characteristics of Lights

Red and green lights may have any phase characteristic, as the color alone is sufficient to show on which side they should be passed. Special marks, when lighted, have a yellow light with any phase characteristic not reserved for white lights of the system. The other types of marks have clearly specified phase characteristics of white light: various quick-flashing phase characteristics for cardinal marks, group flashing (2) for isolated danger marks, and relatively long periods of light for safe water marks.

Some shore lights specifically excluded from the IALA System may coincidentally have characteristics corresponding to those approved for use with the new marks. Care is needed to ensure that such lights are not misinterpreted.

519. IALA Lateral Marks

Lateral marks are generally used for well-defined channels; they indicate the port and starboard hand sides of the route to be followed, and are used in conjunction with a **conventional direction of buoyage**.

This direction is defined in one of two ways:

1. **Local direction of buoyage** is the direction taken by the mariner when approaching a harbor, river estuary, or other waterway from seaward.
2. **General direction of buoyage** is determined by the buoyage authorities, following a clockwise direction around continental land-masses, given in sailing directions, and, if necessary, indicated on charts by a large open arrow symbol.

In some places, particularly straits open at both ends, the local direction of buoyage may be overridden by the general direction.

Along the coasts of the United States, the characteristics assume that proceeding "from seaward" constitutes a clockwise direction: a southerly direction along the Atlantic coast, a westerly direction along the Gulf of Mexico coast, and a northerly direction along the Pacific coast. On the Great Lakes, a westerly and northerly direction is taken as being "from seaward" (except on Lake Michigan, where a southerly direction is used). On the Mississippi and Ohio Rivers and their tributaries, the characteristics of aids to navigation are determined as proceeding from sea toward the head of navigation. On the Intracoastal Waterway, proceeding in a generally southerly direction along the Atlantic coast, and in a generally westerly direction along the gulf coast, is considered as proceeding "from seaward."

520. IALA Cardinal Marks

A **cardinal mark** is used in conjunction with the compass to indicate where the mariner may find the best navigable water. It is placed in one of the four quadrants (north, east, south, and west), bounded by the true bearings

NW-NE, NE-SE, SE-SW, and SW-NW, taken from the point of interest. A cardinal mark takes its name from the quadrant *in which it is placed*.

The mariner is safe if he passes north of a north mark, east of an east mark, south of a south mark, and west of a west mark.

A cardinal mark may be used to:

1. Indicate that the deepest water in an area is on the named side of the mark.
2. Indicate the safe side on which to pass a danger.
3. Emphasize a feature in a channel, such as a bend, junction, bifurcation, or end of a shoal.

Topmarks

Black double-cone topmarks are the most important feature, by day, of cardinal marks. The cones are vertically placed, one over the other. The arrangement of the cones is very logical: North is two cones with their points up (as in "north-up"). South is two cones, points down. East is two cones with bases together, and west is two cones with points together, which gives a wineglass shape. "West is a Wineglass" is a memory aid.

Cardinal marks carry topmarks whenever practicable, with the cones as large as possible and clearly separated.

Colors

Black and yellow horizontal bands are used to color a cardinal mark. The position of the black band, or bands, is related to the points of the black topmarks.

N	Points up	Black above yellow
S	Points down	Black below yellow
W	Points together	Black, yellow above and below
E	Points apart	Yellow, black above and below

Shape

The shape of a cardinal mark is not significant, but buoys must be pillars or spars.

Lights

When lighted, a cardinal mark exhibits a white light; its characteristics are based on a group of quick or very quick flashes which distinguish it as a cardinal mark and indicate its quadrant. The distinguishing quick or very quick flashes are:

North—Uninterrupted
East—three flashes in a group
South—six flashes in a group followed by a long flash
West—nine flashes in a group

As a memory aid, the number of flashes in each group can be associated with a clock face: 3 o'clock—E, 6 o'clock—S, and 9 o'clock—W.

The long flash (of not less than 2 seconds duration), immediately following the group of flashes of a south cardinal mark, is to ensure that its six flashes cannot be mistaken for three or nine.

The periods of the east, south, and west lights are, respectively, 10, 15, and 15 seconds if quick flashing; and 5, 10, and 10 seconds if very quick flashing.

Quick flashing lights flash at a rate between 50 and 79 flashes per minute, usually either 50 or 60. Very quick flashing lights flash at a rate between 80 and 159 flashes per minute, usually either 100 or 120.

It is necessary to have a choice of quick flashing or very quick flashing lights in order to avoid confusion if, for example, two north buoys are placed near enough to each other for one to be mistaken for the other.

521. IALA Isolated Danger Marks

An **isolated danger mark** is erected on, or moored on or above, an isolated danger of limited extent which has navigable water all around it. The extent of the surrounding navigable water is immaterial; such a mark can, for example, indicate either a shoal which is well offshore or an islet separated by a narrow channel from the coast.

Position

On a chart, the position of a danger is the center of the symbol or sounding indicating that danger; an isolated danger buoy may therefore be slightly displaced from its geographic position to avoid overprinting the two symbols. The smaller the scale, the greater this offset will be. At very large scales the symbol may be correctly charted.

Topmark

A black double-sphere topmark is, by day, the most important feature of an isolated danger mark. Whenever practicable, this topmark will be carried with the spheres as large as possible, disposed vertically, and clearly separated.

Color

Black with one or more red horizontal bands are the colors used for isolated danger marks.

Shape

The shape of an isolated danger mark is not significant, but a buoy will be a pillar or a spar.

Light

When lighted, a white flashing light showing a group of two flashes is used to denote an isolated danger mark. As a memory aid, associate two flashes with two balls in the topmark.

522. IALA Safe Water Marks

A **safe water mark** is used to indicate that there is navigable water all around the mark. Such a mark may be used as a center line, mid-channel, or landfall buoy.

Color

Red and white vertical stripes are used for safe water marks, and distinguish them from the black-banded, danger-marking marks.

Shape

Spherical, pillar, or spar buoys may be used as safe water marks.

Topmark

A single red spherical topmark will be carried, whenever practicable, by a pillar or spar buoy used as a safe water mark.

Lights

When lighted, safe water marks exhibit a white light. This light can be occulting, isophase, a single long flash, or Morse "A." If a long flash (i.e. a flash of not less than 2 seconds) is used, the period of the light will be 10 seconds. As a memory aid, remember a single flash and a single sphere topmark.

523. IALA Special Marks

A **special mark** may be used to indicate a special area or feature which is apparent by referring to a chart, sailing directions, or notices to mariners. Uses include:

1. Ocean Data Acquisition System (ODAS) buoys
2. Traffic separation marks
3. Spoil ground marks
4. Military exercise zone marks
5. Cable or pipeline marks, including outfall pipes
6. Recreation zone marks

Another function of a special mark is to define a channel within a channel. For example, a channel for deep draft vessels in a wide estuary, where the limits of the channel for normal navigation are marked by red and green lateral buoys, may have its boundaries or centerline marked by yellow buoys of the appropriate lateral shapes.

Color

Yellow is the color used for special marks.

Shape

The shape of a special mark is optional, but must not conflict with that used for a lateral or a safe water mark. For example, an outfall buoy on the port hand side of a channel could be can-shaped but not conical.

Topmark

When a topmark is carried it takes the form of a single yellow X.

Lights

When a light is exhibited it is yellow. It may show any phase characteristic except those used for the white lights of cardinal, isolated danger, and safe water marks. In the case of ODAS buoys, the phase characteristic used is group-flashing with a group of five flashes every 20 seconds.

524. IALA New Dangers

A newly discovered hazard to navigation not yet shown on charts, included in sailing directions, or announced by a *Notice to Mariners* is termed a **new danger**. The term covers naturally occurring and man-made obstructions.

Marking

A new danger is marked by one or more cardinal or lateral marks in accordance with the IALA system rules. If the danger is especially grave, at least one of the marks will be duplicated as soon as practicable by an identical mark until the danger has been sufficiently identified.

Lights

If a lighted mark is used for a new danger, it must exhibit a quick flashing or very quick flashing light. If a cardinal mark is used, it must exhibit a white light; if a lateral mark, a red or green light.

Racons

The duplicate mark may carry a Racon, Morse coded D, showing a signal length of 1 nautical mile on a radar display.

525. Chart Symbols and Abbreviations

Spar buoys and spindle buoys are represented by the same symbol; it is slanted to distinguish them from upright beacon symbols. The abbreviated description of the color of a buoy is given under the symbol. Where a buoy is colored in bands, the colors are indicated in sequence from the top. If the sequence of the bands is not known, or if the buoy is striped, the colors are indicated with the darker color first.

Topmarks

Topmark symbols are solid black except if the topmark is red.

Lights

The period of the light of a cardinal mark is determined by its quadrant and its flash characteristic (either quick-flashing or a very quick-flashing). The light's period is less important than its phase characteristic. Where space on charts is limited, the period may be omitted.

Light Flares

Magenta light-flares are normally slanted and inserted with their points adjacent to the position circles at the base of the symbols so the flare symbols do not obscure the topmark symbols.

Radar Reflectors

According to IALA rules, radar reflectors are not charted, for several reasons. First, all important buoys are fitted with radar reflectors. It is also necessary to reduce the size and complexity of buoy symbols and associated legends. Finally, it is understood that, in the case of cardinal buoys, buoyage authorities place the reflector so that it cannot be mistaken for a topmark.

The symbols and abbreviations of the IALA Maritime Buoyage System may be found in *U.S. Chart No. 1* and in foreign equivalents.

526. Description of the U.S. Aids to Navigation System

In the United States, the U.S. Coast Guard has incorporated the major features of the IALA system with the existing infrastructure of buoys and lights as explained below.

Colors

Under this system, green buoys mark a channel's port side and obstructions which must be passed by keeping the buoy on the port hand. Red buoys mark a channel's starboard side and obstructions which must be passed by keeping the buoy on the starboard hand.

Red and green horizontally banded **preferred channel buoys** mark junctions or bifurcations in a channel or obstructions which may be passed on either side. If the topmost band is green, the preferred channel will be followed by keeping the buoy on the port hand. If the topmost band is red, the preferred channel will be followed by keeping the buoy on the starboard hand.

Red and white vertically striped safe water buoys mark a fairway or mid-channel.

Reflective material is placed on buoys to assist in their detection at night with a searchlight. The color of the reflective material agrees with the buoy color. Red or green reflective material may be placed on preferred channel (junction) buoys; red if topmost band is red, or green if the topmost band is green. White reflective material is used on safe water buoys. Special purpose buoys display yellow reflective material. Warning or regulatory buoys display orange reflective horizontal bands and a warning symbol. Intracoastal Waterway buoys display a yellow reflective square, triangle, or horizontal strip along with the reflective material coincident with the buoy's function.

Shapes

Certain unlighted buoys are differentiated by shape. Red buoys and red and green horizontally banded buoys with the topmost band red are cone-shaped buoys called **nuns**. Green buoys and green and red horizontally banded buoys with the topmost band green are cylinder-shaped buoys called **cans**.

Unlighted red and white vertically striped buoys may be pillar shaped or spherical. Lighted buoys, sound buoys, and spar buoys are not differentiated by shape to indicate the side on which they should be passed. Their purpose is indicated not by shape but by the color, number, or light characteristics.

Numbers

All solid colored buoys are numbered, red buoys bearing even numbers and green buoys bearing odd numbers. (Note that this same rule applies in IALA System A also.) The numbers increase from seaward upstream or toward land. No other colored buoys are numbered; however, any buoy may have a letter for identification.

Light Colors

Red lights are used only on red buoys or red and green horizontally banded buoys with the topmost band red. Green lights are used only on the green buoys or green and red horizontally banded buoys with the topmost band green. White lights are used on both "safe water" aids showing a Morse Code "A" characteristic and on Information and Regulatory aids.

Light Characteristics

Lights on red buoys or green buoys, if not occulting or isophase, will generally be regularly flashing (Fl). For ordinary purposes, the frequency of flashes will be not more than 50 flashes per minute. Lights with a distinct cautionary significance, such as at sharp turns or marking dangerous obstructions, will flash not less than 50 flashes but not more than 80 flashes per minute (quick flashing, Q). Lights on preferred channel buoys will show a series of group flashes with successive groups in a period having a different number of flashes - composite group flashing (or a quick light in which the sequence of flashes is interrupted by regularly repeated eclipses of constant and long duration). Lights on safe water buoys will always show a white Morse Code "A" (Short-Long) flash recurring at the rate of approximately eight times per minute.

Daylight Controls

Lighted buoys have a special device to energize the light when darkness falls and to de-energize the light when day breaks. These devices are not of equal sensitivity; therefore all lights do not come on or go off at the same time. Mariners should ensure correct identification of aids during twilight periods when some light aids to navigation are on while others are not.

Special Purpose Buoys

Buoys for special purposes are colored yellow. White buoys with orange bands are for informational or regulatory purposes. The shape of special purpose buoys has no significance. They are not numbered, but they may be lettered. If lighted, special purpose buoys display a yellow light usually with fixed or slow flash characteristics. Information and regulatory buoys, if lighted, display white lights.

BEACONS

527. Definition and Description

Beacons are fixed aids to navigation placed on shore or on pilings in relatively shallow water. If unlighted, the beacon is referred to as a **daybeacon**. A daybeacon is identified by the color, shape, and number of its **dayboard**. The simplest form of daybeacon consists of a single pile with a dayboard affixed at or near its top. See Figure 527. Daybeacons may be used to form an unlighted range.

Dayboards identify aids to navigation against daylight backgrounds. The size of the dayboard required to make the aid conspicuous depends upon the aid's intended range.

Most dayboards also display numbers or letters for identification. The numbers, letters, and borders of most dayboards have reflective tape to make them visible at night.

The detection, recognition, and identification distances vary widely for any particular dayboard. They depend upon the luminance of the dayboard, the Sun's position, and the local visibility conditions.

Figure 527. Daybeacon.

SOUND SIGNALS

528. Types of Sound Signals

Most lighthouses and offshore light platforms, as well as some minor light structures and buoys, are equipped with sound-producing devices to help the mariner in periods of low visibility. Charts and *Light Lists* contain the information required for positive identification. Buoys fitted with bells, gongs, or whistles actuated by wave motion may produce no sound when the sea is calm. Sound signals are not designed to identify the buoy or beacon for navigation purposes. Rather, they allow the mariner to pass clear of the buoy or beacon during low visibility.

Sound signals vary. The navigator must use the *Light List* to determine the exact length of each blast and silent interval. The various types of sound signals also differ in tone, facilitating recognition of the respective stations.

Diaphones produce sound with a slotted piston moved back and forth by compressed air. Blasts may consist of a high and low tone. These alternate-pitch signals are called "two-tone." Diaphones are not used by the Coast Guard, but the mariner may find them on some private navigation aids.

Horns produce sound by means of a disc diaphragm operated pneumatically or electrically. Duplex or triplex horn units of differing pitch produce a chime signal.

Sirens produce sound with either a disc or a cup-shaped rotor actuated electrically or pneumatically. Sirens are not used on U.S. navigation aids.

Whistles use compressed air emitted through a circumferential slot into a cylindrical bell chamber.

Bells and gongs are sounded with a mechanically operated hammer.

529. Limitations of Sound Signals

As aids to navigation, sound signals have serious limitations because sound travels through the air in an unpredictable manner.

It has been clearly established that:

1. Sound signals are heard at greatly varying distances and that the distance at which a sound signal can be heard may vary with the bearing and timing of the signal.
2. Under certain atmospheric conditions, when a sound signal has a combination high and low tone, it is not unusual for one of the tones to be inaudible. In the case of sirens, which produce a varying tone, portions of the signal may not be heard.
3. When the sound is screened by an obstruction, there are areas where it is inaudible.
4. Operators may not activate a remotely controlled sound aid for a condition unobserved from the controlling station.
5. Some sound signals cannot be immediately started.
6. The status of the vessel's engines and the location of the observer both affect the effective range of the aid.

These considerations justify the utmost caution when navigating near land in a fog. A navigator can never rely on sound signals alone; he should continuously man both the radar and fathometer. He should place lookouts in positions where the noises in the ship are least likely to interfere with hearing a sound signal. The aid upon which a sound signal rests is usually a good radar target, but collision with the aid or the danger it marks is always a possibility.

Emergency signals are sounded at some of the light and fog signal stations when the main and stand-by sound signals are inoperative. Some of these emergency sound signals are of a different type and characteristic than the main sound signal. The characteristics of the emergency sound signals are listed in the *Light List*.

The mariner should never assume:

1. That he is out of ordinary hearing distance because he fails to hear the sound signal.
2. That because he hears a sound signal faintly, he is far from it.
3. That because he hears it clearly, he is near it.
4. That the distance from and the intensity of a sound on any one occasion is a guide for any future occasion.
5. That the sound signal is not sounding because he does not hear it, even when in close proximity.
6. That the sound signal is in the direction the sound appears to come from.

MISCELLANEOUS U.S. SYSTEMS

530. Intracoastal Waterway Aids to Navigation

The Intracoastal Waterway (ICW) runs parallel to the Atlantic and Gulf of Mexico coasts from Manasquan Inlet on the New Jersey shore to the Texas/Mexican border. It follows rivers, sloughs, estuaries, tidal channels, and other natural waterways, connected with dredged channels where necessary. Some of the aids marking these waters are marked with yellow; otherwise, the marking of buoys and beacons follows the same system as that in other U.S. waterways.

Yellow symbols indicate that an aid marks the Intracoastal Waterway. Yellow triangles indicate starboard hand aids, and yellow squares indicate port hand aids when following the ICW's conventional direction of buoyage. Non-lateral aids such as safe water, isolated danger, and front range boards are marked with a horizontal yellow band. Rear range boards do not display the yellow band. At a junction with a federally-maintained waterway, the preferred channel mark will display a yellow triangle or square as appropriate. Junctions between the ICW and privately maintained waterways are not marked with preferred channel buoys.

531. Western Rivers System

Aids to navigation on the Mississippi River and its tributaries above Baton Rouge generally conform to the lateral system of buoyage in use in the rest of the U.S. The following differences are significant:

1. Buoys are not numbered.
2. The numbers on lights and daybeacons do not have lateral significance; they indicate the mileage from a designated point, normally the river mouth.
3. Flashing lights on the left side proceeding upstream show single green or white flashes while those on the right side show group flashing red or white flashes.
4. Diamond shaped crossing daymarks are used to indicate where the channel crosses from one side of the river to the other.

532. The Uniform State Waterway Marking System (USWMS)

This system was developed jointly by the U.S. Coast Guard and state boating administrators to assist the small craft operator in those state waters marked by participating states. The **USWMS** consists of two categories of aids to navigation. The first is a system of aids to navigation, generally compatible with the Federal lateral system of buoyage, supplementing the federal system in state waters. The other is a system of regulatory markers to warn small craft operators of dangers or to provide general information.

On a well-defined channel, red and black buoys are established in pairs called **gates**; the channel lies between the buoys. The buoy which marks the left side of the channel viewed looking upstream or toward the head of navigation is black; the buoy which marks the right side of the channel is red.

In an irregularly-defined channel, buoys may be staggered on alternate sides of the channel, but they are spaced at sufficiently close intervals to mark clearly the channel lying between them.

Where there is no well-defined channel or where a body of water is obstructed by objects whose nature or location is such that the obstruction can be approached by a vessel from more than one direction, aids to navigation having cardinal significance may be used. The aids conforming to the cardinal system consist of three distinctly colored buoys as follows:

1. A white buoy with a red top must be passed to the south or west of the buoy.
2. A white buoy with a black top must be passed to the north or east of the buoy.
3. A buoy showing alternate vertical red and white stripes indicates that an obstruction to navigation extends from the nearest shore to the buoy and that the vessel must not pass between the buoy and the nearest shore.

The shape of buoys has no significance under the USWMS.

Regulatory buoys are colored white with orange horizontal bands completely around them. One band is at the top of the buoy and a second band just above the waterline of the buoy so that both orange bands are clearly visible.

Geometric shapes colored orange are placed on the white portion of the buoy body. The authorized geometric shapes and meanings associated with them are as follows:

1. A vertical open faced diamond shape means danger.
2. A vertical open faced diamond shape with a cross centered in the diamond means that vessels are excluded from the marked area.
3. A circular shape means that vessels in the marked area are subject to certain operating restrictions.
4. A square or rectangular shape indicates that directions or information is written inside the shape.

Regulatory markers consist of square and rectangular shaped signs displayed from fixed structures. Each sign is white with an orange border. Geometric shapes with the same meanings as those displayed on buoys are centered on the sign boards. The geometric shape displayed on a regulatory marker tells the mariner if he should stay well clear of the marker or if he may approach the marker in order to read directions.

533. Private Aids to Navigation

A **private navigation aid** is any aid established and maintained by entities other than the Coast Guard.

The Coast Guard must approve the placement of private navigation aids. In addition, the District Engineer, U.S. Army Corps of Engineers, must approve the placement of any structure, including aids to navigation, in the navigable waters of the U.S.

Private aids to navigation are similar to the aids established and maintained by the U.S. Coast Guard; they are specially designated on the chart and in the *Light List*. In some cases, particularly on large commercial structures, the aids are the same type of equipment used by the Coast Guard. Although the Coast Guard periodically inspects some private navigation aids, the mariner should exercise special caution when using them.

In addition to private aids to navigation, numerous types of construction and anchor buoys are used in various oil drilling operations and marine construction. These buoys are not charted, as they are temporary, and may not be lighted well or at all. Mariners should give a wide berth to drilling and construction sites to avoid the possibility of fouling moorings. This is a particular danger in offshore oil fields, where large anchors are often used to stabilize the positions of drill rigs in deep water. Up to eight anchors may be placed at various positions as much as a mile from the drill ship. These positions may or may not be marked by buoys. Such operations in the U.S. are announced in the *Local Notice to Mariners*.

534. Protection by Law

It is unlawful to impair the usefulness of any navigation aid established and maintained by the United States. If any vessel collides with a navigation aid, it is the legal duty of the person in charge of the vessel to report the accident to the nearest U.S. Coast Guard station.

CHAPTER 6

COMPASSES

INTRODUCTION

600. Changes in Compass Technologies

This chapter discusses the major types of compasses available to the navigator, their operating principles, their capabilities, and limitations of their use. As with other aspects of navigation, technology is rapidly revolutionizing the field of compasses. Amazingly, after at least a millennia of constant use, it is now possible (however advisable it may or may not be aboard any given vessel) to dispense with the traditional magnetic compass.

For much of maritime history the only heading reference for navigators has been the magnetic compass. A great deal of effort and expense has gone into understanding the magnetic compass scientifically and making it as accurate as possible through elaborate compensation techniques.

The introduction of the electro-mechanical gyrocompass relegated the magnetic compass to backup status for many large vessels. Later came the development of inertial navigation systems based on gyroscopic principles. The interruption of electrical power to the gyrocompass or inertial navigator, mechanical failure, or its physical destruction would instantly elevate the magnetic compass to primary status for most vessels.

New technologies are both refining and replacing the magnetic compass as a heading reference and navigational tool. Although a magnetic compass for backup is certainly advisable, today's navigator can safely avoid nearly all of the effort and expense associated with the binnacle-mounted magnetic compass, its compensation, adjustment, and maintenance.

Similarly, electro-mechanical gyrocompasses are being supplanted by far lighter, cheaper, and more dependable ring laser gyrocompasses. These devices do not operate on the principle of the gyroscope (which is based on Newton's laws of motion), but instead rely on the principles of electromagnetic energy and wave theory.

Magnetic flux gate compasses, while relying on the earth's magnetic field for reference, have no moving parts and can compensate themselves, adjusting for both deviation and variation to provide true heading, thus completely eliminating the process of compass correction.

To the extent that one depends on the magnetic compass for navigation, it should be checked regularly and adjusted when observed errors exceed certain minimal limits, usually a few degrees for most vessels. Compensation of a magnetic compass aboard vessels expected to rely on it offshore during long voyages is best left to professionals. However, this chapter will present enough material for the competent navigator to do a passable job.

Whatever type of compass is used, it is advisable to check it periodically against an error free reference to determine its error. This may be done when steering along any range during harbor and approach navigation, or by aligning any two charted objects and finding the difference between their observed and charted bearings. When navigating offshore, the use of azimuths and amplitudes of celestial bodies will also suffice, a subject covered in Chapter 17.

MAGNETIC COMPASSES

601. The Magnetic Compass and Magnetism

The principle of the present day magnetic compass is no different from that of the compasses used by ancient mariners. The magnetic compass consists of a magnetized needle, or an array of needles, allowed to rotate in the horizontal plane. The superiority of present day magnetic compasses over ancient ones results from a better knowledge of the laws of magnetism which govern the behavior of the compass and from greater precision in design and construction.

Any magnetized piece of metal will have regions of concentrated magnetism called **poles**. Any such magnet will have at least two poles of opposite polarity. Magnetic force (flux) lines connect one pole of such a magnet with the other pole. The number of such lines per unit area represents the intensity of the magnetic field in that area.

If two magnets are placed close to each other, the like poles will repel each other and the unlike poles will attract each other.

Magnetism can be either **permanent** or **induced**. A bar having permanent magnetism will retain its magnetism when it is removed from a magnetizing field. A bar having induced magnetism will lose its magnetism when removed from the magnetizing field. Whether or not a bar will retain its magnetism on removal from the magnetizing field will depend on the strength of that field, the degree of hardness of the iron (retentivity), and upon the amount of physical stress applied to the bar while in the magnetizing field. The harder the iron, the more permanent will be the magnetism acquired.

602. Terrestrial Magnetism

Consider the Earth as a huge magnet surrounded by lines of magnetic flux connecting its two **magnetic poles**. These magnetic poles are near, but not coincidental with, the Earth's geographic poles. Since the north seeking end of a compass needle is conventionally called the **north pole**, or **positive pole**, it must therefore be attracted to a **south pole**, or **negative pole**.

Figure 602a illustrates the Earth and its surrounding magnetic field. The flux lines enter the surface of the Earth at different angles to the horizontal at different magnetic latitudes. This angle is called the **angle of magnetic dip**, θ, and increases from 0° at the magnetic equator to 90° at the magnetic poles. The total magnetic field is generally considered as having two components: H, the horizontal component; and Z, the vertical component. These components change as the angle θ changes, such that H is at its maximum at the magnetic equator and decreases in the direction of either pole, while Z is zero at the magnetic equator and increases in the direction of either pole.

Since the magnetic poles of the Earth do not coincide with the geographic poles, a compass needle in line with the Earth's magnetic field will not indicate true north, but magnetic north. The angular difference between the true meridian (great circle connecting the geographic poles) and the magnetic meridian (direction of the lines of magnetic flux) is called **variation**. This variation has different values at different locations on the Earth. These values of magnetic variation may be found on pilot charts and on the compass rose of navigational charts.

The poles are not geographically static. They are known to migrate slowly, so that variation for most areas undergoes a small annual change, the amount of which is also noted on charts. Figure 602b and Figure 602c show magnetic dip and variation for the world. Up-to-date information on geomagnetics is available at http://geomag.usgs.gov/dod.html.

603. Ship's Magnetism

A ship under construction or repair will acquire permanent magnetism due to hammering and vibration while sitting stationary in the Earth's magnetic field. After launching, the ship will lose some of this original magnetism as a result of vibration and pounding in varying magnetic fields, and will eventually reach a more or less stable magnetic condition. The magnetism which remains is the **permanent magnetism** of the ship.

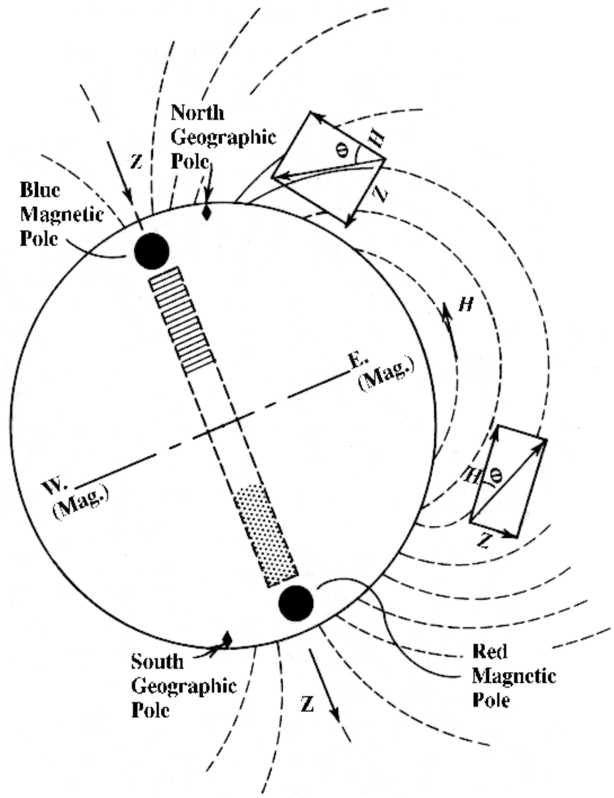


Figure 602a. Terrestrial magnetism.

In addition to its permanent magnetism, a ship acquires **induced magnetism** when placed in the Earth's magnetic field. The magnetism induced in any given piece of soft iron is a function of the field intensity, the alignment of the soft iron in that field, and the physical properties and dimensions of the iron. This induced magnetism may add to, or subtract from, the permanent magnetism already present in the ship, depending on how the ship is aligned in the magnetic field. The softer the iron, the more readily it will be magnetized by the Earth's magnetic field, and the more readily it will give up its magnetism when removed from that field.

The magnetism in the various structures of a ship, which tends to change as a result of cruising, vibration, or aging, but which does not alter immediately so as to be properly termed induced magnetism, is called **subpermanent magnetism**. This magnetism, at any instant, is part of the ship's permanent magnetism, and consequently must be corrected by permanent magnet correctors. It is the principal cause of deviation changes on a magnetic compass. Subsequent reference to permanent magnetism will refer to the apparent permanent magnetism which includes the existing permanent and subpermanent magnetism.

A ship, then, has a combination of permanent, subpermanent, and induced magnetism. Therefore, the ship's

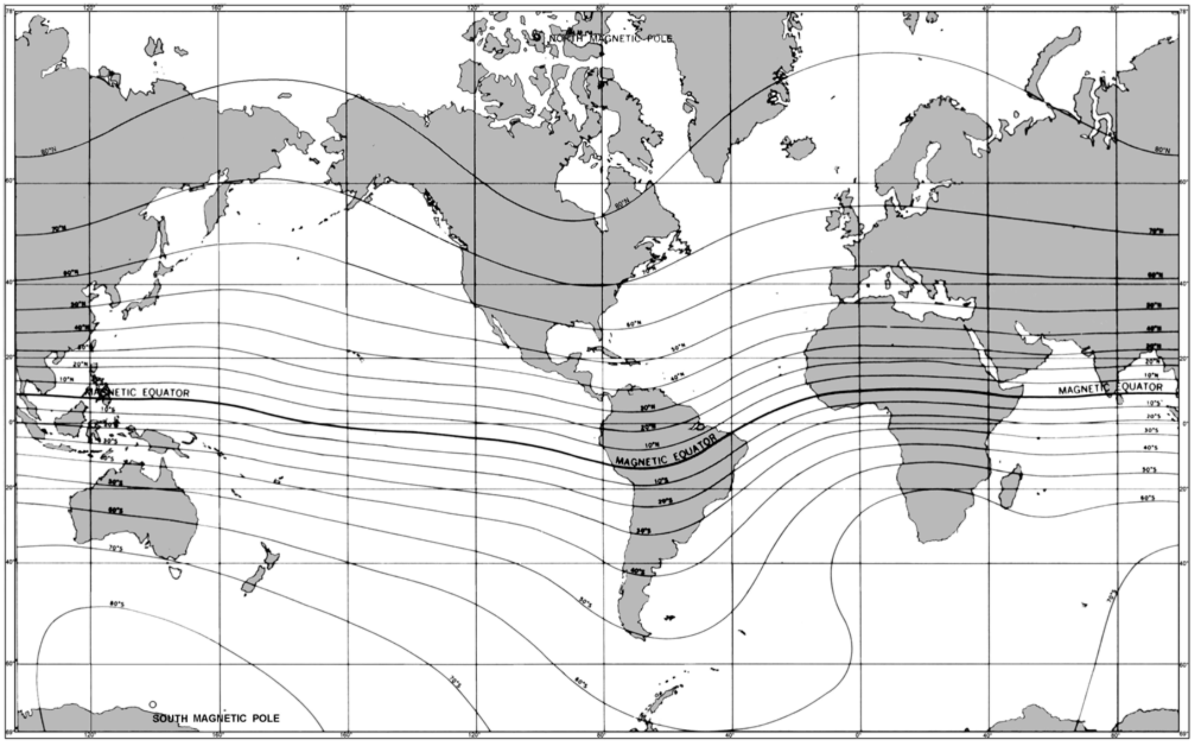


Figure 602b. Magnetic dip for the world.

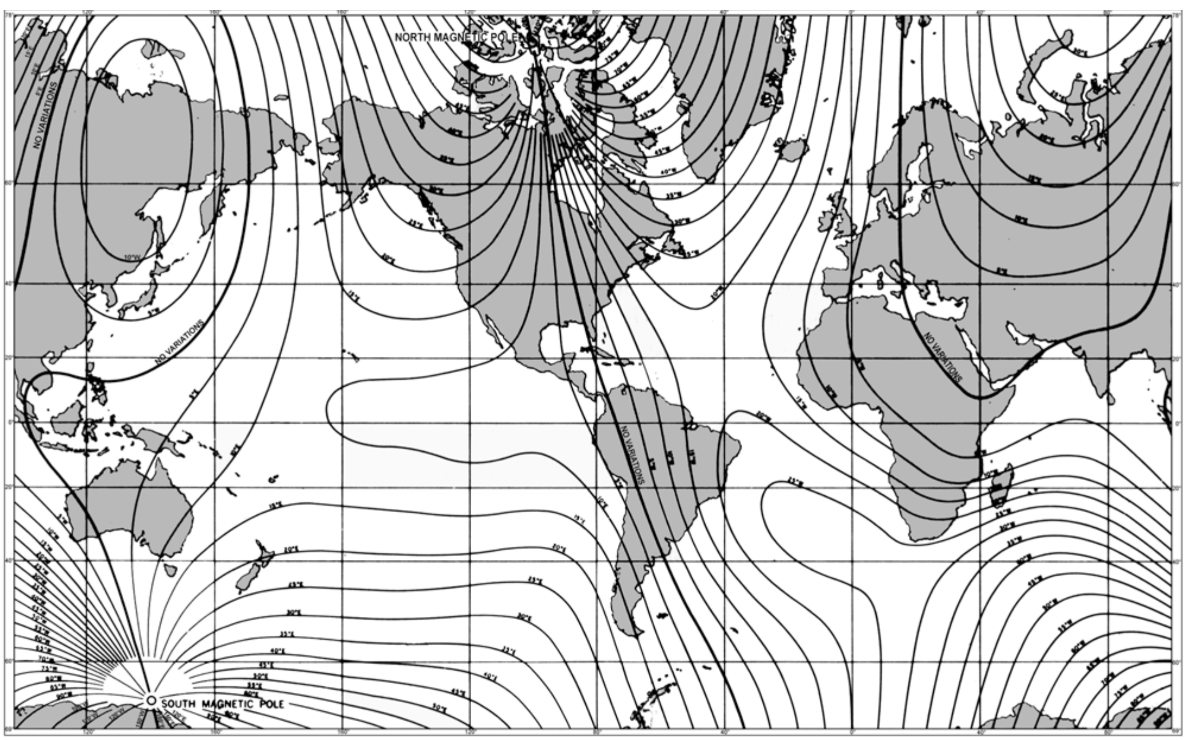


Figure 602c. Magnetic variation for the world.

apparent permanent magnetic condition is subject to change from deperming, shocks, welding, and vibration. The ship's induced magnetism will vary with the Earth's magnetic field strength and with the alignment of the ship in that field.

604. Magnetic Adjustment

A narrow rod of soft iron, placed parallel to the Earth's horizontal magnetic field, H, will have a north pole induced in the end toward the north geographic pole and a south pole induced in the end toward the south geographic pole. This same rod in a horizontal plane, but at right angles to the horizontal Earth's field, would have no magnetism induced in it, because its alignment in the magnetic field precludes linear magnetization, if the rod is of negligible cross section. Should the rod be aligned in some horizontal direction between those headings which create maximum and zero induction, it would be induced by an amount which is a function of the angle of alignment. However, if a similar rod is placed in a vertical position in northern latitudes so as to be aligned with the vertical Earth's field Z, it will have a south pole induced at the upper end and a north pole induced at the lower end. These polarities of vertical induced magnetization will be reversed in southern latitudes.

The amount of horizontal or vertical induction in such rods, or in ships whose construction is equivalent to combinations of such rods, will vary with the intensity of H and Z, heading, and heel of the ship.

The magnetic compass must be corrected for the vessel's permanent and induced magnetism so that its operation approximates that of a completely nonmagnetic vessel. Ship's magnetic conditions create magnetic compass deviations and sectors of sluggishness and unsteadiness. **Deviation** is defined as deflection right or left of the magnetic meridian caused by magnetic properties of the vessel. Adjusting the compass consists of arranging magnetic and soft iron **correctors** near the compass so that their effects are equal and opposite to the effects of the magnetic material in the ship.

The total permanent magnetic field effect at the compass may be broken into three components, mutually 90° to each other, as shown in Figure 604a.

The vertical permanent component tilts the compass card, and, when the ship rolls or pitches, causes oscillating deflections of the card. Oscillation effects which accompany roll are maximum on north and south compass headings, and those which accompany pitch are maximum on east and west compass headings.

The horizontal B and C components of permanent magnetism cause varying deviations of the compass as the ship swings in heading on an even keel. Plotting these deviations against compass heading yields the sine and cosine curves shown in Figure 604b. These deviation curves are called semicircular curves because they reverse direction by 180°.

A vector analysis is helpful in determining deviations

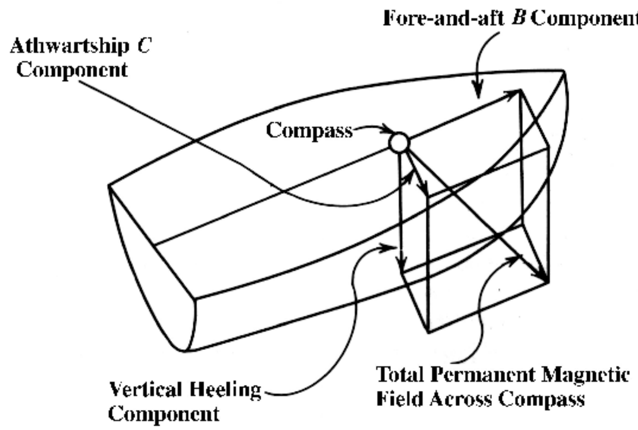


Figure 604a. Components of permanent magnetic field.

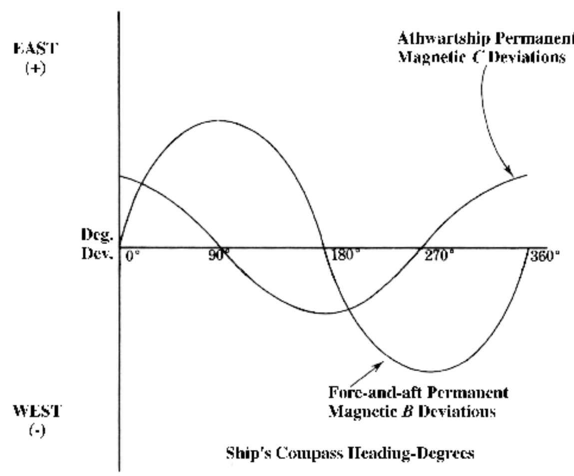


Figure 604b. Permanent magnetic deviation effects.

or the strength of deviating fields. For example, a ship as shown in Figure 604c on an east magnetic heading will subject its compass to a combination of magnetic effects; namely, the Earth's horizontal field H, and the deviating field B, at right angles to the field H. The compass needle will align itself in the resultant field which is represented by the vector sum of H and B, as shown. A similar analysis will reveal that the resulting directive force on the compass would be maximum on a north heading and minimum on a south heading because the deviations for both conditions are zero. The magnitude of the deviation caused by the permanent B magnetic field will vary with different values of H; hence, deviations resulting from permanent magnetic fields will vary with the magnetic latitude of the ship.

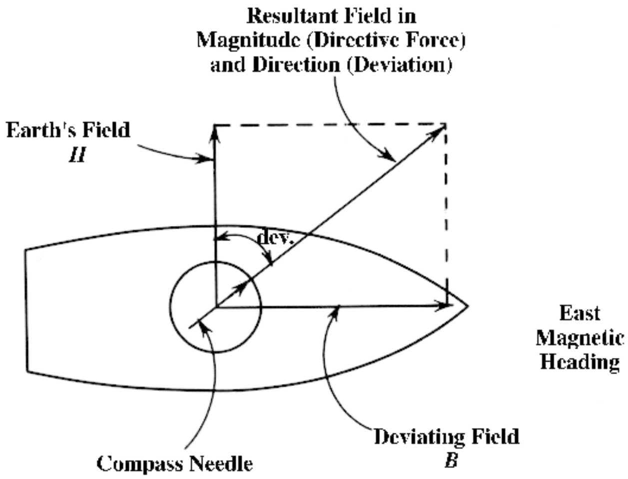


Figure 604c. General force diagram.

605. Effects of Induced Magnetism

Induced magnetism varies with the strength of the surrounding field, the mass of metal, and the alignment of the metal in the field. Since the intensity of the Earth's magnetic field varies over the Earth's surface, the induced magnetism in a ship will vary with latitude, heading, and heeling angle.

With the ship on an even keel, the resultant vertical induced magnetism, if not directed through the compass itself, will create deviations which plot as a semicircular deviation curve. This is true because the vertical induction changes magnitude and polarity only with magnetic latitude and heel, and not with heading of the ship. Therefore, as long as the ship is in the same magnetic latitude, its vertical induced pole swinging about the compass will produce the same effect on the compass as a permanent pole swinging about the compass.

The Earth's field induction in certain other unsymmetrical arrangements of horizontal soft iron create a constant A deviation curve. In addition to this magnetic A error, there are constant A deviations resulting from: (1) physical misalignments of the compass, pelorus, or gyro; (2) errors in calculating the Sun's azimuth, observing time, or taking bearings.

The nature, magnitude, and polarity of these induced effects are dependent upon the disposition of metal, the symmetry or asymmetry of the ship, the location of the binnacle, the strength of the Earth's magnetic field, and the angle of dip.

Certain heeling errors, in addition to those resulting from permanent magnetism, are created by the presence of both horizontal and vertical soft iron which experience changing induction as the ship rolls in the Earth's magnetic field. This part of the heeling error will change in magnitude proportional to changes of magnetic latitude of the ship. Oscillation effects associated with rolling are maximum on north and south headings, just as with the permanent magnetic heeling errors.

606. Adjustments and Correctors

Since some magnetic effects are functions of the vessel's magnetic latitude and others are not, each individual effect should be corrected independently. Furthermore, to make the corrections, we use (1) permanent magnet correctors to compensate for permanent magnetic fields at the compass, and (2) soft iron correctors to compensate for induced magnetism. The compass binnacle provides support for both the compass and its correctors. Typical large ship binnacles hold the following correctors:

1. Vertical permanent **heeling magnet** in the central vertical tube
2. Fore-and-aft **B permanent magnets** in their trays
3. Athwartship **C permanent magnets** in their trays
4. Vertical soft iron **Flinders bar** in its external tube
5. Soft iron **quadrantal spheres**

The heeling magnet is the only corrector which corrects for both permanent and induced effects. Therefore, it may need to be adjusted for changes in latitude if a vessel permanently changes its normal operating area. However, any movement of the heeling magnet will require readjustment of other correctors.

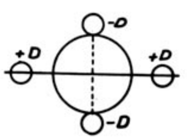
Fairly sophisticated magnetic compasses used on smaller commercial craft, larger yachts, and fishing vessels, may not have soft iron correctors or B and C permanent magnets. These compasses are adjusted by rotating magnets located inside the base of the unit, adjustable by small screws on the outside. A non-magnetic screwdriver is necessary to adjust these compasses. Occasionally one may find a permanent magnet corrector mounted near the compass, placed during the initial installation so as to remove a large, constant deviation before final adjustments are made. Normally, this remains in place for the life of the vessel.

Figure 606 summarizes all the various magnetic conditions in a ship, the types of deviation curves they create, the correctors for each effect, and headings on which each corrector is adjusted. When adjusting the compass, always apply the correctors symmetrically and as far away from the compass as possible. This preserves the uniformity of magnetic fields about the compass needle.

Occasionally, the permanent magnetic effects at the location of the compass are so large that they overcome the Earth's directive force, H. This condition will not only create sluggish and unsteady sectors, but may even freeze the compass to one reading or to one quadrant, regardless of the heading of the ship. Should the compass become so frozen, the polarity of the magnetism which must be attracting the compass needles is indicated; hence, correction may be effected simply by the application of permanent magnet

Coefficient	Type deviation curve	Compass headings of maximum deviation	Causes of such errors	Correctors for such errors	Magnetic or compass headings on which to apply correctors
A	Constant.	Same on all.	Human-error in calculations _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ Physical-compass, gyro, pelorus alignment _ _ _ _ _ _ _ _ _ Magnetic-unsymmetrical arrangements of horiz. soft iron.	Check methods and calculations Check alignments Rare arrangement of soft iron rods.	Any.
B	Semicircular $\sin\phi$.	090° 270°	Fore-and-aft component of permanent magnetic field _ _ _ _ _ Induced magnetism in unsymmetrical vertical iron forward or aft of compass.	Fore-and-aft *B* magnets Flinders bar (forward or aft)	090° or 270°.
C	Semicircular $\cos\phi$.	000° 180°	Athwartship component of permanent magnetic field- - - - - - - Induced magnetism in unsymmetrical vertical iron port or starboard of compass.	Athwartship *C* magnets Flinders bar (port or starboard)	000° or 180°.
D	Quadrantral $\sin 2\phi$.	045° 135° 225° 315°	Induced magnetism in all symmetrical arrangements of horizontal soft iron.	Spheres on appropriate axis. (athwartship for +D) (fore and aft for -D) *See sketch a*	045°, 135°, 225°, or 315°.
E	Quadrantral $\cos 2\phi$.	000° 090° 180° 270°	Induced magnetism in all unsymmetrical arrangements of horizontal soft iron.	Spheres on appropriate axis. (port fwd.-stb'd for +E) (stb'd fwd.-port aft for -E) *See sketch b*	000°, 090°, 180°, or 270°.
Heeling	Oscillations with roll or pitch. Deviations with constant list.	000° } roll 180° 090° } pitch 270°	Change in the horizontal component of the induced or permanent magnetic fields at the compass due to rolling or pitching of the ship.	Heeling magnet (must be readjusted for latitude changes).	090° or 270° with dip needle. 000° or 180° while rolling.

Deviation = $A + B\sin\phi + C\cos\phi + D\sin 2\phi + E\cos 2\phi$ (ϕ = compass heading)



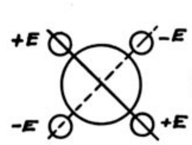


Figure 606. Summary of compass errors and adjustments.

correctors to neutralize this magnetism. Whenever such adjustments are made, the ship should be steered on a heading such that the unfreezing of the compass needles will be immediately evident. For example, a ship whose compass is frozen to a north reading would require fore-and-aft B corrector magnets with the positive ends forward in order to neutralize the existing negative pole which attracted the compass. If made on an east heading, such an adjustment would be evident when the compass card was freed to indicate an east heading.

607. Reasons for Correcting Compass

There are several reasons for correcting the errors of a magnetic compass, even if it is not the primary directional reference:

1. It is easier to use a magnetic compass if the deviations are small.
2. Even known and fully compensated deviation introduces error because the compass operates sluggishly and unsteadily when deviation is present.
3. Even though the deviations are compensated for, they will be subject to appreciable change as a function of heel and magnetic latitude.

Theoretically, it doesn't matter what the compass error is as long as it is known. But a properly adjusted magnetic compass is more accurate in all sea conditions, easier to steer by, and less subject to transient deviations which could result in deviations from the ship's chosen course.

Therefore, if a magnetic compass is installed and meant to be relied upon, it behooves the navigator to attend carefully to its adjustment. Doing so is known as "swinging ship."

608. Adjustment Check-off List

While a professional compass adjuster will be able to obtain the smallest possible error curve in the shortest time, many ship's navigators adjust the compass themselves with satisfactory results. Whether or not a "perfect" adjustment is necessary depends on the degree to which the magnetic compass will be relied upon in day-to-day navigation. If the magnetic compass is only used as a backup compass, removal of every last possible degree of error may not be worthwhile. If the magnetic compass is the only steering reference aboard, as is the case with many smaller commercial craft and fishing vessels, it should be adjusted as accurately as possible.

Prior to getting underway to swing ship, the navigator must ensure that the process will proceed as expeditiously as possible by preparing the vessel and compass. The following tests and adjustment can be done at dockside, assuming that the compass has been installed and maintained properly. Initial installation and adjustment should be done by a professional compass technician during commissioning.

1. Check for bubbles in the compass bowl. Fluid may be added through the filling plug if necessary. Large bubbles indicate serious leakage, indicating that the compass should be taken to a professional compass repair facility for new gaskets.

2. Check for free movement of gimbals. Clean any dust or dirt from gimbal bearings and lubricate them as recommended by the maker.

3. Check for magnetization of the quadrantal spheres by moving them close to the compass and rotating them. If the compass needle moves more than 2 degrees, the spheres must be annealed to remove their magnetism. Annealing consists of heating the spheres to a dull red color in a non-magnetic area and allowing them to cool slowly to ambient temperature.

4. Check for magnetization of the Flinders bar by inverting it, preferably with the ship on an E/W heading. If the compass needle moves more than 2 degrees the Flinders bar must be annealed.

5. Synchronize the gyro repeaters with the master gyro so courses can be steered accurately.

6. Assemble past documentation relating to the compass and its adjustment. Have the ship's degaussing folder ready.

7. Ensure that every possible metallic object is stowed for sea. All guns, doors, booms, and other movable gear should be in its normal seagoing position. All gear normally turned on such as radios, radars, loudspeakers, etc. should be on while swinging ship.

8. Have the International Code flags Oscar-Quebec ready to fly.

Once underway to swing ship, the following procedures will expedite the process. Choose the best helmsman aboard and instruct him to steer each course as steadily and precisely as possible. Each course should be steered steadily for at least two minutes before any adjustments are made to remove Gaussin error. Be sure the gyro is set for the mean speed and latitude of the ship.

The navigator (or compass adjuster if one is employed) should have a pelorus and a table of azimuths prepared for checking the gyro, but the gyrocompass will be the primary steering reference. Normally the adjuster will request courses and move the magnets as he feels necessary, a process much more an art than a science. If a professional adjuster is not available, use the following sequence:

1. If there is a sea running, steer course 000° and adjust the heeling magnet to decrease oscillations to a minimum.

2. Come to course 090°. When steady on course 090°, for at least two minutes, insert, remove, or move fore-and-aft B magnets to remove ALL deviation.

3. Come to a heading of 180°. Insert, remove, or move athwartships C magnets to remove ALL deviation.

4. Come to 270° and move the B magnets to remove one half of the deviation.

5. Come to 000° and move the C magnets to remove one half of the deviation.

6. Come to 045° (or any intercardinal heading) and move the quadrantal spheres toward or away from the compass to minimize any error.

7. Come to 135° (or any intercardinal heading 90° from the previous course) and move the spheres in or out to remove one half of the observed error.

8. Steer the ship in turn on each cardinal and intercardinal heading around the compass, recording the error at each heading called for on the deviation card. If plotted, the errors should plot roughly as a sine curve about the 0° line.

If necessary, repeat steps 1-8. There is no average error, for each ship is different, but generally speaking, errors of more than a few degrees, or errors which seriously distort the sine curve, indicate a magnetic problem which should be addressed.

Once the compass has been swung, tighten all fittings and carefully record the placement of all magnets and correctors. Finally, swing for residual degaussed deviations with the degaussing circuits energized and record the deviations on the deviation card. Post this card near the chart table for ready reference by the navigation team.

Once properly adjusted, the magnetic compass deviations should remain constant until there is some change in the magnetic condition of the vessel resulting from magnetic treatment, shock, vibration, repair, or structural changes. Transient deviations are discussed below.

609. Sources of Transient Error

The ship must be in seagoing trim and condition to properly compensate a magnetic compass. Any movement of large metal objects or the energizing of any electrical equipment in the vicinity of the compass can cause errors. If in doubt about the effect of any such changes, temporarily move the gear or cycle power to the equipment while observing the compass card while on a steady heading. Preferably this should be done on two different headings 90° apart, since the compass might be affected on one heading and not on another.

Some magnetic items which cause deviations if placed too close to the compass are as follows:

1. Movable guns or weapon loads
2. Magnetic cargo
3. Hoisting booms
4. Cable reels
5. Metal doors in wheelhouse
6. Chart table drawers
7. Movable gyro repeater
8. Windows and ports
9. Signal pistols racked near compass
10. Sound powered telephones
11. Magnetic wheel or rudder mechanism
12. Knives or tools near binnacle
13. Watches, wrist bands, spectacle frames
14. Hat grommets, belt buckles, metal pencils
15. Heating of smoke stack or exhaust pipes
16. Landing craft

Some electrical items which cause variable deviations if placed too close to the compass are:

1. Electric motors
2. Magnetic controllers
3. Gyro repeaters
4. Nonmarried conductors
5. Loudspeakers
6. Electric indicators
7. Electric welding
8. Large power circuits
9. Searchlights or flashlights
10. Electrical control panels or switches
11. Telephone headsets
12. Windshield wipers
13. Rudder position indicators, solenoid type
14. Minesweeping power circuits
15. Engine order telegraphs
16. Radar equipment
17. Magnetically controlled switches
18. Radio transmitters
19. Radio receivers
20. Voltage regulators

Another source of transient deviation is the **retentive error**. This error results from the tendency of a ship's structure to retain induced magnetic effects for short periods of time. For example, a ship traveling north for several days, especially if pounding in heavy seas, will tend to retain some fore-and-aft magnetism induced under these conditions. Although this effect is transient, it may cause slightly incorrect observations or adjustments. This same type of error occurs when ships are docked on one heading for long periods of time. A short shakedown, with the ship on other headings, will tend to remove such errors. A similar sort of residual magnetism is left in many ships if the degaussing circuits are not secured by the correct reversal sequence.

A source of transient deviation somewhat shorter in duration than retentive error is known as **Gaussin error**. This error is caused by eddy currents set up by a changing number of magnetic lines of force through soft iron as the ship changes heading. Due to these eddy currents, the induced magnetism on a given heading does not arrive at its normal value until about 2 minutes after changing course.

Deperming and other magnetic treatment will change the magnetic condition of the vessel and therefore require compass readjustment. The decaying effects of deperming can vary. Therefore, it is best to delay readjustment for several days after such treatment. Since the magnetic fields used for such treatments are sometimes rather large at the compass locations, the Flinders bar, compass, and related equipment should be removed from the ship during these operations.

DEGAUSSING (MAGNETIC SILENCING) COMPENSATION

610. Degaussing

A steel vessel has a certain amount of **permanent magnetism** in its "hard" iron and **induced magnetism** in its "soft" iron. Whenever two or more magnetic fields occupy the same space, the total field is the vector sum of the individual fields. Thus, near the magnetic field of a vessel, the total field is the combined total of the Earth's field and the vessel's field. Not only does the Earth's field affect the vessel's, the vessel's field affects the Earth's field in its immediate vicinity.

Since certain types of explosive mines are triggered by the magnetic influence of a vessel passing near them, a vessel may use a degaussing system to minimize its magnetic field. One method of doing this is to neutralize each component of the field with an opposite field produced by electrical cables coiled around the vessel. These cables, when energized, counteract the permanent magnetism of the vessel, rendering it magnetically neutral. This has severe effects on magnetic compasses.

A unit sometimes used for measuring the strength of a magnetic field is the **gauss**. Reducing of the strength of a magnetic field decreases the number of gauss in that field. Hence, the process is called **degaussing**.

The magnetic field of the vessel is completely altered when the degaussing coils are energized, introducing large deviations in the magnetic compass. This deviation can be removed by introducing an equal and opposite force with energized coils near the compass. This is called **compass compensation**. When there is a possibility of confusion with compass adjustment to neutralize the effects of the natural magnetism of the vessel, the expression **degaussing compensation** is used. Since compensation may not be perfect, a small amount of deviation due to degaussing may remain on certain headings. This is the reason for swinging the ship with degaussing off and again with it on, and why there are two separate columns in the deviation table.

611. A Vessel's Magnetic Signature

A simplified diagram of the distortion of the Earth's magnetic field in the vicinity of a steel vessel is shown in Figure 611a. The field strength is directly proportional to the line spacing density. If a vessel passes over a device for detecting and recording the strength of the magnetic field, a certain pattern is traced. Figure 611b shows this pattern. Since the magnetic field of each vessel is different, each produces a distinctive trace. This distinctive trace is referred to as the vessel's **magnetic signature**.

Several **degaussing stations** have been established in major ports to determine magnetic signatures and recommend the currents needed in the various degaussing coils to render it magnetically neutral. Since a vessel's induced magnetism varies with heading and magnetic latitude, the current settings of the coils may sometimes need to be changed. A **degaussing folder** is provided to the vessel to indicate these changes and to document other pertinent information.

A vessel's permanent magnetism changes somewhat with time and the magnetic history of the vessel. Therefore, the data in the degaussing folder should be checked periodically at the magnetic station.

612. Degaussing Coils

For degaussing purposes, the total field of the vessel is divided into three components: (1) vertical, (2) horizontal fore-and-aft, and (3) horizontal athwartships. The positive (+) directions are considered downward, forward, and to port, respectively. These are the normal directions for a vessel headed north or east in north latitude.

Each component is opposed by a separate degaussing field just strong enough to neutralize it. Ideally, when this has been done, the Earth's field passes through the vessel smoothly and without distortion. The opposing degaussing fields are produced by direct current flowing in coils of wire. Each of the degaussing coils is placed so that the field it produces is directed to oppose one component of the ship's field.

The number of coils installed depends upon the magnetic characteristics of the vessel, and the degree of safety desired. The ship's permanent and induced magnetism may be neutralized separately so that control of induced magnetism can be varied as heading and latitude change, without disturbing the fields opposing the vessel's permanent field. The principal coils employed are the following:

Main (M) coil. The M coil is horizontal and completely encircles the vessel, usually at or near the waterline. Its function is to oppose the vertical component of the vessel's combined permanent and induced fields. Generally the induced field predominates. Current in the M-coil is varied or reversed according to the change of the induced component of the vertical field with latitude.

Forecastle (F) and **quarterdeck (Q) coils**. The F and Q coils are placed horizontally just below the forward and after thirds (or quarters), respectively, of the weather deck. These coils, in which current can be individually adjusted, remove much of the fore-and-aft component of the ship's permanent and induced fields. More commonly, the combined F and Q coils consist of two parts; one part the FP and QP coils, to take care of the permanent fore-and-aft field, and the other part, the FI and QI coils, to neutralize the induced fore-and-aft field. Generally, the forward and after coils of each type are connected in series, forming a split-coil installation and designated FP-QP coils and FI-QI coils. Current in the FP-QP coils is generally constant, but in the FI-QI coils is varied according to the heading and magnetic latitude of the vessel. In split-coil installations, the coil designations are often called simply the P-coil and I-coil.

Longitudinal (L) coil. Better control of the fore-and-aft components, but at greater installation expense, is provided by placing a series of vertical, athwartship coils along the length of the ship. It is the field, not the coils, which is longitudinal. Current in an L coil is varied as with the FI-QI coils. It is maximum on north and south headings, and zero on east and west headings.

Athwartship (A) coil. The A coil is in a vertical fore-and-aft plane, thus producing a horizontal athwartship field which neutralizes the athwartship component of the vessel's field. In most vessels, this component of the permanent field is small and can be ignored. Since the A-coil neutralizes the induced field, primarily, the current is changed with magnetic latitude and with heading, maximum on east or west headings, and zero on north or south headings.

The strength and direction of the current in each coil is indicated and adjusted at a control panel accessible to the navigator. Current may be controlled directly by rheostats at the control panel or remotely by push buttons which operate rheostats in the engine room.

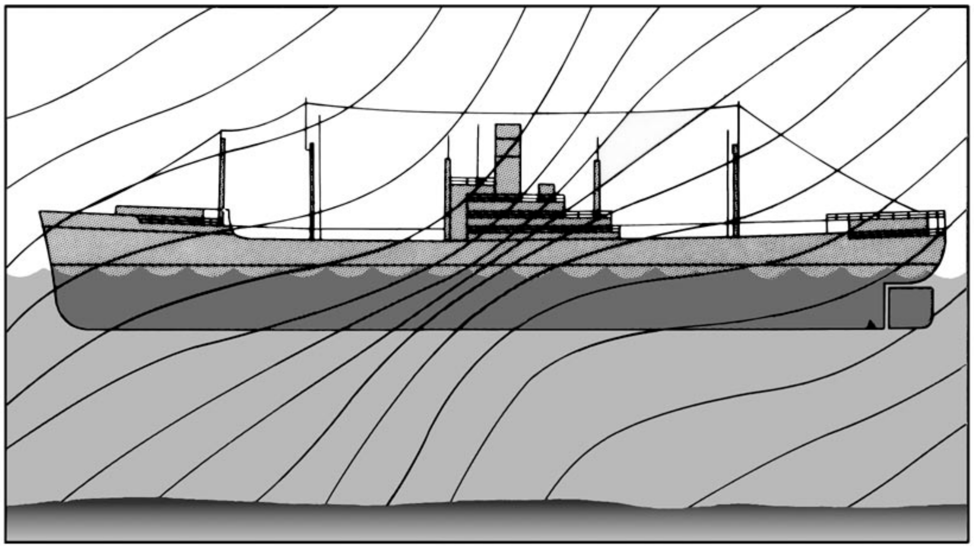


Figure 611a. Simplified diagram of distortion of Earth's magnetic field in the vicinity of a steel vessel.

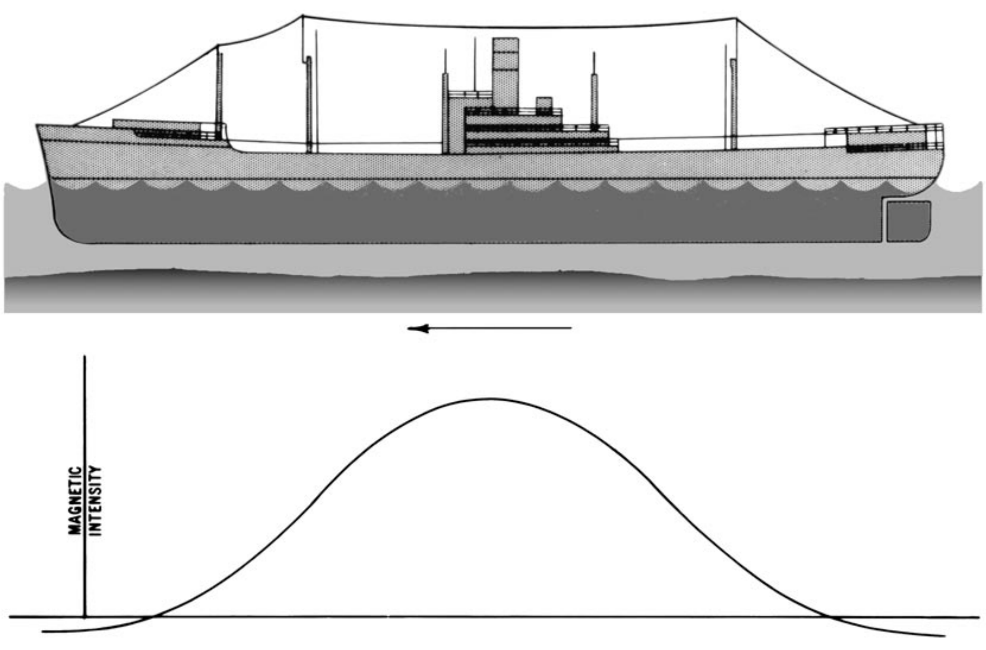


Figure 611b. A simplified signature of a vessel of Figure 611a.

Appropriate values of the current in each coil are determined at a degaussing station, where the various currents are adjusted until the vessel's magnetic signature is made as flat as possible. Recommended current values and directions for all headings and magnetic latitudes are set forth in the vessel's degaussing folder. This document is normally kept by the navigator, who must see that the recommended settings are maintained whenever the degaussing system is energized.

613. Securing The Degaussing System

Unless the degaussing system is properly secured, residual magnetism may remain in the vessel. During degaussing compensation and at other times, as recommended in the degaussing folder, the "reversal" method is used. The steps in the reversal process are as follows:

1. Start with maximum degaussing current used since the system was last energized.
2. Decrease current to zero and increase it in the opposite direction to the same value as in step 1.
3. Decrease the current to zero and increase it to three-fourths maximum value in the original direction.
4. Decrease the current to zero and increase it to one-half maximum value in the opposite direction.
5. Decrease the current to zero and increase it to one-fourth maximum value in the original direction.
6. Decrease the current to zero and increase it to one-eighth maximum value in the opposite direction.
7. Decrease the current to zero and open switch.

614. Magnetic Treatment Of Vessels

In some instances, degaussing can be made more effective by changing the magnetic characteristics of the vessel by a process known as **deperming**. Heavy cables are wound around the vessel in an athwartship direction, forming vertical loops around the longitudinal axis of the vessel. The loops are run beneath the keel, up the sides, and over the top of the weather deck at closely spaced equal intervals along the entire length of the vessel. Predetermined values of direct current are then passed through the coils. When the desired magnetic characteristics have been acquired, the cables are removed.

A vessel which does not have degaussing coils, or which has a degaussing system which is inoperative, can be given some temporary protection by a process known as **flashing**. A horizontal coil is placed around the outside of the vessel and energized with large predetermined values of direct current. When the vessel has acquired a vertical field of permanent magnetism of the correct magnitude and polarity to reduce to a minimum the resultant field below the vessel for the particular magnetic latitude involved, the cable is removed. This type protection is not as satisfactory as that provided by degaussing coils because it is not adjustable for various headings and magnetic latitudes, and also because the vessel's magnetism slowly readjusts following treatment.

During magnetic treatment all magnetic compasses and Flinders bars should be removed from the ship. Permanent adjusting magnets and quadrantal correctors are not materially affected, and need not be removed. If it is impractical to remove a compass, the cables used for magnetic treatment should be kept as far as practical from it.

615. Degaussing Effects

The degaussing of ships for protection against magnetic mines creates additional effects upon magnetic compasses, which are somewhat different from the permanent and induced magnetic effects. The degaussing effects are electromagnetic, and depend on:

1. Number and type of degaussing coils installed.
2. Magnetic strength and polarity of the degaussing coils.
3. Relative location of the different degaussing coils with respect to the binnacle.
4. Presence of masses of steel, which would tend to concentrate or distort magnetic fields in the vicinity of the binnacle.
5. The fact that degaussing coils are operated intermittently, with variable current values, and with different polarities, as dictated by necessary degaussing conditions.

616. Degaussing Compensation

The magnetic fields created by the degaussing coils would render the vessel's magnetic compasses useless unless compensated. This is accomplished by subjecting the compass to compensating fields along three mutually perpendicular axes. These fields are provided by small **compensating coils** adjacent to the compass. In nearly all installations, one of these coils, the **heeling coil**, is horizontal and on the same plane as the compass card, providing a vertical compensating field. Current in the heeling coil is adjusted until the vertical component of the total degaussing field is neutralized. The other compensating coils provide horizontal fields perpendicular to each other. Current is varied in these coils until their resultant field is equal and opposite to the horizontal component of the degaussing field. In early installations, these horizontal fields were directed fore-and-aft and athwartships by placing the coils around the Flinders bar and the quadrantal spheres. Compactness and other advantages are gained by placing the coils on perpendicular axes extending 045°-225° and 315°-135° relative to the heading. A frequently used compensating installation, called the **type K**, is shown in Figure 616. It consists of a heeling coil extending completely around the top of the binnacle, four **intercardinal coils**, and three control boxes. The intercardinal coils are named for their positions relative to the compass when the vessel is on a heading of north, and also for the compass headings on which the current in the coils is adjusted to the correct amount for compensation. The NE-SW coils operate together as one set, and the NW-SE coils operate as another. One control box is provided for each set, and one for the heeling coil.

The compass compensating coils are connected to the power supply of the degaussing coils, and the currents passing through the compensating coils are adjusted by series resistances so that the compensating field is equal to the degaussing field. Thus, a change in the degaussing currents is accompanied by a proportional change in the compensating currents. Each coil has a separate winding for each degaussing circuit it compensates.

Degaussing compensation is carried out while the vessel is moored at the shipyard where the degaussing coils are installed. This process is usually carried out by civilian professionals, using the following procedure:

Step 1. The compass is removed from its binnacle and a dip needle is installed in its place. The M coil and heeling coil are then energized, and the current in the heeling coil is adjusted until the dip needle indicates the correct value for the magnetic latitude of the vessel. The system is then secured by the reversing process.

Step 2. The compass is replaced in the binnacle. With auxiliary magnets, the compass card is deflected until the compass magnets are parallel to one of the compensating coils or set of coils used to produce a horizontal field. The compass magnets are then perpendicular to the field produced by that coil. One of the degaussing circuits producing a horizontal field, and its compensating winding, are then energized, and the current in the compensating winding is adjusted until the compass reading returns to the value it had before the degaussing circuit was energized. The system is then secured by the reversing process. The process is repeated with each additional circuit used to create a horizontal field. The auxiliary magnets are then removed.

Step 3. The auxiliary magnets are placed so that the compass magnets are parallel to the other compensating coils or set of coils used to produce a horizontal field. The procedure of step 2 is then repeated for each circuit producing a horizontal field.

When the vessel gets under way, it proceeds to a suitable maneuvering area. The vessel is then steered so that the compass magnets are parallel first to one compensating coil or set of coils, and then the other. Any needed adjustment is made in the compensating circuits to reduce the error to a minimum. The vessel is then swung for residual deviation, first with degaussing off and then with degaussing on, and the correct current settings determined for each heading at the magnetic latitude of the vessel. From the values thus obtained, the “DG OFF” and “DG ON” columns of the deviation table are filled in. If the results indicate satisfactory compensation, a record is made of the degaussing coil settings and the resistance, voltages, and currents in the compensating coil circuits. The control boxes are then secured.

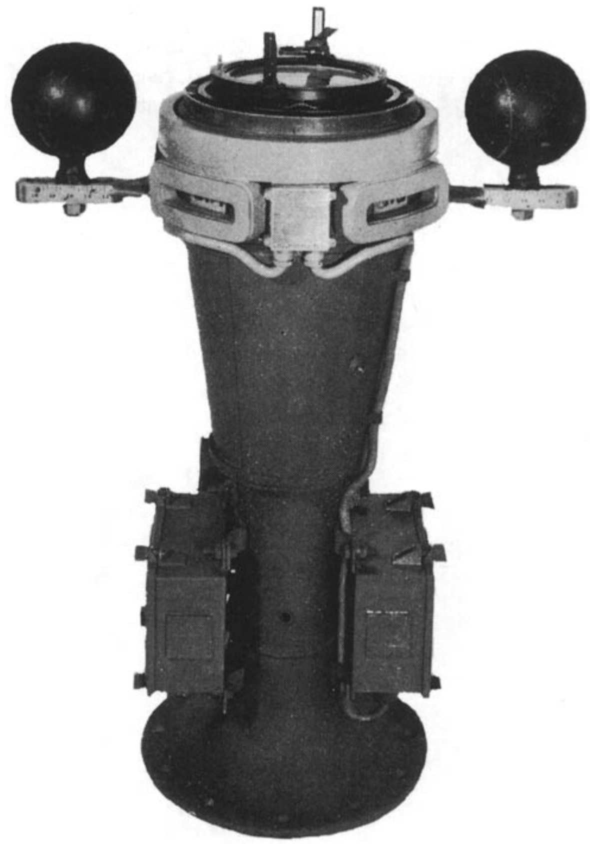


Figure 616. Type K degaussing compensation installation.

Under normal operating conditions, the settings do not need to be changed unless changes are made in the degaussing system, or unless an alteration is made in the length of the Flinders bar or the setting of the quadrantal spheres. However, it is possible for a ground to occur in the coils or control box if the circuits are not adequately protected from moisture. If this occurs, it should be reflected by a change in deviation with degaussing on, or by a decreased installation resistance. Under these conditions, compensation should be done again. If the compass will be used with degaussing on before the ship can be returned to a shipyard where the compensation can be made by experienced personnel, the compensation should be made at sea on the actual headings needed, rather than by deflection of the compass needles by magnets. More complete information related to this process is given in the degaussing folder.

If a vessel has been given magnetic treatment, its magnetic properties have changed, necessitating readjustment of each magnetic compass. This is best delayed for several days to permit the magnetic characteristics of the vessel to settle. If compensation cannot be delayed, the vessel should be swung again for residual deviation after a few days. Degaussing compensation should not be made until after compass adjustment has been completed.

GYROCOMPASSES

617. Principles of the Gyroscope

A gyroscope consists of a spinning wheel or rotor contained within gimbals which permit movement about three mutually perpendicular axes, known as the **horizontal axis**, the **vertical axis**, and the **spin axis**. When spun rapidly, assuming that friction is not considered, the gyroscope develops **gyroscopic inertia**, tending to remain spinning in the same plane indefinitely. The amount of gyroscopic inertia depends on the angular velocity, mass, and radius of the wheel or rotor.

When a force is applied to change alignment of the spin axis of a gyroscope, the resultant motion is perpendicular to the direction of the force. This tendency is known as precession. A force applied to the center of gravity of the gyroscope will move the entire system in the direction of the force. Only a force that tends to change the axis of rotation produces precession.

If a gyroscope is placed at the equator with its spin axis pointing east-west, as the earth turns on its axis, gyroscopic inertia will tend to keep the plane of rotation constant. To the observer, it is the gyroscope which is seen to rotate, not the earth. This effect is called the horizontal earth rate, and is maximum at the equator and zero at the poles. At points between, it is equal to the cosine of the latitude.

If the gyro is placed at a geographic pole with its spin axis horizontal, it will appear to rotate about its vertical axis. This is the vertical earth rate. At all points between the equator and the poles, the gyro appears to turn partly about its horizontal and partly about its vertical axis, being affected by both horizontal and vertical earth rates. In order to visualize these effects, remember that the gyro, at whatever latitude it is placed, is remaining aligned in space while the earth moves beneath it.

618. Gyrocompass Operation

The gyrocompass depends upon four natural phenomena: gyroscopic inertia, precession, earth's rotation, and gravity. To make a gyroscope into a gyrocompass, the wheel or rotor is mounted in a sphere, called the gyrosphere, and the sphere is then supported in a vertical ring. The whole is mounted on a base called the phantom. The gyroscope in a gyrocompass can be pendulous or non-pendulous, according to design. The rotor may weigh as little as half a kilogram to over 25 kg.

To make it seek and maintain true north, three things are necessary. First, the gyro must be made to stay on the plane of the meridian. Second, it must be made to remain horizontal. Third, it must stay in this position once it reaches it regardless of what the vessel on which it is mounted does or where it goes on the earth. To make it seek the meridian, a weight is added to the bottom of the vertical ring, causing it to swing on its vertical axis, and thus seek to align itself horizontally. It will tend to oscillate, so a second weight is added to the side of the sphere in which the rotor is contained, which dampens the oscillations until the gyro stays on the meridian. With these two weights, the only possible position of equilibrium is on the meridian with its spin axis horizontal.

To make the gyro seek north, a system of reservoirs filled with mercury, known as mercury ballistics, is used to apply a force against the spin axis. The ballistics, usually four in number, are placed so that their centers of gravity exactly coincide with the CG of the gyroscope. Precession then causes the spin axis to trace an ellipse, one ellipse taking about 84 minutes to complete. (This is the period of oscillation of a pendulum with an arm equal to the radius of the earth.) To dampen this oscillation, the force is applied, not in the vertical plane, but slightly to the east of the vertical plane. This causes the spin axis to trace a spiral instead of an ellipse and eventually settle on the meridian pointing north.

619. Gyrocompass Errors

The total of the all the combined errors of the gyrocompass is called **gyro error** and is expressed in degrees E or W, just like variation and deviation. But gyro error, unlike magnetic compass error, and being independent of Earth's magnetic field, will be constant in one direction; that is, an error of one degree east will apply to all bearings all around the compass.

The errors to which a gyrocompass is subject are speed error, latitude error, ballistic deflection error, ballistic damping error, quadrantal error, and gimballing error. Additional errors may be introduced by a malfunction or incorrect alignment with the centerline of the vessel.

Speed error is caused by the fact that a gyrocompass only moves directly east or west when it is stationary (on the rotating earth) or placed on a vessel moving exactly east or west. Any movement to the north or south will cause the compass to trace a path which is actually a function of the speed of advance and the amount of northerly or southerly heading. This causes the compass to tend to settle a bit off true north. This error is westerly if the vessel's course is northerly, and easterly if the course is southerly. Its magnitude depends on the vessel's speed, course, and latitude. This error can be corrected internally by means of a cosine cam mounted on the underside of the azimuth gear, which removes most of the error. Any remaining error is minor in amount and can be disregarded.

Tangent latitude error is a property only of gyros with mercury ballistics, and is easterly in north latitudes and westerly in south latitudes. This error is also corrected internally, by offsetting the lubber's line or with a small movable weight attached to the casing.

Ballistic deflection error occurs when there is a marked change in the north-south component of the speed. East-west accelerations have no effect. A change of course or speed also results in speed error in the opposite direction, and the two tend to cancel each other if the compass is properly designed. This aspect of design involves slightly offsetting the ballistics according to the operating latitude, upon which the correction is dependent. As latitude changes, the error becomes apparent, but can be minimized by adjusting the offset.

Ballistic damping error is a temporary oscillation introduced by changes in course or speed. During a change in course or speed, the mercury in the ballistic is subjected to centrifugal and acceleration/deceleration forces. This causes a torquing of the spin axis and subsequent error in the compass reading. Slow changes do not introduce enough error to be a problem, but rapid changes will. This error is counteracted by changing the position of the ballistics so that the true vertical axis is centered, thus not subject to error, but only when certain rates of turn or acceleration are exceeded.

