DMSTTIAC

Defense Modeling, Simulation and Tactical Technology Information Analysis Center

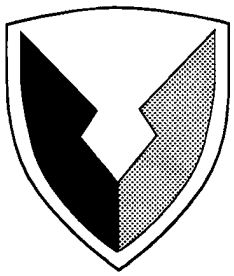
DMSTTIAC SOAR 96-03

De Physica Belli

An Introduction to Lanchestrial Attrition Mechanics Part Two

Bruce W. Fowler, Ph.D.
Technical Director and Deputy
Advanced Systems Concepts Office
Research, Development and Engineering Center
U.S. Army Missile Command
and
Adjunct Associate Professor
Department of Physics
University of Alabama in Huntsville

Published by:
DMSTTIAC
IIT Research Institute
7501 S. Memorial Parkway, Suite 104
Huntsville, AL 35802



September 1996

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	State-of-the-Art-Review; September 1996

4. TITLE AND SUBTITLE
De Physica Belli: An Introduction to Lanchestrian Attrition Mechanics, Part II

5. FUNDING NUMBERS
DAAH01-95-C-0310

6. AUTHOR(S)
Bruce W. Fowler, PhD.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
IIT Research Institute/DMSTTIAC
10 West 35th Street
Chicago, IL 60616-3799

8. PERFORMING ORGANIZATION REPORT NUMBER
DMSTTIAC SOAR 96-03

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Missile Command
ATTN: AMSMI-SW
Redstone Arsenal, AL 35898-5222

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This document is available only from DMSTTIAC, IIT Research Institute, 10 West 35th Street, Chicago, IL 60616-3799.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Physics and War are inextricable. While it is often recognized that war is often the generator of progress in physics, the role of physics in war is less often recognized. This work is dedicated to examination of the physics inherent in some of the processes of war. The framework for this examination is the Lanchester model of attrition. Part I of this work is primarily concerned with the basic foundation of this model . This volume, which comprises Part II, presents the mathematical and assumptive basis of the conjugate to the Lanchester theory, Bonder-Farrell Theory. It reviews the basic theory originally presented by these fathers of modern Lanchester Theory, and then the fundamental physical principles of the combat processes which lead to the basic definitions and descriptions of weapon systems and their use on the battlefield. Part III (to be published) will cover heterogeneous Lanchester theory and aggregation: and Part IV will cover special and advanced topics.

14. SUBJECT TERMS Attrition Theory, Lanchester, Lanchester Attrition Theory, Quadratic Lanchester Equations, Mixed Lanchester Equations, Attrition Differential Equations, Quantified Judgement Model, Epstein Model, Osipov's 3/2 Law, Constant Rate Attrition, Exponential Attrition, Helmbold's Attrition, Time Dependent Attrition, Range Dependent Attrition, Stochastic Duels, Stochastic Attrition, Stochastic Lanchester, Linear Attrition, Attrition Coefficients, Attrition Rates, One-on-one, Force-on-Force

15. NUMBER OF PAGES
248

16. PRICE CODE
$25.00

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unclassified

NOTICES

State of the Art Review. This state-of-the-art review has been published by the Defense Modeling and Simulation Tactical Technology Information Analysis Center (DMSTTIAC) as a service to both defense and non-defense agencies, academia and industry. DMSTTIAC is a DoD Information Analysis Center administered by the Defense Technical Information Center and operated by IIT Research Institute under contract DAAH01-95-C-0310. DMSTTIAC is funded by the Defense Technical Information Center (DTIC) and the Defense Modeling and Simulation Office (DMSO). The Director of DMSTTIAC is Mr. Hunter Chockley. The Contracting Officer is Ms. Cheryl Monteney, Defense Supply Center Columbus (DSCC), Columbus, Ohio. The Technical Monitor is Mr. Chalmer D. George, and the Alternate is Mr. Howard C. Race, AMC-Smart Weapon Management Office (SWMO), Attn: AMSMI-SW, Redstone Arsenal, Alabama 35898-5222.

Reproduction and Handling. Unlimited Distribution

De Physica Belli

An Introduction to Lanchestrian Attrition Mechanics

Bruce W. Fowler, Ph.D., MEL-1

Technical Director and Deputy
Advanced Systems Concepts Office
Research, Development, and Engineering Center
U. S. Army Missile Command

and

Adjunct Associate Professor
Department of Physics
University of Alabama in Huntsville

1 September 1996

This Page Intentionally Left Blank

Foreword

Warfare is as old as recorded human history. War has been especially prevalent in the last 500 years with the increasing conflict between large nation states. A great amount of analysis and thought has been given to the "Art of War". Nine principles of War have been defined: Objective, Offensive, Mass, Economy of Force, Maneuver, Unity of Command, Security, Surprise, and Simplicity. Despite these accepted principles, the science of war has remained elusive. Since World War II, investigators have searched for a theory on the physics of war--"De Physica Belli". Efforts have been more successful with the prominent rise of Operations Research as an analysis tool to assist combat operations. Dr. Bruce W. Fowler uses these modem analytical tools to seek the answer to the following question in this report--"Is there any scientific basis to describe the physics of war?" This report provides the answer to this question. His approach to a physics of war is the application of Lanchestrian attrition mechanics which first appeared in theory in the early 1900's.

Dr. Fowler continues with his consideration of Lanchester Attrition Theory with Part II of "De Physica Belli". In this Part, he considers the underlying theory of Attrition Rate Coefficients. The fundamental basis of this connection of engineering and physical principles and practice is Bonder-Farrell Theory. Upon this theoretical basis, Dr. Fowler reviews the physics and models that comprise this conjugate theory. This review includes the both terrain and weather, basic optical principles, and sensors. Next, the basics of search and detection theories are covered, followed by a consideration of the union of their methodologies. This if followed by an overview of the physics and models of weapons of various types, ranging from the primitive rock through modern smart weapons. Finally, the statistical mechanics of processes is reviewed to provide the theoretical basis for combining all of these topics in the calculation of attrition rate coefficients and their incorporation in modern attrition differential equations.

"De Physica Belli" is intended to be a general reference and introduction to attrition theory suitable for the combat soldier, the student-soldier. or the military analyst. The manuscript succeeds in that respect and provides a good overall summary of the state-of-practice in attrition theory through 1990. However, given the great advances in modeling, simulation and computational power since 1990, it would not be surprising to see future updates to this work. The mathematical tools of complexity theory, fractal dimensions, fuzzy logic, information theory and the power of scientific visualization of data in interactive computer simulations may offer new and exciting insights into the physics of war. These new developments will most certainly provide opportunities to conduct experiments in the science of warfare that go beyond the limitations inherent in the analysis of historical data.

Part II:

Conjugate Theory to the Lanchester Attrition Theory

This Page Intentionally Left Blank

Table of Contents

Table of Contents
(continued)

Table of Contents
(continued)

Table of Contents
(continued)

Table of Contents
(continued)

This Page Intentionally Left Blank

Chapter 0

Preface to Part II

Welcome to Part II! We are now in a position to build on the foundation that we laid in Part I. That epic collective, a work of some five years of research, collection, writing, and calculating with an interlude of effectively two years duration while I attended the U.S. Army War College, basically addressed three major areas:

- the mathematical and assumptive foundation of Lanchester attrition theory;
- the connectivity of Lanchester attrition theory with its counterparts and with higher resolution simulation; and
- an overview with some analysis of the historical data that contributes to the perceived and demonstrated level of validity of Lanchester attrition theory.

Of necessity, this presentation could neither cover all of the relevant material nor be completely satisfactory to my intended audience. Thus, the scholar will find that work deficient in the exhaustiveness of citation that an encyclopedia of Lanchester attrition theory would encompass; the soldier finds too much mathematics and too little military theory and, particularly, practice; the student will find too little precision and too much military anecdotage; and the analyst will find deficiencies between the presented material and the state-of- the-art.

That lack of satisfaction will continue in Part II. Herein we now consider the conjugate theory of attrition rate coefficients. Of necessity, this material is fundamentally technical, drawing deeply from probability theory and physics. The basic framework here will be the processes of combat executed by people, machines, and organizations. Because of inherent variations in the populations of these entities, and the interaction of the external environment (including Clausewitz's friction,) these processes must be considered to be implicitly stochastic. In keeping with our goal of serving as large a population as possible, the level of treatment of probability and statistics will be as simple as possible. This will likely not be completely satisfactory to either advanced Operations Research students or professionals. In general, I will try to keep

to a level consonant with basic understanding of the principles of the materials presented and common technical and analytical practice. The soldier may likely find this level too detailed.

This same philosophy will be applied in the presentation of physics material. The professional engineer and scientist will find this level of treatment simplistic and occasionally even misleading. My goal here is to provide an opportunity of understanding of the basic interplay between the physics and combat processes - elaboration and improvement to (or advancing) the state-of-the-art is left to professional excellence on their part.

Despite this admitted superficiality and simplicity, the material must be technical and fundamentally mathematical in character. This will likely be dissatisfactory for the soldier. I regret this necessity, but just as an engineer must study the mechanics of his discipline so too is similar study of value to the soldier. Where possible to the limits of my knowledge and experience, the military implications of the theory will be presented.

Overall, the intent is to provide a minimalist basis for the calculation of attrition rate coefficients, drawing a balance between the rigor of theoretical representation and the sensitivity impact of that rigor on the attrition rate coefficients. I well recognize that the superficial level of treatment here offers significant potential for misuse if applied unquestioning to the calculation of attrition rate coefficients. I also anticipate that such application will be made by many of the readers of this work either for learning or professional purposes. Assuring that these calculations will capture basic phenomenology and provide reasonable accuracy for learning and analysis is the metric of balance.

As in Part I, I want to acknowledge the mentorship and friendship of those who have contributed to my technical knowledge and skills: my co-worker, Don Peterson of the MICOM Advanced Systems Concepts Office; my doctoral advisor, Dr. C. C. Sung of the Physics Department of the University of Alabama in Huntsville; and colleague, Dr. Donald Gaver of the Operations Research Department of the Naval Postgraduate School.

Chapter 22

Bonder-Farrell Theory

22.1 Introduction

Compared to basic Lanchester attrition theory, [1][2] the basic conjugate theory of attrition rates is a relatively recent development.[3][4][5] Despite this relative youth, the theory is so compelling that an enormous amount of effort has been devoted to the use of this theory. Unfortunately, much of this development of the theory has been shrouded in classified military documents and the proprietary business base of Bonder' company.

The basic theory is freely available, however, and its use so direct and obvious at a simplistic level that we may easily expose it to the student. While the models that we discuss in this and later chapters may not be the most accurate or the most valid, they will cover the basic concepts and capture essential, in not the most exact, details. Happily, this lack of exactitude is not a compelling defect. Both the basic nature of rate theories (as we have seen in an earlier chapter,) and the nature of the conjugate theory, which is very forgiving of these deficiencies by its very nature, allow us to capture much of reality without the price of complete faithfulness.

22.2 Bonder's Equation

The fundamental embodiment of Bonder-Farrell Theory is contained in one fundamental equation, called Bonder"s equation. To arrive at this equation, we briefly return to consideration of the Ironman development.

If we have combat between two forces, one initially comprised of A_0 units and the other comprised of one invulnerable unit, the attrition differential equation describing the combat is

$$\frac{dA}{dt} = -\alpha B, \tag{1}$$

There is only one differential equation since, by definition, $\beta = 0$, and $B = 1$.

The solution to this differential equation is

$$A(t) = A_0 - \alpha t. \tag{2}$$

If we now specify that at time t, $A(t)$ = a identically, and at time $t+\tau$, $A(t+\tau) = a - 1$, then we have the obvious situation,

$$a = A_0 - \alpha\tau$$

$$a - 1 = A_0 - \alpha(t - \tau) \tag{3}$$

If we subtract the second equation from the first, then we get

$$1 = \alpha\tau. \tag{4}$$

This is a result that we have seen before, but now based on the development of rate theory that we saw in an earlier chapter, we may recognize as the expectation value of the interevent killing time. With this recognition and identification, equation 4 is Bonder's equation.

That is, for the quadratic Lanchester attrition differential equation, the attrition rate coefficient/function is the inverse of the expected time to kill (attrit) one unit.

As with most powerful theories, the elegance and simplicity of the theory have a measure of complexity. If we can describe the subprocesses that comprise the attrition process, we may calculate the rate of this attrition albeit that such calculations may be complex and complicated.

22.3 The Harmonic Mean

If we think of this definition of the attrition rate coefficient as

$$\alpha = \frac{1}{\langle \Delta t \rangle} \tag{5}$$

where t is the expectation value of the time for one firer to kill one target, then we may be struck that this is a strange way to define the attrition rate coefficient. We have seen that the basis for why this is so in the earlier chapter on rate theory and we know that the reason for this is that the fundamental quantity that describes the processes are the times to complete the processes and not the rates of the processes (which are derived from the times.)

Luckily, there is a statistical basis for just exactly this sort of thing, and it is known as the Harmonic Mean. If we take a collection of data of times $t_{i,j}$(such as we had in the chapter on rate theory,) then we have interevent times,

$$\Delta t_{i,j} \equiv t_{i,j+1} - t_{i,j} \tag{6}$$

which clearly imply interevent rate coefficients (if we may think of such things,) as

$$\alpha_{i,j} = \frac{1}{\Delta t_{i,j}} \tag{7}$$

and we may think of an expected rate coefficient in a linear (mathematical rather than a Lanchester) sense as

$$\langle \alpha \rangle = \sum_{i,j} \alpha_{i,j} \tag{8}$$

As we have seem, this tends to give us a rate coefficient that does not reproduce the processes (in a rate sense) very well because the rate is not the fundamental measure of the processes.

If instead, we define the expectation value of the attrition rate coefficient as

$$\frac{1}{\langle \alpha \rangle} = \sum_{i,j} \frac{1}{\alpha_{i,j}} \tag{9}$$

then by a little algebra, we may rewrite this as

$$\langle \alpha \rangle = \frac{1}{\sum_{i,j} \frac{1}{\alpha_{i,j}}} \tag{10}$$

which is just the definition of the Harmonic Mean. If we return to the definition of the $\alpha_{i,j}$, equation 7, and substitute it into this equation, then we reproduce the form of Bonder's equation

$$\langle \alpha \rangle = \frac{1}{\sum_{i,j} \Delta t_{i,j}}. \tag{11}$$

Some consideration in the literature has dealt with what the proper form of the attrition rate coefficient should be. We may summarize this in a very simple sense. If we erroneously think of the attrition rate coefficient as a fundamental rather than as a derived quantity, then the expectation of the attrition rate coefficient is indeed a harmonic expectation (or mean.) If we think of the interevent time as the fundamental quantity (which it is,) and the attrition rate coefficient as its inverse, then the expected attrition rate is just the inverse of the expected interevent time (and not the opposite.) This is a valuable lesson for statistical thinking. We need to ultimately be concerned with the statistics of the fundamental rather than the derived quantities.

22.4 Linear Attrition

This leaves us with one last consideration in this brief but vital chapter. This consideration is to repeat the Ironman Analysis for Linear Lanchestrian attrition

If we again adopt the assumptions of the Ironman, that the Blue force consists of one invulnerable unit-element, then the attrition differential equation is

$$\frac{dA}{dt} = -\alpha A, \tag{12}$$

If we now rewrite this equation and integrate it from time t to time t + τ,

$$\int_t^{t+\tau} \frac{dA(t')}{A(t')} = -\alpha \int_t^{t+\tau} dt'. \tag{13}$$

we get the solution

$$\ln\left(\frac{A(t+\tau)}{A(t)}\right) = -\alpha\tau. \tag{14}$$

We now specify that

$$\begin{aligned} A(t) &= a \\ A(t+\tau) &= a-1 \end{aligned} \tag{15}$$

and substitute this into equation 14 to get

$$\ln\left(\frac{a-1}{a}\right) = -\alpha\tau \tag{16}$$

which we may immediately rewrite as

$$\ln\left(1-\frac{1}{a}\right) = -\alpha\tau \tag{17}$$

If a $\gg$1, then we may use the standard expansion for the logarithm and rewrite equation 17 as

$$\frac{1}{a} \simeq \alpha\tau \tag{18}$$

We may further conveniently rewrite this equation as

$$\alpha \simeq \frac{1}{A(t)\tau} \tag{19}$$

from which we may say that the linear Lanchester attrition rate coefficient is approximately the inverse of the product of the force strength (of the targets) and the expected time for one firer to kill one target (so long as A(t) is large compared to one.)

Equation 19 could be the equivalent to Bonder's Equation for linear attrition. This would seem not to be the case however, and we may see this if we go back to the attrition rate differential equations. If we do this combining the linear attrition differential equation for Red,

$$\frac{dA}{dt} = -\alpha AB, \tag{20}$$

and equation 19, then we get

$$\begin{aligned} \frac{dA}{dt} &= -\frac{AB}{A\tau} \\ &= -\frac{B}{\tau}, \end{aligned} \tag{21}$$

which is the quadratic attrition differential equation. Clearly then, this is not the result that we want for defining linear attrition rate coefficients.

How do we correct this situation? What should the definition be? We may see this if we go back to equation 14 and rewrite it as

$$A(t+\tau) = A(t)\, e^{-\alpha\tau}. \tag{22}$$

If we now change our definition so that

$$A(t+\tau) = \frac{A(t)}{e}, \tag{23}$$

which gives us a definition of the linear attrition rate coefficient as

$$1 = \alpha\tau, \tag{24}$$

which is identical to Bonder's equation in form but now has a different meaning - it is the time to reduce the enemy force strength to e^{-1} (approximately 37%) of its previous value.

This is a good news - bad news story. It is good that the same equation form defines the attrition rate coefficients for both quadratic and linear attrition. It is bad that the definitions are different and therefore we have no consistent formalism for calculating attrition rates.

Actually, this will prove to be a mostly good news situation, as we shall see after we investigate the mechanics of the attrition subprocesses.

22.5 The Original Problem

The original problem addressed by Bonder dealt with the firing of a tank gun against a (tank) target. (The original work was sponsored by the U.S. Army's Armor School at Ft. Knox, KY.) At that time, gun fire control (laying and aiming) was accomplished using an optical telescope (gunsight) augmented by a reticle.

In this type of fire control, the gun (presumed to be boresighted with the telescope - a source of error,) was visually aimed at the target. Each successive shot was re-aimed. If the gun (and the telescope) had to be de-pointed from the target for reload, then each shot was essentially a first shot. (Happily this is not generally the case for tank gunnery except possibly for very high angle shots, but is commonly the case for artillery. Special mechanisms for returning the gun

to its original pointing have been developed and deployed, thus making each successive shot essentially (with error) a repeated shot.[1])

A key factor in aiming was whether the previous shot against the target (presuming a kill did not occur,) hit or missed. If the round hit, then its impact point could usually be observed and the accuracy of the next shot thereby increased (remember, we are speaking probabilistically here.) If the round missed, its impact point probably could not be observed (due to the flat, essentially horizontal trajectory,) and the accuracy of the next shot might decrease.

In this case, the expected time to kill a target may be calculated as

$$\tau = t_a + t_1 - t_h + \frac{t_h + t_f}{p_{k|h}} + \frac{t_m + t_f}{p_{k|m}} \left(\frac{1 - p_{h|h}}{p_{k|h}} + p_{h|h} - p_1 \right), \tag{25}$$

where:

t_a = time to acquire the target,
t_1 = time after acquisition to fire the first round,
t_h = time to fire a succeeding round given the preceding round was a hit,
t_m = time to fire a succeeding round given the preceding round was a miss,
t_f = time of flight of the round,
p_1 = probability of hit for the first round,
$p_{h|h}$ = probability of hit of a succeeding round given the preceding round hit,
$p_{h|m}$ = probability of hit of a succeeding round given the preceding round missed, and
$p_{k|h}$ = probability of kill given a hit. All of the times are expected values.

The derivation of this expression is given in the literature. We do not reproduce it here except to note that a key feature is to note that in a sequence of n shots, the order of hits and misses is not important, only the number of occurrences of each in the sequence.

As a preliminary to discussion later, we do note that several of the variables in this expression are dependent on range and other factors. In particular, the acquisition and flight times, probabilities of hit and (possibly) kill, are range dependent.

22.6 References

[1] Lanchester, Frederick W. **Aircraft in Warfare: The Dawn of the Fourth Age**, Constable and Company, LTD., London, 1916.

[2] Osipov, M., "The Effect of the Quantitative Strength of Fighting-Sides on the Losses", *Voennie Shornik*, 3-7, 1918, Drs. Robert L. Helmbold and Allan Rehm, trans., U.S. Army Concepts Analysis Agency, Bethesda, MD, Research Paper CAA-RP-91-2, September 1991.

[1] We shall designate the first shot fired at a target by that name, and subsequent shots as repeated shot. Depending on the system of tracking the round, if any, accuracy may stay constant, increase, or decrease (usually due to the firer coming under fire or exhaustion.)

[3] Bonder, Seth, "The Lanchester Attrition-Rate Coefficient", *Operations Research*, 15 221-232, 1967.

[4] Bonder, Seth, "The Mean Lanchester Attrition Rate", *Operations Research*, 18 179-181, 1970.

[5] Bonder, Seth, and Robert Ferrell, eds., "Development of Analytical Models of Battalion Force Activities", Systems Research Laboratory, Department of Industrial Engineering, The University of Michigan, Ann Arbor, MI, September 1970 (AD 71 4677), and references therein.

This Page Intentionally Left Blank

Chapter 23

Terrain and Weather

23.1 Introduction

In his bestselling history book, **The Birth of the Modern**,[1] Paul Johnson advances the argument that the social structure that we think of as the modern world evolved, in a punctuated sense, not around the turn of the twentieth century, but during the period 1815 - 1830. It is not our purpose here to restate his thesis, nor to advance or refute it, but to use it to illustrate the impact of terrain and weather on warfare.

At the beginning of this period, two major battles were fought that were the culminations of an era of warfare, the Battle of New Orleans and the Battle of Waterloo. The first was militarily almost meaningless, but politically crucial; the second militarily important, but politically anti-climactic. The Battle of New Orleans, actually fought after the conclusion of the War of 1812 by the Treaty of Ghent, provided the emotional culmination of the war that enabled political and social conditions which permitted both the Americans and the British governments and peoples from mutual competition between ex-colony and ex- ruler to increasing mutual cooperation and other pursuits of competition. For the Americans, this was eastward expansion; what became later known as Manifest Destiny. For the British, this was worldwide naval presence, if not domination; what became known as the Second British Empire.

The Battle of Waterloo, also fought after the "conclusion" of the Napoleonic Wars was the emotional termination both of a strange coda-like campaign of morbid Napoleonic reemergence and of the political and social dynamics of the Napoleonic Age. It ushered in an era of protracted European "peace" (Balance of Power), and new cultural conditions (Romanticism), that enabled European technological and social development for much of the century.

This is all very nice, but what does it have to do with the analysis of warfare? In particular, what does it have to do with terrain and weather? These two battles were effectively the last ones in a continuous period of warfare where terrain and weather were the dominant, even determining, factors in where, when, and how battles were fought and wars prosecuted. Some of this dominance is demonstrated in both of these battles.

In the case of the Battle of New Orleans, the city itself, both as the geographic gateway to the Mississippi River interior of the North American continent and as the economic focus of both (then) Western America and the Caribbean, was at once both the political and operational Center of Gravity. This focused the military situation, simplified the American military problem by effectively negating the British advantage of seaborne mobility that had permitted them to conduct painful but militarily ineffective raids from Washington to Mobile.

Although the site of the battle was largely determined by terrain at the operational level, its outcome was influenced, if not determined, by weather. On the day of the battle, there was a heavy ground fog. This had the effect of severely restricting the British commander's view of the battlefield. Even though the terrain was relatively open, the fog has two important tactical effects. The British commander was unable to make use of his long range rocket artillery because the visual restrictions of the fog prevented stand off targeting and long range damage assessment. Likewise, he was unable to either see or communicate rapidly with his cavalry which had the mission of turning the American flank. Losing patience with his cavalry and the use of his technology advantage, the British commander settled on an ill advised option; a frontal infantry assault on a basis of firepower equality against a well prepared defensive position.

The battlefield at Waterloo is not particularly strategically located, except to shield the approach to Brussels. It is also about a day's march from Quatre Bras which is as close as the battle could have been fought. The terrain is hilly and somewhat wooded so that a commander would be unable to see the whole battlefield from any one spot. Further, there was a violent storm the night before the battle, drenching the ground and making movement difficult. As a result, the battle did not begin until midmorning. As a result, Napoleon did not have enough time to defeat Wellington before Blucher arrived to decisively reinforce the allied forces.

These situations are typical of battle prior to the explosion of technology changed some things. (As we shall see.) Up to this point, commanders first fought the terrain and the weather, and then the enemy. Today, the effects of terrain and weather, especially for mechanized forces, are mitigated by technology although extreme weather conditions may still halt (or delay) military activities and dismounted force still largely operate much as they have since war began.

To conclude this description, we note that these two battles were the end of an unbroken series that technology did not mitigate the effects of terrain and weather. By the next major conflict, the American Civil War, technology had changed the effects of terrain and weather. The railroad and the telegraph had accelerated strategic and operational mobility and communications. Further, in a quirk of usage, the conoidal bullet drove artillery off the battlefield and terrain made it an indirect fire arm.

23.2 Terrain

Terrain has two primary effects of combat: it obstructs movement and it limits the range of

vision.[1] Both have adverse impact on optempo. The obstruction of movement is of secondary importance to a theory of attrition and a detailed treatment of an engineering and/or tactical theory of movement is therefore beyond the scope of this book. We shall therefore limit ourselves to a discussion of some basic physics of vehicle movement, and of the historical evidence appertaining to tactical movement. For greater detail on the simulation of terrain and movement, the student is referred to the handook of Youngren.[2]

Alternately, the limitation of visual range is crucially important to attrition, and while the basic theory is simple, its effects are far reaching and shall be frequently seen in later chapters of this book. Thus, while we shall spend considerable attention to terrain effects on attrition, we will first examine simple models of terrain.

23.2.1 Terrain Models

If we are dealing with a relatively small area on the earth's surface, then we may approximately represent terrain in a rectilinear $x-y-z$ coordinate system. The general functional form is

$$z = f(x, y), \tag{1}$$

where:

z is the elevation, and

x and y are perpendicular to elevation.

In this model, $f(x, y)$ defines a surface that (for ground vehicles,) units and elements move on. Of course, air vehicles operate above this surface, so $f(x, y)$ represents the minimum elevation that they can operate at.

If we must operate over a larger area, so that the (approximately) spherical shape of the earth must be taken into account, then we must use a spherical coordinate system, with origin at the Earth's center. This terrain model has the form

$$r = g(\theta, \phi), \tag{2}$$

where:

r is the distance from the Earth's center and

θ, ϕ are Longitude and Latitude.

If this model represents the land (solid) surface of the Earth, then a supplemental model

$$r = h(\theta, \phi, t), \tag{3}$$

represents the sea (liquid) surface of the Earth. This supplemental model is time dependent to account for the tidal motion of the seas.

Which of these models do we use? In practice, we use both, but selectively. The Earth is a large place and one must travel considerable distance for the effects of its sphericity to be

[1]Before some wag makes the comment that this doesn;t apply to naval problems, let me mention that weather tends toinduce terrain (waves) at sea.

appreciable. How far is this? To use a commonplace analog, that distance is about the width of a time zone.

Now we must translate this into terms of attrition and mobility. If the distances we are to consider travel over or between units, and weapons ranges, are less than the width of a time zone, then the flat earth model is valid. This is the case for most of the tactical situations that we want to consider. Only when we want to consider cumulative operational or strategic movement, or long range weapons (e.g., such as those that come under the SALT treaty (i.e., range $>$ 500 km,)) do we need to worry about the spherical reality of the Earth. Thus, we may concentrate on the "flat" earth model as a general case of interest.

Despite the fact that we have depicted the "flat" earth model as continuous in equation 1, the sheer size of the data required for such a representation is prohibitive. In practice, therefore, the most common data representation of terrain is quantized in increments of x and y. Thus, the most common representation of terrain data is in the form

$$z_{i,j} = f\left(x_i, y_j\right), \tag{4}$$

so that if we know the origin (x_0, y_0) and the increments Δx and Δy, then we only need to catalog the $z_{i,j}$.

In practice, one of the resolution measures of a combat simulation is the resolution and scope of the terrain represented in the simulation. The smaller the size of Δx and Δy, the higher the terrain resolution and the greater the value of $n\Delta x\ m\Delta y$, where n and m are the number of x, y terrain increments represented, the wider the scope of the simulation.

Obviously, the greater the value of nm, the greater the data storage required in the simulation (and the greater the number of terrain related calculations). This factor is thus important in terms of the size of the computer hardware that the simulation can be executed on and the time it takes to execute the simulation. Thus, the scope and resolution of the terrain data bases are important criteria in the design of combat simulations.

There are two approaches to addressing these design criteria: speed (or time) and storage (or size). The speed approach seeks to have simulations execute within some standardized bounds (e.g., four hours of combat time per hour of computer time,) so that terrain resolution tends to be fixed and storage requirements are accommodated (again within bounds,) for the terrain scope required for different scenarios.

The storage approach tends to fix the number of terrain data that are stored and to vary the resolution to accommodate different scenarios.

The speed approach is intended to permit a schedule of executions and is therefore important when a large number of executions of the simulation is important. In this approach, obviously, the size of the supporting computer equipment must not be a primary consideration in design. This design approach tends to be associated with stochastic simulations where repeated executions are necessary for statistical significance.

The storage approach is driven by consideration of the size of the supporting computer equipment. There is greater variation in execution speed and thus this design approach tends to be

associated with deterministic simulations that do not require repeated executions.

This points up another difference between the two types of models/ simulations. Deterministic simulations tend to execute faster than stochastic simulations both in the actual and effective senses. Because the mathematical calculations tend to be numerically faster for deterministic simulations than for stochastic simulations, the deterministic simulations tend to execute a scenario faster, so they are faster actually. Further since stochastic simulations must be executed several times to assure statistical significance, deterministic simulations tend to be effectively faster as well.

If we put this in the context of fixed computer assets; that is, fixed in terms of program size and calculation speed; then the former tends to be the critical size for deterministic simulations while the latter tends to be the critical limit for stochastic simulations. In this context, deterministic simulations, being relatively unconcerned with calculation speed, tend to be complex with large inherent data sets that push the program size limits of the computer. On the other hand, stochastic simulations, being completely driven by calculation speed, tend to be relatively simple with small data sets.

We must comment that this tends to be true of all types of simulations, not just combat simulations. The size and complexity of stochastic simulations (or equivalently, the scope of their use,) tends to be limited by effective execution time, while the size and scope of deterministic simulations tends to be limited by computer size (or equivalently, computer accessibility.)

Having now completed our philosophical digression on the nature of combat simulations *vis a vis* terrain modeling, we may now return to the basic question. How are terrain models used?

Given that we have quantified the elevation representation of terrain, i.e., the $z_{i,j}$, then how do we calculate the elevation of a unit (or element) located at coordinates x, y?

The simplistic representation that we may use is a grid rectangle representation. In this representation, the point x_i, y_j where the elevation point $z_{i,j}$ is defined sits at the center of a rectangle with sides Δx and Δy. In this representation, the elevation of the unit at x, y is that of the elevation data point that it is closest to. In principle, we could calculate the square of the distance of the x, y point from each center,

$$d_{i,j}^2 = (x_i - x)^2 + (y_j - y)^2 , \tag{5}$$

for each of the i, j in the terrain data set, but there is an easier (and more economical) way, since this approach would require numerous arithmetic operations.[2] Luckily, there is a more economical way. If we compute the two quantities

$$I = \frac{x - x_0}{\Delta x} \tag{6}$$

[2]There being mn i,j pairs in the terrain data set, there are mn calculations of distances, equation 5. Coupled with this are $mn - 1$ comparisons to determine the minimum distance and temporary storage of the minimum distance values of i, j. This gives a minimum of $2mn - 1$ calculations.

$$J = \frac{y - y_0}{\Delta y},$$

and round correctly, then the resulting values are the appropriate values of i and j.

One of the disadvantages of this method may be seen by examining Figure 1, which is a terrain map based on this model. The student may immediately note the step-like changes in elevation. As we shall see, this can have adverse effects on the calculation of LOS.[3]

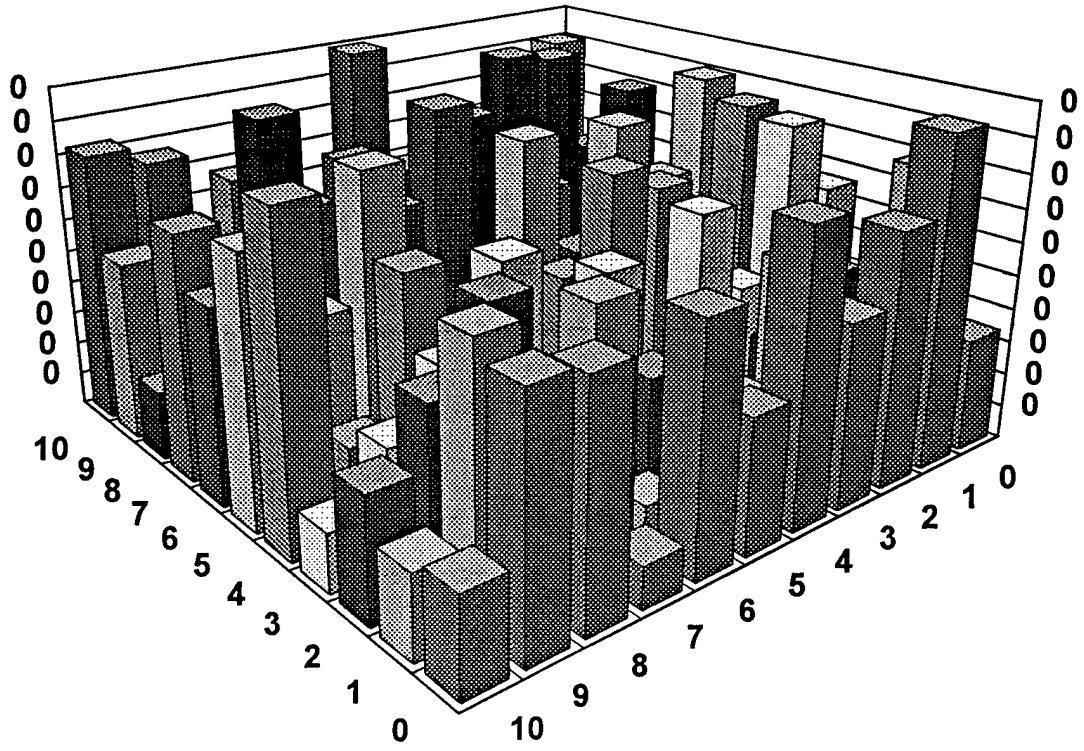


Figure 1: Grid Rectangle Terrain Model

A slightly more costly, but vastly more accurate model is the bi-linear functional model. This model makes use of the I, J given by equation 6, but always rounded down. The resulting values $[I]$ and $[J]$ (the rounded values of I, J,) define the (relative) origin used in the model. The elevation predicted by this model has the functional form

[3]This sentance points up a common problem with acronyms. How does one pluralize (possessivize) acronyms where the last word in the phrase that the acronym represents is not one which changes? The ploral of lines-of-sight is lines-of-sighr. Is the plural of LOS LOSs or LsOS? Isn't this a ridiculous consideration? Clearly the answer is to treat acronyms of this type as collectives (i.e., the plural of LOS is LOS,) and let the reader/student infer the difference from context. But do we apply this rule to all acronyms or just these?

$$z(x,y) = a + b\chi + c\eta + d\chi\eta, \tag{7}$$

where:

$$\begin{aligned} k &\equiv [I] \\ l &\equiv [J] \\ \chi &\equiv \frac{x - k\Delta x}{\Delta x} \\ \eta &\equiv \frac{y - l\Delta y}{\Delta y} \\ a &\equiv z_{k,l} \\ b &\equiv z_{k+1,l} - z_{k,l} \\ c &\equiv z_{k,l+1} - z_{k,l} \\ d &\equiv z_{k+1,l+1} - z_{k+1,l} - z_{k.l+1} + z_{k,l}. \end{aligned} \tag{8}$$

A map generated using this model is shown in Figure 2. The superiority is readily evident. Studies comparing these two models in terms of the results of variable terrain resolution combat simulations have shown that the number of elevations in the terrain data set of the grid rectangle model must be 16 times as large as the one for the bi-linear model to achieve equivalent accuracies. That is, the linear resolution $\Delta x, \Delta y$ of the data in the grid rectangle model must be $\frac{1}{4}^{th}$ that of the data in the bi-linear model.

Finally, we shall consider one more model. This model, developed by Price[3] is based on curve fitting data on a non-linear basis. The basic algorithm calls for the selection of several of the key terrain features, usually local maximum heights, as the centers of fitting functions. Usually these fitting functions are general bi-normal functions. If there are k fitting functions $f_k(x,y)$, then the elevation at a point x,y within the total area represented is

$$z(x,y) = MAX\left\{f_k(x,y)\right\}. \tag{9}$$

This functional form is used to define the error quantity that drives the iterated non-linear curve fit.

This fitted model continues the trend that we have already seen. It can reduce the amount of program stored data (to basically 5k data points) at the price of more numerical calculation to use the model. It has the further disadvantage of requiring a skilled preprocessing analysis to curve fit the data to the fitting functions. For simulations that must be executed in a computer size constrained environment, (Price originally developed the model in the early days of 8-16Kb RAM desktop computers,) this model can be exceptionally accurate.

There are other types of terrain models, but these three models capture the basic ideas and merits of these models. We shall therefore not concern ourselves with any further terrain models beyond a discussion of data sets and supplemental models.

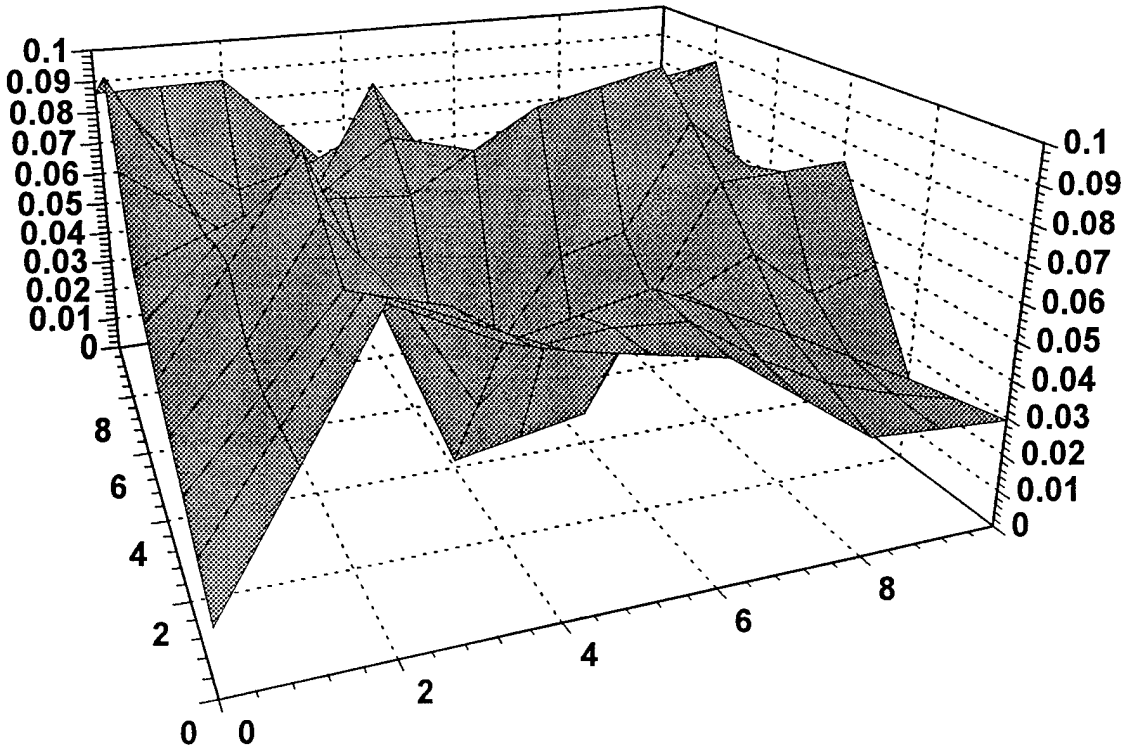


Figure 2: Bi-linear Interpolation Terrain Model

Terrain data sets are derived from topographic maps (or *visa versa*?). In the United States, the data sets for combat simulations are a monopoly; they all come from the Defense Mapping Agency (DMA). In years past, these data sets were somewhat hard to come by, so studies tended to be limited to only certain geographic regions. Recently however, as the result of advances in technology, DMA has generated a world wide database of terrain elevations and published this database on CD-ROM. Of course, data sets at very high resolution $1\text{m} < x, y < 100\text{m}$ are still difficult to obtain, but their applications tend to be limited as well.[4]

Of course, terrain elevation data is not the whole story. It is also desirable to know the locations of roads, streams and lakes, and vegetation. These generally exist in the form of supplemental data sets. Roads are generally represented by straight line segments with the database consisting of the x, y coordinates of line segments and a taxonomic variable indicating the type of road in terms of surface and width or carrying capacity. Since, in general, the end of one line segment is the beginning of another, and road type continues extensively, an obvious economy is possible.

Streams (and rivers) are represented in a similar way except that in addition to width, depth,

[4]Of course, vegetation and, in particular, urban data are still limited.

bottom conditions, and flow rate (possibly as functions of time of year and flood state) are included. Lakes may similarly be characterized either by negative depth (being laid on top of the terrain elevation model,) or positive depth (interpreted below the elevation model.)

Vegetation and terrain type are often characterized as taxonomic variables paired with the elevation data. These may indicate the type of vegetation in terms of mean height and passability, and type of soil and its roughness, Obviously, vegetation height may, depending on type, be added to the basic elevation.[5]

23.2.2 Terrain and Mobility

Both men and ground vehicles interact with terrain in similar manners; they exert a force (gravity) on the surface, compressing it. This has somewhat different effects on the two. Four factors tend to determine the effect of terrain on movement: softness of the terrain (actually hardness in a materials sense); the adhesion of the soil; its coefficient of friction; and the local slope of the terrain.

If the soil is too hard, damage may be done to the human muscular and skeletal systems, resulting in an accidental failure rate (process). If the soil is too soft, the man will sink in, necessitating effort to extricate and lift the feet, doubly slowing the rate of movement by the time required for the extra care in movement, and the additional expenditure of energy. If the soil is especially adhesive (as is the case with some types of mud,) then additional mass will accumulate on the feet, further slowing movement.

The coefficient of friction is important since there is a slight natural sliding component to walking. If the coefficient of friction is too low, then the terrain is slippery, control is difficult, and slipping may occur. If the coefficient is too high, commonly associated with the muddy conditions described above, but otherwise also possible, then additional care and effort is required. Both conditions will slow movement.

The local slope of the terrain is also a factor that unites with friction to slow movement (even disrupt it.) On an increasing slope, additional energy must be expended, slowing movement. Too low a coefficient of friction will further slow, or even preclude movement. On a decreasing slope, the situation is even worse since gravity now tends to accelerate rather than decelerate the walker, and friction is the force which (still) impedes motion, making control possible. If the force of friction is too large for the slope, then gravity overwhelms control, and the downward slope becomes impassable. It is thus possible for a slope to be open to movement going up, but not down.

The situation with vehicles is similar. While the vehicle has a much larger "footprint", it also has a much larger mass. Further, the exertion of "stopping" force components is much more limited for tires and treads than for feet. Thus, the vehicle may compress the soil more, creating a deeper hole, and unlike the man, it cannot just lift its foot - it must drive out of the hole. This requires at least the relative elevation of the vehicle with the expenditure of energy, with the

[5]Terrain and vegetation roughness are also important for lower atmosphere transport and diffusion. We shall briefly consider this in a later chapter on optical countermeasures.

effect of slowing motion and reducing range.

Besides the similar effects of friction and slope, local small variations in elevation longer than the wheel size but (approximately) smaller than the vehicle can couple harmonically or impulsively with the vehicle's suspension system, further reducing speed and range. Characteristics of this sort may be general, or highly direction specific but capturing them in a terrain data base is generally difficult because of the resolution scales involved.

What does this mean in practice? In many combat simulations, it means that vehicle (and walking) speed is idealized by some flat terrain value (with road speed as a separate value,) that is degraded by soil condition and local slope. This degradation will in general be a simple, usually empirical formula.

23.2.3 Fractal Terrain

A detailed consideration of the mathematical theory of fractal dimensionality is beyond the scope of this work.[4] We may however, consider a special application of this theory in the context that we have established for terrain models.

Before proceeding, we must comment that the algorithm to be presented has a stochastic component. Thus it will produce a different result each time it is executed. If consistency is a necessary feature of the application that this algorithm will be used to support, then it will be necessary to generate the terrain data points and store them. The amount of data that must be stored is, of course, a direct function of the final resolution of terrain data generated.

The algorithm is similar to the bi-linear model in that it starts with terrain data of some resolution and interpolates this data to some higher resolution. The key feature to this algorithm is that it adds a stochastic fractal component to this interpolation. This has the effect of making the interpolated terrain have a more natural "feel". This algorithm thus adds a component that more naturally represents the finer structure of actual terrain. If we view a map of this terrain, it will look more natural than the equivalent bi-linear or grid rectangle model.

There us a price to be paid for this "feel", over and above the cost of the data storage. While the data will be more realistic in feel, it will not be faithful to the actual terrain. Even if we select the fractal parameter to be the same as the actual terrain, that is, to have the same fractal dimensionality, the stochastic component will only ensure that the generated terrain has the same statistical properties as the actual terrain. It will not have the same structure as the actual terrain. Thus while the generated terrain data will have a more natural appearance, it will not be the same as the actual terrain.

If this is the case, then why would we want to use this technique? Clearly, if we have actual terrain data of the desired resolution, then we would not. Similarly, if our computer equipment does not have the storage capacity to support the mass of resulting data, we would not want to use the algorithm. If the terrain data base that we do have of the region is not of succinctly high resolution and data of that resolution is not available; if we have the storage capacity; and if we are not going to make comparison to the actual terrain (as in comparing simulation executions to field exercises;) then we may choose to use the algorithm to more "naturally" interpolate the

actual data that we do have to higher resolution.

Let us now consider the algorithm itself. Since each iteration of the algorithm doubles the linear resolution ($x, y \rightarrow \frac{\Delta x}{2}\frac{\Delta y}{2}$) and thus increases the number of data points, we have only to describe one application of the algorithm.

Given the initial terrain data set $z_{i,j}, i = 0..m, j = 0..n$, defined at points $x = i\Delta x$, $y = j\Delta y$, then we immediately map these into a fractal interpolated data set $z'_{k,l}$ with defined points

$$z'_{2i,2j} = z_{i,j}, \tag{10}$$

which are the even-even points. We may then define the even-odd and odd- even points by the interpolations

$$\begin{aligned} z'_{2i+1,2j} &= \frac{z_{i,j} + z_{i+1,j}}{2} + \frac{\mu}{2\sqrt{2^H}}, \\ z'_{2i,2j+1} &= \frac{z_{i,j} + z_{i,j+1}}{2} + \frac{\mu}{2\sqrt{2^H}}, \end{aligned} \tag{11}$$

where

μ is a normally distributed random variable, and

H is a fractal parameter that is related to the fractal dimension D by

$$D = 3 - H. \tag{12}$$

If H = 1, then the result is a Drunkard's Walk.

The odd-odd points are given by the interpolation

$$z'_{2i+1,2j+1} = \frac{z_{i,j} + z_{i+1,j} + z_{i,j+1} + z_{i+1,j+1}}{4} + \frac{\mu}{2\sqrt{2^H}} \tag{13}$$

The algorithm is used by calculating all of the interpolated points using equations 11 and 13, generating a different particular value (draw) of the random variable μ for each interpolation. We may approximate these values from a uniformly distributed random variable R by the formula

$$R \simeq \frac{1}{2}\left(1 \pm \sqrt{1 - e^{-\frac{2\mu^2}{\pi\sigma^2}}}\right). \tag{14}$$

where the + branch and a positive value of μ are used if $R > 0.5$, negative in both cases if $R < 0.5$. (This formula has been described in an appendix of Volume I.)

There is one more wrinkle to this algorithm. Every time it is repeated to increase the number of terrain data points by four, the denominator of the factor μ in the interpolation equations is multiplied by $\sqrt{2^H}$.

23.2.4 Terrain and Vision

From an attrition standpoint, the most important effect of terrain is to restrict the LOS between observer (searcher) and target. This effect is present even for a bald spherical earth.

Let us consider an observer with sensors at height h_O and a target of height h_T, both located on a bald, but spherical earth of radius R_e.[6] The relevant LOS are described by right triangles inscribed on the sphere. For example, the observer has LOS for the whole height of the target out to a range d defined by a right triangle with sides d and R_e, and hypotenuse $R_e + h_O$. By the Pythagorean theorem, the range d is given by

$$\begin{aligned} d &= \sqrt{(R_e + h_O)^2 - R_e^2} \\ &= \sqrt{2R_e h_O + h_O^2}. \end{aligned} \tag{15}$$

For an observer height of 2m, $d \sim 5$ km. If we ignore the h_O^2 factor, then this equation is approximately

$$d \simeq 3.5\sqrt{h_O}, \tag{16}$$

where h_O is in m, and d is in km.

By the same type of geometry, the target is still partly visible out to a range d' given by

$$d' = \sqrt{(R_e + h_O)^2 - R_e^2} + \sqrt{(R_e + h_T)^2 - R_e^2}, \tag{17}$$

which for $h_T = 2$ m, $d' \sim 10$ km. Similarly the range at which half of the target is in LOS is

$$d'' = \sqrt{(R_e + h_O)^2 - R_e^2} + \sqrt{\left(R_e + \frac{h_T}{2}\right)^2 - R_e^2}, \tag{18}$$

which for the conditions outlined is $d'' \sim 7.5$ km.

Admittedly, this is for a bald, spherical earth with no terrain height variations. When we add these variations, the situation becomes more complex. Variations in terrain height may block LOS between observer and target. There are two approaches to this problem. One is deterministic and depends on using an explicit representation of terrain, such as a terrain data base; the other is stochastic, and depends only on having statistical knowledge of the terrain.