Quadrantal error has two causes. The first occurs if the center of gravity of the gyro is not exactly centered in the phantom. This causes the gyro to tend to swing along its heavy axis as the vessel rolls in the sea. It is minimized by adding weight so that the mass is the same in all directions from the center. Without a long axis of weight, there is no tendency to swing in one particular direction. The second source of quadrantal error is more difficult to eliminate. As a vessel rolls in the sea, the apparent vertical axis is displaced, first to one side and then the other. The vertical axis of the gyro tends to align itself with the apparent vertical. On northerly or southerly courses, and on easterly or westerly courses, the compass precesses equally to both sides and the resulting error is zero. On intercardinal courses, the N-S and E-W precessions are additive, and a persistent error is introduced, which changes direction in different quadrants. This error is corrected by use of a second gyroscope called a floating ballistic, which stabilizes the mercury ballistic as the vessel rolls, eliminating the error. Another method is to use two gyros for the directive element, which tend to precess in opposite directions, neutralizing the error.

Gimballing error is caused by taking readings from the compass card when it is tilted from the horizontal plane. It applies to the compass itself and to all repeaters. To minimize this error, the outer ring of the gimbal of each repeater should be installed in alignment with the fore-and-aft line of the vessel. Of course, the lubber's line must be exactly centered as well.

620. Using the Gyrocompass

Since a gyrocompass is not influenced by magnetism, it is not subject to variation or deviation. Any error is constant and equal around the horizon, and can often be reduced to less than one degree, thus effectively eliminating it altogether. Unlike a magnetic compass, it can output a signal to repeaters spaced around the vessel at critical positions.

But it also requires a constant source of stable electrical power, and if power is lost, it requires several hours to settle on the meridian again before it can be used. This period can be reduced by aligning the compass with the meridian before turning on the power.

The directive force of a gyrocompass depends on the amount of precession to which it is subject, which in turn is dependent on latitude. Thus the directive force is maximum at the equator and decreases to zero at the poles. Vessels operating in high latitudes must construct error curves based on latitudes because the errors at high latitudes eventually overcome the ability of the compass to correct them.

The gyrocompass is typically located below decks as close as possible to the center of roll, pitch and yaw of the ship, thus minimizing errors caused by the ship's motion. Repeaters are located at convenient places throughout the ship, such as at the helm for steering, on the bridge wings for taking bearings, in after steering for emergency steering, and other places. The output can also be used to drive course recorders, autopilot systems, plotters, fire control systems, and stabilized radars. The repeaters should be checked regularly against the master to ensure they are all in alignment. The repeaters on the bridge wing used for taking bearings will likely be equipped with removable bearing circles, azimuth circles, and telescopic alidades, which allow one to sight a distant object and see its exact gyrocompass bearing.

ELECTRONIC COMPASSES

621. New Direction Sensing Technologies

The magnetic compass has serious limitations, chiefly that of being unable to isolate the earth's magnetic field from all others close enough to influence it. It also indicates magnetic north, whereas the mariner is most interested in true north. Most of the work involved with compensating a traditional magnetic compass involves neutralizing magnetic influences other than the earth's, a complicated and inexact process often involving more art than science. Residual error is almost always present even after compensation. Degaussing complicates the situation immensely.

The electro-mechanical gyrocompass has been the standard steering and navigational compass since the early 20th century, and has provided several generations of mariners a stable and reliable heading and bearing reference. However, it too has limitations: It is a large, expensive, heavy, sensitive device that must be mounted according to rather strict limitations. It requires a stable and uninterrupted supply of electrical power; it is sensitive to shock, vibration, and environmental changes; and it needs several hours to settle after being turned on.

Fortunately, several new technologies have been developed which promise to greatly reduce or eliminate the limitations of both the mechanical gyroscope and traditional magnetic compasses. Sometimes referred to as "electronic compasses," the digital flux gate magnetic compass and the ring laser gyrocompass are two such devices. They have the following advantages:

1. Solid state electronics, no moving parts
2. Operation at very low power
3. Easy backup power from independent sources
4. Standardized digital output
5. Zero friction, drift, or wear
6. Compact, lightweight, and inexpensive
7. Rapid start-up and self-alignment
8. Low sensitivity to vibration, shock, and temperature changes
9. Self-correcting

Both types are being installed on many vessels as the primary directional reference, enabling the decommissioning of the traditional magnetic compasses and the avoidance of periodic compensation and maintenance.

622. The Flux Gate Compass

The most widely used sensor for digital compasses is the flux-gate magnetometer, developed around 1928. Initially it was used for detecting submarines, for geophysical prospecting, and airborne mapping of earth's magnetic fields.

The most common type, called the second harmonic device, incorporates two coils, a primary and a secondary, both wrapped around a single highly permeable ferromagnetic core. In the presence of an external magnetic field, the core's magnetic induction changes. A signal applied to the primary winding causes the core to oscillate. The secondary winding emits a signal that is induced through the core from the primary winding. This induced signal is affected by changes in the permeability of the core and appears as an amplitude variation in the output of the sensing coil. The signal is then demodulated with a phase-sensitive detector and filtered to retrieve the magnetic field value. After being converted to a standardized digital format, the data can be output to numerous remote devices, including steering compasses, bearing compasses, emergency steering stations, and autopilots.

Since the influence of a ship's inherent magnetism is inversely proportional to the square of the distance to the compass, it is logical that if the compass could be located at some distance from the ship, the influence of the ship's magnetic field could be greatly reduced. One advantage of the flux gate compass is that the sensor can be located remotely from the readout device, allowing it to be placed at a position as far as possible from the hull and its contents, such as high up on a mast, the ideal place on most vessels.

A further advantage is that the digital signal can be processed mathematically, and algorithms written which can correct for observed deviation once the deviation table has been determined. Further, the "table," in digital format, can be found by merely steering the vessel in a full circle. Algorithms then determine and apply corrections that effectively flatten the usual sine wave pattern of deviation. The theoretical result is zero observed compass deviation.

Should there be an index error (which has the effect of skewing the entire sine wave below or above the zero degree axis of the deviation curve) this can be corrected with an index correction applied to all the readings. This problem is largely confined to asymmetric installations such as aircraft carriers. Similarly, a correction for variation can be applied, and with GPS input (so the system knows where it is with respect to the isogonic map) the variation correction can be applied automatically, thus rendering the output in true degrees, corrected for both deviation and variation.

It is important to remember that a flux gate compass is still a magnetic compass, and that it will be influenced by large changes to the ship's magnetic field. Compensation should be accomplished after every such change. Fortunately, as noted, compensation involves merely steering the vessel in a circle in accordance with the manufacturer's recommendations.

Flux-gate compasses from different manufacturers share some similar operational modes. Most of them will have the following:

SET COURSE MODE: A course can be set and "remembered" by the system, which then provides the helmsman a graphic steering aid, enabling him to see if the ship's head is right or left of the set course, as if on a digital "highway." Normal compass operation continues in the background.

DISPLAY RESPONSE DAMPING: In this mode, a switch is used to change the rate of damping and update of the display in response to changes in sea condition and vessel speed.

AUTO-COMPENSATION: This mode is used to determine the deviation curve for the vessel as it steams in a complete circle. The system will then automatically compute correction factors to apply around the entire compass, resulting in zero deviation at any given heading. This should be done after every significant change in the magnetic signature of the ship, and within 24 hours of entering restricted waters.

CONTINUOUS AUTO-COMPENSATION: This mode, which should normally be turned OFF in restricted waters and ON at sea, runs the compensation algorithm each time the ship completes a 360 degree turn in two minutes. A warning will flash on the display in the OFF mode.

PRE-SET VARIATION: In effect an index correction, pre-set variation allows the application of magnetic variation to the heading, resulting in a true output (assuming the unit has been properly compensated and aligned). Since variation changes according to one's location on the earth, it must be changed periodically to agree with the charted variation unless GPS input is provided. The GPS position input is used in an algorithm which computes the variation for the area and automatically corrects the readout.

U.S. Naval policy approves the use of flux gate compasses and the decommissioning, but not removal, of the traditional binnacle mounted compass, which should be clearly marked as "Out of Commission" once an approved flux gate compass is properly installed and tested.

623. The Ring Laser Gyrocompass

The ring laser had its beginnings in England, where in the 1890's two scientists, Joseph Larmor and Sir Oliver Lodge (also one of the pioneers of radio), debated the possibility of measuring rotation by a ring interferometer. Some 15 years later, a French physicist, Georges Sagnac, fully described the phenomenon which today bears his name, the Sagnac Effect. This principle states that if two beams of light are sent in opposite directions around a "ring" or polyhedron and steered so as to meet and combine, a standing wave will form around the ring. If the wave is observed from any point, and that point is then moved along the perimeter of the ring, the wave form will change in direct relationship to the direction and velocity of movement.

It wasn't until 1963 that W. Macek of Sperry-Rand Corporation tested and refined the concept into a useful research device. Initially, mirrors were used to direct light around a square or rectangular pattern. But such mirrors must be made and adjusted to exceptionally close tolerances to allow useful output, and must operate in a vacuum for best effect. Multilayer dielectric mirrors with a reflectivity of 99.9999 percent were developed. The invention of laser light sources and fiber-optics has enabled the production of small, light, and dependable ring laser gyros. Mirror-based devices continue to be used in physics research.

The ring laser gyrocompass (RLG) operates by measuring laser-generated light waves traveling around a fiber-optic ring. A beam splitter divides a beam of light into two counter-rotating waves, which then travel around the fiber-optic ring in opposite directions. The beams are then recombined and sent to an output detector. In the absence of rotation, the path lengths will be the same and the beams will recombine in phase. If the device has rotated, there will be a difference in the length of the paths of the two beams, resulting in a detectable phase difference in the combined signal. The signal will vary in amplitude depending on the amount of the phase shift. The amplitude is thus a measurement of the phase shift, and consequently, the rotation rate. This signal is processed into a digital readout in degrees. This readout, being digital, can then be sent to a variety of devices which need heading information, such as helm, autopilot, and electronic chart systems.

A single ring laser gyroscope can be used to provide a one-dimensional rotational reference, exactly what a compass needs. The usefulness of ring laser gyrocompasses stems from that fact that they share many of the same characteristics of flux gate compasses. They are compact, light, inexpensive, accurate, dependable, and robust. The ring laser device is also quite immune to magnetic influences which would send a traditional compass spinning hopelessly, and might adversely affect even the remotely mounted flux gate compass.

Ring laser gyroscopes can also serve as the stable elements in an inertial guidance system, using three gyros to represent the three degrees of freedom, thus providing both directional and position information. The principle of operation is the same as for mechanical inertial navigation devices, in that a single gyro can measure any rotation about its own axis. This implies that its orientation in space about its own axis will be known at all times. Three gyros arranged along three axes each at 90 degrees to the others can measure accelerations in three dimensional space, and thus track movement over time.

Inertial navigation systems based on ring lasers have been used in aircraft for a number of years, and are becoming increasingly common in maritime applications. Uses include navigation, radar and fire control systems, precise weapons stabilization, and stabilization of directional sensors such as satellite antennas.

CORRECTING AND UNCORRECTING THE COMPASS

624. Ship's Heading

Ship's heading is the angle, expressed in degrees clockwise from north, of the ship's fore-and-aft line with respect to the true meridian or the magnetic meridian. When this angle is referred to the true meridian, it is called a **true heading**. When this angle is referred to the magnetic meridian, it is called a **magnetic heading**. Heading, as indicated on a particular compass, is termed the ship's compass heading by that compass. It is essential to specify every heading as true (T), magnetic (M), or compass. Two abbreviations simplify recording of compass directions. The abbreviation PGC refers to "per gyro compass," and PSC refers to "per steering compass." The steering compass is the one being used by the helmsman or autopilot, regardless of type.

625. Variation And Deviation

Variation is the angle between the magnetic meridian and the true meridian at a given location. If the northerly part of the magnetic meridian lies to the right of the true meridian, the variation is easterly. Conversely, if this part is to the left of the true meridian, the variation is westerly. The local variation and its small **annual change** are noted on the compass rose of all navigational charts. Thus the true and magnetic headings of a ship differ by the local variation.

As previously explained, a ship's magnetic influence will generally cause the compass needle to deflect from the magnetic meridian. This angle of deflection is called **deviation**. If the north end of the needle points east of the magnetic meridian, the deviation is easterly; if it points west of the magnetic meridian, the deviation is westerly.

626. Heading Relationships

A summary of heading relationships follows:

1. **Deviation** is the difference between the compass heading and the magnetic heading.
2. **Variation** is the difference between the magnetic heading and the true heading.
3. The algebraic sum of deviation and variation is the **compass error**.

The following simple rules will assist in correcting and uncorrecting the compass:

1. Compass least, error east; compass best, error west.
2. When correcting, add easterly errors, subtract westerly errors (Remember: "Correcting Add East").
3. When uncorrecting, subtract easterly errors, add westerly errors.

Some typical correction operations follow:

Compass	Deviation	Magnetic	Variation	True
		-> +E, -W		
358°	5°E	003°	6°E	009°
120°	1°W	119°	3°E	122°
180°	6°E	186°	8°W	178°
240°	5°W	235°	7°W	228°
		+W, -E <-		

Figure 626. Examples of compass correcting.

Use the memory aid "Can Dead Men Vote Twice, At Elections" to remember the conversion process (Compass, Deviation, Magnetic, Variation, True; Add East). When converting compass heading to true heading, add easterly deviations and variations and subtract westerly deviations and variations.

The same rules apply to correcting gyrocompass errors, although gyro errors always apply in the same direction. That is, they are E or W all around the compass.

Complete familiarity with the correcting of compasses is essential for navigation by magnetic or gyro compass. The professional navigator who deals with them continually can do them in his head quickly and accurately.

CHAPTER 7

DEAD RECKONING

DEFINITION AND PURPOSE

700. Definition and Use

Dead reckoning is the process of determining one's present position by projecting course(s) and speed(s) from a known past position, and predicting a future position by projecting course(s) and speed(s) from a known present position. The DR position is only an approximate position because it does not allow for the effect of leeway, current, helmsman error, or compass error.

Dead reckoning helps in determining sunrise and sunset; in predicting landfall, sighting lights and predicting arrival times; and in evaluating the accuracy of electronic positioning information. It also helps in predicting which celestial bodies will be available for future observation. But its most important use is in projecting the position of the ship into the immediate future and avoiding hazards to navigation.

The navigator should carefully tend his DR plot, update it when required, use it to evaluate external forces acting on his ship, and consult it to avoid potential navigation hazards. A fix taken at each DR position will reveal the effects of current, wind, and steering error, and allow the navigator to stay on track by correcting for them.

The use of DR when an Electronic Charts Display and Information System (ECDIS) is the primary plotting method will vary with the type of system. An ECDIS allows the display of the ship's heading projected out to some future position as a function of time, the display of waypoint information, and progress toward each waypoint in turn.

Until ECDIS is proven to provide the level of safety and accuracy required, the use of a traditional DR plot on paper charts is a prudent backup, especially in restricted waters. The following procedures apply to DR plotting on the traditional paper chart.

CONSTRUCTING THE DEAD RECKONING PLOT

Maintain the DR plot directly on the chart in use. DR at least two fix intervals ahead while piloting. If transiting in the open ocean, maintain the DR at least four hours ahead of the last fix position. Maintaining the DR plot directly on the chart allows the navigator to evaluate a vessel's future position in relation to charted navigation hazards. It also allows the conning officer and captain to plan course and speed changes required to meet any operational commitments.

This section will discuss how to construct the DR plot.

701. Measuring Courses and Distances

To measure courses, use the chart's compass rose nearest to the chart area currently in use. Transfer course lines to and from the compass rose using parallel rulers, rolling rulers, or triangles. If using a parallel motion plotter (PMP), simply set the plotter at the desired course and plot that course directly on the chart. Transparent plastic navigation plotters that align with the latitude/longitude grid may also be used.

The navigator can measure direction at any convenient place on a Mercator chart because the meridians are parallel to each other and a line making an angle with any one makes the same angle with all others. One must measure direction on a conformal chart having nonparallel meridians at the meridian closest to the area of the chart in use. The only common nonconformal projection used is the gnomonic; a gnomonic chart usually contains instructions for measuring direction.

Compass roses may give both true and magnetic directions. True directions are on the outside of the rose; magnetic directions are on the inside. For most purposes, use true directions.

Measure distances using the chart's latitude scale. Although not technically true, assuming that one minute of latitude equals one nautical mile introduces no significant error. Since the Mercator chart's latitude scale expands as latitude increases, on small scale charts one must measure distances on the latitude scale closest to the area of interest, that is, at the same latitude, or directly to the side. On large scale charts, such as harbor charts, one can use either the latitude scale or the distance scale provided. To measure long distances on small-scale charts, break the distance into a number of segments and measure each segment at its mid-latitude.

702. Plotting and Labeling the Course Line and Positions

Draw a new **course line** whenever restarting the DR. Extend the course line from a fix in the direction of the ordered course. Above the course line place a capital C followed by the ordered course in degrees true. Below the course line, place a capital S followed by the speed in knots. Label all course lines and fixes immediately after plotting them because a conning officer or navigator can easily misinterpret an unlabeled line or position.

Enclose a fix from two or more Lines of Position (LOP's) by a small circle and label it with the time to the nearest minute, written horizontally. Mark a DR position with a semicircle and the time, written diagonally. Mark an **estimated position (EP)** by a small square and the time, written horizontally. Determining an EP is covered later in this chapter.

Express the time using four digits without punctuation, using either zone time or Greenwich Mean Time (GMT), according to procedure. Label the plot neatly, succinctly, and clearly.

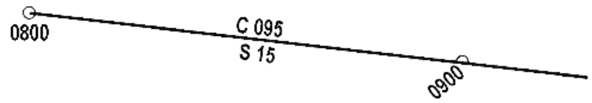


Figure 702. A course line with labels.

Figure 702 illustrates this process. The navigator plots and labels the 0800 fix. The conning officer orders a course of 095°T and a speed of 15 knots. The navigator extends the course line from the 0800 fix in a direction of 095°T. He calculates that in one hour at 15 knots he will travel 15 nautical miles. He measures 15 nautical miles from the 0800 fix position along the course line and marks that point on the course line with a semicircle. He labels this DR with the time. Note that, by convention, he labels the fix time horizontally and the DR time diagonally.

THE RULES OF DEAD RECKONING

703. Plotting the DR

Plot the vessel's DR position:

1. At least every hour on the hour.
2. After every change of course or speed.
3. After every fix or running fix.
4. After plotting a single line of position.

Figure 703 illustrates applying these rules. Clearing the harbor at 0900, the navigator obtains a last visual fix. This is called **taking departure**, and the position determined is called the **departure**. At the 0900 departure, the conning officer orders a course of 090°T and a speed of 10 knots. The navigator lays out the 090°T course line from the departure.

At 1000, the navigator plots a DR position according to the rule requiring plotting a DR position at least every hour on the hour. At 1030, the conning officer orders a course change to 060°T. The navigator plots the 1030 DR position in accordance with the rule requiring plotting a DR position at every course and speed change. Note that the course line changes at 1030 to 060°T to conform to the new course. At 1100, the conning officer changes course back to 090°T. The navigator plots an 1100 DR due to the course change. Note that, regardless of the course change, an 1100 DR would have been required because of the "every hour on the hour" rule.

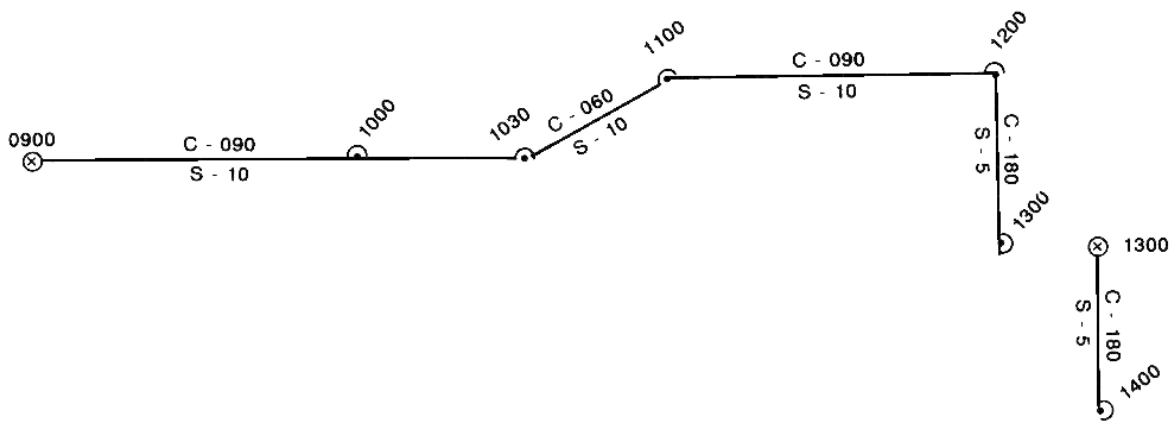


Figure 703. A typical dead reckoning plot.

At 1200, the conning officer changes course to 180°T and speed to 5 knots. The navigator plots the 1200 DR. At 1300, the navigator obtains a fix. Note that the fix position is offset to the east from the DR position. The navigator determines set and drift from this offset and applies this set and drift to any DR position from 1300 until the next fix to determine an estimated position. He also resets the DR to the fix; that is, he draws the 180°T course line from the 1300 fix, not the 1300 DR.

704. Resetting the DR

Reset the DR plot to each fix or running fix in turn. In addition, consider resetting the DR to an inertial estimated position, if an inertial system is installed.

If a navigator has not taken a fix for an extended period of time, the DR plot, not having been reset to a fix, will accumulate time-dependent errors. Over time that error may become so significant that the DR will no longer show the ship's position with acceptable accuracy. If the vessel is equipped with an inertial navigator, the navigator should consider resetting the DR to the inertial estimated position. Some factors to consider when making this determination are:

(1) Time since the last fix and availability of fix information. If it has been a short time since the last fix and fix information may soon become available, it may be advisable to wait for the next fix to reset the DR.

(2) Dynamics of the navigation situation. If, for example, a submerged submarine is operating in the Gulf Stream, fix information is available but operational considerations may preclude the submarine from going to periscope depth to obtain a fix. Similarly, a surface ship with an inertial navigator may be in a dynamic current and suffer a temporary loss of electronic fix equipment. In either case, the fix information will be available shortly but the dynamics of the situation call for a more accurate assessment of the vessel's position. Plotting an inertial EP and resetting the DR to that EP may provide the navigator with a more accurate assessment of the navigation situation.

(3) Reliability and accuracy of the fix source. If a submarine is operating under the ice, for example, only the inertial EP fixes may be available for weeks at a time. Given a high prior correlation between the inertial EP and highly accurate fix systems such as GPS, and the continued proper operation of the inertial navigator, the navigator may decide to reset the DR to the inertial EP.

DEAD RECKONING AND SHIP SAFETY

Properly maintaining a DR plot is important for ship safety. The DR allows the navigator to examine a future position in relation to a planned track. It allows him to anticipate charted hazards and plan appropriate action to avoid them. Recall that the DR position is only approximate. Using a concept called **fix expansion** compensates for the DR's inaccuracy and allows the navigator to use the DR more effectively to anticipate and avoid danger.

705. Fix Expansion

Often a ship steams in the open ocean for extended periods without a fix. This can result from any number of factors ranging from the inability to obtain celestial fixes to malfunctioning electronic navigation systems. Infrequent fixes are particularly common on submarines. Whatever the reason, in some instances a navigator may find himself in the position of having to steam many hours on DR alone.

The navigator must take precautions to ensure that all hazards to navigation along his path are accounted for by the approximate nature of a DR position. One method which can be used is **fix expansion**.

Fix expansion takes into account possible errors in the DR calculation caused by factors which tend to affect the vessel's actual course and speed over the ground. The navigator considers all such factors and develops an expanding "error circle" around the DR plot. One of the basic assumptions of fix expansion is that the various individual effects of current, leeway, and steering error combine to cause a cumulative error which increases over time, hence, the concept of expansion. While the errors may in fact cancel each other out, the worst case is that they will all be additive, and this is what the navigator must anticipate.

Errors considered in the calculation of fix expansion encompass all errors that can lead to DR inaccuracy. Some of the most important factors are current and wind, compass or gyro error, and steering error. Any method which attempts to determine an error circle must take these factors into account. The navigator can use the magnitude of set and drift calculated from his DR plot. See Article 707. He can obtain the current's estimated magnitude from pilot charts or weather reports. He can determine wind speed from weather instruments. He can determine compass error by comparison with an accurate standard or by obtaining an azimuth of the Sun. The navigator determines the effect each of these errors has on his course and speed over ground, and applies that error to the fix expansion calculation.

As noted previously, error is a function of time; it grows as the ship proceeds along the track without obtaining a fix. Therefore, the navigator must incorporate his calculated errors into an **error circle** whose radius grows with time. For example, assume the navigator calculates that all the various sources of error can create a cumulative position error of no more than 2 nm. Then his fix expansion error circle would grow at that rate; it would

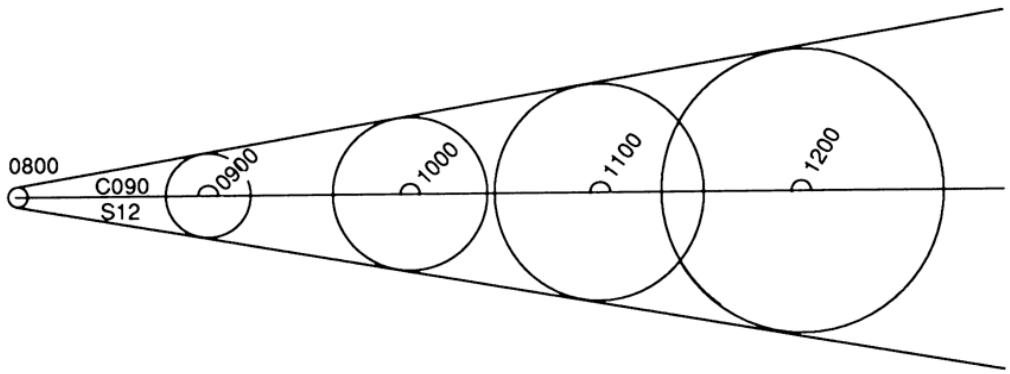


Figure 705. Fix expansion. All possible positions of the ship lie between the lines tangent to the expanding circles. Examine this area for dangers.

be 2 nm after the first hour, 4 nm after the second, and so on.

At what value should the navigator start this error circle? Recall that a DR is laid out from every fix. All fix sources have a finite absolute accuracy, and the initial error circle should reflect that accuracy. Assume, for example, that a satellite navigation system has an accuracy of 0.5 nm. Then the initial error circle around that fix should be set at 0.5 nm.

First, enclose the fix position in a circle, the radius of which is equal to the accuracy of the system used to obtain the fix. Next, lay out the ordered course and speed from the fix position. Then apply the fix expansion circle to the hourly DR's, increasing the radius of the circle by the error factor each time. In the example given above, the DR after one hour would be enclosed by a circle of radius 2.5 nm, after two hours 4.5 nm, and so on. Having encircled the four hour DR positions with the error circles, the navigator then draws two lines originating tangent to the original error circle and simultaneously tangent to the other error circles. The navigator then closely examines the area between the two tangent lines for hazards to navigation. This technique is illustrated in Figure 705.

The fix expansion encompasses the total area in which the vessel could be located (as long as all sources of error are considered). If any hazards are indicated within the cone, the navigator should be especially alert for those dangers. If, for example, the fix expansion indicates that the vessel may be standing into shoal water, continuously monitor the fathometer. Similarly, if the fix expansion indicates that the vessel might be approaching a charted obstruction, post extra lookouts.

The fix expansion may grow at such a rate that it becomes unwieldy. Obviously, if the fix expansion grows to cover too large an area, it has lost its usefulness as a tool for the navigator, and he should obtain a new fix by any available means.

DETERMINING AN ESTIMATED POSITION

An estimated position (EP) is a DR position corrected for the effects of leeway, steering error, and current. This section will briefly discuss the factors that cause the DR position to diverge from the vessel's actual position. It will then discuss calculating set and drift and applying these values to the DR to obtain an estimated position. It will also discuss determining the estimated course and speed made good.

706. Factors Affecting DR Position Accuracy

Tidal current is the periodic horizontal movement of the water's surface caused by the tide-affecting gravitational forces of the Moon and Sun. **Current** is the horizontal movement of the sea surface caused by meteorological, oceanographic, or topographical effects. From whatever its source, the horizontal motion of the sea's surface is an important dynamic force acting on a vessel.

Set refers to the current's direction, and **drift** refers to the current's speed. **Leeway** is the leeward motion of a vessel due to that component of the wind vector perpendicular to the vessel's track. Leeway and current combine to produce the most pronounced natural dynamic effects on a transiting vessel. Leeway especially affects sailing vessels and high-sided vessels.

In addition to these natural forces, relatively small helmsman and steering compass error may combine to cause additional error in the DR.

707. Calculating Set and Drift and Plotting an Estimated Position

It is difficult to quantify the errors discussed above individually. However, the navigator can easily quantify their cumulative effect by comparing simultaneous fix and DR positions. If there are no dynamic forces acting on the vessel and no steering error, the DR position and the fix position will coincide. However, they seldom do so. The fix is offset from the DR by the vector sum of all the errors.

Note again that this methodology provides no means to determine the magnitude of the individual errors. It simply provides the navigator with a measurable representation of their combined effect.

When the navigator measures this combined effect, he often refers to it as the "set and drift." Recall from above that these terms technically were restricted to describing current effects. However, even though the fix-to-DR offset is caused by effects in addition to the current, this text will follow the convention of referring to the offset as the set and drift.

The set is the direction from the DR to the fix. The drift is the distance in miles between the DR and the fix divided by the number of hours since the DR was last reset. This is true regardless of the number of changes of course or speed since the last fix. The prudent navigator calculates set and drift at every fix.

To calculate an EP, draw a vector from the DR position in the direction of the set, with the length equal to the product of the drift and the number of hours since the last reset. See Figure 707. From the 0900 DR position the navigator draws a set and drift vector. The end of that vector marks the 0900 EP. Note that the EP is enclosed in a square and labeled horizontally with the time. Plot and evaluate an EP with every DR position.

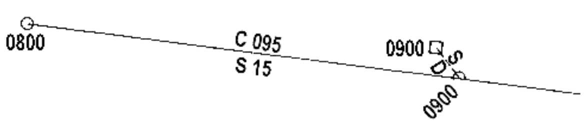


Figure 707. Determining an estimated position.

708. Estimated Course and Speed Made Good

The direction of a straight line from the last fix to the EP is the **estimated track made good**. The length of this line divided by the time between the fix and the EP is the **estimated speed made good**.

Solve for the estimated track and speed by using a vector diagram. See the example problems below and refer to Figure 708a.

Example 1: *A ship on course 080°, speed 10 knots, is steaming through a current having an estimated set of 140° and drift of 2 knots.*

Required: *Estimated track and speed made good.*

Solution: *See Figure 708a. From A, any convenient point, draw AB, the course and speed of the ship, in direction 080°, for a distance of 10 miles.*

From B draw BC, the set and drift of the current, in direction 140°, for a distance of 2 miles.

The direction and length of AC are the estimated track and speed made good.

Answers: *Estimated track made good 089°, estimated speed made good 11.2 knots.*

To find the course to steer at a given speed to make good a desired course, plot the current vector from the origin, A, instead of from B. See Figure 708b.

Example 2: *The captain desires to make good a course of 095° through a current having a set of 170° and a drift of 2.5 knots, using a speed of 12 knots.*

Required: *The course to steer and the speed made good.*

Solution: *See Figure 708b. From A, any convenient point, draw line AB extending in the direction of the course to be made good, 095°.*

From A draw AC, the set and drift of the current.

Using C as a center, swing an arc of radius CD, the speed through the water (12 knots), intersecting line AB at D.

Measure the direction of line CD, 083.5°. This is the course to steer.

Measure the length AD, 12.4 knots. This is the speed made good.

Answers: *Course to steer 083.5°, speed made good 12.4 knots.*

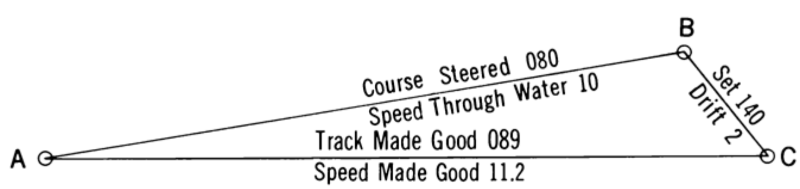


Figure 708a. Finding track and speed made good through a current.

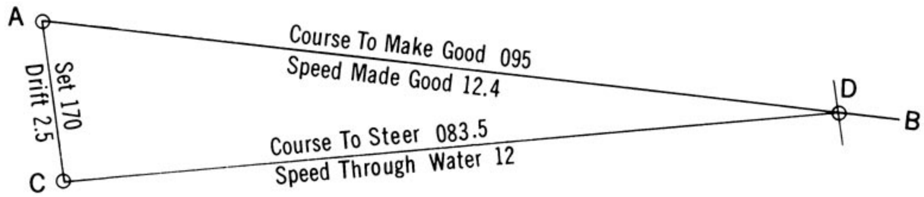


Figure 708b. Finding the course to steer at a given speed to make good a given course through a current.

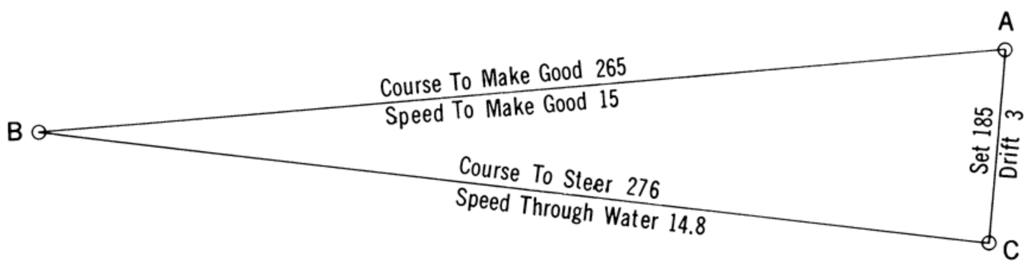


Figure 708c. Finding course to steer and speed to use to make good a given course and speed through the current.

To find the course to steer and the speed to use to make good a desired course and speed, proceed as follows:
See Figure 708c.

Example 3: *The captain desires to make good a course of 265° and a speed of 15 knots through a current having a set of 185° and a drift of 3 knots.*

Required: *The course to steer and the speed to use.*

Solution: *See Figure 708c. From A, any convenient point, draw AB in the direction of the course to be made good, 265° and for length equal to the speed to be made good, 15 knots.*

From A draw AC, the set and drift of the current.

Draw a straight line from C to B. The direction of this line, 276°, is the required course to steer; and the length, 14.8 knots, is the required speed.

Answers: *Course to steer 276°, speed to use 14.8 kn.*

CHAPTER 8

PILOTING

DEFINITION AND PURPOSE

800. Introduction

Piloting involves navigating a vessel in restricted waters and fixing its position as precisely as possible at frequent intervals. More so than in other phases of navigation, proper preparation and attention to detail are important. This chapter will discuss a piloting methodology designed to ensure that procedures are carried out safely and efficiently. These procedures will vary from vessel to vessel according to the skills and composition of the piloting team. It is the responsibility of the navigator to choose the procedures applicable to his own situation, to train the piloting team in their execution, and to ensure that duties are carried out properly.

These procedures are written primarily from the perspective of the military navigator, with some notes included where civilian procedures might differ. This set of procedures is designed to minimize the chance of error and maximize safety of the ship.

The military navigation team will nearly always consist of several more people than are available to the civilian navigator. Therefore, the civilian navigator must streamline these procedures, eliminating certain steps, doing only what is essential to keep his ship in safe water.

The navigation of civilian vessels will therefore proceed differently than for military vessels. For example, while the military navigator might have bearing takers stationed at the gyro repeaters on the bridge wings for taking simultaneous bearings, the civilian navigator must often take and plot them himself. While the military navigator will have a bearing book and someone to record entries for each fix, the civilian navigator will simply plot the bearings on the chart as they are taken and not record them at all.

If the ship is equipped with an ECDIS, it is reasonable for the navigator to simply monitor the progress of the ship along the chosen track, visually ensuring that the ship is proceeding as desired, checking the compass, sounder and other indicators only occasionally. If a pilot is aboard, as is often the case in the most restricted of waters, his judgement can generally be relied upon explicitly, further easing the workload. But should the ECDIS fail, the navigator will have to rely on his skill in the manual and time-tested procedures discussed in this chapter.

While an ECDIS is the legal equivalent of a paper chart and can be used as the primary plot, an ECS, (non-ECDIS compliant electronic chart system) cannot be so used. An ECS may be considered as an additional resource used to ensure safe navigation, but cannot be relied upon for performing all the routine tasks associated with piloting. The individual navigator, with knowledge of his vessel, his crew, and the capabilities they possess, must make a professional judgement as to how the ECS can support his efforts to keep his ship in safe water. The navigator should always remember that reliance on any single navigation system courts disaster. An ECS does not relieve the navigator of maintaining a proper and legal plot on a paper chart.

PREPARATION

801. Plot Setup

The navigator's job begins well before getting underway. Much advance preparation is necessary to ensure a safe and efficient voyage. The following steps are representative:

Ensure the plotting station(s) have the following instruments:

- **Dividers:** Dividers are used to measure distances between points on the chart.

- **Compasses:** Compasses are used to plot range arcs for radar LOP's. **Beam compasses** are used when the range arc exceeds the spread of a conventional compass. Both should be available at both plots.

- **Plotters:** Several types of plotters are available. The preferred device for large vessels is the parallel motion plotter (PMP) used in conjunction with a drafting table. Otherwise, use a transparent protractor plotter, or triangles, parallel rulers or rolling rulers in conjunction with the chart's compass rose. Finally, the plotter can use a one arm protractor. The plotter should use the device with which he can work the most quickly and accurately.

- **Sharpened Pencils and Erasers:** Ensure an adequate supply of pencils is available.

- **Fischer Radar Plotting Templates:** Fischer plotting is covered in Chapter 13. The plotting templates for this technique should be stacked near the radar repeater.

- **Time-Speed-Distance Calculator:** Given two of the three unknowns (between time, speed, and distance), this calculator allows for rapid computation of the third.

- **Tide and Current Graphs:** Post the tide and current graphs near the primary plot for easy reference during the transit. Give a copy of the graphs to the conning officer and the captain.

Once the navigator verifies the above equipment is in place, he tapes down the charts on the chart table. If more than one chart is required for the transit, tape the charts in a stack such that the plotter works from the top to the bottom of the stack. This minimizes the time required to shift the chart during the transit. If the plotter is using a PMP, align the arm of the PMP with any meridian of longitude on the chart. While holding the PMP arm stationary, adjust the PMP to read 000.0°T. This procedure calibrates the PMP to the chart in use. Perform this alignment every time the piloting team shifts charts.

Be careful not to fold under any important information when folding the chart on the chart table. Ensure the chart's distance scale, the entire track, and all important warning information are visible.

Energize and test all electronic navigation equipment, if not already in operation. This includes the radar and the GPS receiver. Energize and test the fathometer. Ensure the entire electronic navigation suite is operating properly prior to entering restricted waters.

802. Preparing Charts and Publications

- **Assemble Required Publications.** These publications should include *Coast Pilots*, *Sailing Directions*, *USCG Light Lists*, *NIMA Lists of Lights*, *Tide Tables*, *Tidal Current Tables*, *Notice to Mariners*, and *Local Notice to Mariners*. Often, for military vessels, a port will be under the operational direction of a particular squadron; obtain that squadron's port Operation Order. Civilian vessels should obtain the port's harbor regulations. These publications will cover local regulations such as speed limits and bridge-to-bridge radio frequency monitoring requirements. Assemble and review the Broadcast Notice to Mariners file.

- **Select and Correct Charts.** Choose the largest scale chart available for the harbor approach or departure. Often, the harbor approach will be too long to be represented on only one chart. For example, three charts are required to cover the waters from the Naval Station in Norfolk to the entrance of the Chesapeake Bay. Therefore, obtain all the charts required to cover the entire passage. Using the *Notice to Mariners*, verify that these charts have been corrected through the latest *Notice to Mariners*. Check the *Local Notice to Mariners* and the Broadcast Notice to Mariners file to ensure the chart is fully corrected. Annotate on the chart or a chart correction card all the corrections that have been made; this will make it easier to verify the chart's correction status prior to its next use. Naval ships may need to prepare three sets of charts. One set is for the primary plot, the second set is for the secondary plot, and the third set is for the conning officer and captain. Civilian vessels will prepare one set.

- **Mark the Minimum Depth Contour:** Determine the minimum depth of water in which the vessel can safely operate and outline that depth contour on the chart. Do this step before doing any other harbor navigation planning. Highlight this outline in a bright color so that it clearly stands out. Carefully examine the area inside the contour and mark the isolated shoals less than the minimum depth which fall inside the marked contour. Determine the minimum depth in which the vessel can operate as follows:

  Minimum Depth = Ship's Draft – Height of Tide + Safety Margin + Squat. (See Article 804 and Article 818.)

  Remember that often the fathometer's transducer is not located at the section of the hull that extends the furthest below the waterline. Therefore, the indicated depth of water is that below the fathometer transducer, not the depth of water below the vessel's deepest draft.

- **Highlight Selected Visual Navigation Aids (NAVAIDS).** Circle, highlight and label the main navigational aids on the chart. Consult the applicable *Coast Pilot* or *Sailing Directions* to determine a port's best NAVAIDS if the piloting team has not visited the port previously. These aids can be lighthouses, piers, shore features, or tanks; any prominent feature that is displayed on the chart can be used as a NAVAID. Label critical buoys, such as those marking a harbor entrance or a traffic separation scheme. Verify charted lights against the *Light List* or the *List of Lights* to confirm the charted information is correct. This becomes most critical when attempting to identify a light at night. Label NAVAIDS succinctly and clearly. Ensure everyone in the navigation team refers to a NAVAID using the same terminology. This will reduce confusion between the bearing taker, the bearing recorder, and plotter.

- **Highlight Selected Radar NAVAIDS.** Highlight radar NAVAIDS with a triangle instead of a circle. If the NAVAID is suitable for either visual or radar piloting, it can be highlighted with either a circle or a triangle.

- **Plot the Departure/Approach Track.** This process is critical for ensuring safe pilotage. Consult the *Fleet Guide* and *Sailing Directions* for recommendations on the best track to use. Look for any information or regulations published by the local harbor authority. Lacking any of this information, locate a channel or safe route on the chart and plot the vessel's track. Most U.S. ports have well-defined channels marked with buoys. Carefully check the intended track to ensure a sufficient depth of water under the keel will exist for the entire passage. If the scale of the chart permits, lay the track out to the starboard side of the channel to allow for any vessel traffic proceeding in the opposite direction. Many channels are marked by natural or man-made ranges. The bearings of these ranges should be measured to the nearest 0.1° or noted from the *Light List*, and this value should be marked on the chart. Not only are ranges useful in keeping a vessel on track, they are invaluable for determining gyro error. See Article 807.

- **Label the Departure/Approach Track.** Label the track course to the nearest 0.5°. Similarly, label the distance of each track leg. Highlight the track courses for easy reference while piloting. Often a navigator might plan two separate tracks. One track would be for use during good visibility and the other for poor visibility. Considerations might include concern for the number of turns (fewer turns for poor visibility) or proximity to shoal water (smaller margin for error might be acceptable in good visibility). In this case, label both tracks as above and appropriately mark when to use each track.

- **Use Advance and Transfer to Find Turning Points.** The distance the vessel moves along its original course from the time the rudder is put over until the new course is reached is called **advance**. The distance the vessel moves perpendicular to the original course during the turn is called **transfer**.The track determined above does not account for these. See Figure 802a. Use the advance and transfer characteristics of the vessel to determine when the vessel must put its rudder over to gain the next course. From that point, fair in a curve between the original course and the new course. Mark the point on the original course where the vessel must put its rudder over as the **turning point**. See Figure 802b.

- **Plot Turn Bearings and Ranges.** A **turn bearing** is a predetermined bearing to a charted object from the track point at which the rudder must be put over in order to make a desired turn. In selecting a NAVAID for a turn bearing, find one as close to abeam as possible at the turning point, and if possible on the inside elbow of the turn. Account for advance and transfer and label the bearing to the nearest 0.1°. A **turn range** is similar, but taken as a radar range to a prominent object ahead or astern. Ideally, both can be used, one as a check against the other.

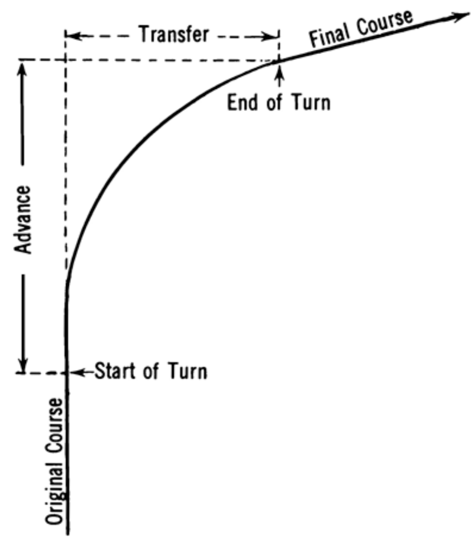


Figure 802a. Advance and transfer.

Example: *Figure 802b illustrates using advance and transfer to determine a turn bearing. A ship proceeding on course 100° is to turn 60° to the left to come on a range which will guide it up a channel. For a 60° turn and the amount of rudder used, the advance is 920 yards and the transfer is 350 yards.*

Required: *The bearing of flagpole "FP." when the rudder is put over.*

Solution:

1. *Extend the original course line, AB.*
2. *At a perpendicular distance of 350 yards, the transfer, draw a line A'B' parallel to the original course line AB. The point of intersection, C, of A'B' with the new course line is the place at which the turn is to be completed.*
3. *From C draw a perpendicular, CD, to the original course line, intersecting at D.*
4. *From D measure the advance, 920 yards, back along the original course line. This locates E, the point at which the turn should be started.*
5. *The direction of "FP." from E, 058°, is the bearing when the turn should be started.*

Answer: *Bearing 058°.*

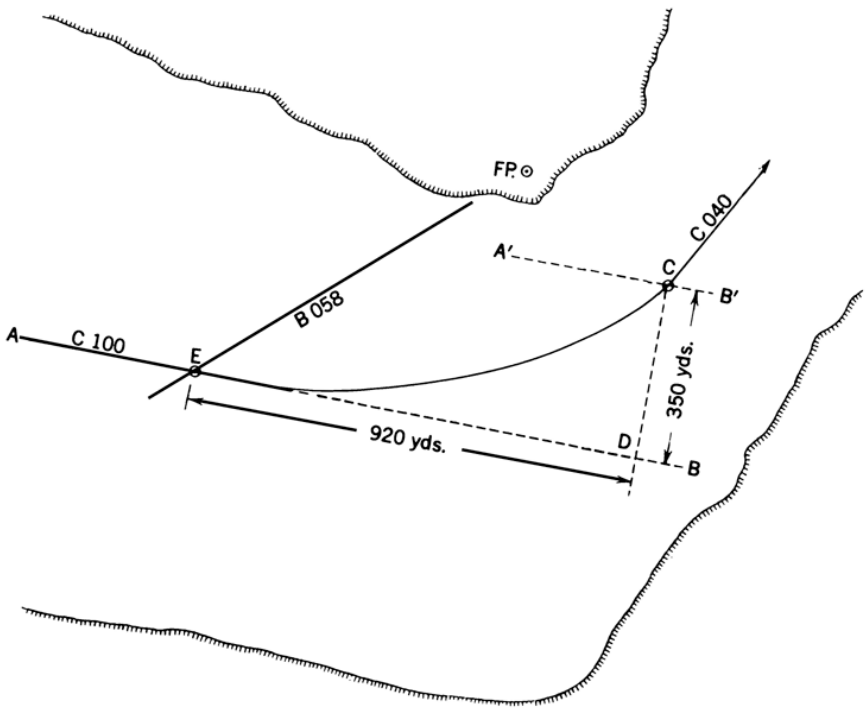


Figure 802b. Allowing for advance and transfer.

- **Plot a Slide Bar for Every Turn Bearing:** If the ship is off track immediately prior to a turn, a plotting technique known as the **slide bar** can quickly revise a turn bearing. See Figure 802c. A slide bar is a line drawn parallel to the new course through the turning point on the original course. The navigator can quickly determine a new turn bearing by dead reckoning ahead from the vessel's last fix position to where the DR intersects the slide bar. The revised turn bearing is simply the bearing from that intersection point to the turn bearing NAVAID. Draw the slide bar with a different color from that used for the track in order to see the slide bar clearly.

- **Label Distance to Go from Each Turn Point:** At each turning point, label the distance to go until either the ship moors (inbound) or the ship clears the harbor (outbound). For an inbound transit, a vessel's captain is usually more concerned about time of arrival, so assume a speed of advance and label each turn point with time to go until mooring.

- **Plot Danger Bearings:** Danger bearings warn a navigator he may be approaching a navigational hazard too closely. See Figure 802d. Vector AB indicates a vessel's intended track. This track passes close to the indicated shoal. Draw a line from the NAVAID H tangent to the shoal. The bearing of that tangent line measured from the ship's track is 074.0°T. In other words, as long as NAVAID H bears *less than* 074°T as the vessel proceeds down its track, the vessel will not ground on the shoal. Hatch the side of the bearing line on the side of the hazard and label the danger bearing NMT (no more than) 074.0°T. For an added margin of safety, the line does not have to be drawn exactly tangent to the shoal. Perhaps, in this case, the navigator might want to set an error margin and draw the danger bearing at 065°T from NAVAID H. Lay down a danger bearing from any appropriate NAVAID in the vicinity of any hazard to navigation. Ensure the track does not cross any danger bearing.

- **Plot Danger Ranges:** The danger range is analogous to the danger bearing. It is a standoff range from an object to prevent the vessel from approaching a hazard too closely.

- **Label Warning and Danger Soundings:** To determine the danger sounding, examine the vessel's proposed track and note the minimum expected sounding. The minimum expected sounding is the difference between the shallowest water expected on the transit and the vessel's maximum draft. Set 90% of this difference as the warning sounding and 80% of this difference as the danger sounding. There may be peculiarities about local conditions that will cause the navigator to choose another method of setting warning and danger soundings. Use the above method if no

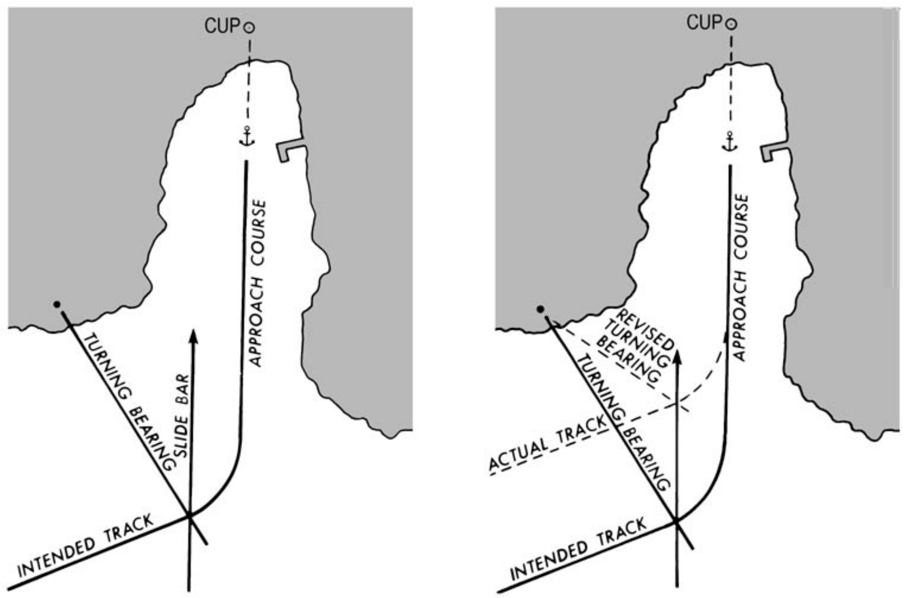


Figure 802c. The slide bar technique.

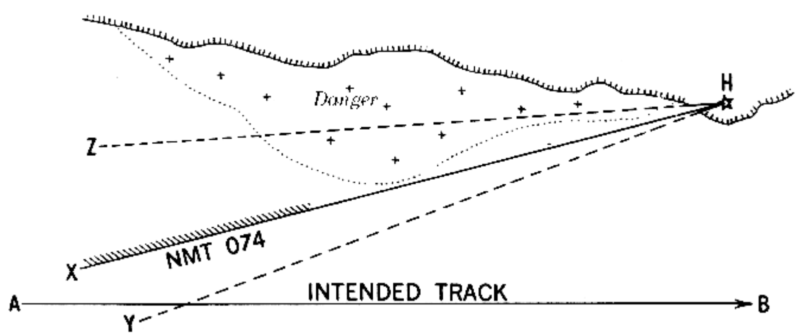


Figure 802d. A danger bearing, hatched on the dangerous side, labeled with the appropriate bearing.

other means is more suitable. For example: A vessel draws a maximum of 20 feet, and it is entering a channel dredged to a minimum depth of 50 feet. Set the warning and danger soundings at 0.9 (50ft. - 20ft) = 27ft and 0.8 (50ft. - 20ft.) = 24ft., respectively. Re-evaluate these soundings at different intervals along the track, when the minimum expected sounding may change. Carefully label the points along the track between which these warning and danger soundings apply.

- **Label Demarcation Line:** Clearly label the point on the ship's track where the Inland and International Rules of the Road apply. This is applicable only when piloting in U.S. ports.
- **Mark Speed Limits Where Applicable:** Often a harbor will have a local speed limit in the vicinity of piers, other vessels, or shore facilities. Mark these speed limits and the points between which they are applicable on the chart.
- **Mark the Point of Pilot Embarkation:** Some ports require vessels over a certain size to embark a pilot. If this is the case, mark the point on the chart where the pilot is to embark.
- **Mark the Tugboat Rendezvous Point:** If the vessel requires a tug to moor, mark the tug rendezvous point on the chart.
- **Mark the Chart Shift Point:** If more than one chart

will be required to complete the passage, mark the point where the navigator should shift to the next chart.

- **Harbor Communications:** Mark the point on the chart where the vessel must contact harbor control. Also mark the point where a vessel must contact its parent squadron to make an arrival report (military vessels only).

- **Tides and Currents:** Mark the points on the chart for which the tides and currents were calculated.

803. Records

Ensure the following records are assembled and personnel assigned to maintain them:

- **Bearing Record Book:** The bearing recorders for the primary and secondary plots should record all the bearings used on their plot during the entire transit. The books should clearly list what NAVAIDS are being used and what method of navigation was being used on their plot. In practice, the primary bearing book will contain mostly visual bearings and the secondary bearing book will contain mostly radar ranges and bearings.

- **Fathometer Log:** In restricted waters, monitor soundings continuously and record soundings every five minutes in the fathometer log. Record all fathometer settings that could affect the sounding display.

- **Deck Log:** This log is the legal record of the passage. Record all ordered course and speed changes. Record all the navigator's recommendations and whether the navigator concurs with the actions of the conning officer. Record all buoys passed, and the shift between different Rules of the Road. Record the name and embarkation of any pilot. Record who has the conn at all times. Record any casualty or important event. The deck log combined with the bearing log should constitute a complete record of the passage.

804. Tides and Currents

Determining the tidal and current conditions of the port is crucial. This process is covered in depth in Chapter 9. In order to anticipate early or late transit, plot a graph of the tidal range for the 24-hour period centered on the scheduled time of arrival or departure. Depending on a vessel's draft and the harbor's depth, some vessels may be able to transit only at high tide. If this is this case, it is critically important to determine the time and range of the tide correctly.

The magnitude and direction of the current will give the navigator some idea of the **set and drift** the vessel will experience during the transit. This will allow him to plan in advance for any potential current effects in the vicinity of navigational hazards.

While printed tide tables can be used for predicting and plotting tides, it is far more efficient to use a computer with appropriate software, or the internet, to compute tides and print out the graphs. These graphs can be posted on the bridge at the chart table for ready reference, and copies made for others involved in the piloting process. NOAA tide tables for the U.S. are available at the following site: http://co-ops.nos.noaa.gov/tp4days.html. Always remember that tide tables give predicted data, but that actual conditions may be quite different due to weather or other natural phenomena.

805. Weather

The navigator should obtain a weather report covering the route which he intends to transit. This will allow him to prepare for any adverse weather by stationing extra lookouts, adjusting speed for poor visibility, and preparing for radar navigation. If the weather is thick, consider standing off the harbor until it clears.

The navigator can receive weather information any number of ways. Military vessels may receive weather reports from their parent squadrons prior to coming into port. Marine band radio carries continuous weather reports. Many vessels are equipped with weather facsimile machines. Some navigators carry cellular phones to reach shoreside personnel and harbor control; these can also be used to get weather reports from NOAA weather stations. If the ship is using a weather routing service for the voyage, it should provide forecasts when asked. Finally, if the vessel has an internet connection, this is an ideal source of weather data. NOAA weather data can be obtained at: http://www.nws.noaa.gov. However he obtains the information, the navigator should have a good idea of the weather before entering piloting waters.

806. The Piloting Brief

Assemble the entire navigation team for a piloting brief prior to entering or leaving port. The vessel's captain and navigator should conduct the briefing. All navigation and bridge personnel should attend. The pilot, if he is already on board, should also attend. If the pilot is not onboard when the ship's company is briefed, the navigator should immediately brief him when he embarks. The pilot must know the ship's maneuvering characteristics before entering restricted waters. The briefing should cover, as a minimum, the following:

- **Detailed Coverage of the Track Plan:** Go over the planned route in detail. Use the prepared and approved chart as part of this brief. Concentrate especially on all the NAVAIDS and soundings which are being used to indicate danger. Cover the buoyage system in use and the port's major NAVAIDS. Point out the radar NAVAIDS for the radar operator. Often, a *Fleet Guide* or *Sailing Directions* will have pictures of a port's NAVAIDS. This is especially important for the piloting party that has never transited a particular port before. If no pictures are available, consider stationing a photographer to take some for submission to NIMA.

- **Harbor Communications:** Discuss the bridge-to bridge radio frequencies used to raise harbor control. Discuss what channel the vessel is supposed to monitor on its passage into port and the port's communication protocol.

- **Duties and Responsibilities:** Each member of the piloting team must have a thorough understanding of his duties and responsibilities. He must also understand how his part fits into the whole. The radar plotter, for example, must know if radar will be the primary or secondary source of fix information. The bearing recorder must know what fix interval the navigator is planning to use. Each person must be thoroughly briefed on his job; there is little time for questions once the vessel enters the channel.

807. Evolutions Prior to Piloting

The navigator should always accomplish the following evolutions prior to piloting:

- **Testing the Shaft on the Main Engines in the Astern Direction:** This ensures that the ship can answer a backing bell. If the ship is entering port, no special precautions are required prior to this test. If the ship is tied up at the pier preparing to get underway, exercise extreme caution to ensure no way is placed on the ship while testing the main engines.

- **Making the Anchor Ready for Letting Go:** Make the anchor ready for letting go and station a watchstander in direct communications with the bridge at the anchor windlass. Be prepared to drop anchor immediately when piloting if required to keep from drifting too close to a navigational hazard.

- **Calculate Gyro Error:** An error of greater than 1.0° T indicates a gyro problem which should be investigated prior to piloting. There are several ways to determine gyro error:

  1. Compare the gyro reading with a known accurate heading reference such as an inertial navigator. The difference in the readings is the gyro error.

  2. Mark the bearing of a charted range as the range NAVAID's come into line and compare the gyro bearing with the charted bearing. The difference is the gyro error.

  3. Prior to getting underway, plot a dockside fix using at least three lines of position. The three LOP's should intersect at a point. Their intersecting in a "cocked hat" indicates a gyro error. Incrementally adjust each visual bearing by the same amount and direction until the fix plots as a pinpoint. The total correction required to eliminate the cocked hat is the gyro error.

  4. Measure a celestial body's azimuth or amplitude, or Polaris' azimuth with the gyro, and then compare the measured value with a value computed from the *Sight Reduction Tables* or the *Nautical Almanac*. These methods are covered in detail in Chapter 17.

Report the magnitude and direction of the gyro error to the navigator and captain. The direction of the error is determined by the relative magnitude of the gyro reading and the value against which it is compared. When the compass is least, the error is east. Conversely, when the compass is best, the error is west. See Chapter 6.

808. Inbound Voyage Planning

The vessel's planned estimated time of arrival (ETA) at its mooring determines the vessel's course and speed to the harbor entrance. Arriving at the mooring site on time may be important in a busy port which operates its port services on a tight schedule. Therefore, it is important to plan the arrival accurately. Take the desired time of arrival at the mooring and subtract from that the time it will take to navigate to it from the entrance. The resulting time is when you must arrive at the harbor entrance. Next, measure the distance between the vessel's present location and the harbor entrance. Determine the speed of advance (SOA) the vessel will use to make the transit to the harbor. Use the distance to the harbor and the SOA to calculate what time to leave the present position to make the mooring ETA, or what speed must be made good to arrive on time.

Consider these factors which might affect this decision:

- **Weather:** This is the single most important factor in harbor approach planning because it directly affects the vessel's SOA. The thicker the weather, the more slowly the vessel must proceed. Therefore, if heavy fog or rain is in the forecast, the navigator must allow more time for the transit.

- **Mooring Procedures:** The navigator must take more than distance into account when calculating how long it will take him to pilot to his mooring. If the vessel needs a tug, that will increase the time needed. Similarly, picking up or dropping off a pilot adds time to the transit. It is better to allow a margin for error when trying to add up all the time delays caused by these procedures. It is always easier to avoid arriving early by slowing down than it is to make up lost time by speeding up.

- **Shipping Density:** Generally, the higher the shipping density entering and exiting the harbor, the longer it will take to proceed into the harbor entrance safely.

TRANSITION TO PILOTING

809. Stationing the Piloting Team

At the appropriate time, station the piloting team. Allow plenty of time to acclimate to the navigational situation and if at night, to the darkness. The number and type of personnel available for the piloting team depend on the vessel. A Navy warship, for example, has more people available for piloting than a merchant ship. Therefore, more than one of the jobs listed below may have to be filled by a single person. The piloting team should consist of:

- **The Captain:** The captain is ultimately responsible for the safe navigation of his vessel. His judgment regarding navigation is final. The piloting team acts to support the captain, advising him so he can make informed decisions on handling his vessel.

- **The Pilot:** The pilot is usually the only member of the piloting team not a member of the ship's company. The piloting team must understand the relationship between the pilot and the captain. The pilot is perhaps the captain's most important navigational advisor. Generally, the captain will follow his recommendations when navigating an unfamiliar harbor. The pilot, too, bears some responsibility for the safe passage of the vessel; he can be censured for errors of judgment which cause accidents. However, the presence of a pilot in no way relieves the captain of his ultimate responsibility for safe navigation. The piloting team works to support and advise the captain.

- **The Officer of the Deck (Conning Officer):** In Navy piloting teams, neither the pilot or the captain usually has the **conn**. The officer having the conn directs the ship's movements by rudder and engine orders. Another officer of the ship's company usually fulfills this function. The captain can take the conn immediately simply by issuing an order to the helm should an emergency arise. The conning officer of a merchant vessel can be either the pilot, the captain, or another watch officer. In any event, the officer having the conn must be clearly indicated in the ship's deck log at all times. Often a single officer will have the deck and the conn. However, sometimes a junior officer will take the conn for training. In this case, different officers will have the deck and the conn. The officer who retains the deck retains the responsibility for the vessel's safe navigation.

- **The Navigator:** The vessel's navigator is the officer directly responsible to the ship's captain for the safe navigation of the ship. He is the captain's principal navigational advisor. The piloting team works for him. He channels the required information developed by the piloting team to the ship's conning officer on recommended courses, speeds, and turns. He also carefully looks ahead for potential navigational hazards and makes appropriate recommendations. He is the most senior officer who devotes his effort exclusively to monitoring the navigation picture. The captain and the conning officer are concerned with all aspects of the passage, including contact avoidance and other necessary ship evolutions (making up tugs, maneuvering alongside a small boat for personnel transfers, engineering evolutions, and coordinating with harbor control via radio, for example). The navigator, on the other hand, focuses solely on safe navigation. It is his job to anticipate dangers, keep himself appraised of the navigation situation at all times, and manage the team.

- **Bearing Plotting Team:** This team consists, ideally, of three persons. The first person measures the bearings. The second person records the bearings in an official record book. The third person plots the bearings. The more quickly and accurately this process is completed, the sooner the navigator has an accurate picture of the ship's position. The bearing taker should be an experienced individual who has traversed the port before and who is familiar with the NAVAIDS. He should take his round of bearings as quickly as possible, beam bearings first, minimizing any time delay errors in the resulting fix. The plotter should also be an experienced individual who can quickly and accurately lay down the required bearings. The bearing recorder can be one of the junior members of the piloting team.

- **The Radar Operator:** The radar operator has one of the more difficult jobs of the team. The radar is as important for collision avoidance as it is for navigation. Therefore, this operator must often "time share" the radar between these two functions. Determining the amount of time spent on these functions falls within the judgment of the captain and the navigator. If the day is clear and the traffic heavy, the captain may want to use the radar mostly for collision avoidance. As the weather worsens, obscuring visual NAVAIDS, the importance of radar for safe navigation increases. The radar operator must be given clear guidance on how the captain and navigator want the radar to be operated.

- **Plot Supervisors:** On many military ships, the piloting team will consist of two plots: the primary plot and the secondary plot. The navigator should designate the type of navigation that will be employed on the primary plot. All other fix sources should be plotted on the secondary plot. The navigator can function as the primary plot supervisor. A senior, experienced individual should be employed as a secondary plot supervisor. The navigator should frequently compare the positions plotted on both plots as a check on the primary plot.

  There are three major reasons for maintaining a primary and secondary plot. First, as mentioned above, the secondary fix sources provide a good check on the accuracy of visual piloting. Large discrepancies between visual and radar positions may point out a problem with the visual fixes that the navigator might not otherwise suspect. Secondly, the navigator often must change the primary means of navigation during the transit. He may initially designate visual bearings as the primary fix method only to have a sudden storm or fog obscure the visual NAVAIDS. If he shifts the primary fix means to radar, he has a track history of the correlation between radar and visual fixes. Finally, the piloting team often must shift charts several times during the transit. When the old chart is taken off the plotting table and before the new chart is secured, there is a period of time when no chart is in use. Maintaining a secondary plot eliminates this complication. Ensure the secondary plot is not shifted prior to getting the new primary plot chart down on the chart table. In this case, there will always be a chart available on which to pilot. Do not consider the primary chart shifted until the new chart is properly secured and the plotter has transferred the last fix from the original chart onto the new chart.

- **Satellite Navigation Operator:** This operator normally works for the secondary plot supervisor. GPS accuracy with Selective Availability (SA) on is not sufficient for navigating restricted waters; but with SA off, GPS can support harbor navigation, in which case it should be considered as only one aid to navigation, not as a substitute for the entire process. If the team loses visual bearings in the channel and no radar NAVAIDS are available, GPS may be the most accurate fix source available. The navigator must have some data on the comparison between satellite positions and visual positions over the history of the passage to use satellite positions effectively. The only way to obtain this data is to plot satellite positions and compare these positions to visual positions throughout the harbor passage.

- **Fathometer Operator:** Run the fathometer continuously and station an operator to monitor it. Do not rely on audible alarms to key your attention to this critically important piloting tool. The fathometer operator must know the warning and danger soundings for the area the vessel is transiting. Most fathometers can display either total depth of water or depth under the keel. Set the fathometer to display depth under the keel. The navigator must check the sounding at each fix and compare that value to the charted sounding. A discrepancy between these values is cause for immediate action to take another fix and check the ship's position.

810. Harbor Approach (Inbound Vessels Only)

The piloting team must make the transition from coastal navigation to piloting smoothly as the vessel approaches restricted waters. There is no rigid demarcation between coastal navigation and piloting. Often visual NAVAIDS are visible miles from shore where Loran and GPS are easier to use. The navigator should take advantage of this overlap when approaching the harbor. Plotting Loran, GPS, and visual fixes concurrently ensures that the piloting team has correctly identified NAVAIDS and that the different types of systems are in agreement. Once the vessel is close enough to the shore such that sufficient NAVAIDS (at least three with sufficient bearing spread) become visible, the navigator should order visual bearings only for the primary plot and shift all other fixes to the secondary plot, unless the decision has been made to proceed with ECDIS as the primary system.

Take advantage of the coastal navigation and piloting overlap to shorten the fix interval gradually. The navigator must use his judgment in adjusting fix intervals. If the ship is steaming inbound directly towards the shore, set a fix interval such that two fix intervals lie between the vessel and the nearest danger. Upon entering restricted waters, the piloting team should be plotting visual fixes at three minute intervals.

Commercial vessels with GPS and/or Loran C, planning the harbor transit with a pilot, will approach a coast differently. The transition from ocean to coastal to harbor approach navigation will proceed as visual aids and radar targets appear and are plotted. With GPS or ECDIS operating and a waypoint set at the pilot station, only a few fixes are necessary to verify that the GPS position is correct. Once the pilot is aboard, the captain/pilot team may elect to navigate visually, depending on the situation.

TAKING FIXES WHILE PILOTING

Safe navigation while piloting requires frequent fixing of the ship's position. If ECDIS is the primary navigation system in use, this process is automatic, and the role of the navigator is to monitor the progress of the vessel, cross-check the position occasionally, and be alert for any indication that the system is not operating optimally.

If an ECS is in use, it should be considered only a supplement to the paper navigation plot, which legally must still be maintained. As long as the manual plot and the ECS plot are in agreement, the ECS is a valuable tool which shows the navigator where the ship is at any instant, not two or three minutes ago when the last fix was taken. It cannot legally take the place of the paper chart and the manual plot, but it can provide an additional measure of assurance that the ship is in safe water and alert the navigator to a developing dangerous situation before the next round of bearings or ranges.

The next several articles will discuss the three major manual methods used to fix a ship's position when piloting: crossing lines of position, copying satellite or Loran data, or advancing a single line of position. Using one method does not exclude using other methods. The navigator must obtain as much information as possible and employ as many of these methods as necessary.

811. Types of Fixes

While the intersection of two LOP's constitutes a **fix** under one definition, and only an estimated position by another, the prudent navigator will always use at least three LOP's if they are available, so that an error is apparent if they don't meet in a point. Some of the most commonly used methods of obtaining LOP's are discussed below:

- **Fix by Bearings:** The navigator can take and plot bearings from two or more charted objects. This is the most common and often the most accurate way to fix a vessel's position. Bearings may be taken directly to charted objects, or tangents of points of land. See Figure 811a. The intersection of these lines constitutes a fix. A position taken by bearings to buoys should not be considered a fix, but an estimated position (EP), because buoys swing about their watch circle and may be out of position.

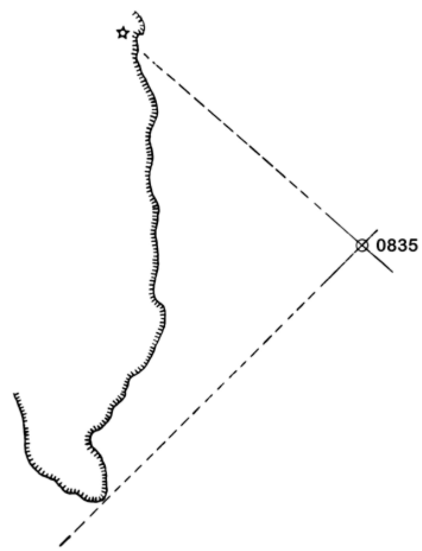


Figure 811a. A fix by two bearing lines.

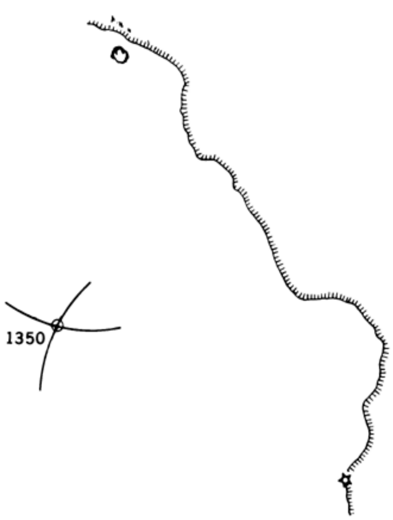


Figure 811b. A fix by two radar ranges.

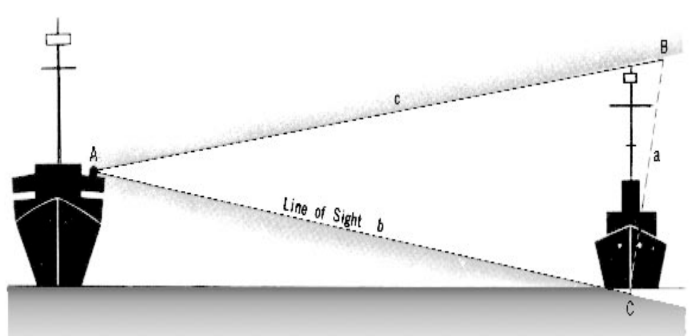


Figure 811c. Principle of stadimeter operation.

- **Fix by Ranges:** The navigator can plot a fix consisting of the intersection of two or more range arcs from charted objects. He can obtain an object's range in several ways:

  1. **Radar Ranges:** See Figure 811b. The navigator may take ranges to two fixed objects. The intersection of the range arcs constitutes a fix. He can plot ranges from any point on the radar scope which he can correlate on his chart. Remember that the shoreline of low-lying land may move many yards in an area of large tidal range, and swampy areas may be indistinct.

  2. **Stadimeter Ranges:** Given a known height of a NAVAID, one can use a stadimeter to determine its range. See Figure 811c for a representation of the geometry involved. Generally, stadimeters contain a height scale on which is set the height of the object. The observer then directs his line of sight through the stadimeter to the base of the object being observed. Finally, he adjusts the stadimeter's range index until the object's top reflection is "brought down" to the visible horizon. Read the object's range off the range index.

  3. **Sextant Vertical Angles:** Measure the vertical angle from the top of the NAVAID to the waterline below the NAVAID. Enter Table 16 to determine the distance of the NAVAID. The navigator must know the height of the NAVAID above sea level to use this table; it can be found in the *Light List*.

  4. **Sonar Ranges:** If the vessel is equipped with a sonar suite, the navigator can use sonar echoes to determine ranges to charted underwater objects. It may take some trial and error to set the active signal strength at a value that will give a strong return and still not cause excessive reverberation. Check local harbor restrictions on energizing active sonar. Avoid active sonar transmissions in the vicinity of divers.

- **Fix by Bearing and Range:** This is a hybrid fix of LOP's from a bearing and range to a single object. The radar is the only instrument that can give simultaneous range and bearing information to the same object. (A sonar system can also provide bearing and range information, but sonar bearings are far too inaccurate to use in piloting.) Therefore, with the radar, the navigator can obtain an instantaneous fix from only one NAVAID. This unique fix is shown in Figure 811d. This makes the radar an extremely useful tool for the piloting team. The radar's characteristics make it much more accurate determining range than determining bearing; therefore, two radar ranges are preferable to a radar range and bearing.

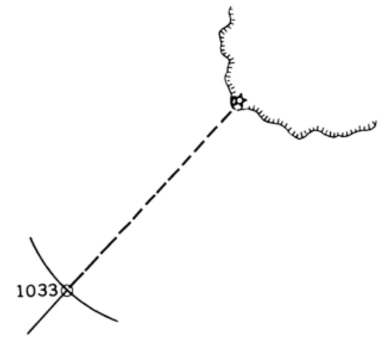


Figure 811d. A fix by range and bearing of a single object.

- **Fix by Range Line and Distance:** When the vessel comes in line with a range, plot the bearing to the range (while checking compass error in the bargain) and cross this LOP with a distance from another NAVAID. Figure 811e shows this fix.

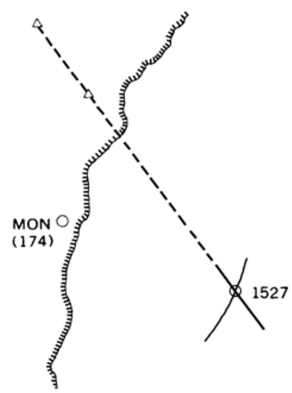


Figure 811e. A fix by a range and distance.

812. The Running Fix

When only one NAVAID is available from which to obtain bearings, use a technique known as the **running fix**. Use the following method:

1. Plot a bearing to a NAVAID (LOP 1).
2. Plot a second bearing to a NAVAID (either the same NAVAID or a different one) at a later time (LOP 2).
3. Advance LOP 1 to the time when LOP 2 was taken.
4. The intersection of LOP 2 and the advanced LOP 1 constitute the running fix.

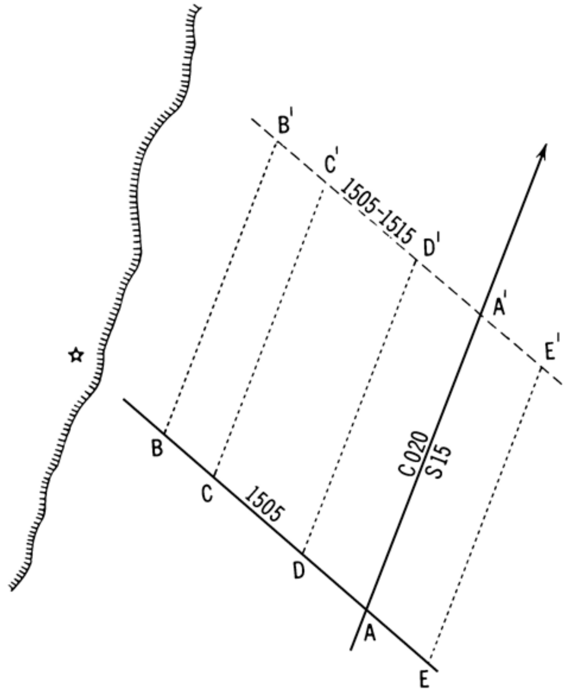


Figure 812a. *Advancing a line of position.*

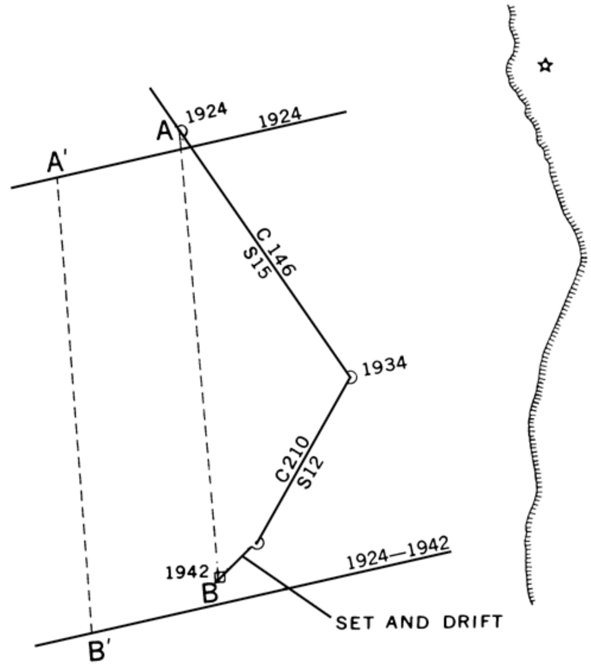


Figure 812b. Advancing a line of position with a change in course and speed, allowing for set and drift.

Figure 812a represents a ship proceeding on course 020°, speed 15 knots. At 1505, the plotter plots an LOP to a lighthouse bearing 310°. The ship can be at any point on this 1505 LOP. Some possible points are represented as points A, B, C, D, and E in Figure 812a. Ten minutes later the ship will have traveled 2.5 miles in direction 020°. If the ship was at A at 1505, it will be at A' at 1515. However, if the position at 1505 was B, the position at 1515 will be B'. A similar relationship exists between C and C', D and D', E and E'. Thus, if any point on the original LOP is moved a distance equal to the distance run in the direction of the motion, a line through this point parallel to the original line of position represents all possible positions of the ship at the later time. This process is called **advancing** a line of position. Moving a line back to an earlier time is called **retiring** a line of position.

When advancing a line of position, account for course changes, speed changes, and set and drift between the two bearing lines. Three methods of advancing an LOP are discussed below:

Method 1: See Figure 812b. To advance the 1924 LOP to 1942, first apply the best estimate of set and drift to the 1942 DR position and label the resulting position point B. Then, measure the distance between the dead reckoning position at 1924 (point A) and point B. Advance the LOP a distance equal to the distance between points A and B. Note that LOP A'B' is in the same direction as line AB.

Method 2: See Figure 812c. Advance the NAVAIDS position on the chart for the course and distance traveled by the vessel and draw the line of position from the NAVAIDS advanced position. This is the most satisfactory method for advancing a circle of position.

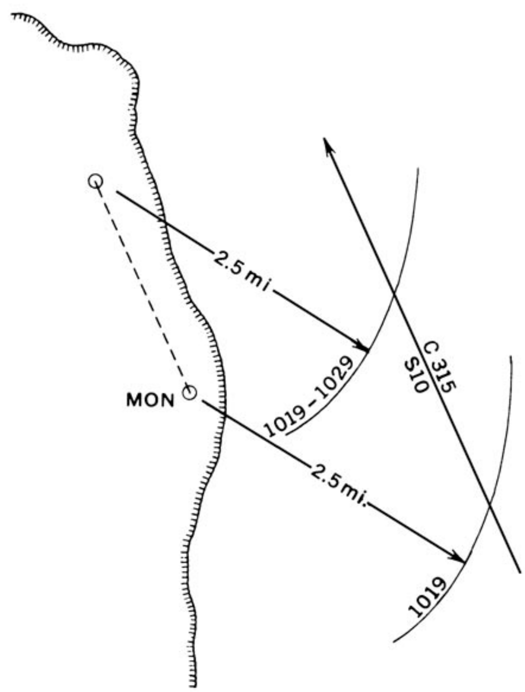


Figure 812c. Advancing a circle of position.

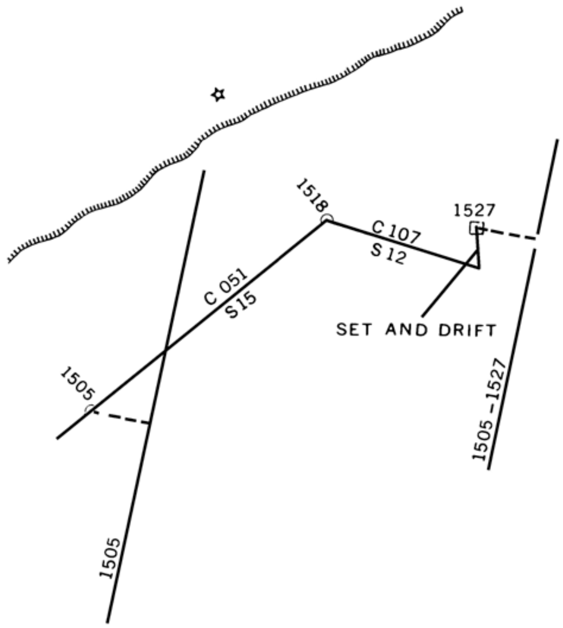


Figure 812d. Advancing a line of position by its relation to the dead reckoning.

Method 3: See Figure 812d. To advance the 1505 LOP to 1527, first draw a correction line from the 1505 DR position to the 1505 LOP. Next, apply a set and drift correction to the 1527 DR position. This results in a 1527 estimated position (EP). Then, draw from the 1527 EP a correction line of the same length and direction as the one drawn from the 1505 DR to the 1505 LOP. Finally, parallel the 1505 bearing to the end of the correction line as shown.

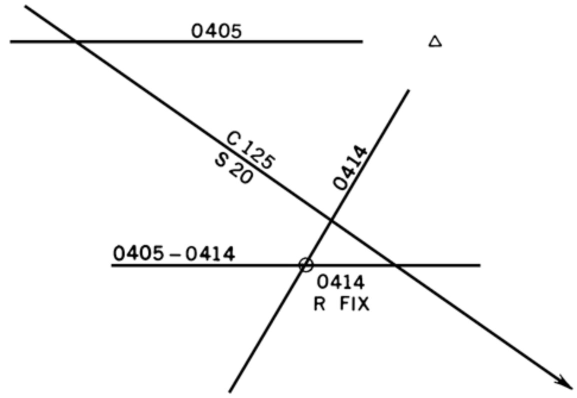


Figure 812e. A running fix by two bearings on the same object.

Label an advanced line of position with both the time of observation and the time to which the line is adjusted.

Figure 812e through Figure 812g demonstrate three running fixes. Figure 812e illustrates the case of obtaining a running fix with no change in course or speed between taking two bearings on the same NAVAID. Figure 812f illustrates a running fix with changes in a vessel's course and speed between taking two bearings on two different objects. Finally, Figure 812g illustrates a running fix obtained by advancing range circles of position using the second method discussed above.

PILOTING PROCEDURES

The previous section discussed the methods for fixing the ship's position. This section discusses integrating the manual fix methods discussed above, and the use of the fathometer, into a piloting procedure. The navigator must develop his piloting procedure to meet several requirements. He must obtain enough information to fix the position of the vessel without question. He must also plot and evaluate this information. Finally, he must relay his evaluations and recommendations to the vessel's conning officer. This section examines some considerations to ensure the navigator accomplishes all these requirements quickly and effectively. Of course, if ECDIS is the primary plot, manual methods as discussed here are for backup use.

813. Fix Type and Fix Interval

The preferred piloting fix is taken from visual bearings from charted fixed NAVAIDS. Plot visual bearings on the primary plot and plot all other fixes on the secondary plot. If poor visibility obscures visual NAVAIDS, shift to radar piloting on the primary plot. If neither visual or radar piloting is available, consider standing off until the visibility improves.

The interval between fixes in restricted waters should usually not exceed three minutes. Setting the fix interval at three minutes optimizes the navigator's ability to assimilate and evaluate all available information. He must relate it to charted navigational hazards and to his vessel's intended track. It should take a well trained plotting team no more than 30 seconds to measure, record, and plot three bearings to three separate NAVAIDS. The navigator should spend the majority of the fix interval time interpreting the information, evaluating the navigational situation, and making recommendations to the conning officer.

If three minutes goes by without a fix, inform the captain and try to plot a fix as soon as possible. If the delay was caused by a loss of visibility, shift to radar piloting. If the delay was caused by plotting error, take another fix. If the navigator cannot get a fix down on the plot for several more minutes, consider slowing or stopping the ship until its position can be fixed. Never continue a passage through

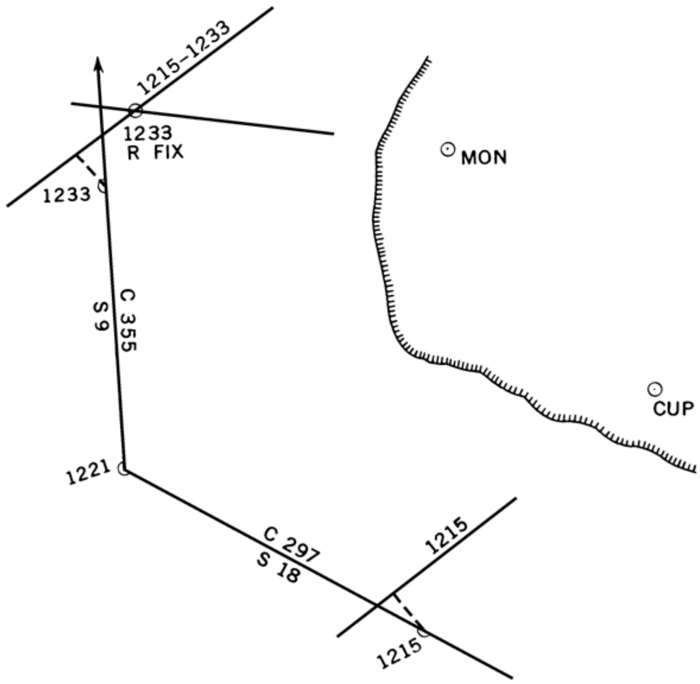


Figure 812f. A running fix with a change of course and speed between observations on separate landmarks.

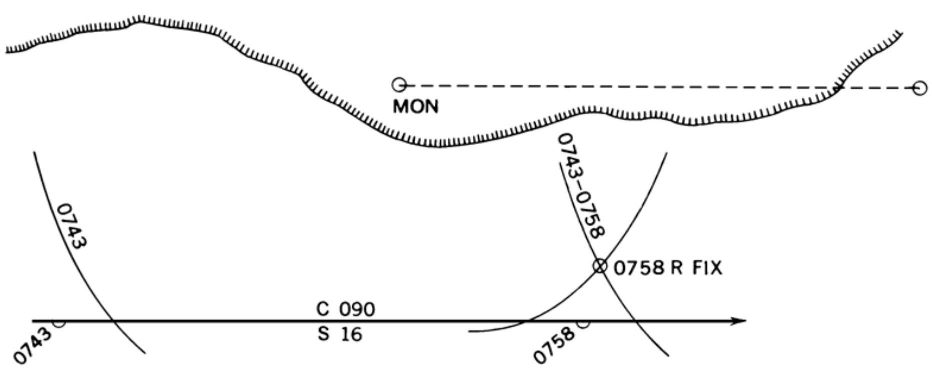


Figure 812g. A running fix by two circles of position.

restricted waters if the vessel's position is uncertain.

The secondary plot supervisor should maintain the same fix interval as the primary plot. Usually, this means he should plot a radar fix every three minutes. He should plot other fix sources (Loran and GPS fixes, for example) at an interval sufficient for making meaningful comparisons between fix sources. Every third fix interval, he should pass a radar fix to the primary plot for comparison with the visual fix. He should inform the navigator how well all the fix sources plotted on the secondary plot are tracking.

814. The Piloting Routine

Following a cyclic routine ensures the timely and efficient processing of data and forms a smoothly functioning piloting team. It quickly yields the information which the navigator needs to make informed recommendations to the conning officer and captain.

Repeat this routine at each fix interval beginning when the ship gets underway until it clears the harbor (outbound) or when the ship enters the harbor until it is moored (inbound).

The routine consists of the following steps:

1. Take, plot and label a fix.
2. Calculate set and drift from the DR position.
3. Reset the DR from the fix and DR two fixes ahead.

- **Plotting the Fix:** This involves coordination between the navigator, bearing taker(s), recorder, and plotter.

The navigator will call for each fix at the DR time. The bearing taker must measure his bearings as quickly as possible, beam bearings first, fore and aft last, on the navigator's mark. The recorder will write the bearings in the book, and the plotter will plot them immediately.

- **Labeling the Fix:** The plotter should clearly mark a visual fix with a circle or an electronic fix with a triangle. Clearly label the time of each fix. A visual running fix should be circled, marked "R Fix" and labeled with the time of the second LOP. Keep the chart neat and uncluttered when labeling fixes.

- **Dead Reckoning Two Fix Intervals Ahead:** After labeling the fix, the plotter should dead reckon the fix position ahead two fix intervals. The navigator should carefully check the area marked by this DR for any navigational hazards. If the ship is approaching a turn, update the turn bearing as discussed in Article 802.

- **Calculate Set and Drift at Every Fix:** Calculating set and drift is covered in Chapter 7. Calculate these values at every fix and inform the captain and conning officer. Compare the actual values of set and drift with the predicted values from the current graph discussed in Article 804. Evaluate how the current is affecting the vessel's position in relation to the track and recommend courses and speeds to regain the planned track. Because the navigator can determine set and drift only when comparing fixes and DR's plotted for the same time, take fixes exactly at the times for which a DR has been plotted. Repeat this routine at each fix interval beginning when the ship gets underway until it clears the harbor (outbound) or when the ship enters the harbor until she is moored (inbound).

- **Piloting Routine When Turning:** Modify the cyclic routine slightly when approaching a turn. Adjust the fix interval so that the plotting team has a fix plotted approximately one minute before a scheduled turn. This gives the navigator sufficient time to evaluate the position in relation to the planned track, DR ahead to the slide bar to determine a new turn bearing, relay the new turn bearing to the conning officer, and then monitor the turn bearing to mark the turn.

Approximately 30 seconds before the time to turn, train the alidade on the turn bearing NAVAID. Watch the bearing of the NAVAID approach the turn bearing. About 1° away from the turn bearing, announce to the conning officer: "Stand by to turn." Slightly before the turn bearing is indicated, report to the conning officer: "Mark the turn." Make this report slightly before the bearing is reached because it takes the conning officer a finite amount of time to acknowledge the report and order the helmsman to put over the rudder. Additionally, it takes a finite amount of time for the helmsman to turn the rudder and for the ship to start to turn. If the navigator waits until the turn bearing is indicated to report the turn, the ship will turn too late.

Once the ship is steady on the new course, immediately take another fix to evaluate the vessel's position in relation to the track. If the ship is not on the track after the turn, calculate and recommend a course to the conning officer to regain the track.

815. Using the Fathometer

Use the fathometer to determine whether the depth of water under the keel is sufficient to prevent the ship from grounding and to check the actual water depth with the charted water depth at the fix position. The navigator must compare the charted sounding at every fix position with the fathometer reading and report to the captain any discrepancies. Taking continuous soundings in restricted waters is mandatory.

See the discussion of calculating the warning and danger soundings in Article 802. If the warning sounding is received, then slow the ship, fix the ship's position more frequently, and proceed with extreme caution. Ascertain immediately where the ship is in the channel; if the minimum expected sounding was noted correctly, the warning sounding indicates the vessel may be leaving the channel and standing into shoal water. Notify the vessel's captain and conning officer immediately.

If the danger sounding is received, take immediate action to get the vessel back to deep water. Reverse the engines and stop the vessel's forward movement. Turn in the direction of the deepest water before the vessel loses steerageway. Consider dropping the anchor to prevent the ship from drifting aground. The danger sounding indicates that the ship has left the channel and is standing into immediate danger. It requires immediate corrective action by the ship's conning officer, navigator, and captain to avoid disaster.

Many underwater features are poorly surveyed. If a fathometer trace of a distinct underwater feature can be obtained along with accurate position information, send the fathometer trace and related navigational data to NIMA for entry into the Digital Bathymetric Data Base.

PILOTING TO AN ANCHORAGE

816. Choosing an Anchorage

Most U.S. Navy vessels receive instructions in their movement orders regarding the choice of anchorage. Merchant ships are often directed to specific anchorages by harbor authorities. However, lacking specific guidance, the mariner should choose his anchoring positions using the following criteria:

- **Depth of Water:** Choose an area that will provide sufficient depth of water through an entire range of tides. Water too shallow will cause the ship to go aground, and water too deep will allow the anchor to drag.

- **Type of Bottom:** Choose the bottom that will best hold the anchor. Avoid rocky bottoms and select sandy or muddy bottoms if they are available.

- **Proximity to navigational Hazards:** Choose an anchorage as far away as possible from known navigational hazards.

- **Proximity to Adjacent Ships:** Anchor well away from adjacent vessels; ensure that another vessel will not swing over your own anchor on a current or wind shift.

- **Proximity to Harbor Traffic Lanes:** Anchor clear of traffic lanes and ensure that the vessel will not swing into the channel on a current or wind shift.

- **Weather:** Choose an area with the weakest winds and currents.

- **Availability of NAVAIDS:** Choose an anchorage with several NAVAIDS available for monitoring the ship's position when anchored.

817. Navigational Preparations for Anchoring

It is usually best to follow an established procedure to ensure an accurate positioning of the anchor, even when anchoring in an open roadstead. The following procedure is representative. See Figure 817.

Locate the selected anchoring position on the chart. Consider limitations of land, current, shoals, and other vessels when determining the direction of approach. Where conditions permit, make the approach heading into the current. Close observation of any other anchored vessels will provide clues as to which way the ship will lie to her anchor. If wind and current are strong and from different directions, ships will lie to their anchors according to the balance between these two forces and the draft and trim of each ship. Different ships may lie at different headings in the same anchorage depending on the balance of forces affecting them.

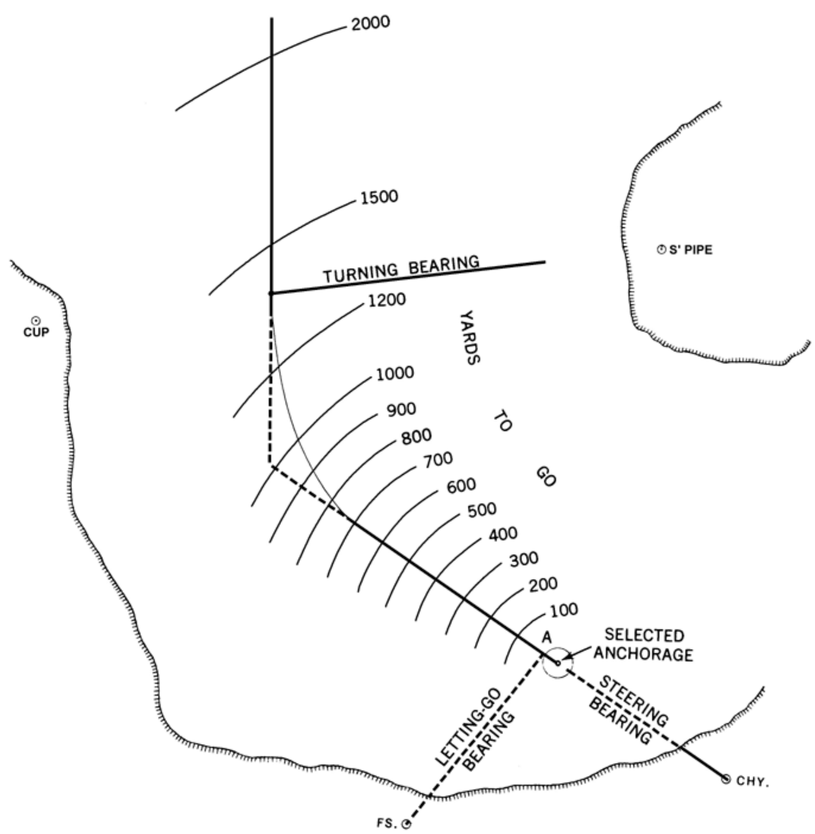


Figure 817. Anchoring.

Approach from a direction with a prominent NAVAID, preferably a range, available dead ahead to serve as a steering guide. If practicable, use a straight approach of at least 1200 yards to permit the vessel to steady on the required course. Draw in the approach track, allowing for advance and transfer during any turns. In Figure 817, the chimney was selected as this steering bearing. A turn range may also be used if a radar-prominent object can be found directly ahead or astern.

Next, draw a circle with the selected position of the anchor as the center, and with a radius equal to the distance between the hawsepipe and pelorus, alidade, or periscope used for measuring bearings. This circle is marked "A" in Figure 817. The intersection of this circle and the approach track is the position of the vessel's bearing-measuring instrument at the moment of letting the anchor go. Select a NAVAID which will be on the beam when the vessel is at the point of letting go the anchor. This NAVAID is marked "FS" in Figure 817. Determine what the bearing to that object will be when the ship is at the drop point and measure this bearing to the nearest 0.1°T. Label this bearing as the letting go bearing.

During the approach to the anchorage, plot fixes at frequent intervals. The navigator must advise the conning officer of any tendency of the vessel to drift from the desired track. The navigator must frequently report to the conning officer the distance to go, permitting adjustment of the speed so that the vessel will be dead in the water or have very slight sternway when the anchor is let go. To aid in determining the distance to the drop point, draw and label a number of range arcs as shown in Figure 817 representing distances to go to the drop point.

At the moment of letting the anchor go, take a fix and plot the vessel's exact position on the chart. This is important in the construction of the swing and drag circles discussed below. To draw these circles accurately, determine the position of the vessel at the time of letting go the anchor as accurately as possible.

Veer the anchor chain to a length equal to five to seven times the depth of water at the anchorage. The exact amount to veer is a function of both vessel type and severity of weather expected at the anchorage. When calculating the scope of anchor chain to veer, take into account the maximum height of tide.

Once the ship is anchored, construct two separate circles around the ship's position when the anchor was dropped. These circles are called the **swing circle** and the **drag circle**. Use the swing circle to check for navigational hazards and use the drag circle to ensure the anchor is holding.

The swing circle's radius is equal to the sum of the ship's length and the scope of the anchor chain released. This represents the maximum arc through which a ship can swing while riding at anchor if the anchor holds. Examine this swing circle carefully for navigational hazards, interfering contacts, and other anchored shipping. Use the lowest height of tide expected during the anchoring period when checking inside the swing circle for shoal water.

The drag circle's radius equals the sum of the hawsepipe to pelorus distance and the scope of the chain released. Any bearing taken to check on the position of the ship should, if the anchor is holding, fall within the drag circle. If a fix falls outside of that circle, then the anchor is dragging. If the vessel has a GPS or Loran system with an off-station alarm, set the alarm at the drag circle radius, or slightly more.

In some cases, the difference between the radii of the swing and drag circles will be so small that, for a given chart scale, there will be no difference between the circles when plotted. If that is the case, plot only the swing circle and treat that circle as both a swing and a drag circle. On the other hand, if there is an appreciable difference in radii between the circles when plotted, plot both on the chart. Which method to use falls within the sound judgment of the navigator.

When determining if the anchor is holding or dragging, the most crucial period is immediately after anchoring. Fixes should be taken frequently, at least every three minutes, for the first thirty minutes after anchoring. The navigator should carefully evaluate each fix to determine if the anchor is holding. If the anchor is holding, the navigator can then increase the fix interval. What interval to set falls within the judgment of the navigator, but the interval should not exceed 30 minutes. If an ECDIS, Loran, or GPS is available, use its off-station alarm feature for an additional safety factor.

NAVIGATIONAL ASPECTS OF SHIP HANDLING

818. Effects Of Banks, Channels, and Shallow Water

A ship moving through shallow water experiences pronounced effects from the proximity of the nearby bottom. Similarly, a ship in a channel will be affected by the proximity of the sides of the channel. These effects can easily cause errors in piloting which lead to grounding. The effects are known as **squat**, **bank cushion**, and **bank suction**. They are more fully explained in texts on shiphandling, but certain navigational aspects are discussed below.

Squat is caused by the interaction of the hull of the ship, the bottom, and the water between. As a ship moves through shallow water, some of the water it displaces rushes under the vessel to rise again at the stern. This causes a venturi effect, decreasing upward pressure on the hull. Squat makes the ship sink deeper in the water than normal and slows the vessel. The faster the ship moves through shallow water, the greater is this effect; groundings on both charted and uncharted shoals and rocks have occurred because of this phenomenon, when at reduced speed the ship could have safely cleared the dangers. When navigating in shallow water, the navigator must reduce speed to avoid squat. If bow and stern waves nearly perpendicular the direction of travel are noticed, and the vessel slows with no change in shaft speed, squat is occurring. Immediately slow the ship to counter it. Squatting occurs in deep water also, but is more pronounced and dangerous in shoal water. The large waves generated by a squatting ship also endanger shore facilities and other craft.

Bank cushion is the effect on a ship approaching a steep underwater bank at an oblique angle. As water is forced into the narrowing gap between the ship's bow and the shore, it tends to rise or pile up on the landward side, causing the ship to sheer away from the bank.

Bank suction occurs at the stern of a ship in a narrow channel. Water rushing past the ship on the landward side exerts less force than water on the opposite or open water side. This effect can actually be seen as a difference in draft readings from one side of the vessel to the other, and is similar to the venturi effect seen in squat. The stern of the ship is forced toward the bank. If the ship gets too close to the bank, it can be forced sideways into it. The same effect occurs between two vessels passing close to each other.

These effects increase as speed increases. Therefore, in shallow water and narrow channels, navigators should decrease speed to minimize these effects. Skilled pilots may use these effects to advantage in particular situations, but the average mariner's best choice is slow speed and careful attention to piloting.

ADVANCED PILOTING TECHNIQUES

819. Assuming Current Values to Set Safety Margins for Running Fixes

Current affects the accuracy of a running fix. Consider, for example, the situation of an unknown head current. In Figure 819a, a ship is proceeding along a coast, on course 250 ° speed 12 knots. At 0920 light A bears 190°, and at 0930 it bears 143°. If the earlier bearing line is advanced a distance of 2 miles (10 minutes at 12 knots) in the direction of the course, the running fix is as shown by the solid lines. However, if there is a head current of 2 knots, the ship is making good a speed of only 10 knots, and in 10 minutes will travel a distance of only 1 $^{2}/_{3}$ miles. If the first bearing line is advanced this distance, as shown by the broken line, the actual position of the ship is at B. This actual position is nearer the shore than the running fix actually plotted. A following current, conversely, would show a position too far from the shore from which the bearing was measured.

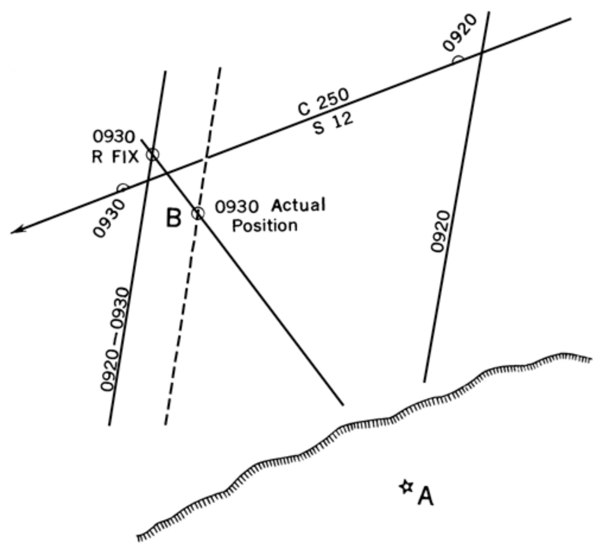


Figure 819a. Effect of a head current on a running fix.

If the navigator assumes a following current when advancing his LOP, the resulting running fix will plot further from the NAVAID than the vessel's actual position. Conversely, if he assumes a head current, the running fix will plot closer to the NAVAID than the vessel's actual position. To ensure a margin of safety when plotting running fix bearings to a NAVAID on shore, always assume the current slows a vessel's speed over ground. This will cause the running fix to plot closer to the shore than the ship's actual position.

When taking the second running fix bearing from a different object, maximize the speed estimate if the second object is on the same side and farther forward, or on the opposite side and farther aft, than the first object was when observed.

All of these situations assume that danger is on the same side as the object observed first. If there is either a head or following current, a series of running fixes based upon a number of bearings of the same object will plot in a straight line parallel to the course line, as shown in Figure 819b. The plotted line will be too close to the object observed if there is a head current and too far out if there is a following current. The existence of the current will not be apparent unless the actual speed over the ground is known. The position of the plotted line relative to the dead reckoning course line is not a reliable guide.

820. Determining Track Made Good by Plotting Running Fixes

A current oblique to a vessel's course will also result in an incorrect running fix position. An oblique current can be detected by observing and plotting several bearings of the same object. The running fix obtained by advancing one

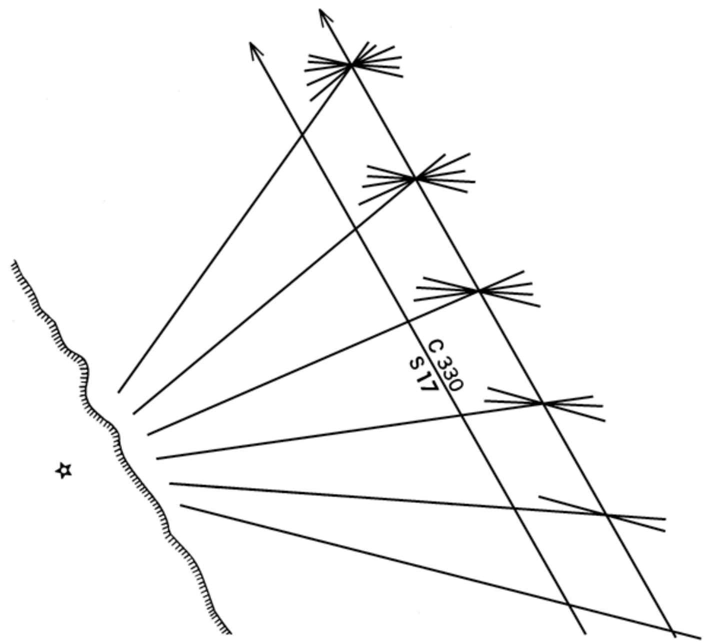


Figure 819b. A number of running fixes with a following current.

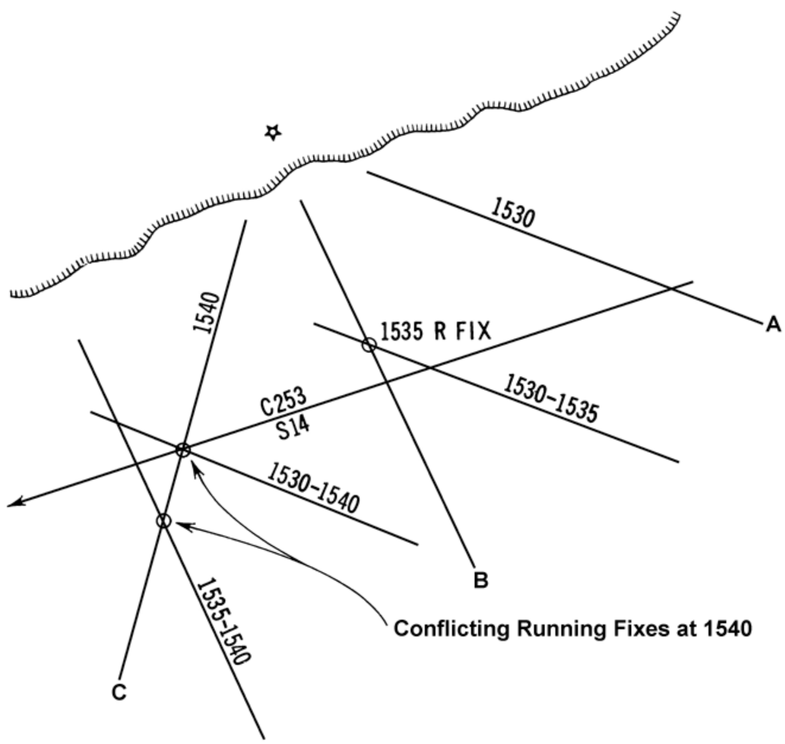


Figure 820a. Detecting the existence of an oblique current, by a series of running fixes.

bearing line to the time of the next one will not agree with the running fix obtained by advancing an earlier line. See Figure 820a. If bearings A, B, and C are observed at five-minute intervals, the running fix obtained by advancing B to the time of C will not be the same as that obtained by advancing A to the time of C, as shown in Figure 820a.

Whatever the current, the navigator can determine the direction of the track made good (assuming constant current and constant course and speed). Observe and plot three bearings of a charted object O. See Figure 820b. Through O draw XY in any direction. Using a convenient scale, determine points A and B so that OA and OB are proportional to the time intervals between the first and second bearings and the second and third bearings, respectively. From A and B draw lines parallel to the second bearing line, intersecting the first and third bearing lines at C and D, respectively. The direction of the line from C and D is the track made good.

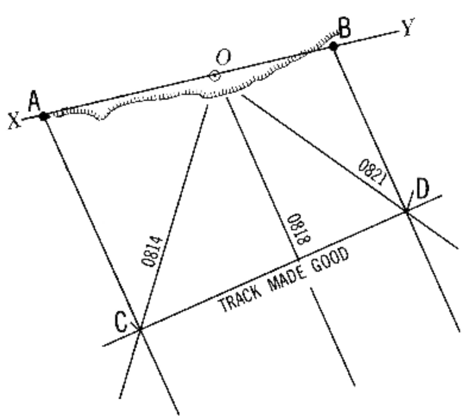


Figure 820b. Determining the track made good.

The distance of the line CD in Figure 820b from the track is in error by an amount proportional to the ratio of the speed made good to the speed assumed for the solution. If a good fix (not a running fix) is obtained at some time before the first bearing for the running fix, and the current has not changed, the track can be determined by drawing a line from the fix, in the direction of the track made good. The intersection of the track with any of the bearing lines is an actual position.

821. Fix by Distance of an Object by Two Bearings (Table 18)

Geometrical relationships can define a running fix. In Figure 821, the navigator takes a bearing on NAVAID D. The bearing is expressed as degrees right or left of course. Later, at B, he takes a second bearing to D; similarly, he takes a bearing at C, when the landmark is broad on the beam. The navigator knows the angles at A, B, and C and the distance run between points. The various triangles can be solved using Table 18. From this table, the navigator can calculate the lengths of segments AD, BD, and CD. He knows the range and bearing; he can then plot an LOP. He can then advance these LOP's to the time of taking the CD bearing to plot a running fix.

Enter the table with the difference between the course and first bearing (angle BAD in Figure 821) along the top of the table and the difference between the course and second bearing (angle CBD) at the left of the table. For each pair of angles listed, two numbers are given. To find the distance from the landmark at the time of the second bearing (BD), multiply the distance run between bearings (in nautical miles) by the first number from Table 18. To find the distance when the object is abeam (CD), multiply the distance run between A and B by the second number from the table. If the run between bearings is exactly 1 mile, the tabulated values are the distances sought.

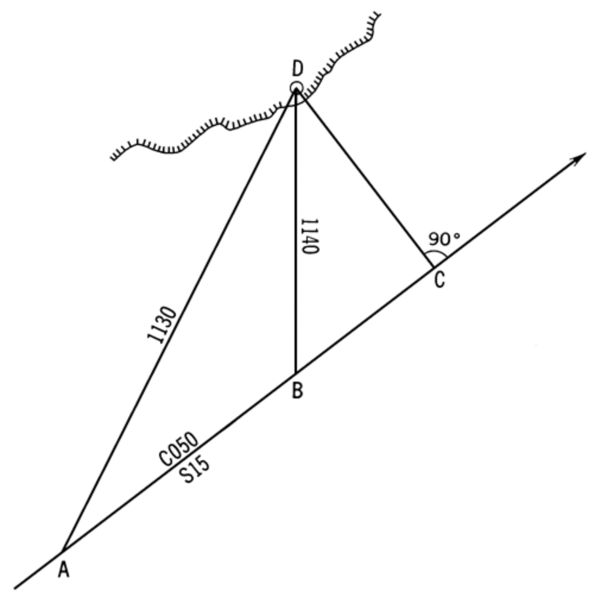


Figure 821. Triangles involved in a Table 18 *running fix.*

Example: *A ship is steaming on course 050°, speed 15 knots. At 1130 a lighthouse bears 024°, and at 1140 it bears 359°.*

Required:

(1) *Distance from the light at 1140.*

(2) *Distance form the light when it is broad on the port beam.*

Solution:

(1) *The difference between the course and the first bearing (050° – 24°) is 26°, and the difference between the course and the second bearing (050° + 360° - 359°) is 51°.*

(2) *From* Table 18*, the two numbers (factors are 1.04 and 0.81, found by interpolation.*

(3) *The distance run between bearings is 2.5 miles (10 minutes at 15 knots).*

(4) T*he distance from the lighthouse at the time of the second bearing is 2.5 × 1.04 = 2.6 miles.*

(5) *The distance from the lighthouse when it is broad on the beam is 2.5 × 0.81 = 2.0 miles.*
***Answer:** (1) D 2.6 mi., (2) D 2.0 mi.*

This method yields accurate results only if the helmsman has steered a steady course and the navigator uses the vessel's speed over ground.

MINIMIZING ERRORS IN PILOTING

822. Common Errors

Piloting requires a thorough familiarity with principles involved, constant alertness, and judgment. A study of groundings reveals that the cause of most is a failure to use or interpret available information. Among the more common errors are:

1. Failure to obtain or evaluate soundings
2. Mis-identification of aids to navigation
3. Failure to use available navigational aids effectively
4. Failure to correct charts
5. Failure to adjust a magnetic compass or keep a table of corrections
6. Failure to apply deviation
7. Failure to apply variation
8. Failure to check gyro and magnetic compass readings regularly
9. Failure to keep a dead reckoning plot
10. Failure to plot new information
11. Failure to properly evaluate information
12. Poor judgment
13. Failure to use information in charts and navigational publications
14. Poor navigation team organization
15. Failure to "keep ahead of the vessel"
16. Failure to have backup navigational methods in place
17. Failure to recognize degradation of electronically obtained LOP's or lat./long. positions

Some of the errors listed above are mechanical and some are matters of judgment. Conscientiously applying the principles and procedures of this chapter will go a long way towards eliminating many of the mechanical errors. However, the navigator must guard against the feeling that in following a checklist he has eliminated all sources of error. A navigator's judgment is just as important as his checklists.

823. Minimizing Errors with a Two Bearing Plot

When measuring bearings from two NAVAIDS, the fix error resulting from an error held constant for both observations is minimized if the angle of intersection of the bearings is 90°. If the observer in Figure 823a is located at point T and the bearings of a beacon and cupola are observed and plotted without error, the intersection of the bearing lines lies on the circumference of a circle passing through the beacon, cupola, and the observer. With constant error, the angular difference between the bearings of the beacon and the cupola is not affected. Thus, the angle formed at point F by the bearing lines plotted with constant error is equal to the angle formed at point T by the bearing lines plotted without error. From geometry it is known that angles having their apexes on the circumference of a circle and that are subtended by the same chord are equal. Since the angles at points T and F are equal and the angles are subtended by the same chord, the intersection at point F lies on the circumference of a circle passing through the beacon, cupola, and the observer.

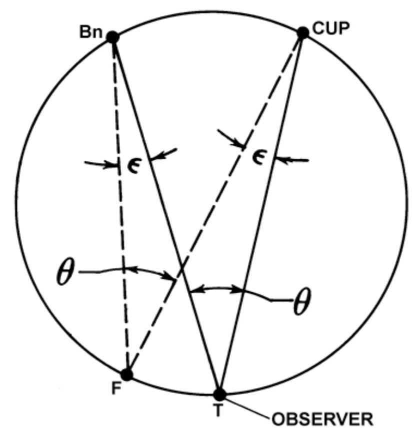


Figure 823a. Two-bearing plot.

Assuming only constant error in the plot, the direction of displacement of the two-bearing fix from the position of the observer is in accordance with the sign (or direction) of the constant error. However, a third bearing is required to determine the direction of the constant error.

Assuming only constant error in the plot, the two-bearing fix lies on the circumference of the circle passing through the two charted objects observed and the observer. The fix error, the length of the chord FT in Figure 823b, depends on the magnitude of the constant error $\in$, the distance between the charted objects, and the cosecant of the angle of cut, angle θ. In Figure 823b,

$$\text{The fix error} = \text{FT} = \frac{\text{BC}\csc\theta}{2}$$

where $\in$ is the magnitude of the constant error, BC is the length of the chord BC, and θ is the angle of the LOP's intersection.

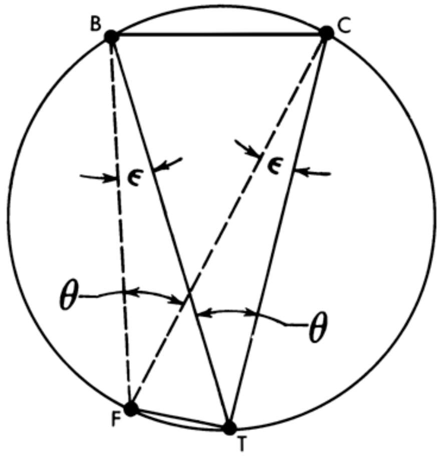


Figure 823b. Two-bearing plot with constant error.

Since the fix error is a function of the cosecant of the angle of intersection, it is least when the angle of intersection is 90°. As illustrated in Figure 823c, the error increases in accordance with the cosecant function as the angle of intersection decreases. The increase in the error becomes quite rapid after the angle of intersection has decreased to below about 30°. With an angle of intersection of 30°, the fix error is about twice that at 90°.

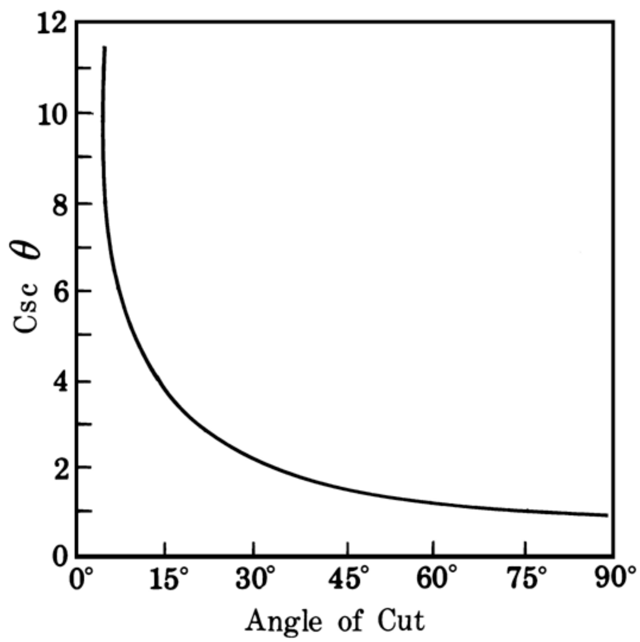


Figure 823c. Error of two-bearing plot.

824. Finding Compass Error by Trial and Error

If several fixes obtained by bearings on three objects produce triangles of error of about the same size, there might be a constant error in observing or plotting the bearings. If applying of a constant error to all bearings results in a pinpoint fix, apply such a correction to all subsequent fixes. Figure 824 illustrates this technique. The solid lines indicate the original plot, and the broken lines indicate each line of position moved 3° in a clockwise direction.

Employ this procedure carefully. Attempt to find and eliminate the error source. The error may be in the gyrocompass, the repeater, or the bearing transmission system. Compare the resulting fix positions with a satellite position, a radar position, or the charted sounding. A high degree of correlation between these three independent positioning systems and an "adjusted" visual fix is further confirmation of a constant bearing error.

TRAINING

825. Piloting Simulators

Civilian piloting training has traditionally been a function of both maritime academies and on-the-job experience. The latter is usually more valuable, because there is no substitute for experience in developing judgment. Military piloting training consists of advanced correspondence courses and formal classroom instruction combined with duties on the bridge. U.S. Navy Quartermasters frequently attend Ship's Piloting and Navigation (SPAN) trainers as a routine segment of shoreside training. Military vessels in general have a much clearer definition of responsibilities, as well as more people to carry them out, than civilian ships, so training is generally more thorough and targeted to specific skills.

Computer technology has made possible the development of computerized **ship simulators**, which allow piloting experience to be gained without risking accidents at sea and without incurring underway expenses. Simulators range from simple micro-computer-based software to a completely equipped ship's bridge with radar, engine controls, 360° horizon views, programmable sea motions, and the capability to simulate almost any navigational situation.

A different type of simulator consists of scale models of ships. The models, actually small craft of about 20-30 feet, have hull forms and power-to-weight ratios similar to various types of ships, primarily supertankers, and the operator pilots the vessel from a position such that his view is from the craft's "bridge." These are primarily used in training pilots and masters in docking maneuvers with exceptionally large vessels.

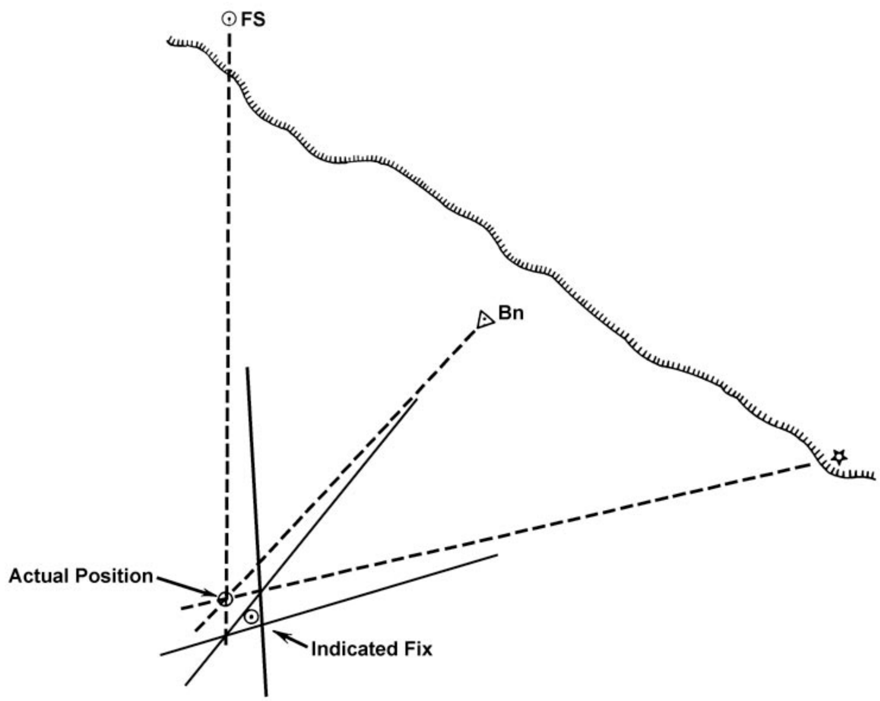


Figure 824. Adjusting a fix for constant error.

The first computer ship simulators came into use in the late 1970s. Several years later the U.S. Coast Guard began accepting a limited amount of simulator time as “sea time” for licensing purposes. They can simulate virtually any conditions encountered at sea or in piloting waters, including land, aids to navigation, ice, wind, fog, snow, rain, and lightning. The system can also be programmed to simulate hydrodynamic effects such as shallow water, passing vessels, current, and tugs.

Virtually any type of vessel can be simulated, including tankers, bulkers, container ships, tugs and barges, yachts, and military vessels. Similarly, any given navigational situation can be modeled, including passage through any chosen harbor, river, or passage, convoy operations, meeting and passing situations at sea and in harbors.

Simulators are used not only to train mariners, but also to test feasibility of port and harbor plans and visual aids to navigation system designs. This allows pilots to “navigate” simulated ships through simulated harbors before construction begins to test the adequacy of channels, turning basins, aids to navigation, and other factors.

A full-capability simulator consists of a ship’s bridge which may have motion and noise/vibration inputs, a programmable visual display system which projects a simulated picture of the area surrounding the vessel in both daylight and night modes, image generators for the various inputs to the scenario such as video images and radar, a central data processor, a human factors monitoring system which may record and videotape bridge activities for later analysis, and a control station where instructors control the entire scenario.

Some simulators are part-task in nature, providing specific training in only one aspect of navigation such as radar navigation, collision avoidance, or night navigation.

While there is no substitute for on-the-job training, simulators are extremely cost effective systems which can be run for a fraction of the cost of an actual vessel. Further, they permit trainees to learn from mistakes with no possibility of an accident, they can model an infinite variety of scenarios, and they permit replay and reassessment of each maneuver.

CHAPTER 9

TIDES AND TIDAL CURRENTS

ORIGINS OF TIDES

900. Introduction

Tides are the periodic motion of the waters of the sea due to changes in the attractive forces of the Moon and Sun upon the rotating Earth. Tides can either help or hinder a mariner. A high tide may provide enough depth to clear a bar, while a low tide may prevent entering or leaving a harbor. Tidal current may help progress or hinder it, may set the ship toward dangers or away from them. By understanding tides and making intelligent use of predictions published in tide and tidal current tables and descriptions in sailing directions, the navigator can plan an expeditious and safe passage through tidal waters.

901. Tide and Current

The rise and fall of tide is accompanied by horizontal movement of the water called tidal current. It is necessary to distinguish clearly between tide and tidal current, for the relation between them is complex and variable. For the sake of clarity mariners have adopted the following definitions: Tide is the vertical rise and fall of the water, and tidal current is the horizontal flow. The tide rises and falls, the tidal current floods and ebbs. The navigator is concerned with the amount and time of the tide, as it affects access to shallow ports. The navigator is concerned with the time, speed, and direction of the tidal current, as it will affect his ship's position, speed, and course.

Tides are superimposed on nontidal rising and falling water levels, caused by weather, seismic events, or other natural forces. Similarly, tidal currents are superimposed upon non-tidal currents such as normal river flows, floods, and freshets.

902. Causes of Tides

The principal tidal forces are generated by the Moon and Sun. The Moon is the main tide-generating body. Due to its greater distance, the Sun's effect is only 46 percent of the Moon's. Observed tides will differ considerably from the tides predicted by equilibrium theory since size, depth, and configuration of the basin or waterway, friction, land masses, inertia of water masses, Coriolis acceleration, and other factors are neglected in this theory. Nevertheless, equilibrium theory is sufficient to describe the magnitude and distribution of the main tide-generating forces across the surface of the Earth.

Newton's universal law of gravitation governs both the orbits of celestial bodies and the tide-generating forces which occur on them. The force of gravitational attraction between any two masses, m_1 and m_2, is given by:

$$F = \frac{Gm_1m_2}{d^2}$$

where d is the distance between the two masses, and G is a constant which depends upon the units employed. This law assumes that m_1 and m_2 are point masses. Newton was able to show that homogeneous spheres could be treated as point masses when determining their orbits.

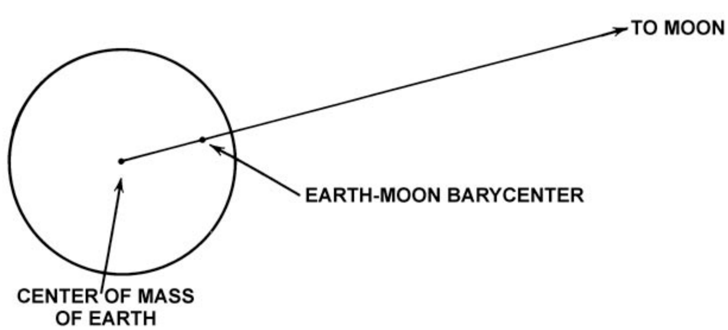


Figure 902a. Earth-Moon barycenter.

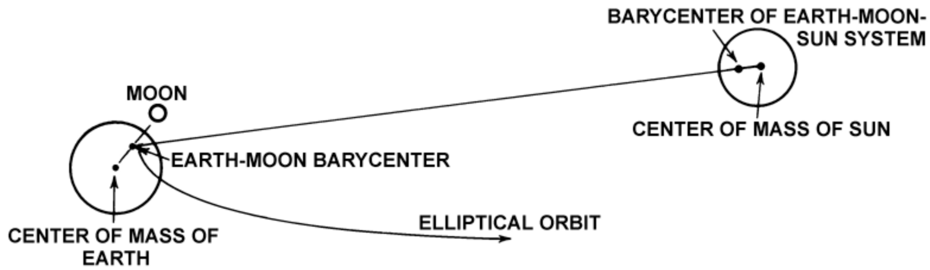


Figure 902b. Orbit of Earth-Moon barycenter (not to scale).

However, when computing differential gravitational forces, the actual dimensions of the masses must be taken into account.

Using the law of gravitation, it is found that the orbits of two point masses are conic sections about the **barycenter** of the two masses. If either one or both of the masses are homogeneous spheres instead of point masses, the orbits are the same as the orbits which would result if all of the mass of the sphere were concentrated at a point at the center of the sphere. In the case of the Earth-Moon system, both the Earth and the Moon describe elliptical orbits about their barycenter if both bodies are assumed to be homogeneous spheres and the gravitational forces of the Sun and other planets are neglected. The Earth-Moon barycenter is located 74/100 of the distance from the center of the Earth to its surface, along the line connecting the Earth's and Moon's centers. See Figure 902a.

Thus the center of mass of the Earth describes a very small ellipse about the Earth-Moon barycenter, while the center of mass of the Moon describes a much larger ellipse about the same barycenter. If the gravitational forces of the other bodies of the solar system are neglected, Newton's law of gravitation also predicts that the Earth-Moon barycenter will describe an orbit which is approximately elliptical about the barycenter of the Sun-Earth-Moon system. This barycentric point lies inside the Sun. See Figure 902b.

903. The Earth-Moon-Sun System

The fundamental tide-generating force on the Earth has two interactive but distinct components. The tide-generating forces are differential forces between the gravitational attraction of the bodies (Earth-Sun and Earth-Moon) and the centrifugal forces on the Earth produced by the Earth's orbit around the Sun and the Moon's orbit around the Earth. Newton's Law of Gravitation and his Second Law of Motion can be combined to develop formulations for the differential force at any point on the Earth, as the direction and magnitude are dependent on where you are on the Earth's surface. As a result of these differential forces, the tide generating forces F_{dm} (Moon) and F_{ds} (Sun) are inversely proportional to the cube of the distance between the bodies, where:

$$F_{dm} = \frac{GM_mR_e}{d_m^3}; \quad F_{ds} = \frac{GM_sR_e}{d_s^3}$$

where M_m is the mass of the Moon and M_s is the mass of the Sun, R_e is the radius of the Earth and d is the distance to the Moon or Sun. This explains why the tide-generating force of the Sun is only 46/100 of the tide-generating force of the Moon. Even though the Sun is much more massive, it is also much farther away.

Using Newton's second law of motion, we can calculate the differential forces generated by the Moon and the Sun affecting any point on the Earth. The easiest calculation is for the point directly below the Moon, known as the **sublunar point**, and the point on the Earth exactly opposite, known as the **antipode**. Similar calculations are done for the Sun.

If we assume that the entire surface of the Earth is covered with a uniform layer of water, the differential forces may be resolved into vectors perpendicular and parallel to the surface of the Earth to determine their effect. See Figure 903a.

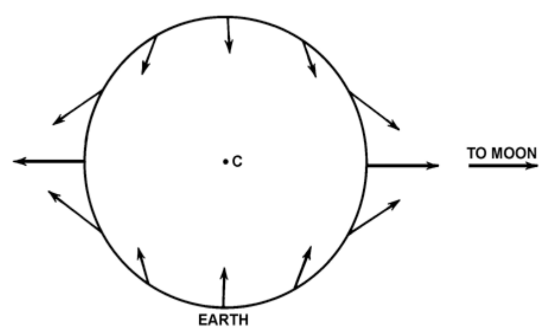


Figure 903a. Differential forces along a great circle connecting the sublunar point and antipode.

The perpendicular components change the mass on which they are acting, but do not contribute to the tidal effect. The horizontal components, parallel to the Earth's surface, have the effect of moving the water in a horizontal

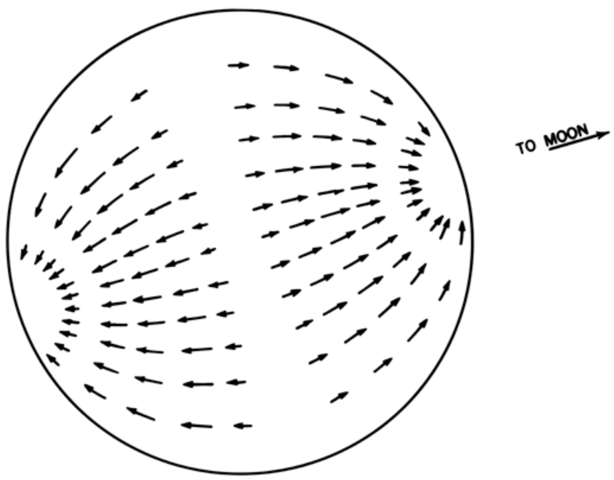


Figure 903b. Tractive forces across the surface of the Earth.

direction toward the sublunar and antipodal points until an equilibrium position is found. The *horizontal* components of the differential forces are the principal tide-generating forces. These are also called **tractive** forces. Tractive forces are zero at the sublunar and antipodal points and along the great circle halfway between these two points. Tractive forces are maximum along the small circles located 45° from the sublunar point and the antipode. Figure 903b shows the tractive forces across the surface of the Earth.

Equilibrium will be reached when a bulge of water has formed at the sublunar and antipodal points such that the tractive forces due to the Moon's differential gravitational forces on the mass of water covering the surface of the Earth are just balanced by the Earth's gravitational attraction (Figure 903c).

Now consider the effect of the rotation of the Earth. If the declination of the Moon is 0°, the bulges will lie on the equator. As the Earth rotates, an observer at the equator will note that the Moon transits approximately every 24 hours and 50 minutes. Since there are two bulges of water on the equator, one at the sublunar point and the other at the antipode, the observer will also see two high tides during this interval with one high tide occurring when the Moon is overhead and another high tide 12 hours 25 minutes later when the observer is at the antipode. He will also experience a low tide between each high tide. The theoretical range of these equilibrium tides at the equator will be less than 1 meter.

In theory, the heights of the two high tides should be equal at the equator. At points north or south of the equator, an observer would still experience two high and two low tides, but the heights of the high tides would not be as great as they are at the equator. The effects of the declination of the Moon are shown in Figure 903d, for three cases, A, B, and C.

A. When the Moon is on the plane of the equator, the forces are equal in magnitude at the two points on the same parallel of latitude and 180° apart in longitude.
B. When the Moon has north or south declination, the forces are unequal at such points and tend to cause an inequality in the two high waters and the two low waters each day.
C. Observers at points X, Y, and Z experience one high tide when the Moon is on their meridian, then another high tide 12 hours 25 minutes later when at X', Y', and Z'. The second high tide is the same at X' as at X. High tides at Y' and Z' are lower than high tides at Y and Z.

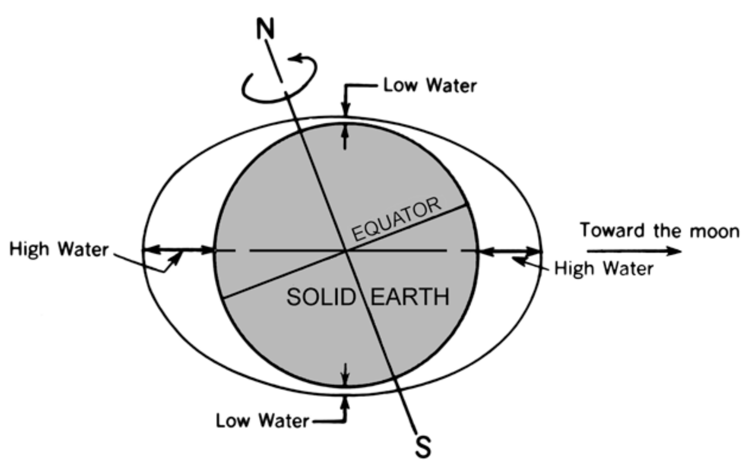


Figure 903c. Theoretical equilibrium configuration due to Moon's differential gravitational forces. One bulge of the water envelope is located at the sublunar point, the other bulge at the antipode.

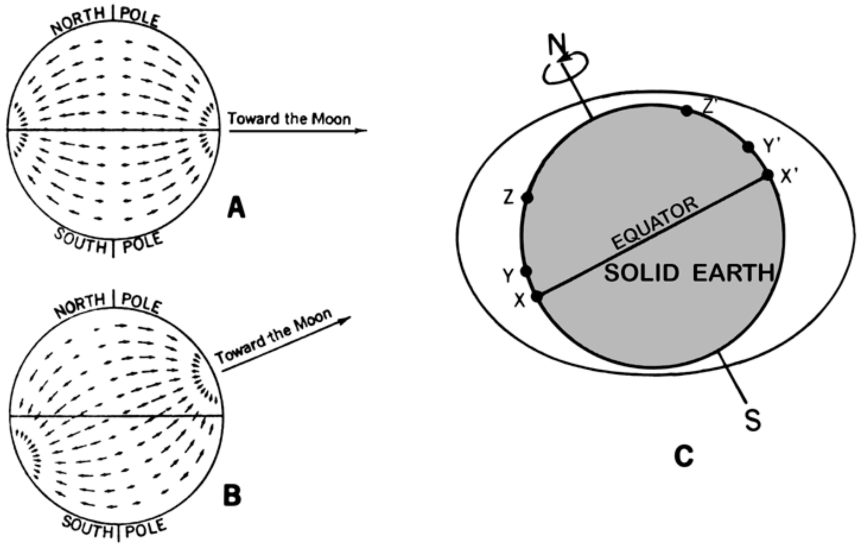


Figure 903d. Effects of the declination of the Moon.

The preceding discussion pertaining to the effects of the Moon is equally valid when discussing the effects of the Sun, taking into account that the magnitude of the solar effect is smaller. Hence, the tides will also vary according to the Sun's declination and its varying distance from the Earth. A second envelope of water representing the equilibrium tides due to the Sun would resemble the envelope shown in Figure 903c except that the heights of the high tides would be smaller, and the low tides correspondingly not as low. The theoretical tide at any place represents the combination of the effects of both the Moon and Sun.

FEATURES OF TIDES

904. General Features

At most places the tidal change occurs twice daily. The tide rises until it reaches a maximum height, called **high tide** or **high water**, and then falls to a minimum level called **low tide** or **low water**.

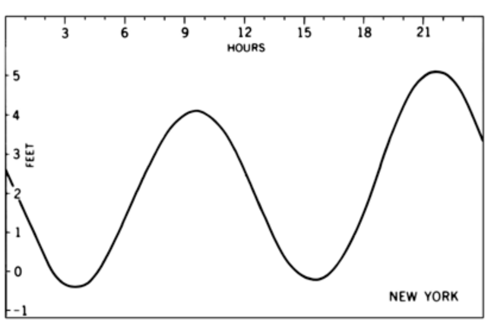


Figure 904. The rise and fall of the tide at New York, shown graphically.

The rate of rise and fall is not uniform. From low water, the tide begins to rise slowly at first, but at an increasing rate until it is about halfway to high water. The rate of rise then decreases until high water is reached, and the rise ceases.

The falling tide behaves in a similar manner. The period at high or low water during which there is no apparent change of level is called **stand**. The difference in height between consecutive high and low waters is the **range**.

Figure 904 is a graphical representation of the rise and fall of the tide at New York during a 24-hour period. The curve has the general form of a variable sine curve.

905. Types of Tide

A body of water has a natural period of oscillation, dependent upon its dimensions. None of the oceans is a single oscillating body; rather each one is made up of several separate oscillating basins. As such basins are acted upon by the tide-producing forces, some respond more readily to daily or diurnal forces, others to semidiurnal forces, and others almost equally to both. Hence, tides are classified as one of three types, semidiurnal, diurnal, or mixed, according to the characteristics of the tidal pattern.

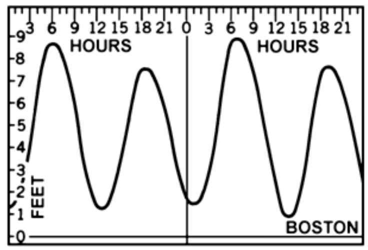


Figure 905a. Semidiurnal type of tide.

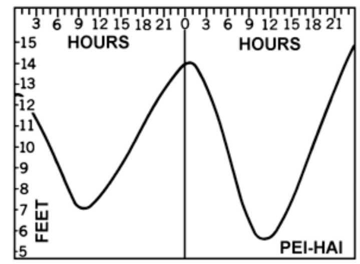


Figure 905b. Diurnal tide.

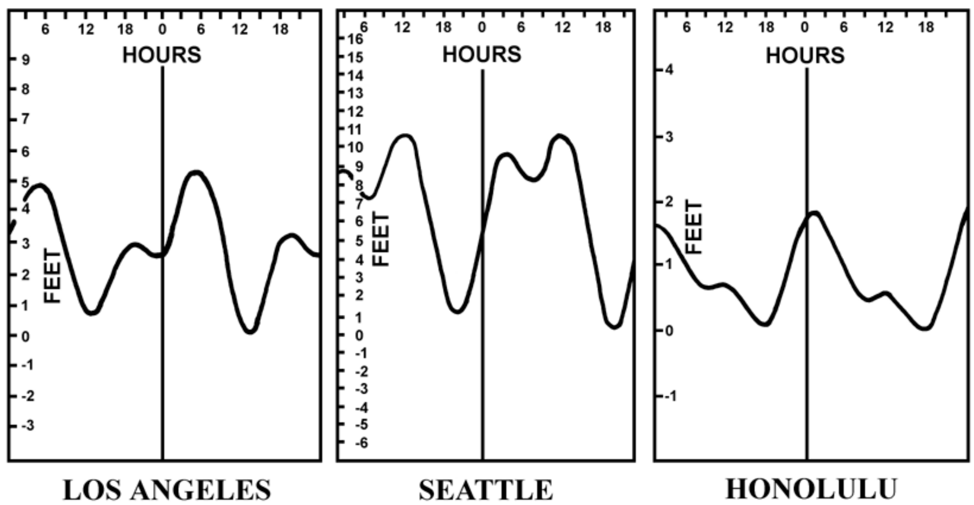


Figure 905c. Mixed tide.

In the **semidiurnal tide**, there are two high and two low waters each tidal day, with relatively small differences in the respective highs and lows. Tides on the Atlantic coast of the United States are of the semidiurnal type, which is illustrated in Figure 905a by the tide curve for Boston Harbor.

In the **diurnal tide**, only a single high and single low water occur each tidal day. Tides of the diurnal type occur along the northern shore of the Gulf of Mexico, in the Java Sea, the Gulf of Tonkin, and in a few other localities. The tide curve for Pei-Hai, China, illustrated in Figure 905b, is an example of the diurnal type.

In the **mixed tide**, the diurnal and semidiurnal oscillations are both important factors and the tide is characterized by a large inequality in the high water heights, low water heights, or in both. There are usually two high and two low waters each day, but occasionally the tide may become diurnal. Such tides are prevalent along the Pacific coast of the United States and in many other parts of the world. Examples of mixed types of tide are shown in Figure 905c. At Los Angeles, it is typical that the inequalities in the high and low waters are about the same. At Seattle the greater inequalities are typically in the low waters, while at Honolulu it is the high waters that have the greater inequalities.

906. Solar Tide

The natural period of oscillation of a body of water may accentuate either the solar or the lunar tidal oscillations. Though as a general rule the tides follow the Moon, the relative importance of the solar effect varies in different areas. There are a few places, primarily in the South Pacific and the Indonesian areas, where the solar oscillation is the more important, and at those places the high and low waters occur at about the same time each day. At Port Adelaide, Australia the solar and lunar semidiurnal oscillations are equal and nullify one another at neaps.

907. Special Tidal Effects

As a wave enters shallow water, its speed is decreased. Since the trough is shallower than the crest, it is retarded more, resulting in a steepening of the wave front. In a few estuaries, the advance of the low water trough is so much retarded that the crest of the rising tide overtakes the low, and advances upstream as a breaking wave called a **bore**. Bores that are large and dangerous at times of large tidal ranges may be mere ripples at those times of the month when the range is small. Examples occur in the Petitcodiac River in the Bay of Fundy, and at Haining, China, in the Tsientang Kaing. The tide tables indicate where bores occur.

Other special features are the **double low water** (as at Hoek Van Holland) and the **double high water** (as at Southampton, England). At such places there is often a slight fall or rise in the middle of the high or low water period. The practical effect is to create a longer period of stand at high or low tide. The tide tables list these and other peculiarities where they occur.

908. Variations in Range

Though the tide at a particular place can be classified as to type, it exhibits many variations during the month (Figure 908a). The range of the tide varies according to the intensity of the tide-producing forces, though there may be a lag of a day or two between a particular astronomic cause and the tidal effect.

The combined lunar-solar effect is obtained by adding the Moon's tractive forces vectorially to the Sun's tractive forces. The resultant tidal bulge will be predominantly lunar with modifying solar effects upon both the height of the tide and the direction of the tidal bulge. Special cases of interest occur during the times of new and full Moon (Figure 908b). With the Earth, Moon, and Sun lying approximately on the same line, the tractive forces of the Sun are acting in the same direction as the Moon's tractive forces (modified by declination effects). The resultant tides are called **spring tides**, whose ranges are greater than average.

Between the spring tides, the Moon is at first and third quarters. At those times, the tractive forces of the Sun are acting at approximately right angles to the Moon's tractive forces. The results are tides called **neap tides**, whose ranges are less than average.

With the Moon in positions between quadrature and new or full, the effect of the Sun is to cause the tidal bulge to either lag or precede the Moon (Figure 908c). These effects are called **priming** and **lagging** the tides.

Thus, when the Moon is at the point in its orbit nearest the Earth (at perigee), the lunar semidiurnal range is increased and **perigean tides** occur. When the Moon is farthest from the Earth (at apogee), the smaller **apogean tides** occur. When the Moon and Sun are in line and pulling together, as at new and full Moon, **spring tides** occur (the term spring has nothing to do with the season of year); when the Moon and Sun oppose each other, as at the quadratures, the smaller **neap tides** occur. When certain of these phenomena coincide, **perigean spring tides** and **apogean neap tides** occur.

These are variations in the semidiurnal portion of the tide. Variations in the diurnal portion occur as the Moon and Sun change declination. When the Moon is at its maximum semi-monthly declination (either north or south), **tropic tides** occur in which the diurnal effect is at a maximum. When it crosses the equator, the diurnal effect is a minimum and **equatorial tides** occur.

When the range of tide is increased, as at spring tides, there is more water available only at high tide; at low tide there is less, for the high waters rise higher and the low waters fall lower at these times. There is more water at neap low water than at spring low water. With tropic tides, there is usually more depth at one low water during the day than at the other. While it is desirable to know the meanings of these terms, the best way of determining the height of the tide at any place and time is to examine the tide predictions for the place as given in the tide tables, which take all these effects into account.

909. Tidal Cycles

Tidal oscillations go through a number of cycles. The shortest cycle, completed in about 12 hours and 25 minutes for a semidiurnal tide, extends from any phase of the tide to the next recurrence of the same phase. During a lunar day (averaging 24 hours and 50 minutes) there are two highs and two lows (two of the shorter cycles) for a semidiurnal tide. The Moon revolves around the Earth with respect to the Sun in a **synodical month** of about 29 1/2 days, commonly called the **lunar month**. The effect of the phase variation is completed in one-half of a synodical month or about 2 weeks as the Moon varies from new to full or full to new.

The effect of the Moon's declination is also repeated in one-half of a **tropical month** of 27 1/3 days, or about every 2 weeks. The cycle involving the Moon's distance requires an **anomalistic month** of about 27 1/2 days. The Sun's declination and distance cycles are respectively a half year and a year in length.

An important lunar cycle, called the **nodal period** or Metonic cycle (after Greek philosopher Meton, fifth century BC, who discovered the phenomenon) is 18.6 years (usually expressed in round figures as 19 years). For a tidal value, particularly a range, to be considered a true mean, it must be either based upon observations extended over this period of time, or adjusted to take account of variations known to occur during the nodal period.

The nodal period is the result of axis of the Moon's rotation being tilted 5 degrees with respect to the axis of the Earth's rotation. Since the Earth's axis is tilted 23.5 degrees with respect to the plane of its revolution around the sun, the combined effect is that the Moon's declination varies from 28.5 degrees to 18.5 degrees in a cycle lasting 18.6 years. For practical purposes, the nodal period can be con-

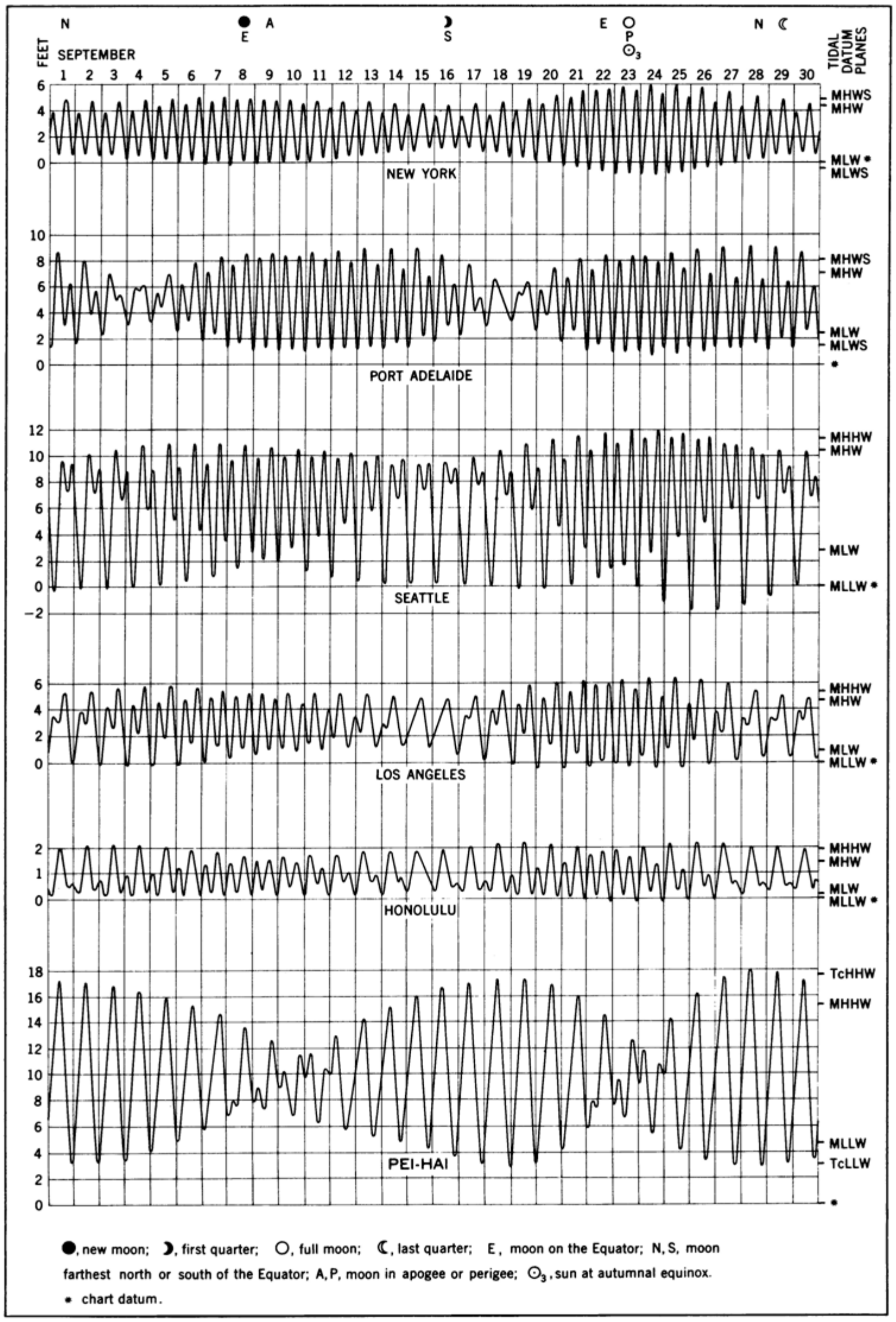


Figure 908a. Monthly tidal variations at various places.

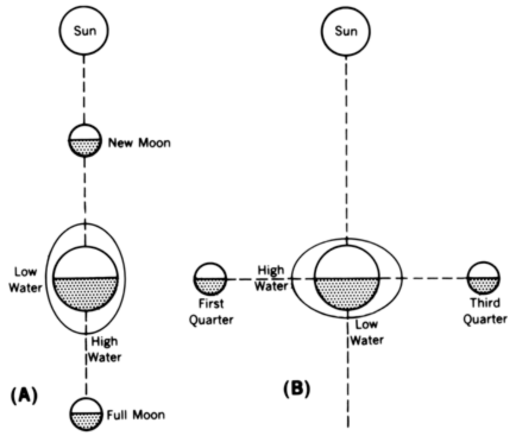


Figure 908b. (A) Spring tides occur at times of new and full Moon. Range of tide is greater than average since solar and lunar tractive forces act in same direction. (B) Neap ties occur at times of first and third quarters. Range of tide is less than average since solar and lunar tractive forces act at right angles.

sidered as the time between the Sun and Moon appearing in precisely the same relative positions in the sky.

910. Time of Tide

Since the lunar tide-producing force has the greatest effect in producing tides at most places, the tides "follow the Moon." Because the Earth rotates, high water lags behind both upper and lower meridian passage of the Moon. The **tidal day**, which is also the lunar day, is the time between consecutive transits of the Moon, or 24 hours and 50 minutes on the average. Where the tide is largely semidiurnal in type, the **lunitidal interval** (the interval between the Moon's meridian transit and a particular phase of tide) is fairly constant throughout the month, varying somewhat with the tidal cycles. There are many places, however, where solar or diurnal oscillations are effective in upsetting this relationship. The interval generally given is the average elapsed time from the meridian transit (upper or lower) of the Moon until the next high tide. This may be called **mean high water lunitidal interval** or **corrected** (or **mean**) **establishment**. The **common establishment** is

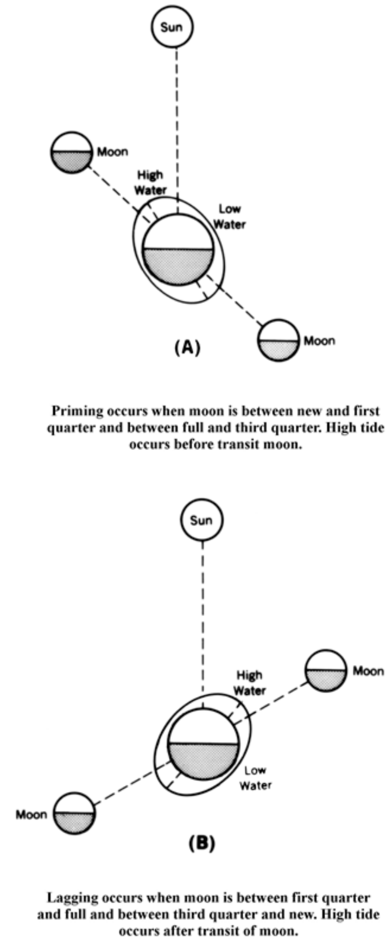


Figure 908c. Priming and lagging the tides.

the average interval on days of full or new Moon, and approximates the mean high water lunitidal interval.