The deterministic approach assumes that we have a terrain representation (data base,) and know the x, y coordinates and heights of the observer and target. We then define the parametric equations of the LOS as

$$\begin{aligned} z(\eta) &= z_O + h_O + (z_T + h_T - z_O - h_O)\,\eta \\ x(\eta) &= x_O + (x_T - x_O)\,\eta \\ y(\eta) &= y_O + (y_T - y_O)\,\eta, \end{aligned} \tag{19}$$

[6] The mean radius of the earth at the equator is approximately 6371011m or about $6x10^6$m.

where η: is a parametric variable on [0,1].

Next we want to know if any terrain between x_O, y_O and x_T, y_T located at point x, y has height z greater than or equal to the LOS. In principle, this means calculating the height of the terrain at all points along the ground trace of the LOS. If we use the bi-linear interpolation model, however, we only have to compare height at a small number of points. Specifically, we only have to check heights on lines between adjacent terrain data points that intersect the ground trace. This is shown in Figure 3. The only comparisons that need to be made are those along lines C-D, B-D, and A-D. Note that all of these lines are connected to point D.

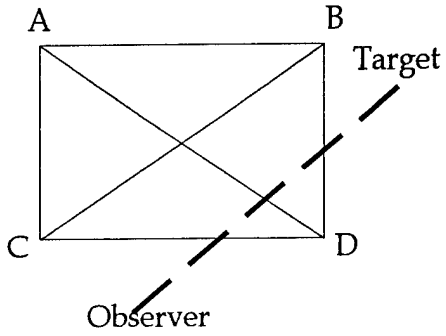


Figure 3: Ground Trace - Bi-linear Model Intersections

It is a simple matter to determine the rectangles that the ground trace of the LOS passes through. For each of the six lines of possible intersection, we may calculate the intersections of the cross lines,

$$\begin{aligned} x &= x_i + (x_{i+1} - x_i)\,\zeta, \\ y &= y_j, y_{j+1}, \end{aligned} \tag{20}$$

the verticals,

$$\begin{aligned} y &= y_j + (y_{j+1} - y_j)\,\zeta, \\ x &= x_i, x_{i+1}, \end{aligned} \tag{21}$$

and the diagonals,

$$\begin{aligned} x &= x_i + (x_{i+1} - x_i)\,\zeta, \\ y &= y_j + (y_{j+1} - y_j)\,\zeta, \end{aligned} \tag{22}$$

and

$$\begin{aligned} x &= x_i + (x_{i+1} - x_i)\,\zeta, \\ y &= y_j - (y_{j+1} - y_j)\,\zeta, \end{aligned} \tag{23}$$

with the ground trace, equations 19. Note that since ζ is on [0.1] to stay within the grid rectangle, then any intersection point with < 0, or > 1 does not lie within the rectangle. Similarly, any computed value of $\zeta < 0$, or $\zeta > 1$ does not lie between observer and target.

Having calculated the values of ζ and η that constitute a point of intersection, we may now compute $z(\eta)$ from equation 19, and $z(x,y)$ from the appropriate bi-linear interpolation. If $z(\eta) - z(x,y) > 0$, then the LOS exists at this point.. If $z(\eta) - z(x,y) \leq 0$, then the LOS is blocked at this point.

The calculation is repeated for all of the ground trace - grid rectangle intersection points between observer and target . If none of the terrain at the intersection points blocks the LOS, then the LOS between observer and target exists. If even one intersection point blocks the LOS, then the LOS does not exist.

It is also useful to compute one other quantity. At each intersection point, assuming LOS is not blocked, we compute the quantity

$$\tau = \frac{z(\eta) - z(x,y)}{\sqrt{(x - x_O)^2 + (y - y_O)^2 + z(\eta)^2}}, \tag{24}$$

and retain the minimum value of this quantity for all intersection points. If the LOS exists, this quantity τ is approximately the angle below the LOS that is visible along the LOS. Thus the quantity

$$H = \sqrt{(x_T - x_O)^2 + (y_T - y_O)^2 + (z_T + h_T - z_O - h_O)^2}, \tag{25}$$

is approximately the height below the LOS that is visible along the LOS at the target. Thus, if $H \geq h_T$, the whole target is visible, while if $H < h_T$, only the top H of the target is visible.[7] This quantity will be important later in calculating probability of detection.

Of course, this is a lot of calculations and is one of the primary reasons that deterministic simulations can have a faster effective execution rate since repeated executions are not necessary.

There is another approach to the problem, and that is a stochastic one. We may assume that the probability of LOS existing has some range (even angle) dependent probability distribution function that has parameters which reflect the type or location of the terrain. if the probability distribution function is $p(r)$, the probability the LOS exists to a range r is just the integral of $p(r)$,

[7]This neglects any left-right hiding of the target. Terrain usually, but not always has a vertical slope sharp enough to hide part of a target in a left-right sense rather than an up-down sense. We have ignored this by assuming either all of the width of the target is exposed or hidden in our simple analysis.

$$P_{LOS}(r) = \int_0^r p(r')\,dr'. \tag{26}$$

We may use $P_{LOS}(r)$ in one of two ways: as a stochastic driver, we may generate a uniform random value R and compute the resulting range r^* from

$$r^* = P_{LOS}^{-1}(R), \tag{27}$$

and if

$$r^{*2} > (x_T - x_O)^2 + (y_T - y_O)^2, \tag{28}$$

then the LOS exists. Alternately, we may use $P_{LOS}(r)$ directly in computing the attrition rate, an operation that we shall examine in a later chapter.

The most common form of $P_{LOS}(r)$ given in the literature is negative exponential,[8]

$$P_{LOS}(r) = e^{-\gamma r}, \tag{29}$$

where γ is a parameter that depends on terrain geography and geometry. Burton and Chu[5] give values for γ of $0.156km^{-1}$ for particularly smooth terrain and $0.381km^{-1}$ for particularly rough terrain. The noted Russian analyst and soldier Chuyev[6] gives the definition

$$k = \frac{\kappa}{h}, \tag{30}$$

where:

h is the height to the target, and

κ is a parameter on [1,3] m km^{-1}.

We note that for nominal target heights that we would consider representative ($h \sim 2.5 - 3m$,) Chuyev's values of γ are larger than Burton and Chu's, perhaps reflecting their different terrain data bases.

Chuyev's form is somewhat more useful since we may use it to infer the expected height of target that is visible. If $P(r,h)$ is the probability that a LOS exists to a target of height h at range r, then

$$\langle h \rangle = h - \int_0^h \frac{dP(r,h')}{dh'} h' dh', \tag{31}$$

is the expected height of target that is visible.We shall also explore this concept in a later chapter when we consider probability of detection. (At that time, we shall draw a distinction between the height of the target h, and the vertical size of the target.)

This effectively concludes our introduction to the effect of terrain on attrition. As we have already noted, terrain has two primary visual effects on attrition: it blocks LOS from observer

[8]Champion (cited in Youngren - see References(cites a Weibull distribution as appropriate. For simplicity, we shall continue to assume a NED.

to target; and it may limit the visible (and vulnerable) size of the target. These two factors will be examined later in conjunction with other factors.

23.2.5 Statistical Terrain and Averaging

Before proceeding, a brief discussion is appropriate on convention and how statistical terrain models may be constructed. First, one must be careful to define how the LOS is to be calculated:

- from observer point to target point on the ground;
- from observer's height to point on the ground;
- from observer's point on the ground to target height;
- from observer's height to target height; or
- some other convention (e.g. observer's height to half target height)?

The first always assures that all of the target may be seen, regardless of height, but may underestimate by ignoring the addition of the observer's height. The second corrects this underestimate, but now penalizes for the target's height. The third not only penalizes for the observer's height, but allows miniscually visable targets to be included. The fourth corrects for both heights but again may allow miniscually visible targets to be included. The fifth tries to correct for this usually by counting all of the observer's height but only half of the target's.

Ideally, what we would like is a distribution $P_{LOS}(r, z, h_o)$ which would consider observer height as well as height above terrain at range as variables. This would allow an effective LOS probability to be calculated by integrating over the target's height. (Why this becomes desirable will become evident later when we consider sensors and detection.)

Ignoring for the moment how that distribution might be formulated, and what the LOS convention is, now consider how a stochastic LOS distribution may be calculated. Assume that we possess terrain data of the form of equation 4. Select two sets of points on this data set: a set of observer locations $\{x_{oi}, y_{oi}\}$; and a set of "target" locations $\{x_{tj}, y_{tj}\}$. These two sets may be as dense as desired up to the set of all points represented by the data set. Then for all possible combinations of locations, determine whether LOS exists or not. Sort the results into bins according to range between locations (at least) and a distribution may be formed.

Another approach requires only the observer locations. At each observer location calculate the range that LOS exists along different azimuths. If there is no prefered direction combat is likely to occur, these values may be used to construct a local LOS distribution or combined to form a global distribution. (Or retained explicitly to form a Dirac delta function distribution.)

Both of these techniques are eminently adaptable to filed experimentation as well. The latter

is somewhat cheaper to implement since if one uses a convention of observer height to ground, the measurements can easily be made with a surveyer's laser range finder. The former method requires the additional use of a target object.

23.3 Weather

Like terrain, weather has two major effects on combat processes: movement and vision.[9] The details of the two effects may interact with terrain or they may be independent. In addition, weather may have some secondary effects on combat. In particular, we note that adverse weather may have deleterious effect on materiel, usually in the form of corrosion or accelerated wear. As we know from peaceful experience, this corrosion may be a direct result of the weather itself, or it may be an indirect result of palliative measures to temper other effects of weather (e.g., salt - NaCl - on icy roads.)

In addition, weather may have a secondary effect on the health and fitness of troops. Again this may be direct from such aliments as respiratory infections or dehydration, or in combination with terrain through such varied aliments as gastrointestinal infections including the age old military scourge dysentery, or problems such as frostbite, immersion foot, or even Athlete's Foot.

23.3.1 Weather and Movement

Just as the best rate of movement can be realized on flat, smooth terrain (with proper frictional properties,) and departures from flatness and smoothness reduce that rate of movement, in general, departures from clear weather reduce the rate of movement. As a rule, the rate of speed for a given mode of movement (man or machine,) is a complicated function of weight, pressure, power (rate of energy use,) soil composition, friction, adhesion, terrain slope and roughness, and other factors. Even this rate is further moderated by a practical trade-off among safe control of movement, acceptable damage and exhaustion rate, and acceptable transit time. In pure movement terms, control is largely a matter of friction, but in the military context, it also extends to either the coordinated movement of the units, or the coordinated occupation of terrain, or both.

One of the fundamental interactions of weather with movement is its effect on friction. When we think of adverse weather, we most often think of rain and snow. The usual effect of these is to decrease friction which decreases the speed of safe control of both vehicles and men (albeit usually less so for the latter.) Additionally, large amounts of rain result in flooding and mud, which further slow movement. Likewise snow has the effect of hiding terrain variations which slows speed and increases damage.

There are exceptions to this. Moderate amounts of rain on sandy soil actually increases soil cohesion, making it a firmer surface for travel and thus increasing movement rate. This is why

[9] Actually, both have a third effect - on communication. This is actually a special case of the effect on vision and we will not spend any appreciable attention on it.

faster movement is possible in moist sand than in dry sand. Similarly, snow is accompanied by reduced temperature which freezes muddy terrain and makes it more trafficable.

In addition, rain and snow and indeed haze and fog reduce the range of vision which slows movement through a reduction of the range of control. Just as the variations in terrain may separate units from sight and control of each other, the reduction of intervisibility by weather may have an adverse effect on control.

Further, weather tends to moderate the effects of direct sunlight and temperature. As we have already noted, too much sun and temperature can result in dehydration and heat prostration; too little can result in frostbite.

In summary the combination of terrain and weather are the key determiners, in a classical sense, of the movement and control of military operations. Only in the last century or so has technology advanced to the point where this determination has weakened.

23.3.2 Weather and Attrition

Just as the reduction of vision by weather reduces the span of control, it reduces the range and effectiveness of weapons. This is the fundamental effect of weather on attrition - the reduction of visible range. It does no good to have a weapon with a maximum range of 3 km when targets cannot be seen to be engaged at ranges greater than 1.5 km (e.g.)

How exactly does this effect attrition? The answer is relatively complicated and we shall arrive at this answer in the next few chapters.

23.4 References

[1] Johnson, Paul, **The Birth of the Modern**, Harper Perennial, New York, 1991.

[2] Youngren, Mark A., series ed., "Military OR Analyst's Handbook: Volume I Terrain, Unit Movement, and Environment", Military Operations Reserach Society, 1994.

[3] Price, James Davis, "An Investigation into Development of a Military Small Unit Tactical Training Simulation", MSOR Thesis, University of Alabama in Huntsville, 1980.

[4] Peitgen, Heinz-Otto, Harmut Jurgens, and Dielmar Saupe, **Chaos and Fractals New Frontiers of Science**, Springer-Verlag, New York, 1992.

[5] Burton, L. S., and W. C. Chu, "Frequency Distribution of Targets Captured by FOV", Hughes Aircraft Company, Interdepartmental Correspondence, 10 March 1982.

[6] Chuyev, Yu. S., **Research of Military Operations**, Moscow, 1970, American Translation JSRS 53366, 15 June 1971.

Chapter 24

Optics I

24.1 Introduction

In this chapter we begin our consideration of the phenomenology that goes into the calculation of attrition rate coefficients. This chapter and the next deal with those aspects of optics that contribute to this. (Albeit, the optics presented here is quite simple, in keeping with the general tone of the book.) In this chapter, we shall be concerned with the basics of light movement while the next chapter will deal with the optics of image transmission. Because there is a well developed textbook literature on optics (unlike attrition,) we will not make any attempt to be complete in our coverage of this subject.[1][2] Both the technical and the military student may avail themselves of these texts to supplement the small amount that we do present here.

Having said that, I will tell the technical student that there will not be enough of either detail or rigor to satisfy here. Similarly, the military student may find too much detail and rigor. What I hope that both will find here is a reasonably integrated picture of what the physical pieces are and how they interact.

In fact, this is the approach for the chapters of this book that deal with the physics that contributes to attrition rate coefficients.

24.2 The Nature of Light

Light is a complicated subject. It may be viewed in a variety of manners and from several directions. First of all, light is an electromagnetic wave. A wave may be thought of as (usually being) a three-dimensional artifact. The waves that we normally think of, those coming from a pebble dropped into a still pond or lake are circular waves. They consist of a direction of propagation (outward from the point the pebble is dropped at) which is one of the three dimensions. The amplitude of the wave, displayed in the sinusoidal variation of the surface may be thought of as having two dimensions perpendicular to the direction of propagation. The

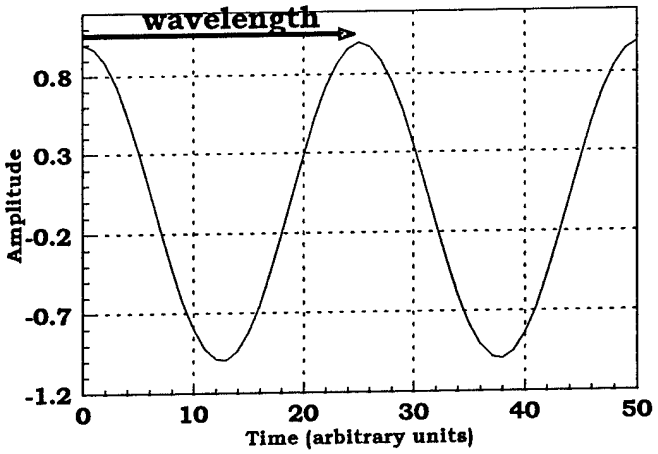


Figure 1: Sinusoidal Wave

surface variation is one of these dimensions, the circular pattern, exemplified by the peaks and valleys of the amplitude, are the other dimension. This is sketched in Figure 1

Similarly, the electromagnetic wave has three dimensions. It has a direction of propagation and it has two dimensions of amplitude; represented by an electric field; and a magnetic field. Both of these are sinusoidally varying and they are nominally 90 degrees out of phase from each other. (This is dictated by energy conservation.)

There are several different types of waves. The two most common are plane and spherical waves. In analogy with the circular waves of our pond example, the terms plane and spherical refer to the surfaces exhibited by the amplitudes. In the first case, if we select a value of amplitude of the electric (or magnetic) field, the surfaces of that amplitude value are planes. In the second case, they are spheres.

We may associate the quantities of wavelength and frequency with electromagnetic waves. Wavelength (λ) is the distance from amplitude peak (or valley) to amplitude peak (or valley) along the direction of propagation.. The frequency of the wave (ν) is the number of wavelengths the wave has per second (e.g.) These two are related to the velocity of propagation (technically its speed - c) by the simple relation

$$\lambda\nu = c. \tag{1}$$

The velocity of propagation, also known as the speed of light, depends on the medium that the wave propagates through. In a vacuum, also known as free space, the speed of light has an approximate value of $3x10^8$ m/sec.

There are other views, as we have said. For example, an electromagnetic wave that has a limited duration in time (and thereby in space,) is often called a pulse. In addition to its inherent structure, the wave may have a magnitude structure called its intensity which is related to the square of the amplitude.

We may also consider the structure of light (a term by which we shall refer to electromagnetic radiation in general, regardless of wavelength/frequency,) as a ray typified by its direction of propagation. A further description that captures the volumetric structure of light is called beam optics. This discipline is most commonly used in describing laser beam propagation. Finally, we may deal with what is called Fourier Optics which we will make use of primarily in considering image propagation.

Lastly, there is one more way of looking at light. Instead of considering light as a wave, we may consider it as particles called photons. These particles individually have an energy given by the simple relationship

$$E = h\nu, \tag{2}$$

where h is known as Planck's constant. Because this picture views light as particles, we may often treat these particles with mechanics albeit that we must normally use Quantum rather than Classical Mechanics.

24.3 Light and Matter

Of course, to be interesting, light must interact with matter. These interactions have a variety of forms. Reflection and refraction are most often associated with light (viewed as a ray) at the interface between two regions of different optical properties, usually characterized by different refractive indices. The refractive index is usually indicated by the variable n, which may be a complex (valued) quantity and is defined by the simple relationship

$$n = \frac{c_0}{c}, \tag{3}$$

where: c_0 = the speed of light in vacuum, and c = the speed of light in the medium.

Reflection is the physical effect at the interface of returning some of the incident light back into the originating medium but in a (usually) different direction. Refraction is the transmission of some of the incident light into the new medium, again usually in a different direction. This is the basis of lenses. The ideas of reflection and refraction are sketched in Figure 2.

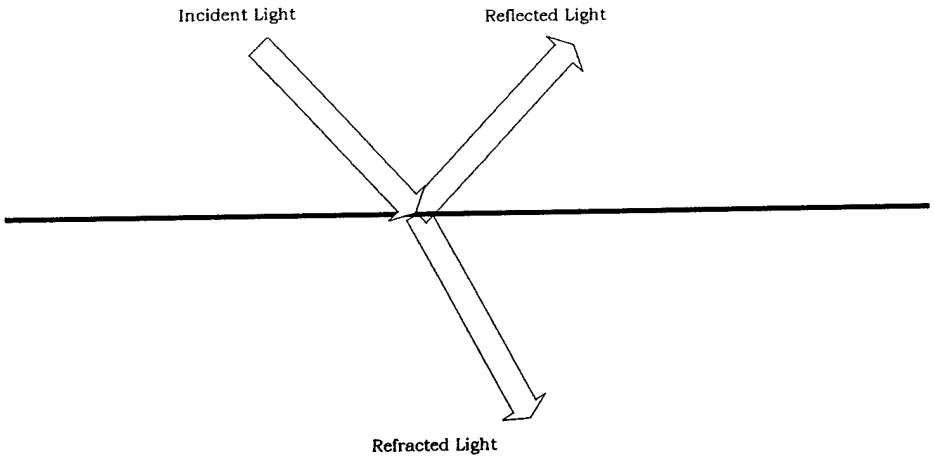


Figure 2: Reflection and Refraction at a Surface

Absorption is just what the name implies, the absorption or consumption of light by matter. Of course, the energy of the light must be conserved so the light must be converted to heat, vibration, or even electric currents ("free" electrons.) The latter is the basis of light detectors.

Scattering is the interaction of light with relatively small (compared to the wavelength of the light) bits of matter. If the scattering is elastic, the interaction changes only the direction of the propagation of the light. In analogy with probability, there is a distribution function (of angles or direction of propagation,) associated with scattering.

Finally, there is emission which as the name suggests, is the generation of light by matter. The most commonly considered form of emission is black body emission which is the light emitted by an idealized piece of matter called a black body. By its nature, black body radiation has a distribution in frequency (or wavelength.) This distribution, in terms of the average energy per second per unit projected source area, unit solid angle, and unit wavelength, is given by

$$L_\lambda = \frac{8\pi h}{\lambda^5} \frac{1}{e^{\frac{hc}{k\lambda T}} - 1}, \tag{4}$$

where:

h is (again) Planck's constant,

k is Boltzmann's constant, and

T is the absolute temperature (in degrees, Kelvin.)

Usually, this equation is presented as a function of frequency, rather than wavelength, but we have presented it here this way to facilitate our discussion. We present a set of representative

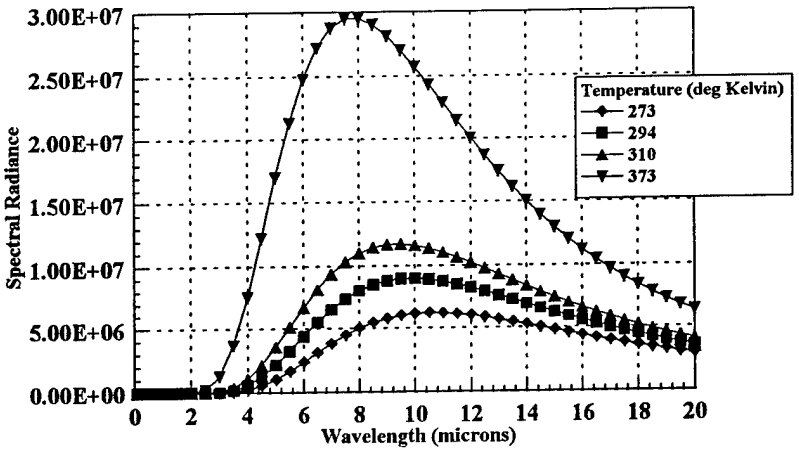


Figure 3: Black Body Curves

plots of this distribution for several temperatures in Figure 3.

(Note that the temperatures are given in degrees Kelvin (absolute metric temperature.) Since the Fahrenheit temperature scale is commonly used in the United States, the conversion from Fahrenheit T_F to Celsius T_C (Centigrade for old folks like myself) is

$$T_C \simeq \frac{5}{9}\left(T_F - 32^\circ\right), \tag{5}$$

and the conversion to absolute, T_K, is

$$T_K \simeq 273^\circ + T_C. \tag{6}$$

I use this "Strange, unscientific" system to facilitate the discussion.) There are four curves in Figure 3, corresponding to absolute temperatures of $273^\circ, 294^\circ, 310^\circ$,and 373°. These are respectively, the freezing point of water, the temperature of a warm room (70o F), the temperature of a human body (98.6°), and the boiling point of water.

As may be seen, there are two features of varying temperature. First, the magnitude of the distribution the area under the curve - (this is clearly not a normalized distribution!) increases

with increasing temperature. This isn't really surprising since the temperature is a measure of the internal energy of the black body. Given that the black body radiates (emits light which is a form of energy,) then we would expect the quantity emitted to increase as the energy of the black body increases.

Second, the peak value of the distribution changes with temperature. Specifically, as temperature increases, the wavelength associated with the peak becomes smaller. (This is one of those times when it is more useful to think in frequency terms. Remember that the higher the frequency of the photon, the greater its energy - equation 2. The greater the internal energy of the black body, the greater the number of high frequency (low wavelength) photons that are emitted.)

We may make two other immediate observations. Once the peak passes a given wavelength, the amount of energy emitted at that wavelength increases with increasing temperature. This fact is important when we consider the total energy emitted over a range of wavelengths. Additionally, the peaks of all of these curves lie in a wavelength region between 8 and 14 microns.

There is one more quantity that we want to consider. The albedo is the fraction of light incident on a body (obviously not necessarily black) that is reflected and emitted by it. Normally, the albedo has a value less than one, but under certain circumstances, it may be greater than one (due to emission and/or inelastic scattering.)

24.4 Light and the Atmosphere

The atmosphere is comprised of gas molecules and particles. Both of these absorb, scatter, and emit light. A key characteristic of the gas absorption is that it obeys the law of Quantum Mechanics almost exclusively. Absorption thus occurs only at wavelengths (frequency picture really) that correspond to allowed quantum mechanical transitions. Normally, this would mean that the atmosphere would only absorb light at very specific wavelengths and over very short wavelength variations. Thus the atmosphere would be only very weakly absorbing in a broad band (wide range of wavelengths) sense and otherwise clear (non-absorbing.)

This isn't quite the case. The atmosphere has a temperature and thus its molecules are in motion. Two important physical processes come into play as a result. These are the Doppler Effect and collisions among molecules. As a result of these two, the atmospheric gas molecules absorb photons over much broader ranges of wavelengths (although still centered about those quantum mechanics dictated wavelengths). Thus instead of very narrow distributions of absorption (essentially Dirac delta functions,) the distributions are broad (essentially gaussian distributions) and the atmosphere absorbs in wavelengths regions of strong and weak absorption.

The regions of weak absorption are often called bands and four of these bands are of primary interest to us. These are shown in Figure 4.[5] The first of these bands, between 0.3 and 0.7 microns wavelength, is known as the visible band. (A micron, often abbreviated by the Greek letter μ, is 10^{-6} m.) The visible spectral band is the region where the human eye operates.

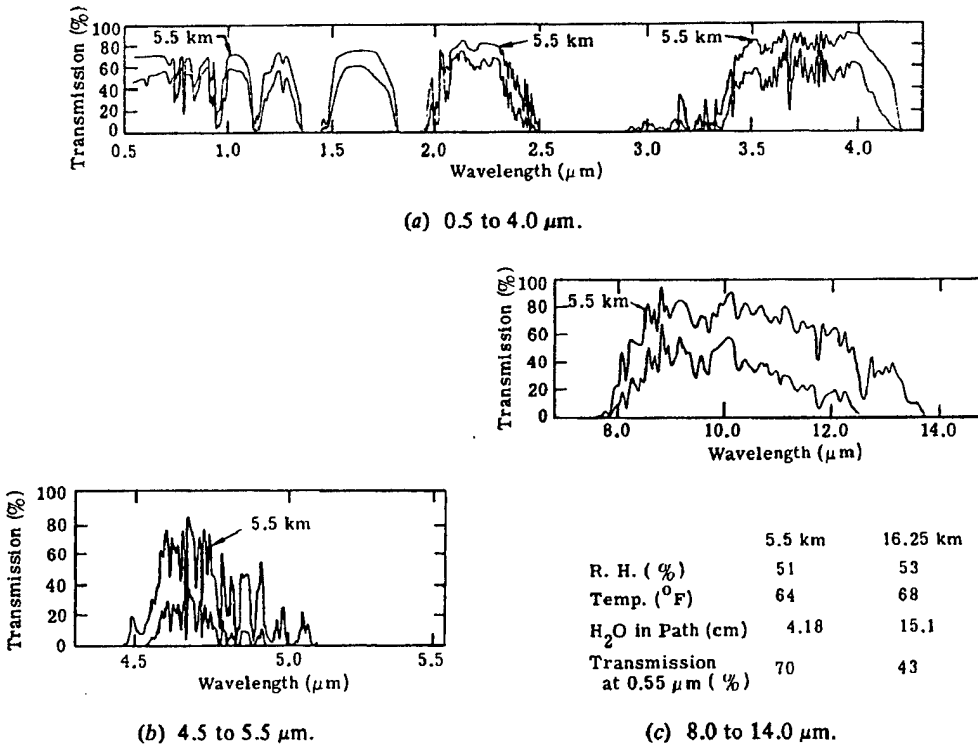


Figure 4: Representative Atmospheric Transmission Spectra

The second band, the near infrared, is between 0.9 and 1.1 . This is the band where sniperscopes operate. The third band, the mid infrared, is between 3 and 5 , and the fourth, the far infrared, is between 8 and 14 . As may be seen, the mid infrared band has a break between about 4.2 and 4.5μ. For longer pathlengths (the figure displays two pathlengths, 5.5, and $16.25km$ in length,) the region of the far infrared band between 12 and 14μ effectively shuts down (has low transmittance,) and this band is also frequently referred to as having a range of $8 - 12\mu$, especially by the engineering communities who are primarily concerned with longer ranges (e.g., aviation and naval).

The mid and far infrared bands are where Forward Looking InfraRed (FLIR) sensors operate. The importance of the far infrared band may be seen by examining Figure 3. Specifically, we want to examine the bottom three curves which correspond to temperatures of 32, 70, 98,6 and 212 degrees Fahrenheit. These temperatures correspond to the freezing point of water, a nominal indoor temperature, the temperature of the human body, and the boiling point of water. Most of the black body light emitted at these temperatures falls in the far (or mid) infrared bands. (Remember, only the emitted light of wavelength in a band - a region of weak atmospheric absorption - travels very far.) Thus, the natural emissive signatures of objects, including people,

are sensed by FLIRs.

Similarly, the emissions of engines (500-1000° Absolute) tends to be in the mid infrared band. Thus the infrared sensors for tracking aircraft usually operate in this band.

We may now consider how the intensity of light varies with range. This is known as the Beer-Lambert-Bouguer Law. Mathematically, this has the form

$$I(x,\lambda) = I_0(\lambda)\, e^{-\alpha(\lambda)x}, \tag{7}$$

where:

x is range,

$\alpha(\lambda)$ is the attenuation coefficient,

$I_0(\lambda)$ is the intensity at zero range and wavelength λ, and

$I(x,\lambda)$ is the intensity at range x and wavelength λ.

The Transmittance, $T(r)$, is defined by

$$T(r) = e^{-\alpha(\lambda)x}, \tag{8}$$

or, as we shall see below, by the averaged attenuation coefficient, as appropriate.

This equation is an approximation. It does not really include scattering. A more general form of equation 7 can be found by solving the differential equation

$$\underrightarrow{k} \bullet \underrightarrow{\nabla} I\left(\underrightarrow{r}, \underrightarrow{k}\right) = -\alpha I\left(\underrightarrow{r}, \underrightarrow{k}\right) + \alpha\omega \int d^3\, \underrightarrow{k} p\left(\underrightarrow{k} \bullet \underrightarrow{k'}\right) I\left(\underrightarrow{r}, \underrightarrow{k'}\right), \tag{9}$$

where:

$\underrightarrow{k}$, $\underrightarrow{k'}$ are (unit) propagation direction vectors,

$\underrightarrow{r}$ is the position vector,

ω is the ratio of the scattering coefficient to the absorption coefficient, and

$p\left(\underrightarrow{k} \bullet \underrightarrow{k'}\right)$ is the scattering distribution function (which is normalized.)[1] We will not spend much effort with this equation for two reasons: no general analytical solution exists, and for short distances, scattering effects are small. The single exception to this is treated late in this chapter.

24.5 Band Wide Transmission

As we have already noted, the Beer-Lambert-Bouguer Law is wavelength dependent. This is important since the attenuation (the combination of absorption and scattering) of the atmosphere is wavelength specific, and, as we have seen for black body emissions, target signature intensity varies with wavelength. A further consideration is that the efficiency of sensors (also known as sensitivity) is wavelength specific. Thus, to be technically correct, we would calculate the signal generated in a sensor at range r from a target as

[1] The • indicates the scalar vector (or "dot") product.

$$\begin{aligned} S &= \int_{\lambda_1}^{\lambda_2} E(\lambda) I(r,\lambda) d\lambda \\ &= \int_{\lambda_1}^{\lambda_2} E(\lambda) I_0(\lambda) e^{-\alpha(\lambda)r} d\lambda, \end{aligned} \tag{10}$$

where: $E(\lambda)$ is the sensor sensitivity.

In many cases, however, the variance of signature $I_0(\lambda)$ and attenuation coefficient $\alpha(\lambda)$,

$$\begin{aligned} \sigma_I^2 &= \left\langle I_0(\lambda)^2 \right\rangle - \langle I_0(\lambda) \rangle^2, \\ \sigma_\alpha^2 &= \left\langle \alpha(\lambda)^2 \right\rangle - \langle \alpha(\lambda) \rangle^2, \end{aligned} \tag{11}$$

where:

$$\langle f(\lambda) \rangle = \frac{1}{\Delta\lambda} \int_{\lambda_1}^{\lambda_2} f(\lambda) d\lambda, \Delta\lambda \equiv \lambda_2 - \lambda_1, \tag{12}$$

are small. As a result, we may, in many cases, approximate the signal equation as

$$S \simeq \langle E(\lambda) \rangle \langle I_0(\lambda) \rangle e^{-\langle \alpha(\lambda) \rangle r} \Delta\lambda, \tag{13}$$

This approximation is generally limited to a short range versus small attenuation trade-off, and a relatively close match between the shapes of the $E(\lambda)$ and $I_0(\lambda)$ curves.

There are other intermediary approximations. If the $E(\lambda)$ and $I_0(\lambda)$ curves do not match well, then we may use the slightly better, but more computational challenging approximation,

$$S \simeq \langle E(\lambda) I_0(\lambda) \rangle e^{-\langle \alpha(\lambda) \rangle r} \Delta\lambda. \tag{14}$$

Alternately, we may use simpler functional approximations for any of these three quantities. One such approach is to calculate approximations for $\langle \alpha(\lambda) \rangle$ that are functions of range. (The interested student is referred to the Proceedings of the Smoke Symposium sponsored by the Project Manager, Smoke in past years.)

As a general approximation consistent with the level that we will be treating most physical phenomena for this book, we will use an approximation of the form of equation 13. We leave more accurate analyses to the student in the work or research place. This leads to a "band averaged" transmittance of

$$\langle T(r) \rangle = e^{-\langle \alpha(\lambda) \rangle r}, \tag{15}$$

which, in general usage, we will not distinguish from "unaveraged" transmittance.[2]

[2]Of necessity, attenuation coefficients estimated from visibilities are "band averaged" over the visual spectral band, the spectral sensitivity of the eye, and the spectral "intensity" of the observed object. Since most considerations are made in the visual spectral band, it is common practice not to distinguish between averaged and "unaveraged" transmittance,

We present selected values of attenuation coefficients for the spectral bands and selected weather conditions in the following section.

24.6 Attenuation Data

In this section, we provide selected attenuation coefficients for sensors operating in the spectral bands already introduced, plus in the millimeter wavelength region corresponding to a frequency of 94 Gigahertz. Greater detail may be found in DOD-HDBK-178,[3] and in the LOWTRAN computer code and documentation.[4]

Atmospheric Condition	Visual	Mid IR	Far IR	mmw
Tropical Hazy	0.8	0.7	0.6	0.2
Tropical Clear	0.2	0.7	0.5	0.2
Midlatitude Summer Hazy	0.8	0.7	0.4	0.1
Midlatitude Summer Clear	0.2	0.5	0.4	0.1
Subarctic Winter Hazy	0.8	0.4	0.1	0.02
Subarctic Winter Clear	0.2	0.3	0.05	0.02

Normal Atmosphere Attenuation Coefficients

The terms Tropical, Midlatitude, and Subarctic are taken from the LOWTRAN terminology and refer approximately to geographic regions. Hazy and Clear refer to the visibility of the atmosphere, 5 and 23 kilometers, respectively. The units, common to all attenuation coefficients unless otherwise specified, are km^{-1}. These coefficients are effectively base case conditions due to atmosphere alone. Attenuation coefficients for adverse weather must be added to these. The coefficients for Rain are

Rain Rate	Visual	Mid IR	Far IR	mmw
1 mm/h	0.4	0.4	0.4	0.4
4 mm/h	0.9	0.9	0.9	1.0
10 mm/h	1.6	1.6	1.6	2.0
25 mm/h	1.2	1.2	1.2	1.7

Attenuation Coefficients for Rain

These rain rates correspond to light, medium, and heavy rains, and a thunderstorm. The units are millimeters of rain per hour (mm/h).In general, the average size of rain drops in a storm is a function of the rain rate - the higher the rate, the larger the drops. The student will note that the coefficients are all the same for the visual and IR bands, primarily because rain drops, regardless of rate, are much larger than the wavelength of light in these bands. This is not the

case for millimeter waves, as evidenced by the variation. The other form of precipitation most encountered is snow:

Snow Rate	Visual	Mid IR	Far IR	mmw
0.5 mm.h	1.7	1.7	1.7	0.03
2 mm/h	2.3	2.3	2.3	0.5
5 mm/5	4.3	4.3	4.3	1.0

Attenuation Coefficients for Snow

Finally, the attenuation coefficients for Fog (visibility < 1 km) are:

Type	Visual	Mid IR	Far IR	mmw
Advection	3.9	4.2	4.5	0.05
Radiation	3.9	5.0	2.9	0.02

Attenuation Coefficients for Fog

The student must be warned that these coefficients are illustrative and representative only. Attenuation due to dust/sand will be considered in the discussion of obscurants in the next chapter. There is generally little difference in the values of attenuation coefficients between the visual and near IR spectral bands, so for most applications (except obscuration,) visual spectral band attenuation coefficients may be used for near IR spectral band calculations.

To be a bit more precise, we must distinguish between two usages of the terms fog, haze, and clear (and the adverbs foggy, hazy, and clear.) In describing meteorological conditions, characterized by visual (or meteorological) ranges (described in a later section,) these terms are generally understood to represent bounding conditions. Specifically, fog refers to visual ranges between 0 and 1 km, haze refers to visual ranges between 1 and 10 km, and clear refers to visual ranges greater than 10 km. Some schemes exist for characterizing these three categories further using terms such as "dense fog", "thin fog", etc. There is little physical reason for greater characterization short of actual statement of the value of visual range.

The second usage is the one we have used predominantly in this section. It is used in the context of defining representative scenarios or conditions for the purpose of comparison of system performance. These two are not inconsistent, since the visual ranges defined in the scenarios conform with the visual range definitions of the first usage. There may be a tendency for the unwary to confuse the labels attached to specific scenarios described above with the more general terminology. My experience is that this can be quickly sorted out from the context of discussion.

24.7 Light Sources

The most fundamental light source in military (and civil) application is the Sun. It provides both direct exterior illumination during the day, and the strongest source of indirect exterior illumination at night by reflection from the Moon. The Sun is a broad spectrum illumination source, which if we neglect the effects of the atmosphere, is essentially a 2000°K black body.

Other important military light sources include the normal civil incandescent light bulb and the fluorescent tube. Both are these are also broad spectrum illumination sources. The incandescent bulb produces light by heating a high resistance wire (called a filament) to a high temperature. Basically, the filament is kept (at the operating voltage and current) at a temperature that produces a black body source peaked in the visual spectral band. The spectral distribution is modified from ideality by the composition of the filament. It is further modified by the "glass" envelope of the bulb, primarily in filtering out the short Ultraviolet wavelengths although it also modifies the InfraRed distribution as well.

The fluorescent tube operates by ionizing a gas in an electric field. These ionized gas atoms deionize by exciting emission from the material coating the inside of the "glass" envelope. Phosphorus compounds are frequently used for this coating material and are quite toxic if introduced into the human body. As with bulbs, the glass envelope of the moderates the spectral distribution of the light emitted. As might be expected, profuse ionization requires greater voltages so fluorescent fixtures include a step-up transformer.

The intent of both incandescent bulbs and fluorescent tubes is to produce light whose spectral distribution approximates that of "natural" sunlight. Thus this light attenuates in a manner similar to sunlight.

The incandescent bulb approach can be extended to longer wavelengths (InfraRed) by using different materials and (generally) lower temperatures and larger areas (to compensate for the lower energy density.) For the mid and far IR spectral bands, the "glass" envelope is often discarded to eliminate absorption. These can be used to provide illumination for sensors operating at these longer wavelengths and have some military application, notably for near IR viewers and missile guidance.

Lasers are increasingly of military importance as sources of illumination in military sensor systems. The term Laser is actually an acronym Light Amplification by the Stimulated Emission of Radiation - LASER. It has come into such common usage in the language that it has become a word rather than an acronym.

Unlike most common sources of illumination such as the Sun (Moon), incandescent bulbs, and fluorescent tubes, lasers are narrow spectrum sources. The popular belief is that lasers are true monochromatic (single wavelength) sources. While the spectral emission of lasers is broadened somewhat by the Doppler effect, this strict monochromaticity is not generally the case. Unless specifically designed to be so, most lasers operate to produce light at several wavelengths although these wavelengths usually have only small differences. Thus most lasers are described in terms of the mean or primary wavelength that they produce light.

Light Sources

By nature, lasers are fundamentally quantum devices, selectively allowing only specific quantum transitions to actually produce light. Since the excitation is usually broad band in form (either light or electric field/discharge), in general lasers have relatively low efficiency of energy conversion. Lasers have been made in a variety of material forms including solids, liquids and gasses. Solid lasers are generally of two types: semiconductor lasers which exploit the band structure of the crystal materials to produce light, and "jewelry" lasers which consist of an active material embedded in a matrix. The original ruby laser falls into this category. Liquid and gas lasers are usually analogs of the solid "jewelry" lasers in a different material state.

Despite the profuseness of laser sources, the two most common military lasers are the Neodymium-Glass (Nd YAG) laser which is a "jewelry" laser operating around 1.06μ and the Carbon Dioxide (CO_2) gaseous laser operating around 10.6μ. (The factor of 10 multiplier here is purely coincidental.) Other than wavelength appropriateness for function, the two primary consideration for military lasers are amenability to ruggedization (filed conditions,) and eye safety (to prevent eye damage by inadvertent viewing of the laser beam.) Thus a desire to select wavelengths that are not focused by the eye.

Light sources have a Field of Illumination (FOI) analogous to the Field of View of sensors (see below). General illumination sources such as incandescent bulbs and fluorescent tubes have FOI that are quite large - usually between 2π (half sphere) and 4π (full sphere) steradians, although fixture geometry may reduce this. Searchlights and flashlights, intended to provide directed illumination under dark conditions, generally have smaller FOI, of the order of a few degrees. Because of their energy inefficiency and construction geometries, lasers are almost always configured to produce exceedingly small FOI, of the order of fractions of degrees. This facilitates their use for range finding and weapon guidance.

Lasers may also offer potential as weapons but we shall defer discussion of this to a later chapter. It suffices for now to mention that this extends the militarily interesting lasers to include both lasers which operate in the visual and mid IR spectral bands.

For the two lasers of primary military interest, Nd YAG and CO_2, atmospheric attenuation is very similar to broad band attenuation both in form and attenuation coefficient, with one exception. That exception is attenuation due to self-absorption of the CO_2 emissions. Thus it may be necessary to increase the value of attenuation coefficients for CO_2 laser emissions under conditions where the atmospheric CO_2 gas concentrations are higher than "normal". Unfortunately, since human beings, motorized vehicles, and weapon usage produces CO_2 as a waste product, there is often additional attenuation under battlefield conditions. Fortunately, this additional attenuation has always been observed to be less than for fog (e.g.,) and since lasers can have very high signal-to-noise ratios at the emitting aperture, this attenuation is usually not troublesome. For most practical purposes, laser transmission can be approximated using the appropriate broad band attenuation coefficients, except for obscurants.

Two other light sources need brief discussion: the Light Emitting Diode (LED); and the Cathode Ray Tube (CRT). The LED is basically an solid state electronic device that emits light when a current is applied. These may be grossly thought of as similar to semiconductor lasers

except that the emitted light is broad band (and thus the devices are more efficient.) LED's may be used as individual light sources ("idiot" lights,) or in groups as a flat panel display of information. Liquid Crystal Displays and Plasma Displays serve a similar function and although there are differences in physics, operate in similar ways.

The CRT is an offshoot of the fluorescent tube. The inside of the viewing surface contains a layer of one or more emissive materials. An electron "gun" inside the tube shoots electrons at the emissive material which becomes excited when hit and decays by emitting light. Different materials emit light of different spectral characteristics, thus making multi-colored displays possible. (This is television.) By keeping the electron beam small in transverse extent, small regions on the viewing surface may be excited. This permits the display of structure in the form of symbols or images.

In general, both LED's and CRT's are used in close, even confined, spaces at short ranges and their degradation by attenuation may be ignored. Both are militarily important primarily for the display of information either in a binary (LED) or structured (CRT) form.

Under clear nocturnal conditions when the Moon is not in the sky, some illumination is provided by starlight. This is a very low intensity light but it is sufficient to permit amplification.

In general, "natural" illumination: Sun; Moon; and stars: varies considerable, as illustrated in the table.

Time	**Situation**	**Luminance OM**	**Illuminance OM**
Day	Clear, sunlit	4	5
	Clear, shady	4	4
	Overcast	3	3
	Heavy Overcast	2	2
Sunset	Overcast	1	1
Sunset + 15 min	Clear	0	1
Sunset + 30 min	Clear	-1	0
Night	Bright Moon	-2	-1
	Clear, Moonless	-3	-3
	Overcast, Moonless	-4	-4

Sky Luminances and Illuminance Orders of Magnitude

The Luminance and Illuminance columns are given as representative Orders of Magnitude (OM), so values are 10^x where x is the column value. Luminance units are candelas per meter^2 and Illuminance values are lumens per meter^2. As may be seen Luminance typically varies by 8 OM while Illuminance typically varies by 9 OM during a day. Nighttime Overcast and Moonless illuminance (luminance) is the result of star light scattered by the clouds and overcast.

24.8 Contrast

A key concept in the simple analysis of imaging is that of contrast. Several definitions of contrast are used but we will only have to deal with two.

Inherent to the idea of contrast is that we have a target (an object to be viewed) and a background (surrounding light environment). Both are assumed to occupy at least part of the field of view of the sensor; both are assumed to possess an inherent intensity, indicated by I_{T0} and I_{B0} for target and background, respectively.

The simple contrast at the target, C_0, is defined by

$$C_0^s \equiv \frac{I_{T0}}{I_{B0}}, \tag{16}$$

which gives rise to a simple contrast at range r from the target that scales as transmittance (on the assumption that background intensity does not vary with range,)

$$C^s(r) = C_0^s e^{-\alpha r}. \tag{17}$$

(Transmittance, $T(r)$, is simply $e^{-\alpha r}$, where we have relaxed the notation on the attenuation coefficients.) This simple contrast transmittance is most often used in describing conditions in the far infrared where there is no direct sunlight and the atmosphere is assumed to essentially be in thermal equilibrium and uniformly mixed.[3]

Alternately, the inherent contrast at the target, designated by C_0^i, is defined by

$$C_0^i = \frac{I_{T0} - I_{B0}}{I_{B0}}. \tag{18}$$

Similarly, the inherent contrast at range r from the target is similarly defined by

$$C_0^i(r) = \frac{I_T(r) - I_B(r)}{I_B(r)}. \tag{19}$$

Obviously, simple and inherent contrast differ by an additive factor of -1.

If background intensity transmits (with range) in the same functional manner as target intensity, then obviously inherent contrast transmits as equation 17 for both types of contrast. On the other hand, if background intensity is uniform regardless of range, but target contrast transmits according to the Beer-Lambert-Bouguer Law, (equation 17 for simple contrast,) and equation 19 reduces to

$$C^i(r) = \frac{I_{T0}e^{-\alpha r} - I_{B0}}{I_{B0}} \tag{20}$$

[3]The terms Transmittance and Contrast Transmittance are somewhat confusing. Transmittance refers to the fraction of light that has reached a range r. Contrast Transmittance is the amount of contrast at range r.

$$= \frac{I_{T0}}{I_{B0}} e^{-\alpha r} - 1$$
$$= C^s(r) - 1.$$

If we now make use of equation 18, then we may further rewrite equation 20 as

$$C^i(r) = C_0^i e^{-\alpha r} - 1 + e^{-\alpha r}, \tag{21}$$

which indicates, as we would expect from the assumption, that

$$C^i(r) \to -1, r \to \infty. \tag{22}$$

As we shall see in the next section, there is a better way of approaching this problem.

24.9 Add the Sky.

For wavelengths up to about 4μ (the symbol common used for a micron,) the Sun is a significant source of light. For longer wavelengths, the atmosphere largely blocks out the Sun's light and the primary source of light is thermal emission. At short wavelengths (visual band), the gases of the atmosphere are strong scatterers, while at the longer wavelengths (far IR or thermal band), they are strong absorbers. In this case, we must consider the scattering of light from the sky into the line of sight (LOS) between target and sensor. We show this situation in Figure 5.

If we represent the intensity of sky light as I_S, and the scattering function of sky light into the LOS as $\beta(\theta)$, then the incremental amount of light $I_L(y)$ scattered into the LOS at range y from the sensor is

$$dI_L(y) = \beta(\theta) I_S e^{-\alpha y} dy. \tag{23}$$

To get the total amount of sky light scattered into the LOS, we need only integrate equation 23 along the LOS,

$$\int_o^x dI_L(y) = \beta(\theta) I_s \int_0^x e^{-\alpha y} dy, \tag{24}$$

which we may immediately see is just

$$I_L(x) = \frac{\beta(\theta)}{\alpha} I_s \left(1 - e^{-\alpha x}\right). \tag{25}$$

If we now combine this equation with equation 18 by adding $I_L(x)$ to both target and background intensities (to get the total intensities with scattered sky light,) then the resulting form of the inherent contrast at range x is

$$C^i(x) = \frac{I_T(x) - I_B(x)}{I_B(x) + I_L(x)}. \tag{26}$$

Add the Sky.

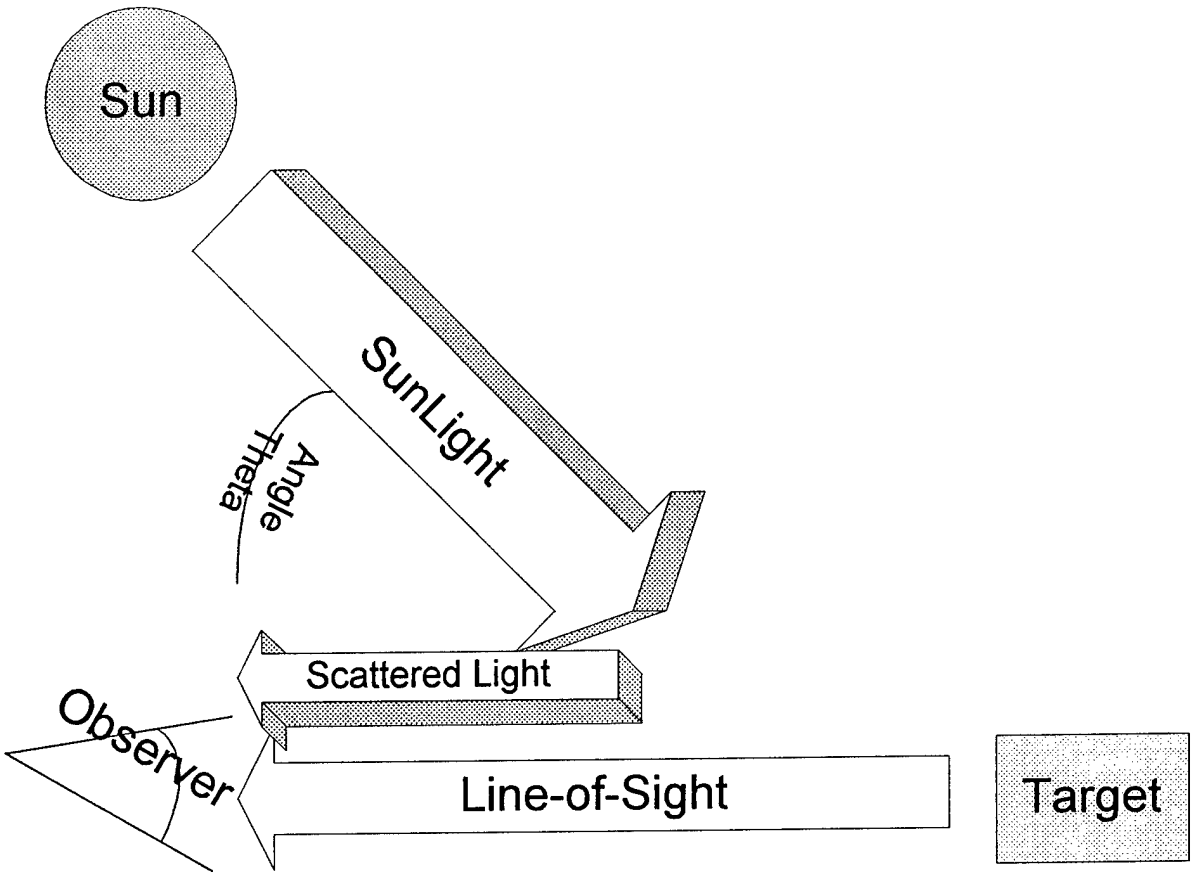


Figure 5: Sun Light Scattering into the Line of Sight

If we now replace each intensity with its functional form, and rearrange slightly, the result is

$$C^i(x) = \frac{I_{T0} - I_{B0}}{I_{B0} + \frac{\beta(\theta)}{\alpha} I_s \left(e^{\alpha x} - 1\right)}. \tag{27}$$

Now, if we make use of equation 18 to replace I_{T0}, the result is

$$C^i(x) = \frac{C_0^i I_{B0}}{I_{B0} + \frac{\beta(\theta)}{\alpha} I_s \left(e^{\alpha x} - 1\right)}. \tag{28}$$

Finally, if we define the sky-to-ground ratio as

$$S_g \equiv \frac{\beta(\theta)}{\alpha} \frac{I_s}{I_{B0}}, \tag{29}$$

then we may rewrite equation 28 as

$$C^i(x) = \frac{C_0^i}{1 + S_g\left(e^{\alpha x} - 1\right)}. \tag{30}$$

This is an important result. For the visual spectral band, this equation describes the transmission of contrast. On a clear day, when the sky light is bright, S_g is relatively large. On a cloudy or overcast day, S_g is not much greater than one. At night S_g is essentially one, and inherent contrast effectively transmits as

$$C^i(x)\mid_{night} = C_0^i e^{-\alpha x}. \tag{31}$$

The same situation applies in the thermal band where sky and ground have essentially the same emissivity and $S_g \sim 1$. In this case, inherent contrast reduces to simple contrast in terms of transmission.[4]

To illustrate this, we present several plots of contrast versus range for an inherent initial contrast of 2, an attenuation coefficient of 0.1 (per unit range), and several sky-to-ground ratios in Figure 6 The effect of sky-to-ground ratio on the transmission of inherent contrast can be clearly seen.

There is one more point to be considered here. We noted earlier that scattering is angle dependent. Let us now consider this effect. We approximate the scattering function as

$$\beta(\theta) \simeq \frac{\alpha}{2}\left(1 + \frac{1}{2}\cos(\theta)\right). \tag{32}$$

In this case, equation 30 becomes

$$C^i(x) \simeq \frac{C_0^i}{1 + \frac{1}{2}\frac{I_g}{I_{B0}}\left(1 + \frac{1}{2}\cos(\theta)\right)\left(e^{\alpha x} - 1\right)}. \tag{33}$$

The angle θ is that between the line to the Sun and points on the LOS. This angle is essentially constant since the Sun is so far away. The geometry is shown in Figure 5. In Figures 7, and 8, we plot the apparent contrast of the same situation as Figure 6, with respect to angle. The sky-to-ground ratios used are 1 and 2, respectively (measured with the sun at zenith.)

Let us now consider the effect of this on combat. Consider the forces arranged in line. In general, the angle of the Sun with one force's LOS will be θ, while the angle of the Sun with the other force's LOS will be $180 - \theta$ (since the LOS are approximately symmetric.) Further assume that the initial inherent contrast of targets is the same for both sides. The geometry of this is shown is Figure 9.

As we can see from this figure, the side that has the Sun in their faces sees a lower target apparent contrast than the side with the Sun over their backs. Turned around, seeing targets

[4]At this point, the student may be getting a bit confused over the "T" words. This seems to be built into the naming system deliberately. To try to clarify things, transmission is the process and transmission is the mathematical factor (function) of transmission. Thus light transmits (active voice) or is transmitted (passive voice) and the fraction of light that arrives is the transmittance.

Add the Sky.

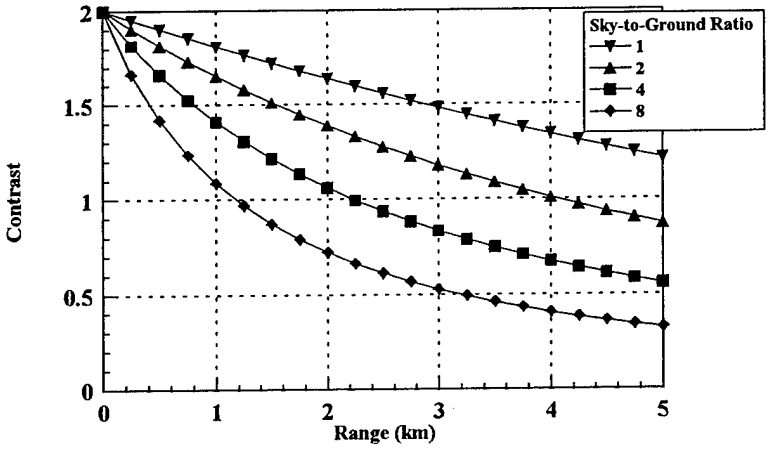


Figure 6: Inherent Contrast versus Range

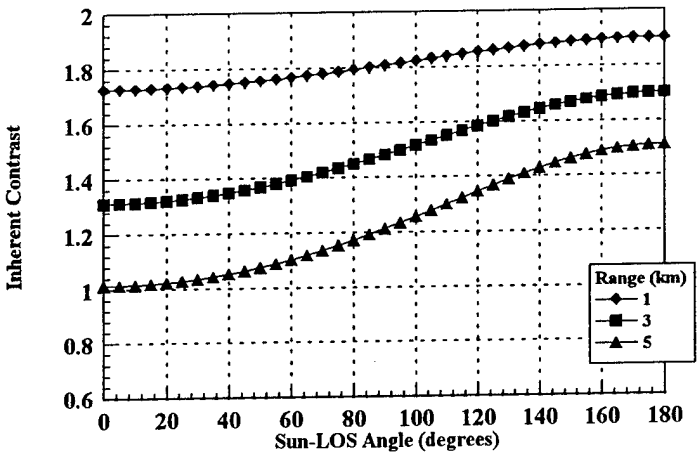


Figure 7: Inherent Contrast versus Sun Angle, Sky-to-Ground = 1

"in" the Sun is more difficult that seeing targets illuminated by the Sun directly. While this is a mere expression, in technical terms of common knowledge, contrast is not the whole story, as we shall see subsequently.

We note that this difference decreases in proportion, and contrast decreases with range, so the effect is most pronounced at short ranges. When we take up probability of detection in a later chapter, we shall see that the leveling of contrast with increasing range is offset by the decrease in probability of detection with range.

Before proceeding, it is useful to consider some typical (representative) Sky-to-Ground ratios. These are given in the table below.

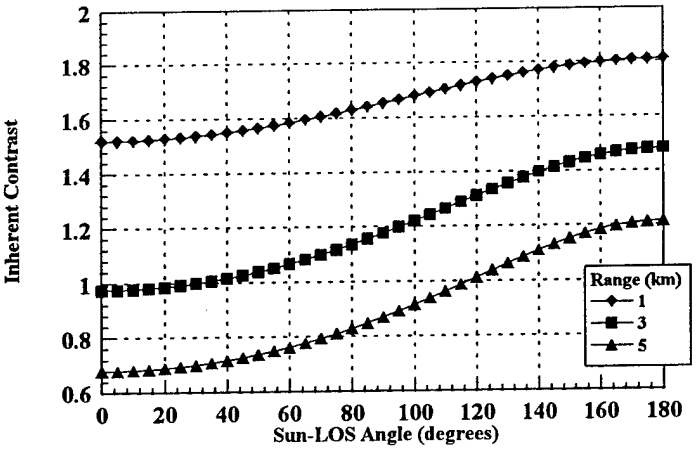


Figure 8: Inherent Contrast versus Sun Angle, Sky-to-Ground = 2

Sky	Ground	S_g
Clear	Forest	5.0
	Desert	1.4
	Snow	0.2
Overcast	Forest	25.0
	Desert	7.0
	Snow	1.0

Typical Sky-to-Ground Ratios

As we have seen, Sky-to-Ground ratio depends on sun angle, so these values should be taken as representing averages.

24.10 The Size of Targets

While we normally measure objects by their physical dimensions, it is convenient to express the

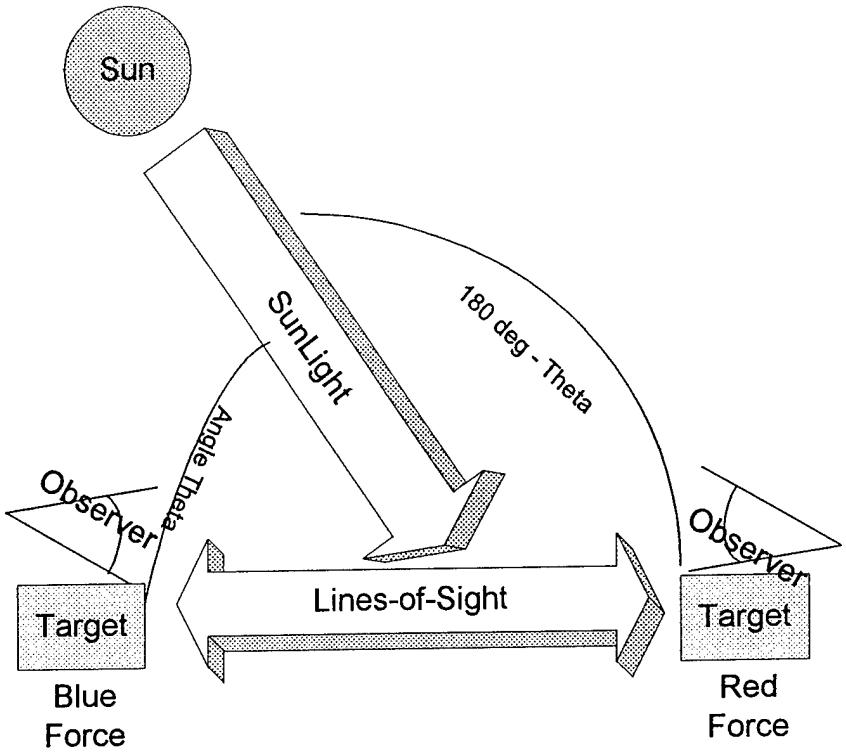


Figure 9: Force Assymetry of Sky-to-Ground Ratio

size of targets in terms of solid angles. Assume the target has a height of h and a width of w as seen by the observer. Further assume that the target is located at a range r from the observer.

Let θ be the vertical angle of the target (as shown in Figure 10.) Similarly, let ϕ be the horizontal angle of the target. We may calculate the value of θ (and ϕ) by noting the right angle relationship of the half angles,

$$\tan\left(\frac{\theta}{2}\right) = \frac{h}{2r}, \tag{34}$$

At long ranges, $r \gg h$, w, so that the fraction on the right hand side of equation 34 is small, and we may replace the tan function with its small angle approximation (a one term Maclaurin's series expansion), so that

$$\theta \simeq \frac{h}{r}, \tag{35}$$

and similarly for ϕ,

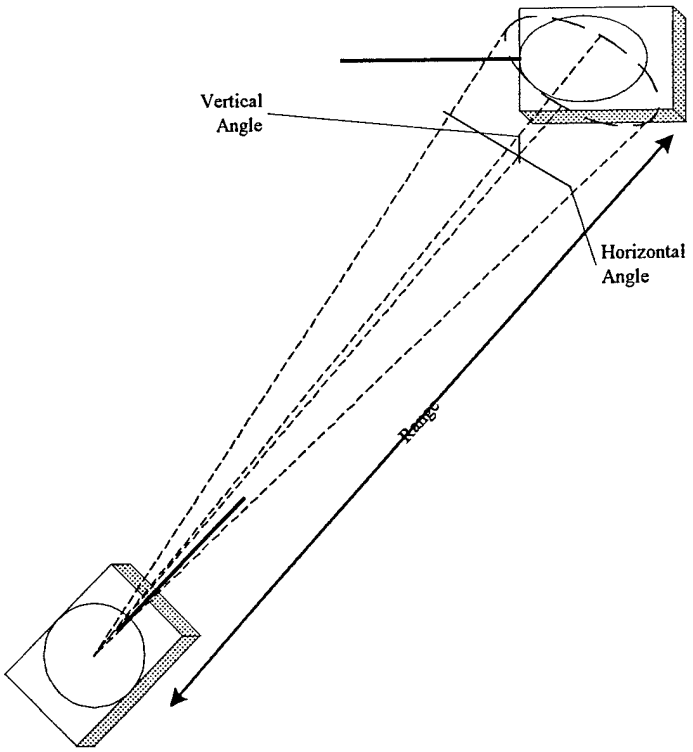


Figure 10: Target Solid Angle

$$\phi \simeq \frac{w}{r}. \tag{36}$$

The solid angle of the target, Ω, is defined on the surface of a unit sphere, so the total solid angle possible is 4π steradians. The differential of the solid angle is

$$\begin{aligned} d\Omega &= \sin(\theta)\, d\theta d\phi \\ &= d\cos(\theta)\, d\phi. \end{aligned} \tag{37}$$

We may calculate the approximate solid angle of the target by integrating over the target's area. To do this, we make a change of variables to target dimensions z (height) and y (width), defined about the center of the target. We further assume that the center of the target is located at $\theta = \frac{\pi}{2}$ in our spherical coordinate system. Then,

$$\sin(\theta)\, d\theta = \frac{dz}{r}, \tag{38}$$

and

$$d\phi = \frac{dy}{r}, \tag{39}$$

so that

$$\begin{aligned} \Omega &= \frac{wh}{r^2} \\ &= \frac{a}{r^2}, \end{aligned} \tag{40}$$

where: a is the area of the target as viewed by the observer. We emphasize that this formalism is valid only when the target dimensions h, w are small compared to the range r.

24.11 The Signatures of Targets and Backgrounds

The term signature has become a common one in military usage with the advent of modern sensor systems. In that context, it generally means an observable (even measurable) characteristic. It usually has two (complementary) connotations: in intelligence circles, signature is often used to connote not just an observable, but a confirming, characteristic of something. In this connotation, a signature at least contributes to the positive identification of the object, and ideally is sufficient to uniquely, positively identify the object. In this connotation, signature is causally linked to identification directly.

The second connotation is more common in usage and is used to mean the observable (although usually optical and acoustic) characteristics of objects of tactical interest, including the backgrounds that military objects may be seen in and against. While this connotation also links signature and detectability/identification, it is through the methodology of search and detection (discussed in a later chapter.) Thus the distinction between the intelligence and tactical connotations is that search is performed to find the former to generate identification while the latter is a quantity that is sensed during search.

We shall tend to use the tactical connotation in this work. Specifically, we shall be concerned here with some of the optical characteristics of objects.

For light wavelengths less than about 4 μ, where the atmosphere effectively cuts off the solar spectrum, the optical signatures of most objects (neglecting, of course, light sources described above,) arises from reflected sunlight (moonlight, starlight,) and the magnitude of their signature is dependent on two quantities: ambient luminance in the environment; and the object's reflectivity. Simply put, reflectivity is the measure of the fraction of incident light that is reflected from an object. Since all of the incident light must be reflected, absorbed, or transmitted, the sum of reflectivity, absorptivity, and transmissivity must be one. These quantities differ somewhat from our previous consideration of the optics of the atmosphere in that these are bulk quantities. Reflection is expressed as a surface condition while absorption and transmission are expressed as volume conditions.

There are two types of reflectivity: specular and diffuse. Specular reflectivity is the optical ideal, and is the basis for the design of simple optical systems of lenses and mirrors. Simply put, if a fraction of an incident ray of light is reflected, the plane angle of the reflected ray with respect to the surface normal has the same magnitude as the plane angle of the incident ray (but on the opposite side of the normal.) It is this relation with the surface normal that prompts reflection description as a surface phenomena even though it depends on the volumetric properties of the two media. Although not usually couched in these terms, reflection is an idealized collision between the incident light and the surface. Momentum transverse to the surface normal is conserved and unchanged in the collision. If we view this at the level of individual photons (which are massless in the sense of normal masses,) they impart no energy to the surface's material unless they are absorbed. Thus energy conservation dictates the geometry of the reflected ray to the incident ray.

For simple materials, there is a simple mathematical relationship for reflectivity:

$$r = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2, \tag{41}$$

where:

r = reflectivity[5],

n_1 = refractive index of the incident medium, and

n_2 = refractive index of the reflecting medium. From this, it is obvious that reflection is a considerable factor only if there is significant difference between the refractive indices of the two media.

Minimization of specular reflectivity is an important consideration in the design and manufacture of military optical devices.

If all reflection were specular, then vision in the world would be very much different from what we are used to - more akin to the conditions a highway (with traffic) at night after a rainstorm (with standing puddles,) where all of the light comes from headlights (no street lights!) Happily, the other type of reflectivity - diffuse - reflects light in all directions away from the surface. This type of reflection phenomena is most common with rough surfaces (that is, not optically smooth.) This type of surface is the norm, especially under field conditions where mud and dirt frequently cover vehicles and equipment. By its very nature (either woven or knitted,) cloth is a rough surface. For this type of reflection, energy (wavelength) is conserved for reflected light but not momentum (although many diffuse reflectors have a strong specular component.) Thus, diffuse reflection spread incident light in all (outward) directions so the objects reflecting the light may be seen.

Since diffuse reflection is the secondary source of the light that we see objects by, it is useful to characterize the environment by representative reflectivities. These are shown in the table below.

[5]The term reflectance may also be found in use. We shall use the terms synonimously with a preference for reflectivity.

Reflector	Minimum	Maximum
Lake or Stream	0.05	0.10
Snow	0.70	0.86
Sand	0.15	0.30
Alabama Clay Soil	0.7	0.15
Illinois Black Earth	0.03	0.05
Conniferous Forest	0.03	0.10
Deciduous Forest	0.05	0.15
Grass	0.10	0.25
Crops	0.07	0.15
Dirt Road	0.03	0.05
Asphalt Road	0.08	0.10
Concrete Road	0.15	0.35
Buildings	0.09	0.25

Representative Reflectivities

The reflectivities of military vehicles (and contemporary field uniforms) vary somewhat but are usually in the range 0.15-0.30. Specific values tend to vary with the surface moisture and dirtiness of the vehicles. For this reason, it is common to parametrize contrasts rather than relectivities in many calculations. This is more due to the difficulty of accurately estimating reflectivity than calculation difficulty. To illustrate this, we return to the definition of target inherent contrast, equation 18, assume that targets and backgrounds are perfectly diffuse reflectors, and that all illumination derives from skylight. Then the inherent intensities of target and background are

$$\begin{aligned} I_{T0} &= r_T I_S, \\ I_{B0} &= r_B I_S, \end{aligned} \tag{42}$$

where: r_T, r_B are the target and background reflectivities, and I_S is the intensity of skylight.. From this, we may immediately calculate the inherent contrast of the target (and cancel common factors of I_S) as

$$C_o^i = \frac{r_T - r_B}{r_B}, \tag{43}$$

so that we may see that the inherent contrast of the target (to its background) is just the percentage difference of the target's reflectivity from that of its background. This is very handy. It means that if we make uniforms bright (and keep them clean), then troops have a high inherent contrast and are easy to see. This maximizes their visibility (and detectability) on a dirty battlefield. Alternately, if the make uniforms look like the background (either OD - Olive Drab, or camouflage[6],) then troops have a small inherent contrast and are hard to see. This minimizes

[6]We will discuss camouflage later in the chapter on search and detection.

their visibility for the battlefield environment for which their uniforms are "tuned".

During the era of bright uniforms, Frederick the Great - Napoleon - Crimean War - Colonial Wars, mobility was low, basically foot (or horse) paced. Concentration of firepower was essentially a matter of concentration of force. Brightly colored uniforms facilitated command, control, and cohesion under these conditions. With technology increasing the range of fire (e.g., conoidal bullet, rifling, breech loading, improved propellants,) and increased literacy (democracy and nationalism,) the battlefield became more dispersed (less dense.) Before, unit survivability (remain together and out-fire your opponent) was the source of individual survivability. On the dispersed battlefield, individual survivability (and effectiveness,) became the source of unit survivability (and effectiveness.) Thus uniforms changed from brightly colored to background colored for reasons that have a firm foundation in basic physics.

It is also interesting to examine these same assumptions in terms of the sky-to-ground ratio, equation 29. That is,

$$S_g = \frac{\beta(\theta)}{\alpha r_B}, \tag{44}$$

in the terms we have defined just above. If the atmosphere were essentially scattering (as it is in the visible spectral band,) and $\beta(\theta)$ were a constant, then we would have a clean inverse relationship between background reflectivity and sky-to-ground. If we examine the tables of representative sky-to-ground ratios and representative reflectivities, we may see that except for snow (which is strongly specular and therefore have a very unconstant β!,) that to within a range of factors of $\sim (\frac{1}{2}, 2)$ (as we would somewhat expect from a Raleigh scattering function,) there is a weak inverse relationship that can be used as an approximation in calculations. As we have already mentioned, the difficulty in using this approach is more due to the problem of estimating reflectivity than of mathematics.

For wavelengths longer than the visible but less than the cut-off, there is little open literature data and we therefore cannot comment further on this here.

For wavelengths longer than the atmospheric cut-off of 4 μ, the source of light is the black body emissions of individual objects. While we can calculate total signatures by integrating equation 4 over the wavelength band of interest, a simpler and usually sufficient approach is to make use of the total integral over all wavelengths. This is know as the Stefan-Boltzmann law and it has the form

$$M = \sigma T^4, \tag{45}$$

where:

M = exitance with units of Watts m^{-2},

σ = Stefan-Boltzmann constant $\simeq 5.7x10^{-8}$ Watts m^{-2} deg(K)$^{-4}$, and

T = absolute temperature.

If we again assume target and background are perfectly diffuse reflectors, that we may scale the fraction of light in the spectral region of consideration as a fraction of the whole, and that

this fraction is the same for both target and background, then the inherent contrast of the target to background for long wavelengths is approximately

$$C_0^i \simeq \frac{T_T^4 - T_B^4}{T_B^4}. \tag{46}$$

If we further assume that the target temperature differs from the background temperature by an amount ΔT, and this quantity is small compared to T_B, ($T_B \gg \Delta T$,) then to first order, equation 46 reduces to

$$C_0^i \simeq \frac{\Delta T}{T_B} \equiv \Delta T\%, \tag{47}$$

where: $\Delta T\%$ is the target-background temperature expressed as a percentage of the background temperature. This is a pure number.

24.12 Visual Range

The last topic that we take up in this chapter is visual range, which is essentially a meteorological measurement. If we consider a black object of well defined dimension and shape placed against a colored or white background, this object has an initial (or inherent) contrast of -1. If we assume that such an object will be barely visible when it has a perceived contrast of ± 0.02 (two percent), then the range at which this occurs is called the visual range.

If we further stipulate that this will be measured when the sky-to-ground ratio is one (e.g., near dawn,) then the apparent contrast transmits as

$$C^i(x) = C_0^i e^{-\alpha x}. \tag{48}$$

From the statement of the problem, we may write this as

$$-0.02 = -1e^{-\alpha r_{vis}}, \tag{49}$$

where r_{vis} is the visible range. This equation may be solved to yield

$$\alpha = \frac{\ln(50)}{r_{vis}}, \tag{50}$$

which is known as Koschmeider's relation. This equation is a useful way of estimating visual attenuation coefficients and is widely used.

24.13 Fields

The term field has a wonderful multiplicity of meanings in our strange mixture of physical and military matters. A field has a definite mathematical and several physical meanings to the

physicist, such as an electric or magnetic field. To the soldier, the field is where one conducts operations and perhaps, fights. It may denote an area of knowledge, as a field of study. In the optics sense that we are considering here, it generally means a solid angle (or a combination of two plane angles, usually called azimuth and elevation.) We have already laid the groundwork for solid angles in the section (above) on target size, and we have discussed the idea of Field of Illumination (FOI) (above) in the section on light sources. As we shall use these "field'" terms they will generally mean a usually small solid angle. Fields greater than about 30° in azimuth and 15° in elevation will be the exception. The most common fields we shall consider are:

- Field of View (FOV) which is the solid angle that a sensor can see,
- Field of Regard (FOR) which is the solid angle that a sensor may consider,
- Field of Illumination (FOI) which is the solid angle that light is projected into, and
- Field of Search (FOS) which is the solid angle that search is conducted over.

There is a precise optical definition of FOV that is not crucial to our discussion here. The term Field of Regard may be context specific. It is sometimes used to designate the part of the total FOV that is actually displayed (for an electronic imaging system (e.g.) It may also be used to designate a field larger than the FOV for autonomous sensor systems (e.g., missile seekers,) and is somewhat of a synonym for FOS in this context. For multiple detector sensors, the Instantaneous FOV (IFOV) is frequently used to designate the FOV of the individual detectors.

24.14 References

[1] Meyer-Arendt, Jurgen R., **Introduction to Classical and Modern Optics**, Prentice-Hall, INC., Englewood Cliffs, NJ, 1972.

[2] Saleh, Bahaa E. A., and Marvin Carl Teich, **Fundamentals of Photonics**, John Wiley & Sons, INC., New York, 1991.

[3] .."Quantitative Description of Obscuration Factors for Electro-Optical and Millimeter Wave Systems", DOD-HDBK-178(ER), 25 July 1986.

[4] Kneizys, F. X., et al., "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN 5", U. S. Air Force Geophysics Laboratory, Hanscom Air Force Base, MA, Report AFGL-TR-80-0067, Feb. 1980.

[5] Wolfe, William L., and George J. Zissis, eds., **The Infrared Handbook**, Environmental Research of Michigan, 1978.

Chapter 25

Optics II

25.1 Introduction

Sometimes, it is necessary to consider not just the bulk transmission of light through the atmosphere, but the actual propagation of the light on a detailed basis. That is the primary concern of this chapter.

25.2 Fourier Transforms

Before considering the detailed propagation of light through the atmosphere, we must first lay some mathematical basis for dealing with this theory. This type of mathematics is part of the general area of integral transforms.[1]

25.2.1 One Dimensional Fourier Transforms

Assume $f(t)$ is a mathematical function that is well defined and behaved everywhere. Define the *Fourier Transform* of $f(t)$ as

$$\begin{aligned} F[f(t)] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \\ &\equiv f(\omega), \end{aligned} \tag{1}$$

where: $i \equiv \sqrt{-1}$.[2][1] The *Inverse Fourier Transform* is

[1] We shall take up another type of integral transform, the Laplace transform, in the chapter on renewal theory.

[2] The knowledgable reader will note that I have made two convention choices here: First, I have chosen to use the notation for the complex unity as i rather than j. The former is commonly used by mathematicians and physicists, while the latter is used by engineers. The common mythology is that physicists use i because they use j as current density while engineers use j because they use i as current. Since we also commonly use i, j as indices

$$F^{-1}\left[f\left(\omega\right)\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f\left(\omega\right)e^{i\omega t}d\omega \equiv f\left(t\right). \tag{2}$$

Note that this form of the Fourier transform is symmetric so that the Fourier transform of $f\left(t\right)$ is $f\left(\omega\right)$ whose Fourier transform is $f\left(t\right)$.Some particularly important properties of Fourier Transforms are Borel's Convolution Theorem,

$$F\left[\int_{-\infty}^{\infty} f_1\left(t'\right)f_2\left(t-t'\right)dt'\right] = F\left[\int_{-\infty}^{\infty} f_1\left(t-t'\right)f_2\left(t'\right)dt'\right] = f_1\left(\omega\right)f_2\left(\omega\right), \tag{3}$$

and Parseval's Theorem,

$$\int_{-\infty}^{\infty} f_1^*\left(t\right)f_2\left(t\right)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} f_1^*\left(\omega\right)f_2\left(\omega\right)d\omega. \tag{4}$$

The Fourier Transform of $f\left(t\right)$, $f\left(\omega\right)$ is sometimes called the *spectrum* of f. By definition, the Fourier Transform of a Dirac delta function $\delta\left(t-t'\right)$ is

$$F\left[\delta\left(t-t'\right)\right] = \frac{1}{\sqrt{2\pi}}e^{-i\omega t'}. \tag{5}$$

Similarly, the Fourier Transform of the nth derivative of a function is

$$F\left[\frac{d^n f\left(t\right)}{dt^n}\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{d^b f\left(t\right)}{dt^n}e^{-i\omega t}dt = \left(i\omega\right)^n f\left(\omega\right), \tag{6}$$

assuming that $f\left(t\right) \rightarrow 0$ in the limit that $t \rightarrow \pm\infty$, otherwise there are other terms in the derivative. If $f\left(t\right)$ is not bounded as $t \rightarrow \pm\infty$, then the Fourier Transform is not defined. This property is very valuable in solving linear differential equations.

25.2.2 Multi-dimensional Fourier Transforms

The Fourier Transform is directly extensible to multiple dimensions. If $\underrightarrow{r}$ is a spatial vector of dimension n, (that is, it has n independent components and directions,) then the Fourier Transform of the scalar function of vector argument, $f\left(\underrightarrow{r}\right)$, is

$$F\left[f\left(\underrightarrow{r}\right)\right] = \frac{1}{\left(2\pi\right)^{\frac{n}{2}}}\int_{-\infty}^{\infty} f\left(\underrightarrow{r}\right)e^{-i\underrightarrow{\omega}\bullet\underrightarrow{r}}d\underrightarrow{r}^n = f\left(\underrightarrow{\omega}\right), \tag{7}$$

for summations, matrices, etc,.it is probably a matter of preference. The second convention is that I have chosen to use the symmetric representation for the Fourier transform. There are at least two other, assymetric conventions possible, but since I prefer to exploit symmetries, I have chosen this convention.

where: $d\underset{\rightarrow}{r}^n \equiv dr_1..dr_n$. The Inverse Fourier Transform is defined in an analogous manner as

$$F^{-1}\left[f\left(\underset{\rightarrow}{\omega}\right)\right] = \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{-\infty}^{\infty} f\left(\underset{\rightarrow}{\omega}\right) e^{i\underset{\rightarrow}{\omega}\bullet\underset{\rightarrow}{r}} d\underset{\rightarrow}{\omega}^n. \tag{8}$$

25.2.3 An Alternate Representation

The optics community, comprised of both physicists and engineers, also commonly employs a different, but still symmetric form of the Fourier Transform. In one dimension, the transform is

$$F\left[f\left(t\right)\right] = \int_{-\infty}^{\infty} f\left(t\right) e^{-2\pi i\nu t} dt, \tag{9}$$

and the inverse transform is

$$F^{-1}\left[f\left(\nu\right)\right] = \int_{-\infty}^{\infty} f\left(\omega\right) e 2\pi^{2\pi i\nu t} d\nu. \tag{10}$$

This form has the advantage of not having a normalization factor to keep up with. On the surface it would appear that the two forms, equations 1 and 2, and equations 9 and 10, are simply related by $\omega = 2\pi\nu$. This is not quite the case. If we substitute this relation into equations 9 and 10, the result is not equations 1 and 2, but an asymmetric transform pair with all of the 2π factor in the inverse transform (arising from $d\omega = 2\pi d\nu$.) We shall not distinguish strongly which of these two transform representations is used in most instances. It may be helpful however, to note that ν is often called frequency, while ω is called angular frequency.

25.3 Linear System Theory

At this point, it is useful to spend a little time discussing Linear System Theory. Simply put in the terminology that we have introduced in Part I, we want to consider a linear system (a process which may be physical, organizational, human behavioral, etc.,) which has time dependent effect. That is, if an input is given to the process at time t, then output is produced at time t' $(\geq t)$. The greater than restriction is a direct result of the assumption of causality. In principle, if the process had a length of l, then in a strict sense, $t' \geq t + \frac{l}{c}$, where c is the speed of light in the process (physical) space. For short distances, this correction may be ignored compared to common delays in many processes (not necessarily physical.) In general, we will want to consider a case where the input (and probably the output) is a function of time. This is depicted in Figure 1.

The input function (or signal as it is often called,) is represented by $f_i\left(t\right)$ and the output function by $f_o\left(t'\right)$. Under the assumption of superposition (which is necessary for subsequent harmonic analysis,) the two are related by the integral

$$f_o\left(t'\right) = \int_{-\infty}^{\infty} h\left(t', t\right) f_i\left(t\right) dt, \tag{11}$$

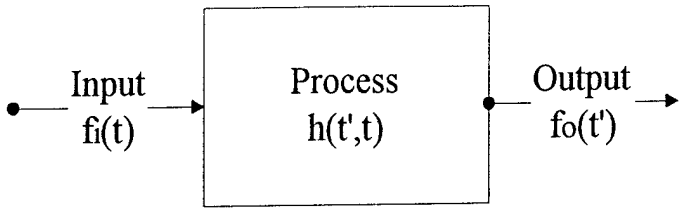


Figure 1: Linear System Diagram

where: $h(t',t)$ is the process effect function. By assumption, $h = 0$ for $t' < t$. If the process' effect is time invariant, then the process (or linear system) is said to be *time* or *shift invariant*.[2] By way of example, if the only effect of the process is a delay (e.g., the transmission of light over a distance l as indicated above,) then $h(t',t) = \delta\left(t' - t - \frac{l}{c}\right)$, and by the properties of the Dirac Delta function, we may evaluate Equation 11, as

$$f_o\left(t + \frac{l}{c}\right) = f_i(t), \tag{12}$$

or more interestingly and useful,

$$f_o(t) = f_i\left(t - \frac{l}{c}\right). \tag{13}$$

A direct result of this is that the process effect function simplifies to be a delay function only. That is, it is a function only of the difference $t' - t$. Thus, equation 11 becomes

$$f_o(t') = \int_{-\infty}^{\infty} h(t' - t) f_i(t)\, dt, \tag{14}$$

where: by definition, $h(x) \equiv 0, x < 0$.

An important result of this is that it permits the defining equation 14, to be made manipulatively and calculationally simpler by use of Fourier Transforms. By virtue of Borel's Convolution Theorem, equation 3, the Fourier Transform of equation 14 is just

$$f_o(\omega) = h(\omega) f_i(\omega). \tag{15}$$

The Fourier Transform of the process effect function, $h(x)$, is called the *Transfer Function* (of the process or linear system.)

25.4 Image Transmission

Now we finally come to the transmission of images through the atmosphere and optical components (i.e., lenses and mirrors). As we may have surmised from the foregoing discussion, this

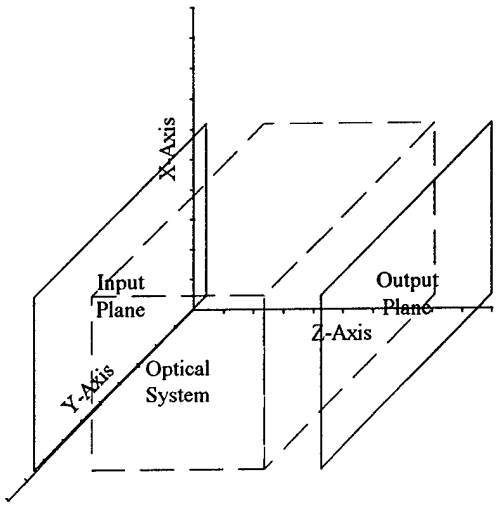


Figure 2: Optical Wave Transmission

is a linear system, generalizing somewhat in that the input and output ate two-dimensional, and the process is three-dimensional. This is depicted in Figure 2.

In this case, there is an input plane that is idealized but generally assumed to be in the far field; an output plane that is usually assumed to be the focal plane of the optical instrument, and an optical system that includes the optical components of the instrument plus the environment between the input and output planes. The usual selection of coordinate system is as shown: the output plane defines the x-y plane, the z axis lies along the optical axis normal to the output plane but usually defined with origin at the object. The orientation of the object plane is imposed by the selection of output plane. The output plane includes both the object to be observed (target) and its background.

In a field sense, the amplitude of the light is given by $U\left(\underrightarrow{r}\right) = U(x, y, z)$. By definition, the amplitude in the input plane is denoted by

$$f(x, y) = U(x, y, 0), \tag{16}$$

while the amplitude in the output plane is denoted by

$$g(x,y) = U(x,y,d), \tag{17}$$

where: d is the distance between input and output planes.

We can construct the free space transmission of the image simply under the assumption that harmonic analysis is possible. The (complex) amplitude of a plane wave is

$$U_{pw}\left(\underrightarrow{r}\right) = Ae^{-i\underrightarrow{k}\bullet\underrightarrow{r}}, \tag{18}$$

where: $\underrightarrow{k}$ is the wavevector with amplitude $\left|\underrightarrow{k}\right| = \frac{2\pi}{\lambda}$, λ being the wavelength of the wave, and A is the complex amplitude of the plane wave. Since by the properties of the Fourier Transform we may write the transform of equation 17 as

$$f(\omega_x,\omega_y) = F^{(x,y)}\left[U(x,y,0)\right], \tag{19}$$

where we have indicated the variables of transform as superscripts, we may infer the general from of U for our input as

$$U(x,y,z) = F^{-1(\omega_x,\omega y)}\left[f(\omega_x,\omega_y)\right]e^{-ik_z z}. \tag{20}$$

By generalizing this, we may form the transfer function of free space as

$$H(\nu_x,\nu_y) = \exp\left[-2\pi i\left(\frac{1}{\lambda^2} - \nu_x^2 - \nu_y^2\right)d\right], \tag{21}$$

where: $k_j = 2\pi\nu_j$, and we have made use of $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$, and the definition of the amplitude of the wave vector. This gives us the general form,

$$g(x,y) = \int\int_{-\infty}^{\infty} H(\nu_x,\nu_y)\, f(\nu_x,\nu_y)\, e^{-2\pi i(\nu_x x + \nu_y y)} d\nu_x d\nu_y, \tag{22}$$

where:

$$f(\nu_x,\nu_y) = \int\int_{-\infty}^{\infty} f(x,y)\, e^{2\pi i(\nu_x x + \nu_y y)} dx dy. \tag{23}$$

There are a variety of transfer functions in use in optics, all going by various, possibly ambiguous or degenerate, names. The transfer function of a point source object is called variously: the Point Spread Function (PSF); the impulse response; Green's function;[3] and the Fraunhofer diffraction pattern.[3] The PSF is computed from the Complex Pupil Function $\Phi(\rho,\theta)$ which is defined by

$$\Psi(\rho,\theta) = P(\rho,\theta)\, e^{2\pi i W(\rho,\theta)}, \tag{24}$$

[3]My Mathematical Methods of Physics teacher at the University of Illinois, Dr. Ting, was adamant that the proper usage was Green Function, not Green's Function. I try to use that form when I am writing, but follow the author's form when I am citing.

where:

$P(\rho,\theta)$ = amplitude transmittance pupil function, and

$W(\rho,\theta)$ = wavefront error pupil function. The reader will note that the Complex Pupil Function is defined in circular coordinates (dictated by the circular symmetry of most geometric optical systems.) The functions may be transformed into normal Cartesian coordinates by the relations,

$$\begin{aligned} x &= \rho\cos(\theta), \\ y &= \rho\sin(\theta). \end{aligned} \tag{25}$$

The inverse relations are, of course,

$$\begin{aligned} \rho &= \sqrt{x^2+y^2}, \\ \theta &= \arctan\left(\frac{y}{x}\right). \end{aligned} \tag{26}$$

The wavefront error pupil function is a measure of the advancement or retardation of the local wavefront relative to the reference. As its name implies, the amplitude transmittance pupil function represents the transmittance of the pupil (i.e., optical system.) For a perfect circular lens of radius 1, the amplitude transmittance pupil function is, by definition,

$$\begin{aligned} P(\rho,\theta) &= 1, \rho \leq 1, \\ &= 0, \rho > 1. \end{aligned} \tag{27}$$

and $W(\rho,\theta) = 0$.These values correspond to ideal (or perfect) conditions that are never realized in practice, so the real world value of these functions are different in practice.

The amplitude of the PSF is calculated from the Complex Pupil Function by the integral

$$\Theta(x,y) = C\int\int_{-\infty}^{\infty} \Phi(\varepsilon,\eta)\exp\left[-\frac{2\pi i}{\lambda R}(x\varepsilon+y\eta)\right] d\varepsilon d\eta, \tag{28}$$

where: R is effectively the distance from object to image, and C is a normalization constant. This equation may also be cast in circular coordinates by introducing specific lens parameters. It is worth noting that the energy distribution is given by

$$E(x,y) = \Theta(x,y)\,\Theta^*(x,y), \tag{29}$$

where: $*$ indicates complex conjugation. The PSF is not generally used for image calculations except for small objects (i.e., points) and for laser sources.

The transfer function most often used for continuous tone, spatially extended objects-images is the Optical Transfer Function (OTF.) A common means of characterizing the OTF is to

use the metaphor of a grating (effectively an alternating bar code pattern) of period p (half of the contiguous space along the pattern for a distance p is black and the other half white or transparent). The grating is oriented at the Target Orientation Angle (TOA) α with respect to the image-object plane y axis. The intent is to align the pattern along the minimum dimension of the target.[4] Since TOA is ambiguous $\pm 180°$, it may be used interchangeably in object and image planes even with inversion of the image in the optical system.

The maximum and minimum values of energy (irradiance) passing through the grating imposed on the object and image planes define the object and image modulation contrasts,

$$C_{m,image} = \frac{\max(E_{image}) - \min(E_{image})}{\max(E_{image}) + \min(E_{image})}, \tag{30}$$

and similarly for $C_{m,object}$. Although we have dropped the x, y functionality of E for brevity, the max,min functions refer to the maximum, minimum values of (in this case) E for all values of x, y in the object, image planes, respectively. (Note that these planes are spatially defined by the FOV of the optical system for the object plane, and the physical size of the optical system focal plane for the image plane.) Modulation is used to eliminate any consideration of average energy, and is a valid concept for linear systems. Thus, it cannot be directly applied to photographic film which is inherently non-linear. Since the modulation is specifically defined for a particular grating spacing and TOA, it is generally represented as a function of spatial frequency (inverse spacing) and TOA.

Another important parameter in image transmission is the *spatial frequency* of the object and image, defined by

$$\nu_{image} = p_{image}^{-1}, \tag{31}$$

and similarly for the object, where p_{image} (object) are the grating periods of the image (object) gratings. Spatial frequency is usually measured in cycles per unit length, although other units may be used in specialized instances and applications. Two commonly used terms are the pixel (picture element) which is the center to center spacing of the detector (display) elements in an array, and the resel which is a half cycle at some spatial frequency. These two terms are often used interchangeably. The contemporary use of the term pixel has come to be associated with display arrays (e.g., monitors) although it is primarily used here as referring to detector arrays (or their processing.)

Optical systems serve as low pass filters which do not pass harmonic components of frequency greater than the cutoff frequency, given by

$$\nu_{co} \equiv \frac{1}{\lambda F}, \tag{32}$$

[4]Thus, the TOA for a ground vehice seen broadside is nominally 0° while that for a standing human figure is nominally 90°. For an artillery piece, seen broadside, the bar pattern would be placed approximately normal to a nominal barrel angle, so a TOA of $\sim \pm 30°$ would be representative.

where: the focal ratio or F number is given by

$$F = \frac{f_i}{D_p}, \tag{33}$$

and f_i is the image (focal) plane focal length (from geometric optics,) and D_p is the entrance pupil diameter of the optical system.

Like the Complex Pupil Function, the Optical Transfer Function consists of two parts,

$$O\left(\nu_{image}, \alpha\right) = M\left(\nu_{image}, \alpha\right) e^{i\Phi(\nu_{image}, \alpha)}, \tag{34}$$

the *Modulation Transfer Function* (MTF), a measure of the shift in grating contrast modulation, defined by

$$M\left(\nu_i, \alpha\right) \equiv \frac{C_{m,i}}{C_{m,o}}, \tag{35}$$

where we have shortened image, object subscripts to i,o, and the Phase Transfer Function (PTF) which is a measure of the shift of the image grating in position relative to the object grating.. A nominal and computationally desirable engineering approximation is that the PTF be zero. The advantage of the OTF is that for a linear system, the MTF of the components are multiplicative and the PTF are additive. For a perfect lens with circular pupil, the MTF is

$$M_{pl}\left(\nu_i\right) = \frac{2}{\pi}\left[\arccos\left(\frac{\nu}{\nu_{co}}\right) - \frac{\nu}{\nu_{co}}\sqrt{1 - \left(\frac{\nu}{\nu_{co}}\right)^2}\right], \tag{36}$$

since the TOA is irrelevant.

We shall make extensive use of the MTF in subsequent discussions of detection of targets.

25.5 Transmission Through Obscurants

In the previous chapter, we discussed the Beer-Lambert-Bouguer Law of transmission,

$$T\left(x\right) = e^{-\alpha x}, \tag{37}$$

where: α was the extinction coefficient per unit length, and x was the target-observer (object-image) range. This form is possible for a couple of reasons. First, the composition of the atmosphere is roughly uniform over even reasonable distances so that the scattering and absorption properties of the atmospheric gasses are fixed except for pressure variations. These are easily corrected by local measurement of α. Additionally, the atmosphere is fairly well mixed so that variations in water vapor concentration (humidity) can be included as bulk corrections. This even applies to meteorological phenomena such as rain, snow, fog, and haze so long as they are spatially uniform over the path.

This is not the case however for situations where optical media are present but are not spatially uniform. This situation can occur naturally with fog (e.g.,) which is either localized in low lying regions of the terrain, or when the fog sits atop an inversion layer of the atmosphere so that paths below the inversion layer are essentially fog free (and clear) but those crossing the inversion layer pass through some amount of fog. The more common problem of military (and commercial/civil) interest is due to the presence of spatially non-uniform clouds of obscurants. We shall defer discussion of the physics of these clouds until a later chapter since they present special problems in the area of attrition modeling, but it is appropriate for now that we consider the mechanics of their effect on transmission.

In general, these clouds are described by a concentration function which we shall designate by $C(x,y,z,t)$, since they are not only spatially non-uniform but are time varying as well. Under the assumption that the material composition of these clouds is optically uniform, that is, a unit mass of cloud material has the same optical properties anywhere in the cloud, then we may define two quantities: the concentration-pathlength, as the path integral,

$$C_L \equiv \int_{observer}^{target} C(x,y,z,t)\, dl, \tag{38}$$

where l is the path from observer to target (or *visa versa;)* and the mass attenuation coefficient, κ, with units of area per mass. The concentration-pathlength is (obviously) a complicated integral and we shall defer detailed consideration of it for a later chapter. It represents the total mass per unit area along the target-observer path. The mass attenuation coefficient is essentially an extinction cross section per unit mass for the material in the cloud (which we have assumed to be a constant.)

The product of concentration-pathlength and mass attenuation coefficient define an obscurant optical depth,

$$\tau \equiv \kappa C_L, \tag{39}$$

which allow us to modify the Beer-Lambert-Bouguer Law to include the effects of optical obscurants on transmission as

$$\begin{aligned} T &= e^{-\alpha x - \tau} \\ &= \exp\left(-\alpha x - \kappa C_L\right). \end{aligned} \tag{40}$$

obviously, this transmission is now not only specifically spatially dependent (on target-observer path), but time dependent as well.

25.6 Imaging through the Atmosphere

As we have already indicated in Optics I, the atmosphere has three optical effects: absorption, scattering, and emission. The sum of the first two is usually referred to as extinction. While it

is beyond the level of this work (and of most engineering efforts,) to consider scattering effects in detail, it is worthwhile at this point to present some of the results of these effects. The MTF for extinction has been determined to have the form,

$$\begin{aligned} M_{ext}(\nu, x) &= \exp\left[-\alpha_s x \left(\frac{\nu}{\nu_{co}}\right)^2\right] \exp\left[\exp\left(-\alpha_s x \left(1 - \frac{\nu}{\nu_{co}}\right)^2\right) - \alpha_a x\right], \nu \leq \nu_{co}, \quad (41) \\ &= \exp\left[-\alpha x\right], \nu > \nu_{co}, \end{aligned}$$

where: α_s, α_a, are the scattering and absorption coefficients, $\alpha \equiv \alpha_s + \alpha_a$.[4] An approximate assymptotic form of this MTF as a gaussian is

$$\begin{aligned} M_{ext}(\nu, x) &\simeq \exp\left[-\alpha_a x - \alpha_s x \left(\frac{\nu}{\nu_{co}}\right)^2\right], \nu \leq \nu_{co}, \quad (42) \\ &\simeq \exp\left[-\alpha x\right], \nu > \nu_{co}. \end{aligned}$$

It must be noted that this MTF is somewhat limited since it considers only first order scattering. It is thus limited to situations where $\alpha_s x < 1$ (approximately,) - thin obscuration. Of course, this will be more useful at longer wavelengths where exyinction is essentially absorption.