In the ocean, the tide may be in the nature of a progressive wave with the crest moving forward, a stationary or standing wave which oscillates in a seesaw fashion, or a combination of the two. Consequently, caution should be used in inferring the time of tide at a place from tidal data for nearby places. In a river or estuary, the tide enters from the sea and is usually sent upstream as a progressive wave so that the tide occurs progressively later at various places upstream.

TIDAL DATUMS

911. Low Water Datums

A tidal datum is a given average tide level from which heights of tides and overhead clearances are measured. It is a vertical datum, but is not the same as vertical geodetic datum, which is a mathematical quantity developed as part of a geodetic system used for horizontal positioning. There are a number of tidal levels of reference that are important to the mariner. See Figure 911.

The most important level of reference is the **sounding datum** shown on charts. The sounding datum is sometimes referred to as the reference plane to distinguish it from vertical geodetic datum. Since the tide rises and falls continually while soundings are being taken during a hy-

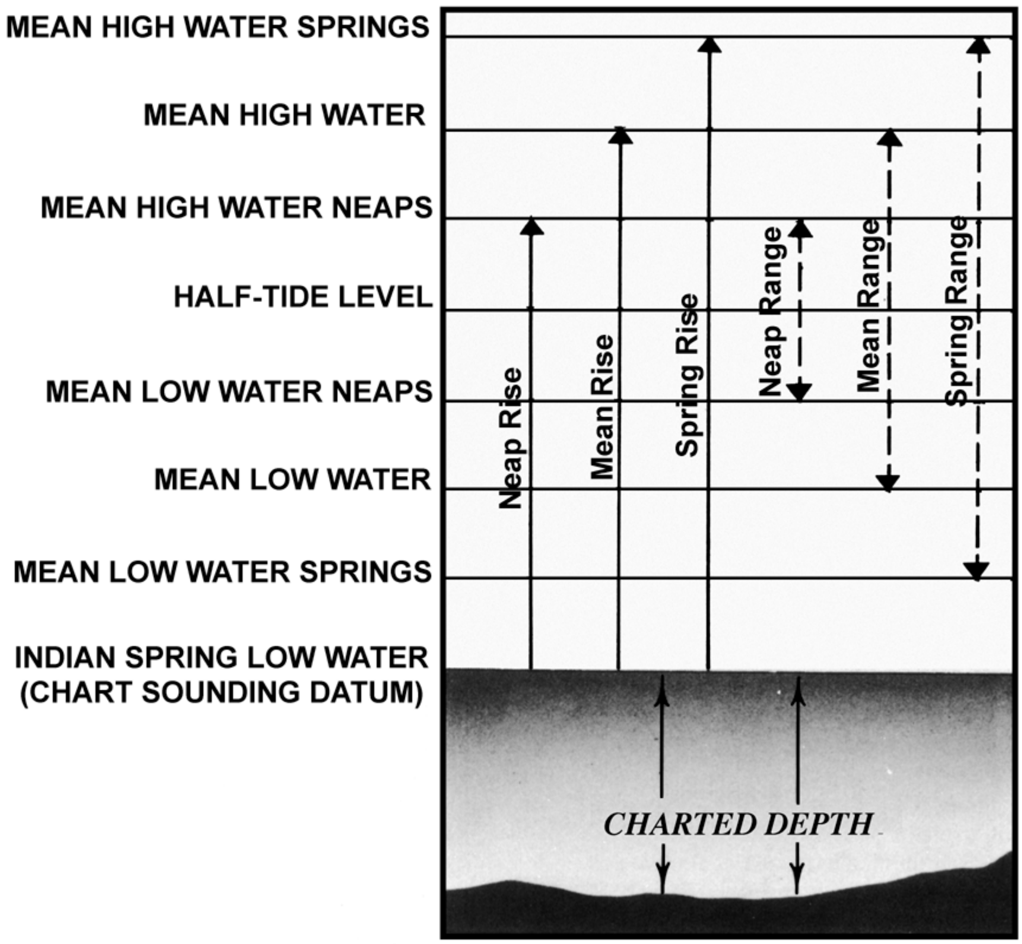


Figure 911. Variations in the ranges and heights of tide where the chart sounding datum is Indian Spring Low Water.

drographic survey, the tide is recorded during the survey so that soundings taken at all stages of the tide can be reduced to a common sounding datum. Soundings on charts show depths below a selected low water datum (occasionally mean sea level), and tide predictions in tide tables show heights above and below the same level. The depth of water available at any time is obtained by adding algebraically the height of the tide at the time in question to the charted depth.

By international agreement, the level used as chart datum should be low enough so that low waters do not fall very far below it. At most places, the level used is one determined from a mean of a number of low waters (usually over a 19 year period); therefore, some low waters can be expected to fall below it. The following are some of the datums in general use.

Mean low water (MLW) is the average height of all low waters at a given place. About half of the low waters fall below it, and half above.

Mean low water springs (MLWS), usually shortened to low water springs, is the average level of the low waters that occur at the times of spring tides.

Mean lower low water (MLLW) is the average height of the lower low waters of each tidal day.

Tropic lower low water (TcLLW) is the average height of the lower low waters (or of the single daily low waters if the tide becomes diurnal) that occur when the Moon is near maximum declination and the diurnal effect is most pronounced. This datum is not in common use as a tidal reference.

Indian spring low water (ISLW), sometimes called **Indian tide plane** or **harmonic tide plane**, is a low water datum that includes the spring effect of the semi-diurnal portion of the tide and the tropic effect of the diurnal portion. It is about the level of lower low water of mixed tides at the time that the Moon's maximum declination coincides with the time of new or full Moon.

Mean lower low water springs (MLLWS) is the average level of the lower of the two low waters on the days of spring tides.

Some lower datums used on charts are determined from tide observations and some are determined arbitrarily and later referred to the tide. Most of them fall close to one or the other of the following two datums.

Lowest normal low water is a datum that approximates the average height of monthly lowest low waters, discarding any tides disturbed by storms.

Lowest low water is an extremely low datum. It conforms generally to the lowest tide observed, or even somewhat lower. Once a tidal datum is established, it is sometimes retained for an indefinite period, even though it might differ slightly from a better determination from later observations. When this occurs, the established datum may be called **low water datum**, **lower low water datum**, etc. These datums are used in a limited area and primarily for river and harbor engineering purposes. Examples are Boston Harbor Low Water Datum and Columbia River Lower Low Water Datum.

Some sounding datums are based on the predicted tide rather than an average of observations. A British sounding datum that may be adopted internationally is the Lowest Astronomical Tide (LAT). LAT is the elevation of the lowest water level predicted in a 19-year period. Canadian coastal charts use a datum of Lower Low Water, Large Tide (LLWLT) which is the average of the lowest low waters, one from each of the 19 years of predictions.

Figure 911 illustrates variations in the ranges and heights of tides in a locality such as the Indian Ocean, where predicted and observed water levels are referenced to a chart sounding datum that will always cause them to be additive relative to the charted depth.

In areas where there is little or no tide, various other datums are used. For the Black Sea for instance, Mean Sea Level (MSL, sometimes referred to as Mean Water Level or MWL) is used, and is the average of the hourly heights observed over a period of time and adjusted to a 19-year period. In the United States, a Low Water Datum (LWD) is used in those coastal areas that have transitioned from tidal to non-tidal (e.g. Laguna Madre, Texas and Pamlico Sound, North Carolina) and is simply 0.5 foot below a computed MLW. For the Great Lakes, the United States and Canada use a separate LWD for each lake, which is designed to ensure that the actual water level is above the datum most of the time during the navigation season. Lake levels vary by several feet over a period of years.

Inconsistencies of terminology are found among charts of different countries and between charts issued at different times.

Large-scale charts usually specify the datum of soundings and may contain a tide note giving mean heights of the tide at one or more places on the chart. These heights are intended merely as a rough guide to the change in depth to be expected under the specified conditions. They should not be used for the prediction of heights on any particular day, which should be obtained from tide tables.

912. High Water Datums

Heights of terrestrial features are usually referred on nautical charts to a high water datum. This gives the mariner a margin of error when passing under bridges, overhead cables, and other obstructions. The one used on charts of the United States, its territories and possessions, and widely used elsewhere, is **mean high water (MHW)**, which is the average height of all high waters over a 19 year period. Any other high water datum in use on charts is likely to be higher than this. Other high water datums are **mean high water springs (MHWS)**, which is the average level of the high waters that occur at the time of spring tides; **mean higher high water (MHHW)**, which is the average height of the higher high waters of each tidal day; and **tropic higher high water (TcHHW)**, which is the average height of the higher high waters (or the single daily high waters if the tide becomes diurnal) that occur when the Moon is near maximum declination and the diurnal effect is most pronounced. A reference merely to "high water" leaves some doubt as to the specific level referred to, for the height of high water varies from day to day. Where the range is large, the variation during a 2 week period may be considerable.

Because there are periodic and apparent secular trends in sea level, a specific 19 year cycle (the **National Tidal Datum Epoch**) is issued for all United States datums. The National Tidal Datum Epoch officially adopted by the National Ocean Service is presently 1960 through 1978. The Epoch is reviewed for revision every 25 years.

TIDAL CURRENTS

913. Tidal and Nontidal Currents

Horizontal movement of water is called **current**. It may be either "tidal" and "nontidal." **Tidal current** is the periodic horizontal flow of water accompanying the rise and fall of the tide. **Nontidal current** includes all currents not due to the tidal movement. Nontidal currents include the permanent currents in the general circulatory system of the oceans as well as temporary currents arising from meteorological conditions. The current experienced at any time is usually a combination of tidal and nontidal currents.

914. General Features

Offshore, where the direction of flow is not restricted by any barriers, the tidal current is rotary; that is, it flows continuously, with the direction changing through all points of the compass during the tidal period. This rotation is caused by the Earth's rotation, and unless modified by local conditions, is clockwise in the Northern Hemisphere and counterclockwise in the Southern Hemisphere. The speed usually varies throughout the tidal cycle, passing through two maximums in approximately opposite directions, and two minimums about halfway between the maximums in time and direction. Rotary currents can be depicted as in Figure 914a, by a series of arrows representing the direction and speed of the current at each hour. This is sometimes called a **current rose**. Because of the elliptical pattern formed by the ends of the arrows, it is also referred to as a **current ellipse**.

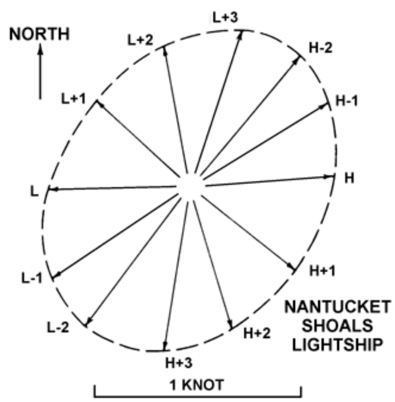


Figure 914a. Rotary tidal current. Times are hours before and after high and low tide at Nantucket Shoals. The bearing and length of each arrow represents the hourly direction and speed of the current.

In rivers or straits, or where the direction of flow is more or less restricted to certain channels, the tidal current is reversing; that is, it flows alternately in approximately opposite directions with an instant or short period of little or no current, called **slack water**, at each reversal of the current. During the flow in each direction, the speed varies from zero at the time of slack water to a maximum, called strength of flood or ebb, about midway between the slacks. Reversing currents can be indicated graphically, as in Figure 914b, by arrows that represent the speed of the current at each hour. The flood is usually depicted above the slack waterline and the ebb below it. The tidal current curve formed by the ends of the arrows has the same characteristic sine form as the tide curve. In illustrations and for certain other purposes it is convenient to omit the arrows and show only the curve.

A slight departure from the sine form is exhibited by the reversing current in a strait that connects two different tidal basins, such as the East River, New York. The tides at the two ends of a strait are seldom in phase or equal in range, and the current, called **hydraulic current**, is generated largely by the continuously changing difference in height of water at the two ends. The speed of a hydraulic current varies nearly as the square root of the difference in

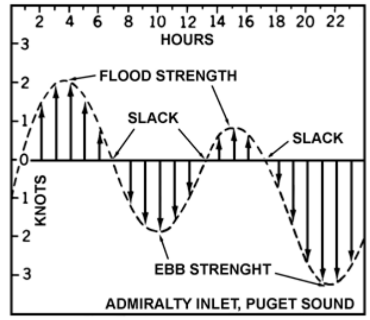


Figure 914b. Reversing tidal current.

height. The speed reaches a maximum more quickly and remains at strength for a longer period than shown in Figure 914b, and the period of weak current near the time of slack is considerably shortened.

The current direction, or **set**, is the direction toward which the current flows. The speed is sometimes called the **drift**. The term "velocity" is often used as the equivalent of "speed" when referring to current, although strictly speaking "velocity" implies direction as well as speed. The term "strength" is also used to refer to speed, but more often to greatest speed between consecutive slack waters. The movement toward shore or upstream is the **flood**, the movement away from shore or downstream is the **ebb**. In a purely semidiurnal current unaffected by nontidal flow, the flood and ebb each last about 6 hours and 13 minutes. But if there is either diurnal inequality or nontidal flow, the durations of flood and ebb may be quite unequal.

915. Types of Tidal Current

Tidal currents, like tides, may be of the **semidiurnal**, **diurnal**, or **mixed** type, corresponding to a considerable degree to the type of tide at the place, but often with a stronger semidiurnal tendency.

The tidal currents in tidal estuaries along the Atlantic coast of the United States are examples of the semidiurnal type of reversing current. Along the Gulf of Mexico coast, such as at Mobile Bay entrance, they are almost purely diurnal. At most places, however, the type is mixed to a greater or lesser degree. At Tampa and Galveston entrances there is only one flood and one ebb each day when the Moon is near its maximum declination, and two floods and two ebbs each day when the Moon is near the equator. Along the Pacific coast of the United States there are generally two floods and two ebbs every day, but one of the floods or ebbs has a greater speed and longer duration than the other, the inequality varying with the declination of the Moon.

The inequalities in the current often differ considerably from place to place even within limited areas, such as adjacent passages in Puget Sound and various passages between the Aleutian Islands. Figure 915a shows several types of re-

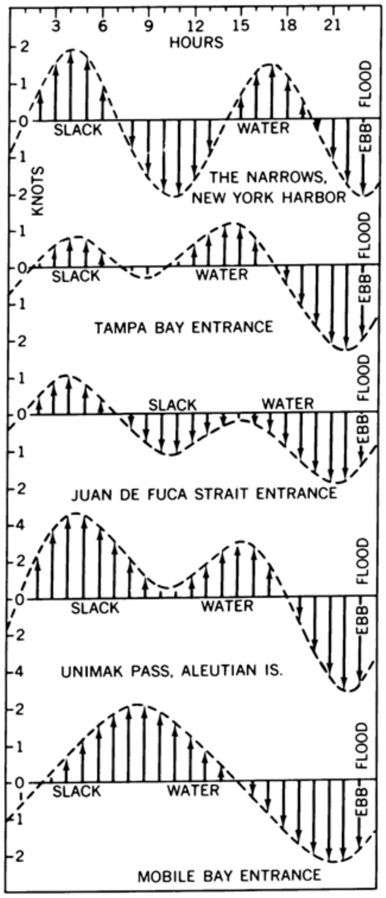


Figure 915a. Several types of reversing current. The pattern changes gradually from day to day, particularly for mixed types, passing through cycles.

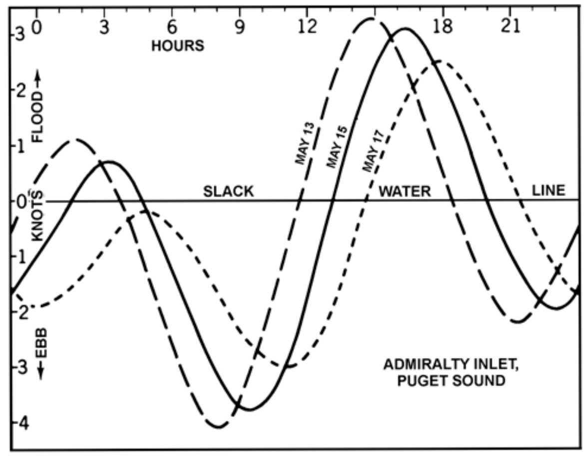


Figure 915b. Changes in a current of the mixed type. Note that each day as the inequality increases, the morning slacks draw together in time until on the 17th the morning flood disappears. On that day the current ebbs throughout the morning.

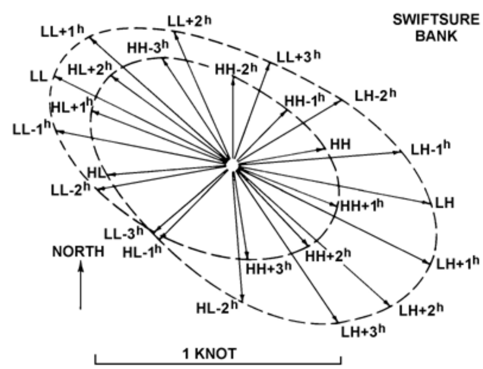


Figure 915c. Rotary tidal current with diurnal inequality. Times are in hours referred to tides (higher high, lower low, lower high, and higher low) at Swiftsure Bank.

versing current. Figure 915b shows how the flood disappears as the diurnal inequality increases at one station.

Offshore rotary currents that are purely semidiurnal repeat the elliptical pattern each tidal cycle of 12 hours and 25 minutes. If there is considerable diurnal inequality, the plotted hourly current arrows describe a set of two ellipses of different sizes during a period of 24 hours and 50 minutes, as shown in Figure 915c, and the greater the diurnal inequality, the greater the difference between the sizes of the two ellipses. In a completely diurnal rotary current, the smaller ellipse disappears and only one ellipse is produced in 24 hours and 50 minutes.

916. Tidal Current Periods and Cycles

Tidal currents have periods and cycles similar to those of the tides, and are subject to similar variations, but flood and ebb of the current do not necessarily occur at the same times as the rise and fall of the tide.

The speed at strength increases and decreases during the 2 week period, month, and year along with the variations in the range of tide. Thus, the stronger spring and perigean currents occur near the times of new and full Moon and near the times of the Moon's perigee, or at times of spring and perigean tides; the weaker neap and apogean currents occur at the times of neap and apogean tides; and tropic currents with increased diurnal speeds or with larger diurnal inequalities in speed occur at times of tropic tides; and equatorial currents with a minimum diurnal effect occur at times of equatorial tides.

As with the tide, a mean value represents an average obtained from a 19 year series. Since a series of current observations is usually limited to a few days, and seldom covers more than a month or two, it is necessary to adjust the observed values, usually by comparison with tides at a nearby place, to obtain such a mean.

917. Effect of Nontidal Flow

The current existing at any time is seldom purely tidal, but usually includes also a nontidal current that is due to drainage, oceanic circulation, wind, or other causes. The method in which tidal and nontidal currents combine is best explained graphically, as in Figure 917a and Figure 917b. The pattern of the tidal current remains unchanged, but the curve is shifted from the point or line from which the currents are measured, in the direction of the nontidal current, and by an amount equal to it. It is sometimes more convenient graphically merely to move the line or point of origin in the opposite direction. Thus, the speed of the current flowing in the direction of the nontidal current is increased by an amount equal to the magnitude of the nontidal current, and the speed of the current flowing in the opposite direction is decreased by an equal amount.

In Figure 917a, a nontidal current is represented both in direction and speed by the vector AO. Since this is greater than the speed of the tidal current in the opposite direction, the point A is outside the ellipse. The direction and speed of the combined tidal and nontidal currents at any time is represented by a vector from A to that point on the curve representing the given time, and can be scaled from the graph. The strongest and weakest currents may no longer be in the directions of the maximum and minimum of the tidal current. If the nontidal current is northwest at 0.3 knot, it may be represented by BO, and all hourly directions and speeds will then be measured from B. If it is 1.0 knot, it will be represented by AO and the actual resultant hourly directions and speeds will be measured from A, as shown by the arrows.

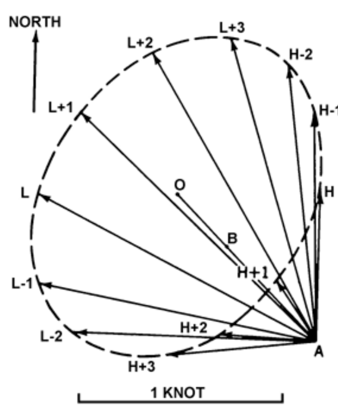


Figure 917a. Effect of nontidal current on the rotary tidal current of Figure 914a.

In a reversing current (Figure 917b), the effect is to advance the time of one slack, and to retard the following one. If the speed of the nontidal current exceeds that of the reversing tidal current, the resultant current flows continuously in one direction without coming to a slack. In this case, the speed varies from a maximum to a minimum and back to a maximum in each tidal cycle. In Figure 917b, the horizontal line A represents slack water if only tidal currents are present. Line B represents the effect of a 0.5 knot nontidal ebb, and line C the effect of a 1.0 knot nontidal ebb. With the condition shown at C there is only one flood each tidal day. If the nontidal ebb were to increase to approximately 2 knots, there would be no flood, two maximum ebbs and two minimum ebbs occurring during a tidal day.

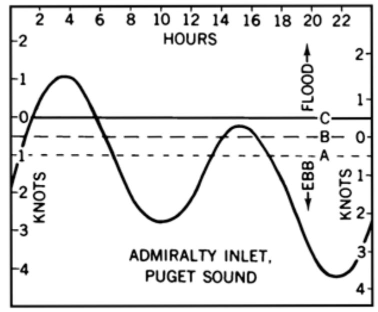


Figure 917b. Effect of nontidal current on the reversing tidal current of Figure 914b.

918. Time of Tidal Current and Time of Tide

At many places where current and tide are both semidiurnal, there is a definite relationship between times of current and times of high and low water in the locality. Current atlases and notes on nautical charts often make use of this relationship by presenting for particular locations, the direction and speed of the current at each succeeding hour after high and low water, at a place for which tide predictions are available.

Where there is considerable diurnal inequality in tide or current, or where the type of current differs from the type of tide, the relationship is not constant, and it may be hazardous to try to predict the times of current from times of tide. Note the current curve for Unimak Pass in the Aleutians in Figure 915a. It shows the current as predicted in the tidal current tables. Predictions of high and low waters in the tide tables might have led one to expect the current to change from flood to ebb in the late morning, whereas actually the current continued to run flood with some strength at that time.

Since the relationship between times of tidal current and tide is not the same everywhere, and may be variable at the same place, one should exercise extreme caution in using general rules. The belief that slacks occur at local high and low tides and that the maximum flood and ebb occur when the tide is rising or falling most rapidly may be approximately true at the seaward entrance to, and in the upper reaches of, an inland tidal waterway. But generally this is not true in other parts of inland waterways. When an inland waterway is extensive or its entrance constricted, the slacks in some parts of the waterway often occur midway between the times of high and low tide. Usually in such waterways the relationship changes from place to place as one progresses upstream, slack water getting progressively closer in time to the local tide maximum until at the head of tidewater (the inland limit of water affected by a tide) the slacks occur at about the times of high and low tide.

919. Relationship Between Speed of Current and Range of Tide

The speed of the tidal current is not necessarily consistent with the range of tide. It may be the reverse. For example, currents are weak in the Gulf of Maine where the tides are large, and strong near Nantucket Island and in Nantucket Sound where the tides are small. However, at any one place the speed of the current at strength of flood and ebb varies during the month in about the same proportion as the range of tide, and this relationship can be used to determine the relative strength of currents on any given day.

920. Variation Across an Estuary

In inland tidal estuaries the time of tidal current varies across the channel from shore to shore. On the average, the current turns earlier near shore than in midstream, where the speed is greater. Differences of half an hour to an hour are not uncommon, but the difference varies and the relationship may be nullified by the effect of nontidal flow.

The speed of the current also varies across the channel, usually being greater in midstream or midchannel than near shore, but in a winding river or channel the strongest currents occur near the concave shore, or the outside corner of the curve. Near the opposite (convex) shore the currents are weak or eddying.

921. Variation with Depth

In tidal rivers the subsurface current acting on the lower portion of a ship's hull may differ considerably from the surface current. An appreciable subsurface current may be present when the surface movement appears to be practically slack, and the subsurface current may even be flowing with appreciable speed in the opposite direction to the surface current.

In a tidal estuary, particularly in the lower reaches where there is considerable difference in density from top to bottom, the flood usually begins earlier near the bottom than at the surface. The difference may be an hour or two, or as little as a few minutes, depending upon the estuary, the location in the estuary, and freshet conditions. Even when the freshwater runoff becomes so great as to prevent the surface current from flooding, it may still flood below the surface. The difference in time of ebb from surface to bottom is normally small but subject to variation with time and location.

The ebb speed at strength usually decreases gradually from top to bottom, but the speed of flood at strength often is stronger at subsurface depths than at the surface.

922. Tidal Current Observations

Observations of current are made with sophisticated electronic **current meters**. Current meters are suspended from a buoy or anchored to the bottom with no surface marker at all. Very sensitive current meters measure and record deep ocean currents; these are later recovered by triggering a release mechanism with a signal from the surface. Untended current meters either record data internally or send it by radio to a base station on ship or land. The period of observation varies from a few hours to as long as 6 months.

TIDE AND CURRENT PREDICTION

923. Tidal Height Predictions

To measure the height of tides, hydrographers select a reference level, sometimes referred to as the reference plane, or vertical datum. This vertical tidal datum is not the same as the vertical geodetic datum. Soundings shown on the largest scale charts are the vertical distances from this datum to the bottom. At any given time the actual depth is this charted depth plus the height of tide. In most places the reference level is some form of low water. But all low waters at a given place are not the same height, and the selected reference level is seldom the lowest tide occurring at the place. When lower tides occur, these are indicated in the tide tables by a negative sign. Thus, at a spot where the charted depth is 15 feet, the actual depth is 15 feet plus the tidal height. When the tide is three feet, the depth is 15 + 3 = 18 feet. When it is -1 foot, the depth is 15 - 1 = 14 feet. The actual depth can be less than the charted depth. In an area where there is a considerable range of tide (the difference between high water and low water), the height of tide might be an important consideration when using soundings to determine if the vessel is in safe water.

The heights given in the tide tables are predictions, and when assumed conditions vary considerably, the predictions shown may be considerably in error. Heights lower than predicted can be anticipated when the atmospheric pressure is higher than normal, or when there is a persistent strong offshore wind. The greater the range of tide, the less reliable are the predictions for both height and current.

924. Tidal Heights

The nature of the tide at any place can best be determined by observation. The predictions in tide tables and the tidal data on nautical charts are based upon detailed observations at specific locations, instead of theoretical predictions.

Tidal elevations are usually observed with a continuously recording gage. A year of observations is the minimum length desirable for determining the harmonic constants used in prediction. For establishing mean sea level and long-term changes in the relative elevations of land and sea, as well as for other special uses, observations have been made over periods of 20, 30, and even 120 years at important locations. Observations for a month or less will establish the type of tide and suffice for comparison with a longer series of observations to determine tidal differences and constants.

Mathematically, the variations in the lunar and solar tide-producing forces, such as those due to changing phase, distance, and declination, are considered as separate constituent forces, and the harmonic analysis of observations reveals the response of each constituent of the tide to its corresponding force. At any one place this response remains constant and is shown for each constituent by **harmonic constants** which are in the form of a phase angle for the time relation and an amplitude for the height. Harmonic constants are used in making technical studies of the tide and in tidal predictions on computers. The tidal predictions in most published tide tables are produced by computer.

925. Meteorological Effects

The foregoing discussion of tidal behavior assumes normal weather conditions. However, sea level is also affected by wind and atmospheric pressure. In general, onshore winds raise the level and offshore winds lower it, but the amount of change varies at different places. During periods of low atmospheric pressure, the water level tends to be higher than normal. For a stationary low, the increase in elevation can be found by the formula

$$R_0=0.01(1010 - P),$$

in which R_0 is the increase in elevation in meters and P is the atmospheric pressure in hectopascals. This is equal approximately to 1 centimeter per hectopascal depression, or about 13.6 inches per inch depression. For a moving low, the increase in elevation is given by the formula

$$R= \frac{R_0}{1 - \frac{C^2}{gh}}$$

in which R is the increase in elevation in feet, R_0 is the increase in meters for a stationary low, C is the rate of motion of the low in feet per second, g is the acceleration due to gravity (32.2 feet per second per second), and h is the depth of water in feet.

Where the range of tide is very small, the meteorological effect may sometimes be greater than the normal tide. Where a body of water is large in area but shallow, high winds can push the water from the windward to the lee shore, creating much greater local differences in water levels than occurs normally, and partially or completely masking the tides. The effect is dependent on the configuration and depth of the body of water relative to the wind direction, strength and duration.

926 Tidal Current Predictions

Tidal currents are due primarily to tidal action, but other causes are often present. The ***Tidal Current Tables*** give the best prediction of total current. Following heavy rains or a drought, a river's current prediction may be considerably in error. Set and drift may vary considerably over different parts of a harbor, because differences in bathymetry from place to place affect current. Since this is usually an area where small errors in a vessel's position are crucial, a knowledge of predicted currents, particularly in reduced visibility, is important. Strong currents occur mostly in narrow passages connecting larger bodies of water. Currents of more than 5 knots are sometimes encountered at the Golden Gate in San Francisco, and currents of more than 13 knots sometimes occur at Seymour Narrows, British Columbia.

In straight portions of rivers and channels, the strongest currents usually occur in the middle of the channel. In curved portions the swiftest currents (and deepest water) usually occur near the outer edge of the curve. Countercurrents and eddies may occur on either side of the main current of a river or narrow passage, especially near obstructions and in bights.

In general, the range of tide and the velocity of tidal current are at a minimum in the open ocean or along straight coasts. The greatest tidal effects are usually encountered in estuaries, bays, and other coastal indentations. A vessel proceeding along an indented coast may encounter a set toward or away from the shore; a similar set is seldom experienced along a straight coast.

PUBLICATIONS FOR PREDICTING TIDES AND CURRENTS

927. *Tide Tables*

Usually, tidal information is obtained from tide and tidal current tables, or from specialized computer software or calculators. However, if these are not available, or if they do not include information at a desired place, the mariner may be able to obtain locally the **mean high water lunitidal interval** or the **high water full and change**. The approximate time of high water can be found by adding either interval to the time of transit (either upper or lower) of the Moon. Low water occurs approximately 1/4 tidal day (about 6^h 12^m) before and after the time of high water. The actual interval varies somewhat from day to day, but approximate results can be obtained in this manner. Similar information for tidal currents (**lunicurrent interval**) is seldom available.

The National Ocean Service (NOS) has traditionally published hard copy tide tables and tidal current tables. Tide and tidal current data continue to be updated by NOS, but hardcopy publication has been transferred to private companies working with NOS data, published on CD-ROM.

Tidal data for various parts of the world is published in 4 volumes by the National Ocean Service. These volumes are:

- Central and Western Pacific Ocean and Indian Ocean
- East Coast of North and South America (including Greenland)
- Europe and West Coast of Africa
- West Coast of North and South America (including the Hawaiian Islands)

A small separate volume, the Alaskan Supplement, is also published.

Each volume has 5 common tables:

- **Table 1** contains a complete list of the predicted times and heights of the tide for each day of the year at a number of places designated as **reference stations**.
- **Table 2** gives tidal differences and ratios which can be used to modify the tidal information for the reference stations to make it applicable to a relatively large number of **subordinate stations**.
- **Table 3** provides information for finding the approximate height of the tide at any time between high water and low water.
- **Table 4** is a sunrise-sunset table at five-day intervals for various latitudes from 76°N to 60°S (40°S in one volume).
- **Table 5** provides an adjustment to convert the local mean time of Table 4 to zone or standard time.

For the East Coast and West Coast volumes, each contains a Table 6, a moonrise and moonset table; Table 7 for conversion from feet to centimeters; Table 8, a table of estimated tide prediction accuracies; a glossary of terms; and an index to stations. Each table is preceded by a complete explanation. Sample problems are given where necessary. The inside back cover of each volume contains a calendar of critical astronomical data to help explain the variations of the tide during each month and throughout the year.

928. Tide Predictions for Reference Stations

For each day, the date and day of week are given, and the time and height of each high and low water are listed in chronological order. Although high and low waters are not labeled as such, they can be distinguished by the relative heights given immediately to the right of the times. If two high tides and two low tides occur each tidal day, the tide is semidiurnal. Since the tidal day is longer than the civil day (because of the revolution of the Moon eastward around the Earth), any given tide occurs later each day. Because of later times of corresponding tides from day to day, certain days have only one high water or only one low water.

929. Tide Predictions for Subordinate Stations

For each subordinate station listed, the following information is given:

1. **Number**. The stations are listed in geographical order and assigned consecutive numbers. Each volume contains an alphabetical station listing correlating the station with its consecutive number to assist in finding the entry in Table 2.
2. **Place**. The list of places includes both subordinate and reference stations; the latter are in bold type.
3. **Position**. The approximate latitude and longitude are given to assist in locating the station. The latitude is north or south, and the longitude east or west, depending upon the letters (N, S, E, W) next above the entry. These may not be the same as those at the top of the column.
4. **Differences**. The differences are to be applied to the predictions for the reference station, shown in capital letters above the entry. Time and height differences are given separately for high and low waters. Where differences are omitted, they are either unreliable or unknown.
5. **Ranges**. Various ranges are given, as indicated in the tables. In each case this is the difference in height between high water and low water for the tides indicated.
6. **Mean tide level**. This is the average between mean low and mean high water, measured from chart datum.

The **time difference** is the number of hours and minutes to be applied to the reference station time to find the time of the corresponding tide at the subordinate station. This interval is added if preceded by a plus sign (+) and subtracted if preceded by a minus sign (-). The results obtained by the application of the time differences will be in the zone time of the time meridian shown directly above the difference for the subordinate station. Special conditions occurring at a few stations are indicated by footnotes on the applicable pages. In some instances, the corresponding tide falls on a different date at reference and subordinate stations.

Height differences are shown in a variety of ways. For most entries, separate height differences in feet are given for high water and low water. These are applied to the height given for the reference station. In many cases a ratio is given for either high water or low water, or both. The height at the reference station is multiplied by this ratio to find the height at the subordinate station. For a few stations, both a ratio and difference are given. In this case the height at the reference station is first multiplied by the ratio, and the difference is then applied. An example is given in each volume of tide tables. Special conditions are indicated in the table or by footnote. For example, a footnote indicates that "Values for the Hudson River above George Washington Bridge are based upon averages for the six months May to October, when the fresh-water discharge is a minimum."

930. Finding Height of Tide at any Time

Table 3 provides means for determining the approximate height of tide at any time. It assumes that plotting height versus time yields a sine curve. Actual values may vary from this. The explanation of the table contains directions for both mathematical and graphical solutions. Though the mathematical solution is quicker, if the vessel's ETA changes significantly, it will have to be done for the new ETA. Therefore, if there is doubt about the ETA, the graphical solution will provide a plot of predictions for several hours and allow quick reference to the predicted height for any given time. This method will also quickly show at what time a given depth of water will occur. Figure 930a shows the OPNAV form used to calculate heights of tides. Figure 930b shows the importance of calculating tides in shallow water.

OPNAV 3530/40 (4-73)
HT OF TIDE

Date	
Location	
Time	
Ref Sta	
HW Time Diff	
LW Time Diff	
HW Ht Diff	
LW Ht Diff	
Ref Sta HW/LW Time	
HW/LW Time Diff	
Sub Sta HW/LW Time	
Ref Sta HW/LW Ht	
HW/LW Ht Diff	
Sub Sta HW/LW Ht	
Duration	**Rise**
	Fall
Time Fm	**Near**
	Tide
Range of Tide	
Ht of Neat Tide	
Corr Table 3	
Ht of Tide	
Charted Depth	
Depth of Water	
Draft	
Clearance	

Figure 930a. OPNAV 3530/40 Tide Form.

931. *Tidal Current Tables*

Tidal Current Tables are somewhat similar to *Tide Tables*, but the coverage is less extensive. NOS publishes 2 volumes on an annual basis: Atlantic Coast of North America, and Pacific Coast of North America and Asia. Each of the two volumes is arranged as follows:

Each volume also contains current diagrams and instructions for their use. Explanations and examples are given in each table.

- **Table 1** contains a complete list of predicted times of maximum currents and slack water, with the velocity of the maximum currents, for a number of reference stations.
- **Table 2** gives differences, ratios, and other information related to a relatively large number of subordinate

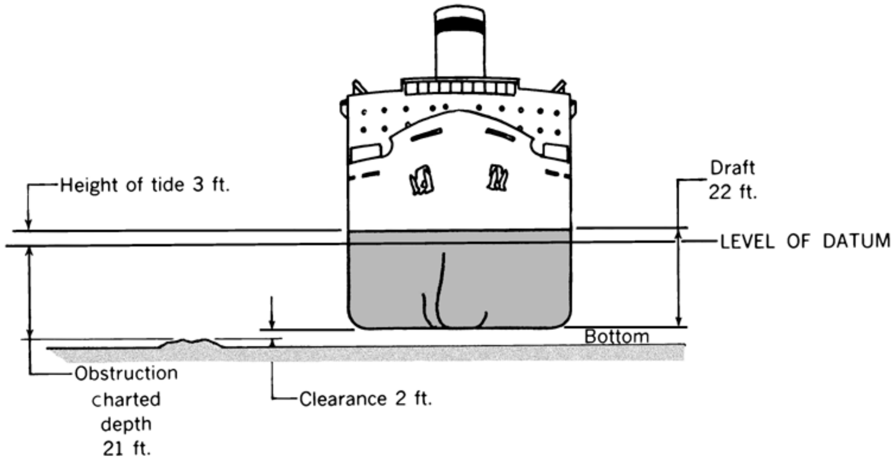


Figure 930b. Height of tide required to pass clear of charted obstruction.

stations.

- **Table 3** provides information to determine the current's velocity at any time between entries in tables 1 and 2.
- **Table 4** gives duration of slack, or the number of minutes the current does not exceed stated amounts, for various maximum velocities.
- **Table 5** (Atlantic Coast of North America only) gives information on rotary tidal currents.

The volumes also contain general descriptive information on wind-driven currents, combination currents, and information such as Gulf Stream currents for the east coast and coastal currents on the west coast.

932. Tidal Current Prediction for Reference Stations

For each day, the date and day of week are given; current information follows. If the cycle is repeated twice each tidal day, currents are semidiurnal. On most days there are four slack waters and four maximum currents, two floods (F) and two ebbs (E). However, since the tidal day is longer than the civil day, the corresponding condition occurs later each day, and on certain days there are only three slack waters or three maximum currents. At some places, the current on some days runs maximum flood twice, but ebbs only once, a minimum flood occurring in place of the second ebb. The tables show this information.

933. Tidal Current Predictions for Subordinate Stations

For each subordinate station listed in Table 2 of the tidal current tables, the following information is given:

1. **Number**: The stations are listed in geographical order and assigned consecutive numbers, as in the tide tables. Each volume contains an alphabetical station listing correlating the station with its consecutive number to assist in locating the entry in Table 2.
2. **Place**: The list of places includes both subordinate and reference stations, the latter given in bold type.
3. **Position**: The approximate latitude and longitude are given to assist in locating the station. The latitude is north or south and the longitude east or west as indicated by the letters (N, S, E, W) next above the entry. The current given is for the center of the channel unless another location is indicated by the station name.
4. **Time difference**: Two time differences are tabulated. One is the number of hours and minutes to be applied to the tabulated times of slack water at the reference station to find the times of slack waters at the subordinate station. The other time difference is applied to the times of maximum current at the reference station to find the times of the corresponding maximum current at the subordinate station. The intervals, which are added or subtracted in accordance with their signs, include any difference in time between the two stations, so that the answer is correct for the standard time of the subordinate station. Limited application and special conditions are indicated by footnotes.
5. **Velocity ratios**: Speed of the current at the subordinate station is the product of the velocity at the reference station and the tabulated ratio. Separate ratios may be given for flood and ebb currents. Special conditions are indicated by footnotes.
6. **Average Speeds and Directions**: Minimum and maximum velocities before flood and ebb are listed for each station, along with the true directions of the flow. Minimum velocity is not always 0.0 knots.

934. Finding Velocity of Tidal Current at any Time

Table 3 of the tidal current tables provides means for determining the approximate velocity at any time. Directions are given in an explanation preceding the table. Figure 934 shows the OPNAV form used for current prediction.

935. Duration of Slack Water

The predicted times of slack water listed in the tidal current tables indicate the instant of zero velocity. There is a period each side of slack water, however, during which the current is so weak that for practical purposes it may be considered negligible. Table 4 of the tidal current tables gives, for various maximum currents, the approximate period of time during which currents not exceeding 0.1 to 0.5 knots will be encountered. This period includes the last of the flood or ebb and the beginning of the following flood or ebb; that is, half of the duration will be before and half after the time of slack water.

When there is a difference between the velocities of the maximum flood and ebb preceding and following the slack for which the duration is desired, it will be sufficiently accurate to find a separate duration for each maximum velocity and average the two to determine the duration of the weak current.

Of the two sub-tables of Table 4, Table A is used for all places except those listed for Table B; Table B is used for just the places listed and the stations in Table 2 which are referred to them.

936. Additional Tide Prediction Publications

NOS also publishes a special *Regional Tide and Tidal Current Table for New York Harbor to Chesapeake Bay*, and a *Tidal Circulation and Water Level Forecast Atlas for Delaware River and Bay.*

937. *Tidal Current Charts*

Tidal Current charts present a comprehensive view of the hourly velocity of current in different bodies of water. They also provide a means for determining the current's velocity at various locations in these waters. The arrows show the direction of the current; the figures give the speed in knots at the time of spring tides. A weak current is defined as less than 0.1 knot. These charts depict the flow of the tidal current under normal weather conditions. Strong winds and freshets, however, may cause nontidal currents, considerably modifying the velocity indicated on the charts.

Tidal Current charts are provided (1994) for Boston Harbor, Charleston Harbor SC, Long Island Sound and Block Island Sound, Narragansett Bay, Narragansett Bay to Nantucket Sound, Puget Sound (Northern Part), Puget Sound (Southern Part), Upper Chesapeake Bay, and Tampa Bay.

OPNAV 3530/40 (4-73)
VEL OF CURRENT

Date	
Location	
Time	
Ref Sta	
Time Diff Stack Water	
Time Diff Max Current	
Vel Ratio Max Flood	
Vel Ratio Max Ebb	
Flood Dir	
Ebb Dir	
Ref Sta Stack Water Time	
Time Diff	
Local Sta Stack Water Time	
Ref Sta Max Current Time	
Time Diff	
Local Sta Max Current Time	
Ref Sta Max Current Vel	
Vel Ratio	
Local Sta Max Current Vel	
Int Between Slack and Desired Time	
Int Between Slack and Max Current	
Max Current	
Factor Table 3	
Velocity	
Direction	

Figure 934. OPNAV 3530/41 Current Form.

The tidal current's velocity varies from day to day as a function of the phase, distance, and declination of the Moon. Therefore, to obtain the velocity for any particular day and hour, the spring velocities shown on the charts must be modified by correction factors. A correction table given in the charts can be used for this purpose.

All of the charts except Narragansett Bay require the use of the annual *Tidal Current Tables*. Narragansett Bay requires use of the annual *Tide Tables*.

938. Current Diagrams

A current diagram is a graph showing the velocity of the current along a channel at different stages of the tidal current cycle. The current tables include diagrams for Martha's Vineyard and Nantucket Sounds (one diagram); East River, New York; New York Harbor; Delaware Bay and River (one diagram); and Chesapeake Bay. These diagrams are no longer published by NOS, but are available privately and remain useful as they are not ephemeral.

On Figure 938, each vertical line represents a given instant identified by the number of hours before or after slack water at The Narrows. Each horizontal line represents a distance from Ambrose Channel entrance, measured along the usually traveled route. The names along the left margin are placed at the correct distances from Ambrose Channel entrance. The current is for the center of the channel opposite these points. The intersection of any vertical line with any horizontal line represents a given moment in the current cycle at a given place in the channel. If this intersection is in a shaded area, the current is flooding; if in an unshaded area, it is ebbing. The velocity can be found by interpolation between the numbers given in the diagram. The given values are averages. To find the value at any time, multiply the velocity found from the diagram by the ratio of maximum velocity of the current involved to the maximum shown on the diagram. If the diurnal inequality is large, the accuracy can be improved by altering the width of the shaded area to fit conditions. The diagram covers 1 1/2 current cycles, so that the right 1/3 duplicates the left 1/3.

Use Table 1 or 2 to determine the current for a single station. The current diagrams are intended for use in either of two ways: to determine a favorable time for passage through the channel and to find the average current to be expected during a passage through the channel. For both of these uses, a number of "velocity lines" are provided. When the appropriate line is transferred to the correct part of the diagram, the current to be encountered during passage is indicated along the line.

If the transferred velocity line is partly in a flood current area, all ebb currents (those increasing the ship's velocity) are given a positive sign (+), and all flood currents a negative sign (-). A separate ratio should be determined for each current (flood or ebb), and applied to the entries for that current. In the Chesapeake Bay, it is common for an outbound vessel to encounter three or even four separate currents during passage. Under the latter condition, it is good practice to multiply each current taken from the diagram by the ratio for the current involved.

If the time of starting the passage is fixed, and the current during passage is desired, the starting time is identified in terms of the reference tidal cycle. The velocity line is then drawn through the intersection of this vertical time line and the horizontal line through the place. The average current is then determined in the same manner as when the velocity line is located as described above.

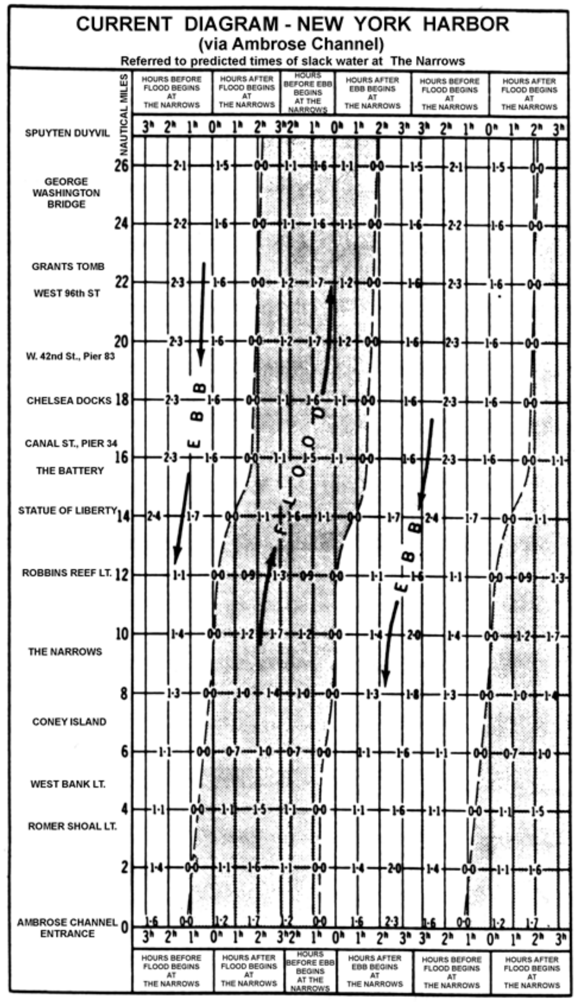


Figure 938. Current diagram for New York Harbor.

939. Computer Predictions

Until recently, tidal predictions were compiled only on mainframe or minicomputers and then put into hardcopy table form for the mariner. There are several types of commercial software available now for personal computers (PC's) that provide digital versions of the NOS tide tables and also graph the tidal heights. The tabular information and graphs can be printed for the desired locations for pre-voyage planning. There are also several types of specialized hand-held calculators and tide clocks that can be used to predict tides for local areas.

Newer versions of PC software use the actual harmonic constants available for locations, the prediction equation, and digital versions of Table 2 in the *Tide Tables* to produce even more products for the navigator's use. Since NOS has published the data, even inexpensive navigation electronics such as handheld GPS receivers and plotters for small craft navigation often include graphic tide tables.

Emerging applications include integration of tidal prediction with positioning systems and vessel traffic systems which are now moving towards full use of GPS. In addition, some electronic chart systems are already able to integrate tide prediction information. Many of these new systems will also use real-time water level and current information. Active research also includes providing predictions of total water level that will include not only the tidal prediction component, but also the weather-related component.

CHAPTER 10

RADIO WAVES

ELECTROMAGNETIC WAVE PROPAGATION

1000. Source of Radio Waves

Consider electric current as a flow of electrons along a conductor between points of differing potential. A **direct current** flows continuously in the same direction. This would occur if the polarity of the electromotive force causing the electron flow were constant, such as is the case with a battery. If, however, the current is induced by the relative motion between a conductor and a magnetic field, such as is the case in a rotating machine called a **generator**, then the resulting current changes direction in the conductor as the polarity of the electromotive force changes with the rotation of the generator's rotor. This is known as **alternating current**.

The energy of the current flowing through the conductor is either dissipated as heat (an energy loss proportional to both the current flowing through the conductor and the conductor's resistance) or stored in an electromagnetic field oriented symmetrically about the conductor. The orientation of this field is a function of the polarity of the source producing the current. When the current is removed from the wire, this electromagnetic field will, after a finite time, collapse back into the wire.

What would occur should the polarity of the current source supplying the wire be reversed at a rate which exceeds the finite amount of time required for the electromagnetic field to collapse back upon the wire? In this case, another magnetic field, proportional in strength but exactly opposite in magnetic orientation to the initial field, will be formed upon the wire. The initial magnetic field, its current source gone, cannot collapse back upon the wire because of the existence of this second electromagnetic field. Instead, it propagates out into space. This is the basic principle of a radio antenna, which transmits a wave at a frequency proportional to the rate of pole reversal and at a speed equal to the speed of light.

1001. Radio Wave Terminology

The magnetic field strength in the vicinity of a conductor is directly proportional to the magnitude of the current flowing through the conductor. Recall the discussion of alternating current above. A rotating generator produces current in the form of a sine wave. That is, the magnitude of the current varies as a function of the relative position of the rotating conductor and the stationary magnetic field used to induce the current. The current starts at zero, increases to a maximum as the rotor completes one quarter of its revolution, and falls to zero when the rotor completes one half of its revolution. The current then approaches a negative maximum; then it once again returns to zero. This cycle can be represented by a sine function.

The relationship between the current and the magnetic field strength induced in the conductor through which the current is flowing is shown in Figure 1001. Recall from the discussion above that this field strength is proportional to the magnitude of the current; that is, if the current is represented by a sine wave function, then so too will be the magnetic field strength resulting from that current. This characteristic shape of the field strength curve has led to the use of the term "wave" when referring to electromagnetic propagation. The maximum displacement of a peak from zero is called the **amplitude**. The forward side of any wave is called the **wave front**. For a non-directional antenna, each wave proceeds outward as an expanding sphere (or hemisphere).

One **cycle** is a complete sequence of values, as from crest to crest. The distance traveled by the energy during one cycle is the **wavelength**, usually expressed in metric units (meters, centimeters, etc.). The number of cycles repeated during unit time (usually 1 second) is the **frequency**. This is given in **hertz** (cycles per second). A kilohertz (kHz) is 1,000 cycles per second. A megahertz (MHz) is 1,000,000 cycles per second. Wavelength and frequency are inversely proportional.

The **phase** of a wave is the amount by which the cycle

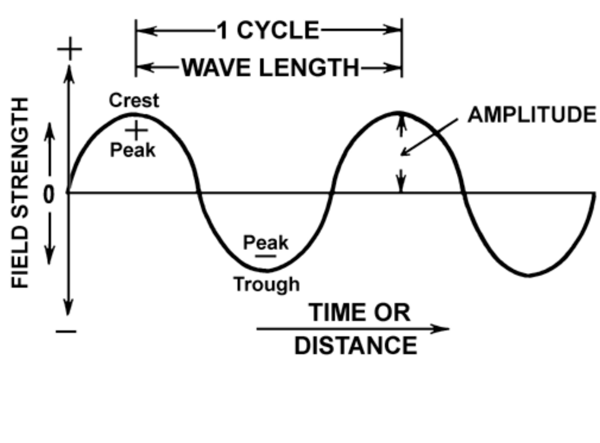


Figure 1001. Radio wave terminology.

has progressed from a specified origin. For most purposes it is stated in circular measure, a complete cycle being considered 360°. Generally, the origin is not important, principal interest being the phase relative to that of some other wave. Thus, two waves having crests 1/4 cycle apart are said to be 90° "out of phase." If the crest of one wave occurs at the trough of another, the two are 180° out of phase.

1002. The Electromagnetic Spectrum

The entire range of electromagnetic radiation frequencies is called the **electromagnetic spectrum**. The frequency range suitable for radio transmission, the **radio spectrum**, extends from 10 kilohertz to 300,000 megahertz. It is divided into a number of bands, as shown in Table 1002.

Below the radio spectrum, but overlapping it, is the audio frequency band, extending from 20 to 20,000 hertz. Above the radio spectrum are heat and infrared, the visible spectrum (light in its various colors), ultraviolet, X-rays, gamma rays, and cosmic rays. These are included in Table 1002. Waves shorter than 30 centimeters are usually called **microwaves**.

Within the frequencies from 1-40 gHz (1,000-40,000 MHz), additional bands are defined as follows:

L-band: 1-2 gHz (1,000-2,000 MHz)
S-band: 2-4 gHz (2,000-4,000 MHz
C-band: 4-8 gHz (4,000-8,000 MHz)
X-band: 8-12.5 gHz (8,000-12,500 MHz)
Lower K-band: 12.5-18 gHz (12,500-18,000 MHz)
Upper K-band: 26.5-40 gHz (26,500-40,000 MHz)

Marine radar systems commonly operate in the S and X bands, while satellite navigation system signals are found in the L-band.

The break of the K-band into lower and upper ranges is necessary because the resonant frequency of water vapor occurs in the middle region of this band, and severe absorption of radio waves occurs in this part of the spectrum.

Band	Abbreviation	Range of frequency	Range of wavelength
Audio frequency	AF	20 to 20,000 Hz	15,000,000 to 15,000 m
Radio frequency	RF	10 kHz to 300,000 MHz	30,000 m to 0.1 cm
Very low frequency	VLF	10 to 30 kHz	30,000 to 10,000 m
Low frequency	LF	30 to 300 kHz	10,000 to 1,000 m
Medium frequency	MF	300 to 3,000 kHz	1,000 to 100 m
High frequency	HF	3 to 30 MHz	100 to 10 m
Very high frequency	VHF	30 to 300 MHz	10 to 1 m
Ultra high frequency	UHF	300 to 3,000 MHz	100 to 10 cm
Super high frequency	SHF	3,000 to 30,000 MHz	10 to 1 cm
Extremely high frequency	EHF	30,000 to 300,000 MHz	1 to 0.1 cm
Heat and infrared*		10^6 to 3.9×10^8 MHz	0.03 to 7.6×10^{-5} cm
Visible spectrum*		3.9×10^8 to 7.9×10^8 MHz	7.6×10^{-5} to 3.8×10^{-5} cm
Ultraviolet*		7.9×10^8 to 2.3×10^{10} MHz	3.8×10^{-5} to 1.3×10^{-6} cm
X-rays*		2.0×10^9 to 3.0×10^{13} MHz	1.5×10^{-5} to 1.0×10^{-9} cm
Gamma rays*		2.3×10^{12} to 3.0×10^{14} MHz	1.3×10^{-8} to 1.0×10^{-10} cm
Cosmic rays*		$>4.8\times10^{15}$ MHz	$<6.2\times10^{-12}$ cm

* Values approximate.

Table 1002. Electromagnetic spectrum.

1003. Polarization

Radio waves produce both electric and magnetic fields. The direction of the electric component of the field is called the **polarization** of the electromagnetic field. Thus, if the electric component is vertical, the wave is said to be "vertically polarized," and if horizontal, "horizontally polarized."

A wave traveling through space may be polarized in any direction. One traveling along the surface of the Earth is always vertically polarized because the Earth, a conductor, short-circuits any horizontal component. The magnetic field and the electric field are always mutually perpendicular.

1004. Reflection

When radio waves strike a surface, the surface reflects them in the same manner as light waves. Radio waves of all frequencies are reflected by the surface of the Earth. The strength of the reflected wave depends upon angle of incidence (the angle between the incident ray and the horizontal), type of polarization, frequency, reflecting properties of the surface, and divergence of the reflected ray. Lower frequencies penetrate the earth's surface more than higher ones. At very low frequencies, usable radio signals can be received some distance below the surface of the sea.

A phase change occurs when a wave is reflected from the surface of the Earth. The amount of the change varies with the conductivity of the Earth and the polarization of the wave, reaching a maximum of 180° for a horizontally polarized wave reflected from sea water (considered to have infinite conductivity).

When direct waves (those traveling from transmitter to receiver in a relatively straight line, without reflection) and reflected waves arrive at a receiver, the total signal is the vector sum of the two. If the signals are in phase, they reinforce each other, producing a stronger signal. If there is a phase difference, the signals tend to cancel each other, the cancellation being complete if the phase difference is 180° and the two signals have the same amplitude. This interaction of waves is called **wave interference**.

A phase difference may occur because of the change of phase of a reflected wave, or because of the longer path it follows. The second effect decreases with greater distance between transmitter and receiver, for under these conditions the difference in path lengths is smaller.

At lower frequencies there is no practical solution to interference caused in this way. For VHF and higher frequencies, the condition can be improved by elevating the antenna, if the wave is vertically polarized. Additionally, interference at higher frequencies can be more nearly eliminated because of the greater ease of beaming the signal to avoid reflection.

Reflections may also occur from mountains, trees, and other obstacles. Such reflection is negligible for lower frequencies, but becomes more prevalent as frequency increases. In radio communication, it can be reduced by using directional antennas, but this solution is not always available for navigational systems.

Various reflecting surfaces occur in the atmosphere. At high frequencies, reflections take place from rain. At still higher frequencies, reflections are possible from clouds, particularly rain clouds. Reflections may even occur at a sharply defined boundary surface between air masses, as when warm, moist air flows over cold, dry air. When such a surface is roughly parallel to the surface of the Earth, radio waves may travel for greater distances than normal The principal source of reflection in the atmosphere is the ionosphere.

1005. Refraction

Refraction of radio waves is similar to that of light waves. Thus, as a signal passes from air of one density to that of a different density, the direction of travel is altered. The principal cause of refraction in the atmosphere is the difference in temperature and pressure occurring at various heights and in different air masses.

Refraction occurs at all frequencies, but below 30 MHz the effect is small as compared with ionospheric effects, diffraction, and absorption. At higher frequencies, refraction in the lower layer of the atmosphere extends the radio horizon to a distance about 15 percent greater than the visible horizon. The effect is the same as if the radius of the Earth were about one-third greater than it is and there were no refraction.

Sometimes the lower portion of the atmosphere becomes stratified. This stratification results in nonstandard temperature and moisture changes with height. If there is a marked temperature inversion or a sharp decrease in water vapor content with increased height, a horizontal radio duct may be formed. High frequency radio waves traveling horizontally within the duct are refracted to such an extent that they remain within the duct, following the curvature of the Earth for phenomenal distances. This is called **super-refraction**. Maximum results are obtained when both transmitting and receiving antennas are within the duct. There is a lower limit to the frequency affected by ducts. It varies from about 200 MHz to more than 1,000 MHz.

At night, surface ducts may occur over land due to cooling of the surface. At sea, surface ducts about 50 feet thick may occur at any time in the trade wind belt. Surface ducts 100 feet or more in thickness may extend from land out to sea when warm air from the land flows over the cooler ocean surface. Elevated ducts from a few feet to more than 1,000 feet in thickness may occur at elevations of 1,000 to 5,000 feet, due to the settling of a large air mass. This is a frequent occurrence in Southern California and certain areas of the Pacific Ocean.

A bending in the horizontal plane occurs when a groundwave crosses a coast at an oblique angle. This is due to a marked difference in the conducting and reflecting properties of the land and water over which the wave travels. The effect is known as **coastal refraction** or **land effect**.

1006. The Ionosphere

Since an atom normally has an equal number of negatively charged electrons and positively charged protons, it is electrically neutral. An **ion** is an atom or group of atoms which has become electrically charged, either positively or negatively, by the loss or gain of one or more electrons.

Loss of electrons may occur in a variety of ways. In the atmosphere, ions are usually formed by collision of atoms with rapidly moving particles, or by the action of cosmic rays or ultraviolet light. In the lower portion of the atmosphere, recombination soon occurs, leaving a small percentage of ions. In thin atmosphere far above the surface of the Earth, however, atoms are widely separated and a large number of ions may be present. The region of numerous positive and negative ions and unattached electrons is called the **ionosphere**. The extent of ionization depends upon the kinds of atoms present in the atmosphere, the density of the atmosphere, and the position relative to the Sun (time of day and season). After sunset, ions and electrons recombine faster than they are separated, decreasing the ionization of the atmosphere.

An electron can be separated from its atom only by the application of greater energy than that holding the electron. Since the energy of the electron depends primarily upon the kind of an atom of which it is a part, and its position relative to the nucleus of that atom, different kinds of radiation may cause ionization of different substances.

In the outermost regions of the atmosphere, the density is so low that oxygen exists largely as separate atoms, rather than combining as molecules as it does nearer the surface of the Earth. At great heights the energy level is low and ionization from solar radiation is intense. This is known as the **F layer**. Above this level the ionization decreases because of the lack of atoms to be ionized. Below this level it decreases because the ionizing agent of appropriate energy has already been absorbed. During daylight, two levels of maximum F ionization can be detected, the F_2 layer at about 125 statute miles above the surface of the Earth, and the F_1 layer at about 90 statute miles. At night, these combine to form a single F layer.

At a height of about 60 statute miles, the solar radiation not absorbed by the F layer encounters, for the first time, large numbers of oxygen molecules. A new maximum ionization occurs, known as the **E layer**. The height of this layer is quite constant, in contrast with the fluctuating F layer. At night the E layer becomes weaker by two orders of magnitude.

Below the E layer, a weak D layer forms at a height of about 45 statute miles, where the incoming radiation encounters ozone for the first time. The D layer is the principal source of absorption of HF waves, and of reflection of LF and VLF waves during daylight.

1007. The Ionosphere and Radio Waves

When a radio wave encounters a particle having an electric charge, it causes that particle to vibrate. The vibrating particle absorbs electromagnetic energy from the radio wave and radiates it. The net effect is a change of polarization and an alteration of the path of the wave. That portion of the wave in a more highly ionized region travels faster, causing the wave front to tilt and the wave to be directed toward a region of less intense ionization.

Refer to Figure 1007a, in which a single layer of the ionosphere is considered. Ray 1 enters the ionosphere at such an angle that its path is altered, but it passes through and proceeds outward into space. As the angle with the horizontal decreases, a critical value is reached where ray 2 is bent or reflected back toward the Earth. As the angle is still further decreased, such as at 3, the return to Earth occurs at a greater distance from the transmitter.

A wave reaching a receiver by way of the ionosphere is called a **skywave**. This expression is also appropriately applied to a wave reflected from an air mass boundary. In common usage, however, it is generally associated with the ionosphere. The wave which travels along the surface of the Earth is called a **groundwave**. At angles greater than the critical angle, no skywave signal is received. Therefore, there is a minimum distance from the transmitter at which skywaves can be received. This is called the **skip distance**, shown in Figure 1007a. If the groundwave extends out for less distance than the skip distance, a skip zone occurs, in which no signal is received.

The critical radiation angle depends upon the intensity of ionization, and the frequency of the radio wave. As the frequency increases, the angle becomes smaller. At frequencies greater than about 30 MHz virtually all of the energy penetrates through or is absorbed by the ionosphere. Therefore, at any given receiver there is a maximum usable frequency if skywaves are to be utilized. The strongest signals are received at or slightly below this frequency. There is also a lower practical frequency beyond which signals are too weak to be of value. Within this band the optimum frequency can be selected to give best results. It cannot be too near the maximum usable frequency because this frequency fluctuates with changes of intensity within the ionosphere. During magnetic storms the ionosphere density decreases. The maximum usable frequency decreases, and the lower usable frequency increases. The band of usable frequencies is thus narrowed. Under extreme conditions it may be completely eliminated, isolating the receiver and causing a radio blackout.

Skywave signals reaching a given receiver may arrive by any of several paths, as shown in Figure 1007b. A signal which undergoes a single reflection is called a “one-hop” signal, one which undergoes two reflections with a ground reflection between is called a “two-hop” signal, etc. A

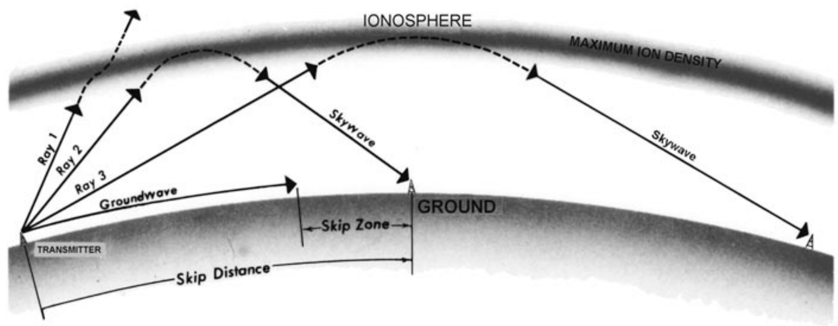


Figure 1007a. The effect of the ionosphere on radio waves.

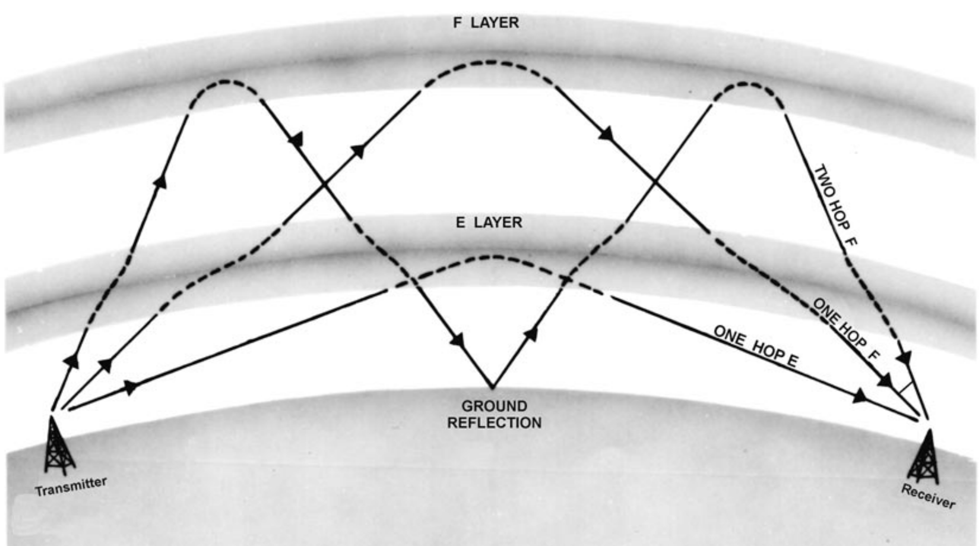


Figure 1007b. Various paths by which a skywave signal might be received.

"multihop" signal undergoes several reflections. The layer at which the reflection occurs is usually indicated, also, as "one-hop E," "two-hop F," etc.

Because of the different paths and phase changes occurring at each reflection, the various signals arriving at a receiver have different phase relationships. Since the density of the ionosphere is continually fluctuating, the strength and phase relationships of the various signals may undergo an almost continuous change. Thus, the various signals may reinforce each other at one moment and cancel each other at the next, resulting in fluctuations of the strength of the total signal received. This is called **fading**. This phenomenon may also be caused by interaction of components within a single reflected wave, or changes in its strength due to changes in the reflecting surface. Ionospheric changes are associated with fluctuations in the radiation received from the Sun, since this is the principal cause of ionization. Signals from the F layer are particularly erratic because of the rapidly fluctuating conditions within the layer itself.

The maximum distance at which a one-hop E signal can be received is about 1,400 miles. At this distance the signal leaves the transmitter in approximately a horizontal direction. A one-hop F signal can be received out to about 2,500 miles. At low frequencies groundwaves extend out for great distances.

A skywave may undergo a change of polarization during reflection from the ionosphere, accompanied by an alteration in the direction of travel of the wave. This is called **polarization error**. Near sunrise and sunset, when rapid changes are occurring in the ionosphere, reception may become erratic and polarization error a maximum. This is called **night effect**.

1008. Diffraction

When a radio wave encounters an obstacle, its energy is reflected or absorbed, causing a shadow beyond the obstacle. However, some energy does enter the shadow area because of diffraction. This is explained by Huygens' principle, which

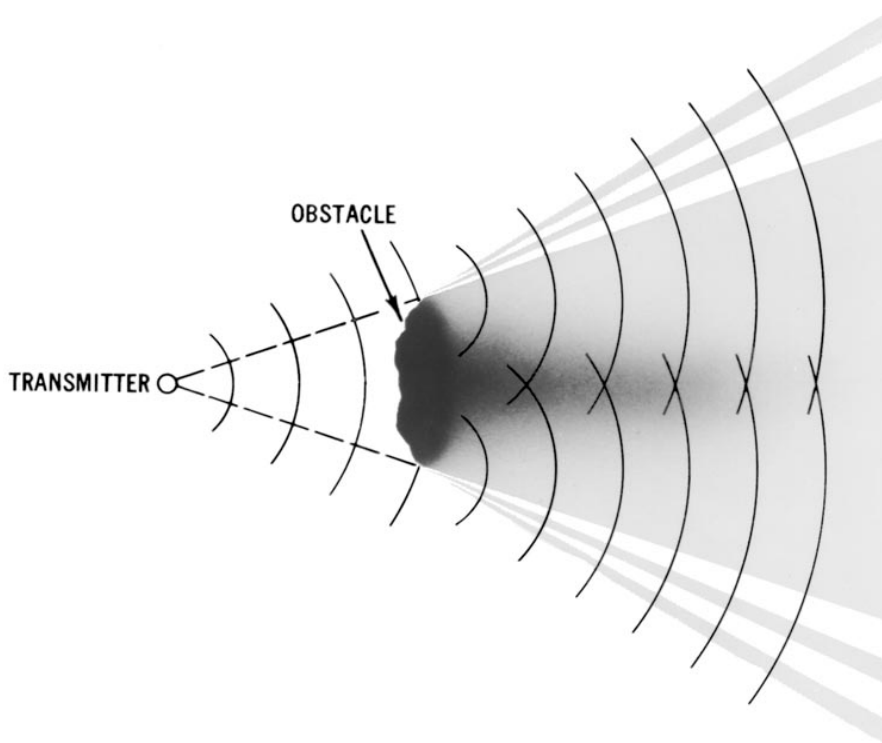


Figure 1008. Diffraction.

states that every point on the surface of a wave front is a source of radiation, transmitting energy in all directions ahead of the wave. No noticeable effect of this principle is observed until the wave front encounters an obstacle, which intercepts a portion of the wave. From the edge of the obstacle, energy is radiated into the shadow area, and also outside of the area. The latter interacts with energy from other parts of the wave front, producing alternate bands in which the secondary radiation reinforces or tends to cancel the energy of the primary radiation. Thus, the practical effect of an obstacle is a greatly reduced signal strength in the shadow area, and a disturbed pattern for a short distance outside the shadow area. This is illustrated in Figure 1008.

The amount of diffraction is inversely proportional to the frequency, being greatest at very low frequencies.

1009. Absorption and Scattering

The amplitude of a radio wave expanding outward through space varies inversely with distance, weakening with increased distance. The decrease of strength with distance is called **attenuation**. Under certain conditions the attenuation is greater than in free space.

A wave traveling along the surface of the Earth loses a certain amount of energy to the Earth. The wave is diffracted downward and absorbed by the Earth. As a result of this absorption, the remainder of the wave front tilts downward, resulting in further absorption by the Earth. Attenuation is greater over a surface which is a poor conductor. Relatively little absorption occurs over sea water, which is an excellent conductor at low frequencies, and low frequency groundwaves travel great distances over water.

A skywave suffers an attenuation loss in its encounter with the ionosphere. The amount depends upon the height and composition of the ionosphere as well as the frequency of the radio wave. Maximum ionospheric absorption occurs at about 1,400 kHz.

In general, atmospheric absorption increases with frequency. It is a problem only in the SHF and EHF frequency range. At these frequencies, attenuation is further increased by scattering due to reflection by oxygen, water vapor, water droplets, and rain in the atmosphere.

1010. Noise

Unwanted signals in a receiver are called **interference**. The intentional production of such interference to obstruct communication is called **jamming**. Unintentional interference is called **noise**.

Noise may originate within the receiver. Hum is usually the result of induction from neighboring circuits carrying alternating current. Irregular crackling or sizzling sounds may be caused by poor contacts or faulty components within the receiver. Stray currents in normal components cause some noise. This source sets the ultimate limit of sensitivity that can be achieved in a receiver. It is the same at any frequency.

Noise originating outside the receiver may be either man-made or natural. Man-made noises originate in electrical appliances, motor and generator brushes, ignition systems, and other sources of sparks which transmit electromagnetic signals that are picked up by the receiving antenna.

Natural noise is caused principally by discharge of static electricity in the atmosphere. This is called **atmospheric noise**, atmospherics, or static. An extreme example is a thunderstorm. An exposed surface may acquire a considerable charge of static electricity. This may be caused by friction of water or solid particles blown against or along such a surface. It may also be caused by splitting of a water droplet which strikes the surface, one part of the droplet requiring a positive charge and the other a negative charge. These charges may be transferred to the surface. The charge tends to gather at points and ridges of the conducting surface, and when it accumulates to a sufficient extent to overcome the insulating properties of the atmosphere, it discharges into the atmosphere. Under suitable conditions this becomes visible and is known as St. Elmo's fire, which is sometimes seen at mastheads, the ends of yardarms, etc.

Atmospheric noise occurs to some extent at all frequencies but decreases with higher frequencies. Above about 30 MHz it is not generally a problem.

1011. Antenna Characteristics

Antenna design and orientation have a marked effect upon radio wave propagation. For a single-wire antenna, strongest signals are transmitted along the perpendicular to the wire, and virtually no signal in the direction of the wire. For a vertical antenna, the signal strength is the same in all horizontal directions. Unless the polarization undergoes a change during transit, the strongest signal received from a vertical transmitting antenna occurs when the receiving antenna is also vertical.

For lower frequencies the radiation of a radio signal takes place by interaction between the antenna and the ground. For a vertical antenna, efficiency increases with greater length of the antenna. For a horizontal antenna, efficiency increases with greater distance between antenna and ground. Near-maximum efficiency is attained when this distance is one-half wavelength. This is the reason for elevating low frequency antennas to great heights. However, at the lowest frequencies, the required height becomes prohibitively great. At 10 kHz it would be about 8 nautical miles for a half-wavelength antenna. Therefore, lower frequency antennas are inherently inefficient. This is partly offset by the greater range of a low frequency signal of the same transmitted power as one of higher frequency.

At higher frequencies, the ground is not used, both conducting portions being included in a dipole antenna. Not only can such an antenna be made efficient, but it can also be made sharply directive, thus greatly increasing the strength of the signal transmitted in a desired direction.

The power received is inversely proportional to the square of the distance from the transmitter, assuming there is no attenuation due to absorption or scattering.

1012. Range

The range at which a usable signal is received depends upon the power transmitted, the sensitivity of the receiver, frequency, route of travel, noise level, and perhaps other factors. For the same transmitted power, both the groundwave and skywave ranges are greatest at the lowest frequencies, but this is somewhat offset by the lesser efficiency of antennas for these frequencies. At higher frequencies, only direct waves are useful, and the effective range is greatly reduced. Attenuation, skip distance, ground reflection, wave interference, condition of the ionosphere, atmospheric noise level, and antenna design all affect the distance at which useful signals can be received.

1013. Radio Wave Spectra

Frequency is an important consideration in radio wave propagation. The following summary indicates the principal effects associated with the various frequency bands, starting with the lowest and progressing to the highest usable radio frequency.

Very Low Frequency (VLF, 10 to 30 kHz): The VLF signals propagate between the bounds of the ionosphere and the Earth and are thus guided around the curvature of the Earth to great distances with low attenuation and excellent stability. Diffraction is maximum. Because of the long wavelength, large antennas are needed, and even these are inefficient, permitting radiation of relatively small amounts of power. Magnetic storms have little effect upon transmission because of the efficiency of the "Earth-ionosphere waveguide." During such storms, VLF signals may constitute the only source of radio communication over great distances. However, interference from atmospheric noise may be troublesome. Signals may be received from below the surface of the sea.

Low Frequency (LF, 30 to 300 kHz): As frequency is increased to the LF band and diffraction decreases, there is greater attenuation with distance, and range for a given power output falls off rapidly. However, this is partly offset by more efficient transmitting antennas. LF signals are most stable within groundwave distance of the transmitter. A wider bandwidth permits pulsed signals at 100 kHz. This allows separation of the stable groundwave pulse from the variable skywave pulse up to 1,500 km, and up to 2,000 km for overwater paths. The frequency for Loran C is in the LF band. This band is also useful for radio direction finding and time dissemination.

Medium Frequency (MF, 300 to 3,000 kHz): Groundwaves provide dependable service, but the range for a given power is reduced greatly. This range varies from about 400 miles at the lower portion of the band to about 15 miles at the upper end for a transmitted signal of 1 kilowatt. These values are influenced, however, by the power of the transmitter, the directivity and efficiency of the antenna, and the nature of the terrain over which signals travel. Elevating the antenna to obtain direct waves may improve the transmission. At the lower frequencies of the band, skywaves are available both day and night. As the frequency is increased, ionospheric absorption increases to a maximum at about 1,400 kHz. At higher frequencies the absorption decreases, permitting increased use of skywaves. Since the ionosphere changes with the hour, season, and sunspot cycle, the reliability of skywave signals is variable. By careful selection of frequency, ranges of as much as 8,000 miles with 1 kilowatt of transmitted power are possible, using multihop signals. However, the frequency selection is critical. If it is too high, the signals penetrate the ionosphere and are lost in space. If it is too low, signals are too weak. In general, skywave reception is equally good by day or night, but lower frequencies are needed at night. The standard broadcast band for commercial stations (535 to 1,605 kHz) is in the MF band.

High Frequency (HF, 3 to 30 MHz): As with higher medium frequencies, the groundwave range of HF signals is limited to a few miles, but the elevation of the antenna may increase the direct-wave distance of transmission. Also, the height of the antenna does have an important effect upon skywave transmission because the antenna has an "image" within the conducting Earth. The distance between antenna and image is related to the height of the antenna, and this distance is as critical as the distance between elements of an antenna system. Maximum usable frequencies fall generally within the HF band. By day this may be 10 to 30 MHz, but during the night it may drop to 8 to 10 MHz. The HF band is widely used for ship-to-ship and ship-to-shore communication.