From an imaging standpoint, the primary considered effect of scattering is blurring of the image. Simply put, this is merely consideration that light which originates from somewhere in the environment (but not necessarily in the FOV,) enters the FOV and spatially appears to arise from the target (or background.) The effect of this extraneous light on the image is called *blur*. The MTF for blur is usually considered to be gaussian,

$$M_{blur}(\nu) = e^{-b\nu^2}, \quad (43)$$

where b is a coefficient reflecting the degree of blurring.[5]

25.7 References

[1] Korn, Granino A., and Theresa M. Korn, **Mathematical Handbook for Scientists and Engineers**, McGraw-Hill Book Company, New York, 1968.

[2] Saleh, Bahaa E. A., and Melvin Carl Tech, **Fundamentals of Photonics**, John Wiley & Sons, INC., New York, 1991.

[3] Wetherell, William B., "The Calculation of Image Quality", Chapter 6 in Robert R. Shannon, and James C. Wyant, eds. **Applied Optics and Optical Engineering**, Volume VIII, Academic Press, New York, 1980.

[4] Sadot, D., N. S. Kopeika, and S. R. Rotman, "Incorporation of Atmospheric Blurring Effects in Target Acquisition Modeling of Thermal Images", Infrared Physics & Technology 36 No. 2, pp. 551-564, 2 Feb 1995.

[5] Ratches, James A., et al., "Night Vision Laboratory Static Performance Model for Thermal Viewing Systems", U.S. Army Electronics Command, Night Vision Laboratory, Ft. Belvoir, VA, ECOM - 7043, April 1975.

Chapter 26

Sensors

26.1 Introduction

In this chapter, we review the various types of sensors commonly in use on the battlefield. By sensors, we mean any means used to perceive the external environment, either directly or indirectly. Most of these sensors make use of electromagnetic radiation, that is, light.. In general, the taxonomy of sensors is dominated by that predilection. Sensors are usually described as imaging or non-imaging, active or passive, and by their spectral region of operation. That is, sensors are distinguished on the basis of whether they contain "optical" elements for forming an image of what is in their Field of View (FOV), or merely concentrating that information into a (focal) point; by whether the sensor system provides its own illumination or whether it takes advantage of illumination provided from some exterior (usually natural) source; and finally, if they are light sensors, the range of wavelengths of light for which they are sensitive and responsive.

We use sensor as a very general term, as will be made evident below. In particular, we include devices which are part of systems that may operate on a fully or partially autonomous basis, such as satellites, missiles, or robots, or continuously or intermittently directed, such as (some) surveillance or target acquisition sensors. Thus, while the scope of consideration is broad, since our concern is how these sensors are part of the processes of combat, there is a fairly continuous path from the basic physics of their operation to their functional operation on the battlefield. This simplifies their consideration.

26.2 The Human Sensors

Our starting place for considering the catalog of sensors is those with the greatest basis of issue. Nominally, all humans are provided with an integrated suite of five sensor subsystems:

- Touch,
- Taste
- Smell
- Hearing,
- Vision.

Because warfare is a fundamental of human nature and behavior,[1] these sensor subsystems are integral to not only the direct gathering of personal and organizational information on the battlefield, but the direct or indirect operation of almost all other sensor systems as well.

26.2.1 Touch, Taste, and Smell

These are the minor senses from a battlefield sense, not because they are unimportant, but because they are not directly relevant to the combat processes affecting most soldiers on the battlefield. In general, these sensors, their operation, and their impact are seldom modeled except at the engineering level of man-machine interface and maintenance task design.

26.2.1.1 Touch

Touch is the perception of pressure, heat, and electric shock (merging with intensity into pain,) on the sensors spread over the surface of the human body. This pressure may be generated either internally (by the human musculature,) or externally (by wind or explosive pressure.) The heat is sensed relative to the temperature environment of the sensor. Thus both pressure and heat are sensed in a non-absolute manner by humans as contrasted to mechanical or electronic pressure or temperature instruments (such as wind gauges or thermometers.)

Touch is important in a military sense for most soldiers in the sense of performing normal (not uniquely military) human activities. Obviously, it is important in the maintenance of physical well being (which is necessary for maintenance of individual and organizational performance,) by avoidance of burns, cuts, abrasions, and similar damage. It is also the basis by which all manual interactions are performed. Lacking touch, the operation of controls, the donning of uniforms, and all general activity is difficult and fraught with peril.

From a uniquely military standpoint, touch is primarily important for those soldiers who must operate in direct interface with the natural environment. Touch is an important military sensation for dismounted infantry and Special Operations Forces (SOF) who spend much of their operational time in "the bosom of nature."

26.2.1.2 Taste

Taste is the perception of the chemical and temperature properties of liquids and solid-liquid mixtures. These sensors are concentrated in the human mouth and the perceptions measured by

these sensors are so relative as to be completely individual although extremes of taste (sweet-sour, e.g.,) can be characterized in a taxonomic sense. In general, there are no electronic or mechanical sensors that replicate taste, so this is a unique human sensor system.

The primary impact of taste in a military sense is morale although it may be important in a somewhat *ad hoc* sense for some maintenance activities. From a morale standpoint, having food and drink which has a "good" taste is important militarily. This presents considerable challenge from several standpoints, including Research and Development, Manufacturing, Logistical, and Operational. No matter how "good" packaged food may be on an occasional basis, and the requirement of long term storability and transportability often degrades "goodness", lack of variety soon cloys and tires. (As an irreverent aside, this is why soldiers and students eating cafeteria food continuously rate the food to be terrible while their visitors (e.g., parents and inspectors,) rate the food to be good.)

As with touch, taste can be an important military sense for dismounted infantry and SOF, primarily from the standpoint of surviving and fighting in the natural environment.

26.2.1.3 Smell

Smell is the perception of the chemical properties of gases. These sensors are believed to be concentrated in the human nasal cavity although there is no commonly agreed theory of smell. These sensations are also individual, although, as with taste, there is a taxonomy of extremes. Also as with taste, the military impact of smell on most soldiers is morale in basis. Efficiency, effectiveness, and satisfaction depend on providing a "good" smelling environment and food. This presents the same type of difficulties as taste considerations plus problems with personal hygiene. Bathing is at best difficult in the field for most troops and is a constant burden for commanders and armies given the requirements for water (heavy!), fuel (to heat the water), special hygiene facilities, and support troops (who don't fight.) Thankfully, smell is also a relative sense operating relative to its local environment.

From a direct combat standpoint, smell is important for dismounted infantry and SOF, perhaps more so that taste. Because smell can be highly sensitive to small concentrations of chemicals, and is a trainable sense, it can serve as an important long range sense for dismounted forces that augments their other sensors.

26.2.2 Hearing

Hearing is the perception of sound, that is pressure waves in the atmosphere. Sound waves are significantly different from electromagnetic (EM) waves in that they are scalar, longitudinal waves while EM waves are vector and transverse. The hardware of hearing consists of:

- an external sound collector, called the *Pinna*, which for young men, and thus the traditional majority of soldiers. has a mean length of 65 mm with a range of 52 – 79 mm; and leads into

- the *Auditory Meatus*, a wave pipe of diameter 0.7 cm and length 2.7 cm; which ends with
- the *Tympanic Membrane* which converts the acoustic pressure wave into mechanical motion. The Tympanic Membrane has a roughly circular area of $0.5 - 0.9$ cm^2, a thickness of about 0.1 mm, and a typical displacement amplitude of 10^{-2} cm (or 0.1 of its thickness). The Tympanic Membrane couples directly into
- the *Middle Ear*, which consists of three bones: the Malleus (Hammer); the Incus (Anvil); and the Stapes (Stirrup); of weights 23, 27, and 215 mg, respectively. which connects directly to
- the *Inner Ear*, through the "oval window" to the Cochlea which has a length of about 3 cm and a width of $0.1 - 0.5$ mm. The cochlea contains some 10^4 hair cells which connect directly to nerve fibers which themselves form the end of
- the *Auditory Nerve* which leads to the brain.

The minimum audible (i.e., threshold) sound pressures of pure tones are shown in Figure 1.[2] These data are representative of young persons, ages $18 - 25$ years, with normal hearing. Measurements were made in laboratory and are given in decibels relative to a pressure of $2x10^{-4}$ dyne cm^{-2}. Maximum thresholds are generally not given, but pain thresholds of $125 - 140$ dB are reported, roughly independent of frequency. (Pain thresholds increase slightly with exposure which explains why rock fans and artillerymen have higher pain thresholds and are harder of hearing.) Loudness is usually measured in reference to a pure tone of frequency 10^3 Hz (1 Hz = 1 cycle-per-second) in units called phons. In this terminology, the pain threshold is about 130 phons.[3]

The frequency range of the normal human is approximately $20 - 2x10^4$ Hz with maximum sensitivity between $2x10^3$ and $4x10^3$ Hz. Defining a Field of "View" for hearing is difficulty given the scattering properties of sound waves, but based on the geometry of the ear, it is at least 150° in azimuth (approximately 90° forward from the horizontal normal to the side of the head to about 30° shy of back), and 120° in elevation (about 60° above and below horizontal).

Hearing has great military significance since speech is one of the two primary modes of communication on the battlefield. Additionally, auditory signals are in common use in the control of weapon systems, ranging widely from the lock-on tone of the STINGER missile through auditory pilot's assistants. Hearing is also a vital diagnostic tool in some forms of maintenance and is a necessary surveillance sense for dismounted forces since it has a limited directionality. This directionality may be degraded in environments where scattering of the acoustic waves is marked, such as close terrain and in fog. The latter may be more due to the breakdown in the role of hearing in cueing visual pointing and search than in actual scattering.

Avoidance of hearing damage is of concern, primarily in the environments of artillerymen and mechanized forces. Blast damage to the actual instrumentality of the ear can also degrade

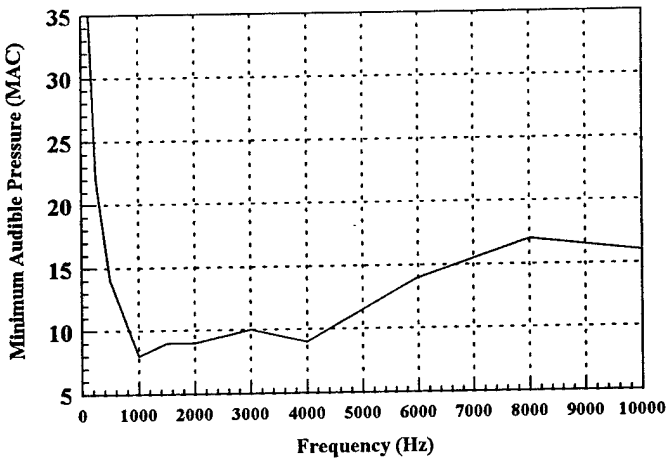


Figure 1: Minimum Audible Sound Pressures

combat performance. For this reason, considerable concern is given to the increased loudness of weapons use in confined spaces, such as rooms. In general, there are two approaches to this: shortening the duration of the sound below about 10^{-4} sec so that the sound is not effectively couplable into the ear; or lengthening the duration of the sound (and reducing the power levels) beyond about 0.5 sec .Both of these approaches present significant challenges to the weapons engineer.

26.2.3 Vision

Vision is the perception of electromagnetic waves. The instrumentality of vision are the eyes which are shown in Figure 2. Light enters the eye, passing through the lens and focusing an image onto the retina. The retina has several layers, containing detector cells named for their shapes rods and cones. The rods and cones react to specific wavelength bands of light and trigger nerve impulses through reaction of light-sensitive chemicals. The optic nerve carries these impulses to the visual center in the brain, where they are interpreted

There are about $75 - 150$ million rods and about 7 million cones in the human retina. Rods perceive only light and dark tones. The rods distinguish outlines or silhouettes of objects even

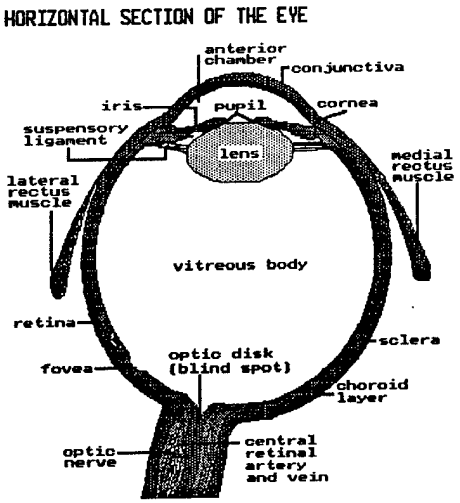


Figure 2: Structure of the Eye

in almost complete darkness. Cones detect the fine lines and points of an image. There are three types of cones absorbing light in blue-violet, green, and yellow-red wavelength bands This gives rise to three types of vision: photopic (cone), mesopic (mixed) and scotopic (rods). Photopic vision is characteristic of adapted vision under luminances of 3 candela-m^{-2} or greater. (A candela is a fundamental unit equal to $\frac{1}{60}$ of the intensity of a square centimeter of a black body at 2045° K.) Scotopic vision is also a light adapted state, corresponding to luminances of $3x10^{-5}$ candela-m^{-2} or less. Mesopic vision is a transitional vision corresponding to luminances between the lower bound of scotopic vision and the upper bound of photopic vision. The approximate spectral sensitivities of scotopic and photopic vision are shown in Figure 3.

This figure also depicts an approximate sensitivity curve for the "standard observer". This curve is an unnormalized gaussian of the form

$$S_{eye}(\lambda) \simeq e^{-\frac{(\lambda-\overline{\lambda})^2}{2\sigma^2}}, \tag{1}$$

where: $\overline{\lambda} \simeq 0.555\mu$, and $\sigma \simeq 4.28x10^{-2}\mu$.

The ability of the eye to image an object depends on the size of the object, its luminance, and

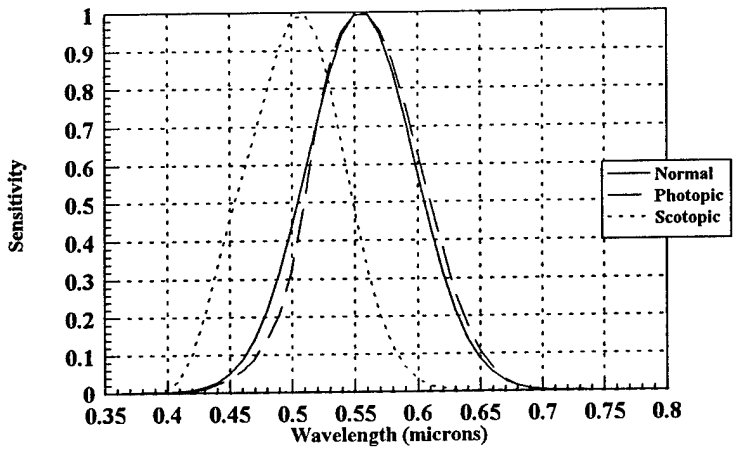


Figure 3: Spectral Sensitivity of the Eye

its background luminance. A detailed exposition of this is beyond our scope, but the interested reader is referred to the work of Blackwell.[4] An approximate curve fit of Threshold Contrast, useful for small objects, but overpredicting threshold contrast for low background luminances (scotopic vision), is

$$C_T\left(r, d_{\min}\right) = e^{a\left(\frac{r}{d_{\min}}\right)^b - c}, \tag{2}$$

where:

r = observer-object range (km),

$d_{\min}$ =minimum perceived object dimension,

$a \simeq 1.93291$,

$b \simeq 0.54186$, and

$c \simeq 5.32681$.[5] A representative plot of Threshold Contrast for Warsaw Pact tanks in fully exposed ($d_{\min} = 2.67$ m) and hull defilade ($d_{\min} = 1.0$ m) situations is shown in Figure 4.

The eye has three Fields of View (FOV): a moderate quality FOV of about 40° in azimuth, and 30° in elevation; a high quality circular FOV (the foveal field) of about 9° in both azimuth

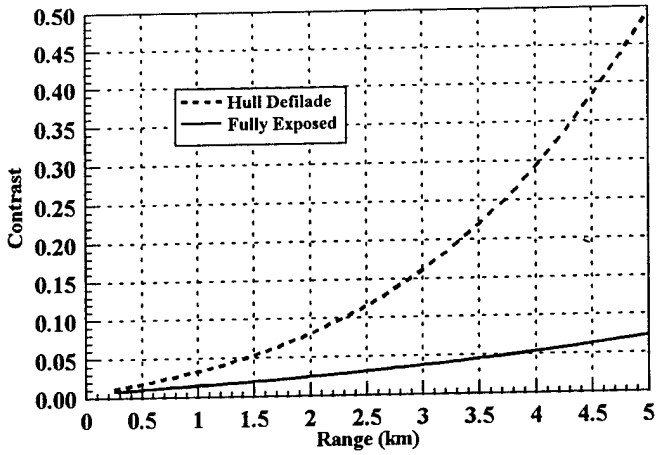


Figure 4: Threshold Contrast

and elevation (center); and a best vision circular FOV of about 3° in both azimuth and elevation. The high quality FOV is used for most visual tasks and is the reason why most sensor displays have 4 : 3 aspect ratios. Color vision covers about 90° of the central field with degradation on the retina edge.[6] A circular FOV of 5° (circular) is commonly used in search engineering calculations with a glimpse time of $0.2-0.3$ sec.[7] The duration of glimpse time is an increasing function of decreasing luminance.

A variety of Modulation Transfer Functions (MTF) for the eye are reported in the literature. Kornfield and Lawson [8] report three of note: an exponential MTF arising from a linear relation between the logarithm of contrast sensitivity and spatial frequency shown by Van Meeteren and Vos,

$$M_{eye}\left(\omega_x\right) \simeq e^{-1.2\omega_x}, \tag{3}$$

a physiological MTF of their own,

$$M_{eye}\left(\omega_x, \omega_y\right) \simeq e^{-\Gamma\left(|\omega_x|+|\omega_y|\right)} - 0.64e^{-\frac{\sigma^2\left(\omega_x^2+\omega_y^2\right)}{2}}, \tag{4}$$

and a phenomenological MTF,

$$M_{eye}(\omega_x, \omega_y) \simeq e^{-\Gamma(|\omega_x|+|\omega_y|)} - M_I(\omega_x) M_I(\omega_y), \tag{5}$$

where:

$$\begin{aligned} M_I(\omega_x) &= e^{-S\sigma\omega_x - \frac{T^2\sigma^2\omega_x^2}{2}}, \omega_x \leq \frac{R}{\sigma}, \\ &= 0.8e^{-\frac{\sigma^2\omega_x^2}{2}}, \omega_x \geq \frac{R}{\sigma}. \end{aligned} \tag{6}$$

The parameters for the above equations are: $\Gamma = 1.2'$, $\sigma = 8'$, $R = 1.0$, $S = 0.262$, and $T = 0.96$. Ratches [15] gives an eye MTF of the form

$$M_{eye}(\omega) = e^{-\frac{\Gamma\omega}{M}}, \tag{7}$$

where: Γ is a light-level dependent parameter and M is the system magnification. This form clearly is simular to that of Van Meeteren and Vos.

We turn now to sensors which are not integral to the human physiognomy.

26.3 Acoustic Sensors

There are considerable similarities between the physics of acoustic (sound) and electromagnetic radiation. The chief differences lie in two main regards: the wave form of the radiation; and the propagation speeds of the radiation. As we have already noted, electromagnetic waves are essentially vector waves of transverse mode - that is, they are described by three magnitudes (direction of propagation and two amplitudes,) and the vibration occurs perpendicular to the direction of propagation. Acoustic waves are essentially scalar waves of longitudinal mode - that is, they are described by a direction of propagation but vibrate along that direction. Other than this, the propagation equations are markedly similar.

The other primary difference is in propagation speed. As we have already noted, the speed of light in vacuum is $\sim 3x10^8 m\,\mathrm{sec}^{-1}$, and the refractive index in air is ~ 1.000292, which must be divided into the vacuum speed to get the air speed (and is thus about 0.03% smaller to first order.) The speed of sound in vacuum, of course, does not exist, since unlike light, sound must have a medium to propagate in. The speed of sound in air is Temperature dependent (alternately pressure dependent,) and has the approximate relationship

$$v_{air} \simeq 331\sqrt{\frac{T_K}{273^\circ}}, \tag{8}$$

where the units are $m\,\mathrm{sec}^{-1}$. Other useful (approximate) sound speeds are given below in the table.[9]

Material	Speed
Water ($20°C$)	1478
Seawater ($20°C$)	1520
Brick	3640
Steel	5500
Wood	4000

As may be seen, the magnitudes of these speeds are $10^{-5} - 10^{-6}$ less than light speeds. This has a decided effect on wavelength which has an lower limit of the mean free path of (e.g.,) air of $10^{-7}m$ as opposed to a minimum wavelength for light of the granularity of free space.[3]

The fundamental acoustic sensor is the microphone. Basically, a microphone is nothing more than a membrane connected to an electric generator. The incident acoustic wave couples with the membrane whose resulting vibrations generate a varying electrical current. This current represents a signal that can be amplified, transmitted, stored, or reproduced with speakers which are essentially microphones in reverse - that is, they are electrical motors which generate vibrations in a membrane which couples with the atmosphere to produce sound. Microphones do not have FOV (since they are not optical sensors,) but they do have an equivalent operational field represented by some solid angle. Similarly, microphones have a sensitivity which represents how well the microphone converts sound of that wavelength (frequency) to electric current Acoustic waves incident upon the microphone within this solid angle and of frequency within the sensitivity band generate corresponding electric signals.

The most common use of microphones in combat are, of course, telephones and radios. Acoustic direction finders have been used for artillery location (by triangulation,) since World War I. These systems have also been used for tracking airplanes with some severe limitations, notably by the air defense forces of the former Soviet Union. The geometry of such a system is shown in Figure 5.

The Acoustic Ranging Sensor consists of an array of four microphones (indicated by circles in the figure.) (We have shown the figure as two-dimensional. Actually we will be working a three-dimensional problem.) If the origin of the coordinate system, the four microphones are displaced by distances $\Delta x, \Delta y$, from the center of the sensor, then they have positions $\underrightarrow{r_i}$, $i = 1..4$, all with identical magnitude. If an artillery piece located at x, y, z ($\underrightarrow{r}$) fires, then the microphones detect the blast at times $t_1..t_4$. If the speed of sound is v_s, then the geometry dictates that

$$t_i - t_f \equiv \Delta t_i = \frac{\left|\underrightarrow{r} - \underrightarrow{r_i}\right|}{v_s}, \tag{9}$$

for each of the four microphones. The actual time of firing, t_f , is, of course, unknown, but because there are four equations, they may be reduced to three equations by subtraction. This eliminates dependence on t_f, and leaves equations of the form,

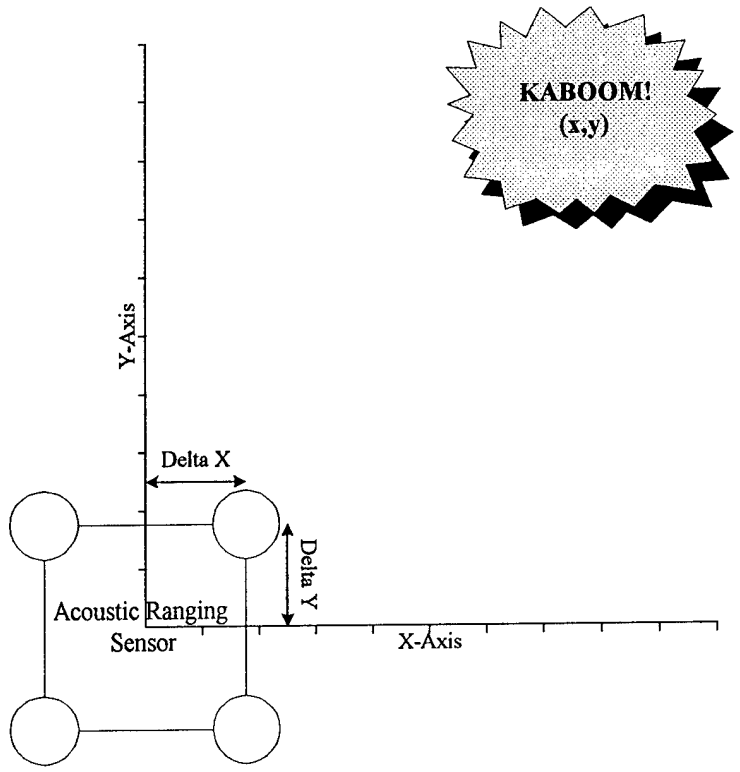


Figure 5: Acoustic Ranging Sensor Geometry

$$\Delta t_{i,j} \equiv \Delta t_i - \Delta t_j, j \neq i. \tag{10}$$

The modulus of the difference in the blast position and microphone position vectors has the form,

$$\begin{aligned} \left| \underrightarrow{r} - \underrightarrow{r_i} \right| &= \sqrt{\underrightarrow{r} \bullet \underrightarrow{r} - 2\,\underrightarrow{r} \bullet \underrightarrow{r_i} + \underrightarrow{r_i} \bullet \underrightarrow{r_i}} \\ &= \sqrt{r^2 - 2\,\underrightarrow{r} \bullet \underrightarrow{r_i} + r_i^2}, \end{aligned} \tag{11}$$

where: $r \equiv \left| \underrightarrow{r} \right|$, and $r_i \equiv \left| \underrightarrow{r_i} \right|$. This may be rewritten as

$$\left| \underrightarrow{r} - \underrightarrow{r_i} \right| = \sqrt{r^2}\sqrt{1 - \frac{2\,\underrightarrow{r} \bullet \underrightarrow{r_i}}{r^2} + \frac{r_i^2}{r^2}}. \tag{12}$$

Since $r \gg r_i$ under field conditions, we may safely ignore the r_i^2 term, and rewrite equation 12 as

$$\left|\underrightarrow{r} - \underrightarrow{r_i}\right| \simeq r\sqrt{1 - \frac{2\,\underrightarrow{r} \bullet \underrightarrow{r_i}}{r^2}}, \tag{13}$$

and expand in a one term Maclaurin series as

$$\begin{aligned}\left|\underrightarrow{r} - \underrightarrow{r_i}\right| &\simeq r - \frac{\underrightarrow{r} \bullet \underrightarrow{r_i}}{r} \\ &= r - \hat{r} \bullet \underrightarrow{r_i},\end{aligned} \tag{14}$$

where: $\hat{r}$ is the unit (magnitude) vector with the same direction as $\underrightarrow{r}$.

If we now substitute equation 14 into equation 10, then we obtain equations of the form

$$\Delta t_{i,j} = \hat{r} \bullet \underrightarrow{r_j} - \hat{r} \bullet \underrightarrow{r_i}, i \neq j. \tag{15}$$

Since there are three equations and three unknowns (the components of $\hat{r}$,) a solution is possible and computationally feasible. What does this solution consist of? Simply put, it is the direction cosines of the position of the blast (fire.) It does not give the exact position of the fire since the magnitude of the position vector cannot be calculated. Since the direction cosines can be converted directly into azimuth and elevation angles, this gives us the azimuth and elevation of the firer relative to the sensor.[1]

Great! Now we know the direction to the shooter, but not where he is. What good is this? Well, if these sensors are used in pairs with a fair distance between them, and their positions are well surveyed (i.e., the sensors know where they are, and today with Global Positioning Satellites (GPS), this is pretty easy,) then with a good map, lines may be drawn from the two sensors along the directions measured by each. The shooter is wherever the lines intersect. This is called triangulation and we show it in Figure 6. The figure displays a simple map with two acoustic sensors (indicated by human figures,) and a shooter (indicated by a bull's eye and an outgoing rocket.) The heading (azimuth) directions relative to the two sensors are drawn as arrows. They meet at the shooter.

This brings us to an historical note. (A bit of James Burke-ism here.[10][11] Burke is a noted historian of technology who is best known for his television series of the same names as his books. Reading these (viewing the series) is highly recommended for both soldier and student interested in how technology interacts in history, especially from a military standpoint as we are interested in.) Up to the period ending with the American Civil War (see Part I,) artillery was a direct fire combat arm. With the development of the conoidal bullet, it was forced to become an

[1]The astute student will note that we haven't talked about about error. In general, we won't except as error naturally occurs in a stochastic manner for the derivation presented. There are threeo fundamental reasons for this. First, our primary purpose in most of our discussion is insight into the physics. Error discussions tend to mask this. Second, most error sources arise from mechanical implementation, which is more properly engineering. Third, and lastly, part of our purpose is connectivity to other sources. Thus, we assume that the diligent student can trace down the error factors.

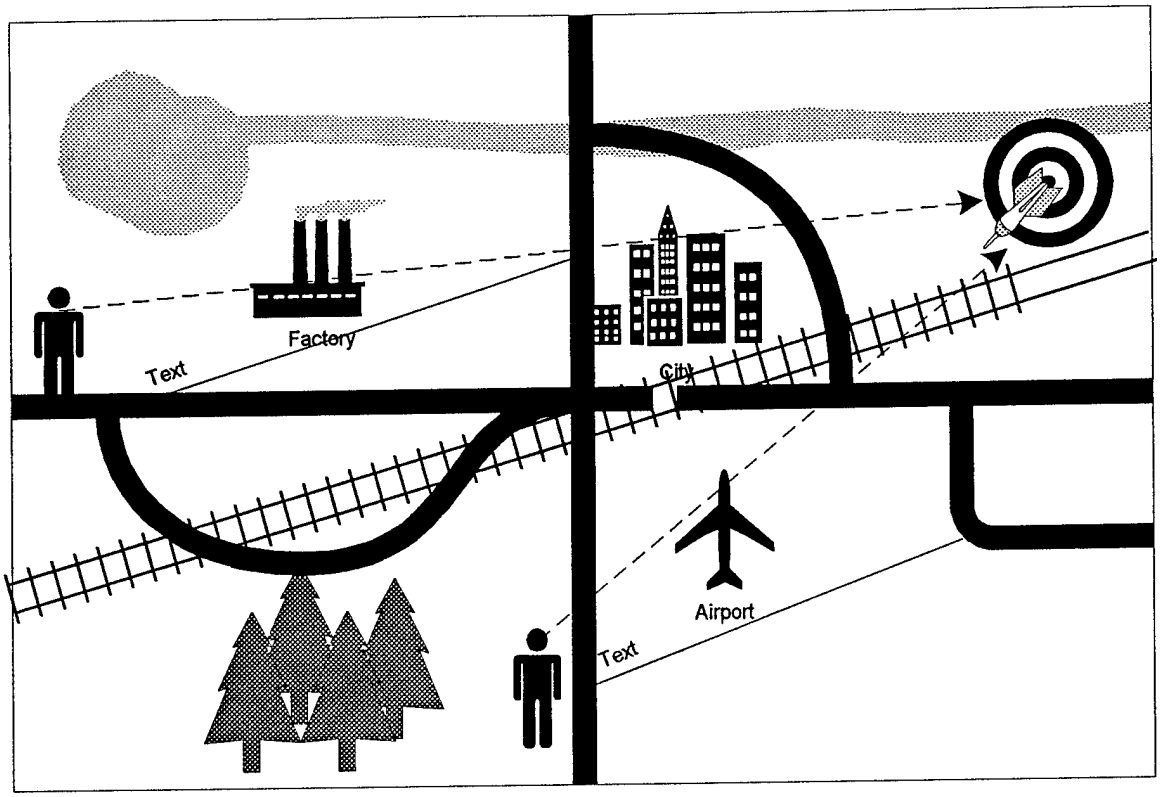


Figure 6: Triangulation Example

indirect fire arm (until the advent of the tank.) This made location difficult. (That was the idea, get it out of sight of infantry fire! and thereby, survivable.) With the advent of effective sound sensors (the telephone and Ma Bell,) and basic electronic amplification (radio and Marconi,) sound detection sensors became possible. The problem lay with the calculation of the azimuths. Equation 15 is really three simultaneous equations. These may be solved iteratively (by guessing an initial solution and repeating successive component calculations until they converge,) or by inverting a $3x3$ matrix. Both of these are laborious and time consuming under World War I field conditions where computational tools were essentially slide rules and paper. Luckily for the sensors, mobility was very poor and they could grind out the calculations before the shooter could move. Unluckily for the shooter, he couldn't move out of the way very fast.

The technical answer to this (for the shooter) was twofold. First, the gun engineers worked to make guns that were longer range (and thus harder to detect since the amplitude of the sound diminished as r^{-2},) and the vehicle engineers (Ford and Christy) matured their technology,

placed the guns on internal combustion powered, tracked (even wheeled) vehicles and gave them greater mobility. The sound engineers weren't asleep however. Courtesy of the communications-electronics community, microphones and amplifiers got better, and by World War II, specialized electronic-mechanical calculators for performing the calculations were available. Sadly, the physical limitations on sound propagation proved to be insurmountable and today, this function is performed by radars (described below.)

Sound direction sensors have not completely disappeared however. As we mentioned earlier, they are still in use for detecting low flying aircraft (below the radar horizon,) albeit chancily. There is promise of these systems enjoying a minor reinvigoration as warning sensors for armored vehicles and (particularly,) teleoperated unmanned vehicles (robots.) Basically, in this application, these sensors serve to provide the vehicle with a sense of hearing so that it can point its visual sensors in the direction where a shot came from. In this case, pairs of sensor stations are not needed (since direction is sufficient,) but they are of value only if a shot was a miss (or didn't kill the vehicle.) Modern electronic filtering technology allows analysis of the frequency spectrum of the shot so that the nature of the shooter can be determined as well.

Another contemporary application of sound sensing is primarily used for surveillance and spying. This application is an active system[2] that marries a laser with a sound detector (microphone.) The laser beam is projected on a vibrating surface (such as a window,) and the reflected signal is detected. Sound produced on the window (e.g., by conversation in a room,) causes the window to vibrate and this vibration frequency modulates the laser return through the mechanism of the Doppler effect. Because of the difficulty of operation and the low data rates, this type of sensor is seldom used tactically except for intelligence gathering at sites of perceived value, usually covertly. (Paranoia Alert! Put thick curtains in front of all windows to decouple the sound from the window.)

The last acoustic sensor that we shall consider here is the best known, at least by anyone who has been exposed to any of the Hollywood cinematic productions dealing with submarines. This is the Sound Navigation And Ranging or SONAR. SONAR is exclusively a naval sensor, used both by submarine and counter-submarine forces. It may be used in either active or passive modes. The passive mode relies on inherent sound emissions from the environment and targets, while the active mode projects a sound beam in a manner analogous to radar. The instrumentality of SONAR has great similarity to radar and we shall not dwell on it here. The literature reports that most active SONAR emissions are around $25kHz$.[3]

The medium of propagation of SONAR is considerably less well behaved that the atmospheric-electromagnetic environment of radar. The sea surface boundary, underwater currents, and thermal variations, as well as the presence of sound producing indigenes (fish), can make the interpretation of SONAR signals difficult. Thus, SONAR operation has a much greater human interpretation component than radar.

SONAR has significant military application for both navigation (especially important on a blind, submerged submarine,) and target acquisition. It is in the latter application that tactical

[2]This is the first active system we have mentioned. See the discussion in the radar section.

acumen becomes important. The passive mode is chancier, relying on sound generated by targets, and thereby often of shorter effective operational range. The active mode, while being deliberate and controllable, and generally of longer range, has the disadvantage (possibly fatal,) of advertising the presence of the searcher.

In general, acoustic sensors are not imaging largely because such images are of low resolution and of limited value. Pseudo-imaging in the form of direction sensing is the practical limit.

26.4 Radar

Outside of the human senses, radar (Radio Detection And Ranging)[3] is probably the best known of the military (and civilian) sensors. Radar has become an ubiquitous part of our daily lives, found in such mundane activities as viewing the weather forecast, or taking a trip in an automobile or an airplane. Alone of sensors, radar has become almost mythic in military history, the Chain Home stations on the eastern coast of England being viewed as crucial in winning the Battle of Britain and thereby victory in World War II. These early radars still constitute a reasonable, albeit simple, model of radar.

Transmit a burst of electromagnetic wave of radio frequency and listen for a return (obviously with a detector). The direction the return comes from indicates the reverse direction to a target (aircraft.) Note the time difference between the transmission and the reception, divide this time by two and multiply by the speed of light and you have the range to the target. That's how a "radar" works. A simple diagram of a radar is shown in Figure 7.

The radar signal is generated and sent to the transmitter which passes it to the antenna. The antenna couples the signal to the environment (essentially free space,) generating the actual electromagnetic wave. This wave travels to the target and some of it is scattered back to the antenna where it is sent to the receiver. The received signal is sent on to a processor which "images" the returns on a display for the operator.

Unfortunately, this simple model is a bit misleading. There are many types of radars and we cannot begin to do much more here than catalog some of the basic types. Before doing this, we have to specify some taxonomy. First, there are the different ways that systems can be configured to use the signal energy. An active system is one which produces its own wave energy (in this case, electromagnetic) as well as detecting it. A passive system uses some other wave energy source and only detects modifications of that energy from the target. Normally, active systems collocate energy source and detector. A special case is bistatic systems which have source and detector at dislocated positions.

Next, we have imaging and non-imaging systems. Imaging systems, as their name indicates, form images of what they view. Non-imaging systems do not form images although they usually

[3]This is another exception in the acronym jungle. Technically, we should always spell this as RADAR since, after all, it is an acronym. The acronym has become such a part of our contemporary vocabulary that it has taken on the life of a proper name. Thus, we shall not capitalize this acronym.

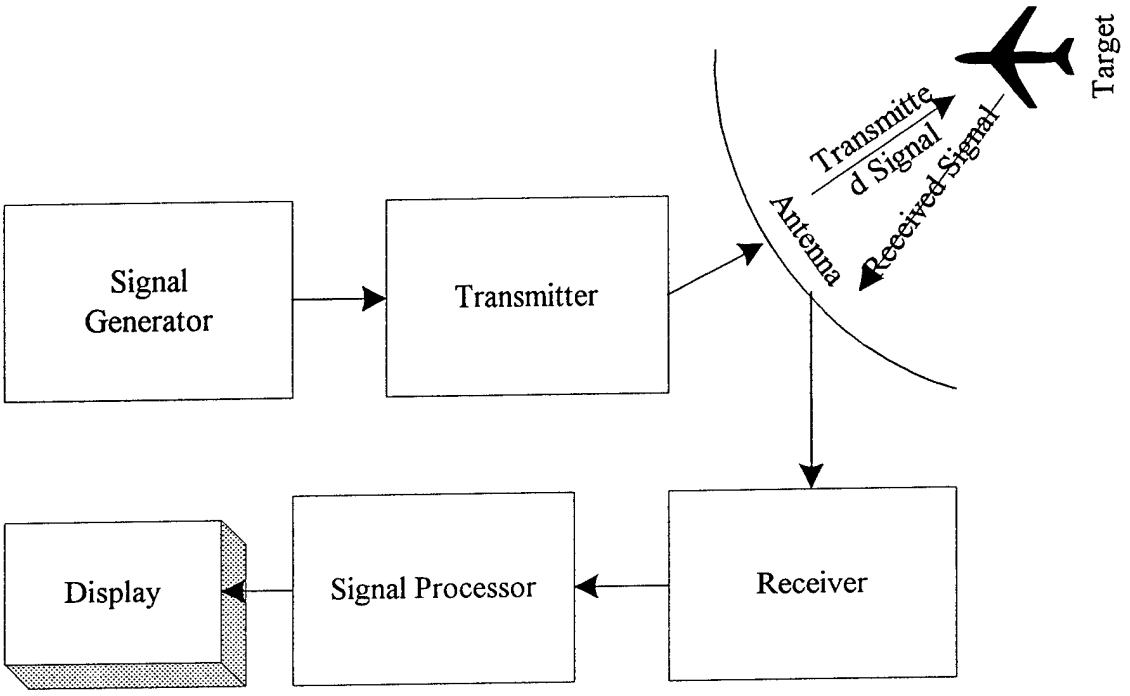


Figure 7: Diagram of a Simple Radar

do measure direction and possibly range to objects in view, In general, imaging systems usually operate at short wavelengths (visual and infrared) for a variety of reasons, but notably because images have resolution comparable to a few wavelengths of the light used. This means that images of interesting physical objects can easily have high resolution in the visual and infrared because these objects are usually vastly larger than a few microns. (This is also why there is a limit to optical microscopy and we must resort to electron microscopy to view objects the size of visual wavelengths and smaller.) At the longer wavelengths of radio, (of order meters,) images of many physical objects would not have much resolution and thereby little information. A secondary, but eminently practical reason is the relatively large size of radar detector elements and thereby the large size of detector arrays (or other instrumentation for imaging) that would be required for image formation. Of course, there are imaging radars (such as Synthetic Aperture Radars - SAR), but these radars are intended to form images of large objects and their occurrence is therefore less profuse than visual imaging sensors (e.g.)

Finally, almost unique to radar sensors, we have the matter of pulse versus continuous sources. For visual and infrared systems, the source of illumination is external. That is, these are primarily passive systems. (There are exceptions.) Radar systems, on the other hand, tend to generate their own illumination since the natural environment does not do so with great inten-

sity. This illumination may be generated continuously or intermittently. Continuous illumination can provide direction information but not range. (Like the acoustic direction sensor described above.) Pulse or burst illumination has a sharp (temporally and spatially) leading edge wave and therefore allows the time difference between transmission and reception to be measured. As indicated above, this permits range to be inferred. (There may also be an engineering problem here. It may be more difficult to build a pulse transmitter of a given energy, than an equal energy continuous transmitter because of switching difficulties. Thus pulse radars may be of lower energy and thereby, as we shall see, of shorter range.)

Another characteristic of radar is the tracking of targets. Usually, they are designed to track targets of considerable speed in contrast to optical sensors which are usually ill suited to such targets. Of course, the issue here is not linear velocity, but angular velocity, which is range dependent. Thus most radar sensors have large FOV/FOR while optical sensors usually have small FOV. Disruption of track is of considerable military importance since, in general, targets which cannot be accurately tracked cannot be engaged.

The differences between FOV and FOR were introduced in the Optics II Chapter. Since radars are active (and thereby ignoring bistatic radars,) FOI = FOV, and FOR > FOV, usually. The normal mode for a radar is to scan the FOR either continuously or discretely. In the former case, the radar searches by continuously moving the FOI over the FOR, usually in some pattern. In the latter case, the radar searches the FOR in FOI sized pieces, dwelling some amount of time in each. Again, some pattern is usually used. We will discuss search and detection in a later chapter.

Other differences aside, the fundamental description of radar performance is the radar range equation (corresponding to the Point Spread Function since the radar is usually non-imaging.)[12],[13] This has the form

$$S = P_t G_t \frac{\sigma}{4\pi r^2} \frac{A_r}{4\pi r^2} e^{-2\alpha r}, \tag{16}$$

where:

S = received signal power (back from the target,)

P_t = power (energy per time) generated by the transmitter,

G_t = gain (efficiency factor) of the transmitting antenna,

r = range to the target,

σ = scattering cross section of the target,

A_r = effective area of the receiving antenna, and

α = atmospheric extinction coefficient at the operating wavelength (often negligible except with heavy rain (large drops) or snow.) This equation openly demonstrates the spherical (rather than plane) wave nature of the light considered.

This equation may be interpreted simply. The transmitter power, P_t, is just the energy per time generated by the transmitter. The gain of the transmitting antenna is a multiplier for the efficiency that this power is actually turned into an electromagnetic wave. The quantity $\frac{P_t G_t}{4\pi r^2}$ is

then the energy per time per unit area that reaches the target. The quantity $\frac{P_t G_t \sigma}{4\pi r^2}$ is the energy per time that is reflected by the target since σ, the scattering cross section, may be seen as the fraction of incident energy that is reflected (scattered) back to the radar. The quantity $\frac{P_t G_t}{4\pi r^2}\frac{\sigma}{4\pi r^2}$ is the energy per time per unit area that arrives back at the radar, and the final quantity $\frac{P_t G_t \sigma}{4\pi r^2}\frac{A_r}{4\pi r^2}$ is just the energy per time (power) that is absorbed by the radar. The factor $e^{-2\alpha r}$ represents the extinction of the energy by the atmosphere along its two transits of the distance r.

A variation of this equation appertains for bistatic radars (where the transmitter and receiver are dislocated,)

$$S = P_t G_t \frac{\sigma}{4\pi r_1^2}\frac{A_r}{4\pi r_2^2} e^{-\alpha(r_1+r_2)}, \tag{17}$$

where: r_1 = range from transmitter to target, and r_2 =range from receiver to target.

The performance of the receiver is usually characterized by a noise level, expressed in terms of an equivalent temperature, as

$$N = kT_s, \tag{18}$$

where: T_s = equivalent noise temperature, and k is Boltzmann's constant. This may be combined with equation 16 to yield a signal-to-noise ratio,

$$\frac{S}{N} = \frac{P_t G_t}{kT_s}\frac{\sigma}{4\pi r^2}\frac{A_r}{4\pi r^2} e^{-2\alpha r}. \tag{19}$$

This quantity is central to the radar theory of detection (discussed in a later chapter.) For some minimum value of signal-to-noise ratio (> 1,) a maximum detectable range can be calculated. We note that when atmospheric extinction can be ignored (which is the usually case,) then equation 19 can easily be solved as

$$r = \left[\frac{P_t G_t}{kT_s}\frac{\sigma}{4\pi}\frac{A_r}{4\pi}\frac{1}{\left(\frac{S}{N}\right)_{\min}}\right]^{\frac{1}{4}}. \tag{20}$$

If extinction cannot be ignored, then the resulting range equation is transcendental (derived from equation 19,)

$$re^{\frac{\alpha r}{2}} = \left[\frac{P_t G_t}{kT_s}\frac{\sigma}{4\pi}\frac{A_r}{4\pi}\frac{1}{\left(\frac{S}{N}\right)_{\min}}\right], \tag{21}$$

which must usually be solved either approximately (by expanding the exponential for small α, or $\ln(r)$ for large α,) or iteratively, usually using equation 20 as a starting guess. These equations are somewhat simplistic in form, especially for radars with elaborate waveforms or spread spectrum capability.

While the first radars generated simple spherical waves, modern waves have been developed which generate waves of more complex structure. (This structure is referred to as the waveform.)

The structure is intended to improve the performance of the radar usually by special consideration of the target composition or the specifics of the operating environment. In addition, the wave may be coded with a (usually) random sequence to facilitate signal processing. Additionally, early radars operated over a fixed spectral (frequency) band. While many radars still perform in this manner, some modern radars operate over very broad bands, or more crucially, operate over varying bands. This is also intended to improve performance by sampling over the value of σ which may have extreme frequency variation.[4]

In addition to atmospheric extinction which is usually not of impact with radar sensors except for heavy rain and snow (ice), unlike visual and infrared sensors, two other effects are of particular interest in radar sensors. These are clutter and jammers. Clutter in this case refers to environmental objects in the sensor FOV which produce signals comparable to the target's. For fast moving targets (such as high performance aircraft,) the effects of clutter may be mitigated by looking at the FOV over a period of time and noting only the differences. Since the clutter will be moving, at best, very slowly compared to the real targets, it is effectively removed. Alternately, of course, the return signal exhibits a frequency shift due to Doppler effect. This allows suppression of returns with small (zero) Doppler shift and thereby elimination of clutter. (Since aircraft flying perpendicular to the wave normal have low Doppler shift, such crossing flight is a frequently used tactical maneuver to break track.)

A special countermeasure of note is chaff, or as originally named by the British, window. Chaff is physical material, usually half or whole wave dipole antennas, dropped by the aircraft. These chaff provide artificial clutter intended to confuse the radar track. Since these are not usually powered, they cannot maintain velocity comparability with the aircraft for a long time due to drag, and soon have velocities dominated by gravitation force. They thus have an easily discernible Doppler signature and are excludable in a manner akin to clutter. Accordingly, the aircraft must continually disseminate chaff for it to be effective as a countermeasure. Waveform structuring may be effective in reducing the effects of clutter and chaff by being shape sensitive.

A special case of note is that of jamming. This involves transmitting a wave of band comparable to the signal returned by the target. The effect of this is to increase the background that the target must be "seen" against. From the standpoint of the radar range equation, equation 19, this has the effect of increasing the effective noise of the system, comparable to our consideration of contrast in the preceding chapter. There are basically two types of jamming: self-screening and escort jamming; and stand-off jamming.

In self-screening jamming, the aircraft carries its own jammer. This has the advantage of providing protection against detection, but since the jammer has weight and volume, it reduces the payload and/or performance of the aircraft. This method is most useful for single aircraft raids or missions. In escort jamming, an aircraft dedicated to jamming accompanies the raiding aircraft. This has the advantage of allowing the raiding aircraft carry their full payload of weapons

[4]Remember, we are dealing with regimes where imaging may not be possible because the characteristic dimensions of targets is of order the wavelength. Thus scattering is Mie rather than optical or Rayleigh. Also, radar convention is to use frequency rather than wavelength terminology since it is non-imaging.[14]

(or sensors,) but has the disadvantage of dedicating an aircraft to jamming. This method is most useful on multi-aircraft raids. Another disadvantage of both these methods is that they increase the risk to the aircraft if their enemy has guided weapons (e.g., missiles) with the capability to home on the jammer.

In stand-off jamming, one or more jamming aircraft remain at some range from the radar while the raiding aircraft close. This method further reduces the vulnerability of the jamming aircraft. This method is most effective when large aircraft (e.g., freight or bomber class) carry large jammers. The greater power of these jammers allow coverage of a greater ground area and therefore several raids may be protected at once.

This analogy comparing jamming to extinction deserves illustration. To do this, we first parametrize the radar equation writing it as

$$\frac{S}{N} = \frac{P_t G_t}{N} \frac{\sigma}{4\pi r^2} \frac{A_r}{4\pi r^2}, \tag{22}$$

where we have set $\alpha = 0$ to reflect an ideal extinction environment (and simplify the calculations.) Now, we prescribe that $\frac{S}{N} = 3$ at $r = 10km$. (This selection of signal-to-noise will be seen to be significant in the next chapter.) This prescription reduces equation 22 to a very simple form,

$$\frac{S}{N} = \frac{3x10^4}{r^4}. \tag{23}$$

Now, we define a Jammer power as

$$S_J = \frac{P_J G_J A_r}{4\pi r^2}, \tag{24}$$

which has the form of the radar equation with analogous definitions. If we parametrize the jammer transmitted power as a fraction f of the transmitted power of the radar, equation 24, reduces to

$$S_J = \frac{f P_t G t A_r}{4\pi r^2}. \tag{25}$$

For self-screening and escort jammers, the r used here is the same as the raiding aircraft, while for stand-off jammers, the r is essentially fixed at some value r_J. We consider each case in turn.

For self-screening and escort jamming, there is considerable simplification. We let $N \rightarrow N + S_J$, and combine equations 22 and 25, within the parametrization already described, to produce the result

$$\frac{S}{N} = \frac{3x10^4\sigma}{r^2\left(\sigma r^2 + 1.2x10^5 f\pi\right)}. \tag{26}$$

If we extend the parametrization to define $\sigma \equiv 1m$, then equations 23 and 26 are functions only of range r and jammer power-gain fraction f. We illustrate this in Figure 8. We have included three jammer powers ($P_J G_J$) parametrized as 1%, 3%, and 10% of the radar power.

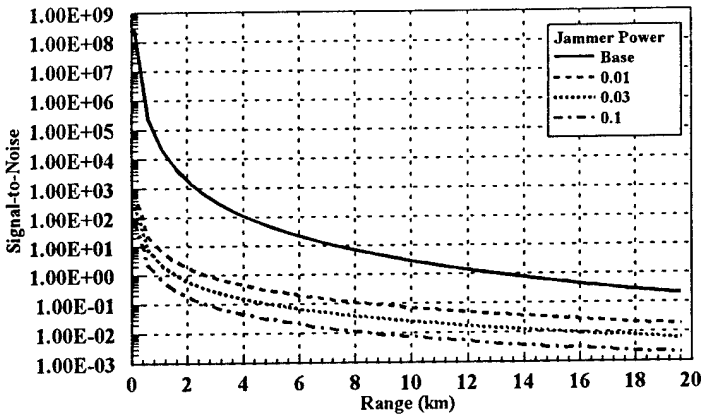


Figure 8: Self-screening and Escourt Jammer Effect Example

These are fairly large jammer powers, but compared to the radar's power are quite nominal. As we may see, however, all of these jammer examples reduce the performance of the radar markedly. Where the radar had a parametrized range of $10km$ without jamming ($\frac{S}{N} = 3$,) with these jamming examples, the equivalent signal-to-noise ranges are reduced to values between 0 and $2km$. Within the context of our extinction analogy, these levels of jamming are equivalent to very dense fog indeed.

We may repeat this procedure for stand-off jamming. In this case, the jammer range is not the same as the raiding aircraft range, but approximately has the constant value r_J. The complement to equation 26 has the form

$$\frac{S}{N} = \frac{3x10^4 \sigma r_J^2}{r^4 \left(\sigma r_J^2 + 1.2x10^5 f\pi\right)}, \tag{27}$$

which we may see is a function only of ranges r and r_J, and jammer power-gain fraction f. If we assign $r_J = 10km$, then we may repeat the calculations presented in figure 8. These are presented in figure 9. Comparison of these two figures illustrates the reduced effectiveness of stand-off jamming (for the same power-gain) as for self-screening or escort jamming.

The operation of radars today is highly automated in terms of signal processing, tracking, and information display, although the latter is still retained to facilitate human operation of the system

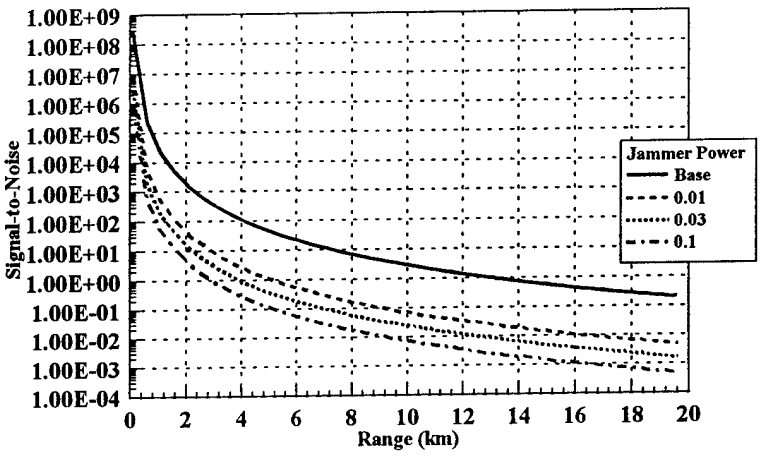


Figure 9: Stanf-0ff Jammer Effect Example

as a whole. Additionally, when the radar is part of a weapon system, typically an air defense weapon system, control of the weapon subsystem may also be highly automated. Tracking is usually performed by some sort of statistical filtering and prediction technique such as a Kalman filter. Discussion of the dynamics of these techniques is beyond the scope of this work and is left to specialized texts.

26.5 Optical Instruments

Since its development in the time of Galileo, the telescope has been an instrument of interest not just for astronomers, but for the military as well. Technological developments and tactical influences have replaced it first with field glasses and then with binoculars. The development in this century of the submarine as a weapon system has also led to the introduction of the periscope which is a combination of a telescope with mirrors (prisms) to permit offset viewing.

The (primary) tactical problem with the telescope has always been its length and fragility. A major improvement was realized with the collapsible telescope that we know today. Interestingly, the collapsible nature of this optical instrument has entered the common English language usage

- telescoping now means collapsible and extensible rather than "far seeing". With the advent of optical glasses with greater refractive index and advances in lens theory to permit manufacture of thicker lenses, the telescope was largely replaced by the field glasses. These have subsequently been replaced by binoculars which incorporate prisms to both invert the image and provide greater optical length. Of course, the length of the telescope is not a problem for the submarine with its periscope since the submarine wishes to observe surface conditions from as great a depth as possible.

The MTF of an optical instrument is essentially the same as that for a simple lens,

$$M_{OPT}(\nu) = \frac{2}{\pi}\left[\arccos(A) - A\sqrt{1-A^2}\right], \tag{28}$$

where:

$A = \frac{\lambda F \nu}{l}$,

λ =wavelength,

F =system F number,

ν =spatial frequency, and

l =system focal length.[15] Note that λ and l must have common units. This MTF is multiplied by a geometric blur MTF,

$$M_{BLUR}(v) = e^{-bv^2}, \tag{29}$$

where b represents the rate of blur degradation, and if the system is not rigidly isolated, by a stabilization MTF which is also Gaussian.

26.6 Electro-Optical Instruments

We now come to the area of indirect view optical systems. These are indirect in the sense that the light that reaches the eye does not arise from the target. Instruments of this type are typified by television. Light arising from the target and background propagate to the system's optics where an image is formed (usually) in the focal plane. A light detector apparatus in the focal plane converts the light of the image into a voltage. This voltage is then processed to generate a structured electrical signal. This electrical signal is then processed and displayed on a CRT or LED panel (LCD, etc.) The "image" on the display screen is then viewed by the eye. A simple diagram of such an instrument is shown in Figure 10.

This scheme has several advantages: first, because the image is reduced to an electrical signal, the optics, detector, and signal generator can be located somewhere different from the signal processor and display. Thus electro-optical instruments can be placed where optical instruments can not. Second, the electrical signal does not have to be viewed by an operator. They may be directly processed and used. This is the basis of autonomous sensors such as missile seekers. Lastly, the light detected does not have to be spectrally compatible with the spectral response of the human eye, or even limited to a single spectral band.

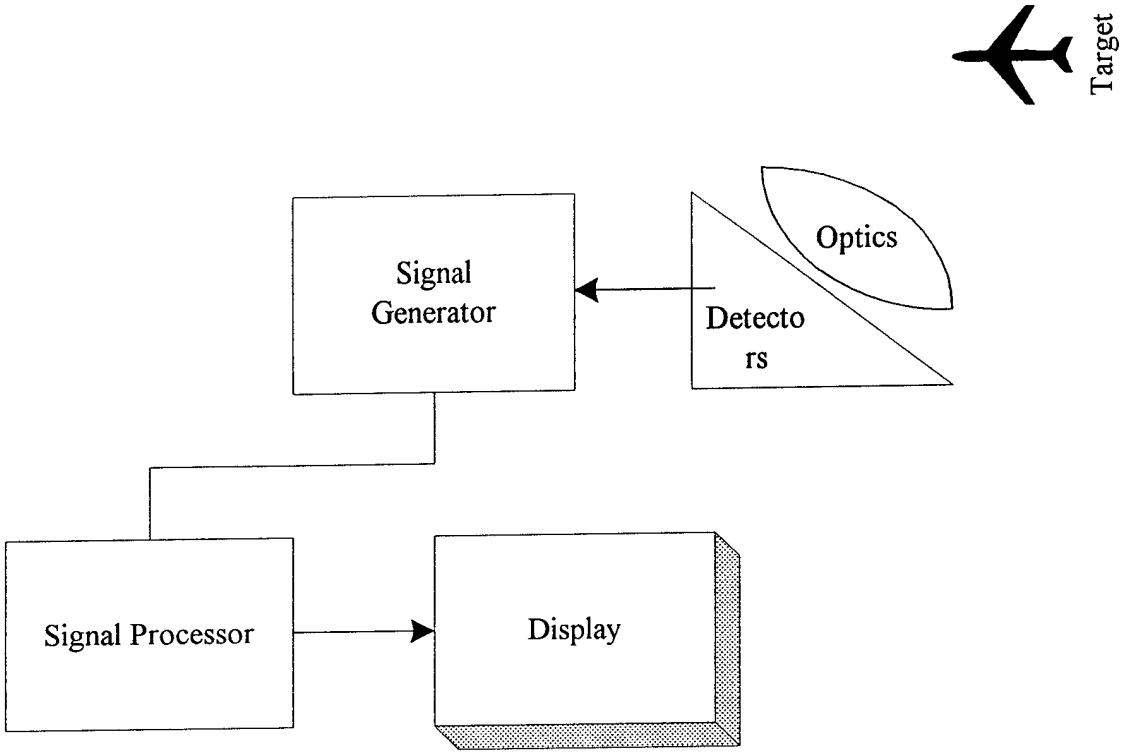


Figure 10: A Simple Electro-Optical Instrument

The latter is used in multi-chromatic television where three detectors, corresponding to red, blue, and green spectral bands, generate three separate signals that are used to form a "colored" image. The former is the basis for long wavelength systems such as night viewers, sniperscopes, and Forward Looking InfraRed (FLIR) sensors. Night viewers and sniperscopes generally operate in the near IR band while FLIRs generally operate in the mid or far IR bands. Additionally, night viewers and sniperscopes may have an integral light source that is either designed and manufactured, or filtered to have no visible band emissions. This allows them to be operated without being visible. Of course, these light sources can still be seen by other sensors operating in the same spectral band and oriented to "see" the FOI. The situation is analogous to that of SONAR in that these instruments can usually be operated either actively (with illumination source) or passively (with ambient light.)

In general, most imaging optical instruments are designed to have FOV greater than the search/detection FOV of the eye, necessitating search of the image. Typically, these FOV will be of the order of $5°x3°$ (azimuth x elevation) in a narrow FOV and $15°x10°$ in wide FOV for search and targeting. Instruments intended for navigation (driving and piloting) usually have at

least 4 times larger FOV - resolution is less important than safe movement.

There is some natural loss of resolution in electro-optical (eo) instruments. This arises from the small number of detectors and display picture elements (pixels,) as compared to the number of rods and cones in the eye. Thus a key engineering consideration in eo instruments is to have as many detectors and pixels as possible. The smaller the IFOV of the detectors, the better, consistent with a trade-off of detector sensitivity. In general, the voltage produced by the detectors is a direct function of the number of photons incident onto the detector, so the goal is small IFOV detectors with high sensitivity. Since sensitivity can usually be increased by cooling the detectors to low (absolute) temperatures, early and many current eo instruments cooled the detector elements to temperatures approaching that of liquid nitrogen. This necessitated considerable cooling gas and a relatively large Dewar for the detectors.

Another factor is the amount of time that the detectors generate voltage before the displayed "image" is redrawn. The longer the time, the more photons that impinge the detectors and thus more voltage (and sensitivity to the image). The trade-off is that if there is movement in the image, this translates into more blurring in the "image".

There are basically two ways of constructing the detector apparatus: scanning and staring. Scanning apparatus either moves the detectors or the focal plane to scan the image across the detectors. In principle, a single detector can be used to scan the image although this results in a very long time between redraws of the "image". This technique is therefore only really useful if the target perceived contrast is very high and tends to limit this approach, which admittedly minimizes the size of the sensor Dewar and cooling gas, to Air Defense seeker applications. A practical compromise was to use a linear array of detectors (the Common Module detectors in most U.S. Army First Generation FLIRs,) that scanned across the image. One disadvantage of this approach is that the MTF is different in the two image directions due to the motion of the detectors. Considerable efforts have been made to develop areal arrays of detectors.

Staring eo instruments must use areal arrays of detectors to operate. Basically, the idea is to completely cover the image area with detectors so that each detector's IFOV always looks at the same piece of the FOV. Sadly, the same sensitivity problems still apply, so that it is necessary to look for some period of time, redraw the "image", and repeat to assure a good trade-off between sensitivity (by time averaging of the image's photons) and blurring. For this reason, scanning, even with areal arrays, is still a useful design technique.

It is possible to infer greater "image" resolution than the number of detectors would imply. If there are nxm detectors in the areal array (or n detectors and the signal is "chopped" m times while scanning across the image,) then the image's structure is averaged across these nm detector IFOV. The detector apparatus thus generates nm voltages. Since normal television (the kind we watch John Wayne movies on!) has a resolution of $400x600$ lines (which equates to pixels in this case,) we would need either a $400x600$ detector areal array, or a 400 detector linear array, "chopped" 600 times (or *visa versa*,) to provide a TV screen worth of "image". Happily, by taking the averages of adjacent groups of 2 and 4 detectors (or the equivalent scanning situation,) the intermediary voltages (image averages) can be approximated. This reduces the number of

detector elements by a factor of four, to $200x300$, which is considerably more practicable.

As we would expect, since the eo instrument consists of several elements, they all have MTFs. In fact, even the electronic components can be represented by MTFs since they can be represented as Linear Systems and the electrical signals have spectra. The MTF of the detector has both spatial and temporal effects due to its finite size and finite integration time, respectively. To account for this temporal effect, Ratches [15] makes an association for scanned instruments between spatial frequency, temporal frequency, and scan rate,

$$f = s\nu, \tag{30}$$

where:

f = temporal frequency (Hz),

s = scan rate (milliradians per sec), and

ν = spatial frequency (cycles per milliradian). (We use cycles here because the quantity of interest is cycles of bar pattern.)

The MTF of the detector has two parts, one temporal and one spatial:

$$M_{\text{det}}(\nu, f) = M_{d|s}(\nu)\, M_{d|t}(f). \tag{31}$$

The spatial part of the MTF is

$$M_{d|s}(\nu) = \frac{\sin(\pi\nu IFOV)}{\pi\nu IFOV}, \tag{32}$$

where $IFOV$ is the instantaneous FOV of a detector element. The observant student will note that this is nothing more than the FT of a step function centered at the origin, indicating uniform sensitivity of the detector element. The temporal part has the approximate form of a resistive-capacitive (RC) circuit,

$$M_{d|t}(f) \simeq \frac{1}{\sqrt{1 + \left(\frac{f}{f^*}\right)^2}}, \tag{33}$$

where: f^* is the $3dB$ (50%) point in the detector's temporal response (frequency). This can be transformed into a spatial frequency representation via equation 30.

The MTF for the electronics is similar to that of the detectors,

$$M_{el}(f) \simeq \frac{1}{\sqrt{1 + \left(\frac{f}{f_0}\right)^2}}, \tag{34}$$

where: f_0 is the 50% point of the electronics response. The MTF for the display is gaussian in the spatial frequency with suitable parameter,

26.7 Systems

If we are interested in the performance of the optical instrument as a system, (and we usually are,) then the quantity of total interest is the system MTF. As we outlined in the Optics II Chapter, MTFs are multiplicative, so that the system MTF is made up of all the MTFs of the parts. For a direct view optical instrument, the system MTF is thus

$$M_{system} = M_{ext} M_{opt} M_{stab} M_{eye}, \tag{35}$$

where: M_{stab} = a gaussian MTF representing the effect of stabilization. If the optical instrument is supported by the user (e.g., binoculars) then the system is stabilized and the insecurity of that stabilization must be considered. This is the opposite of how we usually think of the term. If the system is rigidly stabilized, such as a fixed mount telescope isolated from human motion (breathing, etc.,) then stabilization is still present but is perfect - the MTFs parameter is zero and the MTF is always one! Since the purpose of stabilization is to remove outside motional effect on the imaging instrument, it can be thought of as a necessary evil.

For an indirect view, or eo, instrument, the system MTF reflects the increased complexity,

$$M_{system} = M_{ext} M_{opt} M_{stab} M_{\det} M_{el} M_{\det} M_{eye}. \tag{36}$$

These combinations can become exceedingly complex and hard to manipulate. (That's why gaussian approximations are so popular!) Nonetheless, they serve a useful purpose as we shall see in the next chapter.

26.8 References

[1] Ferrill, Arthur, **The Origins of War From the Stone Age to Alexander the Great**, Thames and Hudson LTD, London, 1985.

[2] Newman, Edwin R., "Speech and Hearing", Chapter 3k in Dwight E. Gray, Ph.D., Coordinating Editor, **American Institute of Physics Handbook**, McGraw-Hill Book Company, New York, 1972.

[3] Ingard, Uno, "Acoustics", Chapter 8 in E. U. Condon, Ph.D., and Hugh Odishaw, Sc.D., **Handbook of Physics**, McGraw-Hill Book Company, New York, 1958.

[4] Blackwell, H. R., "Contrast Thresholds of the Human Eye", Journal of the Optical Society of America 36 no. 11, pp. 624-643, 1946.

[5] Fowler, Bruce W., "Environmental Effects on Combat Performance; A Lanchester Approach", in L.G. Callahan, Jr., and Lawrence Low, **Proceedings of the Workshop on Mod-**

eling, Simulation, and Gaming of Warfare, Callaway Gardens, GA, 2-5 December 1984, Georgia Institute of Technology, Atlanta, GA.

[6] Lloyd, J. M., **Thermal Imaging Systems**, Plenum Press, New York, 1975.

[7] Radio Corporation of America, **Electro-Optics Handbook**, Technical Series EOH-11, Solid State Division, Electro-Optics and Devices, Lancaster, PA 1974.

[8] Kornfield, G. H., and W. R. Lawson, "Visual-Perception Models", Journal of the Optical Society of America 61 no. 6, pp. 811-820, 1971.

[9] Tuma, Jan J., **Handbook of Physical Calculations**, McGraw-Hill Book Company, New York, 1983.

[10] Burke, James, **Connections**, Little, Brown and Company, Boston, 1978.

[11] Burke, James, **The Day The Universe Changed**, Little, Brown and Company, Boston, 1985.

[12] Barton, David K., **Modern Radar System Analysis**, Artech House, Norwood, MA, 1988.

[13] Brookner, Eli., "Fundamentals of Radar Design", Chapter 1 in Eli Brookner, ed., **Radar Technology**, Artech House, Dedham, MA, 1980.

[14] Mie, G., *Ann. Physik* 25 377 (1908), and other references as cited in E. U. Condon, "Molecular Optics", Chapter 6 in E. U. Condon, Ph.D., and Hugh Odishaw, Sc.D., **Handbook of Physics**, McGraw-Hill Book Company, New York, 1958, pp. 6-124,125.

[15] Ratches, James A., et al., "Night Vision Laboratory Static Performance Model for Thermal Viewing Systems", U.S. Army Electronics Command Night Vision Laboratory, Ft. Belvoir, VA, Research and Development Technical Report ECOM-7043, April 1975.

Chapter 27

Search

27.1 Introduction

With this chapter, we shift gears a bit from our intensive discussion of matters optical. We begin to lay the basic foundation for a top level combat attrition process - target acquisition. The final two pieces that will tie all of this together are search and detection. There is a dilemma implict in these two pieces; a basic disconnection. This disconnection will become evident as we cover the basics of search (this chapter) and detection (the next), and shall become a focus of the following chapter that weds the two.

27.2 Simple Search

Search is the process of seeking an object; in our case, a target. In general, this is a very complicated subject, [1], [2], [3], [4], as evidenced by the extensive classical literature, arising from the period of the Second World War and hence. Much of this literature is concerned with optimal methods of search. We shall be content with a very simple subset of this extensive literature, partly because we are dealing with our total subject at a overview level, and partly because we are primarily dealing with search processes which are human based. At best these processes are innate, modified by doctrinal training or by the necessities of engineering implementation of hardware, but human nonetheless.

In keeping with this, we shall be primarily concerned with two simple types of search: random and scanned. For most of what we will be considering, target motion will be relatively slow compared to search speed, so that we will not generally be concerned with the problem of targets moving into regions already searched. These two simple searches represent the two extremes of simple searches in terms of order, and thus provide a reasonable picture of the extremes in search performance that may be expected.

The basic idea of search is simple. A search space is suspected to contain one or more targets.

This search space must be examined to find the target(s). A common assumption is that the space will continue to be searched until a target is found.

The search space is defined by either a solid angle Ω or equivalently, an area, A,[1]. The solid angle that can be searched at any instant is ω (area a.) Thus, the search space consists of N_s individual search subspaces,

$$N_s \simeq \frac{\Omega}{\omega} \simeq \frac{A}{a}. \tag{1}$$

(This is approximate because of the assumption of rectangularity and, more important, even divisibility!) For simplicity, we shall assume the geometry of the search space to be rectangular, although this is not necessary. Thus,

$$N_s = n_x n_y, \tag{2}$$

where: n_x, n_y are the number of subspaces along the edges of the search space. We have assumed here that the subspaces are square although this is neither necessary nor even usual. This assumption does not seriously detract from what we will discuss here.

At any given instant, the searcher can only examine one subspace. The targets are assumed to be randomly placed in the space, frequently with the further restriction of only one target per subspace. Targets are smaller than the subspace. A common convention is to refer to each inspection of a subspace as a *glance* or *glimpse*.[2]

27.3 Simple Search without Detection

We shall start by considering the search process without detection. By this, we mean that the question of uncertainty of detection does not enter into our consideration at this point. Simply put, if a target is located in a search subspace, then the target is found. We shall consider the problem of search with uncertainty of detection below. The problem we consider here is the process by which the subspaces are examined. For simplicity, we shall initially assume that there is only one target in the search space.

27.3.1 Random Search

The basic characterization of search as we have described it above is how individual subspaces are selected for inspection. The simplest scheme for doing this is to select a subspace at random, inspect that subspace, and if the target is not contained in that subspace, then repeat the process. In random search, there is no memory of which subspaces have been searched (we will amend

[1]If the search space is a solid angle Ω, then that angle can be converted into an area by projection at a range r, $A = \Omega r^2$. Conversely, if an area A is to be searched, it represents a solid angle Ω at some effective range r, $\Omega = \frac{A}{r^2}$, assuming, in both cases, $\Omega \ll \pi$.

[2]The term glance seems to be used primarily by the search community while glimpse seems to be used primarily by the detection community,

this later), and thus each time a subspace is selected, the probability of selecting any subspace is the same (assuming the selection is uniformly random.) Thus, the probability of selecting a particular (any) subspace is

$$p_r \simeq \frac{1}{N_s}, \tag{3}$$

which is again approximate because of the assumption of even divisibility. A more exact form is

$$p_r = \frac{\omega}{\Omega} = \frac{a}{A}. \tag{4}$$

Importantly, this is also the probability that the (single) target is present in a subspace.

We now consider individual glances. The probability of finding the target on the first glance is just

$$p_1 = p_r. \tag{5}$$

The probability of finding the target on the second glance is

$$p_2 = (1 - p_r)\, p_r, \tag{6}$$

which is just the conditional probability of not finding the target on the first glance times the conditional probability of finding the target. By abstracting this, we may write the probability of finding the target on the i^{th} glance as

$$p_i = (1 - p_r)^{i-1} p_r. \tag{7}$$

From this, we may now compute the cumulative probability of having found the target after i glances. This is simply the sum of the individual glance probabilities,

$$P_i = \sum_{j=1}^{i} (1 - p_r)^{j-1} p_r. \tag{8}$$

We shall leave as an exercise to the student to demonstrate that $P_i \to 1, i \to \infty$. This result however, is important, since it says that if you keep searching, you will find the target! While this seems obvious, it is a point whose importance will become obvious as we proceed.

Another important quantity is the expected number of glances to find the target? This can be written using equation 7 as

$$\langle i \rangle_{rs} = \sum_{j=1}^{\infty} (1 - p_r)^{j-1} j p_r. \tag{9}$$

This equation can be analytically summed using techniques described in Part I of this work, so we shall only present the result,

$$\langle i \rangle_{rs} = \frac{1}{p_r}. \tag{10}$$

Using equation 3, this is just

$$\langle i \rangle_{rs} \simeq N_s, \tag{11}$$

which simply states that the expected number of glances to find a (single) target using random search (and certain detection) is just the number of search subspaces.

27.3.2 Scanned Search

The other type of simple search that we consider is the scanned search. This is also called a raster search. In this search technique, search typically is initiated in a corner subspace. If the target is found, search ceases, otherwise the left/right/above/below adjacent subspace is selected. Alternately, subspaces along a row or column are selected. We shall take the first for our discussion although the opposite is equally valid.

In this convention, the first subspace selected is row 1, column 1, the next subspace is row 1, column 2, etc. The selection continues until the end of the row is reached, selected, and searched. (All this assumes no finding.) At this point, the scheme either moves up one row in the same column and moves back along the row in the opposite direction, or moves up one row and returns to column 1. Either way, the process is repeated until the target is found or the space is searched (these are equivalent statements.)

We now consider the process of glancing at the space. On the first glance, the probability of finding the target is

$$p_1 \simeq \frac{1}{N_s}. \tag{12}$$

On the second glance, a subspace has been searched, so the probability of finding the target is

$$\begin{aligned} p_2 &\simeq \left(1 - \frac{1}{N_s}\right) \frac{1}{N_s - 1} \\ &\simeq \frac{1}{N_s}. \end{aligned} \tag{13}$$

From this, we may infer that

$$p_i \simeq \frac{1}{N_s}. \tag{14}$$

We may immediately write the cumulative probability of finding the target by the i^{th} glance as

$$P_i = \sum_{j=1}^{i} \frac{1}{N_s} = \frac{i}{N_s}. \tag{15}$$

The expected number of glances to find the target,

$$\langle i \rangle_{ss} = \sum_{j=1}^{N_s} \frac{i}{N_s}, \tag{16}$$

can be solved by use of the finite difference calculus.[5] The factorial function $k^{[l]}$ is defined by

$$k^{[l]} \equiv k\,(k-1)\,..\,(k-l+1)\,. \tag{17}$$

This function has the summation property,

$$\sum_{k=k_1}^{k_2} k^{[l]} = \frac{k^{[l+1]}}{l+1}\,|_{k=k_1}^{k=k_2+1}\,. \tag{18}$$

From this, we may rewrite equation 16 as

$$\langle i \rangle_{ss} = \sum_{j=1}^{N_s} \frac{i^{[1]}}{N_s}, \tag{19}$$

which we may immediately sum as

$$\langle i \rangle_{ss} = \frac{i^{[2]}}{2N_s}\,|_{i=1}^{i=N_s+1} = \frac{(N_s+1)\,N_s - 2}{2N_s}. \tag{20}$$

If $N_s \gg 1$, then this is approximately

$$\langle i \rangle_{ss} \simeq \frac{N_s}{2}, \tag{21}$$

which is half of the value obtained for random search.

27.3.3 Comparison of Random and Scanned Search

As we have seen, the primary difference between these two search techniques is the manner in which subspaces are selected for attention. The random search selects subspaces at random and has no memory; the scanned search has an absolute way of selecting subspaces. The

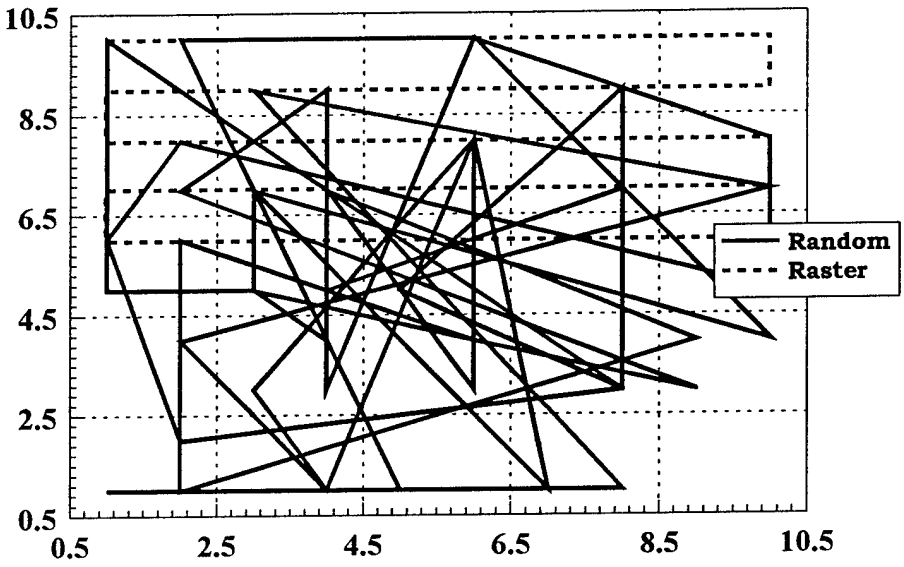


Figure 1: Representative Searches - Two Dimensional View

fundamental differences between these two can be illustrated by examining the way a search space is searched in the two techniques. This is illustrated in figures 1 and 2. (We have cheated a bit here, generating the random search by selecting separately for x,y coordinates, but the illustration still conveys the intent.) The first figure is a two-dimensional overhead view of the search space with the subspace selections shown as a trajectory. This figure shows the difference between the two types of searches, but does not completely convey the confusion of the random search. For this reason, we also present the same information as a three-dimensional trajectory (with glance number as the z coordinate,) which displays the confusion of the random search.

Of course, the random search is less efficient than the scanned search, and we illustrate this in Figure 3. This figure, conducted over a search space of fifty subspaces, shows the cumulative probability of finding the target after a number of glances, equations 7 and 15. The plot is truncated after 50 glances when the scanned search reaches certainty of finding the target. It may be noted that for the scanned search, a cumulative probability of finding of 0.5 is reached after 25 glances, which is what we would expect from equation 21. The 0.5 cumulative probability of finding for the random search is reached after 35 glances, which would seem to be at odds with equation 11. This would seem to indicate that the expected number of glances should be 70 glances except for two things which are really the same. First, the random search cumulative probability is not a simple linear function so this doubling does not apply. Second, as the random search continues, more and more subspaces have been examined, and the probability

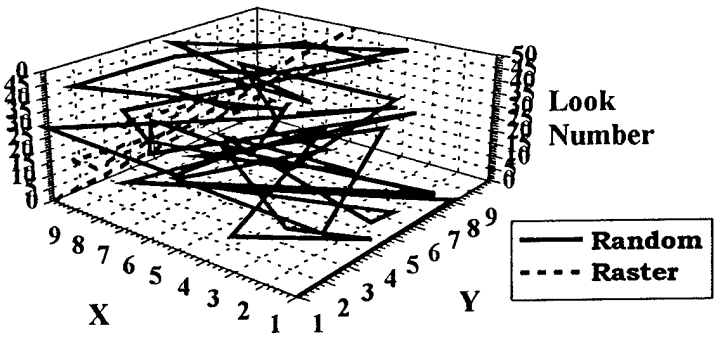


Figure 2: Representative Searches - Three Dimensional

of revisiting a subspace that has already been examined increases. Thus the random search for cumulative probabilities > 0.5 is less efficient than for cumulative probabilities < 0.5. In fact, as one may see, up to a cumulative probability of about 0.2 or about 10 glances, the two techniques are approximately equally efficient. Thereafter they increasingly diverge.

Of course, the scanned search cumulative probability is one for any number of glances $>$ 50, so that eventually, the two search techniques converge again for a large number of glances. Interestingly, it takes about 228 glances to reach a cumulative probability of 0.99, and 342 glances to reach 0.999 for the random search. If we take the latter number as "close enough" to certainty, then (for this example,) the scanned search is almost seven times more efficient than the random search.

Given this vast discrepancy in efficiency, why ever use a random search in preference to a scanned search? Well, there are two reasons (in my mind) for this: first, we are interested in how humans search; and we have not yet considered uncertain detection. As we shall see in the next section, uncertain detection levels the efficiency of the two search techniques.

27.4 Simple Search with Detection

Now we allow the detection process to become uncertain. The search processes still proceed in the same manner as before, but now they have a somewhat different interpretation. The glance probabilities, p_i, may now be interpreted as the probability that the searcher examines the subspace that contains the target. Detection, however, is no longer certain, and we introduce the single glimpse probability of detection, p_d, which is the conditional probability of detecting the target *in a single glimpse* given that the target is located in the glimpse FOV.

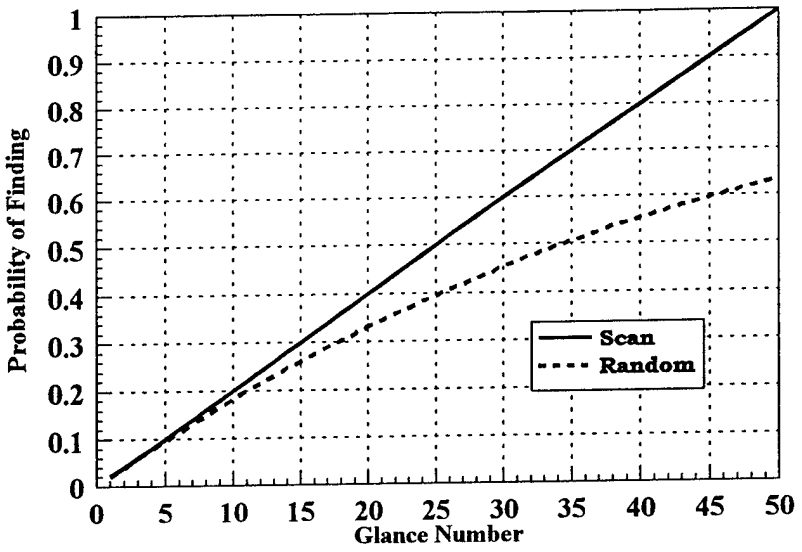


Figure 3: Cumulative Probability of Finding Target

27.4.1 Random Search

As we have indicated, the search process proceeds as before, except that detection of the target, given its presence in the subspace, is no longer certain. Under these new conditions, the probability of detecting the target on the i^{th} glance (glimpse) is simply,[3]

$$p_i = (1 - p_r p_d)^{i-1} p_r p_d, \tag{22}$$

which is just a simple scaling of $p_r \rightarrow p_r p_d$. This scaling carries through in all of the formulations of random search, so that the cumulative probability of detection becomes

$$\begin{aligned} P_i &= \sum_{j=1}^{i} (1 - p_r p_d)^{j-1} p_r p_d \\ &= 1 - (1 - p_r p_d)^i , \end{aligned} \tag{23}$$

and the expected number of glances is

$$\langle i \rangle_{rs} = \frac{1}{p_r p_d} \simeq \frac{N_s}{p_d}. \tag{24}$$

[3]Notice that we now talk about detecting the target rather than just finding it. We are still interested in finding the target, but with the introduction of detection uncertainty, the process becomes dominated by detection.

27.4.2 Scanned Search

Scanned search changes considerably with the introduction of detection uncertainty. The probability of detection on the first glance is just

$$p_1 = \frac{p_d}{N_s}, \tag{25}$$

which is a simple scaling of equation 12. The probability of detection on the second glimpse is similarly

$$p_2 = \left(1 - \frac{p_d}{N_s}\right) \frac{p_d}{N_s - 1}, \tag{26}$$

and on the third glance,

$$p_2 = \left(1 - \frac{p_d}{N_s}\right) \left(1 - \frac{p_d}{N_s - 1}\right) \frac{p_d}{N_s - 2}. \tag{27}$$

These two equations demonstrate a simple fact, the neat cancellation of increasing subspace probabilities (of target presence) does not occur when detection becomes uncertain. Thus, the cumulative probability of detection is the complicated formula

$$P_i = \sum_{j=1}^{i} \frac{p_d}{N_s - j} \prod_{k=0}^{j-1} \left(1 - \frac{p_d}{N_s - k}\right), \tag{28}$$

which is less than obviously analytically summable, does not clearly seem to admit calculating an expected value, and has a singularity when $i = N_s$. This is clearly not a fruitful approach to describing scanned search under conditions of detection uncertainty.

To avoid this problem, we exploit the linear nature of scanned search. If a single target is present, then during a complete scanned search of the space, the target is "seen" exactly once in N_s examinations. The same is true for each subsequent complete scan. If we now treat each complete glance as a "super" glance, then the probability of detecting the target in the first complete scan ("super" glance,) is just

$$q_1 = p_d, \tag{29}$$

and the probability of detecting the target on the i^{th} complete scan is just

$$q_i = (1 - p_d)^{i-1} p_d. \tag{30}$$

We have shifted notation here to q_i to designate "super" glances. Even with this change, we now see that the cumulative probability of detecting the target after i complete scans is the same type of geometric series that we have seen as characteristic of random search,

$$\begin{aligned} Q_i &= \sum_{j=1}^{i} (1-p_d)^{j-1} p_d \\ &= 1-(1-p_d)^i . \end{aligned} \tag{31}$$

This has an expected number of "super" glances to detection of

$$\langle i \rangle_{sg} = \frac{1}{p_d}, \tag{32}$$

which we may multiply by N_s to get the expected number of glances (approximately) as

$$\langle i \rangle_{ss} = \frac{N_s}{p_d}, \tag{33}$$

which is identical to the value for random search, equation 24. Thus, in the approximation of the "super" glances, the two search techniques have equal expected number of glances.

This is an approximation in that we are counting "super" glances, ignoring the fact that somewhere during the i^{th} "super" glance, detection will occur (somewhere between glance 1 and glance N_s,) and search will cease. Thus, we are overcounting the number of glances by a quantity equal to the difference between the glance when detection occurs, and N_s. From our knowledge of the linear form of scanned search without detection, we know that the expected value of this overage is just $\frac{N_s}{2}$.

If we take this as an estimator of the overage, and subtract it from equation 33

$$\begin{aligned} \langle i \rangle'_{ss} &\simeq \frac{N_s}{p_d} - \frac{N_s}{2} \\ &\simeq \frac{N_s}{p_d}\left(1-\frac{p_d}{2}\right), \end{aligned} \tag{34}$$

then we see that as $p_d \rightarrow 1$, this has the proper behavior, equation 21. For nominal value of p_d, say, 0.5 and 0.25, then we may see that scanned search is only $\frac{4}{3}$ and $\frac{8}{7}$ more efficient than random search. Thus, for small single glimpse probabilities of detection, there is little difference in terms of expected number of glances between the two search techniques.

27.5 Conclusion

As we have indicated search is a complicated and extensive subject and we have touched quite lightly upon it. We now abandon our discussion of search for a while to lay the foundation for detection in the next chapter.

27.6 References

[1] Koopman, Bernard Osgood, "The Theory of Search", Parts I-III, Operations Research 4 pp. 324-346, 1956, 5 pp. 503-531, 1956, and 5 pp. 613-626, 1957.

[2] Koopman, Bernard Osgood, **Search and Screening General Principles with Historical Examples**, Pergamon Press, New York, 1980.

[3] Stone, Lawrence D., **Theory of Optimal Search,** Military Applications Section, Operations Research Society of America, 1989.

[4] Washburn, Alan R., **Search and Detection**, Military Applications Section, Operations Research Society of America, 1989.

[5] Levy, H., and F. Lessman, **Finite Difference Equations**, Dover Publications, Inc., New York, 1992.

This Page Intentionally Left Blank

Chapter 28

Detection

28.1 Introduction

In the preceding section, we considered search with detection as a natural adjunct. In this section, we now shift to consider the basic theory of detection with search as an adjunct. Like search, detection work dates primarily from the Second World War. Despite security restrictions, the basic theory developed essentially concurrently in both the radar and imaging communities, albeit with somewhat different terminologies.[1]

This theory postulates that an object to be detected generates a signal (an observable signature) that is viewed in the context of a noise environment. Noise arises from several sources including the target's physical background, and the noise of the sensing system. The signal is treated as a deterministic quantity while the noise is stochastic, possessing both mean and variance, accommodating both deterministic and random noise sources. Detection is then postulated to be a normally distributed stochastic process with probability given by

$$P_d\left(\frac{S}{N}\right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\frac{S}{N}-\mu} e^{-\frac{x^2}{2\sigma^2}} dx, \tag{1}$$

where:

$\frac{S}{N}$ = signal-to-noise ratio,

μ = mean signal-to-noise ratio (50% probability of detection value,) and

σ = standard deviation of signal-to-noise ratio.[1], [2] There is experimental data that indicates for simple situations, $\mu \approx 3.2$, and $\sigma \approx 1$. [3] This applies to human eye detection.

Before proceeding, it is worthwhile to note that we have already provided a useful calculational approximation for equation 1 in Appendix F of Part I. Using this approximation,

[1] Actually, the radar terminology is the more mature and established since it basically dates from the beginnings of radar. Imaging terminology has been revised, notably by the introduction of the MTF.

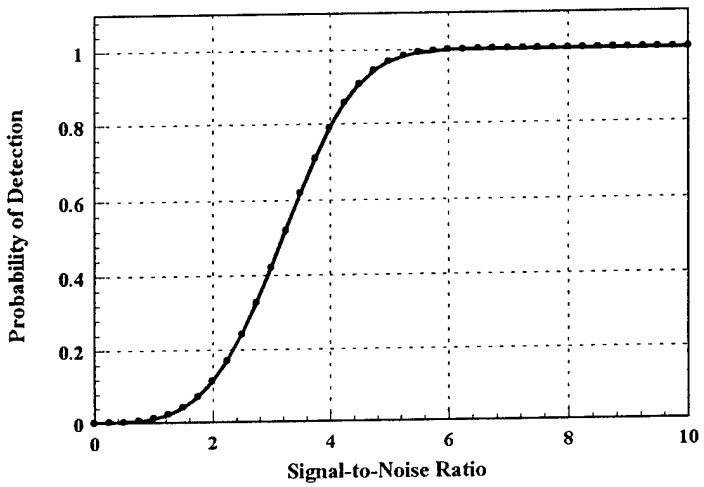


Figure 1: Approximate Normal Probability of Detection Curve

$$P_d\left(\frac{S}{N}\right) \simeq \frac{1 + sign\left(\frac{S}{N} - \mu\right)\sqrt{1 - \exp\left[-\frac{2\left(\frac{S}{N}-\mu\right)^2}{\pi\sigma^2}\right]}}{2}. \tag{2}$$

The behavior of this relationship is presented in Figure 1 where we plot P_d versus $\frac{S}{N}$ for the parameters above. This display the characteristic "S" shape of these probability of detection curves. The student must be advised that this chart is a mirror image of what will usually be seen since $\frac{S}{N}$ usually decreases with range so that P_d increases with decreasing range.

28.2 Radar Detection

Radar, being almost exclusively non-imaging, has developed a detection theory somewhat different from what we have just introduced. Because of the pervasiveness of radars, this theory is typical of non-imaging detection approaches . The actual determination of detection is essentially

performed by the signal processor acting on the basis of the received signals.[2] A key consideration in radar detection engineering is a trade-off between reducing false alarms (detections of false targets) and missing real targets. Because noise is stochastic, a random noise fluctuation may be interpreted by the signal processor as a detection. To offset this, the radar is configured to only accept signals that exceed a threshold value.[4]

In addition, it seems normal to cast the detection process in terms of the voltage the signal produces in the receiver electronics. On physical grounds, this gives

$$v^2 \sim S, \tag{3}$$

which implies all voltages (v) to be positive (or negative) - a simple scaling. Since the electronics is adequately complex to make greater differentiation, a more complicated form,

$$v = sign\,(S - S_0)\sqrt{|S - S_0|}, \tag{4}$$

where S_0 is some reference signal, would seem a more reasonable generalization that is consistent with the statistics.

As we indicated, the false alarm probability is a key parameter in Radar detection. Since the primary source of false alarms is presumed to be noise fluctuations, the calculation of false alarm probability derives from the distribution of noise. The pdf of noise for coherent detection is

$$p_{Nc}\,(v) = \frac{1}{\sqrt{2\pi\,\langle N\rangle}} \exp\left[-\frac{v^2}{2\,\langle N\rangle}\right], \tag{5}$$

which is gaussian with zero mean and standard deviation $\sqrt{\langle N\rangle}$, where $\langle N\rangle$ is the mean noise level. A time moving average of this mean noise level would seem a logical choice for the reference signal. Contrastingly, the pdf of noise for single pulse incoherent detection is

$$p_{Ni}\,(v) = \frac{v}{\langle N\rangle} \exp\left[-\frac{v^2}{2\,\langle N\rangle}\right], \tag{6}$$

which is Rayleigh. The primary difference between coherent and incoherent detection for our purposes is whether pulses are summed before or after being subjected to detection evaluation. A Probability of false alarms, P_{fa}, is calculated using the appropriate noise pdf,

$$P_{fa} = \int_{v_t}^{\infty} p_N\,(v)\,dv. \tag{7}$$

For coherent detection, this gives

[2] In non-imaging systems, the detection is usually done in a signal processor which makes a "decision" to display the detection. The only human detection that is necessary is responding to the display.

$$P_{fa} = \frac{1}{\sqrt{2\pi \langle N \rangle}} \int_{v_t}^{\infty} \exp\left[-\frac{v^2}{2\langle N \rangle}\right] dv, \tag{8}$$

which we may immediately approximate using our integral from Appendix F of Part I as[3]

$$P_{fa} \simeq \frac{1}{2} + \frac{sign(v_t)}{2}\sqrt{1 - \exp\left[-\frac{2v_t^2}{\pi \langle N \rangle}\right]}. \tag{9}$$

This may be solved to yield a value for the threshold signal,

$$v_{tc} \simeq sign(P_{fa} - 0.5)\sqrt{-\frac{\pi \langle N \rangle}{2} \ln\left[1 - (1 - 2P_{fa})^2\right]}, \tag{10}$$

for coherent detection. For incoherent detection, equation 7 is exact, and gives a threshold signal of

$$v_{ti} = \sqrt{-2\langle N \rangle \ln[P_{fa}]}, \tag{11}$$

which implies positive threshold voltages throughout the range.

To calculate the detection probability, signal-plus-noise pdfs are formed. For coherent detection, this pdf has the form

$$p_{SNc}(v) = \frac{1}{\sqrt{2\pi \langle N \rangle}} \exp\left[-\frac{(v + v_s)^2}{2\langle N \rangle}\right], \tag{12}$$

where v_s = received signal voltage. This pdf is gaussian. The corresponding pdf for incoherent detection is

$$p_{SNi}(v) = \frac{v}{\langle N \rangle} \exp\left[-\frac{v^2 + v_s^2}{2\langle N \rangle}\right] I_0\left(\frac{vv_s}{N}\right), \tag{13}$$

where: $I_0(x) = \frac{1}{\pi}\int_0^{\pi} e^{x\cos(\theta)} d\theta$, the modified Bessel function of the first kind and index zero.[5] Probability of detection is calculated using

$$P_d = \int_{v_t}^{\infty} p_{SN}(v)\, dv. \tag{14}$$

For coherent detection, this gives us, using our now familiar approximation,

$$P_{dc} \simeq \frac{1}{2} + \frac{sign(v_t + v_s)}{2}\sqrt{1 - \exp\left[-\frac{2(v_t + v_s)^2}{\pi \langle N \rangle}\right]}. \tag{15}$$

[3]These two equation show why the more general definition of Signal-voltage relation is desirable. Lacking negative voltages, the probability of false alarm can never exceed 0.5.

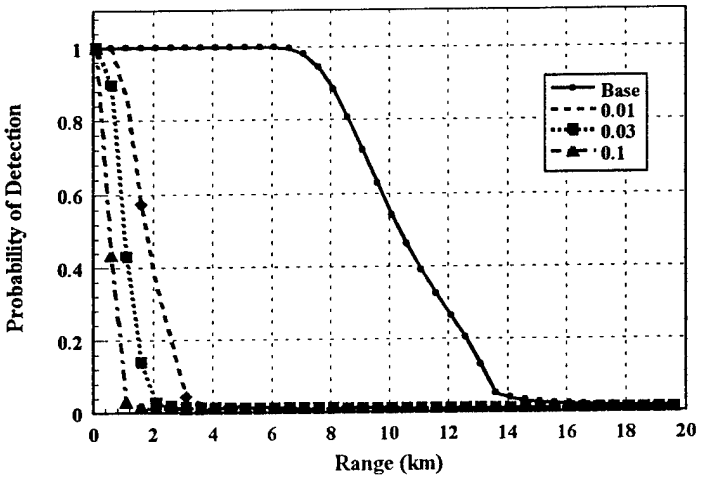


Figure 2: Self-screening and Escourt Jammer Probabilities of Detection

For incoherent detection, a different approximation is necessary since the probability integral is complicated by the Bessel function.[6] This approximation has the form, using our notation, of

$$P_{di} \simeq \frac{1}{2} + \frac{sign\left(\frac{v_t}{\sqrt{\langle N \rangle}} - \sqrt{2\frac{S}{N} - 1}\right)}{2} \sqrt{1 - \exp\left[-\frac{2}{\pi}\left(\frac{v_t}{\sqrt{\langle N \rangle}} - \sqrt{2\frac{S}{N} - 1}\right)^2\right]}, \qquad (16)$$

where: $\frac{S}{N}$ = signal-to-noise ratio as defined in the preceding chapter. We present two examples of coherent detection for the jammer examples of the previous chapter. These are shown in figures 2 and 3 for rather high false alarm probabilities.