Very High Frequency (VHF, 30 to 300 MHz): Communication is limited primarily to the direct wave, or the direct wave plus a ground-reflected wave. Elevating the antenna to increase the distance at which direct waves can be used results in increased distance of reception, even though some wave interference between direct and ground-reflected waves is present. Diffraction is much less than with lower frequencies, but it is most evident when signals cross sharp mountain peaks or ridges. Under suitable conditions, reflections from the ionosphere are sufficiently strong to be useful, but generally they are unavailable. There is relatively little interference from atmospheric noise in this band. Reasonably efficient directional antennas are possible with VHF. The VHF band is much used for communication.

Ultra High Frequency (UHF, 300 to 3,000 MHz): Skywaves are not used in the UHF band because the ionosphere is not sufficiently dense to reflect the waves, which pass through it into space. Groundwaves and ground-reflected waves are used, although there is some wave interference. Diffraction is negligible, but the radio horizon extends about 15 percent beyond the visible horizon, due principally to refraction. Reception of UHF signals is virtually free from fading and interference by atmospheric noise. Sharply directive antennas can be produced for transmission in this band, which is widely used for ship-to-ship and ship-to-shore communication.

Super High Frequency (SHF, 3,000 to 30,000 MHz): In the SHF band, also known as the microwave or as the centimeter wave band, there are no skywaves, transmission being entirely by direct and ground-reflected waves. Diffraction and interference by atmospheric noise are virtually nonexistent. Highly efficient, sharply directive antennas can be produced. Thus, transmission in this band is similar to that of UHF, but with the effects of shorter waves being greater. Reflection by clouds, water droplets, dust particles, etc., increases, causing greater scattering, increased wave interference, and fading. The SHF band is used for marine navigational radar.

Extremely High Frequency (EHF, 30,000 to 300,000 MHz): The effects of shorter waves are more pronounced in the EHF band, transmission being free from wave interference, diffraction, fading, and interference by atmospheric noise. Only direct and ground-reflected waves are available. Scattering and absorption in the atmosphere are pronounced and may produce an upper limit to the frequency useful in radio communication.

1014. Regulation of Frequency Use

While the characteristics of various frequencies are important to the selection of the most suitable one for any given purpose, these are not the only considerations. Confusion and extensive interference would result if every user had complete freedom of selection. Some form of regulation is needed. The allocation of various frequency bands to particular uses is a matter of international agreement. Within the United States, the Federal Communications Commission has responsibility for authorizing use of particular frequencies. In some cases a given frequency is allocated to several widely separated transmitters, but only under conditions which minimize interference, such as during daylight hours. Interference between stations is further reduced by the use of channels, each of a narrow band of frequencies. Assigned frequencies are separated by an arbitrary band of frequencies that are not authorized for use. In the case of radio aids to navigation and ship communications bands of several channels are allocated, permitting selection of band and channel by the user.

1015. Types of Radio Transmission

A series of waves transmitted at constant frequency and amplitude is called a continuous wave (CW). This cannot be heard except at the very lowest radio frequencies, when it may produce, in a receiver, an audible hum of high pitch.

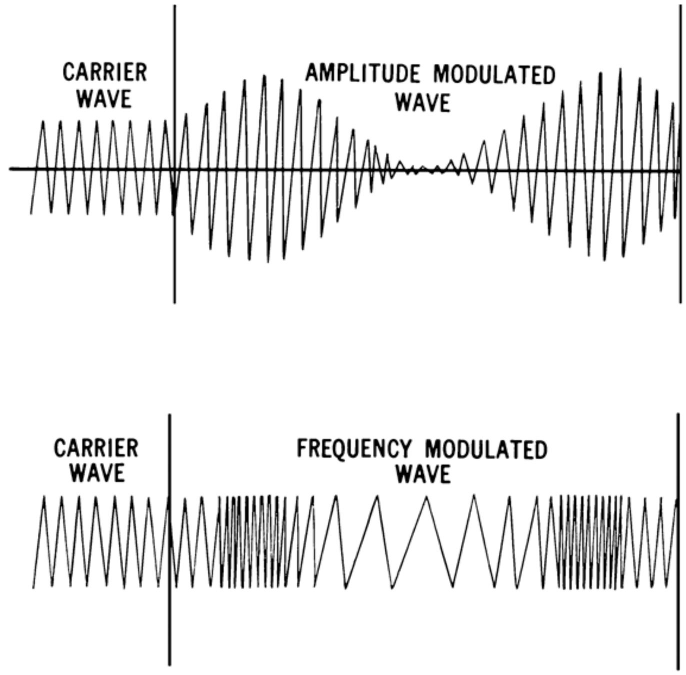


Figure 1015a. Amplitude modulation (upper figure) and frequency modulation (lower figure) by the same modulating wave.

Figure 1015b. Pulse modulation.

Although a continuous wave may be used directly, as in radiodirection finding or Decca, it is more commonly modified in some manner. This is called **modulation**. When this occurs, the continuous wave serves as a carrier wave for information. Any of several types of modulation may be used.

In **amplitude modulation (AM)** the amplitude of the carrier wave is altered in accordance with the amplitude of a modulating wave, usually of audio frequency, as shown in Figure 1015a. In the receiver the signal is demodulated by removing the modulating wave and converting it back to its original form. This form of modulation is widely used in voice radio, as in the standard broadcast band of commercial broadcasting.

If the frequency instead of the amplitude is altered in accordance with the amplitude of the impressed signal, as shown in Figure 1015a, **frequency modulation (FM)** occurs. This is used for commercial FM radio broadcasts and the sound portion of television broadcasts.

Pulse modulation (PM) is somewhat different, there being no impressed modulating wave. In this form of transmission, very short bursts of carrier wave are transmitted, separated by relatively long periods of "silence," during which there is no transmission. This type of transmission, illustrated in Figure 1015b, is used in some common radio navigational aids, including radar and Loran C.

1016. Transmitters

A radio transmitter consists essentially of (1) a power supply to furnish direct current, (2) an oscillator to convert direct current into radio-frequency oscillations (the carrier wave), (3) a device to control the generated signal, and (4) an amplifier to increase the output of the oscillator. For some transmitters a microphone is needed with a modulator and final amplifier to modulate the carrier wave. In addition, an antenna and ground (for lower frequencies) are needed to produce electromagnetic radiation. These components are illustrated in Figure 1016.

1017. Receivers

When a radio wave passes a conductor, a current is induced in that conductor. A radio receiver is a device which senses the power thus generated in an antenna, and transforms it into usable form. It is able to select signals of a single frequency (actually a narrow band of frequencies) from among the many which may reach the receiving antenna. The receiver is able to demodulate the signal and provide adequate amplification. The output of a receiver may be presented audibly by earphones or loudspeaker; or visually on a dial, cathode-ray tube, counter, or other

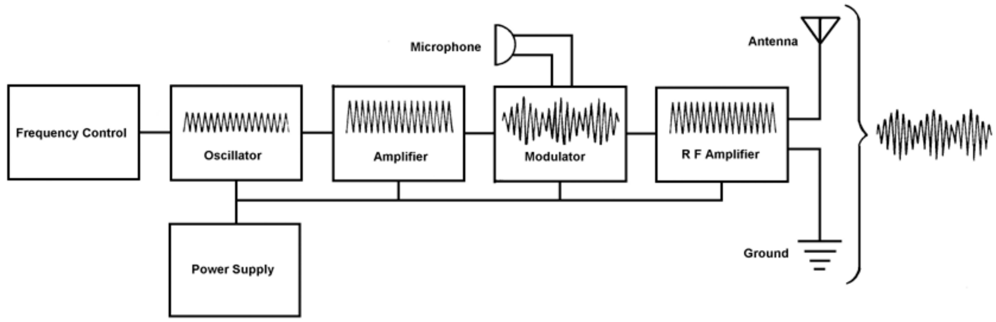


Figure 1016. Components of a radio transmitter.

display. Thus, the useful reception of radio signals requires three components: (1) an antenna, (2) a receiver, and (3) a display unit.

Radio receivers differ mainly in (1) frequency range, the range of frequencies to which they can be tuned; (2) selectivity, the ability to confine reception to signals of the desired frequency and avoid others of nearly the same frequency; (3) sensitivity, the ability to amplify a weak signal to usable strength against a background of noise; (4) stability, the ability to resist drift from conditions or values to which set; and (5) fidelity, the completeness with which the essential characteristics of the original signal are reproduced. Receivers may have additional features such as an automatic frequency control, automatic noise limiter, etc.

Some of these characteristics are interrelated. For instance, if a receiver lacks selectivity, signals of a frequency differing slightly from those to which the receiver is tuned may be received. This condition is called spillover, and the resulting interference is called crosstalk. If the selectivity is increased sufficiently to prevent spillover, it may not permit receipt of a great enough band of frequencies to obtain the full range of those of the desired signal. Thus, the fidelity may be reduced.

A transponder is a transmitter-receiver capable of accepting the challenge of an interrogator and automatically transmitting an appropriate reply.

U.S. RADIO NAVIGATION POLICY

1018. The Federal Radionavigation Plan

The ideal navigation system should provide three things to the user. First, it should be as accurate as necessary for the job it is expected to do. Second, it should be available 100% of the time, in all weather, at any time of day or night. Third, it should have 100% integrity, warning the user and shutting itself down when not operating properly. The mix of navigation systems in the U.S. is carefully chosen to provide maximum accuracy, availability, and integrity to all users, marine, aeronautical, and terrestrial, within the constraints of budget and practicality.

The Federal Radionavigation Plan (FRP) is produced by the U.S. Departments of Defense and Transportation. It establishes government policy on the mix of electronic navigation systems, ensuring consideration of national interests and efficient use of resources. It presents an integrated federal plan for all common-use civilian and military radionavigation systems, outlines approaches for consolidation of systems, provides information and schedules, defines and clarifies new or unresolved issues, and provides a focal point for user input. The FRP is a review of existing and planned radionavigation systems used in air, space, land, and marine navigation. It is available from the National Technical Information Service, Springfield, Virginia, 22161, http://www.ntis.gov.

The first edition of the FRP was released in 1980 as part of a presidential report to Congress. It marked the first time that a joint Department of Transportation/Department of Defense plan had been developed for systems used by both departments. The FRP has had international impact on navigation systems; it is distributed to the International Maritime Organization (IMO), the International Civil Aviation Organization (ICAO), the International Association of Lighthouse Authorities (IALA), and other international organizations.

During a national emergency, any or all of the systems may be temporarily discontinued by the federal government. The government's policy is to continue to operate radionavigation systems as long as the U.S. and its allies derive greater benefit than adversaries. Operating agencies may shut down systems or change signal formats and characteristics during such an emergency.

The plan is reviewed continually and updated biennially. Industry, advisory groups, and other interested parties provide input. The plan considers governmental responsibilities for national security, public safety, and transportation system economy. It is the official source of radionavigation systems policy and planning for the United States. Systems covered by the FRP include GPS, DGPS, WAAS, LAAS, Loran C, TACAN, MLS, VOR/VOR-DME/VORTAC, and ILS.

1019. System Plans

In order to meet both civilian and military needs, the federal government has established a number of different navigation systems. Each system utilizes the latest technology available at the time of implementation and is upgraded as technology and resources permit. The FRP addresses the length of time each system should be part of the system mix. The 2001 FRP sets forth the following system policy guidelines:

RADIOBEACONS: All U.S. marine radiobeacons have been discontinued and most of the stations converted into DGPS sites.

LORAN C: Loran C provides navigation, location, and timing services for both civil and military air, land, and maritime users. It is slated for replacement by GPS, but due to the large number of users, is expected to remain in place indefinitely while its continuation is evaluated. Reasonable notice will be given if the decision is made to terminate it.

GPS: The Global Positioning System, or GPS, will be the nation's primary radionavigation system well into the next century. It is operated by the U.S. Air Force.

1020. Enhancements to GPS

Differential GPS (DGPS): The U.S. Coast Guard operates marine DGPS in U.S. coastal waters. DGPS is a system in which differences between observed and calculated GPS signals are broadcast to users using medium frequencies. DGPS service is available in all U.S. coastal waters including Hawaii, Alaska, and the Great Lakes. It will provide 4-20 meter continuous accuracy. A terrestrial DGPS system is being installed across the United States to bring differential GPS service to land areas.

Wide Area Augmentation System (WAAS): WAAS is a service of the FAA similar to DGPS, and is intended for cross-country and local air navigation, using a series of reference stations and broadcasting correction data through geostationary satellites. WAAS is not optimized for marine use, and while not certified for maritime navigation, may provide additional position accuracy if the signal is unobstructed. Accuracies of a few meters are possible, about the same as with DGPS.

Local Area Augmentation System (LAAS): LAAS is a precision positioning system provided by the FAA for local navigation in the immediate vicinity of airports so equipped. The correctional signals are broadcast on HF radio with a range of about 30 miles. LAAS is not intended or configured for marine use, but can provide extremely accurate position data in a local area.

1021. Factors Affecting Navigation System Mix

The navigator relies on simple, traditional gear, and on some of the most complex and expensive space-based electronic systems man has ever developed. The success of GPS as a robust, accurate, available, and flexible system is rapidly driving older systems off the scene. Several have met their demise already (Transit, Omega, and marine radiobeacons in the U.S.), and the days are numbered for others, as GPS assumes primacy in navigation technology.

In the U.S., the Departments of Defense and Transportation continually evaluate the components which make up the federally provided and maintained radionavigation system. Several factors influence the decision on the proper mix of systems; cost, military utility, accuracy requirements, and user requirements all drive the problem of allocating scarce resources to develop and maintain navigation systems. The decreasing cost of receivers and increasing accuracy of the Global Positioning System increase its attractiveness as the primary navigation method of the future for both military and civilian use, although there are issues of reliability to be addressed in the face of threats to jam or otherwise compromise the system.

Many factors influence the choice of navigation systems, which must satisfy an extremely diverse group of users. International agreements must be honored. The current investment in existing systems by both government and users must be considered. The full life-cycle cost of each system must be considered. No system will be phased out without consideration of these factors. The FRP recognizes that GPS may not meet the needs of all users; therefore, some systems are currently being evaluated independently of GPS. The goal is to meet all military and civilian requirements in the most efficient way possible.

RADIO DIRECTION FINDING

1022. Introduction

The simplest use of radio waves in navigation is radio direction finding, in which a medium frequency radio signal is broadcast from a station at a known location. This signal is omnidirectional, but a directional antenna on a vessel is used to determine the bearing of the station. This constitutes an LOP, which can be crossed with another LOP to determine a fix.

Once used extensively throughout the world, radiobeacons have been discontinued in the U.S. and many other areas. They are now chiefly used as homing devices by local fishermen, and very little of the ocean's surface is covered by any radiobeacon signal. Because of its limited range, limited availability, and inherent errors, radio direction finding is of limited usefulness to the professional navigator.

In the past, when radiobeacon stations were powerful and common enough for routine ocean navigation, correction of radio bearings was necessary to obtain the most accurate LOP's. The correction process accounted for the fact that, while radio bearings travel along great circles, they are most often plotted on Mercator charts. The relatively short range of those stations remaining has made this process obsolete. Once comprising a major part of NIMA *Pub. 117, Radio Navigational Aid*s, radiobeacons are now listed in the back of each volume of the geographically appropriate *List of Lights*.

A Radio Direction Finding Station is one which the mariner can contact via radio and request a bearing. Most of these stations are for emergency use only, and a fee may be involved. These stations and procedures for use are listed in NIMA *Pub. 117, Radio Navigational Aids*.

1023. Using Radio Direction Finders

Depending upon the design of the RDF, the bearings of the radio transmissions are measured as relative bearings, or as both relative and true bearings. The most common type of marine radiobeacon transmits radio waves of approximately uniform strength in all directions. Except during calibration, radiobeacons operate continuously, regardless of weather conditions. Simple combinations of dots and dashes comprising Morse code letters are used for station identification. All radiobeacons superimpose the characteristic on a carrier wave, which is broadcast continuously during the period of transmission. A 10-second dash is incorporated in the characteristic signal to enable users of the aural null type of radio direction finder to refine the bearing.

Bearing measurement is accomplished with a directional antenna. Nearly all types of receiving antennas have some directional properties, but the RDF antenna is designed to be as directional as possible. Simple small craft RDF units usually have a ferrite rod antenna mounted directly on a receiver, with a 360 degree graduated scale. To get a bearing, align the unit to the vessel's course or to true north, and rotate the antenna back and forth to find the exact null point. The bearing to the station, relative or true according to the alignment, will be indicated on the dial. Some small craft RDF's have a portable hand-held combination ferrite rod and compass, with earphones to hear the null.

Two types of loop antenna are used in larger radio direction finders. In one of these, the crossed loop type, two loops are rigidly mounted in such manner that one is placed at 90 degrees to the other. The relative output of the two antennas is related to the orientation of each with respect to the direction of travel of the radio wave, and is measured by a device called a goniometer.

1024. Errors of Radio Direction Finders

RDF bearings are subject to certain errors. Quadrantal error occurs when radio waves arrive at a receiver and are influenced by the immediate shipboard environment.

A radio wave crossing a coastline at an oblique angle experiences a change of direction due to differences in conducting and reflecting properties of land and water known as coastal refraction, sometimes called land effect. It is avoided by not using, or regarding as of doubtful accuracy, bearings which cross a shoreline at an oblique angle.

In general, good radio bearings should not be in error by more than two or three degrees for distances under 150 nautical miles. However, conditions vary considerably, and skill is an important factor. By observing the technical instructions for the equipment and practicing frequently when results can be checked, one can develop skill and learn to what extent radio bearings can be relied upon under various conditions. Other factors affecting accuracy include range, the condition of the equipment, and the accuracy of calibration.

The strength of the signal determines the usable range of a radiobeacon. The actual useful range may vary considerably from the published range with different types of radio direction finders and during varying atmospheric conditions. The sensitivity of a radio direction finder determines the degree to which the full range of a radiobeacon can be utilized. Selectivity varies with the type of receiver and its condition.

CHAPTER 11

SATELLITE NAVIGATION

INTRODUCTION

1100. Development

The idea that led to development of the satellite navigation systems dates back to 1957 and the first launch of an artificial satellite into orbit, Russia's Sputnik I. Dr. William H. Guier and Dr. George C. Wieffenbach at the Applied Physics Laboratory of the Johns Hopkins University were monitoring the famous "beeps" transmitted by the passing satellite. They plotted the received signals at precise intervals, and noticed that a characteristic Doppler curve emerged. Since satellites generally follow fixed orbits, they reasoned that this curve could be used to describe the satellite's orbit. They then demonstrated that they could determine all of the orbital parameters for a passing satellite by Doppler observation of a single pass from a single fixed station. The Doppler shift apparent while receiving a transmission from a passing satellite proved to be an effective measuring device for establishing the satellite orbit.

Dr. Frank T. McClure, also of the Applied Physics Laboratory, reasoned in reverse: If the satellite orbit was known, Doppler shift measurements could be used to determine one's position on Earth. His studies in support of this hypothesis earned him the first National Aeronautics and Space Administration award for important contributions to space development.

In 1958, the Applied Physics Laboratory proposed exploring the possibility of an operational satellite Doppler navigation system. The Chief of Naval Operations then set forth requirements for such a system. The first successful launching of a prototype system satellite in April 1960 demonstrated the Doppler system's operational feasibility.

The **Navy Navigation Satellite System** (**NAVSAT**, also known as **TRANSIT**) was the first operational satellite navigation system. The system's accuracy was better than 0.1 nautical mile anywhere in the world, though its availability was somewhat limited. It was used primarily for the navigation of surface ships and submarines, but it also had some applications in air navigation. It was also used in hydrographic surveying and geodetic position determination.

The transit launch program ended in 1988 and the system was disestablished when the Global Positioning System became operational in 1996.

THE GLOBAL POSITIONING SYSTEM

1101. System Description

The Federal Radionavigation Plan has designated the NAVigation System using Timing And Ranging (NAVSTAR) Global Positioning System (GPS) as the primary navigation system of the U.S. government. GPS is a spaced-based radio positioning system which provides suitably equipped users with highly accurate position, velocity, and time data. It consists of three major segments: a **space segment**, a **control segment**, and a **user segment**.

The space segment comprises some 24 satellites. Spacing of the satellites in their orbits is arranged so that at least four satellites are in view to a user at any time, anywhere on the Earth. Each satellite transmits signals on two radio frequencies, superimposed on which are navigation and system data. Included in this data are predicted satellite ephemeris, atmospheric propagation correction data, satellite clock error information, and satellite health data. This segment normally consists of 21 operational satellites with three satellites orbiting as active spares. The satellites orbit at an altitude of 20,200 km, in six separate orbital planes, each plane inclined 55° relative to the equator. The satellites complete an orbit approximately once every 12 hours.

GPS satellites transmit **pseudorandom noise (PRN)** sequence-modulated radio frequencies, designated L1 (1575.42 MHz) and L2 (1227.60 MHz). The satellite transmits both a **Coarse Acquisition Code** (C/A code) and a **Precision Code** (P code). Both the P and C/A codes are transmitted on the L1 carrier; only the P code is transmitted on the L2 carrier. Superimposed on both the C/A and P codes is the navigation message. This message contains the satellite ephemeris data, atmospheric propagation correction data, and satellite clock bias.

GPS assigns a unique C/A code and a unique P code to each satellite. This practice, known as **code division multiple access (CDMA)**, allows all satellites the use of a common carrier frequency while still allowing the receiver to determine which satellite is transmitting. CDMA also allows for easy user identification of each GPS satellite. Since each satellite broadcasts using its own unique C/A and P code combination, it can be assigned a unique **PRN sequence number**. This number is how a satellite is identified when the GPS control system communicates with users about a particular GPS satellite.

The control segment includes a **master control station** (MCS), a number of monitor stations, and ground antennas located throughout the world. The master control station, located in Colorado Springs, Colorado, consists of equipment and facilities required for satellite monitoring, telemetry, tracking, commanding, control, uploading, and navigation message generation. The monitor stations, located in Hawaii, Colorado Springs, Kwajalein, Diego Garcia, and Ascension Island, passively track the satellites, accumulating ranging data from the satellites' signals and relaying them to the MCS. The MCS processes this information to determine satellite position and signal data accuracy, updates the navigation message of each satellite and relays this information to the ground antennas. The ground antennas then transmit this information to the satellites. The ground antennas, located at Ascension Island, Diego Garcia, and Kwajalein, are also used for transmitting and receiving satellite control information.

The user equipment is designed to receive and process signals from four or more orbiting satellites either simultaneously or sequentially. The processor in the receiver then converts these signals to navigation information. Since GPS is used in a wide variety of applications, from marine navigation to land surveying, these receivers can vary greatly in function and design.

1102. System Capabilities

GPS provides multiple users with accurate, continuous, worldwide, all-weather, common-grid, three-dimensional positioning and navigation information.

To obtain a navigation solution of position (latitude, longitude, and altitude) and time (four unknowns), four satellites must be used. The GPS user measures pseudorange and pseudorange rate by synchronizing and tracking the navigation signal from each of the four selected satellites. Pseudorange is the true distance between the satellite and the user plus an offset due to the user's clock bias. Pseudorange rate is the true slant range rate plus an offset due to the frequency error of the user's clock. By decoding the ephemeris data and system timing information on each satellite's signal, the user's receiver/processor can convert the pseudorange and pseudorange rate to three-dimensional position and velocity. Four measurements are necessary to solve for the three unknown components of position (or velocity) and the unknown user time (or frequency) bias.

The navigation accuracy that can be achieved by any user depends primarily on the variability of the errors in making pseudorange measurements, the instantaneous geometry of the satellites as seen from the user's location on Earth, and the presence of **Selective Availability** (SA). Selective Availability is discussed further below.

1103. Global Positioning System Concepts

GPS measures distances between satellites in orbit and a receiver on Earth, and computes spheres of position from those distances. The intersections of those spheres of position then determine the receiver's position.

The distance measurements described above are done by comparing timing signals generated simultaneously by the satellites' and receiver's internal clocks. These signals, characterized by a special wave form known as the pseudo-random code, are generated in phase with each other. The signal from the satellite arrives at the receiver following a time delay proportional to its distance traveled. This time delay is detected by the phase shift between the received pseudo-random code and the code generated by the receiver. Knowing the time required for the signal to reach the receiver from the satellite allows the receiver to calculate the distance from the satellite. The receiver, therefore, must be located on a sphere centered at the satellite with a radius equal to this distance measurement. The intersection of three spheres of position yields two possible points of receiver position. One of these points can be disregarded since it is hundreds of miles from the surface of the Earth. Theoretically, then, only three time measurements are required to obtain a fix from GPS.

In practice, however, a fourth measurement is required to obtain an accurate position from GPS. This is due to receiver clock error. Timing signals travel from the satellite to the receiver at the speed of light; even extremely slight timing errors between the clocks on the satellite and in the receiver will lead to tremendous range errors. The satellite's atomic clock is accurate to 10^{-9} seconds; installing a clock that accurate on a receiver would make the receiver prohibitively expensive. Therefore, receiver clock accuracy is sacrificed, and an additional satellite timing measurement is made. The fix error caused by the inaccuracies in the receiver clock is reduced by simultaneously subtracting a constant timing error from four satellite timing measurements until a pinpoint fix is reached.

Assuming that the satellite clocks are perfectly synchronized and the receiver clock's error is constant, the subtraction of that constant error from the resulting distance determinations will reduce the fix error until a "pinpoint" position is obtained. It is important to note here that the number of lines of position required to employ this technique is a function of the number of lines of position required to obtain a fix. GPS determines position in three dimensions; the presence of receiver clock error adds an additional unknown. Therefore, four timing measurements are required to solve for the resulting four unknowns.

1104. GPS Signal Coding

Two separate carrier frequencies carry the signal transmitted by a GPS satellite. The first carrier frequency (L1) transmits on 1575.42 MHz; the second (L2) transmits on 1227.60 MHz. The GPS signal consists of three separate messages: the P-code, transmitted on both L1 and L2; the C/A code, transmitted on L1 only; and a navigation data message. The P code and C/A code messages are divided into individual bits known as **chips**. The frequency at which bits are sent for each type of signal is known as the **chipping rate**. The chipping rate for the P-code is 10.23 MHz (10.23 $\times 10^6$ bits per second); for the C/A code, 1.023 MHz (1.023 $\times 10^6$ bits per second); and for the data message, 50 Hz (50 bits per second). The P and C/A codes **phase modulate** the carriers; the C/A code is transmitted at a phase angle of 90° from the P code. The periods of repetition for the C/A and P codes differ. The C/A code repeats once every millisecond; the P-code sequence repeats every seven days.

As stated above the GPS carrier frequencies are phase modulated. This is simply another way of saying that the digital "1's" and "0's" contained in the P and C/A codes are indicated along the carrier by a shift in the carrier phase. This is analogous to sending the same data along a carrier by varying its amplitude (amplitude modulation, or AM) or its frequency (frequency modulation, or FM). See Figure 1104a. In phase modulation, the frequency and the amplitude of the carrier are unchanged by the "information signal," and the digital information is transmitted by shifting the carrier's phase. The phase modulation employed by GPS is known as bi-phase shift keying (BPSK).

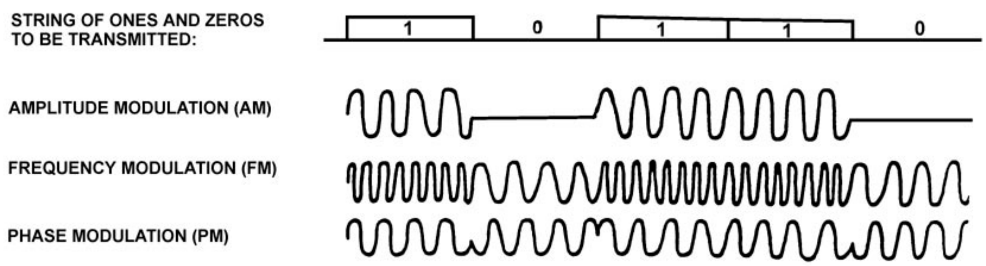


Figure 1104a. Digital data transmission with amplitude, frequency and phase modulation.

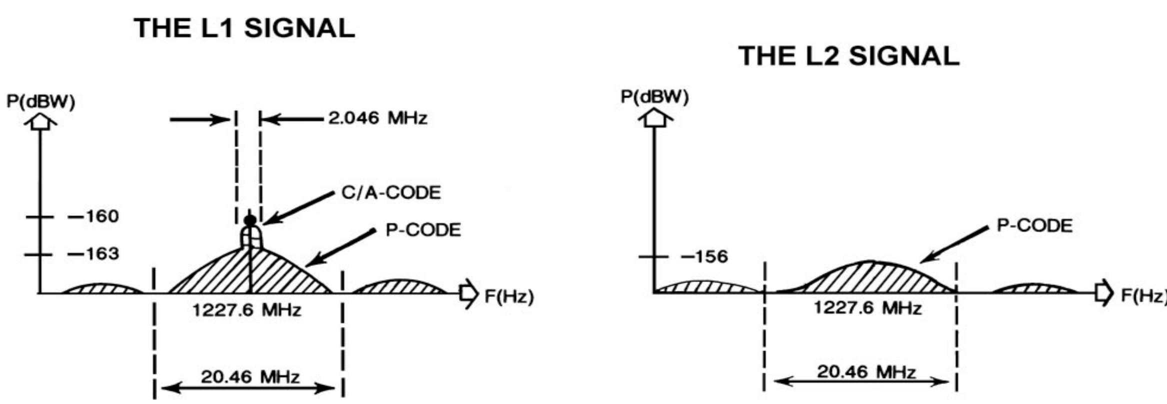


Figure 1104b. Modulation of the L1 and L2 carrier frequencies with the C/A and P code signals.

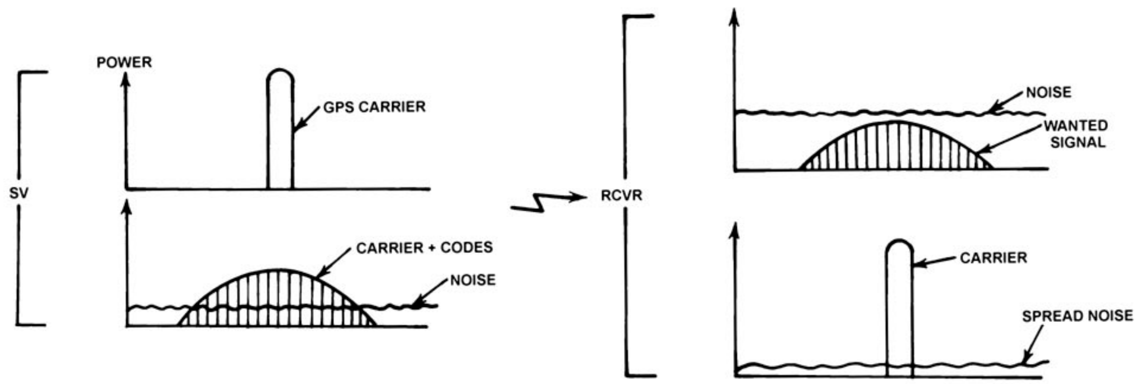


Figure 1104c. GPS signal spreading and recovery from satellite to receiver.

Due to this BPSK, the carrier frequency is "spread" about its center frequency by an amount equal to twice the "chipping rate" of the modulating signal. In the case of the P code, this spreading is equal to (2 × 10.23 MHz) = 20.46 MHz. For the C/A code, the spreading is equal to (2 × 1.023 MHz) = 2.046 MHz. See Figure 1104b. Note that the L1 carrier signal, modulated with both the P code and C/A code, is shaped differently from the L2 carrier, modulated with only the P code. This spreading of the carrier signal lowers the total signal strength below the thermal noise threshold present at the receiver. This effect is demonstrated in Figure 1104c. When the satellite signal is multiplied with the C/A and P codes generated by the receiver, the satellite signal will be collapsed into the original carrier frequency band. The signal power is then raised above the thermal noise level.

The navigation message is superimposed on both the P code and C/A code with a data rate of 50 bits per second (50 Hz.) The navigation message consists of 25 data frames, each frame consisting of 1500 bits. Each frame is divided into five subframes of 300 bits each. It will, therefore, take 30 seconds to receive one data frame and 12.5 minutes to receive all 25 frames. The navigation message contains GPS system time of transmission; a **handover word (HOW)**, allowing the transition between tracking the C/A code to the P code; ephemeris and clock data for the satellite being tracked; and almanac data for the satellites in orbit. It also contains coefficients for ionospheric delay models used by C/A receivers and coefficients used to calculate Universal Coordinated Time (UTC).

1105. The Correlation Process

The correlation process compares the signal received from the satellites with the signal generated by the receiver by comparing the square wave function of the received signal with the square wave function generated by the receiver. The computer logic of the receiver recognizes the square wave signals as either a +1 or a 0 depending on whether the signal is "on" or "off." The signals are processed and matched by using an **autocorrelation function**.

This process defines the necessity for a "pseudo-random code." The code must be repeatable (i.e., non-random) because it is in comparing the two signals that the receiver makes its distance calculations. At the same time, the code must be random for the correlation process to work; the randomness of the signals must be such that the matching process excludes all possible combinations except the combination that occurs when the generated signal is shifted a distance proportional to the received signal's time delay. These simultaneous requirements to be both repeatable (non-random) and random give rise to the description of "pseudo-random"; the signal has enough repeatability to enable the receiver to make the required measurement while simultaneously retaining enough randomness to ensure incorrect calculations are excluded.

1106. Precise Positioning Service and Standard Positioning Service

Two levels of navigational accuracy are provided by the GPS: the **Precise Positioning Service (PPS)** and the **Standard Positioning Service (SPS)**. GPS was designed, first and foremost, by the U.S. Department of Defense as a United States military asset; its extremely accurate positioning capability is an asset access to which the U.S. military may need to limit during time of war to prevent use by enemies. Therefore, the PPS is available only to authorized users, mainly the U.S. military and authorized allies. SPS, on the other hand, is available worldwide to anyone possessing a GPS receiver. Therefore PPS provides

SA/A-S Configuration	SIS Interface Conditions	PPS Users	SPS Users
SA Set to Zero A-S Off	P-Code, no errors C/A-Code, no errors	Full accuracy, spoofable	Full accuracy,* spoofable
SA at Non-Zero Value A-S Off	P-Code, errors C/A-Code, errors	Full accuracy, spoofable	Limited accuracy, spoofable
SA Set to Zero A-S On	Y-Code, no errors C/A-Code, no errors	Full accuracy, Not spoofable**	Full accuracy,*** spoofable
SA at Non-Zero Value A-S On	Y-Code, errors C/A-Code, errors	Full accuracy, Not spoofable**	Limited accuracy, spoofable

* "Full accuracy" defined as equivalent to a PPS-capable UE operated in a similar manner.
** Certain PPS-capable UE do not have P- or Y-code tracking abilities and remain spoofable despite A-S protection being applied
*** Assuming negligible accuracy degradation due to C/A-code operation (but more susceptible to jamming).

Figure 1106. Effect of SA and A-S on GPS accuracy.

a more accurate position than does SPS.

Two cryptographic methods are employed to deny PPS accuracy to civilian users: **selective availability (SA)** and **anti-spoofing (A-S)**. SA operates by introducing controlled errors into both the C/A and P code signals. SA can be programmed to degrade the signals' accuracy even further during time of war, denying a potential adversary the ability to use GPS to nominal SPS accuracy. SA introduces two errors into the satellite signal: (1) The **epsilon error**: an error in satellite ephemeris data in the navigation message; and (2) **clock dither**: error introduced in the satellite atomic clocks' timing. The presence of SA is the largest source of error present in an SPS GPS position measurement. The status of SA, whether off or on, can be checked at the USCG's NAVCEN Web site:

http://www.navcen.uscg.gov.

Anti-spoofing is designed to negate any hostile imitation of GPS signals. The technique alters the P code into another code, designated the Y code. The C/A code remains unaffected. The U.S. employs this technique to the satellite signals at random times and without warning; therefore, civilian users are unaware when this P code transformation takes place. Since anti-spoofing is applied only to the P code, the C/A code is not protected and can be spoofed.

Only users employing the proper cryptographic devices can defeat both SA and anti-spoofing. Without these devices, the user will be subject to the accuracy degradation of SA and will be unable to track the Y code.

GPS PPS receivers can use either the P code or the C/A code, or both, in determining position. Maximum accuracy is obtained by using the P code on both L1 and L2. The difference in propagation delay is then used to calculate ionospheric corrections. The C/A code is normally used to acquire the satellite signal and determine the approximate P code phase. Then, the receiver locks on the P code for precise positioning (subject to SA if not cryptographically equipped). Some PPS receivers possess a clock accurate enough to track and lock on the P code signal without initially tracking the C/A code. Some PPS receivers can track only the C/A code and disregard the P code entirely. Since the C/A code is transmitted on only one frequency, the dual frequency ionosphere correction methodology is unavailable and an ionospheric modeling procedure is required to calculate the required corrections.

SPS receivers, as mentioned above, provide positions with a degraded accuracy. The A-S feature denies SPS users access to the P code when transformed to the Y code. Therefore, the SPS user cannot rely on access to the P code to measure propagation delays between L1 and L2 and compute ionospheric delay corrections. Consequently, the typical SPS receiver uses only the C/A code because it is unaffected by A-S. Since C/A is transmitted only on L1, the dual frequency method of calculating ionospheric corrections is unavailable; an ionospheric modeling technique must be used. This is less accurate than the dual frequency method; this degradation in accuracy is accounted for in the 100-meter accuracy calculation. Figure 1106 presents the effect on SA and A-S on different types of GPS measurements.

1107. GPS Receiver Operations

In order for the GPS receiver to navigate, it has to track satellite signals, make pseudorange measurements, and collect navigation data.

A typical satellite tracking sequence begins with the receiver determining which satellites are available for it to track. Satellite visibility is determined by user-entered predictions of position, velocity, and time, and by almanac information stored internal to the receiver. If no stored almanac information exists, then the receiver must attempt to locate and lock onto the signal from any satellite in view. When the receiver is locked onto a satellite, it can demodulate the navigation message and read the almanac information about all the other satellites in the constellation. A carrier tracking loop tracks the carrier frequency while a code tracking loop tracks the C/A and P code signals. The two tracking loops operate together in an iterative process to acquire and track satellite signals.

The receiver's carrier tracking loop will locally generate an L1 carrier frequency which differs from the satellite produced L1 frequency due to a Doppler shift in the received frequency. This Doppler offset is proportional to the relative velocity along the line of sight between the satellite and the receiver, subject to a receiver frequency bias. The carrier tracking loop adjusts the frequency of the receiver-generated frequency until it matches the incoming frequency. This determines the relative velocity between the satellite and the receiver. The GPS receiver uses this relative velocity to calculate the velocity of the receiver. This velocity is then used to aid the code tracking loop.

The code tracking loop is used to make pseudorange measurements between the GPS receiver and the satellites. The receiver's tracking loop will generate a replica of the targeted satellite's C/A code with estimated ranging delay. In order to match the received signal with the internally generated replica, two things must be done: 1) The center frequency of the replica must be adjusted to be the same as the center frequency of the received signal; and 2) the phase of the replica code must be lined up with the phase of the received code. The center frequency of the replica is set by using the Doppler-estimated output of the carrier tracking loop. The receiver will then slew the code loop generated C/A code though a millisecond search window to correlate with the received C/A code and obtain C/A tracking.

Once the carrier tracking loop and the code tracking loop have locked onto the received signal and the C/A code has been stripped from the carrier, the navigation message is demodulated and read. This gives the receiver other information crucial to a pseudorange measurement. The navigation message also gives the receiver the handover word, the code that allows a GPS receiver to shift from C/A code tracking to P code tracking.

The handover word is required due to the long phase (seven days) of the P code signal. The C/A code repeats every millisecond, allowing for a relatively small search window. The seven day repeat period of the P code requires that the receiver be given the approximate P code phase to narrow its search window to a manageable time. The handover word provides this P code phase information. The handover word is repeated every subframe in a 30 bit long block of data in the navigation message. It is repeated in the second 30 second data block of each subframe. For some receivers, this handover word is unnecessary; they can acquire the P code directly. This normally requires the receiver to have a clock whose accuracy approaches that of an atomic clock. Since this greatly increases the cost of the receiver, most receivers for non-military marine use do not have this capability.

Once the receiver has acquired the satellite signals from four GPS satellites, achieved carrier and code tracking, and has read the navigation message, the receiver is ready to begin making pseudorange measurements. Recall that these measurements are termed *pseudorange* because a receiver clock offset makes them inaccurate; that is, they do not represent the true range from the satellite, only a range biased by a receiver clock error. This clock bias introduces a fourth unknown into the system of equations for which the GPS receiver must solve (the other three being the x coordinate, y coordinate, and z coordinate of the receiver position). Recall from the discussion in Article 1101 that the receiver solves this clock bias problem by making a fourth pseudorange measurement, resulting in a fourth equation to allow solving for the fourth unknown. Once the four equations are solved, the receiver has an estimate of the receiver's position in three dimensions and of GPS time. The receiver then converts this position into coordinates referenced to an Earth model based on the World Geodetic System (1984).

1108. User Range Errors and Geometric Dilution of Precision

There are two formal position accuracy requirements for GPS:

1) The PPS spherical position accuracy shall be 16 meters SEP (spherical error probable) or better.

2) The SPS user two dimensional position accuracy shall be 100 meters 2 drms or better.

Assume that a universal set of GPS pseudorange measurements results in a set of GPS position measurements. The accuracy of these measurements will conform to a normal (i.e. values symmetrically distributed around a mean of zero) probability function because the two most important factors affecting accuracy, the **geometric dilution of precision (GDOP)** and the **user equivalent range error (UERE)**, are continuously variable.

The UERE is the error in the measurement of the pseudoranges from each satellite to the user. The UERE is the product of several factors, including the clock stability, the predictability of the satellite's orbit, errors in the 50 Hz navigation message, the precision of the receiver's correlation process, errors due to atmospheric distortion and the calculations to compensate for it, and the quality of the satellite's signal. The UERE, therefore, is a random error which is the function of errors in both the satellites and the user's receiver.

The GDOP depends on the geometry of the satellites in relation to the user's receiver. It is independent of the quality of the broadcast signals and the user's receiver. Generally speaking, the GDOP measures the "spread" of the satellites around the receiver. The optimum case would be to have one satellite directly overhead and the other three spaced 120° around the receiver on the horizon. The worst GDOP would occur if the satellites were spaced closely together or in a line overhead.

There are special types of DOP's for each of the position and time solution dimensions; these particular DOP's combine to determine the GDOP. For the vertical dimension, the **vertical dilution of precision (VDOP)** describes the effect of satellite geometry on altitude calculations. The **horizontal dilution of precision (HDOP)** describes satellite geometry's effect on position (latitude and longitude) errors. These two DOP's combine

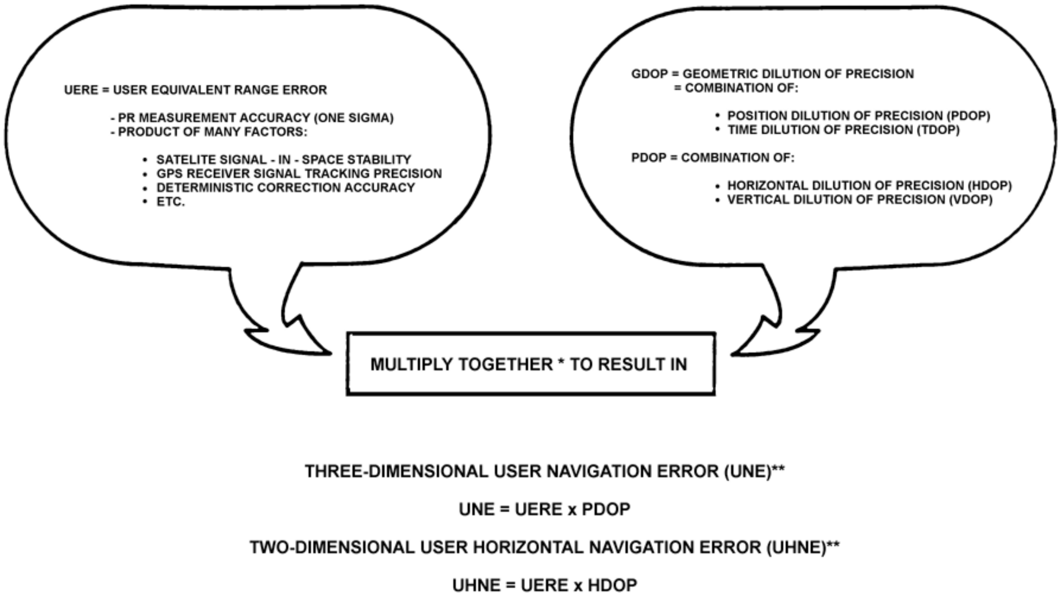


Figure 1108. Position and time error computations.

to determine the **position dilution of precision (PDOP)**. The PDOP combined with the **time dilution of precision (TDOP)** results in the GDOP. See Figure 1108.

1109. Ionospheric Delay Errors

Article 1108 covered errors in GPS positions due to errors inherent in the satellite signal (UERE) and the geometry of the satellite constellation (GDOP). Another major cause of accuracy degradation is the effect of the ionosphere on the radio frequency signals that comprise the GPS signal.

A discussion of a model of the Earth's atmosphere will be useful in understanding this concept. Consider the Earth as surrounded by three layers of atmosphere. The first layer, extending from the surface of the Earth to an altitude of approximately 10 km, is known as the troposphere. Above the troposphere and extending to an altitude of approximately 50 km is the stratosphere. Finally, above the stratosphere and extending to an altitude that varies as a function of the time of day is the **ionosphere**. Though radio signals are subjected to effects which degrade its accuracy in all three layers of this atmospheric model, the effects of the ionosphere are the most significant to GPS operation.

The ionosphere, as the name implies, is that region of the atmosphere which contains a large number of ionized molecules and a correspondingly high number of free electrons. These charged molecules have lost one or more electrons. No atom will loose an electron without an input of energy; the energy input that causes the ions to be formed in the ionosphere comes from the ultraviolet (U-V) radiation of the Sun. Therefore, the more intense the Sun's rays, the larger the number of free electrons which will exist in this region of the atmosphere.

The largest effect that this ionospheric effect has on GPS accuracy is a phenomenon known as **group time delay**. As the name implies, group time delay results in a delay in the time a signal takes to travel through a given distance. Obviously, since GPS relies on extremely accurate timing measurement of these signals between satellites and ground receivers, this group time delay can have a noticeable effect on the magnitude of GPS position error.

The group time delay is a function of several elements. It is inversely proportional to the square of the frequency at which the satellite transmits, and it is directly proportional to the atmosphere's **total electron content (TEC)**, a measure of the degree of the atmosphere's ionization. The general form of the equation describing the delay effect is:

$$\Delta t = \frac{(K \times TEC)}{f^2}$$

where

Δt = group time delay
f = operating frequency
K = constant

Since the Sun's U-V radiation ionizes the molecules in the upper atmosphere, it stands to reason that the time delay value will be highest when the Sun is shining and lowest at night. Experimental evidence has borne this out, showing that the value for TEC is highest around 1500 local time and lowest around 0500 local time. Therefore, the magnitude of the accuracy degradation caused by this effect will be highest during daylight operations. In addition to these daily variations, the magnitude of this time delay error also varies with the seasons; it is highest at the vernal equinox. Finally, this effect shows a solar cycle dependence. The greater the number of sunspots, the higher the TEC value and the greater the group time delay effect. The solar cycle typically follows an eleven year pattern. The next solar cycle will be at a minimum in 2006 and peak again in 2010.

Given that this ionospheric delay introduces a serious accuracy degradation into the system, how does GPS account for it? There are two methods used: (1) the dual frequency technique, and (2) the ionospheric delay method.

1110. Dual Frequency Correction Technique

As the term implies, the dual frequency technique requires the ability to acquire and track both the L1 and L2 frequency signals. Recall from the discussion in Article 1103 that the C/A and P codes are transmitted on carrier frequency L1, but only the P code is transmitted on L2. Recall also that only authorized operators with access to DOD cryptographic material are able to copy the P code. It follows, then, that only those authorized users are able to copy the L2 carrier frequency. Therefore, only those authorized users are able to use the dual frequency correction method. The dual frequency method measures the distance between the satellite and the user based on both the L1 and L2 carrier signal. These ranges will be different because the group time delay for each signal will be different. This is because of the frequency dependence of the time delay error. The range from the satellite to the user will be the true range combined with the range error caused by the time delay, as shown by the following equation:

$$R(f) = R_{actual} + \text{error term}$$

where R(f) is the range which differs from the actual range as a function of the carrier frequency. The dual frequency correction method takes two such range measurements, R(L1) and R(L2). Recall that the error term is a function of a constant divided by the square of the frequency. By combining the two range equations derived from the two frequency measurements, the constant term can be eliminated and one is left with an equation in which the true range is simply a function of the two carrier frequencies and the measured ranges R(L1) and R(L2). This method has two major advantages over the ionospheric model method. (1) It calculates corrections from real-time measured data; therefore, it is more accurate. (2) It alleviates the need to include ionospheric data on the navigation message. A significant portion of the data message is devoted to ionospheric correction data. If the receiver is dual frequency capable, then it does not need any of this data.

The vast majority of maritime users cannot copy dual frequency signals. For them, the ionospheric delay model provides the correction for the group time delay.

1111. The Ionospheric Delay Model

The ionospheric delay model mathematically models the diurnal ionospheric variation. The value for this time delay is determined from a cosinusoidal function into which coefficients representing the maximum value of the time delay (i.e., the amplitude of the cosine wave representing the delay function); the time of day; the period of the variation; and a minimum value of delay are introduced. This model is designed to be most accurate at the diurnal maximum. This is obviously a reasonable design consideration because it is at the time of day when the maximum diurnal time delay occurs that the largest magnitude of error appears. The coefficients for use in this delay model are transmitted to the receiver in the navigation data message. As stated in Article 1110, this method of correction is not as accurate as the dual frequency method; however, for the non-military user, it is the only method of correction available.

1112. Multipath Reflection Errors

Multipath reflection errors occur when the receiver detects parts of the same signal at two different times. The first reception is the direct path reception, the signal that is received directly from the satellite. The second reception is from a reflection of that same signal from the ground or any other reflective surface. The direct path signal arrives first, the reflected signal, having had to travel a longer distance to the receiver, arrives later. The GPS signal is designed to minimize this multipath error. The L1 and L2 frequencies used demonstrate a diffuse reflection pattern, lowering the signal strength of any reflection that arrives at the receiver. In addition, the receiver's antenna can be designed to reject a signal that it recognizes as a reflection. In addition to the properties of the carrier frequencies, the high data frequency of both the P and C/A codes and their resulting good correlation properties minimize the effect of multipath propagation.

The design features mentioned above combine to reduce the maximum error expected from multipath propagation to less than 20 feet.

DIFFERENTIAL GPS

1113. Differential GPS Concept

The discussions above make it clear that the Global Positioning System provides the most accurate positions available to navigators today. They should also make clear that the most accurate positioning information is available to only a small fraction of the using population: U.S. and allied military. For most open ocean navigation applications, the degraded accuracy inherent in selective availability and the inability to copy the precision code presents no serious hazard to navigation. A mariner seldom if ever needs greater than 100 meter accuracy in the middle of the ocean.

It is a different situation as the mariner approaches shore. Typically for harbor approaches and piloting, the mariner will shift to visual piloting. The increase in accuracy provided by this navigational method is required to ensure ship's safety. The 100 meter accuracy of GPS in this situation is not sufficient. Any mariner who has groped his way through a restricted channel in a thick fog will certainly appreciate the fact that even a degraded GPS position is available for them to plot. However, 100 meter accuracy is not sufficient to ensure ship's safety in most piloting situations. In this situation, the mariner needs P code accuracy. The problem then becomes how to obtain the accuracy of the Precise Positioning Service with due regard to the legitimate security concerns of the U.S. military. The answer to this seeming dilemma lies in the concept of **Differential GPS (DGPS)**.

Differential GPS is a system in which a receiver at an accurately surveyed position utilizes GPS signals to calculate timing errors and then broadcasts a correction signal to account for these errors. This is an extremely powerful concept. The errors which contribute to GPS accuracy degradation, ionospheric time delay and selective availability, are experienced simultaneously by both the DGPS receiver and a relatively close user's receiver. The extremely high altitude of the GPS satellites means that, as long as the DGPS receiver is within 100-200 km of the user's receiver, the user's receiver is close enough to take advantage of any DGPS correction signal.

The theory behind a DGPS system is straightforward. Located on an accurately surveyed site, the DGPS receiver already knows its location. It receives data which tell it where the satellite is. Knowing the two locations, it then calculates the theoretical time it should take for a satellite's signal to reach it. It then compares the time that it actually takes for the signal to arrive. This difference in time between the theoretical and the actual is the basis for the DGPS receiver's computation of a timing error signal; this difference in time is caused by all the errors to which the GPS signal is subjected; errors, except for receiver error and multipath error, to which both the DGPS and the user's receivers are simultaneously subject. The DGPS system then broadcasts a timing correction signal, the effect of which is to correct for selective availability, ionospheric delay, and all the other error sources the two receivers share in common.

For suitably equipped users, DGPS results in positions at least as accurate as those obtainable by the Precise Positioning Service. This capability is not limited to simply displaying the correct position for the navigator to plot. The DGPS position can be used as the primary input to an electronic chart system, providing an electronic readout of position accurate enough to pilot safely in the most restricted channel.

WAAS AND LAAS IN MARINE NAVIGATION

1114. WAAS/LAAS for Aeronautical Use

In 1994 the National Telecommunications and Information Administration (NTIA) produced a technical report for the Department of Transportation which concluded that the optimum mix of enhanced GPS systems for overall civilian use would consist of DGPS for marine and terrestrial use and a combined WAAS/LAAS system for air navigation.

The **Wide Area Augmentation System (WAAS)** concept is similar to the DGPS concept, except that correctional signals are sent from geostationary satellites via HF signals directly to the user's GPS receiver. This eliminates the need for a separate receiver and antenna, as is the case with DGPS. WAAS is intended for enroute air navigation, with 25 reference stations widely spaced across the United States, for coverage of the entire U.S. and parts of Mexico and Canada.

The **Local Area Augmentation System (LAAS)** is intended for precision airport approaches, with reference stations located at airports and broadcasting their correction message on VHF radio frequencies.

While many marine GPS receivers incorporate WAAS circuitry (but not the more accurate, shorter-range LAAS), WAAS is not optimized for surface navigation because the HF radio signals are line-of-sight and are transmitted from geostationary satellites. At low angles to the horizon, the WAAS signal may be blocked and the resulting GPS position accuracy significantly degraded with no warning. The DGPS signal, on the other hand, is a terrain-following signal that is unaffected by objects in its path. It simply flows around them and continues on unblocked.

The accuracy of WAAS and DGPS is comparable, on the order of a few meters. WAAS was designed to provide 7 meter accuracy 95% of the time. DGPS was designed to provide 10 meter accuracy 95% of the time, but in actual use one can expect about 1-3 meter accuracy when the user is within 100 miles of he DGPS transmitter. Over 100 miles, DGPS accuracy will commonly degrade by an additional 1 meter per 100 miles from the transmitter site. Both systems have been found in actual use to provide accuracies somewhat better than designed.

The WAAS signal, while not certified for use in the marine environment as is DGPS, can be a very useful navigational tool if its limitations are understood. In open waters of the continental U.S., the WAAS signal can be expected to be available and useful, provided the receiver has WAAS circuitry and is programmed to use the WAAS data. Outside the U.S., or in any area where tall buildings, trees, or other obstructions rise above the horizon, the WAAS signal may be blocked, and the resulting GPS fix could be in error by many meters. Since the highest accuracy is necessary in the most confined waters, WAAS should be used with extreme caution in these areas.

WAAS can enhance the navigator's situational awareness when available, but availability is not assured. Further, a marine receiver will provide no indication when WAAS data is not a part of the fix. [Aircraft GPS receivers may contain Receiver Autonomous Integrity Monitoring (RAIM) software, which does provide warning of WAAS satellite signal failure, and removes the affected signal from the fix solution.]

LAAS data, broadcast on VHF, is less subject to blocking, but is only available in selected areas near airports. Its range is about 30 miles. It is therefore not suitable for general marine navigational use.

NON-U.S. SATELLITE NAVIGATION SYSTEMS

1115. The Galileo System

Since the development of GPS, various European councils and commissions have expressed a need for a satellite navigation system independent of GPS. Economic studies have emphasized this need, and technological studies by the European Space Agency over several years have proven its feasibility. In early 2002 the European Union (EU) decided to fund the development of its new Galileo satellite navigation system. A great deal of preliminary scientific work has already been accomplished, which will enable the full deployment of Galileo over the next few years.

Several factors influenced the decision to develop Galileo, the primary one being that GPS is a U.S. military asset that can be degraded for civilian use on order of the U.S. Government (as is the Russian satellite navigation system GLONASS). Disruption of either system might leave European users without their primary navigation system at a critical time. In contrast, Galileo will be under civilian control and dedicated primarily to civilian use. It is important to note that since GPS has been operational, civilian uses are proliferating far more rapidly than anticipated, to the point that GPS planners are developing new frequencies and enhancements to GPS for civilian use (WAAS and LAAS), SA has been turned off (as of May 1, 2000), and the cost and size of receivers have plummeted.

Plans call for the Galileo constellation to consist of 30 satellites (27 usable and three spares) in three orbital planes, each inclined 56 degrees to the equator. The orbits are at an altitude of 23,616 km (about 12,750 nm). Galileo will be designed to serve higher latitudes than GPS, an additional factor in the EU decision, based on Scandinavian participation.

While U.S. GPS satellites are only launched one at a time, Galileo satellites are being designed with new miniaturization techniques that will allow several to be launched on the same rocket, a far more cost-efficient way to place them in orbit and maintain the constellation.

Galileo will also provide an important feature for civilian use that GPS does not: integrity monitoring. Currently, a civilian GPS user receives no indication that his unit is not receiving proper satellite signals, there being no provision for such notification in the code. However, Galileo will provide such a signal, alerting the user that the system is operating improperly.

The issue of compatibility with GPS is being addressed during ongoing development. Frequency sharing with GPS is under discussion, and it is reasonable to assume that a high degree of compatibility will exist when Galileo is operational. Manufacturers will undoubtedly offer a variety of systems which exploit the best technologies of both GPS and Galileo. Integration with existing shipboard electronic systems such as ECDIS and ECS will be ensured.

The benefit of Galileo for the navigator is that there will be two separate satellite navigation systems to rely on, providing not only redundancy, but also an increased degree of accuracy (for systems that can integrate both systems' signals). Galileo should be first available in 2005, and the full constellation is scheduled to be up by 2008.

1116. GLONASS

The Global Navigation Satellite System (GLONASS), under the control of the Russian military, has been in use since 1993, and is based on the same principles as GPS. The space segment consists of 24 satellites in three orbital planes, the planes separated by 120 degrees and the individual satellites by 45 degrees. The orbits are inclined to the equator at an angle of 64.8 degrees, and the orbital period is about 11hours, 15minutes at an altitude of 19,100 km (10,313 nm). The designed system fix accuracy for civilian use is 100 meters horizontal (95%), 150 meters vertical, and 15 cm/sec. in velocity. Military codes provide accuracies of some 10-20 meters horizontal.

The ground segment of GLONASS lies entirely within the former Soviet Union. Reliability has been an ongoing problem for the GLONASS system, but new satellite designs with longer life spans are addressing these concerns. The user segment consists of various types of receivers that provide position, time, and velocity information.

GLONASS signals are in the L-band, operating in 25 channels with 0.5625 MHz separation in 2 bands: from 1602.5625 MHz to 1615.5 MHz, and from 1240 to 1260 MHz.

CHAPTER 12

LORAN NAVIGATION

INTRODUCTION TO LORAN

1200. History and Role of Loran

The theory behind the operation of hyperbolic navigation systems was known in the late 1930's, but it took the urgency of World War II to speed development of the system into practical use. By early 1942, the British had an operating hyperbolic system in use designed to aid in long-range bomber navigation. This system, named Gee, operated on frequencies between 30 MHz and 80 MHz and employed "master" and "slave" transmitters spaced approximately 100 miles apart. The Americans were not far behind the British in development of their own system. By 1943, the U. S. Coast Guard was operating a chain of hyperbolic navigation transmitters that became Loran A (The term Loran was originally an acronym for LOng RAnge Navigation). By the end of the war, the network consisted of over 70 transmitters providing coverage over approximately 30% of the earth's surface.

In the late 1940's and early 1950's, experiments in low frequency Loran produced a longer range, more accurate system. Using the 90-110 kHz band, Loran developed into a 24-hour-a-day, all-weather radionavigation system named Loran C. From the late 1950's, Loran A and Loran C systems were operated in parallel until the mid 1970's when the U.S. Government began phasing out Loran A. The United States continued to operate Loran C in a number of areas around the world, including Europe, Asia, the Mediterranean Sea, and parts of the Pacific Ocean until the mid 1990's when it began closing its overseas Loran C stations or transferring them to the governments of the host countries. This was a result of the U.S. Department of Defense adopting the Global Positioning System (GPS) as its primary radionavigation service. In the United States, Loran serves the 48 contiguous states, their coastal areas and parts of Alaska. It provides navigation, location, and timing services for both civil and military air, land, and marine users. Loran systems are also operated in Canada, China, India, Japan, Northwest Europe, Russia, Saudi Arabia, and South Korea.

The future role of Loran depends on the radionavigation policies of the countries and international organizations that operate the individual chains. In the United States, the Federal Government plans to continue operating Loran in the short term while it evaluates the long-term need for the system. The U.S. Government will give users reasonable notice if it concludes that Loran is no longer needed or is not cost effective, so that users will have the opportunity to transition to alternative navigation aids and timing services.

Current information on the U.S. Loran system, including Notices to Mariners, may be obtained at the U.S. Coast Guard Navigation Center World Wide Web site at http://www.navcen.uscg.gov/.

LORAN C DESCRIPTION

1201. Summary of Operation

The Loran C (hereafter referred to simply as Loran) system consists of **transmitting stations**, which are placed several hundred miles apart and organized into **chains**. Within a Loran chain, one station is designated as the **master station** and the others as **secondary stations**. Every Loran chain contains at least one master station and two secondary stations in order to provide two lines of position.

The master and secondary stations transmit radio pulses at precise time intervals. A Loran receiver measures the **time difference** (TD) between when the vessel receives the master signal and when it receives each of the secondary signals. When this elapsed time is converted to distance, the locus of points having the same TD between the master and each secondary forms the hyperbolic LOP. The intersection of two or more of these LOP's produces a fix of the vessel's position.

There are two methods by which the navigator can convert this information into a geographic position. The first involves the use of a chart overprinted with a Loran **time delay lattice** consisting of hyperbolic TD lines spaced at convenient intervals. The navigator plots the displayed TD's by interpolating between the lattice lines printed on the chart, manually plots the fix where they intersect and then determines latitude and longitude. In the second method, computer algorithms in the receiver's software convert the TD's to latitude and longitude for display.

As with other computerized navigation receivers, a typical Loran receiver can accept and store **waypoints**.

Waypoints are sets of coordinates that describe either locations of navigational interest or points along a planned route. Waypoints may be entered by visiting the spot of interest and pressing the appropriate receiver control key, or by keying in the waypoint coordinates manually, either as a TD or latitude-longitude pair. If using waypoints to mark a planned route, the navigator can use the receiver to monitor the vessel's progress in relation to the track between each waypoint. By continuously providing parameters such as cross-track error, course over ground, speed over ground, and bearing and distance to next waypoint, the receiver continually serves as a check on the primary navigation plot.

1202. Components of the Loran System

For the marine navigator, the components of the Loran system consist of the land-based transmitting stations, the Loran **receiver** and **antenna**, the **Loran charts.** In addition to the master and secondary transmitting stations themselves, land-based Loran facilities also include the primary and secondary **system area monitor** sites, the **control station** and a precise time reference. The transmitters emit Loran signals at precisely timed intervals. The monitor sites and control stations continually measure and analyze the characteristics of the Loran signals received to detect any anomalies or out-of-specification conditions. Some transmitters serve only one function within a chain (i.e., either master or secondary). However, in many instances, one transmitter transmits signals for each of two adjacent chains. This practice is termed **dual rating**.

Loran receivers exhibit varying degrees of sophistication, but their signal processing is similar. The first processing stage consists of **search and acquisition**, during which the receiver searches for the signal from a particular Loran chain and establishes the approximate time reference of the master and secondaries with sufficient accuracy to permit subsequent settling and tracking.

After search and acquisition, the receiver enters the **settle** phase. In this phase, the receiver searches for and detects the front edge of the Loran pulse. After detecting the front edge of the pulse, it selects the correct cycle of the pulse to track.

Having selected the correct tracking cycle, the receiver begins the **tracking and lock** phase, in which the receiver maintains synchronization with the selected received signals. Once this phase is reached, the receiver displays either the time difference of the signals or the computed latitude and longitude.

1203. The Loran Signal

The Loran signal consists of a series of 100 kHz pulses sent first by the master station and then, in turn, by the secondary stations. Both the shape of the individual pulse and the pattern of the entire pulse sequence are shown in Figure 1203a. As compared to a carrier signal of constant amplitude, pulsed transmission allows the same signal range to be achieved with a lower average output power. Pulsed transmission also yields better signal identification properties and more precise timing of the signals.

The individual sinusoidal Loran pulse exhibits a steep rise to its maximum amplitude within 65 μsec of emission and an exponential decay to zero within 200 to 300 μsec. The signal frequency is nominally defined as 100 kHz; in actuality, the signal is designed such that 99% of the radiated power is contained in a 20 kHz band centered on 100 kHz.

The Loran receiver is programmed to track the signal on the cycle corresponding to the carrier frequency's third positive crossing of the x-axis. This occurrence, termed the **standard zero crossing**, is chosen for two reasons. First, it is late enough for the pulse to have built up sufficient signal strength for the receiver to detect it. Second, it is early enough in the pulse to ensure that the receiver is detecting the transmitting station's ground wave pulse and not its sky wave pulse. Sky wave pulses are affected by atmospheric refraction and if used unknowingly, would introduce large errors into positions determined by a Loran receiver. The pulse architecture described here reduces this major source of error.

Another important parameter of the pulse is the **envelope-to-cycle difference (ECD)**. This parameter indicates how propagation of the signal causes the pulse shape envelope (i.e., the imaginary line connecting the peak of each sinusoidal cycle) to shift in time relative to the zero crossings. The ECD is important because receivers use the precisely shaped pulse envelope to identify the correct zero crossing. Transmitting stations are required to keep the ECD within defined limits. Many receivers display the received ECD as well.

Next, individual pulses are combined into sequences. For the master signal, a series of nine pulses is transmitted, the first eight spaced 1000 μsec apart followed by a ninth transmitted 2000 μsec after the eighth. Secondary stations transmit a series of eight pulses, each spaced 1000 μsec apart. Secondary stations are given letter designations of U, W, X, Y, and Z; this letter designation indicates the order in which they transmit following the master. If a chain has two secondaries, they will be designated Y and Z. If a chain has three secondaries, they are X, Y and Z, and so on. Some exceptions to this general naming pattern exist (e.g., W, X and Y for some 3-secondary chains).

The spacing between the master signal and each of the secondary signals is governed by several parameters as illustrated in Figure 1203b. The general idea is that each of the signals must clear the entire chain coverage area before the next one is transmitted, so that no signal can be received out of order. The time required for the master signal to travel to the secondary station is defined as the average **baseline travel time (BTT),** or **baseline length (BLL)**. To this time interval is added an additional delay defined as the **secondary coding delay (SCD),** or simply **coding delay (CD).** The total of these two delays is termed the **emission delay**

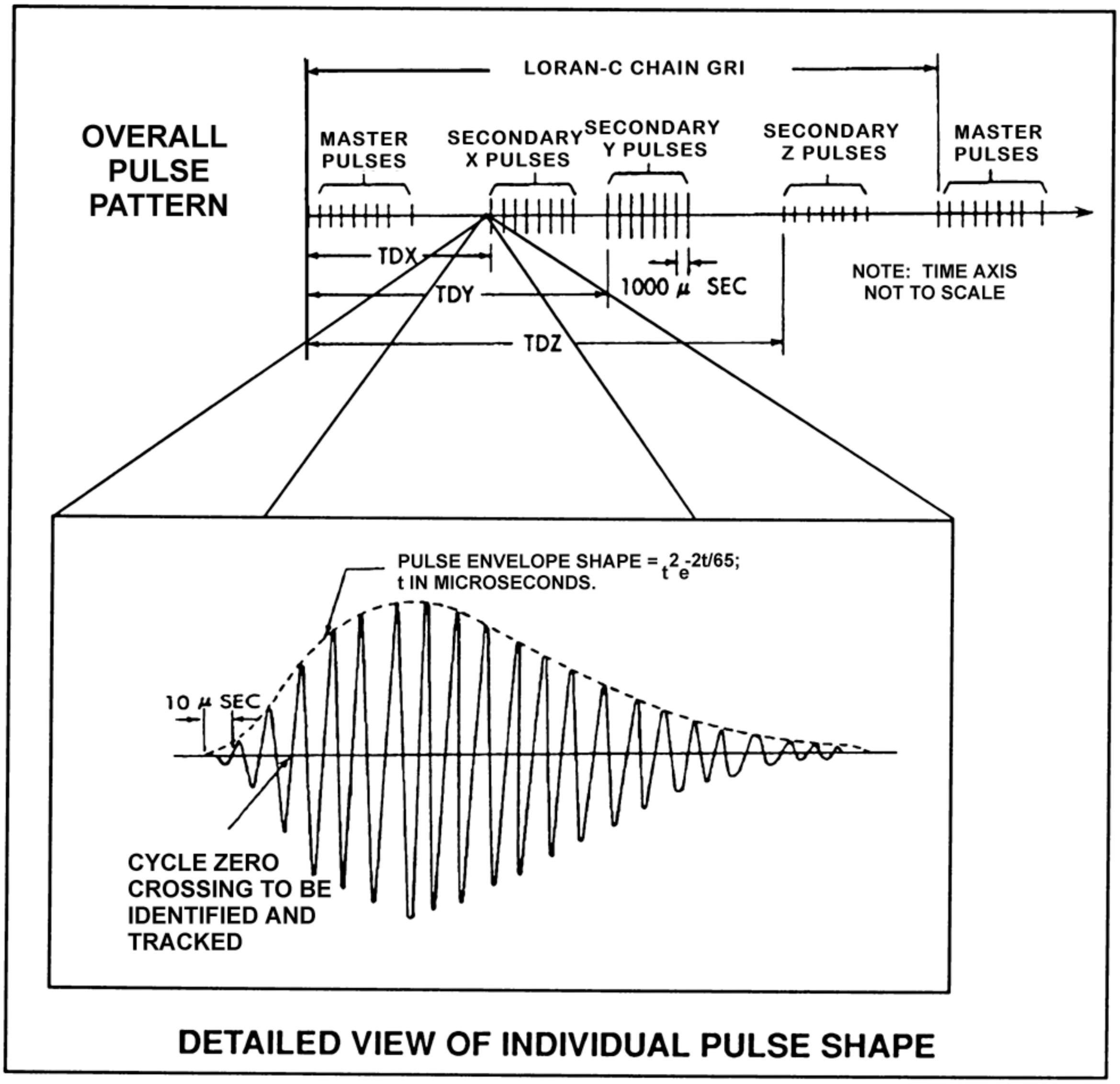


Figure 1203a. Pulse pattern and shape for Loran C transmission.

(ED), which is the exact time interval between the transmission of the master signal and the transmission of the secondary signal. Each secondary station has its own ED value. In order to ensure the proper sequence, the ED of secondary Y is longer than that of X, and the ED of Z is longer than that of Y.

Once the last secondary has transmitted, the master transmits again, and the cycle is repeated. The time to complete this cycle of transmission defines an important characteristic for the chain: the **group repetition interval (GRI)**. The group repetition interval divided by ten yields the chain's numeric designator. For example, the interval between successive transmissions of the master pulse group for the northeast U.S. chain is 99,600 μsec, just less than one tenth of a second. From the definition above, the GRI designator for this chain is defined as 9960. As mentioned previously, the GRI must be sufficiently large to allow the signals from the master and secondary stations in the chain to propagate fully throughout the region covered by the chain before the next cycle of pulses begins.

Two additional characteristics of the pulse group are **phase coding** and **blink coding**. In phase coding, the phase of the 100 kHz carrier signal is reversed from pulse to pulse in a preset pattern that repeats every two GRI's. Phase coding allows a receiver to remove skywave contamination from the groundwave signal. Loran C signals travel away

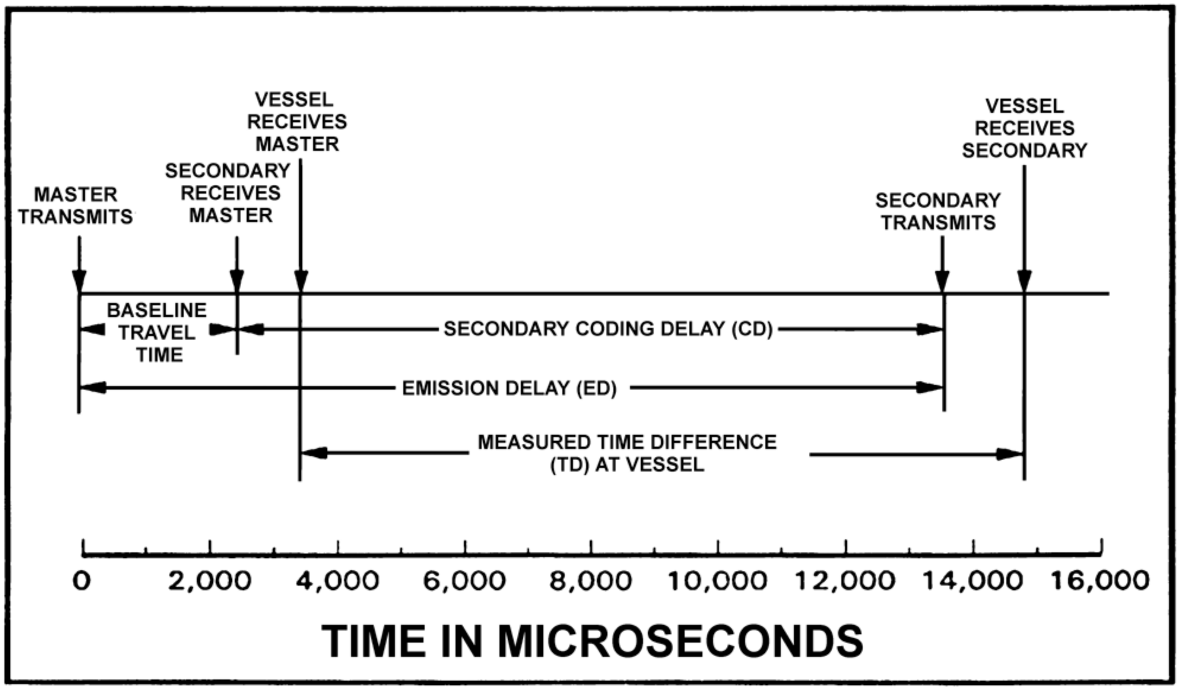


Figure 1203b. The time axis for Loran TD for point "A."

from a transmitting station in all possible directions. Groundwave is the Loran energy that travels along the surface of the earth. Skywave is Loran energy that travels up into the sky. The ionosphere reflects some of these skywaves back to the earth's surface. The skywave always arrives later than the groundwave because it travels a greater distance. The skywave of one pulse can thus contaminate the ground wave of the next pulse in the pulse group. Phase coding ensures that this skywave contamination will always "cancel out" when all the pulses of two consecutive GRI's are averaged together.

Blink coding provides integrity to the received Loran signal. When a signal from a secondary station is out of tolerance and therefore temporarily unsuitable for navigation, the affected secondary station will blink; that is, the first two pulses of the affected secondary station are turned off and on in a repeating cycle, 3.6 seconds off and 0.4 seconds on. The receiver detects this condition and displays it to the operator. When the blink indication is received, the operator should not use the affected secondary station. If a station's signal will be temporarily shut down for maintenance, the Coast Guard communicates this information in a *Notice to Mariners*. The U.S. Coast Guard Navigation Center posts these online at http://www.navcen.uscg.gov/. If a master station is out of tolerance, all secondaries in the affected chain will blink.

Two other concepts important to the understanding of Loran operation are the **baseline** and **baseline extension**. The geographic line connecting a master to a particular secondary station is defined as the station pair baseline. The baseline is, in other words, that part of a great circle on which lie all the points connecting the two stations. The extension of this line beyond the stations to encompass the points along this great circle not lying between the two stations defines the baseline extension. The optimal region for hyperbolic navigation occurs in the vicinity of the baseline, while the most care must be exercised in the regions near the baseline extension. These concepts are further developed in the next few articles.

1204. Loran Theory of Operation

In Loran navigation, the locus of points having a constant difference in distance between an observer and each of two transmitter stations defines a hyperbola, which is a line of position.

Assuming a constant speed of propagation of electromagnetic radiation in the atmosphere, the time difference in the arrival of electromagnetic radiation from the two transmitter sites is proportional to the distance between each of the transmitting sites, thus creating the hyperbola on the earth's surface. The following equations demonstrate this proportionality between distance and time:

Distance=Velocity x Time

or, using algebraic symbols

$d = c \times t$

Therefore, if the velocity (c) is constant, the distance between a vessel and each of two transmitting stations will be directly proportional to the time delay detected at the vessel between pulses of electromagnetic radiation transmitted from the two stations.

An example illustrates the concept. As shown in Figure 1204, let us assume that two Loran transmitting stations, a master and a secondary, are located along with an observer in a Cartesian coordinate system whose units are in nautical miles. We assume further that the master station, designated "M", is located at coordinates (x,y) = (-200,0) and the secondary, designated "X," is located at (x,y) = (+200,0). An observer with a receiver capable of detecting electromagnetic radiation is positioned at any point "A" whose coordinates are defined as (x_a, y_a).

Note that for mathematical convenience, these hyperbola labels have been normalized so that the hyperbola perpendicular to the baseline is labeled zero, with both negative and positive difference values. In actual practice, all Loran TD's are positive.