28.3 Eye Detection

There are two ways of approaching detection of targets by the human eye (either directly or

Eye Detection

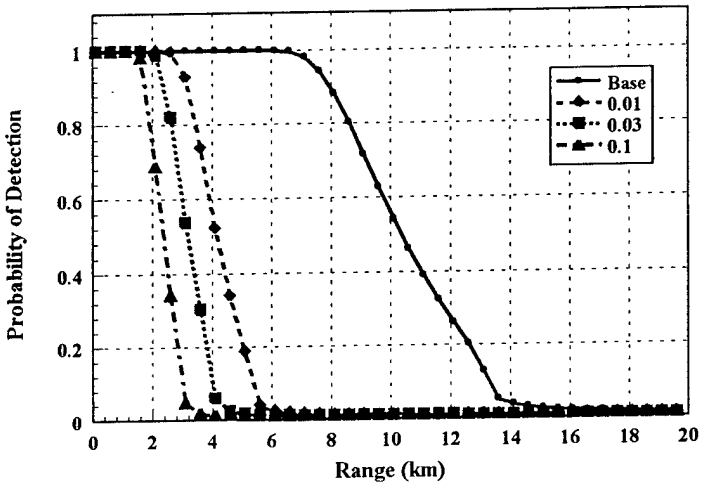


Figure 3: Stand-Off Jammer Probability of Detection Examples

through direct view instruments,) the MTF approach that we have so laboriously laid the foundation for, and the pre-MTF approach introduced during the Second World War. We shall cover both of these approaches, but defer the discussion of the MTF approach till later when we discuss EO Detection.

The basic concepts of imaging detection are usually cast in the framework of a larger process called *Acquisition.* As with any process, the questions of resolution and divisibility depend to a great extent on the circumstances of the analysis and the prejudices and proclivities of the analyst. As we have indicated, Acquisition is one of the primary subprocesses of the Attrition process in our analysis and taxonomy. For now, we shall largely follow the conventions established by Overington and his successors, but with the caveat that when we return to Acquisition in subsequent considerations of the Attrition process, we shall expand the context somewhat in a manner that we believe is still true to the intent of the developers. Nonetheless, the Acquisition that we discuss here will take on the form of a subprocess there.

Overington [7] considered Acquisition to be the process of acquiring information through visual stimuli, specifically in relation to an imaging system that has an output in the form of a

recorded image ("image" in our convention although this is not necessary nor so restricted,) presented to a human observer for direct visual inspection. Thus unlike the discussion immediately above on Radar detection, the phenomenology imposed on this process is inherently human. Acquisition, like Gaul, is divided by Overington into three subcategories: detection; recognition; and identification. These have descriptions:

- Detection is the attainment of awareness by the observer (searcher) of a local and real variation in the image structure;
- Recognition is the first level sorting of detected variations into classes; and
- Identification is recognition of sufficient detain in a detected variation to specify its membership in a specific class.

For example, detection may mean the finding of an object in the image; recognition may mean realization that the object is a military vehicle, and identification may mean realization that the military vehicle is a tank. From an optics standpoint, each subcategory necessitates both greater image resolution and greater light levels detected. If we use the photon picture, then for any image observed over a period of time at least of the order of the eye integration (or glimpse) time, and averaged at the resolution level of the image, there is a mean number of photons $\langle n \rangle$. It is convenient at this juncture to think in terms of the IFOV of individual rods and cones (or for EO systems, picture elements.) If we consider only those rods and cones which receive photons from the image, sum the number of photons received by each of these rods and cones, and divide by the total number of rods and cones, the resulting value is $\langle n \rangle$. Similarly, there is a standard deviation of the number of received photons, and this has a value of $\sqrt{\langle n \rangle}$. This is simply a Poisson distribution.

If these values represent the mean and standard deviations of the image, then the signal-to-noise ratio of any element of the image is

$$\frac{S}{N} = \frac{|n - \langle n \rangle|}{\sqrt{\langle n \rangle}}, \tag{17}$$

which, as we have already postulated, is a measure of the probability that $n - \langle n \rangle$ represents a real deviation in the radiance of an object in the image. Given this, the descriptions of the Acquisition categories can be refined:

- Detection is discrimination of signal from noise;
- Recognition is discrimination of target from clutter; and

- Identification is discrimination of target detail from overall target image.

In the specific military context, much of the basic detection work was begun by Johnson, [8] and continued to this day by his successors. Johnson extended and modified the taxonomy to four subcategories:

1. Detection;
2. Shape Orientation;
3. Shape Recognition; and
4. Detail Recognition.

Equation 17 is noteworthy in that it postulates that the "seeability" of an image element is a function of its difference from the mean image light level. This is somewhat different from what we normally expect from most stochastic processes and accounts for the (not quite) equivalent "seeability" of dark objects against a light background or light objects against a dark background.[4]

From an historical standpoint, the basic statement of probability of detection for the eye was given by Blackwell [9] based on his work commencing during World War II. This statement is couched in terms of perceived contrast, and has the form,

$$P_d = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{C-C_{TH}}{\sigma}} e^{-\frac{x^2}{2}} dx, \tag{18}$$

where: C =perceived contrast, which is a function of range (among other things,) C_{TH} = Threshold Contrast, and the standard deviation σ has a value of 0.05. Of course, this is just a variation on equation 1, and has the approximate form, from equation 2, of

$$P_d(C) \simeq \frac{1 + sign(C - C_{TH}) \sqrt{1 - \exp\left[-\frac{2(C-C_{TH})^2}{\pi\sigma^2}\right]}}{2}. \tag{19}$$

While Threshold Contrast depends on target size, range, and light level, a useful approximation for the restrictive case of bright illumination (clear day) is [13]

$$C_{TH}(r, d_{\min}) \simeq \exp\left[1.933 \left(\frac{r}{d_{\min}}\right)^{0.542} - 5.327\right], \tag{20}$$

[4]The "not quite" arises from the difference in mean light level across the image. This results in a different sensitivity of the eye since it automatically adapts to different overall light levels.

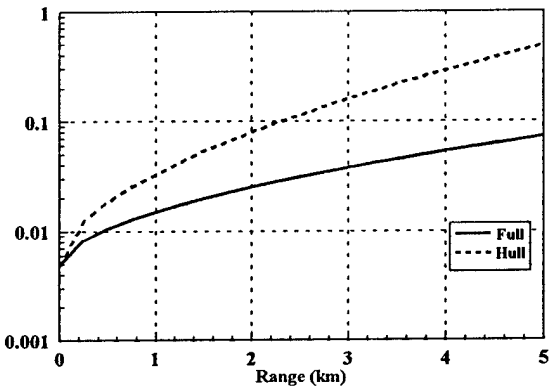


Figure 4: Representative Threshold Contrast Curves

where: r = range from target to observer (km), and $d_{\min}$ = minimum presented dimension of the target (m). Note the difference in units here! If we take as an example a typical Soviet tank (e.g., T-62, T-72,) with frontal dimensions of $2.67m$ x $2.67m$, and specify that in hull defilade (hull hidden from view - only turret visible,) reduces the frontal dimensions to $2.67m$ x $1.00m$, then we may generate representative Threshold contrast curves. These are shown in Figure 4. The label "Full" refers to fully exposed, so that the minimum dimensions used for fully exposed and hull defilade are 2.67 and $1.00m$, respectively.

Blackwell's formula, equation 19, is not used much today, mostly because it tends to underpredict detection probability for probabilities that should be near one. A more useful approximation is [10]

$$\frac{S}{N} \simeq \frac{\frac{C}{C_{TH}} - 1}{0.39}. \tag{21}$$

An example of these two models for the fully exposed and hull defilade examples above, with perceived contrast based on an inherent contrast of 0.2 and an extinction coefficient of $0.2km^{-1}$ (no sky-to-ground ratio) is given in figure 5. The differences in the curves is consistent with the spread of the experimental data. Another useful formula from the same source is the Probability of recognition given detection. This is

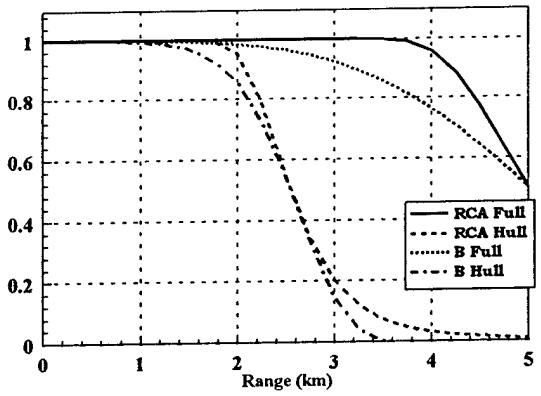


Figure 5: Representative Probability of Detection Curves - RCA and Blackwell Models

$$P_{r|d} \simeq 1 - \exp\left[-\left(\frac{N_r}{2} - 1\right)^2\right], \tag{22}$$

where:

$N_r = \frac{d_{\min}}{\theta r}$,

θ = angular resolution of the sensor, and

r = range. If we assume an FOV of $5°x3°$ with a TV resolution display ($600x400$ lines,) then this gives us an angular resolution of about $0.008°$. A representative pair of curves for Probability of recognition (equation 22 times the probability of detection resulting from equation 21,) are given in figure 6. If we interpret these curves as representing probable ranges where detection (recognition) occurs, then comparison demonstrates that detection generally occurs at longer range than recognition. This is a demonstration of our earlier statement that increasing subcategories of acquisition require greater resolution and more light.

28.4 Electro-Optical Detection

We now come to the more general form of detection theory based on MTFs. The basis of this theory is the work of Johnson.[8] This is based on experimental and theoretical studies of the

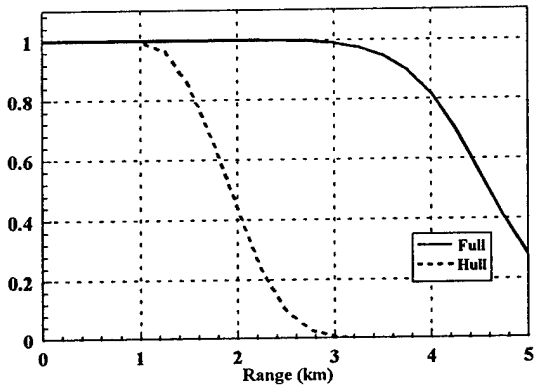


Figure 6: Representative Probability of Recognition Curves - RCA Model

relationship between acquisition subcategories and bar patterns.[11] By observing bar pattern limiting resolution at the threshold of the subcategories, Johnson was able to establish a strong correlation between the number of resolvable cycles (pairs of black and white bar) per minimum target dimension at limiting resolution and probability. In effect, these transforms hold regardless of target perceived contrast and signal-to-noise ratio. This effectively replaces signal-to-noise ratio (or number of photons,) with number of resolvable cycles as the independent variable of the pdf.

Thus, the probability of acquisition subcategory becomes

$$P_{sc} = \frac{1}{\sqrt{2\pi}\sigma_{sc}} \int_{-\infty}^{\frac{n-\mu_{sc}}{\sigma_{sc}}} e^{-\frac{x^2}{2}} dx \qquad (23)$$

$$\simeq \frac{1 + sign\,(n-\mu_{sc})\sqrt{1-\exp\left[-\frac{2(n-\mu_{sc})^2}{\pi\sigma_{sc}^2}\right]}}{2},$$

where:

n = number of (resolution limited) resolvable cycles across target minimum dimension,

μ_{sc} = mean number of (resolution limited) resolvable cycles for an acquisition subcategory, and

σ_{sc} = standard deviation of (resolution limited) resolvable cycles for an acquisition subcategory.

The values of the means and standard deviations are given in the table below:

Subcatagory	Mean	Std. Dev.
Detection	1.0	0.25
Orientation	1.4	0.35
Classification	2.5	1.60
Recognition	4.0	0.80
Identification	6.4	1.50

Resolution Limited Resolvable Cycle Moments

These are commonly called the "Johnson Criteria". It may be noted that an additional subcategory, Classification, has been added.

Within the context of the Johnson Criteria, which are generally applicable to all imaging optical instruments (usually assuming human viewing of the image,) the operant quantity of calculation is now the number of resolvable cycles across the target minimum dimension, n. This has the deceptively simple mathematical form,

$$n = \frac{v_{image} d_{\min}}{r}, \tag{24}$$

where:

v_{image} =spatial frequency of the target, (defined in the Optics II chapter,)

$d_{\min}$ = minimum target dimension, and

r = target-instrument range. In common engineering practice, v_{image} is expressed in units of cycles milliradian^{-1}, so $d_{\min}$ is commonly measured in meters and r is commonly measured in kilometers (so that $\frac{d_{\min}}{r}$ has units of milliradians.) The problem now becomes one of calculating v_{image}.

This is not a trivial problem, but it is simplified by the condition of limiting resolution. It is necessary to find an appropriate relationship between limiting resolution instrument performance and the target's spatial frequency. Happily this can usually be obtained from the instrument engineer. An example of this for a Forward Looking InfraRed (FLIR) instrument is given by Ratches.[12] In this case, the quantity of limiting resolution is Minimum Resolvable Temperature (since contrast is essentially a temperature measure.) The functional form for a First Generation FLIR is

$$T_{MR} = 0.66 \frac{\overline{S} \Delta T_{NE} v}{M(v)} \sqrt{\frac{4}{\pi} \frac{\Delta x \Delta y}{\eta_{ovsc} F_R t_E}} \left(1 + 4v^2 \Delta x^2\right)^{-\frac{1}{4}} \tag{25}$$

where:

T_{MR} = minimum resolvable Temperature,

$\overline{S}$ = signal-to-noise ratio threshold value (given as ≈ 2.25 by Ratches,)

ΔT_{NE} = noise equivalent temperature difference,
v = spatial frequency,
$M(v)$ = MTF of the system,
Δx,Δy = horizontal, vertical detector sizes (milliradians),
η_{ovsc} = overscan ratio,
F_R = Frame Rate, and
t_E = eye integration time (given as $\approx 0.2\,\text{sec}$ by Ratches.)

Since we are interested in the resolution limit, we may simply use the perceived target-background temperature difference as the minimum resolvable temperature, [5]

$$T_{MR} \simeq \Delta T_{tb} e^{-\alpha r}, \tag{26}$$

where: α is the atmospheric extinction coefficient, and r is the target-instrument range. It is now useful to rewrite equation 25 in the form

$$\frac{M(v)}{v}\left(1+4v^2\Delta x^2\right)^{\frac{1}{4}} = \frac{\Delta T_{tb} e^{-\alpha r}}{0.66\bar{S}\Delta T_{NE}}\sqrt{\frac{\pi}{4}\frac{\eta_{ovsc} F_R t_E}{\Delta x \Delta y}} \equiv \Xi, \tag{27}$$

where we have defined the right hand side of the resulting equation to be Ξ for convenience. This equation has all of the spatial frequency dependency on the left hand side, so that the right hand side can be treated as a computed constant for subsequent calculation of v. As we know from our previous discussion, the MTF is a complicated function, so the right hand side of this equation is decidedly non-linear. This makes the calculation of v a problem in non-linear root extraction.

Since many of the MTF component functions are gaussian, the MTF can frequently be approximated as a gaussian,

$$M(v) \simeq e^{-\zeta v^2}, \tag{28}$$

either in toto, or at least, initially. In either case, equation 27 may be approximated as

$$\frac{e^{-\zeta v^2}}{v}\left(1+4v^2\Delta x^2\right)^{\frac{1}{4}} \simeq \Xi. \tag{29}$$

If we take the logarithm of this equation,

$$-\zeta v^2 + \frac{1}{4}\ln\left(1+4v^2\Delta x^2\right) - \ln(v) \simeq \ln(\Xi), \tag{30}$$

which for $v\Delta x \ll 1$, can be expanded to first order,

$$-\zeta v^2 + v^2\Delta x^2 - \ln(v) \simeq \ln(\Xi). \tag{31}$$

[5] Actually, we should calculate the average temperature of the image and use the perceived difference between the target's temperature and this average. If the target is small (occupies a small part of the FOV) and the background is relatively uniform in temperature, then this is a reasonable approximation for the sake of computational simplicity.

If we ignore the logarithm of the spatial frequency, this reduces (with some algebra) to an estimator of the spatial frequency,

$$v^* \simeq \sqrt{\frac{\ln(\Xi)}{\Delta x^2 - \zeta}}, \tag{32}$$

which may either be used to estimate spatial frequency (if all the approximations are valid), or serve as the initial guess for solving equation 27 iteratively.[13]

Happily, this difficult calculation has ceased to be necessary for FLIRs. Ratches [14] has extended his work by consolidating the MTFs effects with the minimum resolvable temperature. (This is essentially a theoretical extension of equations 25-29 onto firmer, physically based ground.) This is written as

$$T_{MR}(v) = T_{MR0} e^{\beta_{sys} v}, \tag{33}$$

where: T_{MR0} = minimum resolvable temperature as $v \to 0$, and β_{sys} = the instrument (system) "extinction" coefficient. As indicated, this extinction coefficient degrades the system's performance (since the exponential is positive, $T_{MR}(v) \geq T_{MR0}$, and thus the system has an increasing minimum resolvable temperature with increasing spatial frequency.) Equation 24 is used to write the spatial frequency as

$$v = \frac{\mu_{sc} r}{d_{\min}}. \tag{34}$$

The signal-to-noise ratio is defined as

$$\frac{S}{N} = \frac{\Delta T_{tb} \sqrt{\frac{\epsilon}{7}}}{T_{MR0}} \exp\left[-\left(\alpha + \frac{\mu \beta_{sys}}{d_{\min}} \right) r \right], \tag{35}$$

where: $\epsilon \equiv \frac{d_{\min}}{d_{\max}}$. The probability of acquisition subcategory is approximated as

$$P_{sc} \simeq \frac{\left(\frac{S}{N}\right)^{2.7+0.7\frac{S}{N}}}{1 + \left(\frac{S}{N}\right)^{2.7+0.7\frac{S}{N}}}, \tag{36}$$

based on a proposal of Kornfeld and Lawson.[15] This signal-to-noise ratio is a normalized one, so that equation 23 may be used but with $\mu_{sc} = 1$, and $\sigma_{sc} \approx 0.5$. This produces close agreement with equation 36 in terms of minimizing squares of differences between the two curves, but does tend to produce higher value for small signal-to-noise ratios. As noted by Ratches, this is not inconsistent with the experimental data, so the student has the option of using either form, bearing in mind that they are *both* approximations.

It is worthwhile at this point to present some representative values and calculations. For targets, we may consider two tanks, presented below:

Target	$d_{min}\,(m)$	ϵ	$\Delta T_{tb}\,(^\circ K)$	C_0^i
Model Tank	3.00	2	2	
Soviet Tank	2.67	1	2	0.2

Tank Target Models for Acquisition Calculations

The characteristics of two representative $(8-10\mu)$ FLIRs are given below:

FLIR	$T_{MR0}\,(^\circ K)$	$\beta_{sys}\left(\frac{cy}{mrad}\right)$
1974	0.0254	0.996
1978	0.0112	0.633

FLIR Models for Acquisition Calculations

The Model Tank and the two FLIRs are taken from the Institute for Defense Analysis/Night Vision Laboratory (IDA/NVL) model.[16] The Soviet Tank model is taken from Fowler.[13] It seems obvious from these tables that the Model Tank is in side (or partial side) view (from the value of ϵ,) while the Soviet Tank is fully frontal. Further, there is evidently considerable technical improvement in FLIR performance from 1974 to 1978. Representative calculations using these two FLIR models with the Model Tank for the stressing case of a radiation fog are presented in figures 7 and 8, representing detection and recognition, respectively. Curves for both approximations to the normal probability integral, equations 23 and 36, are presented. The improved performance of the 1978 FLIR as compared to the 1974 FLIR is obvious, as is the tendency of the original normal probability integral approximation of Appendix F, Part I to predict greater probabilities for small signal-to-noise ratios. The overall shorter ranges for recognition versus detection may also be noted. (The angularity of the curves results from the small number of range values computed for plotting.)

28.5 Detection Comparison and Summary

At this point, it is possible to roll up much of what we have done in a comparison. Specifically, we may compare the performance of the eye with the (arbitrarily selected) 1974 FLIR using the Soviet Tank Model. In this comparison, we consider both fully exposed and hull defilade targets, under weather conditions of Midlatitude Clear, Midlatitide Hazy, and Midlatitude Hazy plus Fog (as described in Optics I.) Contrast is calculated using a sky-to-ground ratio of 4.0. For the sake of comparison, the gaussian probability integral approximation, equations 19 and 23, from Appendix F of Part I, is used. These probability of detection curves are given in figures 9 for a fully exposed target ($d_{\min} = 2.67m$,) and 10 for a hull defilade target ($d_{\min} = 1.0m$.) The curves are coded with a three letter taxonomy ABC, where:

- A = V for visual (eye) spectral band or A = I for InfraRed (FLIR) spectral band;

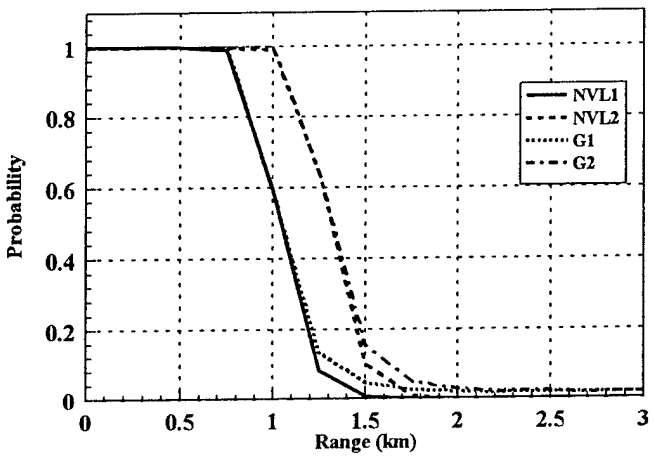


Figure 7: Representative FLIR Probability of Detection Curves

- B = F for fully exposed target or B = H for hull defilade target; and
- C = C for clear, C = H for hazy, or C = F for fog.

These curves readily display the fundamental intensity of interest in the FLIR, it has a vast superiority over the eye under bright, but adverse (even non-adverse) weather. Additionally, it operates as well at night as in daytime (maybe better since ΔT_{tb} is probably larger,) without revealing illumination. This is an enormous promise for military purposes!

From the standpoint of modeling Acquisition for the Attrition process, these simple probability of detection (recognition,..) versus range curves are essentially half of what we need to know about detection. Unfortunately, they also complicate the problem of getting the other half of the information, which is the temporal information.

This is the dilemma that we mentioned in the introduction. These probability versus range curves are something very different from the single glance probabilities of detection we encountered in the section on search - they are infinite time probability of detection curves for static targets. Even ignoring this question of moving targets (which we shall treat later,) the dilemma is

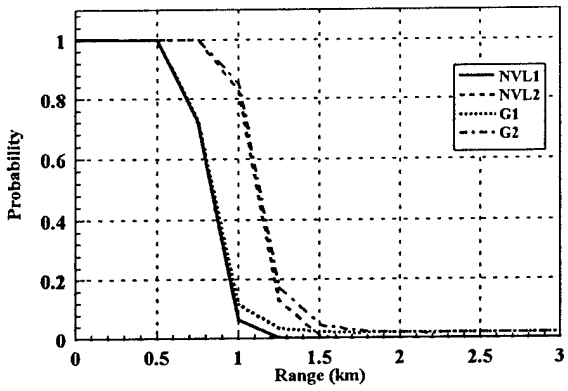


Figure 8: Representative FLIR Probability of Recognition Curves

that search theory, regardless of whether random or scanned search, gives an infinite time probability of detection of one, while detection theory gives infinite time probabilities of detection other than one. We shall now turn to, if not resolving, at least investigating this dilemma.

28.6 References

[1] Lloyd, J. M., Thermal Imaging Systems, Plenum Press, New York, 1975.

[2] Yakushenkov, Yu., Electro-Optical Devices Theory and Design, P. S. Ivanov, trans., Mir Publishers, Moscow, 1983.

[3] Rosell, F. A., and R. H. Wilson, "Recent Psychophysical Experiments and the Display Signal-to-Noise Concept", Chapter 5 in L. M. Biberman, ed., Perception of Displayed Information, Plenum, 1973, as cited in Lloyd.

[4] Barton, David K., Modern Radar Systems Analysis, Artech House, Norwood, MA, 1988.

[5] Abramowitz, Milton, and Irene A. Stegun, eds., Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, U. S. Government Printing Office, Washington,

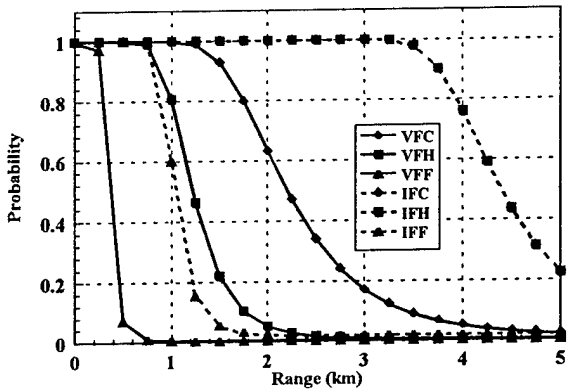


Figure 9: Eye-FLIR Probability of Detection Comparison (Fully Exposed Target)

D.C., May 1968.

[6] North, D. O., "An Analysis of the factors which determine signal/noise discrimination in pulsed carrier systems", Proceedings of the IEEE 51 no. 7, July 1963, pp. 1015-1027.

[7] Overington, J., Vision and Acquisition, Pentech Press, London, 1976.

[8] Johnson, J., "Image Intensifier Symposium", U. S. Army E. R. D. L., Ft. Belvoir, VA, 1958, p. 249.

[9] Blackwell, H. R., "Contrast Thresholds of the Human Eye", Journal of the Optical Society of America 36 pp. 624-643, Nov 1946, "Studies of the Form of Visual Threshold Data", Journal of the Optical Society of America 43 pp. 456–463, June 1953, "Neural Theories of Simple Visual Discrimination", Journal of the Optical Society of America 53 pp. 129-160, Jan 1963.

[10] Radio Corporation of America, Electro-Optics Handbook, Solid State Division, Electro-Optics and Devices, Lancaster, PA, Technical Series EOH-11, 1974.

[11] Wetherell, William B, "The Calculation of Image Quality", Chapter 6 in Robert R. Shannon,

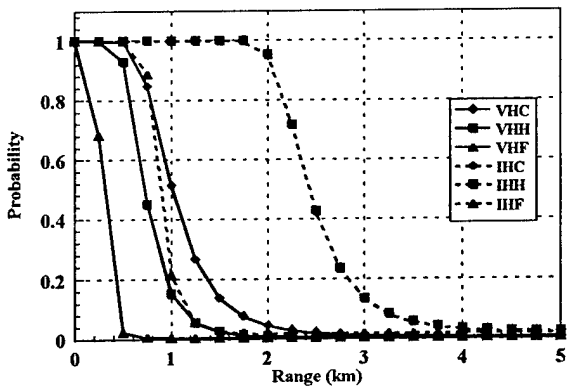


Figure 10: Eye-FLIR Probability of Detection Comparison (Hull Defilade Target)

and James C. Wyant, eds., Applied Optics and Optical Engineering, Vol. VIII, Academic Press, New York, 1980.

[12] Ratches, James A., et al., "Night Vision Laboratory Static Performance Model for Thermal Viewing Systems", U.S. Army Electronics Command Night Vision Laboratory, Ft. Belvoir, VA, Research and Development Technical Report ECOM-7043, April 1975.

[13] Fowler, Bruce W., "Environmental Effects on Combat Performance: A Lanchester Approach", in L. G. Callahan, Jr., and Lawrence Low, eds., Proceedings of the Workshop on Modeling, Simulation, and Gaming of Warfare, Callaway Gardens, GA, 2-5 Dec. 1984, pp. 145-175.

[14] Ratches, J., and C. Nguyen, eds., "FLIR Modeling Workshop I: Performance Summary Measures", CNVEO Report, March 1990.

[15] Kornfeld, G. H., and W. R. Lawson, "Visual-Perception Models", Journal of the Optical Society of America 61 no. 6, June 1971, pp. 811-820.

[16] Seekamp, L. N., Institute for Defense Analysis, Alexandria, VA, IDA Paper P-1419, Sept. 1979, as cited in D. Sadot, N. S. Kopeika, and S. R. Rotman, "Incorporation of Atmospheric

Blurring Effects in Target Acquisition Modeling of Thermal Images", Infrared Physics & Technology 36 No. 2, 1995, pp. 551-564.

[17] Rotman, S. R., E. S. Gordon, and M. K. Kowalczyk, "Modeling human search and target acquisition performance: I. First detection probability in a realistic multitarget scenario", Optical Engineering 28 No. 11, Nov. 1989, pp. 1216-1222.

[18] Rotman, S. R., E. S. Gordon, and M. L. Kowalczyk, "Modeling human search and target acquisition performance: III. Target detection in the presence of obscurants", Optical Engineering 30 No. 6, June 1991, pp. 824-829.

Chapter 29

Search and Detection

29.1 Introduction

With this chapter, we come to a partial conclusion and to a dilemma. The partial conclusion is that with this chapter we will have laid the basic foundation for a top level combat attrition process - target acquisition. The final two pieces that will tie all of this together are here in the form of search and detection. The dilemma is the disconnection between these two pieces. This disconnection will become evident as we proceed, but we shall still cap off the chapter by discussing it.

29.2 Less Simple Search

At this point, we take up the problem of bridging the gap between simple search theory, which admits infinite time probabilities of detection only of one, and simple detection theory, which admits infinite time probabilities of detection for static targets that are not just one. This problem is one that has concerned researchers and practitioners for some time. From the standpoint of the modeler, the problem is that the model(s) are not general enough. One readily identifiable area of inadequacy is in representation of human behavior. Several possible reasons for this non-unity have been suggested, and they are summarized by Rotman et al.[3] These are:

Proposition 1 *If an observer cannot detect a target with* $\left(\frac{S}{N}\right)_1$, *then that same observer cannot detect any target with a smaller signal-to-noise ratio. If each member of the human race has an individual threshold signal-to-noise ratio, and these signal-to-noise ratios thus form a distribution (from population sampling.) Thus, experimental trials (the basis for much of detection theory,) will produce a probability of detection (infinite time) that is proportional to this distribution.*

Proposition 2 *The observer may be unable to discern the target from the background, thus resulting in non-detections.*

Proposition 3 *The observer ceases to search effectively after some time, either due to exhaustion (tiredness) or repetitiveness (boredom.)*

Of these, the first seems eminently plausible, even if it directly confronts and refutes Overington's postulate. While it seems likely that there is such a distribution of sensitivity among the human race, given the extensive dynamic range of the human eye (except for pathological cases,) it is difficult to accept that this distribution can be the whole cause of the detection curve's behavior. This proposition also seems to imply that there will be a set order of detecting targets, which seems contrary to probability theory. Further, the effect of signal-to-noise ratio is already built into Overington's postulate and thus while this proposition adds another phenomena, it does not add a new effect that can be discerned from the available experimental data. Thus, while this proposition seems reasonable, it does not really help resolve the dilemma.

Similarly, the second proposition has an effect which is already incorporated in the basic theory and thereby cannot be used to resolve the dilemma.

The third proposition, however, offers potential for introducing an effect that may help resolve the dilemma. To examine this, we return to random search.

29.3 Random Search with Detection and Stopping

To investigate proposition 3 above, we start by re-examining random search with an additional probability, a probability of stopping the search, because of tiredness, or whatever reason. Initially, we shall assume that this probability is linked to the glance process and is a constant for each glance. We designate this probability of stopping by p_s, and write the cumulative probability of having detected the target by the i^{th} glance as

$$P_i = \sum_{j=1}^{i} (1 - p_r p_d - p_s)^{j-1} p_r p_d. \tag{1}$$

This gives us an infinite number of glances probability of detection of

$$P_\infty = \frac{p_r p_d}{p_r p_d + p_s}, \tag{2}$$

which is less than one for $p_s > 0$.

Since this search technique can terminate by stopping, as well as by detection, we cannot calculate the expected number of glances to detection, but we can calculate the expected number of glances to ending search, either by detection or stopping. This has the value

$$\langle i \rangle_{end} = \frac{1}{p_r p_d + p_s}. \tag{3}$$

Since we cannot distinguish between ending either by detection or by stopping, we shall interpret this value as the mean number of glances to detection, given detection occurs.

Now we want to associate this search technique with the common empirical model of search in conjunction with detection. Since this model is continuous, we analytically sum equation 1,

$$P_i = \frac{p_r p_d}{p_r p_d + p_s}\left[1 - (1 - p_r p_d - p_s)^i\right], \tag{4}$$

and then rewrite it as

$$P_i = \frac{p_r p_d}{p_r p_d + p_s}\left[1 - e^{i\ln(1-p_r p_d - p_s)}\right]. \tag{5}$$

Now, define the parameter, $\xi \equiv -\ln(1 - p_r p_d - p_s) \simeq p_r p_d + p_s$ if all of the probabilities are small. Also, assume that glances occur at a constant rate (and are of a constant duration) so that glance number i may be replaced by a glance rate g times time. This changes equation 5 to

$$P_d(t) \simeq \frac{p_r p_d}{p_r p_d + p_s}\left[1 - e^{-\xi g t}\right]. \tag{6}$$

29.4 Engineering Search-Detection Models

We now turn to engineering search and detection models. We shall cover two such: the Institute for Defense Analysis/Night Vision Laboratory (IDA/NVL) model (also known as the Center for Night Vision and Electro-Optics (CNVEO)) model; and the Radio Corporation of America (RCA) model.

29.4.1 IDA/NVL Model

The IDA/ NVL model is a simple random search model that has non-unitary infinite value given by the Johnson criteria formalism described above. The time dependent detection uses the empirical approximation,

$$P_d(t) \simeq P_{\text{det}}\left[1 - \exp\left(-\frac{P_{\text{det}} t}{3.4}\right)\right], \tag{7}$$

where: $P_{\text{det}} = P_{sc}\,|_{sc=\text{det}}$, as described in the previous chapter. [4] The factor $\frac{P_{\text{det}}}{3.4}(\sim g p_r p_d)$ is the empirical part. This model does not explicitly consider the search process, specifically, the probability of looking at the target.

This model may be compared to our simple search with detection and stopping model. If we equate the parts of equations 6 and 7, then we get

$$\frac{p_r p_d}{p_r p_d + p_s} \simeq P_{\text{det}}, \tag{8}$$

and

$$(p_r p_d + p_s)\, g \simeq \frac{P_{\text{det}}}{3.4}. \tag{9}$$

A bit of algebra gives

$$p_r p_d \simeq \frac{P_{\text{det}}^2}{11.3}, \tag{10}$$

where we have substituted $g \simeq (0.3\,\text{sec})^{-1}$. The lack of a "looking-at" term can be clearly seen from this equation. If $p_r \sim 0.02 - 0.05$, then calculating p_d from this equation can easily result in $p_d > 1$ for $P_{\text{det}} \sim 1$.

29.4.2 RCA Model

The RCA model is also a simple search model, but it arises directly from search theory with a very simple single glimpse probability of detection.[1] The RCA Handbook gives the model as

$$P_d(t) \simeq P_c(C)\left[1 - \exp\left(-\frac{700}{G}\frac{a_t}{A}t\right)\right], \tag{11}$$

where:

$P_d(C)$ = infinite time probability of detection, previous chapter;

G = an image congestion factor, on $[1, 10]$,

a_t = target area; and

A = search area.

Lloyd [2] gives an equivalent form,

$$P_d(t) \simeq P_{\text{det}|look}\left[1 - \exp\left(-\frac{6.9}{K}\frac{\Omega_t}{\Omega_s}t\right)\right], \tag{12}$$

where:

$P_{\text{det}|look}$ = infinite time probability of detection given a look, essentially a Johnson criteria formalism,

K = "an empirical clutter factor proportional to the density of false targets in the scene", on $[0.01, 0.1]$;

Ω_t = target solid angle; and

Ω_s = search solid angle, nominally FOV.

These equivalent models give expected search times of

$$\langle t \rangle_{search} \simeq \frac{K}{6.9}\frac{\Omega_s}{\Omega_t} \simeq \frac{G}{700}\frac{A}{a_t}. \tag{13}$$

We may rewrite this using the glance solid angle ω as

$$\langle t \rangle_{search} \simeq \frac{K}{6.9}\frac{\Omega_s}{\omega}\frac{\omega}{\Omega_t}. \tag{14}$$

Then we may view the factor $\frac{6.9}{K}$ as the efficiency of search-detection, so its inverse is the number of times the target must be looked at for detection; $\frac{\omega}{\Omega_s}$ as the probability of looking at the target, so its inverse is the expected number of search subspaces that must be examined to examine the

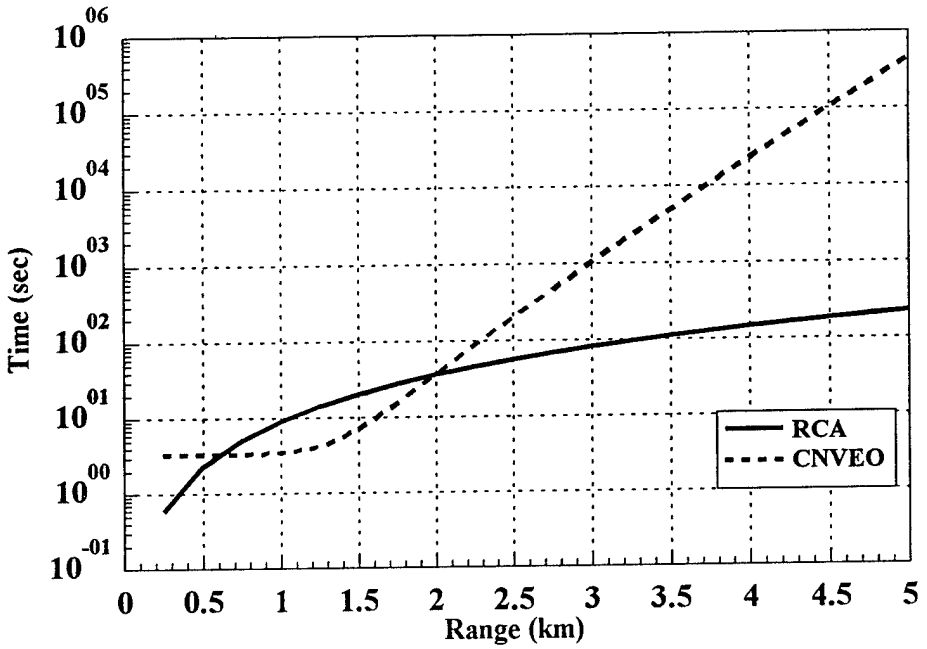


Figure 1: Expected Times to Detect - IDA/NVL and RCA Models

target's subspace; and $\frac{\Omega_t}{\omega}$ is the fraction of the subspace that is target. This latter part represents the functionality of single glimpse detection, albeit in an exceedingly simple manner that we know to be experimentally invalid under general circumstances.

29.4.3 Comparison

Both the IDA/NVL and RCA models are "based" on simple random search with detection, although both have been "fixed up" to have non-unitary infinite time probabilities of detection. Beyond this, the two models are considerably different: the IDA/NVL model grafting on the search parameters empirically while the RCA model uses a very simple single glimpse probability of detection. Accordingly, these two models have very different expected times to detect (stop). For our standard example of a $2.67m$ x $2.67m$ target with inherent contrast of 0.2, we have plotted the expected time to detect in figure 1. Note that the vertical axis, time, is plotted logarithmically. For ranges less than $2km$, there is reasonable agreement between the two times, but they diverge rapidly beyond that range. One reason for this is the more rapid truncation of the normal integral approximation used in the IDA/NVL model (compared to that of Appendix F of Part I.) Whether this corresponds to reality is not certain. Given this vast difference, the analyst is commonly left with uncertainty on how to model detection times at long ranges.

29.5 And Even Less Simple

Human search observed in the "wild" seems to be a mixture of scanned search, random search, and excursion search. In fact, the whole thing is pretty similar to how my wife shops for groceries. The grocery store, of course, is arranged in parallel aisles with shelves on either side. In general, the aisles are wide enough so that two way grocery buggy traffic is feasible (although traffic jams do occur and these may be some analog of false targets that get rejected?) Pushing the grocery buggy up and down the aisles is, of course, nothing more than a raster scan. If we did grocery shopping with two people, one who pushed the buggy at a constant speed, and one who jumps from side to side grasping food (etc.) packages and throwing them into the buggy - without changing its speed, then grocery shopping would be a simple raster scan process.

That's not how my wife shops though. She pushes the buggy down (up) the aisle, moving rapidly through areas whose shelves without interesting groceries, slower through areas where there are occasional interesting groceries that she can grab on the fly without altering her pace. Also, there are two interesting deviations: first, when she gets into an area with a high density of interesting stuff, or where she needs to make a decision on what (if) to get, she tends to stop in the "middle" of the area and walk from the buggy to the shelves and back again with stuff, sometimes five or six times. Once she has depleted the area of her listed items, she then moves on. The second deviation occurs when she remembers something important that didn't get onto her list. In general, she knows exactly where in the store this item is, so she can go immediately to it. Having gotten the item, she either returns to her departure point, or to the other end of the store, scanning backwards, whichever is closer.

This shopping process is an analog of human search. It consists of smooth trajectories, called *saccades*. These can be long (uninteresting objects,) or short (low density of interesting objects). When a highly interesting object is encountered, search enter a *fixation*, A fixation is not a steady gaze but a repeated process of looking-at, looking-away-from the fixation point, rapidly and many times. Under some circumstances, the searcher may spontaneously return to a previously examined fixation point. Thus, the combination of random and scanned search presents an approximation of human search.

This model can be translated into a mathematical model, and has, but we cannot reproduce it here because of information controls.[5]

In addition, we have used very simple models for several parameters, such as contrast threshold and glimpse time. These parameters have greater functionality or even stochastic distributions. We have not reproduced these here for two reasons: first, they don't contribute significantly to understanding the processes; and secondly, they are seldom used in engineering analyses. Nonetheless, they are available for the student who needs them.[6]

29.6 Why All This Fuss?

By now, the reader is likely wondering why we are making all this fuss, anyway? We have

multiple models for getting cumulative probability of detection and expected time to detection, (and presumably other moments as well,) so what else do we need? Why worry about the single glimpse probability of detection anyway?

29.6.1 Moving Targets

Part of the answer is that the detection theory models are essentially static models, so they don't really give any insight into detecting moving targets except by averaging the apparent size of the target over the glimpse time. Human beings are pretty good, courtesy of our hunter firmware, at detecting moving targets. Actually, we are better at detecting moving targets than static targets. Still, neglecting that improvement in detectability, we now examine the cumulative detection probability of a moving target.

To do this, we make a couple of assumptions and simplifications. First, we restrict the motion of the target to seem to be at a constant speed and that the target remain in the FOV. Second, we assume that we know the single glimpse probability of detection, $p_d(r)$, where r = range. Since we really do not care where the target is in the image (in this model,) as long as the background remains consistent, the trajectory is simply

$$r(t) = r_0 - \rho t, \tag{15}$$

where: r_0 = initial range when search begins, or the LOS begins, etc; and ρ = apparent speed of the target (component of velocity in the direction of the sensor.)

For simple random search with detection, we can write a pdf as

$$p \simeq p_r p_t g e^{-p_r p_t g t}, \tag{16}$$

where: g = glimpse rate. If we use the time and range dependent single glimpse probability of detection, then the cumulative probability of detection is just

$$P_d(t) = p_r g \int_0^t p_d(r_0 - \rho t) \exp\left[-p_r g p_d(r_0 - \rho t)\, t\right] dt. \tag{17}$$

Since we do not know the explicit form of the single glimpse probability of detection, we cannot discuss evaluation of this integral except to speculate that if the static probabilities are any indication, we may expect this integral to be exceedingly difficult to evaluate. If the speed is sufficiently slow, then we may expect that a first order approximation of the single glimpse probability may be acceptable. In this case,

$$\begin{aligned} p_d(r_0 - \rho t) &\simeq p_d(r_0) - \frac{dp_d(r)}{dr}\Big|_{r=r_0} \rho t, \\ &\equiv p_{d0} + q_{d0} t, \end{aligned} \tag{18}$$

where the change in sign in the definition results because we expect the space derivative of the single glimpse probability to be negative.

Combining these two equations, we get

$$P_d(t) = p_r g \int_0^t (p_{d0} + q_{d0} t) \exp\left[-p_r g (p_{d0} + q_{d0} t) t\right] dt, \tag{19}$$

which, while difficult, is readily approximable using techniques we have already demonstrated.

29.6.2 Multiple Targets

Another part of the answer to why we are interested in the single glimpse probability of detection is the detection of multiple targets. To set the stage for considering this problem, we shall start with the simplest case possible.

Assume that there are m targets in the FOV,which consists of N search subspaces, and that all of these targets are identical - thus they have identical single glimpse probabilities of detection - p_d. Further assume that the targets are sufficiently far apart than no search subspace contains more than one target.[1]

Let us now consider a state space of $m+1$ states, representing the number of targets that have been detected. Designate by $T_i(t)$ the probability that by time t, i targets have been detected. These probabilities constitute a simple Markov system with defining differential equations,

$$\begin{aligned} \frac{dT_0}{dt} &= -\zeta_0 T_0, \\ \frac{dT_i}{dt} &= \zeta_{i-1} T_{i-1} - \zeta_i T_i, 0 < i < m \\ \frac{dT_m}{dt} &= \zeta_{m-1} T_{m-1}, \end{aligned} \tag{20}$$

where: $\zeta_i \equiv \frac{m-i}{N} p_d g$, and g = the glimpse rate. These equations are almost a Poisson system - they would be if all the ζ_i were identical. Nonetheless, their interpretation and solution method are essentially the same as we have used previously for Poisson systems. Basically, the interpretation is that the no detection state is depleted at a rate ζ_0, which is just the product of the probability of looking at a target, per glance, times the single glance probability of detection, times the glance rate. This decrease in no detection state probability is added to the one-detection state, which is also depleted by the rate of two-detections. This process is repeated until we reach the m-detection state, which cannot be depleted.

There are several further assumptions in this model. First, it assumes that the observer can keep up with the targets that have been detected, and spends no time doing so. Since the image is static, this assumes that the subspaces containing detected targets are subsequently not searched, with no time penalty, so the number of targets to be found is decreased by one. The number

[1]This is not a show stopper, but it does have two effects. First, and more importantly, it maintains the assumption of identical single glimpse probabilities of detection. Second, it assures that there really are m targets to be seen, and not, for example, $m-2$ single targets and 1 double target.

of subspaces is not decremented by one, assuming $N \gg m$. (As we shall see, this neglect is necessary to simplify the mathematics.)

We may immediately solve for the no-detection state probability as

$$T_0(t) = \exp(-\zeta_0 t), \tag{21}$$

the other states are somewhat more complicated. To obtain their solutions, we introduce the same solution assumption as used in the Poisson system,

$$T_i(t) = S_i(t) \exp(-\zeta_i t). \tag{22}$$

This assumed solution has derivative

$$\frac{dT_i}{dt} = -\zeta_i T_i + \exp(-\zeta_i t) \frac{dS_i}{dt}. \tag{23}$$

Substituting this into the rest of equation 20 yields the general condition equation, (after a bit of cancellation)

$$\exp(-\zeta_i t) \frac{dS_i}{dt} = \zeta_{i-1} \exp(-\zeta_{i-1} t) S_{i-1}, \tag{24}$$

which we may rearrange as

$$\frac{dS_i}{dt} = \zeta_{i-1} \exp[-(\zeta_{i-1} - \zeta_i) t] S_{i-1}. \tag{25}$$

By the definition of the rate coefficients,

$$\begin{aligned} \zeta_{i-1} - \zeta_i &= \frac{m-i+1}{N} p_d g - \frac{m-i}{N} p_d g \\ &= \frac{1}{N} p_d g \equiv \zeta, \end{aligned} \tag{26}$$

which is a constant which is equal to the detection rate for a single target (and to the rate coefficient for the $m-1$-detection state.) This important result simplifies equation 25,

$$\frac{dS_i}{dt} = \zeta_{i-1} \exp[-\zeta t] S_{i-1}, \tag{27}$$

which we may immediately rewrite as an integral equation,

$$S_i(t) = \zeta_{i-1} \int_0^t e^{-\zeta t} S_{i-1}(t)\, dt. \tag{28}$$

Since $S_0(t) = 1$ identically by equation 21, we may write the one-detection state reduced solution as

$$S_1(t) = \zeta_0 \int_0^t e^{-\zeta t_0} dt_0. \tag{29}$$

Similarly, the two-detection reduced solution is

$$S_2(t) = \zeta_1 \int_0^t e^{-\zeta t_1} S_1(t_1)\, dt_1, \tag{30}$$

which by use of equation 29 is

$$S_2(t) = \zeta_0 \zeta_1 \int_0^t e^{-\zeta t_1} dt_1 \int_0^{t_1} e^{-\zeta t_0} dt_0. \tag{31}$$

From this, we may infer the general reduced solution form

$$S_i(t) = \left[\prod_{j=0}^{i-1} \zeta_j\right] \int_0^t e^{-\zeta t_{i-1}} dt_{i-1} .. \int_0^{t_1} e^{-\zeta t_0} dt_0. \tag{32}$$

If we now make a change of variables to

$$y = e^{-\zeta t}, \tag{33}$$

then equation 32 becomes

$$S_i(t) = \left(-\frac{1}{\zeta}\right)^i \left[\prod_{j=0}^{i-1} \zeta_j\right] \int_1^y dy_{i-1} .. \int_1^{y_1} dy_0. \tag{34}$$

We may now make use of an integral substitution [7]

$$\int_a^x \int_a^{x_n} .. \int_a^{x_3} \int_a^{x_2} f(x_1)\, dx_1 dx_2 .. dx_{n-1} dx_n = \frac{1}{(n-1)!} \int_a^x (x-z)^{n-1} f(z)\, dz. \tag{35}$$

This changes equation 34 to

$$S_i(t) = \left(-\frac{1}{\zeta}\right)^i \left[\prod_{j=0}^{i-1} \zeta_j\right] \frac{1}{(i-1)!} \int_1^y (y-z)^{i-1} dy, \tag{36}$$

which by a second change of variables $x = y - z$, reduces to

$$\begin{aligned} S_i(t) &= \left(-\frac{1}{\zeta}\right)^i \left[\prod_{j=0}^{i-1} \zeta_j\right] \frac{1}{(i-1)!} \int_0^{y-1} x^{i-1} dx \\ &= \left(-\frac{1}{\zeta}\right)^i \left[\prod_{j=0}^{i-1} \zeta_j\right] \frac{1}{(i-1)!} \frac{x^i}{i} \Big|_{x=0}^{x=y-1} \end{aligned} \tag{37}$$

$$= \left(\frac{1}{\zeta}\right)^i \left[\prod_{j=0}^{i-1} \zeta_j\right] \frac{(1-y)^i}{i!}$$
$$= \left(\frac{1}{\zeta}\right)^i \left[\prod_{j=0}^{i-1} \zeta_j\right] \frac{\left(1-e^{-\zeta t}\right)^i}{i!}.$$

If we now use the definitions of ζ and ζ_i, then we may rewrite this as

$$S_i(t) = \left(\frac{N}{p_d g}\right)^i \left[\prod_{j=0}^{i-1} \frac{m-j}{N} p_d g\right] \frac{\left(1-e^{-\zeta t}\right)^i}{i!}, \tag{38}$$

which we may immediately further reduce to a final form,

$$S_i(t) = \binom{m}{i} \left(1-e^{-\zeta t}\right)^i, 0 < i < m. \tag{39}$$

This gives a set of solutions of

$$\begin{aligned} T_0(t) &= \exp(-\zeta_0 t), \\ T_i(t) &= e^{-\zeta_i t} \binom{m}{i} \left(1-e^{-\zeta t}\right)^i, 0 < i < m, \\ T_m(t) &= \zeta_{m-1} \int_0^t T_{m-1}(t')\, dt'. \end{aligned} \tag{40}$$

(In practice, the easy way to do calculations is to explicitly calculate T_0-T_{m-1}, sum them, and subtract from 1 to get T_m or even do a brute force numeric integration.) The expected number of targets detected may be calculated in the usual manner,

$$\langle i \rangle = \sum_{i=1}^{m} i T_i, \tag{41}$$

which is not obviously summable analytically (unlike a Poisson system.)

We present an example of this in figure 2 using $N = 50$, $p_d = 0.1$, and $g = \frac{1}{0.3\,\mathrm{sec}}$, and $m = 3$ targets.. The characteristic behavior of these types of curves is displayed, especially for the one-detection state - the state probability first increases, due to influx of probability density from the no-detection state, and then decreases, due to outflux of probability density to the two-detection state. In this figure, the two-detection state probability has just peaked, while the three-detection state is still growing. Note the steady, smooth increase in expected number of targets detected.

Now, what about the situation where the targets have different characteristics? That is, instead of all the targets having a common single glimpse probability of detection, let them have individual single glimpse probabilities of detection p_{di}, $i = 1..m$. How do we calculate the cumulative probability of detection over time?

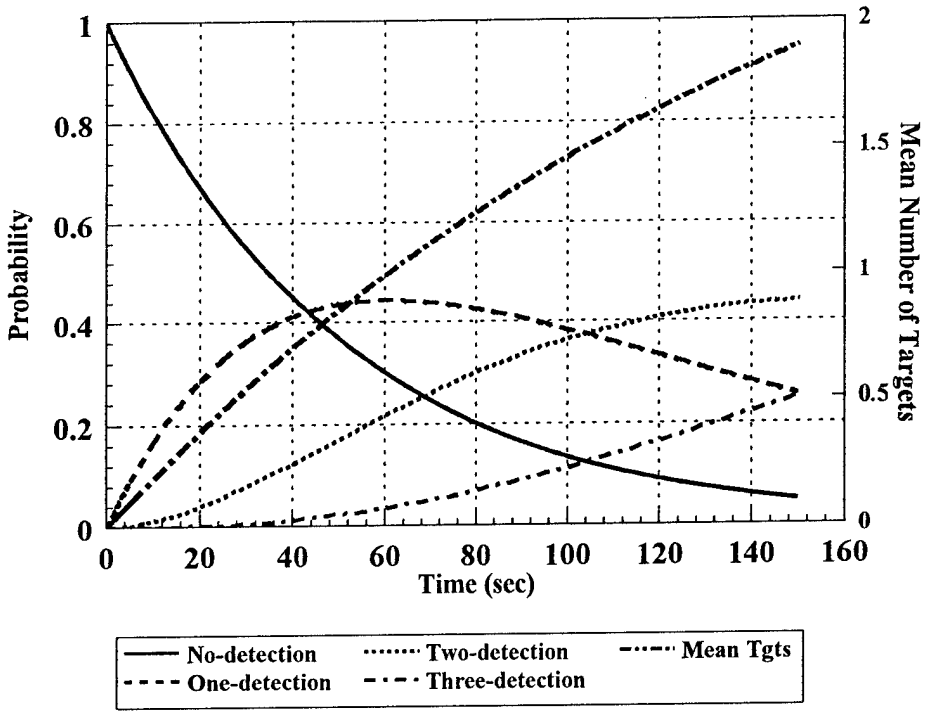


Figure 2: Multiple Target Detection Example - Three Targets

There is no reason why we couldn't develop this problem as a Markov system, establishing a set of states $\{0\}$, $\{i\}$, $\{i,j\}$, $\{i,j,k\}$, which designate no-detection, one-detection, two-detection, three-detection states, respectively. Assuming that we are not interested in the order of detection, then we may establish a convention that $k < j < i$, etc. Even with this convention (restriction,) this results in a very large number of states: 1 no-detection; m one-detection; $\frac{m(m-1)}{2}$ two detection; and in general for the i-detection states, $\binom{m}{i}$. Thus, for our three target example above, there would be a total of eight states!

Having established this system of states, the next step would be to write the evolution differential equations and solve them. The geometry of these states is shown for the example above in figure 3, indicated by circles. The probability density fluxes are indicated by arrows which have been annotated by the single glimpse probability of detection that determines the rate.

This geometry defines the evolution equations. If we define $\psi_i \equiv \frac{p_{dig}}{N}$, where $g =$ the glimpse rate, and $N =$ number of search subspaces, then the evolution differential equations are:

$$\frac{d}{dt}P_{\{No\}}(t) = -\sum_{i=1}^{m}\psi_i P_{\{No\}}(t); \tag{42}$$

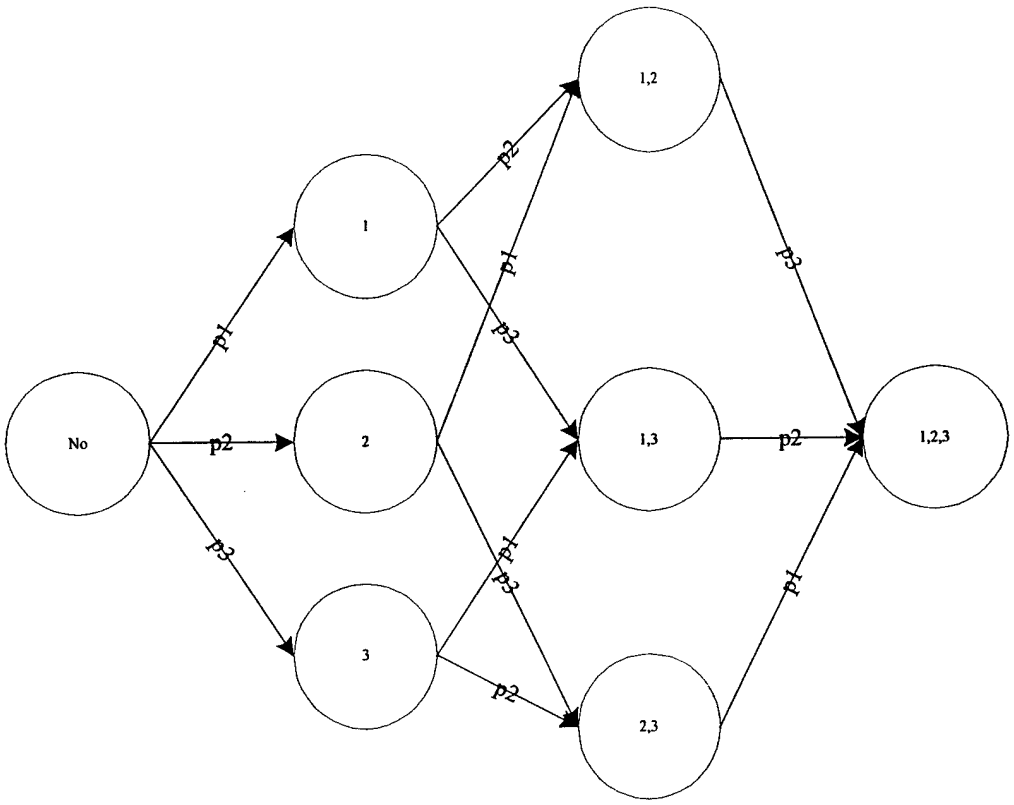


Figure 3: Multiple, Different Target State Geometry - Three Targets

$$\frac{d}{dt}P_{\{i\}}(t) = \psi_i P_{\{No\}}(t) - \sum_{\substack{j=1 \\ j\neq i}}^{m} \psi_j P_{\{i\}}(t);$$

$$\frac{d}{dt}P_{\{i,j\}}(t) = \psi_i P_{\{j\}}(t) + \psi_j P_{\{i\}}(t) - \psi_k P_{\{i,j\}}(t);$$

$$\frac{d}{dt}P_{\{i,j,k\}}(t) = \sum_{\substack{i,j=1 \\ j<i}}^{m} \sum_{\substack{k=1 \\ k\neq i,j}}^{m} \psi_j P_{\{i,j\}}(t);$$

there being one each of the first and last equations, and three each of the middle two. The complexity of this problem is now fairly evident, and while these equations can be solved (since they are linear, Laplace Transforms are a potentially viable approach,) their solution is a more complicated undertaking than we would generally like. Of course, we could always solve them numerically, but again, keeping track of the notation alone makes this a complicated problem. In particular, since the progression from m to $m+1$ targets both increases the number of evolution equations, and is difficult to generalize in writing equations, we would like to find an approximate way of handling this problem.

Since the observer does not know about the targets' identities until they are detected, we may partially aggregate the targets. To do this, we start by defining relative single glimpse probabilities of detection,

$$q_{di} \equiv \frac{p_{di}}{\sum_{j=1}^{m} p_{dj}}, \tag{43}$$

and the mean single glimpse probability of detection

$$\langle p_d \rangle = \frac{\sum_{i=1}^{m} p_{di}}{m}. \tag{44}$$

If we use this mean single glimpse probability of detection in the formalism we previously developed for m identical targets, then we may calculate a set of state probabilities $T_i(t)$, where $i =$ the number of targets detected. Then, we may approximate the state probabilities for the different targets as

$$\begin{aligned} P_{\{No\}}(t) &\simeq T_0(t), \\ P_{\{i\}}(t) &\simeq q_{di} T_1(t), \\ P_{\{i,j\}}(t) &\simeq \left(q_{di} q_{dj}^{\{i\}} + q_{di}^{\{j\}} q_{dj}\right) T_2(t), \end{aligned} \tag{45}$$

and so forth. The quantities $q_{di}^{\{j\}}$ are reduced relative conditional probabilities defined by

$$q_{di}^{\{j\}} \equiv \frac{q_{di}}{\sum_{\substack{k=1 \\ k \neq j}}^{m} q_{dk}}, \tag{46}$$

and we note in passing that if $\{i, j, k, ..m\}$ is the terminating state, then

$$P_{\{i,j,k,..m\}}(t) \simeq T_m(t). \tag{47}$$

29.7 Conclusion

That's enough! As we have already declaimed, search and detection are prolific, intensive, and important fields of study, as evidenced by their profuse literature. We cannot begin to do a respectable job of coverage in one chapter, but we can begin to introduce the subjects and try to convey the idea that the two theoretical bases do not mesh as well as we would like. As a result, various approximations are necessary in the calculations that go into attrition theory.

We shall leave the area of acquisition for now, and turn down a different path, to consider the engagement portion of attrition. Later, the two paths will join when we finally get to consider the basics of attrition itself.

29.8 References

[1] Radio Corporation of America, **Electro-Optics Handbook**, Solid State Division, Electro-Optics and Devices, Lancaster, PA, Technical Series EOH-11, 1974.

[2] Lloyd, J. M., **Thermal Imaging Systems**, Plenum Press, New York, 1975.

[3] Rotman, S. R., E. S. Gordon, and M. K. Kowalczyk, "Modeling human search and target acquisition performance: I. First detection probability in a realistic multitarget scenario", Optical Engineering 28 No. 11, Nov. 1989, pp. 1216-1222.

[4] Rotman, S. R., E. S. Gordon, and M. L. Kowalczyk, "Modeling human search and target acquisition performance: III. Target detection in the presence of obscurants", Optical Engineering 30 No. 6, June 1991, pp. 824-829.

[5] Nicoll, J. F., "A Mathematical Framework for Search: The Neo-Classical Model", Unnumbered IDA paper.

[6] Waldman, Gary, John Wootton, and Greg Hobson, "Visual Detection with Search: An Empirical Model", IEEE Transactions on Systems, Man, and Cybernetics 21 No. 3, May/June 1991, pp. 596-606.

[7] Hildebrand, Francis B., **Methods of Applied Mathematics**, Prentice-Hall, INC., Englewood Cliffs, NJ, 1965, pp. 224-225.

This Page Intentionally Left Blank

Chapter 30

Weapons I

30.1 Introduction

We now come to consideration of the engagement part of the attrition process. The fundamental componentry of engagement are weapons. No other single class of material products of civilization enjoy such an emotional relationship as do weapons. They generate fierce hatred and fear among many elements of the non-military community as the iconology of war itself. Even for those who are not actively averse to war, weaponry generates intense interest, as evidenced by the continued success of museum collections and public displays such as parades and exhibits on the Fourth of July and Armed Forces Day in the United States. Soldiers display the fundamental respect of the workman for his tools in his relationship with these instruments of policy continuance. Only the material accoutrements of religious observation and rite have the care of manufacture and (in olden days,) decoration that is lavished on weaponry.

This robustness and durability is, of course, a direct consequence of military operations. The success orientation of military operations, the importance of obedience and determination that is often the crucial determiner of victory, carries over into the manufactory and maintenance of weaponry. Good soldiers obey intelligently and care for their weapons in a utilitarian manner; bad soldiers are bad because they either obey unthinkingly and irresponsibly or do not obey at all; their weapons either rust or become useless decorations.

It has been argued that war and preparations for war are the generator of technology. Others argue that technology is the generator of war. Regardless, the military establishment has long maintained a love-hate relationship with technology. On the one hand, it can provide the basis for superior capability; on the other, changes in technology engender changes in organization, tactics, and doctrine, which are detrimental to capability. The middle ground is held by the concern that the enemy will adopt technology so that technology adoption maintains defensive parity. Thus, there exists a fundamental dilemma that cannot be resolved at either extreme. There must be a balance between today's capability, which is enhanced by firm doctrine, organization, and tactics and hurt by technology, and tomorrow's capability, which is enhanced by technology

and hurt by rigid doctrine, organization, and tactics.

30.2 Weapons and Physics

If the physics of acquisition was primarily optics, then the physics of engagement is primarily mechanics. The fundamental effective principle of weaponry (in the large) has not changed since the invention (discovery?) of the combat rock by Og the Cave Man in the year Real Long Time Ago. With the exception of certain weapons, notably Chemical, Biological, and Psychological in nature, the mechanics of weaponry has been the transfer of momentum to the target.

30.2.1 Human Powered Weapons

It seems reasonable that the original weapons were those that man was endowed with as part of his physiognomy - hands and teeth. Given his natural aptitude as a tool user however, the progression to fighting tools seems both natural and rapid. While it is unclear whether the original non-corporate weapon was the rock or the club, there being plentiful evidence of rock artifacts but little of wooden ones, the advantages of both over teeth and fist are many: damage is not done to one's own body; greater range is achieved (for the club); greater angular momentum and thereby, greater momentum transfer is possible; and the rigidity is increased and striking area decreased. Thus, the adoption of the rock or club both obeys Patton's dictum ("Let the enemy die for his tribe."), and increases range and momentum transfer. These latter two quantities (along with accuracy) summarize the bulk of weapons evolution.

The soft tissues of the human body can be penetrated with an energy density of about 0.4 Joule cm^{-2}(2 lb-ft in^{-2}).[1] If we consider a (very!) simple collision between a mass m with initial velocity (speed) v with an effective target mass of m', then the energy before the collision is

$$E_b = \frac{mv^2}{2}. \tag{1}$$

If the weapon mass rebounds (as with a club,) then assuming a final velocity of v', the energy imparted to the target is

$$E_t = \frac{m'\left(v - v'\right)^2}{2}, \tag{2}$$

and if we assume the striking area to be a, then the energy density transferred is just

$$\epsilon_t = \frac{m'\left(v - v'\right)^2}{2a}. \tag{3}$$

Let us further assume that $a \sim 20cm^2$, which is a fairly reasonable area for a war club, and that $m' \sim 5kg$. If the rebound speed is half the striking speed, then penetrating damage can occur

for $v > 3.6m\,\mathrm{sec}^{-1}$. This level of speed can be achieved by a healthy human being with some exertion.

The next steps in the evolution of weapons probably were the combination of the club and the rock into the axe, thus combining the added range and angular momentum of the club with the rigidity and smaller striking area of the rock, and the extension of the idea to the lance or pike (unthrown spear),[1] thus giving even greater range, accuracy, and an even smaller striking area. A constant factor of weapons technology (and military evolution) seems to be the relaxation of control. This is evidenced by the next logical weapons evolutions, the (thrown) spear and the bow and arrow.

Development of weapons projected beyond the grasp of the soldier represents as great a leap as the original development of club or rock. A major change in the physics now comes into play. The situation is actually somewhat simpler (in this very simple model) for an inelastic collision such as would be expected for a bullet. In this case, the energy imparted to the target is

$$E_t = \frac{(m + m')\, v'^2}{2}, \tag{4}$$

where v' is now the final velocity of bullet and target. This velocity has the form,

$$v' = v\sqrt{\frac{m}{m + m'}}, \tag{5}$$

which we do not really need if energy is conserved. In this case, and assuming $a \sim 1cm^{-2}$, and $m \sim 5g$, then penetrating damage is done for $v > 12.6m\,\mathrm{sec}^{-1}$. (These values are small for a spear or arrow, and reasonable for a bullet. The physics does not reflect the effects of drag which naturally limits the range of projected weapons.)

30.2.2 Guns

The natural limitations on range and momentum are the strength of the human arm (primarily), so the next evolutionary developments were the development, essentially simultaneously, of the gun and the rocket. A gun is a tube with one closed and one open end. A chemical that burns rapidly is placed against the closed end and a piece of inert material is placed next to the chemical. The chemical is burned and the resulting gas pressure propels the inert material out of the tube.

There are several trade-offs in the technology here. Clearly, the inert material (hereafter, the bullet,) should fit snugly against the tube so that gas pressure is not wasted. This fit however, should not be too snug, first so that the bullet will move, and second, so that the tube does not experience significant damage from the passage of the bullet. The accuracy of the bullet's trajectory is dependent on the tightness of fit and the length of the tube. Since the bullet has

[1]The distinction between lance and pike here is operational. The lance is carried to the enemy by the soldier, using his own momentum of motion. The pike is trhust at the enemy, using the momentum of the thrust.

some room to move sideways in the tube, this sideways motion (momentum) is a source of inaccuracy that is decreased by lengthening the tube.[2]

The burning of the chemical rapidly produces a quantity of gas at high temperature in an enclosed space, producing intense pressure. Ideally, the chemical should burn over the entire duration of the bullet's flight through the tube, but this is difficult to manage in this geometry. Note that this pressure provides a force on the bullet. Since the bullet is moving down the tube, the volume of gas is increasing and thus the pressure decreases. (This is another reason why we want the chemical to burn continuously while the bullet is in the tube - we want a constant pressure.) The trade-off here is that to get enough pressure to get enough velocity, the catastrophic pressure threshold of the tube material must be exceeded (usually.) The trick is to keep the overload of such a short duration that the tube does not fail. There are two regimes of this failure, the immediate failure due to the first generation of pressure, and the long term failure due to prolonged high pressure while the bullet is still in the tube. This latter, plus bending effects, tends to limit the length of gun tubes (a practical engineering limit for military use is about fifty times the diameter of the bullet.)

We may build a simple model of this. Assume that the tube and bullet constitute an isothermal (constant temperature) system, and that the pressure of the gas may be expressed using the ideal gas law,[2]

$$PV = nRT, \tag{6}$$

where:

P = pressure,

V = volume,

n = number of moles of gas,

R = ideal gas constant, and

T = absolute temperature.

Assume the tube (and the bullet) to have a cross sectional area a, and let x be the distance traveled down the tube. Since the system is assumed to be isothermal, T =constant, so we may scale pressure-volume. Since the volume of the gas is just the cross sectional area times the distance traveled,

$$V = ax, \tag{7}$$

the pressure scales as

$$P = P_0 \frac{x_0}{x}, \tag{8}$$

[2]If we think (crudely) of the bullet flying down the tube and bouncing off its walls, with a slight loss of transverse momentum with each bounce, then the transverse momentum is minimized by maximizing the number of bounces. Of course, the preferred solution is no transverse momentum (and no bounces.)

where: P_0 = initial pressure, and x_0 = (small) initial bullet displacement. Now we may develop the equations of motion. Assume that the mass of the bullet is m, and that the tube is rigidly mounted so it does not move. Then by the definition of pressure (= force per area,), and Newton's second law of motion,

$$\begin{aligned} F &= Pa \\ &= m\frac{d^2x}{dt^2}. \end{aligned} \tag{9}$$

Combining these last two equations with a bit of rearrangement yields

$$\frac{d^2x}{dt^2} = \frac{P_0 a x_0}{mx}, \tag{10}$$

which is the equation of motion of the bullet. This differential equation is difficult to solve analytically,[3] but can be integrated numerically using a simple spreadsheet simulation. We reproduce a sample calculation (in arbitrary units) in figure 1. Several behaviors may be noted. Since we have treated the burning as instantaneous, there is no build-up of pressure as there would be in a real gun; therefore, the pressure decays essentially exponentially. Velocity increases essentially logarithmically, as we would expect from equation 8 (actually it is square root of log of x.) Distance does not quite increase quadratically (in fact, becomes essentially linear with time over long distances.)

30.2.3 Rockets

Now we want to consider a rocket. Basically, a rocket is a tube, closed at one end, open at the other. Part (or all) of the tube is filled with a slower burning chemical. A shaped ring of inert material is placed in the open end to shape the gas flow. The gasses from the burning chemical propel the tube.

There are also obvious trade-offs here as well. Since we are flying the whole tube rather than just a bullet, we want to make the tube to be as light as possible and still withstand the pressure of the gas. Since the chemical burns slower and exhausts the gas continuously, the pressure is happily much lower. Unfortunately, when we lighten the tube, we reduce the momentum of the rocket. In fact, this is another trade-off, weight (mass) versus speed. Normally, the need to make the rocket tube as light as possible forces us to use some other means of conveying momentum rather than through the tube itself. Thus, the need for the warhead is born!

There is also an accuracy problem with rockets since they do not have the long tube that a bullet has to initially direct them. For this reason, it becomes desirable for a rocket motor to

[3] Actually, the velocity is easy to obtain as $v = \sqrt{\frac{2P_0 a x_0}{m} \ln\left(\frac{x}{x_0}\right)}$ by simply changing the variable of integration from t to v. Integrating the inverse of this wrt x is an error function of imaginary argument, so recourse to a numerical approach is inescapable. Thus, having been forced to use the numerical hammer, we use it earlier and easier.

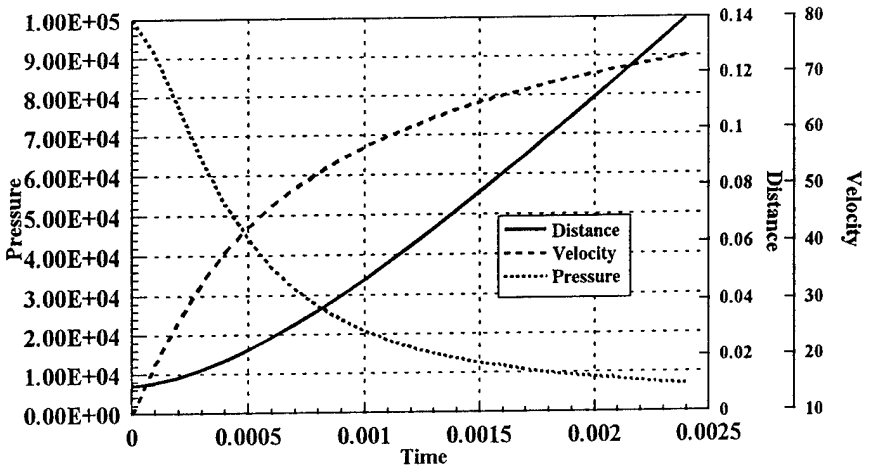


Figure 1: Sample Bullet Motion

burn fairly fast and to place the rocket on a light weight launcher that will initially direct it. (People are terrible launch platforms - more below.) Thus, we would like the motor to burn only while the rocket is on the launcher, just as we wanted for a gun, but now this is realizable. In fact, this is how most human launched rockets are designed today.

We may also build a simple model of a rocket. Let the initial mass of the rocket be

$$m_r = m_t + m_c, \tag{11}$$

where: m_r = mass of the rocket, m_t = mass of tube, and m_c = mass of chemical (propellant.) The ratio

$$f_p \equiv \frac{m_c}{m_r}, \tag{12}$$

is called the propellant mass fraction (although this is usually only referred to the mass of the rocket motor in modern usage,) and the propellant only burns for a time τ so that if the propellant is burned at a constant rate, that rate is

$$\mu \equiv \frac{m_c}{\tau}. \tag{13}$$

Then as a function of time, the mass of the rocket is

$$\begin{aligned} m_r(t) &= m_r - \mu t, t \leq \tau, \\ &= m_r, t \geq \tau. \end{aligned} \tag{14}$$

The force on the rocket due to the exhaust of the burning gas is

$$F = \mu(v_e - v), \tag{15}$$

where: v_e = the exhaust velocity of the gas. This just says that the force exerted is equal to the mass per unit time exhausted times the relative velocity of the gas. This is a simple statement of force as the time rate of change of momentum. The force on the rocket can also be written as

$$F = \frac{d}{dt} m_r(t) v, \tag{16}$$

so that after combining these two equations, and doing a bit of calculus, we get

$$-\mu v + m_r(t) \frac{d}{dt} v = \mu(v_e - v). \tag{17}$$

This reduces after a bit of cancellation and rearrangement to

$$\frac{d}{dt} v = \frac{\mu v_e}{m_r - \mu t}, t \leq \tau, \tag{18}$$

which we may immediately recognize as a logarithm, and integrate as

$$v(t) = v_e \ln\left(\frac{m_r}{m_r - \mu t}\right), t \leq \tau. \tag{19}$$

For $t > \tau$, the velocity has the value

$$v(t) = v_e \ln\left(\frac{m_r}{m_t}\right). \tag{20}$$

We again come to a nasty problem, but one that is a bit more tractable. We may rewrite equation 19 as

$$\frac{dx}{dt} = -v_e \ln\left(\frac{m_r - \mu t}{m_r}\right), t \leq \tau, \tag{21}$$

which may be integrated using [3]

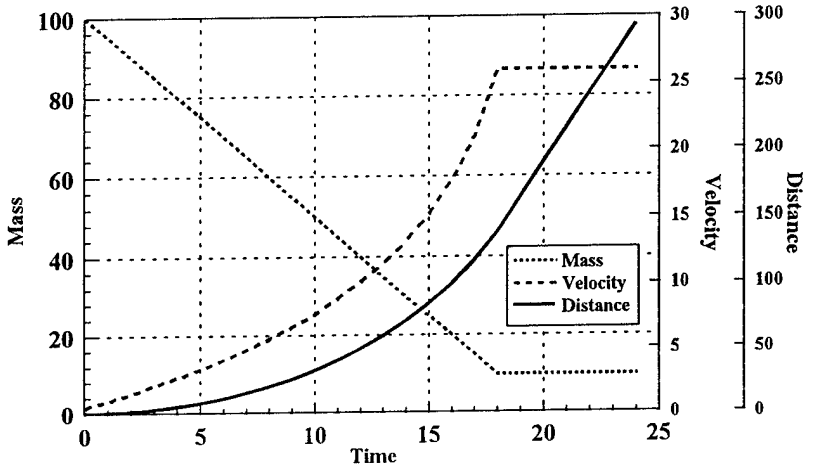


Figure 2: Sample Rocket Motion

$$\int \ln(az+b)\,dz = \frac{1}{a}(az+b)\ln(az+b) - z. \tag{22}$$

Since we may readily have $x = 0$ at $t = 0$, equation 21 becomes

$$x(t) = \frac{v_e m_r}{\mu}\left(1 - \frac{\mu}{m_r}t\right)\ln\left(1 - \frac{\mu}{m_r}t\right) + v_e t, t \leq \tau. \tag{23}$$

As in the tube-bullet example, we can always integrate this numerically using a simple spreadsheet simulation. A sample of this is shown in figure 2. In this example, the chemical (propellant) is 90% of the rocket mass. All other values are arbitrary. The linear decrease of the mass is evident. Note the cut-off of velocity increase when the propellant "burns out". Although not obvious, distance becomes linear at this point.

In general, engineers designate three types of mechanics, called ballistics, for bullets and rockets:

- interior ballistics which deals with motion in the tube/gun/launcher;

- transition ballistics which deals with the exit from the tube/gun/launcher; and
- exterior ballistics which deals with motion outside the tube/gun/launcher.

We shall primarily deal with exterior ballistics.

30.2.4 The See-Saw: Weapons and Armor

Before proceeding to greater consideration of the mechanics and operation of weapon systems, there are a few more evolutionary developments that we need to discuss. The most obvious of these are armor and warheads. Armor is a natural result of observing the fragility of the human body to puncture. The idea is to place some inert material on the target that is sufficiently light weight not to drastically reduce performance and mobility, but tough enough to reduce (even eliminate) the likelihood of puncture. In effect, armor raises the threshold for momentum transfer.

If this were all that it did, then armor would be a simple matter, but it has a more widespread military effect. As one might expect, the technologies of momentum transfer and momentum transfer denial (weapons and armor) do not march in lock step. Therefore, there tend to be period of time when a new weapons technology is introduced and armor is ineffective (bows and arrows at Agincourt), and periods when a new armor technology is introduced and weapons are ineffective (trenches in World War I.) There is a general correlation: when armor is more effective than weaponry, defensive operations tend to be favored over offensive; when weaponry is more effective than armor, offensive operations tend to be favored over defensive. Care must be taken in carrying this correlation too far, since it ignores important considerations such as mobility, logistics, and density of forces. Nonetheless, the relative effectiveness of weapons and armor technologies seem to have a strong influence on the efficacy of military operations.

The other obvious evolutionary advance was warheads. Long range, whether by gun or by rocket dictated a large bullet or rocket, with accompanying burden for crews. The search for loner range thus effectively ended the idea of one man-one (or many) weapons and ushered in the idea of many men - one weapon. It is obviously unrealistic, to say nothing of inefficient, to attack one man (target) with a large weapon (with a large crew), so the idea of putting some variant of the burning chemical in the bullet/rocket to transfer momentum arose. Originally, there were three major implementations of this idea:

- the fragment warhead, which consisted of several small (compared to the bullet/rocket) pieces of inert material (smaller bullets) which were dispensed from the bullet/rocket and relied on the imparted velocity for momentum, should they strike a target;
- the blast warhead which consisted of rapidly burning chemical whose gasses provided an impulsive pressure field to transfer momentum to the targets; and

- the blast-fragment warhead which consisted of a burning chemical, either surrounded by small fragments or encased in a frangible container which burst into small fragments when the chemical was burned.

In practice, these warheads have different effectiveness against different types of targets, depending on the target's vulnerability to momentum transfer. In general, each of these types of warheads is characterized by a lethal volume of effect. Modern warheads have carried on along these lines, refining the form and volume of momentum transfer. These reach extremes with modern High Explosive Anti-Tank (HEAT) and Explosively Formed Penetrator (EFP) warheads which used an explosive (rapidly burning chemical) to either vaporize a heavy metal in a high speed, highly directional jet, or form a bullet and propel it in a highly directional manner. These warheads essentially recapture the idea of one weapon - one target because of their highly limited volume of lethality. This is a direct result of progress in armor technology and the desire for longer range (and accuracy, which is demanded by these warheads.)

30.2.5 Point and Area Weapons

This brings us to a distinction that we have made in Part I, but are now in better shape physically to make the distinction. That distinction is between point and area weapons, which have different interpretations under attrition theory in general, and Lanchester theory in particular. If momentum transfer is directed at a single target in a specific sense, then the weapon is considered to be a point (attack) weapon. While the target may be small or large, the fact that the weapon operates by attacking the target at a point makes it a point weapon. Similarly, if the momentum transfer is directed either at (possibly) several targets in either a specific or a general sense, or against a single target in a general sense, then the weapon is considered to be an area (or volume) weapon.

Actually, most area warheads/weapons are actually volume weapons. That is, their warheads have a lethal effect volume even if they detonate on or under the ground (which changes the boundary conditions and thereby, possibly, the effect.) The human tendency is to visualize things in two dimensions, so it is common to think of weapons'/warheads' effects in terms of a footprint on the ground. Obviously, this does not hold true when we consider their effect against aerial targets.

30.2.6 Weapons that do not transfer momentum

We mentioned earlier that there were some exceptions to momentum transfer. It would be neat to categorically say that Weapons of Mass Destruction are not momentum transfer weapons, but this is not the case.[4] Nuclear weapons have three major lethality mechanisms: pressure, radiation, and heat. Pressure is an area (volume) momentum transfer; radiation is an area (volume) momentum transfer, either in the form of fundamental particles which do damage by

[4]The term Weapon of Mass Destruction is largely a misnomer. Mass cannot be destroyed in the classical sense, but can be converted, by fission or fusion, into energy. This is the basis of nuclear weapons. Chemical and Biological weapons conserve mass. The only way the term makes sense is if destruction is the weapon's effect on people and material, and mass means widespread.

momentum transfer in an essentially classical sense but at the molecular level, or in the form of the electro-magnetic pulse, which transfers the momentum of the wave to the receivers; even heat can be a momentum transfer system in the sense of raising the temperature of the target. Nonetheless, we normally think of the heat component of nuclear weapons as a generator of fire, and thereby not a momentum transfer mechanism. Chemical weapons on the other hand work by causing a chemical reaction in the target. Normally, this means degradation of the human respiratory or nervous system. Similarly, biological weapons also work by degrading the performance of the human body. Both of these types of weapons do not use momentum transfer.

Psychological weapons are intended to degrade performance by degrading the mental or psychological performance of their targets. They also do not operate by momentum transfer.

30.3 Gravity

The force of gravity is at once both the historical genesis of mechanics and one of the two primary forces that shape the projection mechanics of weapons. For most applications that we shall make, the use of a flat earth model is not damning. Occasionally, usually for long range engagements, we shall have to make use of a spherical earth model.

Before discussing gravity, definitions of coordinate systems are in order. The most common systems that we shall use are the spherical and the rectangular (Cartesian). The latter is characterized by three orthogonal axes, commonly labeled x, y, z, and right handed. Vectors are designated by underscored arrows, and components by axis subscripts. Unit vectors are indicated by the symbol $\hat{e}_s$, where the subscript s indicates axis. Thus, in the rectangular coordinate system, a vector $\underrightarrow{v}$ has the form,

$$\underrightarrow{v} = v_x\hat{e}_x + v_y\hat{e}_y + v_z\hat{e}_z, \tag{24}$$

while in the spherical system, it has the form

$$\underrightarrow{v} = v_r\hat{e}_r + v_\theta\hat{e}_\theta + v_\phi\hat{e}_\phi. \tag{25}$$

We shall also use the notation

$$\underrightarrow{v} = \sum_{i=1}^{3} v_i\hat{e}_i. \tag{26}$$

The student should refer to a good introductory text on mathematical physics or classical mechanics for additional details, e.g., Marion.[4] The spherical coordinate system has coordinates r, θ, ϕ. The relationship of the systems is shown in figure 3. The mathematical relationship of the components by

$$r = \sqrt{x^2 + y^2 + z^2}, \tag{27}$$

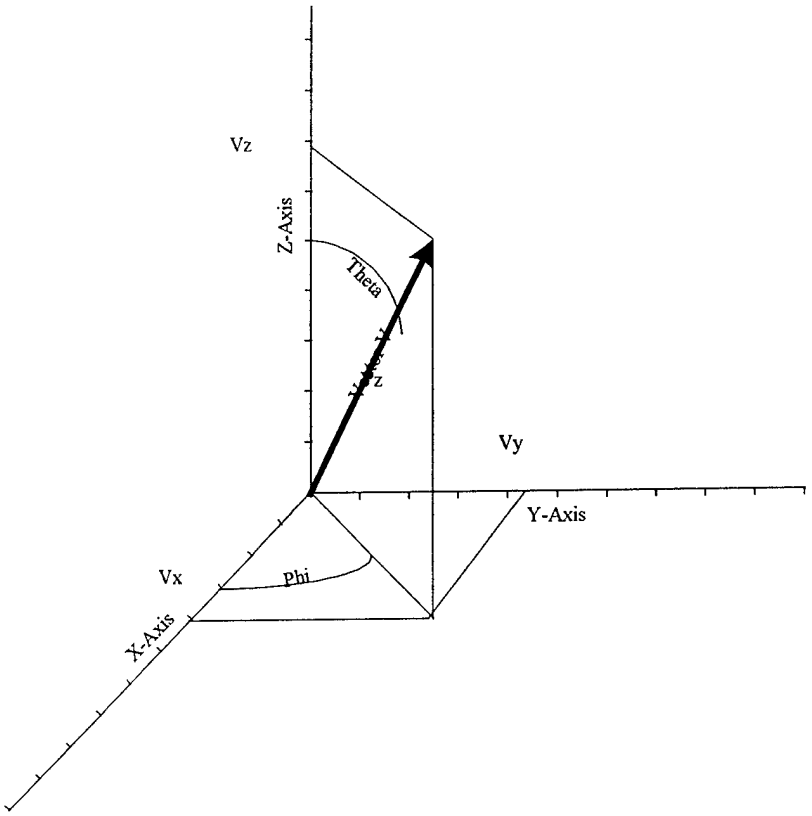


Figure 3: Rectangular and Spherical Coordinate Systems

$$\cos(\theta) = \frac{z}{r},$$
$$\tan(\phi) = \frac{y}{x},$$

and

$$\begin{aligned} x &= r\sin(\theta)\cos(\phi), \\ y &= r\sin(\theta)\sin(\phi), \\ z &= r\cos(\theta). \end{aligned} \tag{28}$$

In spherical coordinates, gravity is a radial force given by

$$F_r = -\frac{MmG}{r^2}, \tag{29}$$

where: M, m = masses of the two interacting bodies, and G = universal gravitational constant. The force is negative under the convention that attractive forces are negative. In general, we

shall normally operate with earth centered spherical coordinate systems, with the z axis through the geographic north pole, and the x axis crossing the Greenwich Meridian.

The more normal coordinate system that we shall use will be an earth surface rectangular system with the z axis positive along increasing elevation. Thus, the z component is the height component. The magnitude in the x-y plane will usually be called the range (except for air defense applications.) In this coordinate system, gravity has the form

$$F_z = -mg, \tag{30}$$

where: g = acceleration due to gravity ($\sim 9.8 m \sec^{-2}$).

Having been through this, it is now worthwhile to review a trivial bullet or rocket (after burn-out) trajectory. We assume short range so that we may use our flat earth system. The Newtonian equations of motion are

$$\begin{aligned} \frac{d^2x}{dt^2} &= 0, \\ \frac{d^2z}{dt^2} &= -mg, \end{aligned} \tag{31}$$

with initial conditions,

$$\begin{aligned} \frac{dx}{dt} \;|\; _{t=0} &= v_x = v\cos(\psi), \\ \frac{dz}{dt} \;|\; _{t=0} &= v_z = v\sin(\psi), \end{aligned} \tag{32}$$

where: ψ = elevation angle. These equations can be integrated trivially, to yield parametric time solutions,

$$\begin{aligned} x(t) &= x_0 + v\cos(\psi)\, t, \\ z(t) &= z_0 + v\sin(\psi)\, t - \frac{gt^2}{2}, \end{aligned} \tag{33}$$

where: x_0, z_0 = the initial position conditions. It is convenient to reduce this to a simple range equation, so we solve the z equation for t,

$$t = \frac{v\sin(\psi) \pm \sqrt{v^2\sin^2(\psi) - 2g\,(z(t) - z_0)}}{g}, \tag{34}$$

where the positive branch is usually the desired one. Since we are interested in impact, $z(t) = 0$. If $z_0 = 0$, then this reduces to

$$t_{impact} = \frac{2v\sin(\psi)}{g}, \tag{35}$$

and the range equation becomes simply

$$\begin{aligned} x(t) &= x_0 + \frac{2v^2\cos(\psi)\sin(\psi)}{g}, \\ &= x_0 + \frac{v^2\sin(2\psi)}{g}. \end{aligned} \tag{36}$$

It is also useful to note the time of maximum elevation, which is simply the time when the upward velocity is zero,

$$t_{z\max} = \frac{v\sin(\psi)}{g}, \tag{37}$$

which gives a maximum elevation of

$$z_{\max} = z_0 + \frac{v^2\sin^2(\psi)}{2g}, \tag{38}$$

and a range of maximum elevation of

$$x_{z\max} = x_0 + \frac{v^2\sin(2\psi)}{2g}. \tag{39}$$

Note that for small angles, these two range equations simplify by using $\sin(2\psi) \simeq 2\psi$. Thus, to hit a target at range $x = r$, the gun/launcher must be elevated at an angle ψ, rather than just pointed at the target. We shall explore range accuracy effects later.

30.4 Drag

Just as the atmosphere extinguishes the intensity of light, it diminishes the momentum of weapons by drag. Drag is simply the resistance of air to the motion of the projectile, its magnitude depends on the velocity of the projectile and its shape. (This is why the conoidal bullet - Minnie ball - provided such an increase in effective range.) Drag has two effects on projectiles. First, it reduces the momentum of the projectile, thus diminishing the amount of momentum available for penetration (at a given range.) Second, by slowing the flight, gravity has more time to act, so that for a given elevation angle, a shorter range is achieved. The force of drag has the mathematical form

$$\underrightarrow{F} = -k\,\underrightarrow{v}\left|\underrightarrow{v}\right|^{n-1}, \tag{40}$$

where the minus sign indicates the force operates against the direction of the projectile's velocity. The parameter k depends on the shape of the projectile, and the value of n depends on its speed. For speeds less than about $24m\,\mathrm{sec}^{-1}$, $n \sim 1$, while for greater speeds, but less than the speed of sound $\sim 330m\,\mathrm{sec}^{-1}$, $n \sim 2$. This generally continues for speeds greater than the speed of sound, but less than about $730m\,\mathrm{sec}^{-1}$, when it appears that $n \sim 1$ again holds.[5]

It is relatively trivial to solve one-dimensional equations of motion with drag, but without gravity. Without gravity, the equation of motion is

$$\frac{dv}{dt} = -\frac{k}{m}v^n, \tag{41}$$

which must be treated separately for $n = 1$, and $n \neq 1$ because of the form of the logarithm. For $n \neq 1$, this readily integrates twice to

$$x(t) = x_0 + \frac{m}{k(n-2)}\left[\left(v_0^{1-n} + \frac{k}{m}(n-1)t\right)^{\frac{n-2}{n-1}} - v_0^{2-n}\right], \tag{42}$$

which seems undefined for $n = 2$, but is really

$$x(t) = x_0 + \frac{m}{k}\ln\left(1 + \frac{kv_0t}{m}\right), n = 2. \tag{43}$$

For $n = 1$, the result is

$$x(t) = x_0 + \frac{mv_0}{k}\left[1 - e^{\frac{kt}{m}}\right]. \tag{44}$$

When we include gravity, the integrals are still doable, albeit we want to limit the values of n to be 1 or 2 exactly. This is left as an exercise for the student for $n = 2$. For $n = 1$, the velocity is

$$v(t) = \left(v_0 + \frac{gm}{k}\right)e^{-\frac{kt}{m}} - \frac{gm}{k}, \tag{45}$$

and the distance is

$$x(t) = x_0 + \left(\frac{v_0m}{k} + g\right)\left(1 - e^{-\frac{kt}{m}}\right) - \frac{gmt}{k}. \tag{46}$$

The noteworthy thing here is that velocity has a terminal value, as $t \to \infty$, of $-\frac{gm}{k}$. This demonstrates the effect of drag in that it can completely counteract gravity. (At least from the acceleration standpoint; you still hit the ground hard! This is why skydiving is possible and why skydivers wear parachutes.)

The effect of drag on speed and distance can be demonstrated using our previous two examples for a bullet and a rocket. These are shown in figures 4 and 5, respectively. The decrease in velocity and distance can readily be seen by comparison with figures 1 and 2, respectively. In both cases, we used $n = 2$ for the drag force even when the velocities were small. This incurs a

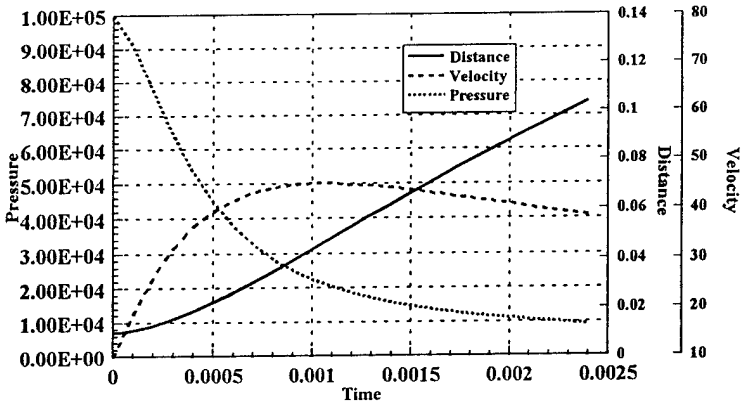


Figure 4: Simple Bullet Motion with Drag

small error that does not detract from the illustration but would be corrected in an engineering calculation with an exact value for k.

As we have already indicated, drag and gravity together act to sharply reduce the range of projectiles. This is illustrated in figures 6 and 7, which depict the trajectories for elevation angles of $15, 30, 45$, and $60°$, with just gravity, and gravity and drag, respectively. The calculations were performed numerically for $n = 2$ using a simple spreadsheet simulation with time as the independent parametric variable. That is why the gravity only curves do not always impact the ground. The differences between impact ranges are dramatic and evident. Note also the asymmetric shapes of the gravity plus drag curves compared to the gravity only curves. Since the projectiles velocity is zero at the top of the curve, drag slows its fall downward, destroying the symmetry (and energy conservation) of the gravity only curves.

In analytical practice, interest is often limited to the accuracy of the projectile, and its range-time-of-flight curve. A example of the latter, using the same data as above is shown in figure 8. These curves depict only the $15°$ elevation trajectories, and are carried to impact. The slower speed and shorter range under drag is evident.

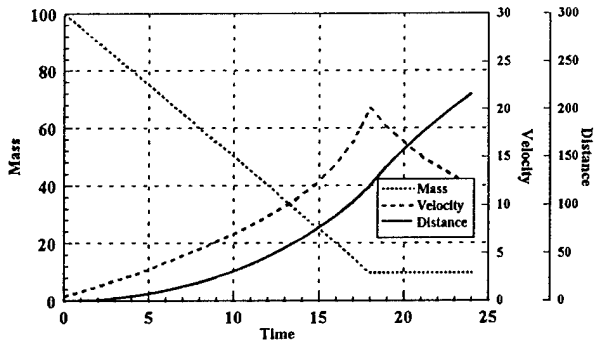


Figure 5: Simple Rocket Motion with Drag

30.5 Range Deviation

We now turn to considering the question of inaccuracy in shooting, either a bullet or a rocket. Our first consideration will be for range error. As a simplification to the mathematics, we shall ignore the effects of drag, recognizing that it will reduce the ranges that we calculate. The simple approach to direct fire shooting is to want the want the projectile to hit the target when the projectile is at the top of its trajectory. To do this, range must be estimated (at least to set the sights.) Call this r_e, and call the true range r_t, and the difference is δr. If we use the small angle approximation, then by equation 39,

$$\begin{aligned} \psi_e &\simeq \frac{g r_e}{v^2}, \\ &= \psi_t + \delta\psi. \end{aligned} \tag{47}$$

If the height of the target to be hit is h, and we want the projectile to strike the target at half this height, then we may take the height equation 38 as

$$\begin{aligned} h_e &\simeq \frac{v^2 \psi_e^2}{2g}, \\ &= h_t + \delta h. \end{aligned} \tag{48}$$

If we substitute equation 47 into equation 48, assume the true range gives the true height, and that the difference if the angles is small, then

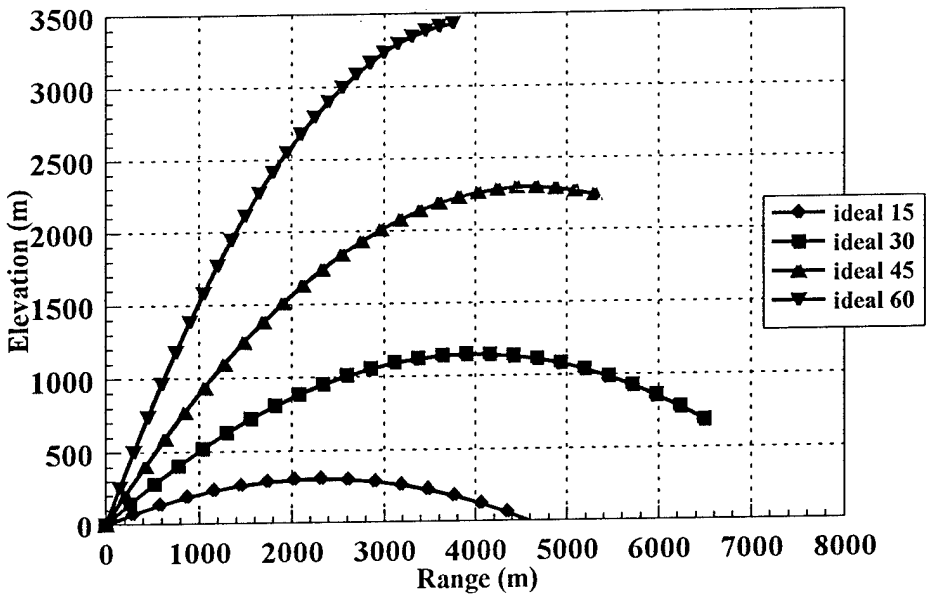


Figure 6: Sample Trajectories with Gravity

$$\delta h \simeq \frac{v^2 \psi_t \delta\psi}{g}. \tag{49}$$

Since

$$\delta\psi = \frac{g\delta r}{v^2}, \tag{50}$$

this reduces to

$$\delta h \simeq \psi_t \delta r. \tag{51}$$

This means that the error in height is directly proportional to the error in range estimation. Admittedly, this is an overestimation error (as we have defined it,) but it is still valid since most people overestimate range consistently, even when trained. To illustrate this effect, assume $h \sim 2m$, so that if $\delta h \sim 1m$, the projectile misses by flying over the target's top. If we further assume a true range of $400m$, and an average speed of $200m\,\mathrm{sec}^{-1}$(not unreasonable for small arms when we include drag,) then we get a value of $\psi_t \sim 0.1 radians$. This gives us a permissible range error $\delta r \sim 10m$. This partly explains why small arms are so alethal and why sergeants always tell their troops to aim high - in other words, underestimate range!