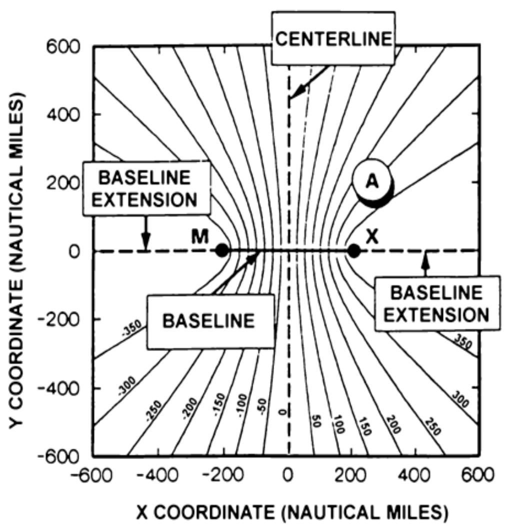


Figure 1204. Depiction of Loran LOP's.

The Pythagorean theorem can be used to determine the distance between the observer and the master station; similarly, one can obtain the distance between the observer and the secondary station:

$$\text{distance}_{am} = [(x_a + 200)^2 + y_a^2]^{0.5}$$

$$\text{distance}_{ax} = [(x_a - 200)^2 + y_a^2]^{0.5}$$

The difference between these distances (D) is:

$$D = \text{distance}_{am} - \text{distance}_{ax}$$

Substituting,

$$D = [(x_a + 200)^2 + y_a^2]^{0.5} - [(x_a - 200)^2 + y_a^2]^{0.5}$$

With the master and secondary stations in known geographic positions, the only unknowns are the two geographic coordinates of the observer.

Each hyperbolic line of position in Figure 1204 represents the locus of points for which (D) is held constant. For example, if the observer above were located at point A (271.9, 200) then the distance between that observer and the secondary station (the point designated "X" in Figure 1204a) would be 212.5 NM. In turn, the observer's distance from the master station would be 512.5 NM. The function D would simply be the difference of the two, or 300 NM. For every other point along the hyperbola passing through A, distance D has a value of 300 NM. Adjacent LOP's indicate where D is 250 NM or 350 NM.

To produce a fix, the observer must obtain a similar hyperbolic line of position generated by another master-secondary pair. Let us say another secondary station "Y" is placed at point (50,500). Mathematically, the observer will then have two equations corresponding to the M-X and M-Y TD pairs:

$$D_1 = [(x_a + 200)^2 + y_a^2]^{0.5} - [(x_a - 200)^2 + y_a^2]^{0.5}$$

$$D_2 = [(x_a + 200)^2 + y_a^2]^{0.5} - [(x_a - 50)^2 + (y_a - 500)^2]^{0.5}$$

Distances D_1 and D_2 are known because the time differences have been measured by the receiver and converted to these distances. The two remaining unknowns, x_a and y_a, may then be solved.

The above example is expressed in terms of distance in nautical miles. Because the navigator uses TD's to perform Loran hyperbolic navigation, let us rework the example for the M-X TD pair in terms of time rather than distance, adding timing details specific to Loran. Let us assume that electromagnetic radiation travels at the speed of light (one nautical mile traveled in 6.18 µsec). The distance from master station M to point A was 512.5 NM. From the relationship just defined between distance and time, it would take a signal (6.18 µsec/NM) × 512.5 NM = 3,167 µsec to travel from the master station to the observer at point A. At the arrival of this signal, the observer's Loran receiver would start the TD measurement. Recall from the general discussion above that a secondary station transmits after an emission delay equal to the sum of the baseline travel time and the secondary coding delay. In this example, the master and the secondary are 400 NM apart; therefore, the baseline travel time is (6.18 µsec/NM) × 400 NM = 2,472 µsec. Assuming a secondary coding delay of 11,000 µsec, the secondary station in this example would transmit (2,472 + 11,000) µsec or 13,472 µsec after the master station. The secondary signal then propagates over a distance 212.5 NM to reach point A, taking (6.18 µsec/NM) × 212.5 NM = 1,313 µsec to do so. Therefore, the total time from *transmission* of the master signal to the *reception* of the secondary signal by the observer at point A is (13,472 + 1,313) µsec = 14,785 µsec.

Recall, however, that the Loran receiver measures the time delay between *reception* of the master signal and the *reception* of the secondary signal. Therefore, the time quantity above must be corrected by subtracting the amount of time required for the signal to travel from the master transmitter to the observer at point A. This amount of time was 3,167 µsec. Therefore, the TD observed at point A in this hypothetical example would be (14,785 - 3,167) µsec or 11,618 µsec. Once again, this time delay is a function of the simultaneous differences in distance between the observer and the two transmitting stations, and it gives rise to a hyperbolic line of position which can be crossed with another LOP to fix the observer's position.

1205. Allowances for Non-Uniform Propagation Rates

The initial calculations above assumed the speed of light in free space; however, the actual speed at which electromagnetic radiation propagates on earth is reduced both by the atmosphere through which it travels and by the conductive surfaces—sea and land—over which it passes. The specified accuracy needed from Loran therefore requires three corrections to the propagation speed of the signal.

The reduction in propagation speed due to the atmosphere is represented by the first correction term: the **Primary Phase Factor (PF)**. Similarly, a **Secondary Phase Factor (SF)** accounts for the reduced propagation speed due to traveling over seawater. These two corrections are transparent to the operator since they are uniformly incorporated into all calculations represented on charts and in Loran receivers.

Because land surfaces have lower conductivity than seawater, the propagation speed of the Loran signal passing over land is further reduced as compared to the signal passing over seawater. A third and final correction, the **Additional Secondary Phase Factor (ASF)**, accounts for the delay due to the land conductivity when converting time delays to distances and then to geographic coordinates. Depending on the mariner's location, signals from some Loran transmitters may have traveled hundreds of miles over land and must be corrected to account for this non-seawater portion of the signal path. Of the three corrections mentioned in this article, this is the most complex and the most important one to understand, and is accordingly treated in detail in Article 1210.

LORAN ACCURACY

1206. Defining Accuracy

Specifications of Loran and other radionavigation systems typically refer to three types of accuracy: **absolute, repeatable** and **relative.**

Absolute accuracy, also termed predictable or geodetic accuracy, is the accuracy of a position with respect to the geographic coordinates of the earth. For example, if the navigator plots a position based on the Loran latitude and longitude (or based on Loran TD's) the difference between the Loran position and the actual position is a measure of the system's absolute accuracy.

Repeatable accuracy is the accuracy with which the navigator can return to a position whose coordinates have been measured previously with the same navigational system. For example, suppose a navigator were to travel to a buoy and note the TD's at that position. Later, suppose the navigator, wanting to return to the buoy, returns to the previously measured TD's. The resulting position difference between the vessel and the buoy is a measure of the system's repeatable accuracy.

Relative accuracy is the accuracy with which a user can measure position relative to that of another user of the same navigation system at the same time. If one vessel were to travel to the TD's determined by another vessel, the difference in position between the two vessels would be a measure of the system's relative accuracy.

The distinction between absolute and repeatable accuracy is the most important one to understand. With the correct application of ASF's and within the **coverage area** defined for each chain, the absolute accuracy of the Loran system varies from between 0.1 and 0.25 nautical miles. However, the repeatable accuracy of the system is much better, typically between 18 and 90 meters (approximately 60 to 300 feet) depending on one's location in the coverage area. If the navigator has been to an area previously and noted the TD's corresponding to different navigational aids (e.g., a buoy marking a harbor entrance), the high repeatable accuracy of the system enables location of the buoy in adverse weather. Similarly, selected TD data for various harbor navigational aids and other locations of interest have been collected and recorded and is generally commercially available. This information provides an excellent backup navigational source to conventional harbor approach navigation.

1207. Limitations to Loran Accuracy

There are limits on the accuracy of any navigational system, and Loran is no exception. Several factors that contribute to limiting the accuracy of Loran as a navigational aid are listed in Table 1207 and are briefly discussed in this article. Even though all these factors except operator error are included in the published accuracy of Loran, the mariner's aim should be to have a working knowledge of each one and minimize any that are under his control so as to obtain the best possible accuracy.

The geometry of LOP's used in a Loran fix is of prime importance to the mariner. Because understanding of this factor is so critical to proper Loran operation, the effects of crossing angles and gradients are discussed in detail in the Article 1208. The remaining factors are briefly explained as follows.

The age of the Coast Guard's Loran transmitting equipment varies from station to station. When some older types of equipment are switched from standby to active and vice versa, a slight timing shift as large as tens of nanoseconds may be seen. This is so small that it is undetectable by most marine receivers, but since all errors accumulate, it should be understood as part of the Loran "error budget."

The effects of actions to control chain timing are similar. The timing of each station in a chain is controlled based on data received at the primary system area monitor site. Signal timing errors are kept as near to zero as possible at the primary site, making the absolute accuracy of Loran generally the best in the vicinity of the primary site. Whenever, due to equipment casualty or to accomplish system maintenance, the control station shifts to the secondary system area monitor site, slight timing shifts may be introduced in parts of the coverage area.

Atmospheric noise, generally caused by lightning, reduces the **signal-to-noise ratio (SNR)** available at the receiver. This in turn degrades accuracy of the LOP. Man-made noise has a similar effect on accuracy. In rare cases, a man-made noise source whose carrier signal frequency or harmonics are near 100 kHz (such as the constant carrier control signals commonly used on high-tension power lines) may also interfere with lock-on and tracking of a Loran receiver. In general, Loran stations that are the closest to the user will have the highest SNR and will produce LOP's with the lowest errors. Geometry, however, remains a key factor in producing a good fix from combined LOP's. Therefore, the best LOP's for a fix may not all be from the very nearest stations.

The user should also be aware that the propagation speed of Loran changes with time as well. Temporal changes may be seasonal, due to snow cover or changing groundwater levels, or diurnal, due to atmospheric and surface changes from day to night. Seasonal changes may be as large as 1 µsec and diurnal changes as large as 0.2 µsec, but these vary with location and chain being used. Passing cold weather fronts may have temporary effects as well.

Disturbances on the sun's surface, most notably solar flares, disturb the earth's atmosphere as well. These Sudden Ionospheric Disturbances (SID's) increase attenuation of radio waves and thus disturb Loran signals and reduce SNR. Such a disturbance may interfere with Loran reception for periods of hours or even longer.

The factors above all relate to the propagated signal before it reaches the mariner. The remaining factors discussed below address the accuracy with which the mariner receives and interprets the signal.

Factor	Has effect on	
	Absolute Accuracy	**Repeatable Accuracy**
Crossing angles and gradients of the Loran LOP's	Yes	Yes
Stability of the transmitted signal (e.g., transmitter effect)	Yes	Yes
Loran chain control parameters (e.g., how closely actual ED is maintained to published ED, which system area monitor is being used, etc.)	Yes	Yes
Atmospheric and man-made ambient electronic noise	Yes	Yes
Factors with temporal variations in signal propagation speed (e.g., weather, seasonal effects, diurnal variations, etc.)	Yes	Yes
Sudden ionospheric disturbances	Yes	Yes
Receiver quality and sensitivity	Yes	Yes
Shipboard electric noise	Yes	Yes
Accuracy with which LOP's are printed on nautical charts	Yes	No
Accuracy of receiver's computer algorithms for coordinate conversion	Yes	No
Operator error	Yes	Yes

Table 1207. Selected Factors that Limit Loran Accuracy.

Receivers vary in precision, quality and sophistication. Some receivers display TD's to the nearest 0.1 μsec; others to 0.01 μsec. Internal processing also varies, whether in the analog "front end" or the digital computer algorithms that use the processed analog signal. By referencing the user manual, the mariner may gain an appreciation for the advantages and limitations of the particular model available, and may adjust operator settings to maximize performance.

The best receiver available may be hindered by a poor installation. Similarly, electronic noise produced by engine and drive machinery, various electric motors, other electronic equipment or even household appliances may hinder the performance of a Loran receiver. The mariner should consult documentation supplied with the receiver for proper installation. Generally, proper installation and placement of the receiver's components will mitigate these problems. In some cases, contacting the manufacturer or obtaining professional installation assistance may be appropriate.

The raw TD's obtained by the receiver must be corrected with ASF's and then translated to position. Whether the receiver performs this entire process or the mariner assists by translating TD's to position manually using a Loran overprinted chart, published accuracies take into account the small errors involved in this conversion process.

Finally, as in all endeavors, operator error when using Loran is always possible. This can be minimized with alertness, knowledge and practice.

1208. The Effects of Crossing Angles and Gradients

The hyperbolic nature of Loran requires the operator to pay special attention to the geometry of the fix, specifically to crossing angles and gradients, and to the possibility of fix ambiguity. We begin with crossing angles.

As discussed above, the TD's from any given master-secondary pair form a family of hyperbolas. Each hyperbola in this family can be considered a line of position; the vessel must be somewhere along that locus of points which forms the hyperbola. A typical family of hyperbolas is shown in Figure 1208a.

Now, suppose the hyperbolic family from the Master-Xray station pair shown in Figure 1204 were superimposed upon the family shown in Figure 1208a. The results would be the hyperbolic lattice shown in Figure 1208b.

As has been noted, Loran LOP's for various chains and secondaries are printed on nautical charts. Each of the sets of LOP's is given a separate color and is denoted by a characteristic set of symbols. For example, an LOP might be designated 9960-X-25750. The designation is read as follows: the chain GRI designator is 9960, the TD is for the Master-Xray pair (M-X), and the time difference along this LOP is 25750 μsec. The chart shows only a limited number of LOP's to reduce clutter on the chart. Therefore, if the observed time delay falls between two charted LOP's, interpolation between them is required to obtain the precise LOP. After having interpolated (if necessary) between two

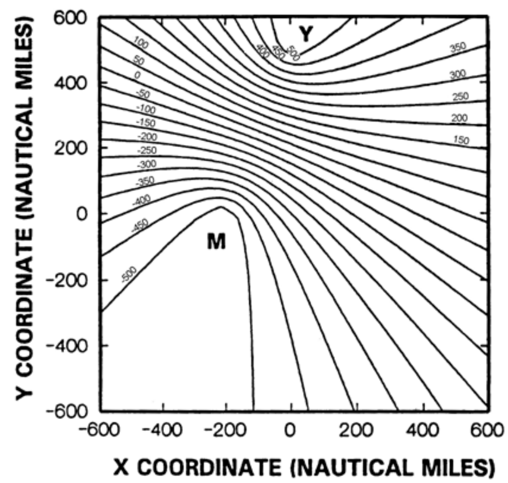


Figure 1208a. A family of hyperbolic lines generated by Loran signals.

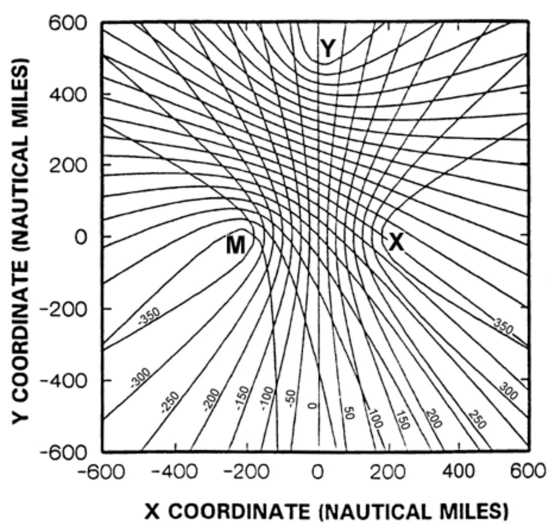


Figure 1208b. A hyperbolic lattice formed by station pairs M-X and M-Y.

TD measurements and plotted the resulting LOP's on the chart, the navigator marks the intersection of the LOP's and labels that intersection as the Loran fix. Note also in Figure 1208b the various angles at which the hyperbolas cross each other.

Figure 1208c shows graphically how error magnitude varies as a function of crossing angle. Assume that LOP 1

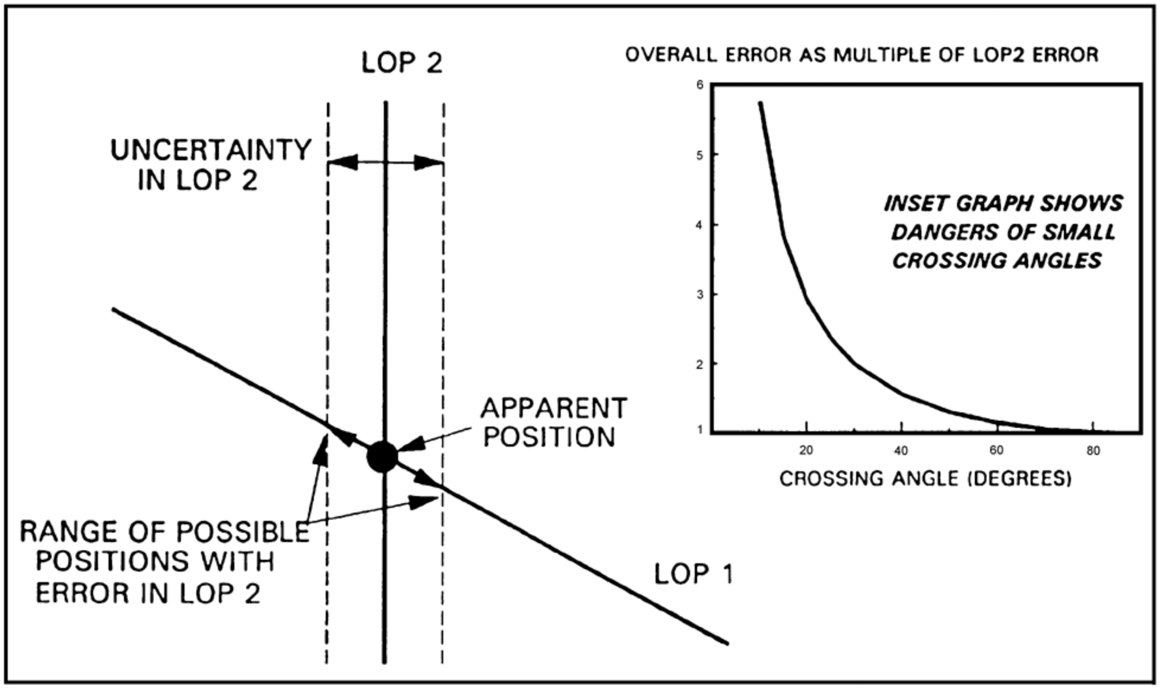


Figure 1208c. Error in Loran LOP's is magnified if the crossing angle is less than 90°.

is known to contain no error, while LOP 2 has an uncertainty as shown. As the crossing angle (i.e., the angle of intersection of the two LOP's) approaches 90°, range of possible positions along LOP 1 (i.e., the position uncertainty or fix error) approaches a minimum; conversely, as the crossing angle decreases, the position uncertainty increases; the line defining the range of uncertainty grows longer. This illustration demonstrates the desirability of choosing LOP's for which the crossing angle is as close to 90° as possible.

The relationship between crossing angle and fix uncertainty can be expressed mathematically:

$$\sin(x) = \frac{\text{LOP error}}{\text{fix uncertainty}}$$

where x is the crossing angle.

Rearranging algebraically,

$$\text{fix uncertainty} = \frac{\text{LOP error}}{\sin(x)}$$

Assuming that LOP error is constant, then position uncertainty is inversely proportional to the sine of the crossing angle. As the crossing angle increases from 0° to 90°, the sine of the crossing angle increases from 0 to 1. Therefore, the error is at a minimum when the crossing angle is 90°, and increases thereafter as the crossing angle decreases.

Understanding and proper use of TD gradients are also important to the navigator. The gradient is defined as the rate of change of distance with respect to TD. Put another way, this quantity is the ratio of the spacing between adjacent Loran TD's (usually expressed in feet or meters) and the difference in microseconds between these adjacent LOP's. For example, if at a particular location two printed TD lines differ by 20 µsec and are 6 NM apart, the gradient is.

$$\text{Gradient} = \frac{6\text{NM} \times 6076\text{ft/NM}}{20\mu\text{sec}} = 1822.8 \text{ ft/}\mu\text{sec}$$

The smaller the gradient, the smaller the distance error that results from any TD error. Thus, the best accuracy from Loran is obtained by using TD's whose gradient is the smallest possible (i.e. the hyperbolic lines are closest together). This occurs along the baseline. Gradients are much larger (i.e. hyperbolic lines are farther apart) in the vicinity of the baseline extension. Therefore, the user should select TD's having the smallest possible gradients.

Another Loran effect that can lead to navigational error in the vicinity of the baseline extension is fix ambiguity. Fix ambiguity results when one Loran LOP crosses another LOP in two separate places. Near the baseline extension, the "ends" of a hyperbola can wrap around so that they cross another LOP twice, once along the baseline, and again along the baseline extension. A third LOP would resolve the ambiguity.

Most Loran receivers have an ambiguity alarm to alert the navigator to this occurrence. However, both fix ambiguity and large gradients necessitate that the navigator avoid using a master-secondary pair when operating in the vicinity of that pair's baseline extension.

1209. Coverage Areas

The 0.25 NM absolute accuracy specified for Loran is valid within each chain's coverage area. This area, whose limits define the maximum range of Loran for a particular chain, is the region in which both accuracy and SNR criteria are met. The National Oceanographic and Atmospheric Administration (NOAA) has generally followed these coverage area limits when selecting where to print particular Loran TD lines on Loran overprinted charts. Coverage area diagrams of each chain are also available online from the U.S. Coast Guard's Navigation Center, currently at http://www.navcen.uscg.gov/ftp/loran/lgeninfo/h-book/loranappendixb.pdf. Other helpful information available at this FTP site includes the Loran C User's Handbook and the Loran C Signal Specification, two key sources of material in this chapter.

One caveat to remember when considering coverage areas is that the 0.25 NM accuracy criteria is modified inside the coverage area in the vicinity of the coastline due to ASF effects. The following article describes this more fully.

1210. Understanding Additional Secondary Factors (ASF's)

Mathematically, calculating the reduction in propagation speed of an electromagnetic signal passing over a land surface of known conductivity is relatively straightforward. In practice, however, determining this Loran ASF correction accurately for use in the real world can be complex.

There are at least four reasons for this complexity. First, the conductivity of ground varies from region to region, so the correction to be applied is different for every signal path. Moreover, ground conductivity data currently available do not take into account all the minor variations within each region. Second, methods used to compute ASF's vary. ASF's can be determined from either a mathematical model based on known approximate ground conductivities, or from empirical time delay measurements in various locations, or a combination of both. Methods incorporating empirical measurements tend to yield more accurate results. One receiver manufacturer may not use exactly the same correction method as another, and neither may use exactly the same method as those incorporated into time differences printed on a particular nautical chart. While such differences are minor, a user unaware of these differences may not obtain the best accuracy possible from Loran. Third, relatively large local variations in ASF variations that cannot fully be accounted for in current ASF models applied to the coverage area as a whole, may be observed in the region within 10 NM of the coast. Over the years, even empirically measured ASF's may change slightly in these areas with the addition of buildings, bridges and other structures to coastal areas. Fourth and finally, ASF's vary seasonally with changes in groundwater levels, snow pack depths and similar factors.

Designers of the Loran system, including Loran receiver manufacturers, have expended a great deal of effort to include ASF's in error calculations and to minimize these effects. Indeed, inaccuracies in ASF modeling are accounted for in published accuracy specifications for Loran. What then does the marine navigator need to know about ASF's beyond this? To obtain the 0.25 NM absolute accuracy advertised for Loran, the answer is clear. One must know *where* in the coverage area ASF's affect published accuracies, and one must know *when* ASF's are being incorporated, both in the receiver and on any chart in use.

With respect to *where* ASF's affect published accuracies, one must remember that local variations in the vicinity of the coastline are the most unpredictable of all ASF related effects because they are not adequately explained by current predictive ASF models. As a result, even though fixes determined by Loran may satisfy the 0.25 NM accuracy specification in these areas, such accuracy is not "guaranteed" for Loran within 10 NM of the coast. Users should also avoid relying solely on the lattice of Loran TD's in inshore areas.

With respect to *when* ASF's are being applied, one should realize that the default mode in most receivers combines ASF's with raw TD measurements. This is because the inclusion of ASF's is required in order to meet the 0.25 NM accuracy criteria. The navigator should verify which mode the receiver is in, and ensure the mode is not changed unknowingly. Similarly, current NOAA Loran overprinted charts of the U.S. incorporate ASF's, and in the chart's margin the following note appears:

> "Loran C correction tables published by the National Imagery and Mapping Agency or others should not be used with this chart. The lines of position shown have been adjusted based on survey data. Every effort has been made to meet the 0.25 nautical mile accuracy criteria established by the U.S. Coast Guard. Mariners are cautioned not to rely solely on the lattices in inshore waters."

The key point to remember there is that the "ASF included" and "ASF not included" modes must not be mixed. In other words, the receiver and any chart in use must handle ASF's in the same manner. If the receiver includes them, any chart in use must also include them. If operating on a chart that does not include ASF's—Loran coverage areas in another part of the world, for example—the receiver must be set to the same mode. If the navigator desires to correct

ASF's manually, tables for U.S. Loran chains are available at http://chartmaker.ncd.noaa.gov/mcd/loranc.htm. These documents also provide a fuller explanation of manual ASF corrections. When viewing ASF tables, remember that although the ASF correction for a single signal is always positive (indicating that the signal is always slowed and never speeded by its passage over land), the ASF correction for a time *difference* may be negative because two signal delays are included in the computation.

The U.S. Government does not guarantee the accuracy of ASF corrections incorporated into Loran receivers by their respective manufacturers. The prudent navigator will regularly check Loran TD's against charted LOP's when in a known position, and will compare Loran latitude and longitude readouts against other sources of position information. Ensuring the proper configuration and operation of the Loran receiver remains the navigator's responsibility.

Up to this point, our discussion has largely focused on correctly understanding and using Loran in order to obtain published accuracies. In some portions of the coverage areas, accuracy levels actually obtainable may be significantly better than these minimum published values. The following articles discuss practical techniques for maximizing the absolute, repeatable and relative accuracy of Loran.

1211. Maximizing Loran's Absolute Accuracy

Obtaining the best possible absolute accuracy from Loran rests primarily on the navigator's selection of TD's, particularly taking into account geometry, SNR and proximity to the baseline and baseline extension. As a vessel transits the coverage area, these factors gradually change and, except for SNR, are not visible on the display panel of the Loran receiver. Most receivers track an entire chain and some track multiple chains simultaneously, but the majority of installed marine receivers still use only two TD's to produce a latitude and longitude. Some receivers monitor these factors and may automatically select the best pair. The best way for the navigator, however, to monitor these factors is by referring to a Loran overprinted chart, even if not actually plotting fixes on it. The alert navigator will frequently reevaluate the selection of TD's during a transit and make adjustments as necessary.

Beyond this advice, two additional considerations may help the navigator maximize absolute accuracy. The first is the realization that Loran TD error is not evenly distributed over the coverage area. Besides the effects of transmitter station location on geometry and fix error, the locations of the primary and secondary monitor sites also have a discernible effect on TD error in the coverage area. As ASF's change daily and seasonally, the Loran control stations continually adjust the emission delay of each secondary station to keep it statistically at its nominal value as observed at the primary monitor site. What this means is that, on average, the Loran TD is more stable and more accurate in the absolute sense in the vicinity of the primary monitor site. The primary system area monitor for stations 9960-M, 9960-X and 9960-Y was placed at the entrance to New York harbor at Sandy Hook, New Jersey for just this reason. A switch by the control station to the secondary monitor site will shift the error distribution slightly within the coverage area, reducing it near the secondary site and slightly increasing it elsewhere. The locations of primary system area monitor sites can be found at the USCG NAVCEN web site.

The second consideration in maximizing absolute accuracy is that most Loran receivers may be manually calibrated using a feature variously called "bias," "offset," "homeport" or a similar term. When in homeport or another known location, the known latitude and longitude (or in some cases, the difference between the current Loran display and the known values) is entered into the receiver. This forces the receiver's position error to be zero at that particular point and time.

The limitation of this technique is that this correction becomes less accurate with the passage of time and with increasing distance away from the point used. Most published sources indicate the technique to be of value out to a distance of 10 to 100 miles of the point where the calibration was performed. This correction does not take into account local distortions of the Loran grid due to bridges, power lines or other such man-made structures. The navigator should evaluate experimentally the effectiveness of this technique in good weather conditions before relying on it for navigation at other times. The bias should also be adjusted regularly to account for seasonal Loran variations; using the same value throughout the year is not the most effective application of this technique. Also, entering an offset into a Loran receiver alters the apparent location of waypoints stored prior to establishing this correction.

Finally, receivers vary in how this feature is implemented. Some receivers save the offset when the receiver is turned off; others zero the correction when the receiver is turned on. Some receivers replace the internal ASF value with the offset, while others add it to the internal ASF values. Refer to the owner's manual for the receiver in use.

1212. Maximizing Loran's Repeatable Accuracy

Many users consider the high repeatable accuracy of Loran its most important characteristic. To obtain the best repeatable accuracy consistently, the navigator should use measured TD's rather than latitude and longitude values supplied by the receiver.

The reason for this lies in the ASF conversion process. Recall that Loran receivers use ASF's to correct TD's. Recall also that the ASF's are a function of the terrain over which the signal must pass to reach the receiver. Therefore, the ASF's for one station pair are different from the ASF's for another station pair because the signals from the different pairs must travel over different terrain to reach the receiver.

This consideration matters because a Loran receiver may not always use the same pairs of TD's to calculate a fix. Suppose a navigator marks the position of a channel buoy by recording its latitude and longitude using the TD pair selected automatically by the Loran receiver. If, on the return trip, the receiver is using a different TD pair, the latitude and longitude readings for the exact same buoy would be slightly different because the new TD pair would be using a different ASF value. By using previously-measured TD's and not previously-measured latitudes and longitudes, this ASF-introduced error is avoided. The navigator should also record the values of all secondary TD's at the waypoint and not just the ones used by the receiver at the time. When returning to the waypoint, other TD's will be available even if the previously used TD pair is not. Recording the time and date the waypoint is stored will also help evaluate the cyclical seasonal and diurnal variations that may have since occurred.

1213. Maximizing Loran's Relative Accuracy

The classical application of relative accuracy involves two users finding the same point on the earth's surface at the same time using the same navigation system. The maximum relative Loran accuracy would be theoretically be achieved by identical receivers, configured and installed identically on identical vessels, tracking the same TD's. In practice, the two most important factors are tracking the same TD's and ensuring that ASF's are being treated consistently between the two receivers. By attending to these, the navigator should obtain relative accuracy close to the theoretical maximum.

Another application of relative accuracy is the current practice of converting old Loran TD's into latitude and longitude for use with GPS and DGPS receivers. Several commercial firms sell software applications that perform this tedious task. One key question posed by these programs is whether or not the Loran TD's include ASF's. The difficulty in answering this question depends on how the Loran TD's were obtained, and of course an understanding of ASF's. If in doubt, the navigator can perform the conversion once by specifying "with" ASF's and once "without," and then carefully choosing which is the valid one, assisted by direct observation underway if needed.

To round out the discussion of Loran, the following article briefly describes present and possible future uses for this system beyond the well-known hyperbolic navigation mode.

NON-HYPERBOLIC USES OF LORAN C

1214. Precise Timing with Loran

Because Loran is fundamentally a precise timing system, a significant segment of the user community uses Loran for the propagation of Coordinated Universal Time (UTC). The accessibility of UTC at any desired location enables such applications as the synchronization of telephone and data networks. The U.S. Coast Guard makes every effort to ensure that each Loran master transmitter station emits its signal within 100 ns of UTC. Because the timing of each secondary station is relative to the master, its timing accuracy derives from that of the master.

The start of each Loran station's GRI periodically coincides with the start of the UTC second. This is termed the Time of Coincidence (TOC). The U.S. Naval Observatory publishes TOC's at http://tycho.usno.navy.mil/loran.html for the benefit of timing users. Because one Loran station is sufficient to provide an absolute timing reference, timing receivers do not typically rely on the hyperbolic mode or use TD's per se.

A noteworthy feature of Loran is that each transmitter station has an independent timing reference consisting of three modern cesium beam oscillators. Timing equipment at the transmitter stations constantly compares these signals and adjusts to minimize oscillator drift. The end result is a nationwide system with a large ensemble of independent timing sources. This strengthens the U.S. technology infrastructure. As another cross-check of Loran time, daily comparisons are made with UTC, as disseminated via GPS.

1215. Loran Time of Arrival (TOA) Mode

With the advent of the powerful digital processors and compact precise oscillators now embedded in user receivers, technical limitations that dictated Loran's hyperbolic architecture decades ago have been overcome. A receiver can now predict in real time the exact point in time a Loran station will transmit its signal, as well as the exact time the signal will be received at any assumed position.

An alternate receiver architecture that takes advantage of these capabilities uses Loran Time of Arrival (TOA) measurement, which are measured relative to UTC rather than to an arbitrary master station's transmission. A receiver operating in TOA mode can locate and track all Loran signals in view, prompting the descriptor "all in view" for this type of receiver. This architecture steps beyond the limitations of using only one Loran chain at a time. As a result, system availability can be improved across all the overlapping coverage areas. Coupled with advanced Receiver Autonomous Integrity Monitor (RAIM) algorithms, this architecture can also add an additional layer of integrity at the user level, independent of Loran blink.

One technical possibility arising out of this new capability is to control the time of transmission of each station independently with direct reference to UTC, rather than by using system area monitors. Such an arrangement could offer the advantage of more uniformly distributing Loran fix errors across the coverage areas. This could in turn more naturally configure Loran for use in an integrated navigation system.

1216. Loran in an Integrated Navigation System

An exponential worldwide increase in reliance on electronic navigation systems, most notably GPS, for positioning and timing has fueled a drive for more robust systems immune from accidental or intentional interference. Even a short outage of GPS, for example, would likely have severe safety and economic consequences for the United States and other nations.

In this environment, integrated navigation systems are attractive options as robust sources of position and time. The ideal integrated navigation system can tolerate the degradation or failure of any component system without degradation as a whole.

Loran offers several advantages to an integrated system based on GPS or DGPS. Although Loran relies on radio propagation and is thus similarly vulnerable to large-scale atmospheric events such as ionospheric disturbances, at 100 kHz it occupies a very different portion of the spectrum than the 1.2 GHz to 1.6 GHz band used by GPS. Loran is a high-power system whose low frequency requires a very large antenna for efficient propagation. Therefore, jamming Loran over a broad area is much more difficult than jamming GPS over the same area. Loran signals are present in urban and natural canyons and under foliage, where GPS signals may be partially or completely blocked. Loran's independent timing source also provides an additional degree of robustness to an integrated system. In short, the circumstances that cause failure or degradation of Loran are very different from those that cause failure or degradation of GPS or DGPS. When the absolute accuracy of Loran is continually calibrated by GPS, the repeatable accuracy of Loran could ensure near-GPS performance of an integrated system in several possible navigation and timing scenarios, for periods of several hours to a few days after a total loss of GPS.

1217. Loran as a Data Transfer Channel

The U.S. Coast Guard has practiced low data rate transmission using Loran signals during various periods since the 1970's. The two primary uses of this capability were Loran chain control and backup military communications. In all cases, the data superimposed on the Loran signal were transparent to the users, who were nearly universally unaware of this dual use.

In the late 1990's, the Northwest European Loran System (NELS) implemented a pulse-position modulation pattern termed Eurofix to provide differential GPS corrections via the Loran signal to certain areas in western and northern Europe. Eurofix successfully incorporated sophisticated data communications techniques to broadcast GPS corrections in real time while allowing traditional Loran users to operate without interruption.

Another possible use of a Loran data transfer channel is to broadcast GPS corrections provided by the U.S. Wide Area Augmentation System (WAAS), which was designed for the benefit of aircraft in the U.S. National Airspace System (NAS). Preliminary tests have shown modulated Loran signals could be successfully used to broadcast WAAS data.

CHAPTER 13

RADAR NAVIGATION

PRINCIPLES OF RADAR OPERATION

1300. Introduction

Radar determines distance to an object by measuring the time required for a radio signal to travel from a transmitter to the object and return. Such measurements can be converted into lines of position (LOP's) comprised of circles with radius equal to the distance to the object. Since marine radars use directional antennae, they can also determine an object's bearing. However, due to its design, a radar's bearing measurement is less accurate than its distance measurement. Understanding this concept is crucial to ensuring the optimal employment of the radar for safe navigation.

1301. Signal Characteristics

In most marine navigation applications, the radar signal is pulse modulated. Signals are generated by a timing circuit so that energy leaves the antenna in very short pulses. When transmitting, the antenna is connected to the transmitter but not the receiver. As soon as the pulse leaves, an electronic switch disconnects the antenna from the transmitter and connects it to the receiver. Another pulse is not transmitted until after the preceding one has had time to travel to the most distant target within range and return. Since the interval between pulses is long compared with the length of a pulse, strong signals can be provided with low average power. The duration or length of a single pulse is called **pulse length**, **pulse duration**, or **pulse width**. This pulse emission sequence repeats a great many times, perhaps 1,000 per second. This rate defines the **pulse repetition rate (PRR)**. The returned pulses are displayed on an indicator screen.

1302. The Display

The radar display is often referred to as the **plan position indicator (PPI)**. On a PPI, the sweep appears as a radial line, centered at the center of the scope and rotating in synchronization with the antenna. Any returned echo causes a brightening of the display screen at the bearing and range of the object. Because of a luminescent coating on the inside of the tube, the glow continues after the trace rotates past the target.

On a PPI, a target's actual range is proportional to its distance from the center of the scope. A moveable cursor helps to measure ranges and bearings. In the "heading-upward" presentation, which indicates relative bearings, the top of the scope represents the direction of the ship's head. In this unstabilized presentation, the orientation changes as the ship changes heading. In the stabilized "north-upward" presentation, gyro north is always at the top of the scope.

1303. The Radar Beam

The pulses of energy comprising the radar beam would form a single lobe-shaped pattern of radiation if emitted in free space. Figure 1303a shows this free space radiation pattern, including the undesirable minor lobes or side lobes associated with practical antenna design.

Although the radiated energy is concentrated into a relatively narrow main beam by the antenna, there is no clearly defined envelope of the energy radiated, although most of the energy is concentrated along the axis of the beam. With the rapid decrease in the amount of radiated energy in directions away from this axis, practical power limits may be used to define the dimensions of the radar beam.

A radar beam's horizontal and vertical beam widths are referenced to arbitrarily selected power limits. The most common convention defines beam width as the angular width between half power points. The half power point corresponds to a drop in 3 decibels from the maximum beam strength.

The definition of the decibel shows this halving of power at a decrease in 3 dB from maximum power. A decibel is simply the logarithm of the ratio of a final power level to a reference power level:

$$dB = 10 \log\left[\frac{P_1}{P_0}\right]$$

where P_1 is the final power level, and P_0 is a reference power level. When calculating the dB drop for a 50% reduction in power level, the equation becomes:

$$dB = 10 \log(.5)$$
$$dB = -3\ dB$$

The radiation diagram shown in Figure 1303b depicts relative values of power in the same plane existing at the same distances from the antenna or the origin of the radar beam. Maximum power is in the direction of the axis of the beam. Power values diminish rapidly in directions away from the axis. The beam width is taken as the angle between the half-power points.

The beam width depends upon the frequency or wavelength of the transmitted energy, antenna design, and the dimensions of the antenna. For a given antenna size (antenna aperture), narrower beam widths result from using shorter wavelengths. For a given wavelength, narrower beam widths result from using larger antennas.

With radar waves being propagated in the vicinity of the surface of the sea, the main lobe of the radar beam is composed of a number of separate lobes, as opposed to the single lobe-shaped pattern of radiation as emitted in free space. This phenomenon is the result of interference between radar waves directly transmitted, and those waves which are reflected from the surface of the sea. Radar waves strike the surface of the sea, and the indirect waves reflect off the surface of the sea. See Figure 1303c. These reflected waves either constructively or destructively interfere with the direct waves depending upon the waves' phase relationship.

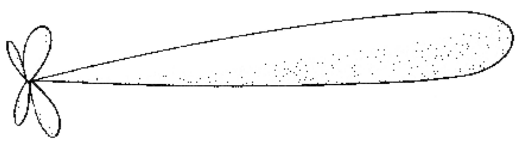


Figure 1303a. Freespace radiation pattern.

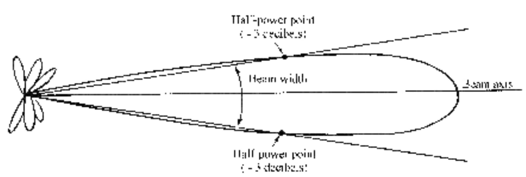


Figure 1303b. Radiation diagram.

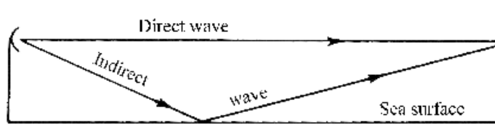


Figure 1303c. Direct and indirect waves.

1304. Diffraction and Attenuation

Diffraction is the bending of a wave as it passes an obstruction. Because of diffraction there is some illumination of the region behind an obstruction or target by the radar beam. Diffraction effects are greater at the lower frequencies. Thus, the radar beam of a lower frequency radar tends to illuminate more of the shadow region behind an obstruction than the beam of a radar of higher frequency or shorter wavelength.

Attenuation is the scattering and absorption of the energy in the radar beam as it passes through the atmosphere. It causes a decrease in echo strength. Attenuation is greater at the higher frequencies or shorter wavelengths.

While reflected echoes are much weaker than the transmitted pulses, the characteristics of their return to the source are similar to the characteristics of propagation. The strengths of these echoes are dependent upon the amount of transmitted energy striking the targets and the size and reflecting properties of the targets.

1305. Refraction

If the radar waves traveled in straight lines, the distance to the radar horizon would be dependent only on the power output of the transmitter and the height of the antenna. In other words, the distance to the radar horizon would be the same as that of the geometrical horizon for the antenna height. However, atmospheric density gradients bend radar rays as they travel to and from a target. This bending is called **refraction**.

The distance to the radar horizon does not limit the distance from which echoes may be received from targets. Assuming that adequate power is transmitted, echoes may be received from targets beyond the radar horizon if their reflecting surfaces extend above it. The distance to the radar horizon is the distance at which the radar rays pass tangent to the surface of the Earth.

The following formula, where h is the height of the antenna in feet, gives the theoretical distance to the radar horizon in nautical miles:

$$d = 1.22\sqrt{h} \ .$$

1306. Factors Affecting Radar Interpretation

Radar's value as a navigational aid depends on the navigator's understanding its characteristics and limitations. Whether measuring the range to a single reflective object or trying to discern a shoreline lost amid severe clutter, knowledge of the characteristics of the individual radar used are crucial. Some of the factors to be considered in interpretation are discussed below:

- **Resolution in Range.** In part A of Figure 1306a, a transmitted pulse has arrived at the second of two targets of insufficient size or density to absorb or reflect all of the energy of the pulse. While the pulse has traveled from the first to the second target, the echo from the first has traveled an equal distance in the opposite direction. At B, the transmitted pulse has continued on beyond the second target, and the two echoes are returning toward the transmitter. The distance between leading edges of the two echoes is twice the distance between targets. The correct distance will be shown on the scope, which is calibrated to show half the distance traveled out and back. At C the targets are closer together and the pulse length has been increased. The two echoes merge, and on the scope they will appear as a single, large target. At D the pulse length has been decreased, and the two echoes appear separated. The ability of a radar to separate targets close together on the same bearing is called **resolution in range**. It is related primarily to pulse length. The minimum distance between targets that can be distinguished as separate is half the pulse length. This (half the pulse length) is the apparent depth or thickness of a target presenting a flat perpendicular surface to the radar beam. Thus, several ships close together may appear as an island. Echoes from a number of small boats, piles, breakers, or even large ships close to the shore may blend with echoes from the shore, resulting in an incorrect indication of the position and shape of the shoreline.

- **Resolution in Bearing**. Echoes from two or more targets close together at the same range may merge to form a single, wider echo. The ability to separate targets close together at the same range is called **resolution in bearing**. Bearing resolution is a function of two variables: beam width and range to the targets. A narrower beam and a shorter distance to the objects both increase bearing resolution.

- **Height of Antenna and Target**. If the radar horizon is between the transmitting vessel and the target, the lower part of the target will not be visible. A large vessel may appear as a small craft, or a shoreline may appear at some distance inland.

- **Reflecting Quality and Aspect of Target**. Echoes from several targets of the same size may be quite different in appearance. A metal surface reflects radio waves more strongly than a wooden surface. A surface perpendicular to the beam returns a stronger echo than a non perpendicular one. A vessel seen broadside returns a stronger echo than one heading directly toward or away. Some surfaces absorb most radar energy rather that reflecting it.

- **Frequency**. As frequency increases, reflections occur from smaller targets.

Atmospheric noise, sea return, and precipitation complicate radar interpretation by producing **clutter**. Clutter is usually strongest near the vessel. Strong echoes can sometimes be detected by reducing receiver gain to eliminate weaker signals. By watching the repeater during several rotations of the antenna, the operator can often discriminate between clutter and a target even when the signal strengths from clutter and the target are equal. At each rotation, the signals from targets will remain relatively stationary on the display while those caused by clutter will appear at different locations on each sweep.

Another major problem lies in determining which features in the vicinity of the shoreline are actually represented by echoes shown on the repeater. Particularly in cases where a low lying shore is being scanned, there may be considerable uncertainty.

A related problem is that certain features on the shore will not return echoes because they are blocked from the radar beam by other physical features or obstructions. This factor in turn causes the chart-like image painted on the scope to differ from the chart of the area.

If the navigator is to be able to interpret the presentation on his radarscope, he must understand the characteristics of radar propagation, the capabilities of his radar set, the reflecting properties of different types of radar targets, and the ability to analyze his chart to determine which charted features are most likely to reflect the transmitted pulses or to be blocked. Experience gained during clear weather comparison between radar and visual images is invaluable.

Land masses are generally recognizable because of the steady brilliance of the relatively large areas painted on the PPI. Also, land should be at positions expected from the ship's navigational position. Although land masses are readily recognizable, the primary problem is the identification of specific land features. Identification of specific features can be quite difficult because of various factors, including distortion resulting from beam width and pulse length, and uncertainty as to just which charted features are reflecting the echoes.

Sand spits and smooth, clear beaches normally do not appear on the PPI at ranges beyond 1 or 2 miles because these targets have almost no area that can reflect energy back to the radar. Ranges determined from these targets are not reliable. If waves are breaking over a sandbar, echoes may be returned from the surf. Waves may, however, break well out from the actual shoreline, so that ranging on the surf may be misleading.

Mud flats and marshes normally reflect radar pulses only a little better than a sand spit. The weak echoes received at low tide disappear at high tide. Mangroves and other thick growth may produce a strong echo. Areas that are indicated as swamps on a chart, therefore, may return either strong or weak echoes, depending on the density type, and size of the vegetation growing in the area.

When sand dunes are covered with vegetation and are well back from a low, smooth beach, the apparent shoreline determined by radar appears as the line of the dunes rather than the true shoreline. Under some conditions, sand dunes may return strong echo signals because the combination of the vertical surface of the vegetation and the horizontal

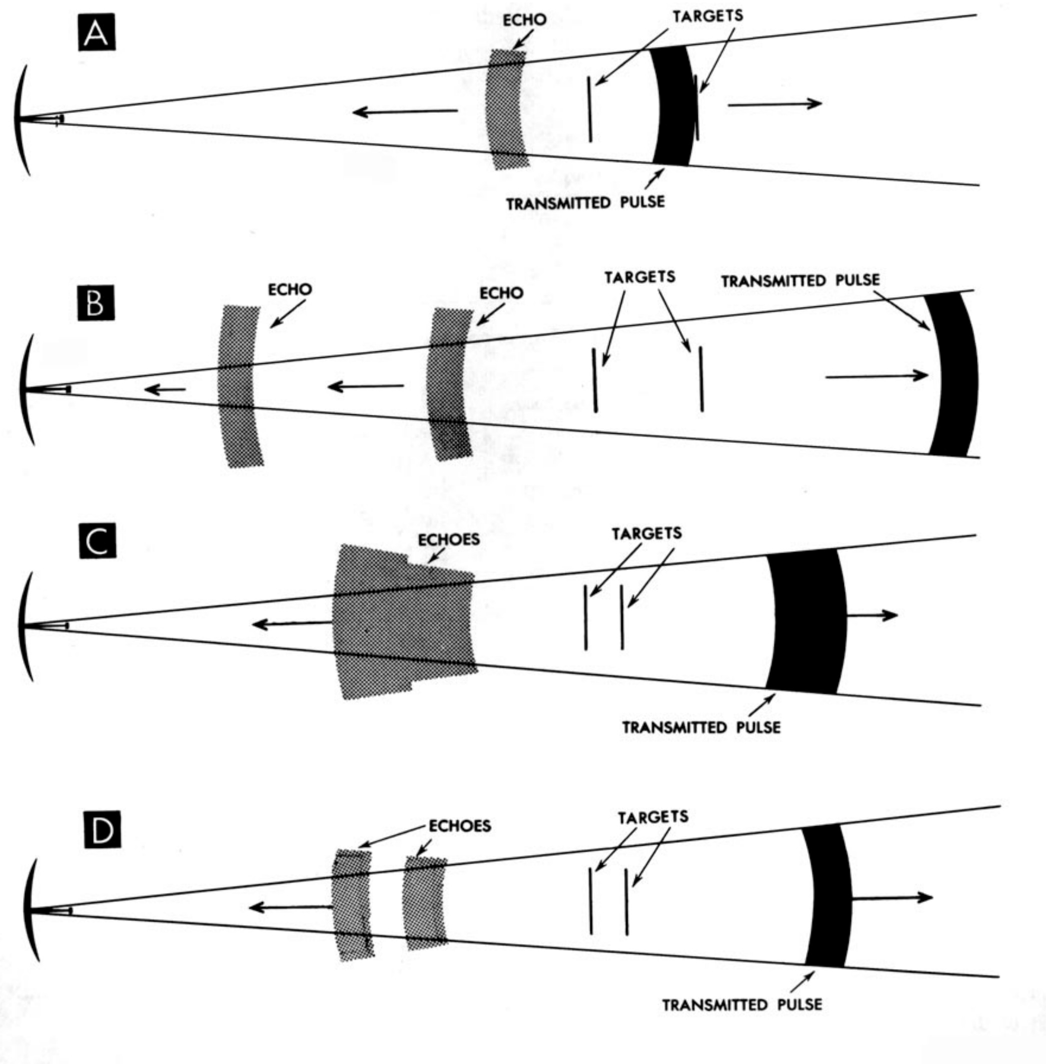


Figure 1306a. Resolution in range.

beach may form a sort of corner reflector.

Lagoons and inland lakes usually appear as blank areas on a PPI because the smooth water surface returns no energy to the radar antenna. In some instances, the sandbar or reef surrounding the lagoon may not appear on the PPI because it lies too low in the water.

Coral atolls and long chains of islands may produce long lines of echoes when the radar beam is directed perpendicular to the line of the islands. This indication is especially true when the islands are closely spaced. The reason is that the spreading resulting from the width of the radar beam causes the echoes to blend into continuous lines. When the chain of islands is viewed lengthwise, or obliquely, however, each island may produce a separate return. Surf breaking on a reef around an atoll produces a ragged, variable line of echoes.

One or two rocks projecting above the surface of the water, or waves breaking over a reef, may appear on the PPI.

If the land rises in a gradual, regular manner from the shoreline, no part of the terrain produces an echo that is stronger than the echo from any other part. As a result, a general haze of echoes appears on the PPI, and it is difficult to ascertain the range to any particular part of the land.

Blotchy signals are returned from hilly ground, because the crest of each hill returns a good echo although the valley beyond is in a shadow. If high receiver gain is used, the pattern may become solid except for the very deep shadows.

Low islands ordinarily produce small echoes. When thick palm trees or other foliage grow on the island, strong echoes often are produced because the horizontal surface of the water around the island forms a sort of corner reflector with the vertical surfaces of the trees. As a result, wooded islands give good echoes and can be detected at a much

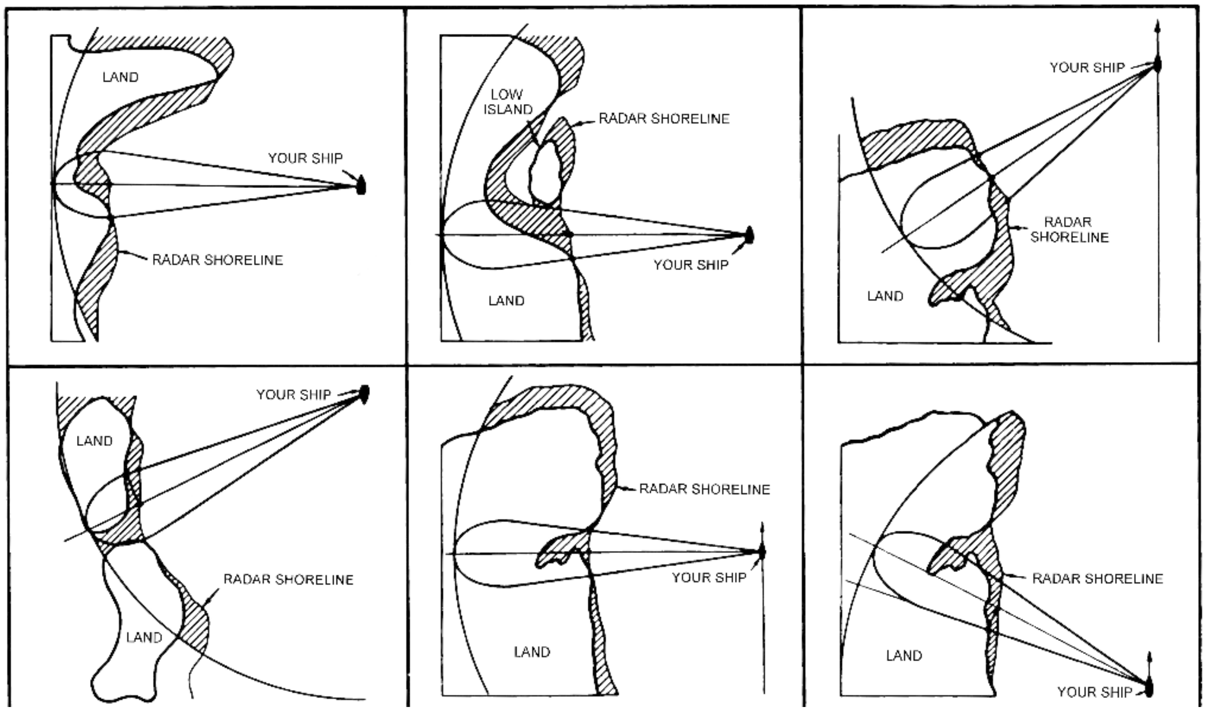


Figure 1306b. Effects of ship's position, beam width, and pulse length on radar shoreline.

greater range than barren islands.

Sizable land masses may be missing from the radar display because of certain features being blocked from the radar beam by other features. A shoreline which is continuous on the PPI display when the ship is at one position, may not be continuous when the ship is at another position and scanning the same shoreline. The radar beam may be blocked from a segment of this shoreline by an obstruction such as a promontory. An indentation in the shoreline, such as a cove or bay, appearing on the PPI when the ship is at one position, may not appear when the ship is at another position nearby. Thus, radar shadow alone can cause considerable differences between the PPI display and the chart presentation. This effect in conjunction with beam width and pulse length distortion of the PPI display can cause even greater differences.

The returns of objects close to shore may merge with the shoreline image on the PPI, because of distortion effects of horizontal beam width and pulse length. Target images on the PPI are distorted angularly by an amount equal to the effective horizontal beam width. Also, the target images always are distorted radially by an amount at least equal to one-half the pulse length (164 yards per microsecond of pulse length).

Figure 1306b illustrates the effects of ship's position, beam width, and pulse length on the radar shoreline. Because of beam width distortion, a straight, or nearly straight, shoreline often appears crescent-shaped on the PPI. This effect is greater with the wider beam widths. Note that this distortion increases as the angle between the beam axis and the shoreline decreases.

Figure 1306c illustrates the distortion effects of radar shadow, beam width, and pulse length. View A shows the actual shape of the shoreline and the land behind it. Note the steel tower on the low sand beach and the two ships at anchor close to shore. The heavy line in view B represents the shoreline on the PPI. The dotted lines represent the actual position and shape of all targets. Note in particular:

1. The low sand beach is not detected by the radar.
2. The tower on the low beach is detected, but it looks like a ship in a cove. At closer range the land would be detected and the cove-shaped area would begin to fill in; then the tower could not be seen without reducing the receiver gain.
3. The radar shadow behind both mountains. Distortion owing to radar shadows is responsible for more confusion than any other cause. The small island does not appear because it is in the radar shadow.
4. The spreading of the land in bearing caused by beam width distortion. Look at the upper shore of the peninsula. The shoreline distortion is greater to the west because the angle between the radar beam and the shore is smaller as the beam seeks out the more westerly shore.
5. Ship No. 1 appears as a small peninsula. Its return has merged with the land because of the beam width

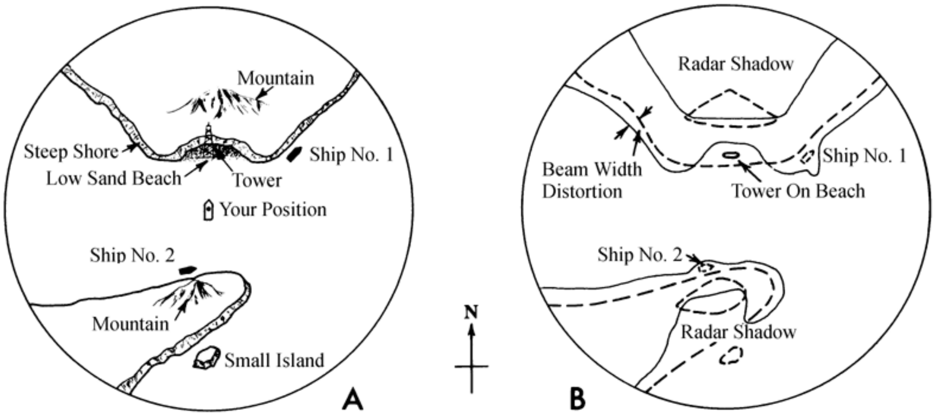


Figure 1306c. Distortion effects of radar shadow, beam width, and pulse length.

distortion.

6. Ship No. 2 also merges with the shoreline and forms a bump. This bump is caused by pulse length and beam width distortion. Reducing receiver gain might cause the ship to separate from land, provided the ship is not too close to the shore. The Fast Time Constant (FTC) control could also be used to attempt to separate the ship from land.

1307. Recognition of Unwanted Echoes

Indirect or false echoes are caused by reflection of the main lobe of the radar beam off ship's structures such as stacks and kingposts. When such reflection does occur, the echo will return from a legitimate radar contact to the antenna by the same indirect path. Consequently, the echo will appear on the PPI at the bearing of the reflecting surface. As shown in Figure 1307a, the indirect echo will appear on the PPI at the same range as the direct echo received, assuming that the additional distance by the indirect path is negligible.

Characteristics by which indirect echoes may be recognized are summarized as follows:

1. Indirect echoes will often occur in shadow sectors.
2. They are received on substantially constant bearings, although the true bearing of the radar contact may change appreciably.
3. They appear at the same ranges as the corresponding direct echoes.
4. When plotted, their movements are usually abnormal.
5. Their shapes may indicate that they are not direct echoes.

Side-lobe effects are readily recognized in that they produce a series of echoes (Figure 1307b) on each side of the main lobe echo at the same range as the latter. Semicircles, or even complete circles, may be produced. Because of the low energy of the side-lobes, these effects will normally occur only at the shorter ranges. The effects may be minimized or eliminated, through use of the gain and anti-clutter controls. Slotted wave guide antennas have largely eliminated the side-lobe problem.

Multiple echoes may occur when a strong echo is received from another ship at close range. A second or third or more echoes may be observed on the radarscope at double, triple, or other multiples of the actual range of the radar contact (Figure 1307c).

Second-trace echoes (multiple-trace echoes) are echoes received from a contact at an actual range greater than the radar range setting. If an echo from a distant target is received after the following pulse has been transmitted, the echo will appear on the radarscope at the correct bearing but not at the true range. Second-trace echoes are unusual, except under abnormal atmospheric conditions, or conditions under which super-refraction is present. Second-trace echoes may be recognized through changes in their positions on the radarscope in changing the pulse repetition rate (PRR); their hazy, streaky, or distorted shape; and the erratic movements on plotting.

As illustrated in Figure 1307d, a target return is detected on a true bearing of 090° at a distance of 7.5 miles. On changing the PRR from 2,000 to 1,800 pulses per second, the same target is detected on a bearing of 090° at a distance of 3 miles (Figure 1307e). The change in the position of the return indicates that the return is a second-trace echo. The actual distance of the target is the distance as indicated on the PPI plus half the distance the radar wave travels between pulses.

Electronic interference effects, such as may occur

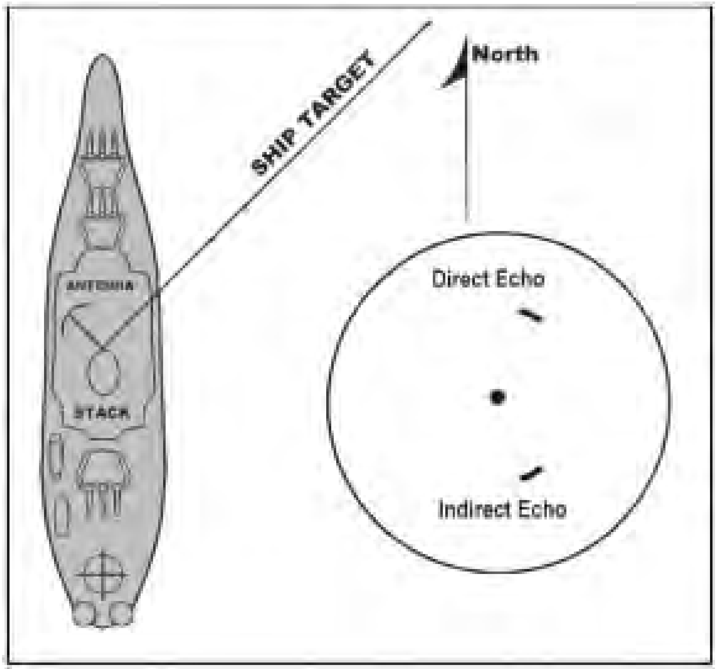


Figure 1307a. Indirect echo.

when near another radar operating in the same frequency band as that of the observer's ship, is usually seen on the PPI as a large number of bright dots either scattered at random or in the form of dotted lines extending from the center to the edge of the PPI.

Interference effects are greater at the longer radar range scale settings. The interference effects can be distinguished easily from normal echoes because they do not appear in the same places on successive rotations of the antenna.

Stacks, masts, samson posts, and other structures, may cause a reduction in the intensity of the radar beam beyond these obstructions, especially if they are close to the radar antenna. If the angle at the antenna subtended by the obstruction is more than a few degrees, the reduction of the intensity of the radar beam beyond the obstruction may produce a blind sector. Less reduction in the intensity of the beam beyond the obstructions may produce shadow sectors. Within a shadow sector, small targets at close range may not be detected, while larger targets at much greater ranges will appear.

Spoking appears on the PPI as a number of spokes or radial lines. Spoking is easily distinguished from interference effects because the lines are straight on all range-scale settings, and are lines rather than a series of dots.

The spokes may appear all around the PPI, or they may be confined to a sector. If spoking is confined to a narrow sector, the effect can be distinguished from a Ramark signal of similar appearance through observation of the steady relative bearing of the spoke in a situation where the bearing of the Ramark signal should change. Spoking indicates a need for maintenance or adjustment. The PPI display may appear as normal sectors alternating with dark sectors. This is usually due to the automatic frequency control being out of adjustment. The appearance of serrated range rings indicates a need for maintenance.

After the radar set has been turned on, the display may not spread immediately to the whole of the PPI because of static electricity inside the CRT. Usually, the static electricity effect, which produces a distorted PPI display, lasts no longer than a few minutes.

Hour-glass effect appears as either a constriction or expansion of the display near the center of the PPI. The expansion effect is similar in appearance to the expanded center display. This effect, which can be caused by a nonlinear time base or the sweep not starting on the indicator at the same instant as the transmission of the pulse, is most apparent when in narrow rivers or close to shore.

The echo from an overhead power cable can be wrongly identified as the echo from a ship on a steady bearing and decreasing range. Course changes to avoid the contact are ineffective; the contact remains on a steady bearing, decreasing range. This phenomenon is particularly apparent for the power cable spanning the Straits of Messina.

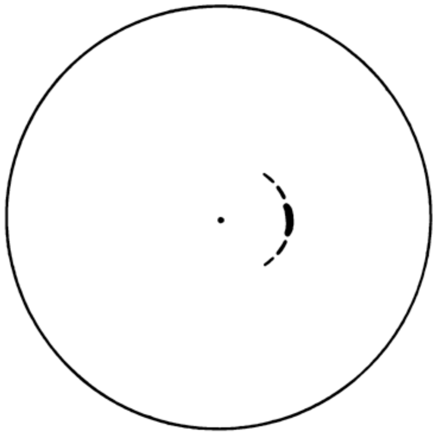


Figure 1307b. Side-lobe effects.

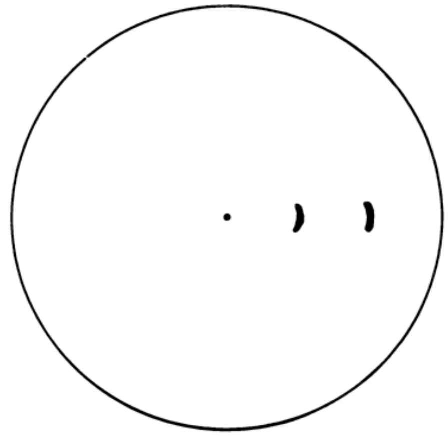


Figure 1307c. Multiple echoes.

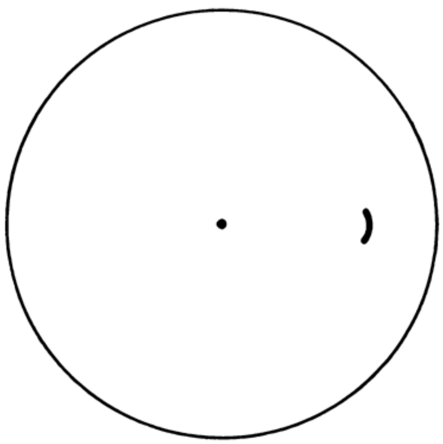


Figure 1307d. Second-trace echo on 12-mile range scale.

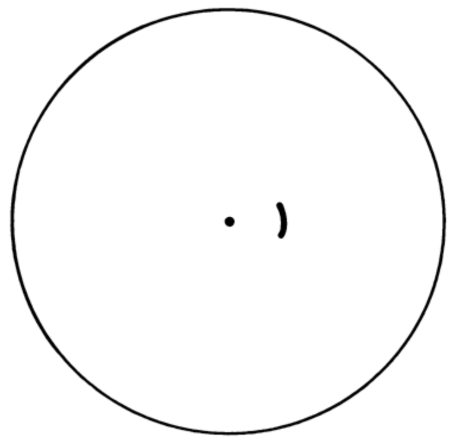


Figure 1307e. Position of second-trace echo on 12-mile range scale after changing PRR.

1308. Aids to Radar Navigation

Radar navigation aids help identify radar targets and increase echo signal strength from otherwise poor radar targets.

Buoys are particularly poor radar targets. Weak, fluctuating echoes received from these targets are easily lost in the sea clutter. To aid in the detection of these targets, **radar reflectors**, designated corner reflectors, may be used. These reflectors may be mounted on the tops of buoys or designed into the structure.

Each corner reflector, shown in Figure 1308a, consists of three mutually perpendicular flat metal surfaces. A radar wave striking any of the metal surfaces or plates will be reflected back in the direction of its source. Maximum energy will be reflected back to the antenna if the axis of the radar beam makes equal angles with all the metal surfaces. Frequently, corner reflectors are assembled in clusters to maximize the reflected signal.

Although radar reflectors are used to obtain stronger echoes from radar targets, other means are required for more positive identification of radar targets. **Radar beacons** are transmitters operating in the marine radar frequency band, which produce distinctive indications on the radarscopes of ships within range of these beacons. There are two general classes of these beacons: **racons**, which provide both bearing and range information to the target, and **ramarks** which provide bearing information only. However, if the ramark installation is detected as an echo on the radarscope, the range will be available also.

A racon is a radar transponder which emits a characteristic signal when triggered by a ship's radar. The signal may be emitted on the same frequency as that of the triggering radar, in which case it is superimposed on the ship's radar display automatically. The signal may be emitted on a separate frequency, in which case to receive the signal the ship's radar receiver must be tuned to the beacon frequency, or a special receiver must be used. In either case, the PPI will be blank except for the beacon signal. However, the only racons in service are "in band"

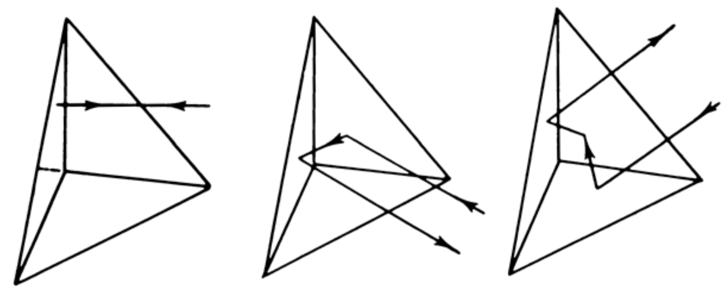


Figure 1308a. Corner reflectors.

beacons which transmit in one of the marine radar bands, usually only the 3-centimeter band.

The racon signal appears on the PPI as a radial line originating at a point just beyond the position of the radar beacon, or as a Morse code signal (Figure 1308b) displayed radially from just beyond the beacon.

A ramark is a radar beacon which transmits either continuously or at intervals. The latter method of transmission is used so that the PPI can be inspected without any clutter introduced by the ramark signal on the scope. The ramark signal as it appears on the PPI is a radial line from the center. The radial line may be a continuous narrow line, a broken line (Figure 1308c), a series of dots, or a series of dots and dashes.

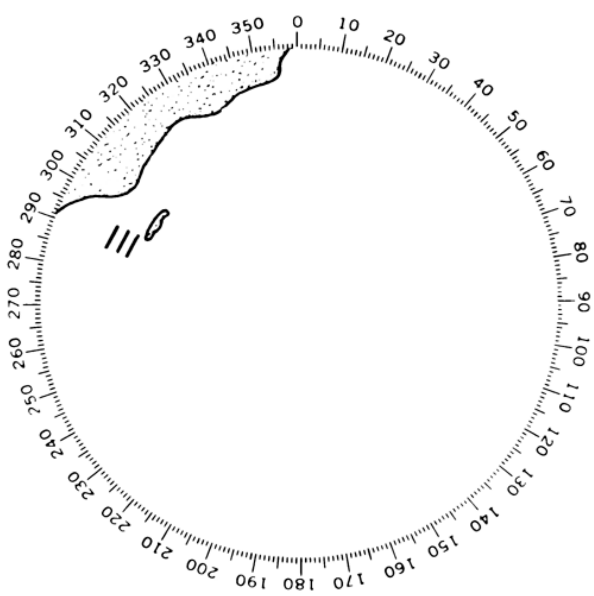


Figure 1308b. Coded racon signal.

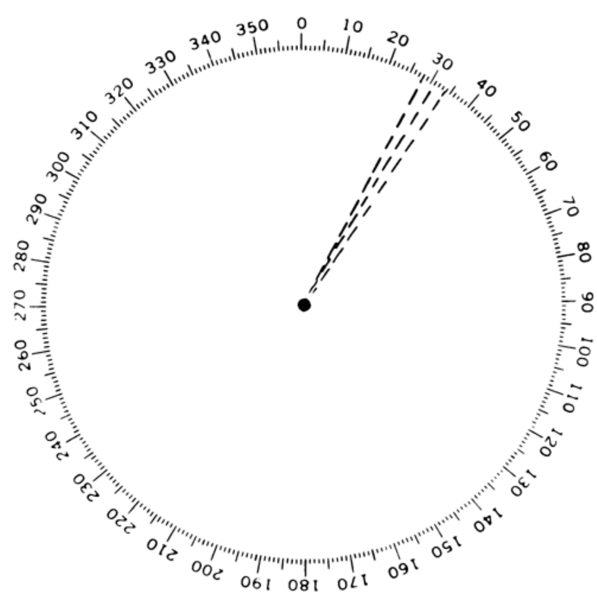


Figure 1308c. Ramark appears as broken radial line.

RADAR PILOTING

1309. Introduction

When navigating in restricted waters, a mariner most often relies on visual piloting to provide the accuracy required to ensure ship safety. Visual piloting, however, requires clear weather; often, mariners must navigate through fog. When weather conditions render visual piloting impossible on a vessel not equipped with ECDIS, radar navigation provides a method of fixing a vessel's position with sufficient accuracy to allow safe passage. See Chapter 8 for a detailed discussion of integrating radar into a piloting procedure.

1310. Fix by Radar Ranges

Since radar can more accurately determine ranges than bearings, the most accurate radar fixes result from measuring and plotting ranges to two or more objects. Measure objects directly ahead or astern first; measure objects closest to the beam last.

This procedure is the opposite to that recommended for taking visual bearings, where objects closest to the beam are measured first; however, both recommendations rest on the same principle. When measuring objects to determine a line of position, measure first those which have the greatest rate of change in the quantity being measured; measure last those which have the least rate of change. This minimizes measurement time delay errors. Since the range of those objects directly ahead or astern of the ship changes more rapidly than those objects located abeam, we measure ranges to objects ahead or astern first.

Record the ranges to the navigation aids used and lay the resulting range arcs down on the chart. Theoretically, these lines of position should intersect at a point coincident with the ship's position at the time of the fix.

Though verifying soundings is always a good practice in all navigation scenarios, its importance increases when piloting using only radar. Assuming proper operation of the fathometer, soundings give the navigator invaluable information on the reliability of his fixes.

1311. Fix by Range and Bearing to One Object

Visual piloting requires bearings from at least two objects; radar, with its ability to determine both bearing and range from one object, allows the navigator to obtain a fix where only a single navigation aid is available. An example of using radar in this fashion occurs in approaching a harbor whose entrance is marked with a single, prominent object such as Chesapeake Light at the entrance of the Chesapeake Bay. Well beyond the range of any land-based visual navigation aid, and beyond the visual range of the light itself, a shipboard radar can detect the light and provide bearings and ranges for the ship's piloting party. Care must be taken that fixes are not taken on any nearby stationary vessel.

This methodology is limited by the inherent inaccuracy associated with radar bearings; typically, a radar bearing is accurate to within about 5° of the true bearing. Therefore, the navigator must carefully evaluate the resulting position, possibly checking it with a sounding. If a visual bearing is available from the object, use that bearing instead of the radar bearing when laying down the fix. This illustrates the basic concept discussed above: radar ranges are inherently more accurate than radar bearings. One must also be aware that if the radar is gyro stabilized and there is a gyro error of more than a degree or so, radar bearings will be in error by that amount.

Prior to using this method, the navigator must ensure that he has correctly identified the object from which the bearing and range are to be taken. Using only one navigation aid for both lines of position can lead to disaster if the navigation aid is not properly identified.

1312. Fix Using Tangent Bearings and Range

This method combines bearings tangent to an object with a range measurement from some point on that object. The object must be large enough to provide sufficient bearing spread between the tangent bearings; often an island or peninsula works well. Identify some prominent feature of the object that is displayed on both the chart and the radar display. Take a range measurement from that feature and plot it on the chart. Then determine the tangent bearings to the feature and plot them on the chart.

Steep-sided features work the best. Tangents to low, sloping shorelines will seriously reduce accuracy, as will tangent bearings in areas of excessively high tides, which can change the location of the apparent shoreline by many meters.

1313. Fix by Radar Bearings

The inherent inaccuracy of radar bearings discussed above makes this method less accurate than fixing position by radar range. Use this method to plot a position quickly on the chart when approaching restricted waters to obtain an approximate ship's position for evaluating radar targets to use for range measurements. Unless no more accurate method is available, this method is not suitable while piloting in restricted waters.

1314. Fischer Plotting

In Fischer plotting, the navigator adjusts the scale of the radar to match the scale of the chart in use. Then he places a clear plastic disk, sized to the radar, on the center of the radar screen and quickly traces the shape of land and location of any navigation aids onto the plastic overlay with a grease pencil. Taking the plastic with the tracings on it to the chart, he matches the features traced from the radar with the chart's features. A hole in the center of the plastic allows the navigator to mark the position of the ship at the time the tracing was done.

RASTER RADARS

1315. Basic Description

Conventional PPI-display radars use a circular **Cathode Ray Tube (CRT)** to direct an electron beam at a screen coated on the inside with phosphorus, which glows when illuminated by the beam. Internal circuitry forms the beam such that a "sweep" is indicated on the face of the PPI. This sweep is timed to coincide with the sweep of the radar's antenna. A return echo is added to the sweep signal so that the screen is more brightly illuminated at a point corresponding to the bearing and range of the target that returned the echo.

The raster radar also employs a cathode ray tube; however, the end of the tube upon which the picture is formed is rectangular, not circular as in the PPI display. The raster radar does not produce its picture from a circular sweep; it utilizes a liner scan in which the picture is "drawn," line by line, horizontally across the screen. As the sweep moves across the screen, the electron beam from the CRT illuminates the picture elements, or **pixels**, on the screen. A pixel is the smallest area of a display that can be excited individually.

In order to produce a sufficiently high resolution, larger raster radars require over 1 million pixels per screen combined with an update rate of 60 or more scans per second. Processing such a large number of pixel elements requires a rather sophisticated computer. One way to lower cost is to slow down the required processing speed. This speed can be lowered to approximately 30 frames per second before the picture develops a noticeable flicker, but the best radars have scan rates of at least 60 scans per second.

Further cost reduction can be gained by using an **interlaced display**. An interlaced display does not draw the entire picture in one pass. On the first pass, it draws every other line; it draws the remaining lines on the second pass. This type of display reduces the number of screens that have to be drawn per unit of time by a factor of two; however, if the two pictures are misaligned, the picture will appear to jitter.

CHAPTER 14

ELECTRONIC CHARTS

INTRODUCTION

1400. The Importance of Electronic Charts

Since the beginning of maritime navigation, the desire of the navigator has always been to answer a fundamental question: "Where, exactly, is my vessel?" To answer that question, the navigator was forced to continually take fixes on celestial bodies, on fixed objects ashore, or using radio signals, and plot the resulting lines of position as a fix on a paper chart. Only then could he begin to assess the safety of the ship and its progress toward its destination. He spent far more time taking fixes, working out solutions, and plotting the results than on making assessments, and the fix only told him where the ship was at the time that fix was taken, not where the vessel was some time later when the assessment was made. He was always "behind the vessel." On the high seas this is of little import. Near shore, it becomes vitally important.

Electronic charts automate the process of integrating real-time positions with the chart display and allow the navigator to continuously assess the position and safety of the vessel. Further, the GPS/DGPS fixes are far more accurate and taken far more often than any navigator ever could. A good piloting team is expected to take and plot a fix every three minutes. An electronic chart system can do it once per second to a standard of accuracy at least an order of magnitude better.

Electronic charts also allow the integration of other operational data, such as ship's course and speed, depth soundings, and radar data into the display. Further, they allow automation of alarm systems to alert the navigator to potentially dangerous situations well in advance of a disaster.

Finally, the navigator has a complete picture of the instantaneous situation of the vessel and all charted dangers in the area. With a radar overlay, the tactical situation with respect to other vessels is clear as well. This chapter will discuss the various types of electronic charts, the requirements for using them, their characteristics, capabilities and limitations.

1401. Terminology

Before understanding what an electronic chart is and what it does, one must learn a number of terms and definitions. We must first make a distinction between official and unofficial charts. Official charts are those, and only those, produced by a government hydrographic office (HO). Unofficial charts are produced by a variety of private companies and may or may not meet the same standards used by HO's for data accuracy, currency, and completeness.

An **electronic chart system (ECS)** is a commercial electronic chart system not designed to satisfy the regulatory requirements of the IMO Safety of Life at Sea (SOLAS) convention. ECS is an aid to navigation and when used on SOLAS regulated vessels is to be used in conjunctions with corrected paper charts.

An **electronic chart display and information system (ECDIS)** is an electronic chart system which satisfies the IMO SOLAS convention carriage requirements for corrected paper charts when used with an ENC or its functional equivalent (e.g. NIMA Digital Nautical Chart.)

An **electronic chart (EC)** is any digitized chart intended for display on a computerized navigation system.

An **electronic chart data base (ECDB)** is the digital database from which electronic charts are produced.

An **electronic navigational chart (ENC)** is an electronic chart issued by a national hydrographic authority designed to satisfy the regulatory requirements for chart carriage.

The **electronic navigation chart database (ENCDB)** is the hydrographic database from which the ENC is produced.

The **system electronic navigation chart (SENC)** is the database created by an ECDIS from the ENC data.

A **raster navigation chart (RNC)**is a raster-formatted chart produced by a national hydrographic office.

A **raster chart display system (RCDS)** is a system which displays official raster-formatted charts on an ECDIS system. Raster charts cannot take the place of paper charts because they lack key features required by the IMO, so that when an ECDIS uses raster charts it operates in the ECS mode.

Overscale and **underscale** refer to the display of electronic chart data at too large and too small a scale, respectively. In the case of overscale, the display is "zoomed in" too close, beyond the standard of accuracy to which the data was digitized. Underscale indicates that larger scale data is available for the area in question. ECDIS provides a warning in either case.

Raster chart data is a digitized "picture" of a chart comprised of millions of "picture elements" or "pixels." All data is in one layer and one format. The video display simply reproduces the picture from its digitized data file. With raster data, it is difficult to change individual elements of the chart since they are not separated in the data file. Raster data files tend to be large, since a data point with associated color and intensity values must be entered for every pixel on the chart.

Vector chart data is data that is organized into many separate files or layers. It contains graphics files and programs to produce certain symbols, points, lines, and areas with associated colors, text, and other chart elements. The programmer can change individual elements in the file and link elements to additional data. Vector files of a given area are a fraction the size of raster files, and at the same time much more versatile. The navigator can selectively display vector data, adjusting the display according to his needs. Vector data supports the computation of precise distances between features and can provide warnings when hazardous situations arise.

1402. Components of ECS's and ECDIS's

The terms ECS and ECDIS encompasses many possible combinations of equipment and software designed for a variety of navigational purposes. In general, the following components comprise an ECS or ECDIS.

• **Computer processor, software, and network:** These subsystems control the processing of information from the vessel's navigation sensors and the flow of information between various system components. Electronic positioning information from GPS or Loran C, contact information from radar, and digital compass data, for example, can be integrated with the electronic chart data.

• **Chart database:** At the heart of any ECS lies a database of digital charts, which may be in either raster or vector format. It is this dataset, or a portion of it, that produces the chart seen on the display screen.

• **System display:** This unit displays the electronic chart and indicates the vessel's position on it, and provides other information such as heading, speed, distance to the next waypoint or destination, soundings, etc. There are two modes of display, **relative** and **true**. In the relative mode the ship remains fixed in the center of the screen and the chart moves past it. This requires a lot of computer power, as all the screen data must be updated and re-drawn at each fix. In true mode, the chart remains fixed and the ship moves across it. The display may also be north-up or course-up, according to the availability of data from a heading sensor such as a digital compass.

• **User interface:** This is the user's link to the system. It allows the navigator to change system parameters, enter data, control the display, and operate the various functions of the system. Radar may be integrated with the ECDIS or ECS for navigation or collision avoidance, but is not required by SOLAS regulations.

1403. Legal Aspects of Using Electronic Charts

Requirements for carriage of charts are found in SOLAS Chapter V, which states in part: "All ships shall carry adequate and up-to-date charts... necessary for the intended voyage." As electronic charts have developed and the supporting technology has matured, regulations have been adopted internationally to set standards for what constitutes a "chart" in the electronic sense, and under what conditions such a chart will satisfy the chart carriage requirement.

An extensive body of rules and regulations controls the production of ECDIS equipment, which must meet certain high standards of reliability and performance. By definition, **only an ECDIS can replace a paper chart.** No system which is not an ECDIS relieves the navigator of the responsibility of maintaining a plot on a corrected paper chart. Neither can the presence of an electronic chart system substitute for good judgement, sea sense, and taking all reasonable precautions to ensure the safety of the vessel and crew.

An electronic chart system should be considered as an aid to navigation, one of many the navigator might have at his disposal to help ensure a safe passage. While possessing revolutionary capabilities, it must be considered as a tool, not an infallible answer to all navigational problems. The rule for the use of electronic charts is the same as for all other aids to navigation: The prudent navigator will never rely completely on any single one.

CAPABILITIES AND PERFORMANCE STANDARDS

1404. ECDIS Performance Standards

The specifications for ECDIS consist of a set of interrelated standards from three organizations, the International Maritime Organization (IMO), the International Hydrographic Organization (IHO), and the International Electrotechnical Commission (IEC). The IMO published a resolution in November 1995 to establish performance standards for the general functionality of ECDIS, and to define the conditions for its replacement of paper charts. It consisted of a 15-section annex and 5 original appendices. Appendix 6 was adopted in 1996 to define the backup requirements for ECDIS. Appendix 7 was adopted in 1998 to define the operation of ECDIS in a raster chart mode. Previous standards related only to vector data.

The IMO performance standards refer to IHO Special

Publication S-52 for specification of technical details pertaining to the ECDIS display. Produced in 1997, the 3rd edition of S-52 includes appendices specifying the issue, updating, display, color, and symbology of official electronic navigational charts (ENC), as well as a revised glossary of ECDIS-related terms. The IMO performance standards also refer to IEC International Standard 61174 for the requirements of type approval of an ECDIS. Published in 1998, the IEC standard defines the testing methods and required results for an ECDIS to be certified as compliant with IMO standards. Accordingly, the first ECDIS was given type approval by Germany's classification society (BSH) in 1999. Since then, several other makes of ECDIS have gained type approval by various classification societies.

The IMO performance standards specify the following general requirements: Display of government-authorized vector chart data including an updating capability; enable route planning, route monitoring, manual positioning, and continuous plotting of the ship's position; have a presentation as reliable and available as an official paper chart; provide appropriate alarms or indications regarding displayed information or malfunctions; and permit a mode of operation with raster charts similar to the above standards.

The performance standards also specify additional functions, summarized as follows:

- Display of system information in three selectable levels of detail
- Means to ensure correct loading of ENC data and updates
- Apply updates automatically to system display
- Protect chart data from any alteration
- Permit display of update content
- Store updates separately and keep records of application in system
- Indicate when user zooms too far in or out on a chart (over- or under-scale) or when a larger scale chart is available in memory
- Permit the overlay of radar image and ARPA information onto the display
- Require north-up orientation and true motion mode, but permit other combinations
- Use IHO-specified resolution, colors and symbols
- Use IEC-specified navigational elements and parameters (range & bearing marker, position fix, own ship's track and vector, waypoint, tidal information, etc.)
- Use specified size of symbols, letters and figures at scale specified in chart data
- Permit display of ship as symbol or in true scale
- Display route planning and other tasks
- Display route monitoring
- Permit display to be clearly viewed by more than one user in day or night conditions
- Permit route planning in straight and curved segments and adjustment of waypoints
- Display a route plan in addition to the route selected for monitoring
- Permit track limit selection and display an indication if track limit crosses a safety contour or a selected prohibited area
- Permit display of an area away from ship while continuing to monitor selected route
- Give an alarm at a selectable time prior to ship crossing a selected safety contour or prohibited area
- Plot ship's position using a continuous positioning system with an accuracy consistent with the requirements of safe navigation
- Identify selectable discrepancy between primary and secondary positioning system
- Provide an alarm when positioning system input is lost
- Provide an alarm when positioning system and chart are based on different geodetic datums
- Store and provide for replay the elements necessary to reconstruct navigation and verify chart data in use during previous 12 hours
- Record the track for entire voyage with at least four hour time marks
- Permit accurate drawing of ranges and bearings not limited by display resolution
- Require system connection to continuous position-fixing, heading and speed information
- Neither degrade nor be degraded by connection to other sensors
- Conduct on-board tests of major functions with alarm or indication of malfunction
- Permit normal functions on emergency power circuit
- Permit power interruptions of up to 45 seconds without system failure or need to reboot
- Enable takeover by backup unit to continue navigation if master unit fails,

Before an IMO-compliant ECDIS can replace paper charts on vessels governed by SOLAS regulations, the route of the intended voyage must be covered completely by ENC data, that ENC data must include the latest updates, the ECDIS installation must be IMO-compliant including the master-slave network with full sensor feed to both units, and the national authority of the transited waters must allow for paperless navigation through published regulations. The latter may also include requirements for certified training in the operational use of ECDIS.

The first type approval was earned in 1999 and since the finalization of the standards in 1998, many manufacturers of ECDIS equipment have gained such certification.

The certifying agency issues a certificate valid for two years. For renewal, a survey is conducted to ensure that systems, software versions, components and materials used comply with type-approved documents and to review possible changes in design of systems, software versions, components, materials performance, and make sure that such changes do not affect the type approval granted.

Manufacturers have been willing to provide type-approved ECDIS to vessel operators, but in a non-compliant installation. Without the geographical coverage of ENC data, the expensive dual-network installation required by ECDIS will not eliminate the requirement to carry a corrected portfolio of paper charts. These partial installations range from approved ECDIS software in a single PC, to ECDIS with its IEC-approved hardware. In these instances, plotting on paper charts continues to be the primary means of navigation. As more ENC data and updates become available, and as governments regulate paperless transits, vessel operators are upgrading their installations to meet full IMO compliance and to make ECDIS the primary means of navigation.

1405. ECS Standards

Although the IMO has declined to issue guidelines on ECS, the Radio Technical Commission for Maritime Services (RTCM) in the United States developed a voluntary, industry-wide standard for ECS. Published in December 1994, the RTCM Standard called for ECS to be capable of executing basic navigational functions, providing continuous plots of own ship position, and providing appropriate indicators with respect to information displayed. The RTCM ECS Standard allows the use of either raster or vector data, and includes the requirement for simple and reliable updating of information, or an indication that the electronic chart information has changed.

In November 2001, RTCM published Version 2.1 of the "RTCM Recommended Standards for Electronic Chart Systems." This updated version is intended to better define requirements applicable to various classes of vessels operating in a variety of areas. Three general classes of vessels are designated:

Large commercial vessels (oceangoing ships)
Small commercial vessels (tugs, research vessels. etc.)
Smaller craft (yachts, fishing boats, etc.)

The intent is that users, manufacturers, and regulatory authorities will have a means of differentiating between the needs of various vessels as relates to ECS. In concept, an ECS meeting the minimum requirements of the RTCM standard should reduce the risk of incidents and improve the efficiency of navigating for many types of vessels.

However, unlike IMO-compliant ECDIS, an ECS is not intended to comply with the up-to-date chart requirements of SOLAS. As such, an ECS must be considered as a single aid to navigation, and should always be used with a corrected chart from a government-authorized hydrographic office.

Initially, IMO regulations require the use of vector data in an ECDIS; raster data does not have the flexibility needed to do what the ECDIS must do. But it soon became clear that the hydrographic offices of the world would not be able to produce vector data for any significant part of the world for some years. Meanwhile, commercial interests were rasterizing charts as fast as they could for the emerging electronic chart market, and national hydrographic offices began rasterizing their own inventories to meet public demand. The result was a rather complete set of raster data for the most heavily travelled waters of the world, while production of man-power intensive vector data lagged far behind. IMO regulations were then amended to allow ECDIS to function in an RCDS mode using official raster data in conjunction with an appropriate portfolio of corrected paper charts. Nations may issue regulations authorizing the use of RCDS and define what constitutes an appropriate folio of paper charts for use in their waters.

In general, an ECS is not designed to read and display the S-57 format, and does not meet the performance standards of either ECDIS or RCDS. But an ECDIS can operate in ECS mode when using raster charts or when using non-S-57 vector charts. When a type-approved ECDIS is installed without being networked to a backup ECDIS, or when it is using non-official ENC data, or ENC data without updates, it can be said to be operating in an ECS mode, and as such cannot be used as a substitute for official, corrected paper charts.

1406. Display Characteristics

While manufacturers of electronic chart systems have designed their own proprietary colors and symbols, the IMO Performance Standard requires that all IMO approved ECDIS follow the International Hydrographic Organization (IHO) Color & Symbol Specifications. These specifications are embodied in the ECDIS Presentation Library. Their development was a joint effort between Canada and Germany during the 1990s. In order for ECDIS to enhance the safety of navigation, every detail of the display should be clearly visible, unambiguous in its meaning, and uncluttered by superfluous information. The unofficial ECS's continue to be free to develop independent of IHO control. In general they seek to emulate the look of the traditional paper chart.

To reduce clutter, the IMO Standard lays down a permanent display base of essentials such as depths, aids to navigation, shoreline, etc., making the remaining information selectable. The navigator may then select only what is essential for the navigational task at hand. A black-background display for night use provides good color contrast without compromising the mariner's night vision. Similarly, a "bright sun" color table is designed to output maximum luminance in order to be daylight visible, and the colors for details such as buoys are made as contrasting as possible.

The symbols for ECDIS are based on the familiar paper chart symbols, with some optional extras such as simplified buoy symbols that show up better at night. Since the ECDIS can be customized to each ship's requirements, new symbols were added such as a highlighted, mariner selectable, safety contour and a prominent isolated danger symbol.

The Presentation Library is a set of colors and symbols together with rules relating them to the digital data of the ENC, and procedures for handling special cases, such as priorities for the display of overlapping objects. Every feature in the ENC is first passed through the look-up table of the Presentation Library that either assigns a symbol or line style immediately, or, for complex cases, passes the object to a symbology procedure. Such procedures are used for objects like lights, which have so many variations that a look-up table for their symbolization would be too long. The Presentation Library includes a Chart 1, illustrating the symbology. Given the IHO S-57 data standards and S-52 display specifications, a waterway should look the same no matter which hydrographic office produced the ENC, and no matter which manufacturer built the ECDIS.

The overwhelming advantage of the vector-based ECDIS display is its ability to remove cluttering information not needed at a given time. By comparison, the paper chart and its raster equivalent is an unchangeable diagram. A second advantage is the ability to orient the display course-up when this is convenient, while the text remains screen-up.

Taking advantage of affordable yet high-powered computers, some ECDIS's now permit a split screen display, where mode of motion, orientation and scale are individually selectable on each panel. This permits, for example, a north-up small-scale overview in true motion alongside a course-up large-scale view in relative motion. Yet another display advantage occurs with zooming, in that symbols and text describing areas center themselves automatically in whatever part of the area appears on the screen. None of these functions are possible with raster charts.

The display operates by a set of rules, and data is arranged hierarchically, For example, where lines overlap, the less important line is not drawn. A more complex rule always places text at the same position relative to the object it applies to, no matter what else may be there. Since a long name or light description will often over-write another object, the only solution is to zoom in until the objects separate from each other. Note that because text causes so much clutter, and is seldom vital for safe navigation, it is written automatically when the object it refers to is on the display, but is an option under the "all other information" display level.

Flexibility in display scale requires some indication of distance to objects seen on the display. Some manufacturers use the rather restrictive but familiar radar range rings to provide this, while another uses a line symbol keyed to data's original scale. The ECDIS design also includes a one-mile scalebar at the side of the display, and an optionally displayed course and speed made good vector for own ship. There may be a heading line leading from the vessel's position indicating her future track for one minute, three minutes, or some other selectable time.

To provide the option of creating manual chart corrections, ECDIS includes a means of drawing lines, adding text and inserting stored objects on the display. These may be saved as user files, called up from a subdirectory, and edited on the display. Once loaded into the SENC, the objects may be selected or de-selected just as with other objects of the SENC.

Display options for ECDIS include transfer of ARPA-acquired targets and radar image overlay. IMO standards for ECDIS require that the operator be able to deselect the radar picture from the chart with a single operator action for fast "uncluttering" of the chart presentation.

1407. Units, Data Layers and Calculations

ECDIS uses the following units of measure:

- **Position:** Latitude and longitude will be shown in degrees, minutes, and decimal minutes, normally based on WGS-84 datum.
- **Depth:** Depths will be indicated in meters and decimeters.
- **Height:** Meters
- **Distance:** Nautical miles and tenths, or meters
- **Speed:** Knots and tenths

ECDIS requires data layers to establish a priority of data displayed. The minimum number of information categories required and their relative priority from highest to lowest are listed below:

- ECDIS warnings and messages
- Hydrographic office data
- *Notice to Mariner*s information
- Hydrographic office cautions
- Hydrographic office color-fill area data
- Hydrographic office on demand data
- Radar information
- User's data
- Manufacturer's data
- User's color-fill area data
- Manufacturer's color-fill area data

As a minimum, an ECDIS system must be able to perform the following calculations and conversions:

- Geographical coordinates to display coordinates, and display coordinates to geographical coordinates.
- Transformation from local datum to WGS-84.
- True distance and azimuth between two geographical positions.
- Geographic position from a known position given distance and azimuth.
- Projection calculations such as great circle and rhumb line courses and distances.

1408. Warnings and Alarms

Appendix 5 of the IMO Performance Standard specifies that ECDIS must monitor the status of its systems continuously, and must provide alarms and indications for certain functions if a condition occurs that requires immediate attention. Indications may be either visual or audible. An alarm must be audible and may be visual as well.

An alarm is required for the following:

- Exceeding cross-track limits
- Crossing selected safety contour
- Deviation from route
- Position system failure
- Approaching a critical point
- Chart on different geodetic datum from positioning system

An alarm or indication is required for the following:

- Largest scale for alarm (indicates that presently loaded chart is too small a scale to activate anti-grounding feature)
- Area with special conditions (means a special type of chart is within a time or distance setting)
- Malfunction of ECDIS (means the master unit in a master-backup network has failed)

An indication is required for the following:

- Chart overscale (zoomed in too close)
- Larger scale ENC available
- Different reference units (charted depths not in meters)
- Route crosses safety contour
- Route crosses specified area activated for alarms
- System test failure

As these lists reveal, ECDIS has been programmed to constantly "know" what the navigation team should know, and to help the team to apply its experience and judgment through the adjustment of operational settings.

This automation in ECDIS has two important consequences: First, route or track monitoring does not replace situational awareness; it only enhances it. The alarm functions, while useful, are partial and have the potential to be in error, misinterpreted, ignored, or overlooked.

Secondly, situational awareness must now include, especially when ECDIS is used as the primary means of navigation, the processes and status of the electronic components of the system. This includes all attached sensors, the serial connections and communication ports and data interfaces, the computer processor and operating system, navigation and chart software, data storage devices, and power supply. Furthermore, these new responsibilities must still be balanced with the traditional matters of keeping a vigilant navigational watch.

ECDIS or not, the windows in the pilothouse are still the best tool for situational awareness. Paradoxically, ECDIS makes the navigator's job both simpler and more complex.

1409. ECDIS Outputs

During the past 12 hours of the voyage, ECDIS must be able to reconstruct the navigation and verify the official database used. Recorded at one minute intervals, the information includes:

- Own ship's past track including time, position, heading, and speed
- A record of official ENC used including source, edition, date, cell and update history

It is important to note that if ECDIS is turned off, such as for chart management or through malfunction, voyage recording ceases, unless a networked backup system takes over the functions of the master ECDIS. In that case, the voyage recording will continue, including an entry in the electronic log for all the alarms that were activated and reset during the switchover. Voyage files consist of logbook files, track files and target files. The file structure is based on the date and is automatically created at midnight for the time reference in use. If the computer system time is used for that purpose, the possibility exists for overwriting voyage files if the system time is manually set back. Allowing GPS time as the system reference avoids this pitfall.

In addition, ECDIS must be able to record the complete track for the entire voyage with time marks at least once every four hours. ECDIS should also have the capability to preserve the record of the previous 12 hours of the voyage. It is a requirement that the recorded information be inaccessible to alteration. Preserving voyage files should follow procedures for archiving data. Unless radar overlay data is being recorded, voyage files tend to be relatively small, permitting backup onto low-capacity media, and purging from system memory at regular intervals. (This form of backing up should not be confused with the network master-slave backup system.)

Adequate backup arrangements must be provided to ensure safe navigation in case of ECDIS failure. This includes provisions to take over ECDIS functions so that an ECDIS failure does not develop into a critical situation, and a means of safe navigation for the remaining part of the voyage in case of complete failure.

1410. Voyage Data Recorder (VDR)

The purpose of the voyage data recorder VDR is to provide accurate historical navigational data in the investigation of maritime incidents. It is additionally useful for system performance monitoring. A certified VDR configuration records all data points, as per IMO Resolution A.861(20) & EC Directive 1999/35/EC. Some of the voyage data can be relayed through ECDIS. A fully IEC compliant data capsule passes fire and immersion tests.

The implementation of a secure "black box" and comprehensive Voyage Data Recorder (VDR) is now a carriage requirement on passenger and Ro-Ro vessels over 3000 GT (1600 GRT) engaged in international passages. Existing vessels must be retrofitted by July 2004, and all vessels built after July 2002 must be fitted with a VDR. Retrofit regulations for other vessels built before July 2002 are still in development. Non-RO-RO passenger vessels built before July 2002 may be exempted from carriage where an operator can show that interfacing a VDR with the existing equipment on the ship is unreasonable and impracticable. The European Union requires that all RO-RO ferries or high speed craft engaged on a regular service in European waters (domestic or international) be fitted with a VDR if built before February 2003, and otherwise retrofitted by July 2004.

VDR features include:

- Radar video capture: Radar video is captured and compressed every 15 seconds to comply with IEC performance standards.
- I/O subsystem: To collect a wide variety of data types, a sensor interface unit provides signal conditioning for all analog, digital and serial inputs. All data is converted and transmitted to a data acquisition unit via an ethernet LAN.
- Audio compression: An audio module collects analog signals from microphone preamplifiers. The data is digitized and compressed to meet Lloyds of London 24-hour voice storage requirements.
- Integral uninterruptible power supply (UPS) IEC requires a UPS backup for all components of the data acquisition unit and for the data capsule to provide two hours continuous recording following a blackout.
- Hardened fixed data capsule: IEC 61996 compliant data capsules fitted with ethernet connections provide fast download as well as fast upload to satellite links.
- Remote data recovery and shoreside playback: Options available in several systems.
- Annual system certification: The IMO requires that the VDR system, including all sensors, be subjected to an annual performance test for certification.

DATA FORMATS

1411. Official Vector Data

How ECDIS operates depends on what type of chart data is used. ENC's (electronic navigational charts) and RNC's (raster nautical charts) are approved for use in ECDIS. By definition both ENC's and RNC's are issued under the authority of national hydrographic offices (HO's). ECDIS functions as a true ECDIS when used with corrected ENC data, but ECDIS operates in the less functional raster chart display system (RCDS) mode when using corrected RNC data. When ECDIS is used with non-official vector chart data (corrected or not), it operates in the ECS mode.

In vector charts, hydrographic data is comprised of a series of files in which different layers of information are stored or displayed. This form of "intelligent" spatial data is obtained by digitizing information from existing paper charts or by storing a list of instructions that define various position-referenced features or objects (e.g., buoys, lighthouses, etc.). In displaying vector chart data on ECDIS, the user has considerable flexibility and discretion regarding the amount of information that is displayed.

An ENC is vector data conforming to the IHO S-57 ENC product specification in terms of content, structure and format. An ENC contains all the chart information necessary for safe navigation and may contain supplementary information in addition to that contained in the paper chart. In general, an S-57 ENC is a structurally layered data set designed for a range of hydrographic applications. As defined in IHO S-57 Edition 3, the data is comprised of a series of points, lines, areas, features, and objects. The minimum size of a data set is a cell, which is a spherical rectangle (i.e., bordered by meridians and latitudes). Adjacent cells do not overlap. The scale of the data contained in the cell is dependent upon the navigational purpose (e.g., general, coastal, approach, harbor).

Under S-57, cells have a standard format but do not have a standard coverage size. Instead, cells are limited to 5mb of data. S-57 cells are normally copy protected and therefore require a permit before use is allowed. These permits are delivered as either a file containing the chart permits or as a code. In both cases the first step is to install the chart permit into the ECDIS. Some hydrographic offices deliver S-57 cells without copy protection and therefore permits are not required.

Any regional agency responsible for collecting and distributing S-57 data, such as PRIMAR for Northern Europe, will also maintain data consistency. National hydrographic offices are responsible for producing S-57 data for their own country area. Throughout the world HO's have been slow to produce sufficient quantities of ENC data. This is due to the fact that the standards evolved over several years, and that vector data is much harder to collect than raster data.

In 1996 the IHO S-57 data standard and IHO S-52 specifications for chart content and display were "frozen." It took three versions of S-57 before the issue was finally settled as to what actually comprises an ENC (i.e., ENC Product Specification) and what is required for updating (ENC Updating Profile). The ENC Test Dataset that the International Electrotechnical Commission (IEC) requires for use in conjunction with IEC Publication 61174 (IEC 1997) was finalized by IHO in 1998. It was not possible to conduct ECDIS type-approval procedures without a complete and validated IHO ENC Test Dataset.

Major areas of ENC coverage now include most of Canadian and Japanese waters, the Baltic and North Sea, and important waterways such as the Straits of Malacca, Singapore Strait, and the Straits of Magellan (Chile).

At the same time, many countries including the United States, are stepping up their production of ENC's where issues of port security require the collection of baseline data of submerged hazards. In the U.S., NOAA plans to complete its portfolio of large-scale charts of 42 ports in ENC format by mid-2003, with smaller scale chart completion by 2005. As the chart cells are completed, the data is being made available on the World Wide Web at no cost. Beginning in 2003, NOAA will post critical notice to mariner corrections without restrictions in monthly increments. At that point the status of NOAA's available ENC data will be changed from provisional to official.

ENC data is currently available from the HO's of most Northern European countries, Japan, Korea, Hong Kong, Singapore, Canada, Chile, and the United States, although the coverage and updating process is incomplete. Most ENC is available only through purchase, permits or licensing.

1412. Vector Data Formats Other Than IHO S-57

The largest of the non-S-57 format databases is the Digital Nautical Chart (DNC). The National Imagery and Mapping Agency (NIMA) produces the content and format for the DNC according to a military specification. This allows compatibility among all U.S. Defense Department assets. The DNC is a vector-based digital product that portrays significant maritime features in a format suitable for computerized marine navigation. The DNC is a general-purpose global database designed to support marine navigation and Geographic Information System (GIS) applications. DNC data is only available to the U.S. military and selected allies. It is designed to conform to the IMO Performance Standard and IHO specifications for ECDIS.

Several commercial manufacturers have developed vector databases beyond those that have been issued by official hydrographic offices. These companies are typically manufacturers of ECDIS or ECS equipment or have direct relationships with companies that do, and typically have developed data in proprietary format in order to provide options to raster charts in the absence of ENC data. HO-issued paper charts provide the source data for these formats, although in some cases non-official paper charts are used. In some cases, ECS manufacturers provide a regular updating and maintenance service for their vector data, resulting in added confidence and satisfaction among users. The manufacturer's source of the updates is through HO's. Hence, these two particular non-official formats allows for a very high degree of confidence and satisfaction among mariners using this data.

ECS systems sometimes apply rules of presentation similar to officially specified rules. Thus information is displayed or removed automatically according to scale level to manage clutter. The same indications pertinent to overscaling ENC apply to private vector data. Since the chart data is not ENC, the systems must display that non-official status when used in an ECDIS.

1413. Raster Data

Raster navigational chart (RNC) data is stored as picture elements (pixels). Each pixel is a minute component of the chart image with a defined color and brightness level. Raster-scanned images are derived by scanning paper charts to produce a digital photograph of the chart. Raster data are far easier to produce than vector data, but raster charts present many limitations to the user.

The official raster chart formats are:

ARCS (British Admiralty)
Seafarer (Australia)
BSB (U.S., NOAA/Maptech)

These charts are produced from the same raster process used to print paper charts. They are accurate representations of the original paper chart with every pixel geographically referenced. Where applicable, horizontal datum shifts are included with each chart to enable referencing to WGS84. This permits compatibility with information overlaid on the chart. *Note: Not all available charts have WGS84 shift information*. Extreme caution is necessary if the datum shift cannot be determined exactly.

Raster nautical charts require significantly larger amounts of memory than vector charts. Whereas a world portfolio of more than 7500 vector charts may occupy about 500mb, a typical coastal region in raster format may consist of just 40 charts and occupy more than 1000mb of memory.

For practical purposes, most of a portfolio of raster charts should be left on the CD and not loaded into the ECDIS hard drive unless one is route planning or actually sailing in a given region. Of course, updates can only be performed on charts that are loaded onto the hard drive.

Certain non-official raster charts are produced that cover European and some South American waters. These are scanned from local paper charts. Additionally, some ECDIS and ECS manufacturers also produce raster charts in proprietary formats.

In 1998 the IMO's Maritime Safety Committee (MSC 70) adopted the Raster Chart Display System (RCDS) as Appendix 7 to the IMO Performance Standards. The IMO-IHO Harmonization Group on ECDIS (HGE) considered this issue for over three years. Where IHO S-57 Ed. 3 ENC data coverage is not available, raster data provided by official HO's can be used as an interim solution. But this RCDS mode does not have the full functionality of an otherwise IMO-compliant ECDIS using ENC data. Therefore, RCDS does not meet SOLAS requirements for carriage of paper charts, meaning that when ECDIS equipment is operated in the RCDS mode, it must be used together with an appropriate portfolio of corrected paper charts.

Some of the limitations of RCDS compared to ECDIS include:

- Chart features cannot be simplified or removed to suit a particular navigational circumstance or task.
- Orientation of the RCDS display to course-up may affect the readability of the chart text and symbols since these are fixed to the chart image in a north-up orientation.
- Depending on the source of the raster chart data, different colors may be used to show similar chart information, and there may be differences between colors used during day and night time.
- The accuracy of the raster chart data may be less than that of the position-fixing system being used.
- Unlike vector data, charted objects on raster charts do not support any underlying information.
- RNC data will not trigger automatic alarms. (However, some alarms can be generated by the RCDS from user-inserted information.).
- Soundings on raster charts may be in fathoms and feet, rather than meters.

The use of ECDIS in RCDS mode can only be considered as long as there is a backup folio of appropriate up-to-date paper charts.

INTEGRATED BRIDGE SYSTEMS

1414. Description

An Integrated Bridge System (IBS) is a combination of equipment and software which uses interconnected controls and displays to present a comprehensive suite of navigational information to the mariner. Rules from classification societies such as Det Norske Veritas (DNV) specify design criteria for bridge workstations. Their rules define tasks to be performed, and specify how and where equipment should be sited to enable those tasks to be performed. Equipment carriage requirements are specified for ships according to the requested class certification or notation. Publication IEC 61029 defines operational and performance requirements, methods of testing, and required test results for IBS.

Classification society rules address the total bridge system in four parts: technical system, human operator, man/machine interface, and operational procedures. The DNV classifies IBS with three certifications: NAUT-C covers bridge design; W1-OC covers bridge design, instrumentation and bridge procedures; W1 augments certain portions of W1-OC.

An IBS generally consists of at least:

- Dual ECDIS installation – one serving master and the other as backup and route planning station
- Dual radar/ARPA installation
- Conning display with a concentrated presentation of navigational information (the master ECDIS)
- DGPS positioning
- Ship's speed measuring system
- Auto-pilot and gyrocompass system
- Full GMDSS functionality

Some systems include full internal communications, and a means of monitoring fire control, shipboard status alarms, and machinery control. Additionally, functions for the loading and discharge of cargo may also be provided.

An IBS is designed to centralize the functions of monitoring collision and grounding risks, and to automate navigation and ship control. Control and display of component systems are not simply interconnected, but often share a proprietary language or code. Several instruments and indicators are considered essential for safe and efficient performance of tasks, and are easily readable at the navigation workstation, such as heading, rudder angle, depth, propeller speed or pitch, thruster azimuth and force, and speed and distance log.

Type approval by Det Norske Veritas for the DNV-W1-ANTS (Automatic Navigation and Track-Keeping

System) certification is given to ship bridge systems designed for one-man watch (W1) in an unbounded sea area. DNV also provides for the other two class notations, NAUT-C and W1-OC. The W1 specifications require the integration of:

- CDIS (providing the functions of safety-contour checks and alarms during voyage planning and execution)
- Manual and automatic steering system (including software for calculation, execution and adjustments to maintain a pre-planned route, and including rate of turn indicator)
- Automatic Navigation and Track-keeping System (ANTS)
- Conning information display
- Differential GPS (redundant)
- Gyrocompass (redundant)
- Radar (redundant) and ARPA
- Central alarm panel
- Wind measuring system
- Internal communications systems
- GMDSS
- Speed over ground (SOG) and speed through water (STW or Doppler log)
- Depth sounder (dual transducer >250m)
- Course alteration warnings and acknowledgment
- Provision to digitize paper charts for areas not covered by ENC data

The W1 classification requires that maneuvering information be made available on the bridge and presented as a pilot card, wheelhouse poster, and maneuvering booklet. The information should include characteristics of speed, stopping, turning, course change, low-speed steering, course stability, trials with the auxiliary maneuvering device, and man-overboard rescue maneuvers.

The W1-OC and W1 classifications specify responsibilities of ship owner and ship operator, qualifications, bridge procedures, and particular to W1, a requirement for operational safety standards. The W1 operational safety manual requires compliance with guidelines on bridge organization, navigational watch routines, operation and maintenance of navigational equipment, procedures for arrival and departure, navigational procedures for various conditions of confinement and visibility, and system fallback procedures. Both classifications also require compliance with a contingency and emergency manual, including organization, accident, security, evacuation, and other related issues.

MILITARY ECDIS

1415. ECDIS-N

In 1998, the U.S. Navy issued a policy letter for a naval version of ECDIS, ECDIS-N, and included a performance standard that not only conforms to the IMO Performance Standards, but extends it to meet unique requirements of the U.S. Department of Defense.

A major difference from an IMO-compliant ECDIS is the requirement that the ECDIS-N SENC must be the Digital Nautical Chart (DNC) issued by the National Imagery and Mapping Agency (NIMA). The DNC conforms to the U.S. DoD standard Vector Product Format (VPF), an implementation of the NATO DIGEST C Vector Relational Format. All of NIMA's nautical, aeronautical, and topographic vector databases are in VPF to ensure interoperability between DoD forces.

In the United States, NIMA produces the Digital Nautical Chart (DNC). It is a vector database of significant maritime features that can be used with shipboard integrated navigation systems such as ECDIS, ECDIS-N, or other types of geographic information systems. NIMA has been working closely with the U.S. Navy to help facilitate a transition from reliance on paper charts to electronic chart navigation using the DNC. The U.S. Navy plans to have all of its surface and sub-surface vessels using DNC's by 2004. NIMA has produced the DNC to support worldwide navigation requirements of the U.S. Navy and U.S. Coast Guard.

To ensure that the DNC data would not be manipulated or inadvertently altered when used by different military units, a decision was made to produce a specific data software product that must be used in a "direct read" capability. As such, a DNC is really a system electronic navigational chart (SENC) that contains specified data and display characteristics. Control of the SENC provides the military with interoperability across deployed systems, which is particularly important when integrated with military data layers.

1416. Navigation Sensor System Interface (NAVSSI)

The Navigation Sensor System Interface (NAVSSI) contains the U.S. Navy's version of ECDIS, and also has significant additional capabilities for the Navy's defense missions. NIMA's Vector Product Format (VPF) DNC's are used in conjunction with NAVSSI. NAVSSI performs three important functions:

- Navigation Safety: NAVSSI distributes real time navigation data to the navigation team members to ensure navigation safety.
- Weapons System Support: NAVSSI provides initial-

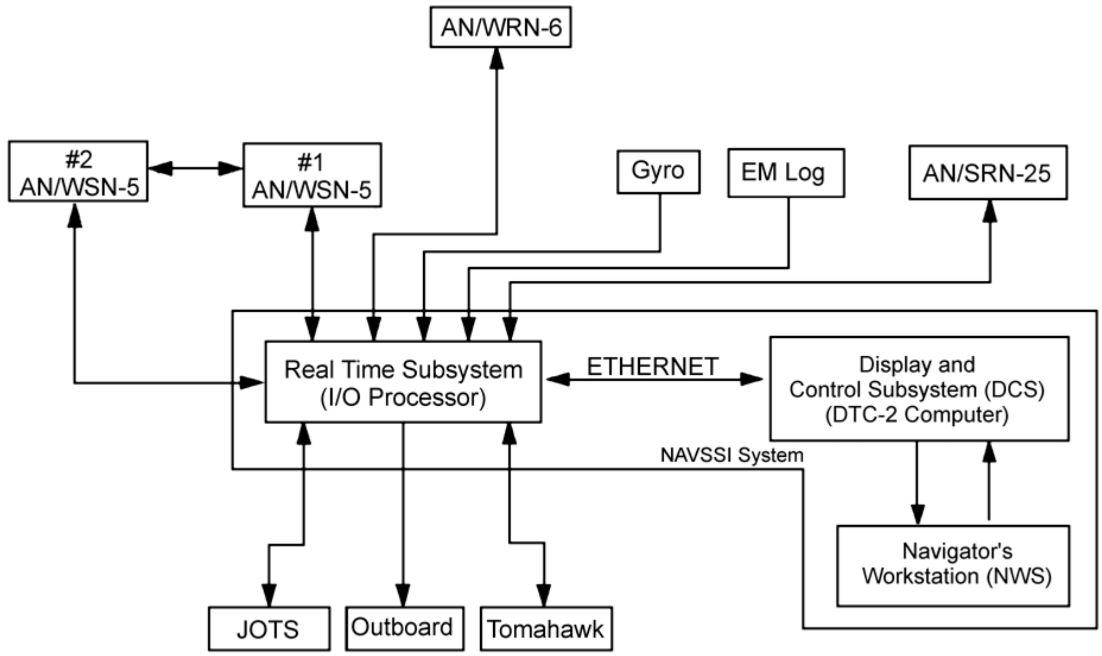


Figure 1416. Block diagram of NAVSSI.

ization data for weapons systems.

- Battlegroup Planning: NAVSSI provides a workstation for battlegroup planning.

The navigational function of NAVSSI, therefore, is only one of several tasks accomplished by the system. The navigational portion of NAVSSI complies with the IMO/IHO ECDIS standards for content and function.

The heart of NAVSSI is the Real Time Subsystem (RTS). The RTS receives, processes and distributes navigational data to the navigation display, weapons systems, and other networked vessels. This ensures that all elements of a battlegroup have the same navigational picture. Inputs come from GPS, Loran, inertial navigation systems, compass, and speed log. The bridge display consists of a monitor and control panel, while the RTS is mounted below decks. DNC's are contained in the **Display and Control Subsystem (DCS)** typically mounted in the chartroom with a monitor on the bridge. This is unlike many current commercial systems which house all hardware and software in a single unit on the bridge. A separate NAVSSI software package supports operator interface, waypoint capability, collision and grounding avoidance features, and other aspects of an ECDIS.

Figure 1416 illustrates a basic block diagram of the NAVSSI system. The RTS takes inputs from the inertial navigators, the GPS in PPS mode, the compass, the EM Log, and the SRN-25. The RTS distributes navigation information to the various tactical applications requiring navigation input, and it communicates via fiber optic network with the DCS. The DCS exchanges information with the Navigator's Workstation.

1417. The Digital Nautical Chart

NAVSSI uses the Digital Nautical Chart (DNC) as its chart database. The DNC is in Vector Product Format (VPF) and is based on the contents of the traditional paper harbor, approach, coastal and general charts produced by NIMA and NOS.

Horizontal datum is WGS 84 (NAD 83 in the U. S. is equivalent). There are three vertical datums. Topographic features are referenced to Mean Sea Level and the shore line is referenced to Mean High Water. Hydrography is referenced to a low water level suitable for the region. All measurements are metric.

The DNC portfolio consists of 29 CD-ROM's and provides global coverage between 84 degrees N and 81 degrees S. This comprises some 4,820 charts group into five libraries based on scale:

General: (>1:500K)
Coastal: (1:75K - 1: 500K)
Approach (1:25K - 1:75K)
Harbor (1 <1:50K)
Browse Index (1:3,100,000)

DNC data is layered together into 12 related feature classes:

- Cultural Landmarks
- Earth Cover
- Inland Waterways
- Relief
- Landcover
- Port Facilities
- Aids to Navigation
- Obstructions
- Hydrography
- Environment
- Limits
- Data Quality

Content is generally the same as on a paper chart. The data is stored in libraries; each library represents a different level of detail. The libraries are then stored on CD-ROM and organized as tiles according to the World Geodetic Reference System (GEOREF) tiling scheme.

A subset of the DNC is known as Tactical Ocean Data (TOD). TOD data is bathymetric in nature and intended for Naval operations.

There are 6 levels of TOD:

Level 0 - OPAREA charts
Level 1 - Bottom Contour
Level 2 - Bathymetric Navigation Planning Charts
Level 3 - Shallow Water
Level 4 - Hull Integrity Test Charts
Level 5 - Strategic Straits Charts

1418. Warship ECDIS (WECDIS)

A Warship ECDIS is an ECDIS approved by international authorities for warship use, which, while meeting the operating standards of ECDIS, may not conform exactly to ECDIS specifications.

Performance Standards for “Warship” ECDIS (WECDIS) were approved by the North Atlantic Treaty Organization (NATO) in 1999 and issued as STANAG 4564. The core functionality of WECDIS is an IMO-compliant ECDIS. Beyond the minimum performance requirements for ECDIS, WECDIS has the ability to use a variety of geospatial data from both civilian and military sources. For navigational data, WECDIS uses both IHO S-57 ENC data and data conforming to NATO Digital Geographic Information Exchange (DIGEST) Standards. This latter includes such products as Vector Product Format (VPF) and Digital Nautical Chart (DNC).

In addition to core navigation information (IHO S-57 ENC and VPF-DNC), WECDIS will also use Additional Navigation Information (ANI) provided by government hydrographic offices and military sources. Specific types of ANI data include Raster Navigational Charts (RNC’s), such as Admiralty Raster Chart Service (ARCS) or NOAA’s raster charts distributed and updated by Maptech, Inc. The ability to use different types of navigational data from a variety sources is often referred to as “multi-fuel.”

CORRECTING ELECTRONIC CHARTS

1419. ECDIS Correction Systems

ECDIS software creates a database from the ENC data called the system electronic navigational chart (SENC) and from this selects information for display. The ECDIS software meanwhile receives and processes serial data from navigational sensors and displays that textual and graphical information simultaneously with the SENC information.

It is the SENC that is equivalent to up-to-date charts, as stated by the Performance Standards. As originally conceived, ECDIS was designed to use internationally standardized and officially produced vector data called the ENC (electronic navigational chart). Only when using ENC data can ECDIS create an SENC, and thereby function in the ECDIS mode.

Updates for ENC are installed into the ECDIS separate from the ENC data itself. For the mariner, this involves activating a special utility accompanying the ECDIS and following the on-screen prompts. Within this same utility, update content and update log files in textual form can be viewed. Once the ECDIS software itself is reactivated, the update information is accessed in conjunction with the ENC data and the SENC database is created.

Just as ENC and updates are transformed into the SENC, so too are other data types accessed and combined. The user has the option to add lines, objects, text and links to other files supported by application. Referred to in the Performance Standards as data added by the mariner, these notes function as layers on the displayed chart. The user can select all or parts of the layers for display to keep clutter to a minimum. The mariner’s own layers, however, must be called into the SENC from stored memory. As a practical matter, not only must the mariner take care to associate file names with actual content, such as with manually created chart corrections, but also must realize that the files themselves do not have the tamper-proof status that ENC and official updates have.

Within the SENC resides all the information available for the display. The Presentation Library rules such as Standard Display and Display Base define what levels of information from the SENC can be shown.

An ENC updating profile is contained within the IHO S-57 Edition 3.0 specification. This enables the efficient addition, removal or replacement of any line, feature, object or area contained within the ENC dataset. Guidance on the means and process for ENC updating is provided in IHO S-52, Appendix 1. In terms of what is called for in the IMO Performance Standards, an ENC dataset being used in an ECDIS must also have an ENC updating service. This permits the ENC and the SENC to be corrected for the intended voyage, and thus achieves an important component of SOLAS compliance.

Accordingly, ECDIS must be capable of accepting official updates to the ENC data provided in conformity with IHO standard. Updated cells are stored in a file and transmitted by e-mail, floppy disk or CD-ROM, or satellite. For example, PRIMAR charts and updates are delivered on two CD's: the Base CD contains the PRIMAR database at the time indicated on the label and the second CD contains the updates for those charts. But the update CD also contains new charts issued since the base CD was printed. Since the operator must acquire the files and then initiate the update functions of the ECDIS software, this form of updating is referred to as semi-automatic.

Generally, ECDIS will reject updates if the update issuing authority is different from the cell issuing authority. It will also reject corrupted update files and files with an incorrect extension. ECDIS checks that updates are applied in the right sequence. If one update is missing the next update is rejected. An update CD-ROM should contain all available updates for all S57 cells. Generally, ECDIS will automatically run all updates in the right order for all cells.

For S-57 data, the content of updates in text form can be viewed from within the utility that permits the management of chart data. Generally it can only be run when ECDIS is terminated. ECDIS is also capable of showing or hiding S-57 updates on a given chart or cell. The update must first be installed via the chart utility. After restarting ECDIS, and after loading into the display the particular chart with the correction, the correction should be manually accepted. That enables the function in S-57 chart options to show or hide the symbol indicating the location of the correction.

NIMA DNC Corrections

NIMA has produced the DNC Vector Product Format Database Update (VDU) to support worldwide DNC navigation requirements of the U.S. Navy, the U.S. Coast Guard, and certain allies. NIMA does not distribute DNC to other than U.S. government agencies and foreign governments having data exchange agreements with NIMA. The DNC maintenance system will be able to apply new source materials such as bathymetry, imagery, *Notice to Mariners*, local notices, new foreign charts, etc. for inclusion in the DNC database.

The VDU system works by performing a binary comparison of the corrected chart with the previous version. The differences are then written to a binary "patch" file with instructions as to its exact location. The user then applies this patch by specifying the proper path and filename on his own ship. Every new change incorporates all previous changes, so the navigator is assured that, having received the latest change, he has all changes issued to date.

File sizes are small enough to support bandwidth limitations of ships at sea and requires only one-way communication. Patch files are posted every four weeks. Authorized commands may access DNC's and the associated VDU files through the NIMA Gateway:

OSIS http://osis.nima.mil/gidbe/index.htm
SIPRNET http://www.nima.smil.mil/products/dnc1
JWICS http://www.nima.is.gov/products/dnc1

The VDU patch files are posted to the World Wide Web monthly at:

http://www.nima.mil/dncpublic/

A separate layer within DNC provides the user with identification of where changes have been made during the updating process.

British RCS Corrections

For the British RCS system, updates for all 2700 charts affected by *Admiralty Notice to Mariners* are compiled and placed on a weekly ARCS Update CD-ROM. Applying the corrections is only semi-automatic (not fully automatic), but it is also error-free, and each CD-ROM provides cumulative updates. The CD-ROM's are available through chart agents.

NOAA Corrections

In the U.S., NOAA has contracted with Maptech, Inc. to provide updating of all NOS raster charts using information from the USCG, NIMA and the Canadian Hydrographic Service (CHS). Maptech uses a "patch technique" to update only those parts of a given chart identified as needing correction. The method compares the existing chart file and its corrected counterpart on a pixel-by-pixel basis. The software creates a "difference file" that is associated with the existing raster file to which it applies. This difference file is then compressed so that a typical patch contains only a few kilobytes of data. Ninety-nine percent are under 10kb. Typical downloads for a chart take 15 seconds to 5 minutes depending on modem speed.

The raster chart is updated as the patch file alters the pixels on the original chart. Update patches are available by download, and are cumulative for the all the charts packed on a given source folio CD. Further refinement will permit the separate storage of the RNC and update patches, so that as the patch is applied dynamically in real time, the user will be able to view the correction. The dynamic patching is similar to ENC updating in that the original chart data is not altered. Presently the service is a subscription service with weekly updates at a nominal cost. Information is available at http://chartmaker.ncd.noaa.gov.

Commercial Systems

There are a variety of ECS's available for small craft, often found aboard fishing vessels, tugs, research vessels, yachts, and other craft not large enough to need SOLAS equipment but wanting the best in navigation technology. Given that these systems comprise a single aid to navigation and do not represent a legal chart in any sense, it is probably not a critical point that correction systems for these products are not robust enough to support regular application of changes.

In fact, often the only way to make changes is to purchase new editions, although the more sophisticated ones allow the placement of electronic "notes" on the chart. The data is commonly stored on RAM chips of various types, and cannot be changed or without re-programming the chip from a CD-ROM or disk containing the data. If the data is on CD-ROM, a new CD-ROM is the update mechanism, and they are, for the most part, infrequently produced. Users of these systems are required to maintain a plot on a corrected paper chart.

USING ELECTRONIC CHARTS

1420. Digital Chart Accuracy

As is the case with any shipboard gear, the user must be aware of the capabilities and limitations of digital charts. The mariner should understand that nautical chart data displayed possess inherent accuracy limitations. Because digital charts are necessarily based primarily on paper charts, many of these limitations have migrated from the paper chart into the electronic chart. Electronic chart accuracy is, for the most part, dependent on the accuracy of the features being displayed and manipulated. While some ECDIS and ECS have the capability to use large-scale data produced from recent hydrographic survey operations (e.g., dredged channel limits or pier/terminal facilities) most raster and vector-based electronic chart data are derived from existing paper charts.

Twenty years ago, mariners were typically obtaining position fixes using radar ranges, visual bearings or Loran. Generally, these positioning methods were an order of magnitude less accurate than the horizontal accuracy of the survey information portrayed on the chart. For example, a three-line fix that results in an equilateral triangle with sides two millimeters in length at a chart scale of 1:20,000 represents a triangle with 40-meter sides in real-world coordinates.

A potential source of error is related to the system configuration, rather than the accuracy of electronic chart data being used. All ECDIS's and most ECS's enable the user to input the vessel's dimensions and GPS antenna location. On larger vessels, the relative position of the GPS antenna aboard the ship can be a source of error when viewing the "own-ship" icon next to a pier or wharf.

In U.S. waters, the Coast Guard's DGPS provides a horizontal accuracy of +/-10 meters (95 percent). However, with selective availability off, even the most basic GPS receiver in a non-differential mode may be capable of providing better than 10 meter horizontal accuracy. In actual operation, accuracies of 3-5 meters are being achieved. As a result, some mariners have reported that when using an electronic chart while moored alongside a pier, the vessel icon plots on top of the pier or out in the channel.

Similarly, some mariners transiting a range that marks the centerline of a channel report that the vessel icon plots along the edge or even outside of the channel. Mariners now expect, just as they did 20 years ago, that the horizontal accuracy of their charts will be as accurate as the positioning system available to them. Unfortunately, any electronic chart based on a paper chart, whether it is raster or vector, is not able to meet this expectation.

The overall horizontal accuracy of data portrayed on paper charts is a combination of the accuracy of the underlying source data and the accuracy of the chart compilation process. Most paper charts are generalized composite documents compiled from survey data that have been collected by various sources over a long period of time. A given chart might encompass one area that is based on a lead line and sextant hydrographic survey conducted in 1890, while another area of the same chart might have been surveyed in the year 2000 with a full-coverage shallow-water multibeam system. In the U.S., agencies have typically used the most accurate hydrographic survey instrumentation available at the time of the survey.

While survey positioning methods have changed over the years, standards have generally been such that surveys were conducted with a positioning accuracy of better than 0.75 millimeters at the scale of the chart. Therefore, on a 1:20,000-scale chart, the survey data was required to be accurate to 15 meters. Features whose positions originate in the local notice to mariners, reported by unknown source, are usually charted with qualifying notations like position approximate (PA) or position doubtful (PD). The charted positions of these features, if they do exist, may be in error by miles.

As of 2002, over 50 percent of the depth information found on U.S. charts is based on hydrographic surveys conducted before 1940. Surveys conducted many years ago with lead lines or single-beam echo sounders sampled only a tiny percentage of the ocean bottom. Hydrographers were unable to collect data between the sounding lines. Depending on the water depth, these lines may have been spaced at

50, 100, 200 or 400 meters. As areas are re-surveyed and full-bottom coverage is obtained, uncharted features, some dangerous to navigation, are discovered quite often. These features were either: 1) not detected on prior surveys, 2) objects such as wrecks that have appeared on the ocean bottom since the prior survey or 3) the result of natural changes that have occurred since the prior survey.

In a similar manner, the shoreline found on most U.S. charts is based on photogrammetric or plane table surveys that are more than 20 years old. In major commercial harbors, the waterfront is constantly changing. New piers, wharves, and docks are constructed and old facilities are demolished. Some of these man-made changes are added to the chart when the responsible authority provides as-built drawings. However, many changes are not reported and therefore do not appear on the chart. Natural erosion along the shoreline, shifting sand bars and spits, and geological subsidence and uplift also tend to render the charted shoreline inaccurate over time.

Another component of horizontal chart accuracy involves the chart compilation process. For example, in the U.S. before NOAA's suite of charts was scanned into raster format in 1994, all chart compilation was performed manually. Projection lines were constructed and drawn by hand and all plotting was done relative to these lines. Cartographers graphically reduced large scale surveys or engineering drawings to chart scale. Very often these drawings were referenced to state plane or other local coordinate systems. The data would then be converted to the horizontal datum of the chart (e.g., the North American 1927 (NAD 27) or the North American Datum 1983 (NAD 83). In the late 1980's and early 1990's, NOAA converted all of its charts to NAD 83. In accomplishing this task, averaging techniques were used and all of the projection lines were redrawn.

When NOAA scanned its charts and moved its cartographic production into a computer environment, variations were noted between manually constructed projection lines and those that were computer generated. All of the raster charts were adjusted or warped so that the manual projection lines conformed to the computer-generated projection. In doing so, all information displayed on the chart was moved or adjusted.

Similar processes take place during NIMA’s digital chart production, but involving more complexity, since NIMA cartographers must work with a variety of different datums in use throughout the world, and with hydrographic data from hundreds of official and unofficial sources. While much of NIMA’s incoming data was collected to IHO standards during hydrographic surveys, many sources are questionable at best, especially among the older data.

Today, when survey crews and contractors obtain DGPS positions on prominent shoreline features and compare those positions to the chart, biases may be found that are on the order of two millimeters at the scale of the chart (e.g., 20 meters on 1:10,000-scale chart). High accuracy aerial photography reveals similar discrepancies between the true shoreline and the charted shoreline. It stands to reason that other important features such as dredged channel limits and navigational aids also exhibit these types of biases. Unfortunately, on any given chart, the magnitude and the direction of these discrepancies will vary by unknown amounts in different areas of the chart. Therefore, no systematic adjustment can easily be performed that will improve the inherent accuracy of the paper or electronic chart.

Some mariners have the misconception that because charts can be viewed on a computer, the information has somehow become more accurate than it appears on paper. Some mariners believe that vector data is more accurate than paper or raster data. Clearly, if an electronic chart database is built by digitizing a paper chart, it can be no more accurate than the paper chart.

The most accurate way to create an ENC is to re-compile the chart from all of the original source material. Unfortunately, the process is far too labor intensive. In the U.S., NOAA has used original source material where possible to compile navigation critical information such as aids to navigation and channel limits. The remaining data are being digitized from the largest scale paper charts.

Once ENC’s are compiled, they may be enhanced with higher-accuracy data over time. High-resolution shoreline data may be incorporated into the ENC’s as new photogrammetric surveys are conducted. Likewise, depths from new hydrographic surveys will gradually supersede depths that originated from old surveys.

1421. Route Planning and Monitoring

Presumably, route planning takes place before the voyage begins, except in situations where major changes in the route are called for while the ship is underway. In either case, both ECDIS and ECS will allow the display of the smallest scale charts of the operating area and the selection of waypoints from those charts. ECDIS requires a warning that a chosen route crosses a safety contour or prohibited area; ECS will not necessarily do so. If the data is raster, this function is not possible. Once the waypoints are chosen, they can be saved as a route in a separate file for later reference and output to the autopilot.

It is a good idea to zoom in on each waypoint if the chart scale from which it is selected is very small, so that the navigational picture in the area can be seen at a reasonable scale. Also, if a great circle route is involved, the software may be able to enter the waypoints directly from the great circle route file. If not, they will have to be entered by hand.

During route monitoring, ECDIS must show own ship’s position whenever the display covers that area. Although the navigator may chose to “look-ahead” while in route monitoring, it must be possible to return to own ship’s position with a single operator action. Key information provided during route monitoring includes a continuous indication of vessel position, course, and speed. Additional information that ECDIS or ECS can provide includes distance right/left of intended track, time-to-turn, distance-to-turn, position and time of "wheel-over", and past track history.

As specified in Appendix 5 of the IMO Performance Standard, ECDIS must provide an indication of the condition of the system and its components. An alarm must be provided if there is a condition that requires immediate attention. An indication can be visual, while an alarm must either be audible, or both audible and visual.

The operator can control certain settings and functions, some of the most important of which are the parameters for certain alarms and indications, including:

- Cross-track error: Set the distance to either side of the track the vessel can stray before an alarm sounds. This will depend on the phase of navigation, weather, and traffic.
- Safety contour: Set the depth contour line which will alert the navigator that the vessel is approaching shallow water.
- Course deviation: Set the number of degrees off course the vessel's heading should be allowed to stray before an alarm sounds.
- Critical point approach: Set the distance before approaching each waypoint or other critical point that an alarm will sound.
- Datum: Set the datum of the positioning system to the datum of the chart, if different.

1422. Waypoints and Routes

In the route planning mode, the ECS or ECDIS will allow the entry of waypoints as coordinates of latitude and longitude, or the selection of waypoints by moving a cursor around on the charts. It will allow the creation and storage of numerous pre-defined routes, which can be combined in various ways to create complex voyages.

For example, one might define a route from the inner harbor to the outer harbor of a major port, a route for each of two or more channels to the sea, and several more for open sea routes to different destinations. These can then be combined in different ways to create comprehensive routes that will comprise entire dock-to-dock voyages. They may also be run in reverse for the return trip.

When selecting waypoints, take care to leave any aids to navigation marking the route well to one side of the course. Many navigational software programs contain databases listing the location of the aids to navigation in the United States and other countries. This list should NOT be used to create routes, because the accuracy of today's navigation systems is good enough that to do so invites a collision with any aid whose actual position is entered as a waypoint. Always leave a prudent amount of room between the waypoint and the aid.

Some published routes exist, also a feature of certain software programs. The wise navigator will not use these until he has verified the exact position of each waypoint using the best scale chart. Using pre-programmed routes from an unknown source is the same as letting someone else navigator your vessel. Such a route may pass over shoal water, under a bridge, or through an area that your own vessel might find hazardous. Always check each waypoint personally.

Many electronic chart systems will also allow the coupling of the navigation system to the autopilot. Technically, it is possible to turn the navigation of the vessel over to the autopilot almost as soon as the vessel is underway, allowing the autopilot to make the course changes according to each waypoint. While this may be possible for small craft in most inland, harbor and harbor approach situations, the larger the vessel, the less advisable this practice is, because autopilots do not take advance and transfer into account. The large ship under autopilot control will not anticipate the turn in a channel, and will not begin the turn until the antenna of the positioning system, presumably GPS and often located in the stern of the ship, is at the exact waypoint. By this time it is too late, for the turn should likely have been started at least two ship lengths previous. It is perfectly prudent to allow autopilot control of course changes for vessels in the open sea if the proper parameters for maximum rudder angle have been set.

1423. Training and Simulation

In 2001, the IMO issued guidelines for training with ECDIS simulation. The guidelines stipulate that ECDIS training should include simulation of live data streams, as well as ARPA and Automated Information System (AIS) target information, and a Voyage Data Recorder (VDR) interface. But the IMO has not specifically required ECDIS training other than as a general substitution in the Standards of Training, Certification, and Watchkeeping (STCW) 95 code for navigation with paper charts.

Also in 2001, the USCG approved the country's first STCW-compliant five day ECDIS training course in the U.S. Long-term STCW 95 training and education programs are presently in development. The two levels of competency defined by STCW are operational (OIC or 3rd mate / 2nd mate) and management (CCM or 1st officer / Master). It is likely that for mariners sailing since August 1998, training and education in navigation at both the OIC and CCM levels will include the five day competency-based ECDIS training course.

Accordingly, certified training in the operational use of ECDIS should consist of a five day course making use of simulation equipment for a real-time operating environment appropriate for tasks in navigation, watchkeeping and maneuvering. The primary goal is that the trainee should be able to smoothly operate the ECDIS equipment, use all of its navigational functions, select and assess all relevant information, respond correctly in the case of a malfunction, describe common errors of interpretation and describe potential errors of displayed data. The trainee should follow structured practice in the following: setting up and maintaining the display; operational use of electronic charts including updating, route monitoring, route planning, handling alarms; work with motion parameters and position correction; work with log records and voyage files; and operate interfaces with radar, ARPA, AIS transponders, and VDR's.

CHAPTER 15

NAVIGATIONAL ASTRONOMY

PRELIMINARY CONSIDERATIONS

1500. Definitions

The science of Astronomy studies the positions and motions of celestial bodies and seeks to understand and explain their physical properties. Navigational astronomy deals with their coordinates, time, and motions. The symbols commonly recognized in navigational astronomy are given in Table 1500.

Celestial Bodies

☉	Sun	☉̲ ☾̲	Lower limb
☾	Moon	⊖ ☾	Center
☿	Mercury	☉̄ ☾̄	Upper limb
♀	Venus	●	New moon
⊕	Earth	◐	Crescent moon
♂	Mars	◐	First quarter
♃	Jupiter	◐	Gibbous moon
♄	Saturn	○	Full moon
⛢	Uranus	◑	Gibbous moon
♆	Neptune	◑	Last quarter
♇	Pluto	◑	Crescent moon
☆	Star		
☆-P	Star-planet altitude correction (altitude)		

Miscellaneous Symbols

y	Years	✳	Interpolation impractical
m	Months	°	Degrees
d	Days	′	Minutes of arc
h	Hours	″	Seconds of arc
m	Minutes of time	☌	Conjunction
s	Seconds of time	☍	Opposition
▬	Remains below horizon	□	Quadrature
▭	Remains above horizon	☊	Ascending node
////	Twilight all night	☋	Descending node

Signs of the Zodiac

♈	Aries (vernal equinox)	♎	Libra (autumnal equinox)
♉	Taurus	♏	Scorpius
♊	Gemini	♐	Sagittarius
♋	Cancer (summer solstice)	♑	Capricornus (winter solstice)
♌	Leo	♒	Aquarius
♍	Virgo	♓	Pisces

Table 1500. Astronomical symbols.

1501. The Celestial Sphere

Looking at the sky on a dark night, imagine that celestial bodies are located on the inner surface of a vast, Earth-centered sphere (Figure 1501). This model is useful since we are only interested in the relative positions and motions of celestial bodies on this imaginary surface. Understanding the concept of the celestial sphere is most important when discussing sight reduction in Chapter 20.

1502. Relative and Apparent Motion

Celestial bodies are in constant motion. There is no fixed position in space from which one can observe absolute motion. Since all motion is relative, the position of the observer must be noted when discussing planetary motion. From the Earth we see apparent motions of celestial bodies on the celestial sphere. In considering how planets follow their orbits around the Sun, we assume a hypothetical observer at some distant point in space. When discussing the rising or setting of a body on a local horizon, we must locate the observer at a particular point on the Earth because the setting Sun for one observer may be the rising Sun for another.

Motion on the celestial sphere results from the motions in space of both the celestial body and the Earth. Without special instruments, motions toward and away from the Earth cannot be discerned.

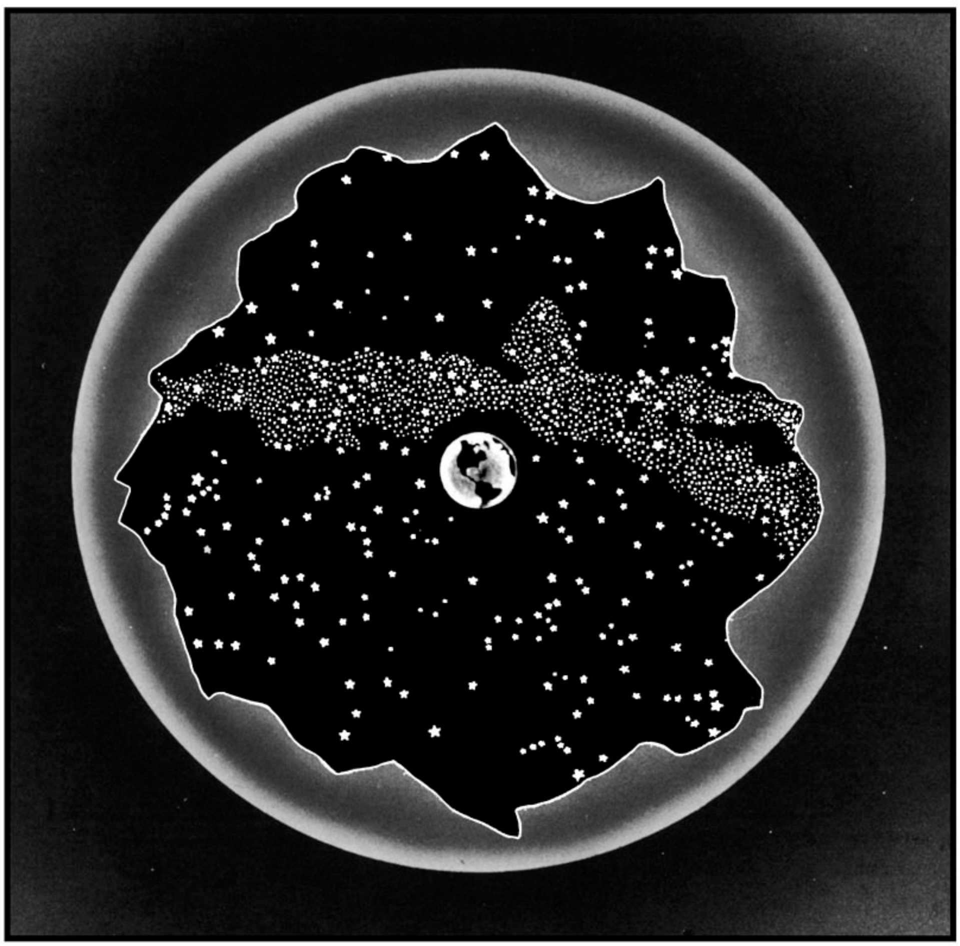


Figure 1501. The celestial sphere.

1503. Astronomical Distances

We can consider the celestial sphere as having an infinite radius because distances between celestial bodies are so vast. For an example in scale, if the Earth were represented by a ball one inch in diameter, the Moon would be a ball one-fourth inch in diameter at a distance of 30 inches, the Sun would be a ball nine feet in diameter at a distance of nearly a fifth of a mile, and Pluto would be a ball half an inch in diameter at a distance of about seven miles. The nearest star would be one-fifth of the actual distance to the Moon.

Because of the size of celestial distances, it is inconvenient to measure them in common units such as the mile or kilometer. The mean distance to our nearest neighbor, the Moon, is 238,855 miles. For convenience this distance is sometimes expressed in units of the equatorial radius of the Earth: 60.27 Earth radii.

Distances between the planets are usually expressed in terms of the **astronomical unit (AU)**, the mean distance between the Earth and the Sun. This is approximately 92,960,000 miles. Thus the mean distance of the Earth from the Sun is 1 AU. The mean distance of Pluto, the outermost known planet in our solar system, is 39.5 A.U. Expressed in astronomical units, the mean distance from the Earth to the Moon is 0.00257 A.U.

Distances to the stars require another leap in units. A commonly-used unit is the **light-year**, the distance light travels in one year. Since the speed of light is about 1.86×10^5 miles per second and there are about 3.16×10^7 seconds per year, the length of one light-year is about 5.88×10^{12} miles. The nearest stars, Alpha Centauri and its neighbor Proxima, are 4.3 light-years away. Relatively few stars are less than 100 light-years away. The nearest galaxies, the Clouds of Magellan, are 150,000 to 200,000 light years away. The most distant galaxies observed by astronomers are several billion light years away.

1504. Magnitude

The relative brightness of celestial bodies is indicated by a scale of stellar **magnitudes**. Initially, astronomers divided the stars into 6 groups according to brightness. The 20 brightest were classified as of the first magnitude, and the dimmest were of the sixth magnitude. In modern times, when it became desirable to define more precisely the limits of magnitude, a first magnitude star was considered 100 times brighter than one of the sixth magnitude. Since the fifth root of 100 is 2.512, this number is considered the **magnitude ratio**. A first magnitude star is 2.512 times as bright as a second magnitude star, which is 2.512 times as bright as a third magnitude star,. A second magnitude is $2.512 \times 2.512 = 6.310$ times as bright as a fourth magnitude star. A first magnitude star is 2.512^{20} times as bright as a star of the 21st magnitude, the dimmest that can be seen through a 200-inch telescope.

Brightness is normally tabulated to the nearest 0.1 magnitude, about the smallest change that can be detected by the unaided eye of a trained observer. All stars of magnitude 1.50 or brighter are popularly called “first magnitude” stars. Those between 1.51 and 2.50 are called “second magnitude” stars, those between 2.51 and 3.50 are called “third magnitude” stars, etc. Sirius, the brightest star, has a magnitude of –1.6. The only other star with a negative magnitude is Canopus, –0.9. At greatest brilliance Venus has a magnitude of about –4.4. Mars, Jupiter, and Saturn are sometimes of negative magnitude. The full Moon has a magnitude of about –12.6, but varies somewhat. The magnitude of the Sun is about –26.7.

THE UNIVERSE

1505. The Solar System

The **Sun**, the most conspicuous celestial object in the sky, is the central body of the solar system. Associated with it are at least nine principal **planets** and thousands of asteroids, comets, and meteors. Some planets have moons.

1506. Motions of Bodies of the Solar System

Astronomers distinguish between two principal motions of celestial bodies. Rotation is a spinning motion about an axis within the body, whereas revolution is the motion of a body in its orbit around another body. The body around which a celestial object revolves is known as that body’s primary. For the satellites, the primary is a planet. For the planets and other bodies of the solar system, the primary is the Sun. The entire solar system is held together by the gravitational force of the Sun. The whole system revolves around the center of the Milky Way galaxy (Article 1515), and the Milky Way is in motion relative to its neighboring galaxies.

The hierarchies of motions in the universe are caused by the force of gravity. As a result of gravity, bodies attract each other in proportion to their masses and to the inverse square of the distances between them. This force causes the planets to go around the sun in nearly circular, elliptical orbits.

In each planet’s orbit, the point nearest the Sun is called the **perihelion**. The point farthest from the Sun is called the **aphelion**. The line joining perihelion and aphelion is called the **line of apsides**. In the orbit of the Moon, the point nearest the Earth is called the **perigee**, and that point farthest from the Earth is called the **apogee**. Figure 1506 shows the orbit of the Earth (with exaggerated eccentricity), and the orbit of the Moon around the Earth.

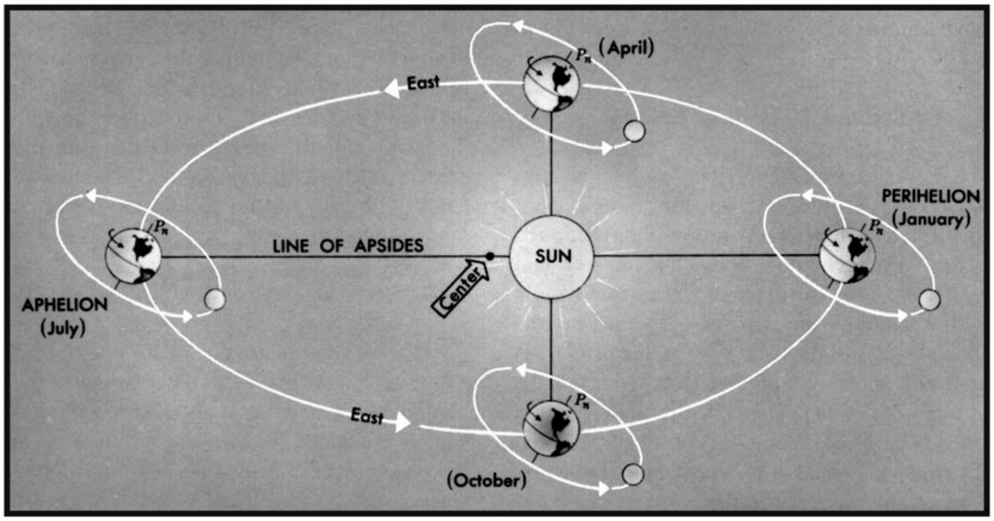


Figure 1506. Orbits of the Earth and Moon.

1507. The Sun

The Sun dominates our solar system. Its mass is nearly a thousand times that of all other bodies of the solar system combined. Its diameter is about 865,000 miles. Since it is a star, it generates its own energy through a thermonuclear reaction, thereby providing heat and light for the entire solar system.

The distance from the Earth to the Sun varies from 91,300,000 at perihelion to 94,500,000 miles at aphelion. When the Earth is at perihelion, which always occurs early in January, the Sun appears largest, 32.6' of arc in diameter. Six months later at aphelion, the Sun's apparent diameter is a minimum of 31.5'.

Observations of the Sun's surface (called the **photosphere**) reveal small dark areas called **sunspots**. These are areas of intense magnetic fields in which relatively cool gas (at 7000°F.) appears dark in contrast to the surrounding hotter gas (10,000°F.). Sunspots vary in size from perhaps 50,000 miles in diameter to the smallest spots that can be detected (a few hundred miles in diameter). They generally appear in groups. See Figure 1507. Large sunspots can be seen without a telescope if the eyes are protected.

Surrounding the photosphere is an outer **corona** of very hot but tenuous gas. This can only be seen during an eclipse of the Sun, when the Moon blocks the light of the photosphere.

The Sun is continuously emitting charged particles, which form the **solar wind**. As the solar wind sweeps past the Earth, these particles interact with the Earth's magnetic field. If the solar wind is particularly strong, the interaction can produce magnetic storms which adversely affect radio signals on the Earth. At such times the auroras are particularly brilliant and widespread.

The Sun is moving approximately in the direction of Vega at about 12 miles per second, or about two-thirds as fast as the Earth moves in its orbit around the Sun.

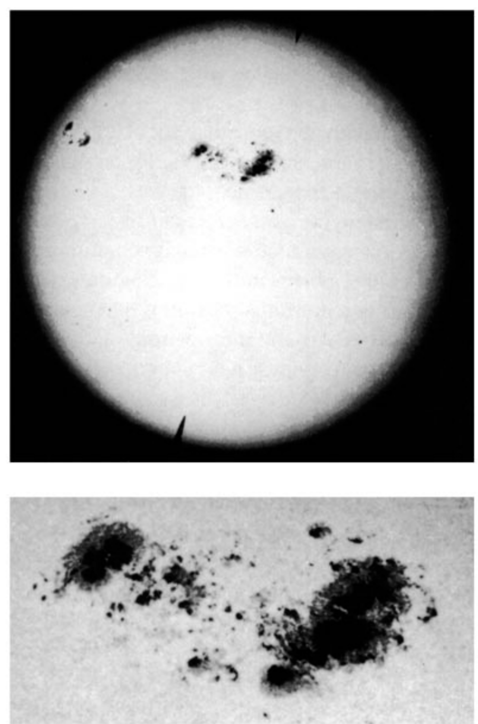


Figure 1507. Whole solar disk and an enlargement of the great spot group of April 7, 1947. Courtesy of Mt. Wilson and Palomar Observatories.

1508. The Planets

The principal bodies orbiting the Sun are called **planets**. Nine principal planets are known: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune, and Pluto. Of these, only four are commonly used for celestial navigation: Venus, Mars, Jupiter, and Saturn.

Except for Pluto, the orbits of the planets lie in nearly the same plane as the Earth's orbit. Therefore, as seen from the Earth, the planets are confined to a strip of the celestial sphere near the **ecliptic**, which is the intersection of the mean plane of the Earth's orbit around the Sun with the celestial sphere.

The two planets with orbits smaller than that of the Earth are called **inferior planets**, and those with orbits larger than that of the Earth are called **superior planets**. The four planets nearest the Sun are sometimes called the inner planets, and the others the outer planets. Jupiter, Saturn, Uranus, and Neptune are so much larger than the others that they are sometimes classed as major planets. Uranus is barely visible to the unaided eye; Neptune and Pluto are not visible without a telescope.

Planets can be identified in the sky because, unlike the stars, they do not twinkle. The stars are so distant that they are point sources of light. Therefore the stream of light from a star is easily scattered in the atmosphere, causing the twinkling effect. The naked-eye planets, however, are close enough to present perceptible disks. The broader stream of light from a planet is not easily disrupted.

The orbits of many thousands of tiny minor planets or asteroids lie chiefly between the orbits of Mars and Jupiter. These are all too faint to be seen with the naked eye.

1509. The Earth

In common with other planets, the Earth **rotates** on its axis and **revolves** in its orbit around the Sun. These motions are the principal source of the daily apparent motions of other celestial bodies. The Earth's rotation also causes a deflection of water and air currents to the right in the Northern Hemisphere and to the left in the Southern Hemisphere. Because of the Earth's rotation, high tides on the open sea lag behind the meridian transit of the Moon.

For most navigational purposes, the Earth can be considered a sphere. However, like the other planets, the Earth is approximately an **oblate spheroid**, or **ellipsoid of revolution**, flattened at the poles and bulged at the equator. See Figure 1509. Therefore, the polar diameter is less than the equatorial diameter, and the meridians are slightly elliptical, rather than circular. The dimensions of the Earth are recomputed from time to time, as additional and more precise measurements become available. Since the Earth is not exactly an ellipsoid, results differ slightly when equally precise and extensive measurements are made on different parts of the surface.

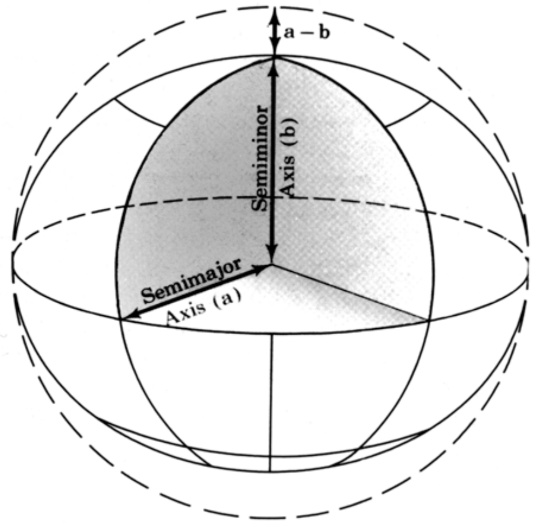


Figure 1509. Oblate spheroid or ellipsoid of revolution.

1510. Inferior Planets

Since Mercury and Venus are inside the Earth's orbit, they always appear in the neighborhood of the Sun. Over a period of weeks or months, they appear to oscillate back and forth from one side of the Sun to the other. They are seen either in the eastern sky before sunrise or in the western sky after sunset. For brief periods they disappear into the Sun's glare. At this time they are between the Earth and Sun (known as **inferior conjunction**) or on the opposite side of the Sun from the Earth (**superior conjunction**). On rare occasions at inferior conjunction, the planet will cross the face of the Sun as seen from the Earth. This is known as a **transit of the Sun**.

When Mercury or Venus appears most distant from the Sun in the evening sky, it is at greatest eastern elongation. (Although the planet is in the western sky, it is at its easternmost point from the Sun.) From night to night the planet will approach the Sun until it disappears into the glare of twilight. At this time it is moving between the Earth and Sun to inferior conjunction. A few days later, the planet will appear in the morning sky at dawn. It will gradually move away from the Sun to western elongation, then move back toward the Sun. After disappearing in the morning twilight, it will move behind the Sun to superior conjunction. After this it will reappear in the evening sky, heading toward eastern elongation.

Mercury is never seen more than about 28° from the Sun. For this reason it is not commonly used for navigation. Near greatest elongation it appears near the western horizon after sunset, or the eastern horizon before sunrise. At these times it resembles a first magnitude star and is sometimes reported as a new or strange object in the sky. The interval during which it appears as a morning or evening star can

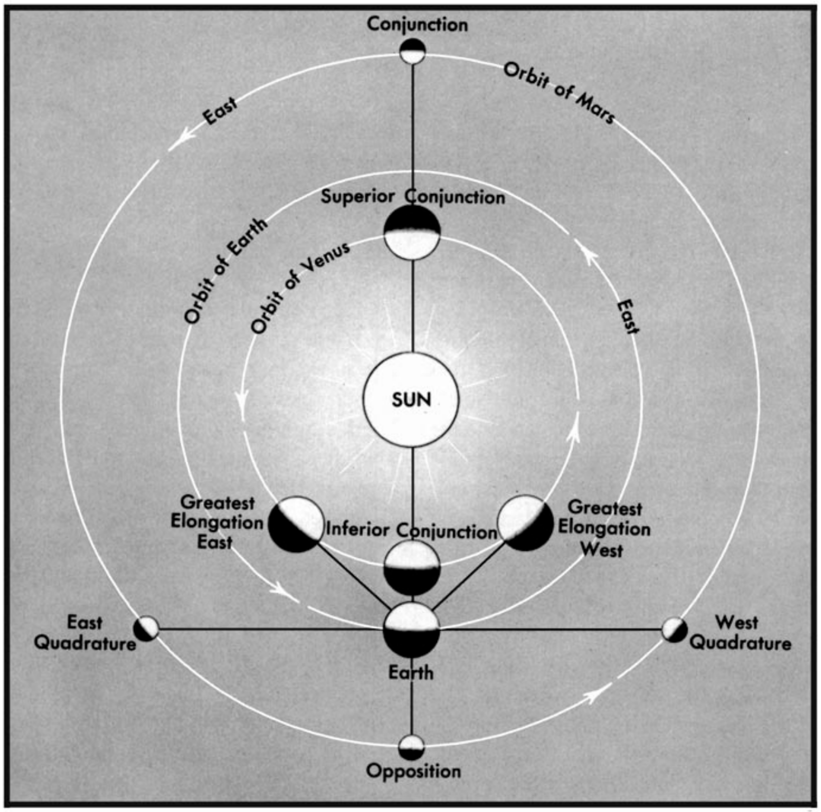


Figure 1510. Planetary configurations.

vary from about 30 to 50 days. Around inferior conjunction, Mercury disappears for about 5 days; near superior conjunction, it disappears for about 35 days. Observed with a telescope, Mercury is seen to go through phases similar to those of the Moon.

Venus can reach a distance of 47° from the Sun, allowing it to dominate the morning or evening sky. At maximum brilliance, about five weeks before and after inferior conjunction, it has a magnitude of about –4.4 and is brighter than any other object in the sky except the Sun and Moon. At these times it can be seen during the day and is sometimes observed for a celestial line of position. It appears as a morning or evening star for approximately 263 days in succession. Near inferior conjunction Venus disappears for 8 days; around superior conjunction it disappears for 50 days. When it transits the Sun, Venus can be seen by the naked eye as a small dot about the size of a group of Sunspots. Through strong binoculars or a telescope, Venus can be seen to go through a full set of phases.

1511. Superior Planets

As planets outside the Earth's orbit, the superior planets are not confined to the proximity of the Sun as seen from the Earth. They can pass behind the Sun (conjunction), but they cannot pass between the Sun and the Earth. Instead we see them move away from the Sun until they are opposite the Sun in the sky (**opposition**). When a superior planet is near conjunction, it rises and sets approximately with the Sun and is thus lost in the Sun's glare. Gradually it becomes visible in the early morning sky before sunrise. From day to day, it rises and sets earlier, becoming increasingly visible through the late night hours until dawn. Approaching opposition, the planet will rise in the late evening, until at opposition, it will rise when the Sun sets, be visible throughout the night, and set when the Sun rises.

Observed against the background stars, the planets normally move eastward in what is called **direct motion**. Approaching opposition, however, a planet will slow down, pause (at a stationary point), and begin moving westward (**retrograde motion**), until it reaches the next stationary point and resumes its direct motion. This is not because the planet is moving strangely in space. This relative, observed motion results because the faster moving Earth is catching up with and passing by the slower moving superior planet.

The superior planets are brightest and closest to the Earth at opposition. The interval between oppositions is known as the **synodic period**. This period is longest for the closest planet, Mars, and becomes increasingly shorter for the outer planets.

Unlike Mercury and Venus, the superior planets do not go through a full cycle of phases. They are always full or highly gibbous.

Mars can usually be identified by its orange color. It can become as bright as magnitude –2.8 but is more often between –1.0 and –2.0 at opposition. Oppositions occur at intervals of about 780 days. The planet is visible for about 330 days on either side of opposition. Near conjunction it is lost from view for about 120 days. Its two satellites can only be seen in a large telescope.

Jupiter, largest of the known planets, normally outshines Mars, regularly reaching magnitude –2.0 or brighter at opposition. Oppositions occur at intervals of about 400 days, with the planet being visible for about 180 days before and after opposition. The planet disappears for about 32 days at conjunction. Four satellites (of a total 16 currently known) are bright enough to be seen with binoculars. Their motions around Jupiter can be observed over the course of several hours.

Saturn, the outermost of the navigational planets, comes to opposition at intervals of about 380 days. It is visible for about 175 days before and after opposition, and disappears for about 25 days near conjunction. At opposition it becomes as bright as magnitude +0.8 to –0.2. Through good, high powered binoculars, Saturn appears as elongated because of its system of rings. A telescope is needed to examine the rings in any detail. Saturn is now known to have at least 18 satellites, none of which are visible to the unaided eye.

Uranus, **Neptune** and **Pluto** are too faint to be used for navigation; Uranus, at about magnitude 5.5, is faintly visible to the unaided eye.

1512. The Moon

The **Moon** is the only satellite of direct navigational interest. It revolves around the Earth once in about 27.3 days, as measured with respect to the stars. This is called the **sidereal month**. Because the Moon rotates on its axis with the same period with which it revolves around the Earth, the same side of the Moon is always turned toward the Earth. The cycle of phases depends on the Moon's revolution with respect to the Sun. This synodic month is approximately 29.53 days, but can vary from this average by up to a quarter of a day during any given month.

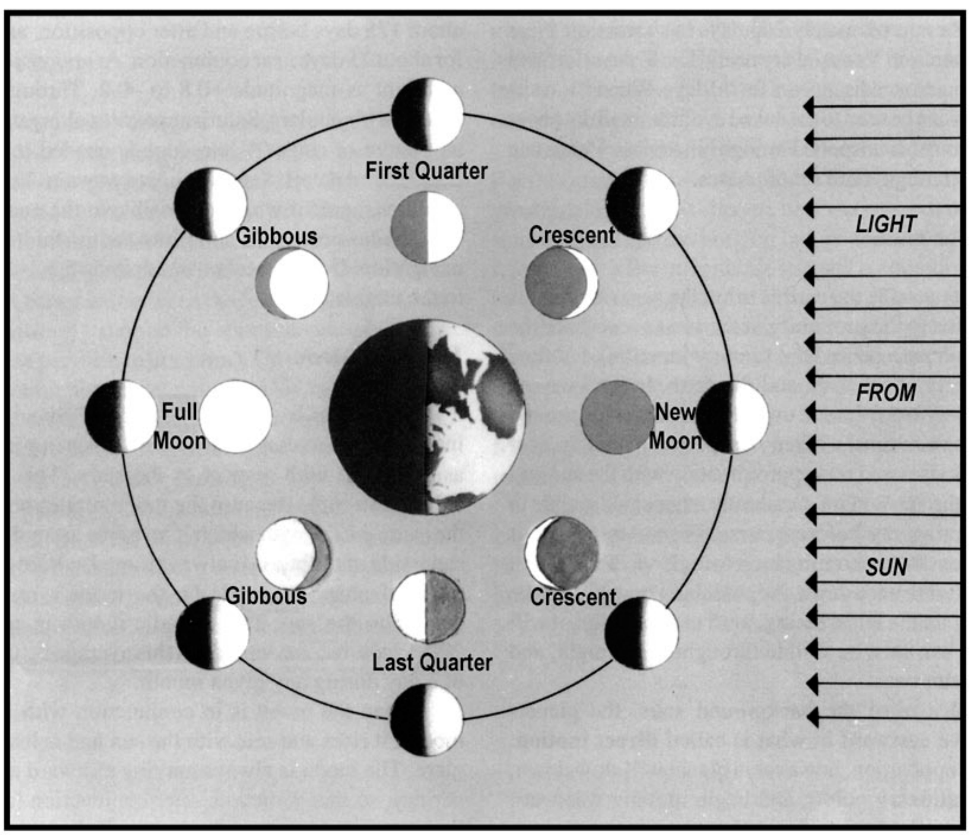


Figure 1512. Phases of the Moon. The inner figures of the Moon represent its appearance from the Earth.

When the Moon is in conjunction with the Sun (new Moon), it rises and sets with the Sun and is lost in the Sun's glare. The Moon is always moving eastward at about 12.2° per day, so that sometime after conjunction (as little as 16 hours, or as long as two days), the thin lunar crescent can be observed after sunset, low in the west. For the next couple of weeks, the Moon will **wax**, becoming more fully illuminated. From day to day, the Moon will rise (and set) later, becoming increasingly visible in the evening sky, until (about 7 days after new Moon) it reaches first quarter, when the Moon rises about noon and sets about midnight. Over the next week the Moon will rise later and later in the afternoon until full Moon, when it rises about sunset and dominates the sky throughout the night. During the next couple of weeks the Moon will **wane**, rising later and later at night. By last quarter (a week after full Moon), the Moon rises about midnight and sets at noon. As it approaches new Moon, the Moon becomes an increasingly thin crescent, and is seen only in the early morning sky. Sometime before conjunction (16 hours to 2 days before conjunction) the thin crescent will disappear in the glare of morning twilight.

At full Moon, the Sun and Moon are on opposite sides of the ecliptic. Therefore, in the winter the full Moon rises early, crosses the celestial meridian high in the sky, and sets late; as the Sun does in the summer. In the summer the full Moon rises in the southeastern part of the sky (Northern Hemisphere), remains relatively low in the sky, and sets along the southwestern horizon after a short time above the horizon.

At the time of the autumnal equinox, the part of the ecliptic opposite the Sun is most nearly parallel to the horizon. Since the eastward motion of the Moon is approximately along the ecliptic, the delay in the time of rising of the full Moon from night to night is less than at other times of the year. The full Moon nearest the autumnal equinox is called the **Harvest Moon**; the full Moon a month later is called the **Hunter's Moon**. See Figure 1512.

1513. Comets and Meteors

Although **comets** are noted as great spectacles of nature, very few are visible without a telescope. Those that become widely visible do so because they develop long, glowing tails. Comets are swarms of relatively small solid bodies held together by gravity. Around the nucleus, a gaseous head or coma and tail may form as the comet approaches the Sun. The tail is directed away from the Sun, so that it follows the head while the comet is approaching the Sun, and precedes the head while the comet is receding. The total mass of a comet is very small, and the tail is so thin that stars can easily be seen through it. In 1910, the Earth passed through the tail of Halley's comet without noticeable effect.

Compared to the well-ordered orbits of the planets, comets are erratic and inconsistent. Some travel east to west and some west to east, in highly eccentric orbits inclined at any angle to the ecliptic. Periods of revolution range from about 3 years to thousands of years. Some comets may speed away from the solar system after gaining velocity as they pass by Jupiter or Saturn.

The short-period comets long ago lost the gasses needed to form a tail. Long period comets, such as Halley's comet, are more likely to develop tails. The visibility of a comet depends very much on how close it approaches the Earth. In 1910, Halley's comet spread across the sky (Figure 1513). Yet when it returned in 1986, the Earth was not well situated to get a good view, and it was barely visible to the unaided eye.

Meteors, popularly called **shooting stars**, are tiny, solid bodies too small to be seen until heated to incandescence by air friction while passing through the Earth's atmosphere. A particularly bright meteor is called a **fireball**. One that explodes is called a **bolide**. A meteor that survives its trip through the atmosphere and lands as a solid particle is called a **meteorite**.

Vast numbers of meteors exist. An estimated average of some 1,000,000 meteors large enough to be seen enter the Earth's atmosphere each hour, and many times this number undoubtedly enter, but are too small to attract attention. The cosmic dust they create falls to earth in a constant shower.

Meteor showers occur at certain times of the year when the Earth passes through **meteor swarms**, the scattered remains of comets that have broken up. At these times the number of meteors observed is many times the usual number.

A faint glow sometimes observed extending upward approximately along the ecliptic before sunrise and after sunset has been attributed to the reflection of Sunlight from quantities of this material. This glow is called **zodiacal light**. A faint glow at that point of the ecliptic 180° from the Sun is called the **gegenschein** or **counterglow**.

1514. Stars

Stars are distant Suns, in many ways resembling our own. Like the Sun, stars are massive balls of gas that create their own energy through thermonuclear reactions.

Although stars differ in size and temperature, these differences are apparent only through analysis by astronomers. Some differences in color are noticeable to the unaided eye. While most stars appear white, some (those of lower temperature) have a reddish hue. In Orion, blue Rigel and red Betelgeuse, located on opposite sides of the belt, constitute a noticeable contrast.

The stars are not distributed uniformly around the sky. Striking configurations, known as **constellations**, were noted by ancient peoples, who supplied them with names and myths. Today astronomers use constellations—88 in all—to identify areas of the sky.

Under ideal viewing conditions, the dimmest star that can be seen with the unaided eye is of the sixth magnitude. In the entire sky there are about 6,000 stars of this

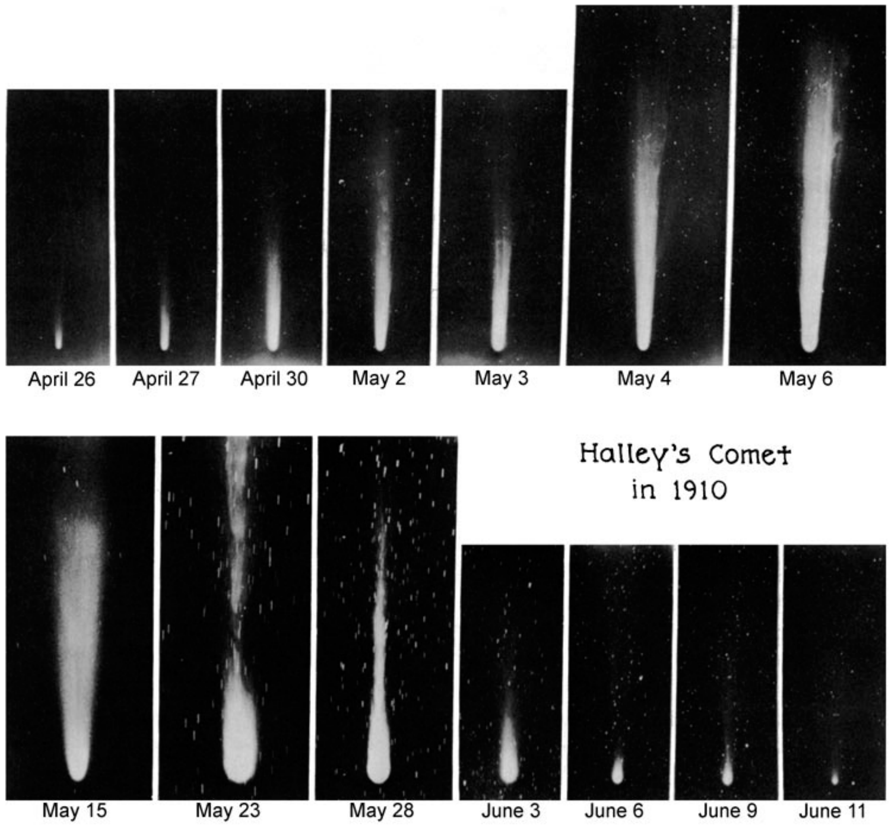


Figure 1513. Halley's Comet; fourteen views, made between April 26 and June 11, 1910.
Courtesy of Mt. Wilson and Palomar Observatories.

magnitude or brighter. Half of these are below the horizon at any time. Because of the greater absorption of light near the horizon, where the path of a ray travels for a greater distance through the atmosphere, not more than perhaps 2,500 stars are visible to the unaided eye at any time. However, the average navigator seldom uses more than perhaps 20 or 30 of the brighter stars.

Stars which exhibit a noticeable change of magnitude are called **variable stars**. A star which suddenly becomes several magnitudes brighter and then gradually fades is called a **nova**. A particularly bright nova is called a **supernova**.

Two stars which appear to be very close together are called a **double star**. If more than two stars are included in the group, it is called a **multiple star**. A group of a few dozen to several hundred stars moving through space together is called an **open cluster**. The Pleiades is an example of an open cluster. There are also spherically symmetric clusters of hundreds of thousands of stars known as **globular clusters**. The globular clusters are all too distant to be seen with the naked eye.

A cloudy patch of matter in the heavens is called a **nebula**. If it is within the galaxy of which the Sun is a part, it is called a **galactic nebula**; if outside, it is called an **extragalactic nebula**.

Motion of a star through space can be classified by its vector components. That component in the line of sight is called **radial motion**, while that component across the line of sight, causing a star to change its apparent position relative to the background of more distant stars, is called **proper motion**.

1515. Galaxies

A **galaxy** is a vast collection of clusters of stars and clouds of gas. In a galaxy the stars tend to congregate in groups called **star clouds** arranged in long spiral arms. The spiral nature is believed due to revolution of the stars about the center of the galaxy, the inner stars revolving more rapidly than the outer ones (Figure 1515).

The Earth is located in the Milky Way galaxy, a slowly spinning disk more than 100,000 light years in diameter. All the bright stars in the sky are in the Milky Way. However, the most dense portions of the galaxy are seen as the great, broad band that glows in the summer nighttime sky. When we look toward the constellation Sagittarius, we are looking toward the

Figure 1515. Spiral nebula Messier 51, In Canes Venetici. Satellite nebula is NGC 5195.
Courtesy of Mt. Wilson and Palomar Observatories.

center of the Milky Way, 30,000 light years away.

Despite their size and luminance, almost all other galaxies are too far away to be seen with the unaided eye. An exception in the northern hemisphere is the Great Galaxy (sometimes called the Great Nebula) in Andromeda, which appears as a faint glow. In the southern hemisphere, the Large and Small Magellanic Clouds (named after Ferdinand Magellan) are the nearest known neighbors of the Milky Way. They are approximately 1,700,000 light years distant. The Magellanic Clouds can be seen as sizable glowing patches in the southern sky.

APPARENT MOTION

1516. Apparent Motion due to Rotation of the Earth

Apparent motion caused by the Earth's rotation is much greater than any other observed motion of celestial bodies. It is this motion that causes celestial bodies to appear to rise along the eastern half of the horizon, climb to maximum altitude as they cross the meridian, and set along the western horizon, at about the same point relative to due west as the rising point was to due east. This apparent motion along the daily path, or **diurnal circle**, of the body is approximately parallel to the plane of the equator. It would be exactly so if rotation of the Earth were the only motion and the axis of rotation of the Earth were stationary in space.

The apparent effect due to rotation of the Earth varies with the latitude of the observer. At the equator, where the equatorial plane is vertical (since the axis of rotation of the Earth is parallel to the plane of the horizon), bodies appear to rise and set vertically. Every celestial body is above the horizon approximately half the time. The celestial sphere as seen by an observer at the equator is called the right sphere, shown in Figure 1516a.

For an observer at one of the poles, bodies having constant declination neither rise nor set (neglecting precession of the equinoxes and changes in refraction), but circle the sky, always at the same altitude, making one complete trip around the horizon each day. At the North Pole the motion is clockwise, and at the South Pole it is counterclockwise. Approximately half the stars are always above the horizon and the other half never are. The parallel sphere at the poles is illustrated in Figure 1516b.

Between these two extremes, the apparent motion is a combination of the two. On this oblique sphere, illustrated in Figure 1516c, circumpolar celestial bodies remain above the horizon during the entire 24 hours, circling the elevated celestial pole each day. The stars of Ursa Major (the Big Dipper) and Cassiopeia are circumpolar for many observers in the United States.

An approximately equal part of the celestial sphere remains below the horizon during the entire day. For example, Crux is not visible to most observers in the United States. Other bodies rise obliquely along the eastern horizon, climb to maximum altitude at the celestial meridian, and set along the western horizon. The length of time above the horizon and the altitude at meridian transit vary with both the latitude of the observer and the declination of the body. At the polar circles of the Earth even the Sun becomes circumpolar. This is the land of the midnight Sun, where the Sun does not set during part of the summer and does not rise during part of the winter.

The increased obliquity at higher latitudes explains why days and nights are always about the same length in the tropics, and the change of length of the day becomes greater as latitude increases, and why twilight lasts longer in higher latitudes. Evening twilight starts at sunset, and morning twilight ends at sunrise. The darker limit of twilight occurs when the center of the Sun is a stated number of degrees below the celestial horizon. Three kinds of twilight are

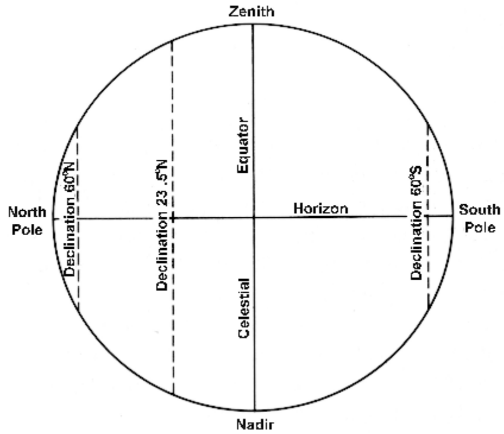


Figure 1516a. The right sphere.

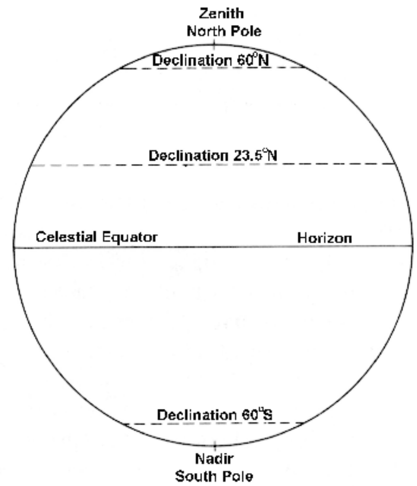


Figure 1516b. The parallel sphere.

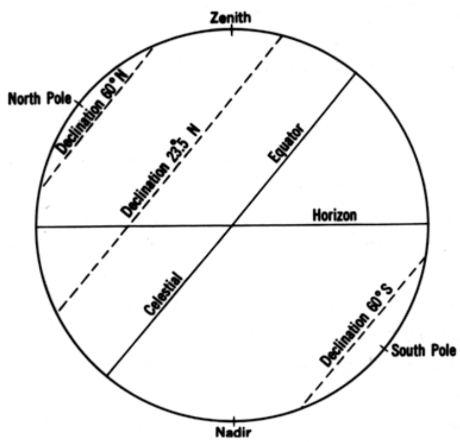


Figure 1516c. The oblique sphere at latitude 40°N.

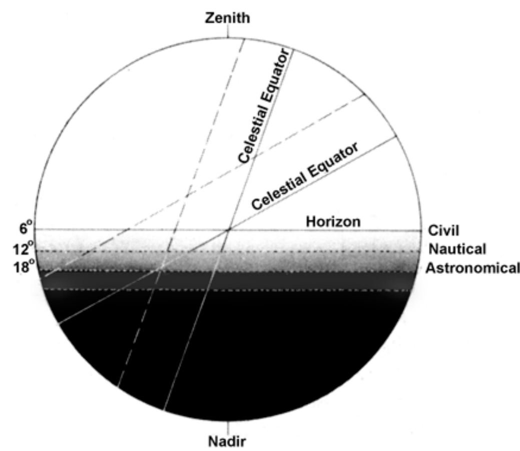


Figure 1516d. The various twilight at latitude 20°N and latitude 60°N.

Twilight	*Lighter limit*	*Darker limit*	*At darker limit*
civil	–0°50'	–6°	Horizon clear; bright stars visible
nautical	–0°50'	–12°	Horizon not visible
astronomical	–0°50'	–18°	Full night

Table 1516. Limits of the three twilights.

defined: civil, nautical and astronomical. See Table 1516.

The conditions at the darker limit are relative and vary considerably under different atmospheric conditions.

In Figure 1516d, the twilight band is shown, with the darker limits of the various kinds indicated. The nearly vertical celestial equator line is for an observer at latitude 20°N. The nearly horizontal celestial equator line is for an observer at latitude 60°N. The broken line in each case is the diurnal circle of the Sun when its declination is 15°N. The relative duration of any kind of twilight at the two latitudes is indicated by the portion of the diurnal circle between the horizon and the darker limit, although it is not directly proportional to the relative length of line shown since the projection is orthographic. The duration of twilight at the higher latitude is longer, proportionally, than shown. Note that complete darkness does not occur at latitude 60°N when the declination of the Sun is 15°N.

1517. Apparent Motion due to Revolution of the Earth

If it were possible to stop the rotation of the Earth so that the celestial sphere would appear stationary, the effects of the revolution of the Earth would become more noticeable. In one year the Sun would appear to make one complete trip around the Earth, from west to east. Hence, it would seem to move eastward a little less than 1° per day. This motion can be observed by watching the changing position of the Sun among the stars. But since both Sun and stars generally are not visible at the same time, a better way is to observe the constellations at the same time each night. On any night a star rises nearly four minutes earlier than on the previous night. Thus, the celestial sphere appears to shift westward nearly 1° each night, so that different constellations are associated with different seasons of the year.

Apparent motions of planets and the Moon are due to a combination of their motions and those of the Earth. If the rotation of the Earth were stopped, the combined apparent motion due to the revolutions of the Earth and other bodies would be similar to that occurring if both rotation and revolution of the Earth were stopped. Stars would appear nearly stationary in the sky but would undergo a small annual cycle of change due to aberration. The motion of the Earth in its orbit is sufficiently fast to cause the light from stars to appear to shift slightly in the direction of the Earth's motion. This is similar to the effect one experiences when walking in vertically-falling rain that appears to come from ahead due to the observer's own forward motion. The apparent direction of the light ray from the star is the vector difference of the motion of light and the motion of the Earth, similar to that of apparent wind on a moving vessel. This effect is most apparent for a body perpendicular to the line of travel of the Earth in its orbit, for which it reaches a maximum value of 20.5". The effect of aberration can be noted by comparing the coordinates (declination and sidereal hour angle) of various stars throughout the year. A change is observed in some bodies as the year progresses, but at the end of the year the values have returned almost to what they were at the beginning. The reason they do not return exactly is due to proper motion and precession of the equinoxes. It is also due to nutation, an irregularity in the motion of the Earth due to the disturbing effect of other celestial bodies, principally the Moon. Polar motion is a slight wobbling of the Earth about its axis of rotation and sometimes wandering of the poles. This motion, which does not exceed 40 feet from the mean position, produces slight variation of latitude and longitude of places on the Earth.

1518. Apparent Motion due to Movement of other Celestial Bodies

Even if it were possible to stop both the rotation and revolution of the Earth, celestial bodies would not appear stationary on the celestial sphere. The Moon would make one revolution about the Earth each sidereal month, rising in the west and setting in the east. The inferior planets would appear to move eastward and westward relative to the Sun, staying within the zodiac. Superior planets would appear to make one revolution around the Earth, from west to east, each sidereal period.

Since the Sun (and the Earth with it) and all other stars are in motion relative to each other, slow apparent motions would result in slight changes in the positions of the stars relative to each other. This space motion is, in fact, observed by telescope. The component of such motion across the line of sight, called proper motion, produces a change in the apparent position of the star. The maximum which has been observed is that of Barnard's Star, which is moving at the rate of 10.3 seconds per year. This is a tenth-magnitude star, not visible to the unaided eye. Of the 57 stars listed on the daily pages of the almanacs, Rigil Kentaurus has the greatest proper motion, about 3.7 seconds per year. Arcturus, with 2.3 seconds per year, has the greatest proper motion of the navigational stars in the Northern Hemisphere. In a few thousand years proper motion will be sufficient to materially alter some familiar configurations of stars, notably Ursa Major.

1519. The Ecliptic

The **ecliptic** is the path the Sun appears to take among the stars due to the annual revolution of the Earth in its orbit. It is considered a great circle of the celestial sphere, inclined at an angle of about 23°26' to the celestial equator, but undergoing a continuous slight change. This angle is called the **obliquity of the ecliptic**. This inclination is due to the fact that the axis of rotation of the Earth is not perpendicular to its orbit. It is this inclination which causes the Sun to appear to move north and south during the year, giving the Earth its seasons and changing lengths of periods of daylight.

Refer to Figure 1519a. The Earth is at perihelion early

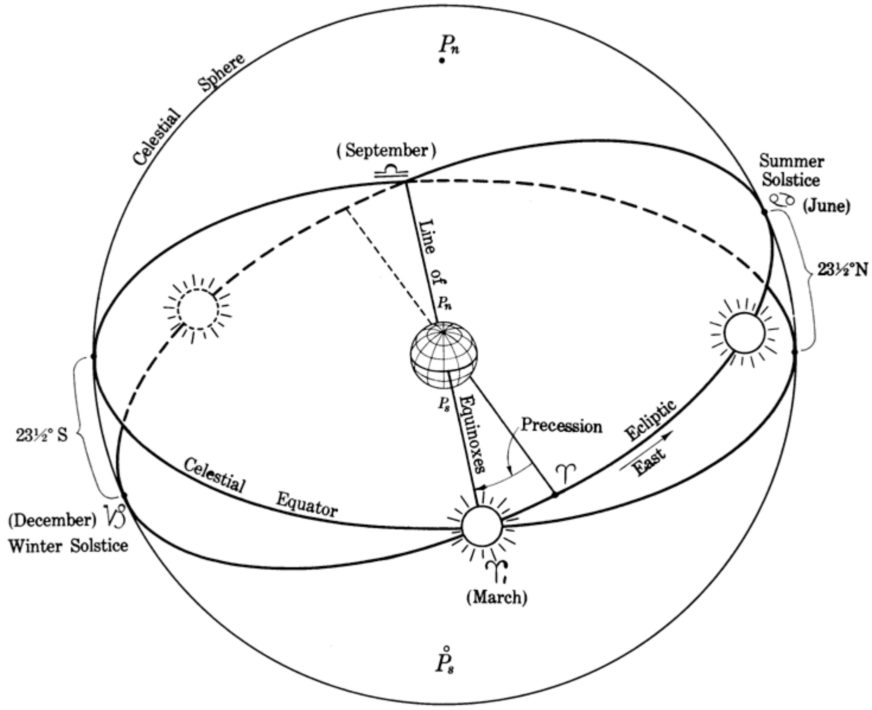


Figure 1519a. Apparent motion of the Sun in the ecliptic.

in January and at aphelion 6 months later. On or about June 21, about 10 or 11 days before reaching aphelion, the northern part of the Earth's axis is tilted toward the Sun. The north polar regions are having continuous Sunlight; the Northern Hemisphere is having its summer with long, warm days and short nights; the Southern Hemisphere is having winter with short days and long, cold nights; and the south polar region is in continuous darkness. This is the **summer solstice**. Three months later, about September 23, the Earth has moved a quarter of the way around the Sun, but its axis of rotation still points in about the same direction in space. The Sun shines equally on both hemispheres, and days and nights are the same length over the entire world. The Sun is setting at the North Pole and rising at the South Pole. The Northern Hemisphere is having its autumn, and the Southern Hemisphere its spring. This is the **autumnal equinox**. In another three months, on or about December 22, the Southern Hemisphere is tilted toward the Sun and conditions are the reverse of those six months earlier; the Northern Hemisphere is having its winter, and the Southern Hemisphere its summer. This is the **winter solstice**. Three months later, when both hemispheres again receive equal amounts of Sunshine, the Northern Hemisphere is having spring and the Southern Hemisphere autumn, the reverse of conditions six months before. This is the **vernal equinox**.

The word "equinox," meaning "equal nights," is applied because it occurs at the time when days and nights are of approximately equal length all over the Earth. The word "solstice," meaning "Sun stands still," is applied because the Sun stops its apparent northward or southward motion and momentarily "stands still" before it starts in the opposite direction. This action, somewhat analogous to the "stand" of the tide, refers to the motion in a north-south direction only, and not to the daily apparent revolution around the Earth. Note that it does not occur when the Earth is at perihelion or aphelion. Refer to Figure 1519a. At the time of the vernal equinox, the Sun is directly over the equator, crossing from the Southern Hemisphere to the Northern Hemisphere. It rises due east and sets due west, remaining above the horizon for approximately 12 hours. It is not exactly 12 hours because of refraction, semidiameter, and the height of the eye of the observer. These cause it to be above the horizon a little longer than below the horizon. Following the vernal equinox, the northerly declination increases, and the Sun climbs higher in the sky each day (at the latitudes of the United States), until the summer solstice, when a declination of about 23°26' north of the celestial equator is reached. The Sun then gradually retreats southward until it is again over the equator at the autumnal equinox, at about 23°26' south of the celestial equator at the winter solstice, and back over the celestial equator again at the next vernal equinox.

The Earth is nearest the Sun during the northern hemisphere winter. It is not the distance between the Earth and Sun that is responsible for the difference in temperature during the different seasons, but the altitude of the Sun in the sky and the length of time it remains above the horizon.

During the summer the rays are more nearly vertical, and hence more concentrated, as shown in Figure 1519b. Since the Sun is above the horizon more than half the time, heat is being added by absorption during a longer period than it is being lost by radiation. This explains the lag of the seasons. Following the longest day, the Earth continues to receive more heat than it dissipates, but at a decreasing proportion. Gradually the proportion decreases until a balance is reached, after which the Earth cools, losing more heat than it gains. This is analogous to the day, when the highest temperatures normally occur several hours after the Sun reaches maximum altitude at meridian transit. A similar lag occurs at other seasons of the year. Astronomically, the seasons begin at the equinoxes and solstices. Meteorologically, they differ from place to place.

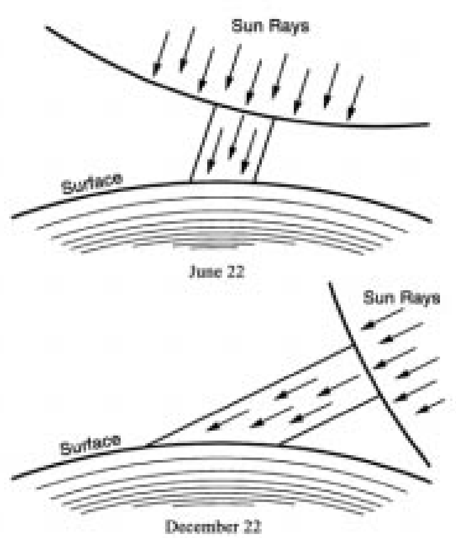


Figure 1519b. Sunlight in summer and winter. Winter sunlight is distributed over a larger area and shines fewer hours each day, causing less heat energy to reach the Earth.

Since the Earth travels faster when nearest the Sun, the northern hemisphere (astronomical) winter is shorter than its summer by about seven days.

Everywhere between the parallels of about 23°26'N and about 23°26'S the Sun is directly overhead at some time during the year. Except at the extremes, this occurs twice: once as the Sun appears to move northward, and the second time as it moves southward. This is the **torrid zone**. The northern limit is the **Tropic of Cancer**, and the southern limit is the **Tropic of Capricorn**. These names come from the constellations which the Sun entered at the solstices when the names were first applied more than 2,000 years ago. Today, the Sun is in the next constellation toward the west because of precession of the equinoxes. The parallels about 23°26' from the poles, marking the approximate limits of the circumpolar Sun, are called **polar circles**, the one in the Northern Hemisphere being the **Arctic Circle** and the one in the Southern Hemisphere the **Antarctic Circle**. The areas inside the polar circles are the north and south **frigid zones**. The regions between the frigid zones and the torrid zones are the north and south **temperate zones**.

The expression "vernal equinox" and associated expressions are applied both to the times and points of occurrence of the various phenomena. Navigationally, the vernal equinox is sometimes called the **first point of Aries** (symbol ♈) because, when the name was given, the Sun entered the constellation Aries, the ram, at this time. This point is of interest to navigators because it is the origin for measuring **sidereal hour angle**. The expressions March equinox, June solstice, September equinox, and December solstice are occasionally applied as appropriate, because the more common names are associated with the seasons in the Northern Hemisphere and are six months out of step for the Southern Hemisphere.

The axis of the Earth is undergoing a precessional motion similar to that of a top spinning with its axis tilted. In about 25,800 years the axis completes a cycle and returns to the position from which it started. Since the celestial equator is 90° from the celestial poles, it too is moving. The result is a slow westward movement of the equinoxes and solstices, which has already carried them about 30°, or one constellation, along the ecliptic from the positions they occupied when named more than 2,000 years ago. Since sidereal hour angle is measured from the vernal equinox, and declination from the celestial equator, the coordinates of celestial bodies would be changing even if the bodies themselves were stationary. This westward motion of the equinoxes along the ecliptic is called **precession of the equinoxes**. The total amount, called **general precession**, is about 50 seconds of arc per year. It may be considered divided into two components: precession in right ascension (about 46.10 seconds per year) measured along the celestial equator, and precession in declination (about 20.04" per year) measured perpendicular to the celestial equator. The annual change in the coordinates of any given star, due to precession alone, depends upon its position on the celestial sphere, since these coordinates are measured relative to the polar axis while the precessional motion is relative to the ecliptic axis.

Due to precession of the equinoxes, the celestial poles are slowly describing circles in the sky. The north celestial pole is moving closer to Polaris, which it will pass at a distance of approximately 28 minutes about the year 2102. Following this, the polar distance will increase, and eventually other stars, in their turn, will become the Pole Star.

The precession of the Earth's axis is the result of gravitational forces exerted principally by the Sun and Moon on the Earth's equatorial bulge. The spinning Earth responds to these forces in the manner of a gyroscope. Regression of the nodes introduces certain irregularities known as **nutation** in the precessional motion. See Figure 1519c.

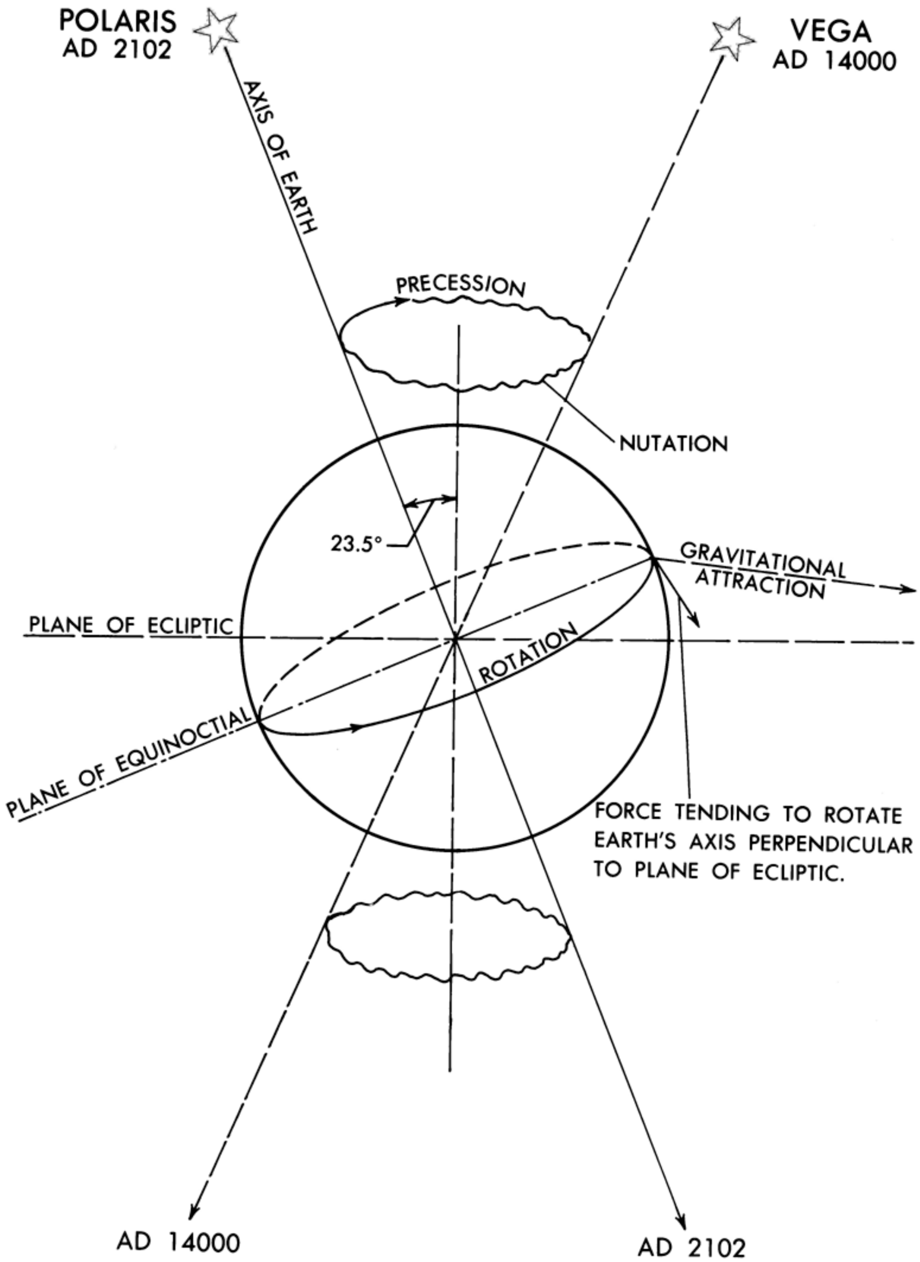


Figure 1519c. Precession and nutation.

1520. The Zodiac

The **zodiac** is a circular band of the sky extending 8° on each side of the ecliptic. The navigational planets and the Moon are within these limits. The zodiac is divided into 12 sections of 30° each, each section being given the name and symbol ("sign") of a constellation. These are shown in Figure 1520. The names were assigned more than 2,000 years ago, when the Sun entered Aries at the vernal equinox, Cancer at the summer solstice, Libra at the autumnal equinox, and Capricornus at the winter solstice. Because of precession, the zodiacal signs have shifted with respect to the constellations. Thus at the time of the vernal equinox, the Sun is said to be at the "first point of Aries," though it is in the constellation Pisces.

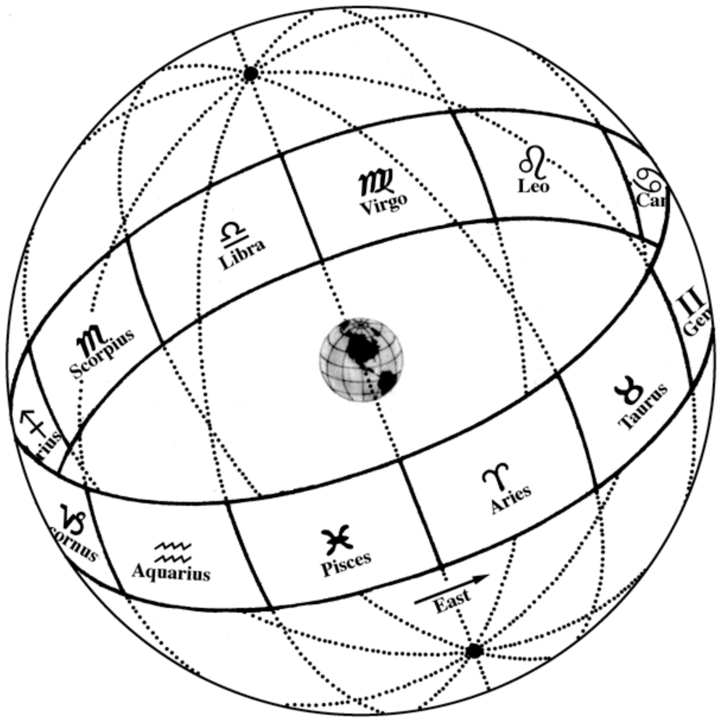


Figure 1520. The zodiac.

1521. Time and the Calendar

Traditionally, astronomy has furnished the basis for measurement of time, a subject of primary importance to the navigator. The **year** is associated with the revolution of the Earth in its orbit. The **day** is one rotation of the Earth about its axis.

The duration of one rotation of the Earth depends upon the external reference point used. One rotation relative to the Sun is called a **solar day**. However, rotation relative to the apparent Sun (the actual Sun that appears in the sky) does not provide time of uniform rate because of variations in the rate of revolution and rotation of the Earth. The error due to lack of uniform rate of revolution is removed by using a fictitious **mean Sun**. Thus, mean solar time is nearly equal to the average apparent solar time. Because the accumulated difference between these times, called the **equation of time**, is continually changing, the period of daylight is shifting slightly, in addition to its increase or decrease in length due to changing declination. Apparent and mean Suns seldom cross the celestial meridian at the same time. The earliest sunset (in latitudes of the United States) occurs about two weeks before the winter solstice, and the latest sunrise occurs about two weeks after winter solstice. A similar but smaller apparent discrepancy occurs at the summer solstice.

Universal Time is a particular case of the measure known in general as mean solar time. Universal Time is the mean solar time on the Greenwich meridian, reckoned in days of 24 mean solar hours beginning with 0 hours at midnight. Universal Time and sidereal time are rigorously related by a formula so that if one is known the other can be found. Universal Time is the standard in the application of astronomy to navigation.

If the vernal equinox is used as the reference, a **sidereal day** is obtained, and from it, **sidereal time**. This indicates the approximate positions of the stars, and for this reason it is the basis of star charts and star finders. Because of the revolution of the Earth around the Sun, a sidereal day is about 3 minutes 56 seconds shorter than a solar day, and there is one more sidereal than solar days in a year. One mean solar day equals 1.00273791 mean sidereal days. Because of precession of the equinoxes, one rotation of the Earth with respect to the stars is not quite the same as one rotation with respect to the vernal equinox. One mean solar day averages 1.0027378118868 rotations of the Earth with respect to the stars.

In tide analysis, the Moon is sometimes used as the reference, producing a **lunar day** averaging 24 hours 50 minutes (mean solar units) in length, and lunar time.

Since each kind of day is divided arbitrarily into 24 hours, each hour having 60 minutes of 60 seconds, the length of each of these units differs somewhat in the various kinds of time.

Time is also classified according to the terrestrial meridian used as a reference. **Local time** results if one's own meridian is used, **zone time** if a nearby reference meridian is used over a spread of longitudes, and **Greenwich** or **Universal Time** if the Greenwich meridian is used.

The period from one vernal equinox to the next (the cycle of the seasons) is known as the **tropical year**. It is approximately 365 days, 5 hours, 48 minutes, 45 seconds, though the length has been slowly changing for many centuries. Our calendar, the Gregorian calendar, approximates the tropical year with a combination of common years of 365 days and leap years of 366 days. A leap year is any year divisible by four, unless it is a century year, which must be divisible by 400 to be a leap year. Thus, 1700, 1800, and 1900 were not leap years, but 2000 was. A critical mistake was made by John Hamilton Moore in calling 1800 a leap year, causing an error in the tables in his book, *The Practical Navigator*. This error caused the loss of at least one ship and was later discovered by Nathaniel Bowditch while writing the first edition of *The New American Practical Navigator*.

See Chapter 18 for an in-depth discussion of time.

1522. Eclipses

If the orbit of the Moon coincided with the plane of the ecliptic, the Moon would pass in front of the Sun at every new Moon, causing a solar eclipse. At full Moon, the Moon would pass through the Earth's shadow, causing a lunar eclipse. Because of the Moon's orbit is inclined 5° with respect to the ecliptic, the Moon usually passes above or below the Sun at new Moon and above or below the Earth's shadow at full Moon. However, there are two points at which the plane of the Moon's orbit intersects the ecliptic. These are the **nodes** of the Moon's orbit. If the Moon passes one of these points at the same time as the Sun, a **solar eclipse** takes place. This is shown in Figure 1522.

The Sun and Moon are of nearly the same apparent size to an observer on the Earth. If the Moon is at perigee, the Moon's apparent diameter is larger than that of the Sun, and its shadow reaches the Earth as a nearly round dot only a few miles in diameter. The dot moves rapidly across the Earth, from west to east, as the Moon continues in its orbit. Within the dot, the Sun is completely hidden from view, and a total eclipse of the Sun occurs. For a considerable distance around the shadow, part of the surface of the Sun is obscured, and a **partial eclipse** occurs. In the line of travel of the shadow a partial eclipse occurs as the round disk of the Moon appears to move slowly across the surface of the Sun, hiding an ever-increasing part of it, until the total eclipse occurs. Because of the uneven edge of the mountainous Moon, the light is not cut off evenly. But several last illuminated portions appear through the valleys or passes between the mountain peaks. These are called **Baily's Beads**.

A total eclipse is a spectacular phenomenon. As the last light from the Sun is cut off, the solar **corona**, or envelope of thin, illuminated gas around the Sun becomes visible. Wisps of more dense gas may appear as **solar prominences**. The only light reaching the observer is that diffused by the atmosphere surrounding the shadow. As the Moon appears to continue on across the face of the Sun, the Sun finally emerges from the other side, first as Baily's Beads, and then as an ever widening crescent until no part of its surface is obscured by the Moon.

The duration of a total eclipse depends upon how nearly the Moon crosses the center of the Sun, the location of the shadow on the Earth, the relative orbital speeds of the Moon and Earth, and (principally) the relative apparent diameters of the Sun and Moon. The maximum length that can occur is a little more than seven minutes.

If the Moon is near apogee, its apparent diameter is less than that of the Sun, and its shadow does not quite reach the Earth. Over a small area of the Earth directly in line with the Moon and Sun, the Moon appears as a black disk almost covering the surface of the Sun, but with a thin ring of the Sun around its edge. This **annular eclipse** occurs a little more often than a total eclipse.

If the shadow of the Moon passes close to the Earth, but not directly in line with it, a partial eclipse may occur without a total or annular eclipse.

An eclipse of the Moon (or **lunar eclipse**) occurs when the Moon passes through the shadow of the Earth, as shown in Figure 1522. Since the diameter of the Earth is about $3^1/_2$ times that of the Moon, the Earth's shadow at the distance of the Moon is much larger than that of the Moon. A total eclipse of the Moon can last nearly $1^3/_4$ hours, and some part of the Moon may be in the Earth's shadow for almost 4 hours.

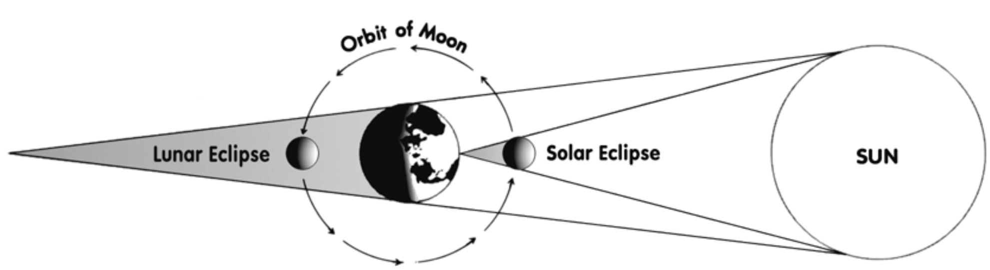


Figure 1522. Eclipses of the Sun and Moon.

During a total solar eclipse no part of the Sun is visible because the Moon is in the line of sight. But during a lunar eclipse some light does reach the Moon, diffracted by the atmosphere of the Earth, and hence the eclipsed full Moon is visible as a faint reddish disk. A lunar eclipse is visible over the entire hemisphere of the Earth facing the Moon. Anyone who can see the Moon can see the eclipse.

During any one year there may be as many as five eclipses of the Sun, and always there are at least two. There may be as many as three eclipses of the Moon, or none. The total number of eclipses during a single year does not exceed seven, and can be as few as two. There are more solar than lunar eclipses, but the latter can be seen more often because of the restricted areas over which solar eclipses are visible.

The Sun, Earth, and Moon are nearly aligned on the line of nodes twice each eclipse year of 346.6 days. This is less than a calendar year because of **regression of the nodes**. In a little more than 18 years the line of nodes returns to approximately the same position with respect to the Sun, Earth, and Moon. During an almost equal period, called the **saros**, a cycle of eclipses occurs. During the following saros the cycle is repeated with only minor differences.

COORDINATES

1523. Latitude And Longitude

Latitude and **longitude** are coordinates used to locate positions on the Earth. This article discusses three different definitions of these coordinates.

Astronomic latitude is the angle (ABQ, Figure 1523) between a line in the direction of gravity (AB) at a station and the plane of the equator (QQ'). **Astronomic longitude** is the angle between the plane of the celestial meridian at a station and the plane of the celestial meridian at Greenwich. These coordinates are customarily found by means of celestial observations. If the Earth were perfectly homogeneous and round, these positions would be consistent and satisfactory. However, because of deflection of the vertical due to uneven distribution of the mass of the Earth, lines of equal astronomic latitude and longitude are not circles, although the irregularities are small. In the United States the prime vertical component (affecting longitude) may be a little more than 18", and the meridional component (affecting latitude) as much as 25".

Geodetic latitude is the angle (ACQ, Figure 1523) between a normal to the spheroid (AC) at a station and the plane of the geodetic equator (QQ'). **Geodetic longitude** is the angle between the plane defined by the normal to the spheroid and the axis of the Earth and the plane of the geodetic meridian at Greenwich. These values are obtained when astronomical latitude and longitude are corrected for deflection of the vertical. These coordinates are used for charting and are frequently referred to as **geographic latitude** and **geographic longitude**, although these expressions are sometimes used to refer to astronomical latitude.

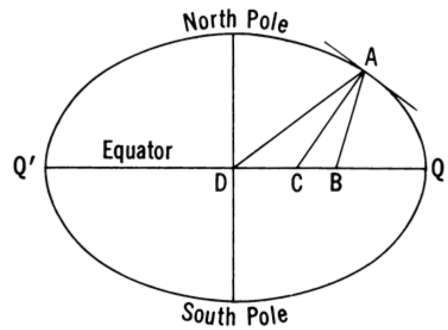


Figure 1523. Three kinds of latitude at point A.

Geocentric latitude is the angle (ADQ, Figure 1523) at the center of the ellipsoid between the plane of its equator (QQ') and a straight line (AD) to a point on the surface of the Earth. This differs from geodetic latitude because the Earth is a spheroid rather than a sphere, and the meridians are ellipses. Since the parallels of latitude are considered to be circles, geodetic longitude is geocentric, and a separate expression is not used. The difference between geocentric and geodetic latitudes is a maximum of about 11.6' at latitude 45°.

Because of the oblate shape of the ellipsoid, the length of a degree of geodetic latitude is not everywhere the same, increasing from about 59.7 nautical miles at the equator to about 60.3 nautical miles at the poles. The value of 60 nautical miles customarily used by the navigator is correct at about latitude 45°.

MEASUREMENTS ON THE CELESTIAL SPHERE

1524. Elements of the Celestial Sphere

The **celestial sphere** (Article 1501) is an imaginary sphere of infinite radius with the Earth at its center (Figure 1524a). The north and south celestial poles of this sphere are located by extension of the Earth's axis. The **celestial equator** (sometimes called **equinoctial**) is formed by projecting the plane of the Earth's equator to the celestial sphere. A **celestial meridian** is formed by the intersection of the plane of a terrestrial meridian and the celestial sphere. It is the arc of a great circle through the poles of the celestial sphere.

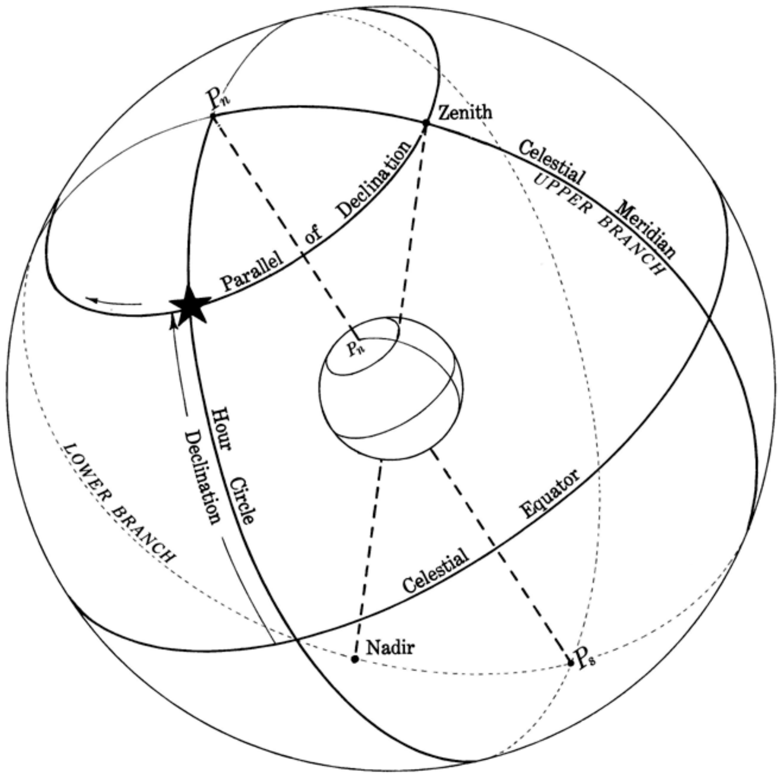


Figure 1524a. Elements of the celestial sphere. The celestial equator is the primary great circle.

The point on the celestial sphere vertically overhead of an observer is the **zenith**, and the point on the opposite side of the sphere vertically below him is the **nadir**. The zenith and nadir are the extremities of a diameter of the celestial sphere through the observer and the common center of the Earth and the celestial sphere. The arc of a celestial meridian between the poles is called the **upper branch** if it contains the zenith and the **lower branch** if it contains the nadir. The upper branch is frequently used in navigation, and references to a celestial meridian are understood to mean only its upper branch unless otherwise stated. Celestial meridians take the names of their terrestrial counterparts, such as 65° west.

An **hour circle** is a great circle through the celestial poles and a point or body on the celestial sphere. It is similar to a celestial meridian, but moves with the celestial sphere as it rotates about the Earth, while a celestial meridian remains fixed with respect to the Earth.

The location of a body on its hour circle is defined by the body's angular distance from the celestial equator. This distance, called **declination**, is measured north or south of the celestial equator in degrees, from 0° through 90°, similar to latitude on the Earth.

A circle parallel to the celestial equator is called a **parallel of declination**, since it connects all points of equal declination. It is similar to a parallel of latitude on the Earth. The path of a celestial body during its daily apparent revolution around the Earth is called its **diurnal circle**. It is not actually a circle if a body changes its declination. Since the declination of all navigational bodies is continually changing, the bodies are describing flat, spherical spirals as they circle the Earth. However, since the change is relatively slow, a diurnal circle and a parallel of declination are usually considered identical.

A point on the celestial sphere may be identified at the intersection of its parallel of declination and its hour circle. The parallel of declination is identified by the declination.

Two basic methods of locating the hour circle are in use. First, the angular distance west of a reference hour circle through a point on the celestial sphere, called the vernal equinox or first point of Aries, is called **sidereal hour angle (SHA)** (Figure 1524b). This angle, measured eastward from the vernal equinox, is called **right ascension** and is usually expressed in time units.

The second method of locating the hour circle is to indicate its angular distance west of a celestial meridian (Figure 1524c). If the Greenwich celestial meridian is used as the reference, the angular distance is called **Greenwich hour angle (GHA)**, and if the meridian of the observer, it is called **local hour angle (LHA)**. It is

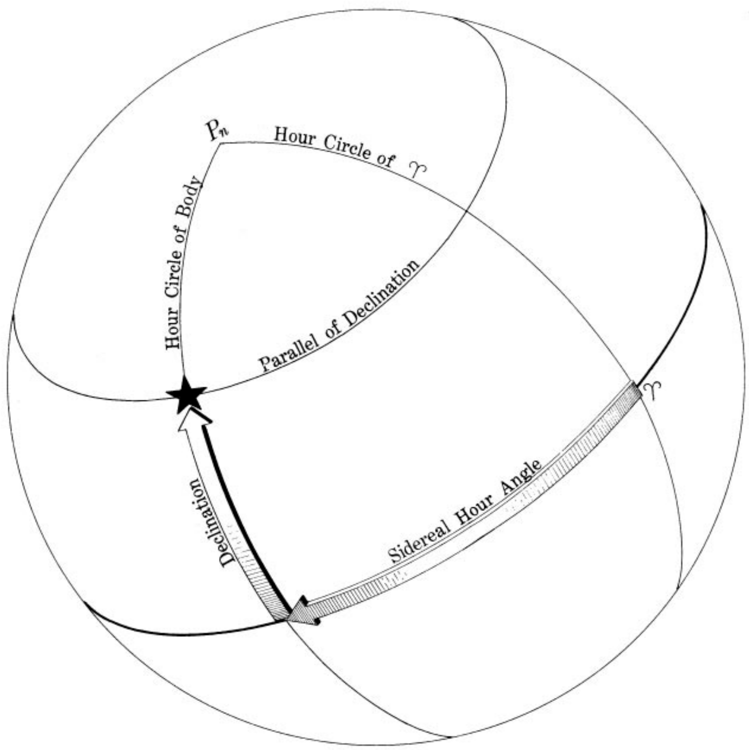


Figure 1524b. A point on the celestial sphere can be located by its declination and sidereal hour angle.

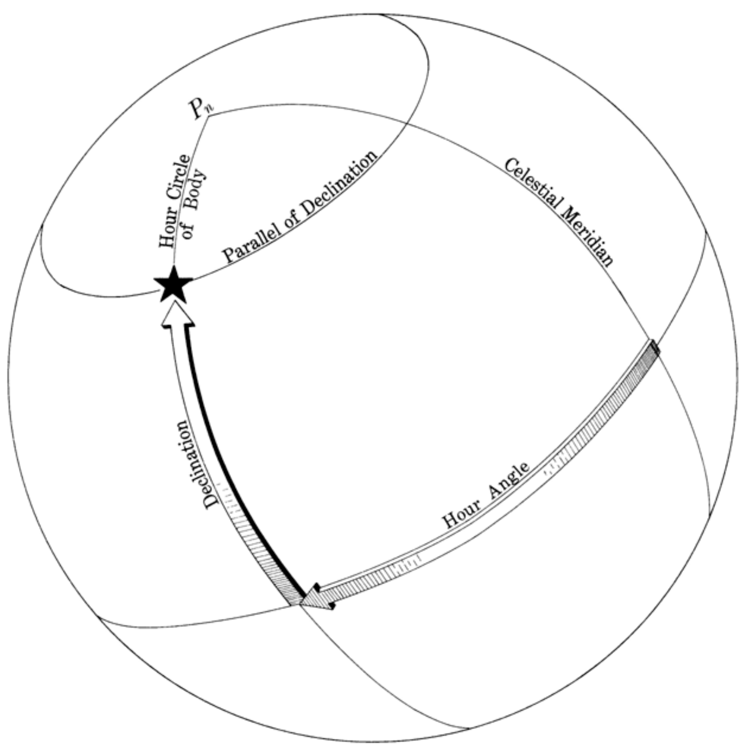


Figure 1524c. A point on the celestial sphere can be located by its declination and hour angle.

sometimes more convenient to measure hour angle either eastward or westward, as longitude is measured on the Earth, in which case it is called **meridian angle** (designated “t”).

A point on the celestial sphere may also be located using altitude and azimuth coordinates based upon the horizon as the primary great circle instead of the celestial equator.

COORDINATE SYSTEMS

1525. The Celestial Equator System of Coordinates

The familiar graticule of latitude and longitude lines, expanded until it reaches the celestial sphere, forms the basis of the celestial equator system of coordinates. On the celestial sphere latitude becomes declination, while longitude becomes sidereal hour angle, measured from the vernal equinox.

Declination is angular distance north or south of the celestial equator (d in Figure 1525a). It is measured along an hour circle, from 0° at the celestial equator through 90° at the celestial poles. It is labeled N or S to indicate the direction of measurement. All points having the same declination lie along a parallel of declination.

Polar distance (p) is angular distance from a celestial pole, or the arc of an hour circle between the celestial pole and a point on the celestial sphere. It is measured along an hour circle and may vary from 0° to 180°, since either pole may be used as the origin of measurement. It is usually considered the complement of declination, though it may be either 90° – d or 90° + d, depending upon the pole used.

Local hour angle (LHA) is angular distance west of the local celestial meridian, or the arc of the celestial equator between the upper branch of the local celestial meridian and the hour circle through a point on the celestial sphere, measured westward from the local celestial meridian, through 360°. It is also the similar arc of the parallel of declination and the angle at the celestial pole, similarly measured. If the Greenwich (0°) meridian is used as the reference, instead of the local meridian, the expression **Greenwich hour angle (GHA)** is applied. It is sometimes convenient to measure the arc or angle in either an easterly or westerly direction from the local meridian, through 180°, when it is called **meridian angle (t)** and labeled E or W to indicate the direction of measurement. All bodies or other points having the same hour angle lie along the same hour circle.

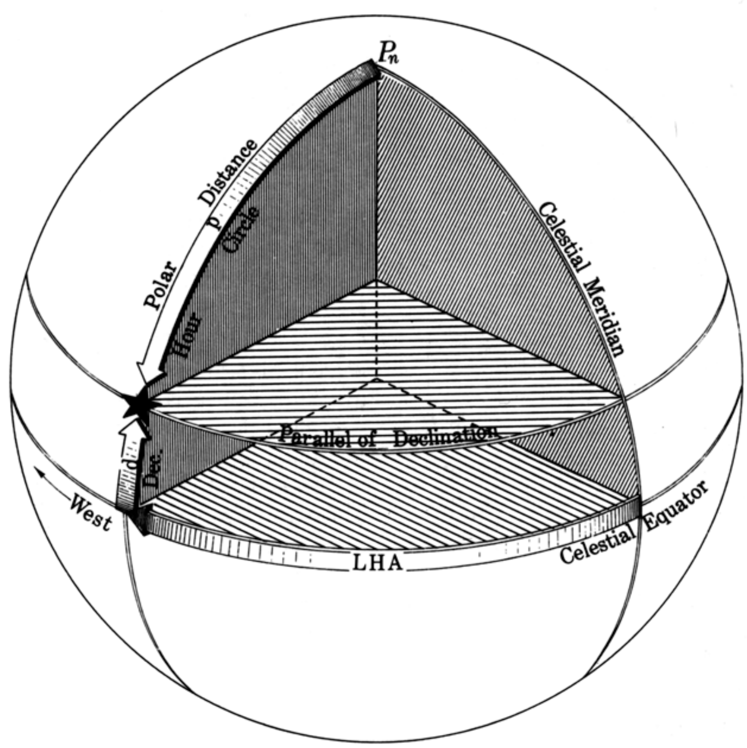


Figure 1525a. The celestial equator system of coordinates, showing measurements of declination, polar distance, and local hour angle.

Because of the apparent daily rotation of the celestial sphere, hour angle continually increases, but meridian angle increases from 0° at the celestial meridian to 180°W, which is also 180°E, and then decreases to 0° again. The rate of change for the mean Sun is 15° per hour. The rate of all other bodies except the Moon is within 3' of this value. The average rate of the Moon is about 15.5°.

As the celestial sphere rotates, each body crosses each branch of the celestial meridian approximately once a day. This crossing is called **meridian transit** (sometimes called culmination). It may be called **upper transit** to indicate crossing of the upper branch of the celestial meridian, and **lower transit** to indicate crossing of the lower branch.

The **time diagram** shown in Figure 1525b illustrates the relationship between the various hour angles and meridian angle. The circle is the celestial equator as seen from above the South Pole, with the upper branch of the observer's meridian (P_sM) at the top. The radius P_sG is the Greenwich meridian; P_s ♈ is the hour circle of the vernal equinox. The Sun's hour circle is to the east of the observer's meridian; the Moon's hour circle is to the west of the observer's meridian Note that when LHA is less than 180°, t is numerically the same and is labeled W, but that when LHA is greater than 180°, t = 360° – LHA and is labeled E. In Figure 1525b arc GM is the longitude, which in this case is west. The relationships shown apply equally to other arrangements of radii, except for relative magnitudes of the quantities involved.

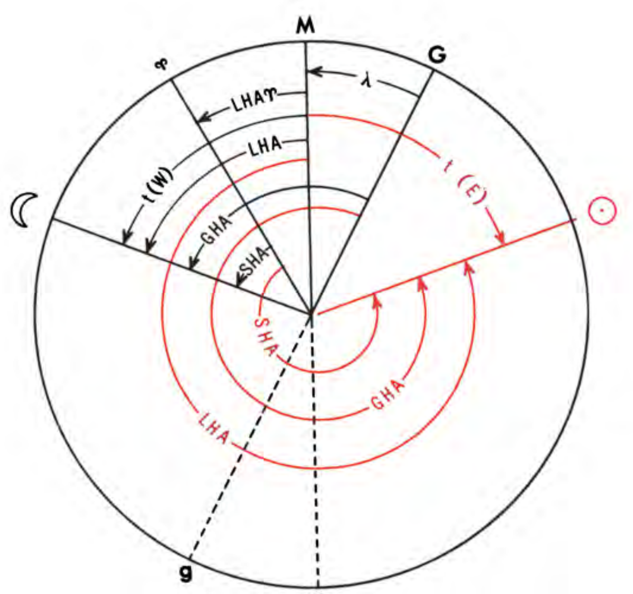


Figure 1525b. Time diagram.

1526. The Horizons

The second set of celestial coordinates with which the navigator is directly concerned is based upon the horizon as the primary great circle. However, since several different horizons are defined, these should be thoroughly understood before proceeding with a consideration of the horizon system of coordinates.

The line where Earth and sky appear to meet is called the **visible** or **apparent horizon**. On land this is usually an irregular line unless the terrain is level. At sea the visible horizon appears very regular and is often very sharp. However, its position relative to the celestial sphere depends primarily upon (1) the refractive index of the air and (2) the height of the observer's eye above the surface.

Figure 1526 shows a cross section of the Earth and celestial sphere through the position of an observer at A. A straight line through A and the center of the Earth O is the vertical of the observer and contains his zenith (Z) and nadir (Na). A plane perpendicular to the true vertical is a horizontal plane, and its intersection with the celestial sphere is a horizon. It is the **celestial horizon** if the plane passes through the center of the Earth, the **geoidal horizon** if it is tangent to the Earth, and the **sensible horizon** if it passes through the eye of the observer at A. Since the radius of the Earth is considered negligible with respect to that of the celestial sphere, these horizons become superimposed, and most measurements are referred only to the celestial horizon. This is sometimes called the **rational horizon**.

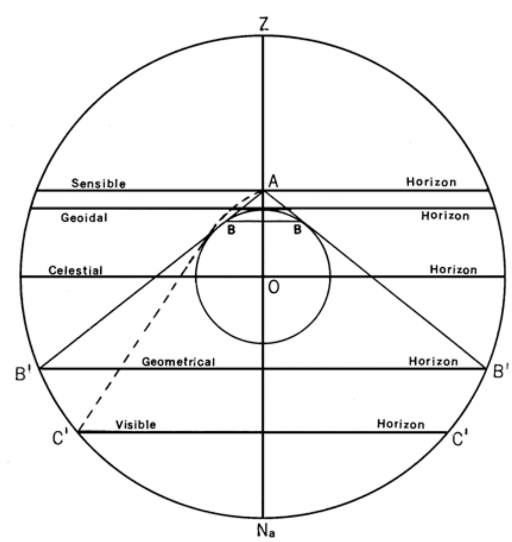


Figure 1526. The horizons used in navigation.

If the eye of the observer is at the surface of the Earth, his visible horizon coincides with the plane of the geoidal horizon; but when elevated above the surface, as at A, his eye becomes the vertex of a cone which is tangent to the

Earth at the small circle BB, and which intersects the celestial sphere in B'B', the **geometrical horizon**. This expression is sometimes applied to the celestial horizon.

Because of refraction, the visible horizon C'C' appears above but is actually slightly below the geometrical horizon as shown in Figure 1526. In Figure 1525b the Local hour angle, Greenwich hour angle, and sidereal hour angle are measured westward through 360°. Meridian angle (t) is measured eastward or westward through 180° and labeled E or W to indicate the direction of measurement.

For any elevation above the surface, the celestial horizon is usually above the geometrical and visible horizons, the difference increasing as elevation increases. It is thus possible to observe a body which is above the visible horizon but below the celestial horizon. That is, the body's altitude is negative and its zenith distance is greater than 90°.

1527. The Horizon System of Coordinates

This system is based upon the celestial horizon as the primary great circle and a series of secondary vertical circles which are great circles through the zenith and nadir of the observer and hence perpendicular to his horizon (Figure 1527a). Thus, the celestial horizon is similar to the equator, and the vertical circles are similar to meridians, but with one important difference. The celestial horizon and vertical circles are dependent upon the position of the observer and hence move with him as he changes position, while the primary and secondary great circles of both the geographical and celestial equator systems are independent of the observer. The horizon and celestial equator systems coincide for an observer at the geographical pole of the Earth and are mutually perpendicular for an observer on the equator. At all other places the two are oblique.

The vertical circle through the north and south points of the horizon passes through the poles of the celestial equator system of coordinates. One of these poles (having the same name as the latitude) is above the horizon and is called the **elevated pole**. The other, called the **depressed pole**, is below the horizon. Since this vertical circle is a great circle through the celestial poles, and includes the zenith of the observer, it is also a celestial meridian. In the horizon system it is called the **principal vertical circle**. The vertical circle through the east and west points of the horizon, and hence perpendicular to the principal vertical circle, is called the **prime vertical circle**, or simply the **prime vertical**.

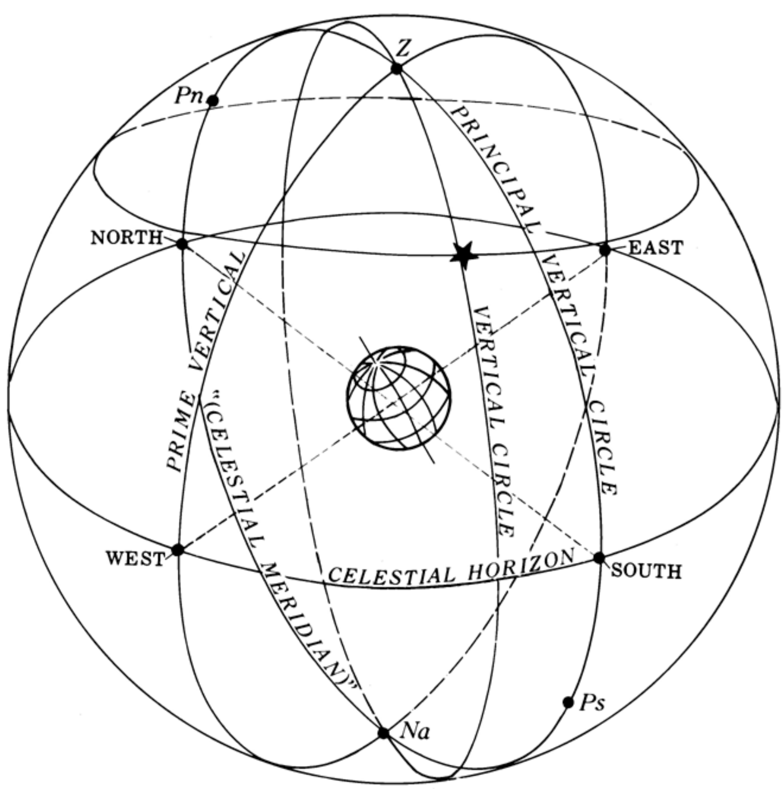


Figure 1527a. Elements of the celestial sphere. The celestial horizon is the primary great circle.