This formula is still approximately correct even if we catch the target while the projectile is dropping. It is generally more useful to rewrite it in the form

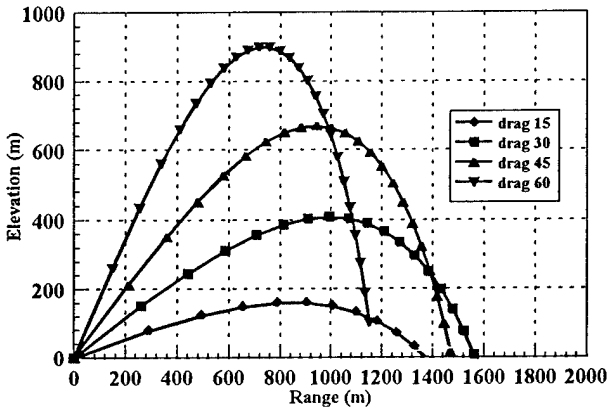


Figure 7: Sample Trajectories with Gravity and Drag

$$\delta h \simeq \frac{gr}{v^2}\delta r, \tag{52}$$

which shows why accurate rangefinders are necessary for long range fire. This formula is primarily of interest in direct fire where the target is engaged horizontally rather than in indirect fire where the target is engaged vertically.

30.6 Angular Deviation

We mentioned earlier that the motion of the bullet or rocket in the tube produced a exit velocity that was other than ideal. Additionally, it is hard to control rocket burning so that mal-alignment of the rocket thrust may also cause deviation from the intended path. We shall now consider the mathematical description of these deviations.

We adopt a rectangular coordinate system where the z axis is positive out from the tube and aligned along its axis and the x axis has a component in the local vertical. The y axis is parallel to the ground (assuming the ground to be flat.) The tube is elevated at an angle ψ that under ideal conditions would carry the projectile to the aim point on the target. (We assume that the azimuth heading of the tube is also ideal.) Because of tube or bullet/rocket variations, the

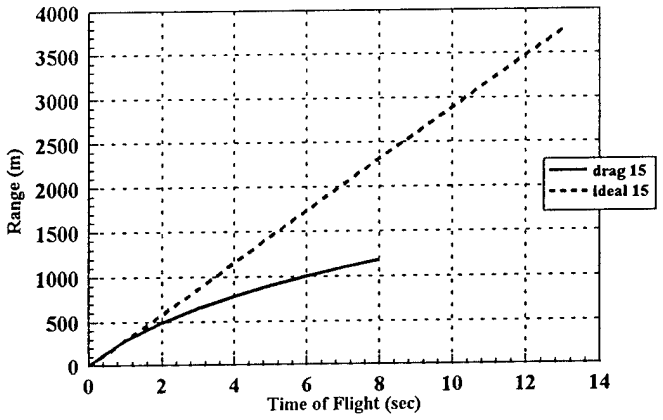


Figure 8: Simple Range-Time-of-Flight Curves

velocity vector of the projectile is not aligned with the z axis, but is very close. The projections of the velocity vector into the x-z and y-z planes of the coordinate system produce angles ω_x and ω_y with respect to the z axis. This geometry is shown in figure 9. For a velocity magnitude v, the components are then

$$\begin{aligned} v'_x &= v\sin(\omega_x), \\ v'_y &= v\sin(\omega_y), \\ v'_z &= v\sqrt{1-\sin^2(\omega_x)-\sin^2(\omega_y)}, \end{aligned} \tag{53}$$

where the primes indicate the tube coordinate system. If the deviations really are small, then the small angle approximation holds, and these may be approximated as

$$\begin{aligned} v'_x &\simeq v\omega_x, \\ v'_y &\simeq v\omega_y, \\ v'_z &\simeq v\sqrt{1-\omega_x^2-\omega_y^2} \\ &\simeq v. \end{aligned} \tag{54}$$

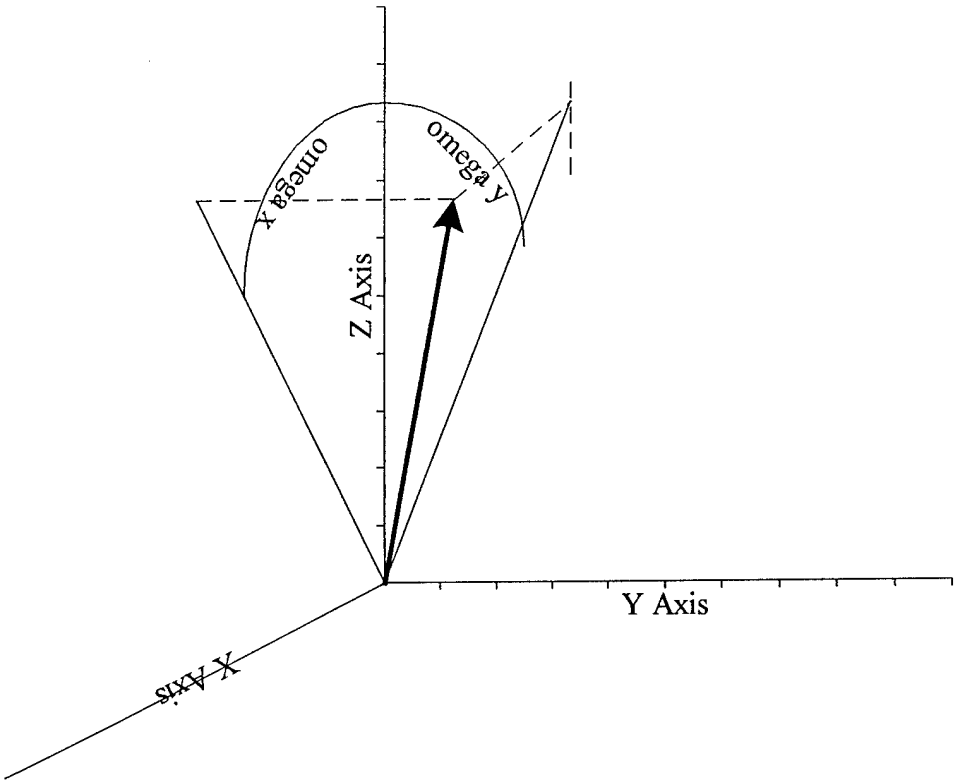


Figure 9: Exit Velocity Coordinate System

Thus, in the limit that the small angle approximation holds, the deviation of the trajectory from the ideal is directly proportional to the angles ω_x and ω_y. We must now consider their effect in the coordinate system (frame of reference) of the launcher, so we must now consider these motion under two conditions - direct and indirect fire.

30.6.1 Trajectories and Impact Points

The direct fire case is really a special case of the indirect fire case, but takes on special moment because the elevation angle is small. For these considerations, we shall use a coordinate system where the z axis is positive along increasing elevation, the x axis lies along increasing range, and the y axis is cross-range. The two coordinate systems are related (in the direction we want,) by

$$\begin{aligned} x &= -x' \sin(\psi) + z' \cos(\psi), \\ y &= -y', \\ z &= x' \cos(\psi) + z' \sin(\psi). \end{aligned} \tag{55}$$

Using equations 54, this gives us velocity components in the tube system of

$$\begin{aligned} v_x &\simeq -v\omega_x \sin(\psi) + v\cos(\psi), \\ v_y &\simeq -v\omega_y, \\ v_z &\simeq v\omega_x \cos(\psi) + v\sin(\psi), \end{aligned} \tag{56}$$

Before we solve the trajectory differential equations, we need to calculate the ideal (non-deviated) velocity vector. We designate this by $\underrightarrow{u'} = u'_{z'}\hat{e}_{z'}$, which has magnitude v.. This gives ideal velocity vector components in the tube system as

$$\begin{aligned} u_x &= v\cos(\psi), \\ u_y &= 0, \\ u_z &= v\sin(\psi). \end{aligned} \tag{57}$$

The trajectory equations are

$$\frac{d^2}{dt^2}\underrightarrow{r(t)} = -g\hat{e}_z, \tag{58}$$

with initial conditions

$$\begin{aligned} \underrightarrow{r}(t)\Big|_{t=0} &= \underrightarrow{0}, \\ \frac{d}{dt}\underrightarrow{r}(t)\Big|_{t=0} &= \underrightarrow{v}, \end{aligned} \tag{59}$$

for the deviated trajectory. For the ideal trajectory, $\underrightarrow{v}$ is replaced with $\underrightarrow{u}$. This gives us a trajectory of

$$\begin{aligned} x &= v_x t_i, \\ y &= v_y t_i, \\ z &= v_z t_i - \frac{g t_i^2}{2}, \end{aligned} \tag{60}$$

where t_i = time of impact. Since $z = 0$ at impact, and we are not interested in the $t_i = 0$ root of the last of equations 60, then

$$t_i = \frac{2v_z}{g}. \tag{61}$$

This gives the values of the impact point as

$$\begin{aligned} x &= \frac{2v_z v_x}{g}, \\ y &= \frac{2v_z v_y}{g}, \end{aligned} \tag{62}$$

by substitution.

We may now make use of the ideal and deviated velocity components to write the ideal and deviated (indicated by subscripts i and d) impact points. The ideal impact points are

$$\begin{aligned} x_i &= \frac{2v^2 \sin(\psi)\cos(\psi)}{g} \\ &= \frac{v^2}{g}\sin(2\psi), \\ y_i &= 0. \end{aligned} \tag{63}$$

The deviated impact points are

$$\begin{aligned} x_d &= \frac{2v^2\left(\sin(\psi)\cos(\psi) - \omega_x \sin^2(\psi)\right)}{g}, \\ y_d &= -\frac{2v^2 \sin(\psi)\omega_y}{g}, \\ z_d &= \frac{2v^2\omega_x \sin(\psi)\cos(\psi)}{g}. \end{aligned} \tag{64}$$

To make use of these, we must calculate the deviation of the impact point from the ideal,

$$\begin{aligned} \delta x &= x_d - x_i, \\ \delta y &= y_d - y_i, \\ \delta y &= y_d - y_i. \end{aligned} \tag{65}$$

This gives

$$\begin{aligned} \delta x &\simeq -\frac{2v^2\omega_x \sin^2(\psi)}{g}, \\ \delta y &\simeq -\frac{2v^2 \sin(\psi)\omega_y}{g} \\ \delta z &\simeq \frac{2v^2\omega_x \sin(\psi)\cos(\psi)}{g}. \end{aligned} \tag{66}$$

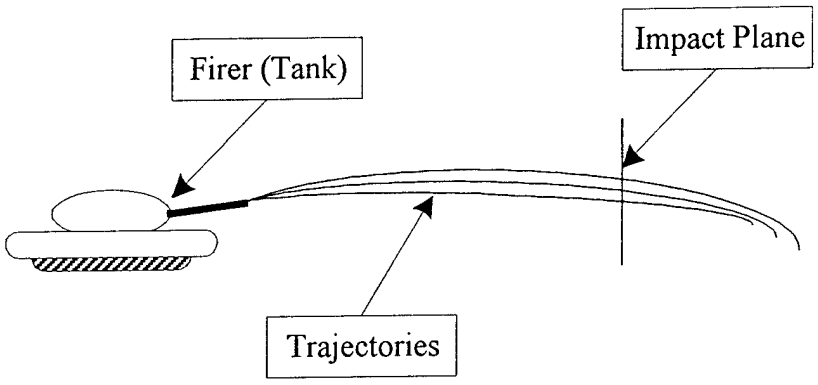


Figure 10: Direct Fire Vertical Dispersion

It is useful to define the range as

$$r \equiv \frac{2v^2 \sin(\psi) \cos(\psi)}{g}, \tag{67}$$

which is just x_i. This allows us to rewrite equations 66 in terms of the range to the ideal impact point,

$$\begin{aligned} \delta x &\simeq -r\omega_x \tan(\psi), \\ \delta y &\simeq -r\omega_y \sec(\psi), \\ \delta z &\simeq r\omega_x. \end{aligned} \tag{68}$$

30.6.2 Direct Fire

We now specifically treat the problem of direct fire. In a military since, this is fire where the target is directly observable (LOS exists), and the effect of the fire can usually be observed as well. Until the American Civil War, this was the province of infantry, artillery (except mortars - maybe,) and cavalry. With the invention of the conoidal bullet, artillery was banished behind the hill until World War I when the tank was developed. Even with modern sighting and fire control systems, there is still a basic need for these weapon system to be sighted, aimed at the target, by an operator (gunner for guns.) This leads to a mathematical simplification of the impact point deviations above. The situation is shown in figure 10 for the vertical component, and figure 11 for the horizontal component. The vertical trajectory components curve due to gravity.

In direct fire, the target is viewed as having vertical and width dimensions, but depth is not considered. As a result, the δx deviation is unimportant, being merely a minor delay or advance in the time of impact. Further, because of the problem of range error described above, muzzle

Direct Fire

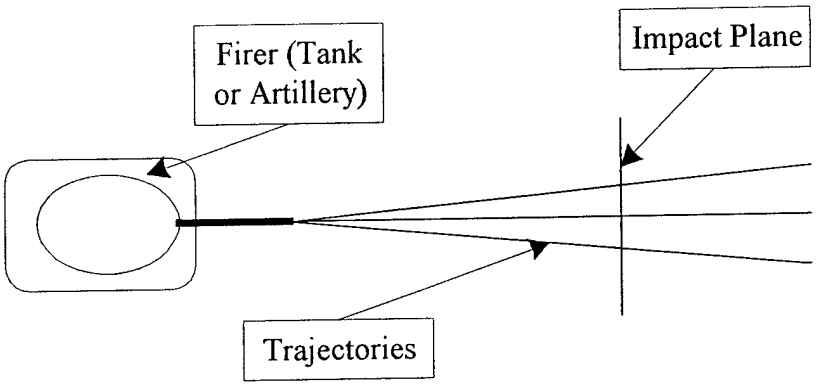


Figure 11: Direct or Indirect Fire Horizontal Dispersion

(and average) velocities are kept as high as possible. This translates into a situation where elevation angle is small, and the small angle approximation may be applied. In this case, the impact point deviations become

$$\begin{aligned} \delta y &\simeq -r\omega_y, \\ \delta z &\simeq r\omega_x. \end{aligned} \tag{69}$$

Note that each deviation depends on only one angle ω_x or ω_y, and that they are linear. If we now make the leap of belief that for a collection of identical weapons, there are pdfs of the angles ω_x and ω_y, then we are now in a position to rewrite them as pdfs of deviation. If p_{ω_x} = pdf of ω_x, defined by

$$p_{\omega_x}(\omega_x) = \left.\frac{dP_{\omega_x}(z)}{dz}\right|_{z=\omega_x}, \tag{70}$$

where: $P_1(z)$ = probability that $z \leq \omega_x$, then by equation 69,

$$\left.\frac{dP_{\delta z}(z)}{dz}\right|_{z=\delta z} = \left.\frac{dP_{\omega_x}(z)}{dz}\right|_{x=\omega_x} \left|\frac{d\omega_x}{d\delta z}\right|. \tag{71}$$

This gives

$$p_{\delta z}(\delta z) = p_{\omega_x}\left(\frac{\delta z}{r}\right)\frac{1}{r}. \tag{72}$$

In particular, if the pdfs are even functions, and in particular gaussians, which they are normally modeled as, then this transition is particularly painless. If the angle pdfs are of the form,

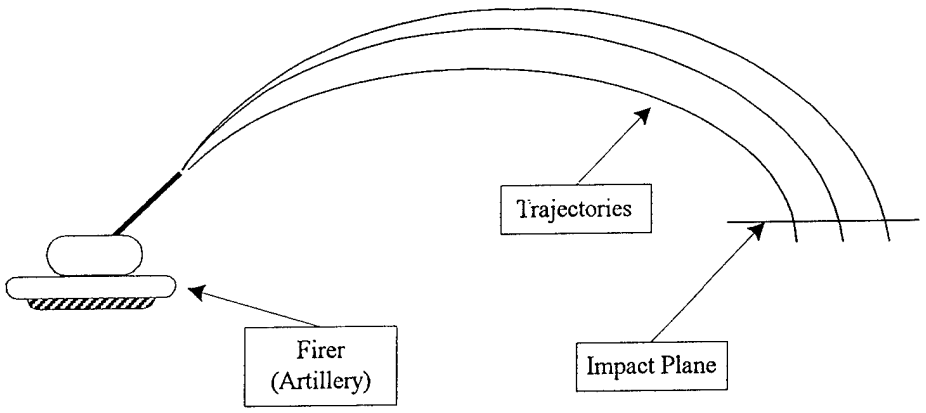


Figure 12: Indirect Fire Range Dispersion

$$p_{\omega_x}(\omega_x) = \frac{1}{\sqrt{2\pi}\alpha_x} \exp\left[-\frac{\omega_x^2}{2\alpha_x^2}\right], \tag{73}$$

and similarly for ω_y, then the pdf for direct fire is just

$$p(y_d, z_d) = \frac{1}{2\pi\sigma_y\sigma_z} \exp\left[-\frac{(y_d - y_i)^2}{2\sigma_y^2} - \frac{(z_d - z_i)^2}{2\sigma_z^2}\right], \tag{74}$$

where the standard deviations have the simple form:

$$\begin{aligned} \sigma_y &= \alpha_y r, \\ \sigma_z &= \alpha_x r. \end{aligned} \tag{75}$$

It may be noted that we have not considered drag, wind gusts, etc. in this development. Quite frequently, these additional sources of deviation are either added into the angle standard deviations, or complicate those functions. We shall return to this pdf later.

30.6.3 Indirect Fire.

In indirect fire, the province of artillery and mortars (whoever they belong to, and this seems to vary a lot among armies,) is characterized by consideration of where the round lands on the ground. In this case, we are not interested in the elevation deviation δz since, again, it represents a minor variation in arrival time, but we are interested in the range and cross range deviations, δx and δy. This is depicted for the cross range (horizontal) deviation in figure 11, and for the range deviation in figure 12.

Unlike for the direct fire case, we cannot apply the small angle approximation to the elevation angle. Thus, the deviations, written in range form like equation 63 are

$$\begin{aligned} \delta x &\simeq -r\omega_x \tan(\psi), \\ \delta y &= -r\omega_y \sec(\psi), \end{aligned} \tag{76}$$

which, as in the direct fire case, are linear and simple in the deviation angles. Development of the pdfs proceeds in the same manner as above, and results in the same functional form as equation 74,

$$p(x_d, y_d) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{(x_d - x_i)^2}{2\sigma_x^2} - \frac{(y_d - y_i)^2}{2\sigma_y^2}\right], \tag{77}$$

except that now,

$$\begin{aligned} \sigma_x &\simeq \alpha_x r \tan(\psi), \\ \sigma_y &\simeq \alpha_y r \sec(\psi), \end{aligned} \tag{78}$$

the standard deviations are functions of elevation angle. This means that the range and cross range standard deviations do not simply scale the angular deviation standard deviations. We shall examine this in detail below.

30.6.4 Probability Density Contours

Before proceeding, we note that the contours of constant probability density (curve of the negative of the exponent set equal to some value,) are ellipses centered on the ideal impact point. For the direct fire case, the axes of the ellipse are aligned along the vertical and width dimensions. While the size of these ellipses grows with range, they preserve the shape of the equivalent ellipses for the deviation angles ω_x and ω_y. If, for example, these two angles' distributions have the same standard deviation (which ignoring gravity tip-off and wind gusts, they should have,) then the contours of constant probability density for the deviation angles would be circles. Whatever the shape, it is preserved when going to the direct fire case. Even more interestingly, because of the small angle approximation, the size of the standard deviations do not depend on elevation angle.

The situation changes when we go to the indirect fire case. The contours of constant probability density are still ellipses centered on the ideal impact point, but their axes are now aligned along the range and cross range directions. What is more important, not only do they now depend on elevation angle, but the functional dependence is different. Thus the shape of the indirect fire impact point deviations' contour of constant probability density is different from that of the deviation angles. To illustrate this, we plot several such contours in figure 13 The deviation angle contour is the circle, which arises from assumption the $\omega_x = \omega_y$. The four elliptical contours correspond to elevation angles of $15, 30, 45$,and $60°$, with the first defining

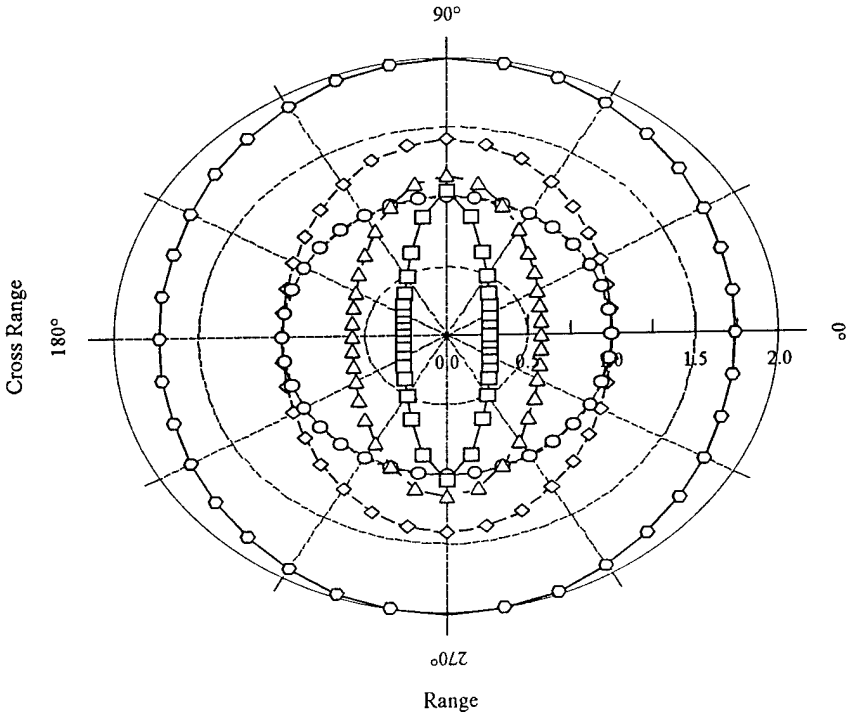


Figure 13: Representative Probability Density Contours

the smallest and the last defining the largest. Range has been held constant in calculating these contours. The deviation angle contour's size should not be used for comparison since its units are different (radians versus meters.)

Having established that the indirect fire contours of constant probability density alter the shape of the equivalent curves for the angle dispersions, the next question is, how do we compute these? This question is at the root of the concept of the Circular Error Probable (CEP) which is the size of the contour that corresponds to a probability of 0.5 for a circular pdf. A circular pdf has $\sigma_x = \sigma_y$, so that the pdf becomes

$$p(x,y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2+y^2}{2\sigma^2}\right]. \tag{79}$$

Since $x^2 + y^2 = \rho^2$ in a circular coordinate system, this reduces the pdf to

$$p(\rho,\theta) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{\rho^2}{2\sigma^2}\right], \tag{80}$$

which is exactly integrable since the Jacobean of the transformation is

$$J(x,y:\rho,\theta) = \left| \begin{array}{cc} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \theta} \end{array} \right|, \tag{81}$$

where the || indicates both determinant and absolute value. This modifies the probability integral to

$$\begin{aligned} P_{hit}(\rho') &= \frac{1}{2\pi\sigma^2}\int_0^{2\pi} d\theta \int_0^{\rho'} \rho d\rho \exp\left[-\frac{\rho^2}{2\sigma^2}\right] \\ &= \frac{1}{\sigma^2}\int_0^{\rho'} \rho d\rho \exp\left[-\frac{\rho^2}{2\sigma^2}\right], \end{aligned} \tag{82}$$

which is exactly integrable with the value,

$$P_{hit}(\rho') = 1 - \exp\left[-\frac{\rho'^2}{2\sigma^2}\right]. \tag{83}$$

Note that this form of the integral applies only to circular areas, centered on the ideal impact point. This equation can be solved for the radius circle that corresponds to a given probability as

$$\rho' = \sigma\sqrt{-2\ln(1-P_{hit})}, \tag{84}$$

so that the CEP has a radius,

$$\rho_{CEP} = \sigma\sqrt{-2\ln(0.5)}. \tag{85}$$

This points up the general approach to doing these probability integrals - finding a coordinate system. that matches the shape of the target area, and transforming to it. A somewhat more general approach which accommodates the differences in the standard deviations is an elliptical coordinate system defined by

$$\begin{aligned} x &= \sigma_x \rho\cos(\theta), \\ y &= \sigma_y \rho\sin(\theta), \end{aligned} \tag{86}$$

which has a Jacobean

$$J(x,y:\rho,\theta) = \sigma_x\sigma_y\rho. \tag{87}$$

This maps the pdfs exponent

$$\begin{aligned} \frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} &= \frac{\left(\sigma_x \rho \cos\left(\theta\right)\right)^2}{2\sigma_x^2} + \frac{\left(\sigma_y \rho \sin\left(\theta\right)\right)^2}{2\sigma_y^2} \\ &= \frac{\rho^2}{2}, \end{aligned} \tag{88}$$

and transforms the probability integral into the same form as equation 82 and results in the same form as equation 83. This allows solution for a radius ρ which then generates an ellipse of constant probability density using the coordinate system definitions above which are also the definition of an ellipse in circular coordinates.

30.7 Probability of Hit

It is a considerable misrepresentation to talk about probability of hit, P_{hit}, separate from probability of kill, P_{kill}. Since kill depends on momentum transfer, generally, penetration, and most targets had spatially varying vulnerability to penetration (i.e., amount of momentum transfer necessary for penetration,) then the proper approach to calculating P_{kill} would be to calculate the probability of kill at each point on the target (given a level of momentum,) $p_{kill}\left(x, y\right)$, and then to average with respect to the probability of hitting each point. That is,

$$P_{kill} = \int_{tgt} p_{hit}\left(x, y\right) p_{kill}\left(x, y\right) dxdy, \tag{89}$$

which is an area integral because, regardless of whether direct or indirect fire, the target presents a perceived area. (This falls apart for certain non-momentum transfer weapons such as the thermal shock of nuclear weapons.) If we define P_{hit} as

$$P_{hit} = \int_{tgt} p_{hit}\left(x, y\right) dxdy, \tag{90}$$

then we may reliably define a conditional probability of kill given a hit,

$$P_{kill|hit} \equiv \frac{P_{kill}}{P_{hit}}. \tag{91}$$

This is very different from a probability of kill (!) given a hit defined as

$$P^*_{kill|hit} = \frac{\int_{tgt} p_{kill}\left(x, y\right) dxdy}{A_{tgt}}, \tag{92}$$

where: $A_{tgt} = \int_{tgt} dxdy$. This probability is approximately valid only when $p_{kill}\left(x, y\right)$is relatively constant, changing much slower than $p_{hit}\left(x, y\right)$. For hard targets (e.g., tanks,) with limited vulnerable area over the total area, this quantity is spurious.

Nonetheless, since P_{hit} is calculable, it is common practice to separate these two. So long as the distinction above is realized, then only limited harm is done.

Having said all that, how do we calculate P_{hit}? The answer is that we use the pdfs defined above for impact point deviations.

For direct fire, we adopt a new two-dimensional rectangular coordinate system with origin at the ideal aim point. In keeping with our normal convention, the vertical axis is z and the horizontal is y. If we use the pdf defined assuming gaussian pdfs for the deviation angles, equation 74, then the hit probability is

$$P_{hit} = \frac{1}{2\pi\sigma_y\sigma_z}\int_{tgt}\exp\left[-\frac{(y-y_i)^2}{2\sigma_y^2}-\frac{(z-z_i)^2}{2\sigma_z^2}\right]dydz. \tag{93}$$

If the target is approximately rectangular, and the aim point (idealized impact point) is at the center of the target, then this becomes

$$P_{hit} = \frac{1}{2\pi\sigma_y\sigma_z}\int_{-\frac{y_t}{2}}^{\frac{y_t}{2}}\exp\left[-\frac{y^2}{2\sigma_y^2}\right]dy\int_{-\frac{z_t}{2}}^{\frac{z_t}{2}}\exp\left[-\frac{(z-z_i)^2}{2\sigma_z^2}\right]dz. \tag{94}$$

These integrals can be approximated by use of our (favorite?) gaussian integral approximation from Appendix F of Part I. Because of the symmetry, this becomes simply

$$P_{hit} \simeq \sqrt{1-\exp\left[-\frac{y_t^2}{2\pi\sigma_y^2}\right]}\sqrt{1-\exp\left[-\frac{z_t^2}{2\pi\sigma_z^2}\right]}. \tag{95}$$

If we now use the range dependent form of the standard deviations, this becomes

$$P_{hit} \simeq \sqrt{1-\exp\left[-\frac{y_t^2}{2\pi r^2\omega_y^2}\right]}\sqrt{1-\exp\left[-\frac{z_t^2}{2\pi r^2\omega_z^2}\right]}, \tag{96}$$

which provides us with a range dependent probability of hit. If the target is also square, with area A_t, and the deviation angle standard deviations are the same, this further reduces to

$$P_{hit} \simeq 1-\exp\left[-\frac{A_t}{2\pi r^2\omega^2}\right]. \tag{97}$$

To examine this, we calculate the probability of hit for two classes of weapons against our model fully exposed and hull defilade tanks ($2.67m$ x $2.67m$ and $2.67m$ x $1.0m$). The weapons are a modern tank and a Light Anti-Tank (LAW) rocket. Both are assumed to have identical deviation angle standard deviations of $0.5mrad$ and $15mrad$, respectively. (A mrad is a milliradian, a thousandth of a radian. This is a frequently used angle unit since when multiplied by a range in kilometers, the result is in meters.) This is shown in figure 14. These two systems represent the extremes of direct fire anti-tank weapons.[6]

For indirect fire, calculations are generally similar, except that equations 77 and 77 appertain. They give a range and elevation angle dependent pdf,

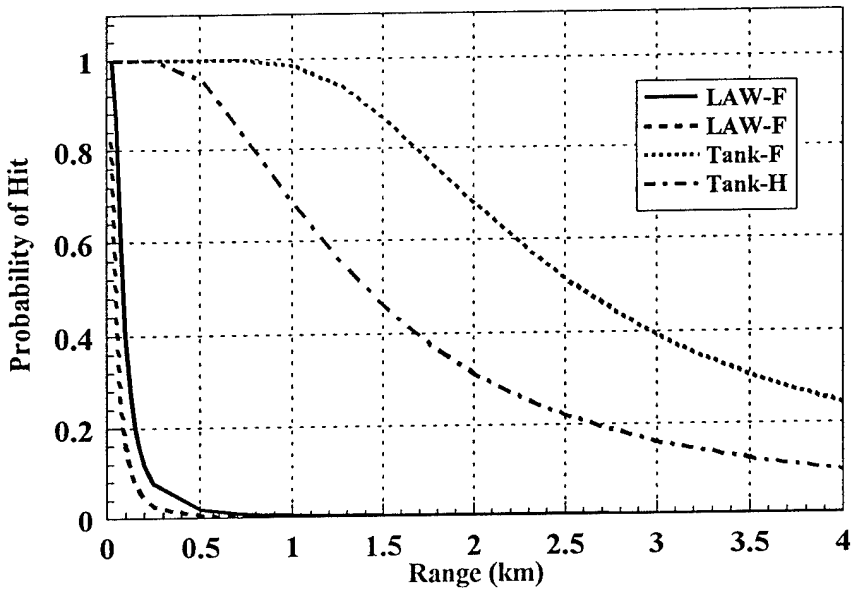


Figure 14: Representative Direct Fire Probability of Hit

$$p_{hit}(x_d, y_d) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{(x_d - x_i)^2}{2\alpha_x^2 r^2 \tan^2(\psi)} - \frac{(y_d - y_i)^2}{2\alpha_y^2 r^2 \sec^2(\psi)}\right]. \quad (98)$$

30.8 Kills

A matter of special consideration is the presence of multiple targets. In this case, it is useful to proceed directly to expected kills per round fired. If the probability of kill, $p_{kill}(x', y')$ is a function of the distance, x', y',[5] from the actual impact point, and the density of targets is $\rho_{tgt}(x", y")$, then the expected kills per round fired is

$$\left\langle \frac{kills}{round} \right\rangle = \int\int_{-\infty}^{\infty} dx_d dy_d p_{hit}(x_d, y_d) \int\int_{-\infty}^{\infty} dx' dy' p_{kill}(x' - x_d, y' - y_d)\, \rho_{tgt}(x', y'). \quad (99)$$

If $\rho_{tgt}(x', y') \sim const$, then this just reduces to

[5]This is a function of either the ground distance or the volume distance from the effective impact point. The latter is a distinction for air bursts, or earth penetrating bursts. In this case, the probability of hit must be modified to the ideal height $\neq 0$.

$$\left\langle \frac{kills}{round} \right\rangle = \rho_{tgt} \int \int_{-\infty}^{\infty} dx' dy' p_{kill}(x', y'), \tag{100}$$

because we don't care (for this calculation,) where the round lands since by assumption the density of targets is constant. A frequently used form of lethality function is

$$p_{kill}(R) = \begin{cases} p_{kill}, R \leq R' \\ p_{kill} e^{-\kappa(R-R')}, R > R' \end{cases}, \tag{101}$$

which is just a disk of radius R' of constant (normally unitary) lethality, and a decaying lethality component outside. Other variations are also used, but all represent an inner region of certain level of lethality (nominally unitary,) and an outer region of decaying lethality. This form of the lethality function reduces the kills per round calculation to

$$\begin{aligned} \left\langle \frac{kills}{round} \right\rangle &= \rho_{tgt} p_{kill} \int_0^{2\pi} d\theta \left[\int_0^{R'} dr r + \int_{R'}^{\infty} dr r e^{-\kappa(r_d - R')} \right] \\ &= 2\pi \rho_{tgt} p_{kill} \left[\frac{R'^2}{2} + \frac{1 - \kappa R'}{\kappa^2} \right]. \end{aligned} \tag{102}$$

We may also use this formalism to compute the spatial losses per round. To do this, we modify equation 99 to

$$\Delta \rho_{tgt}(x', y') = \int \int_{-\infty}^{\infty} dx_d dy_d p_{hit}(x_d, y_d) p_{kill}(x' - x_d, y' - y_d) \rho_{tgt}(x', y'). \tag{103}$$

If we define a lethality function as

$$L(x', y') \equiv \int \int_{-\infty}^{\infty} dx_d dy_d p_{hit}(x_d, y_d) p_{kill}(x' - x_d, y' - y_d), \tag{104}$$

which is just the expected probability of kill at any point, then we may consider the effect of multiple round fire. To do this, we note that if the initial force strength density is $\rho_{tgt}(x', y')$, then after one round, the force strength density after one shot is

$$\rho_{tgt}(x', y')|_{1rnd} = (1 - L) \rho_{tgt}(x', y'), \tag{105}$$

where L is given by equation 104. The force strength density after two shots is then

$$\rho_{tgt}(x', y')|_{2rnd} = (1 - L)^2 \rho_{tgt}(x', y'), \tag{106}$$

and the force strength density after n shots is just

$$\rho_{tgt}(x', y')|_{nrnd} = (1 - L)^n \rho_{tgt}(x', y'). \tag{107}$$

If we now define the loss function as

$$\begin{aligned} \lambda(x',y') &\equiv -\ln(1-L(x',y')) \\ &\simeq L(x',y'), \end{aligned} \tag{108}$$

where the approximation is valid for $L(x',y') \ll 1$ for all x', y', then we find a force strength equation of the form

$$\rho_{tgt}(x',y',n) = e^{-\lambda(x',y')n}\rho_{tgt}(x',y',0). \tag{109}$$

As we did in the detection case with glimpses, then we may define a mean rate of fire such that $n = \varsigma t$, then this equations becomes

$$\rho_{tgt}(x',y',t) = e^{-\lambda(x',y')\varsigma t}\rho_{tgt}(x',y',0), \tag{110}$$

which we may differentiate wrt time to yield a simple attrition differential equation,

$$\frac{d}{dt}\rho_{tgt}(x',y',t) = -\lambda(x',y')\,\varsigma\rho_{tgt}(x',y',t). \tag{111}$$

It is notable that even though this equation is for indirect fire, it appears to be an exponential attrition differential equation. This results from the restriction to a single ideal aim point. In effect this equation represents the fire from only one element of firer.

To address this situation, we need to address the points of multiple firers and aggregation. We address the latter first. If the variation in $\rho_{tgt}(x',y')$ is greater than the variation in $L(x',y')$, and assume that the force strength density (approximately constant area of coverage,) adjusts rapidly and maintains the same approximate shape, then we may approximately define an average lethality as

$$\overline{L} \simeq \frac{= \iint_{-\infty}^{\infty} dx_d dy_d p_{hit}(x_d,y_d) \iint_{-\infty}^{\infty} dx'dy' p_{kill}(x'-x_d, y'-y_d)\,\rho_{tgt}(x',y')}{\iint_{-\infty}^{\infty} dx'dy'\rho_{tgt}(x',y')}, \tag{112}$$

which we note is really a function of x_i, y_i. If we further assume that for multiple fires into an area, then we may postulate a distribution of ideal aim points, and if we average $\overline{L}$ with respect to this distribution, then we may calculate a multiple firer average lethality $\widehat{L}$ that allows us to define an average loss $\widehat{\lambda}$ in analog with equation 108 which reduces equation 111 to

$$\frac{d}{dt}\rho_{tgt}(x',y',t) = -\widehat{\lambda}\varsigma\rho_{tgt}(x',y',t)\,F, \tag{113}$$

where: $F =$ number of firers. This is clearly a linear attrition differential equation. (Of course, we could have used equation 112 as the basis for our averaging, but this way is cleaner.) Integrating equation 113 wrt space reduces to

$$\frac{d}{dt}T = -\widehat{\lambda}\varsigma TF, \tag{114}$$

where: $T = \iint_{-\infty}^{\infty} dx'dy'\rho_{tgt}(x',y')$.

30.9 References

[1] Gabriel, Richard A., and Karen S. Metz, **From Sumer to Rome - The Military Capabilities of Ancient Armies**, Greenwood Press, New York, 1991.

[2] Halliday, David, and Robert Resnick, **Physics**, John Wiley & Sons, Inc., New York, 1966, pp. 572 ff.

[3] Prudnikov, A. P., Yu. A. Brychkov, and O. I. Marichev, **Integrals and Series Volume I: Elementary Functions**, Nauka, 1981, N. M. Queen, trans., Gordon and Breach Science Publishers, New York, 1986, p. 244, Integral 1.6.5.4.

[4] Marion, Jerry B., **Classical Dynamics of Particles and Systems**, Academic Press, New York, 1965.

[5] Marion, Jerry B., *loc. cit.*, pp. 64-74.

[6] Fowler, Bruce W., "Environmental Effects on Combat Performance: A Lanchester Approach", in L. G. Callahan, Jr., and Lawrence Low, eds., **Proceedings of the Workshop on Modeling, Simulation, and Gaming of Warfare**, Callaway Gardens, GA, 2-5 Dec 1984, pp. 145-175.

This Page Intentionally Left Blank

Chapter 31

Weapons II

31.1 Introduction

In this chapter, we now take up the gauntlet of modern weapon systems, primarily those which are guided in some fashion. This will be the principal theme of this chapter and in keeping with our tone thus far, we shall retain a top level discussion. Unfortunately, this will tend to preclude much of the mechanics of these systems since that mechanics becomes exceedingly complicated and thus lies beyond the scope of our discussion. For more detailed discussion of these systems, the reader is initially referred to general texts such as Lee, et al., [1], and from there to advanced texts and reports.

31.2 Why Guidance?

The basic motivation for guidance is found in the preceding chapter, Weapons I. Because bullets and rockets have inaccuracies that increase with range, attriting small targets at long range becomes increasingly difficult. We may view the expected number of projectile necessary to attrit a target at range as

$$\overline{n} \simeq \frac{1}{P_h(r)\, P_{k|h}}, \tag{1}$$

where:

$P_h(r)$ = probability of hit at range r, and

$P_{k|h}$ = probability of kill given a hit.

If we take the special case of circular distribution as representative,

$$P_h(r) \simeq 1 - \exp\left[-\frac{A_t}{2\pi r^2 \omega^2}\right], \tag{2}$$

which to first order is

$$P_h(r) \simeq \frac{A_t}{2\pi r^2 \omega^2}. \tag{3}$$

If we now combine equations 1 and 3, then we get

$$\overline{n} \simeq \frac{2\pi r^2 \omega^2}{A_t P_{k|h}}, \tag{4}$$

which indicates that the expected number of shots to kill a target at range scales directly as the square of the range and inversely as the size of the target. Thus, if it takes two shots to kill a target at one kilometer, it takes eight shots to kill at two kilometers!

This is the motivation for guided weapons - to find a way to increase the accuracy of the projectile at range. There are basically two ways of implementing guidance: command guidance and terminal homing guidance, and there are two forms of implementation of guidance: projectile guidance and projectile warhead guidance. We shall briefly discussed in turn.

31.3 Command Guidance

In most implementations, command guidance is a push type of guidance in the sense that the guidance pushes the projectile to the target. In general, the guidance is effected by a combination of an electro-optical mechanism at the launcher and a combination of electro-optical and mechanical mechanisms in the projectile.

31.3.1 Inertial Guidance

The fundamental baseline of guidance is usually inertial guidance. In this form of guidance, gyroscopes are incorporated into the body of the projectile along with some mechanical means of steering the projectile, usually either movable fins or small reactive jets or rockets. A special case of this is incorporation of one of these means in the nozzle of a rocket to give limited pointing capability for the exhaust. This is known as Thrust Vector Control.

The gyroscopes provide a trajectory reference for the projectile. Deviations from a programmed trajectory are sensed and corrections to the actual trajectory are made. In practice, this type of guidance reduces the effects of post launch influences on the trajectory so that at best it reduces the inaccuracies to about the level of the launcher induced inaccuracies. Thus, this type of guidance is primarily of value in short range direct fire and moderate range indirect fire applications.

An approximation to this type of guidance may be gained dynamically by spinning the projectile. This type of guidance has always been dominated by the cost and complexity of the gyroscopes and was introduced during World War II with an initial application to anti-aircraft shells. Despite this, inertial guidance is usually incorporated in modern guided projectile as part of a mixed guidance system because of the advantages it offers in smoothing the trajectory.

31.3.2 Differential Guidance

The basic idea of command guidance is to direct the projectile towards the target. The most common and oldest forms of this are guidance techniques that guide the projectile towards a trajectory that impacts the target. The implementation of this is to sense the locations of the target and projectile in real time, calculate what the ideal trajectory should be, and guide the projectile back to this ideal trajectory (if the projectile has deviated from the ideal trajectory.) The simplest form of this is called Command to Line-of-Sight (CLOS) guidance, which uses the LOS to the target to define the ideal trajectory. There are three implementations of CLOS guidance: Manual CLOS, where both target and projectile must be tracked by the weapon system operator; Semi-Automatic CLOS (SACLOS), where only the target must be manually tracked, but the projectile is automatically tracked; and Automatic CLOS (ACLOS), where both target and projectile are automatically tracked. Most second generation Anti-Tank Guided Missiles[1] (ATGMs) use SACLOS guidance. The common implementation is to place an optical or radar light source on the rear of the missile to facilitate tracking. This is sketched in figure 1.

The CLOS technique normally only works for missiles that have boost-coast propulsion. Missiles that either boost or sustain all the way to the target must be guided to the target along a shaped trajectory to prevent the exhaust plume's smoke or thermal emissions from obscuring the target to the tracker on the launcher. Generally, differential commanded projectiles are quite accurate and have a probability of hit that is fairly consistent. The idealized form of this probability is [2]

$$P_{hit}(r) = \begin{cases} 0, r < r_{rin}, \\ p_{hit}, r_{\min} \leq r \leq r_{\max}, \\ 0, r > r_{\max}. \end{cases} \tag{5}$$

Most modern systems realize this idealized form almost exactly, but some older systems suffer some degradation at ranges close to $r_{\max}$. In general, the minimum range occurs for three reasons: allowing the missile to reach sufficient range to (i) complete boost; (ii) stabilize its trajectory past launch transients; and (iii) achieve a sufficient range to safely arm the warhead fuze. The maximum range generally occurs because of loss of aerodynamic stability and/or degradation of tracking and guidance accuracy of the electro-optical system.

31.3.3 Beam Guidance

The other common form of command guidance is beam guidance. In this scheme, a light (usually laser) or radar beam is projected to the target. The missile, using backwards looking detectors, senses the beam and follows it to the target. This scheme has the advantage of negating the need to track the missile, but suffers from the disadvantage of decreasing beam energy density with range. An approach to mitigating this is to reduce the size of the beam as the missile flies

[1]Usually, the distinction is made that unguided rockets, possibly neglecting inertially guided rockets, are called rockets. Guided rockets (or guided powered projectiles such as small airplanes,) are usually called missiles.

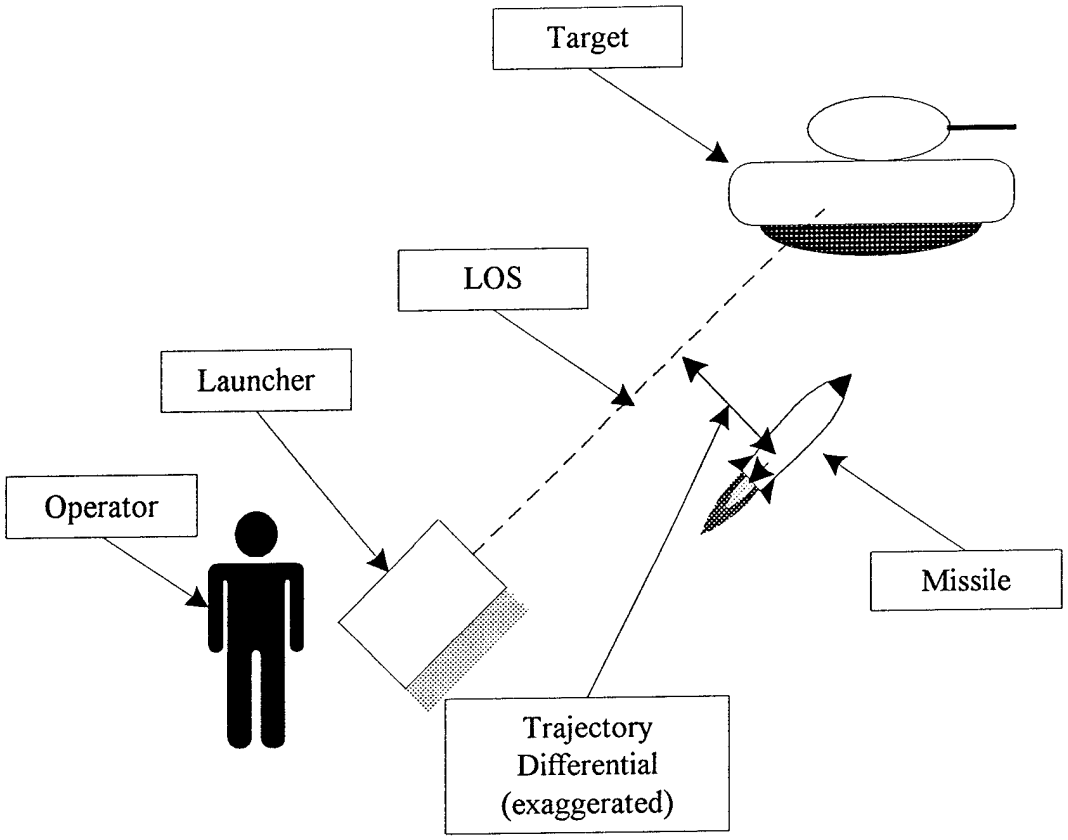


Figure 1: Differential Command Guidance

so that the diameter of the beam at the rear of the missile is essentially constant (and thus has constant energy density.) This is sketched in figure 2.

The general form of the probability of hit is similar to that for CLOS.

31.3.4 Mechanics and Accuracy

Because of the nature of command guidance, the trajectories of command guided projectiles are highly non-ballistic. Ideal CLOS trajectories are straight lines between launcher/controller and target. The only curvature thus induced into these trajectories is due to the motion of target or (less commonly) launcher/controller. These are not energy efficient trajectories since thrust and lift are used to maintain this shape of trajectory. For this reason, the trajectories of these guided projectiles have a very simple mathematical form that owes little to basic mechanics. (Indeed, the engineering computation of trajectories for these projectiles incorporates the laws of Newtonian Mechanics, but these are effectively overwhelmed by the stringency of guidance and control.)

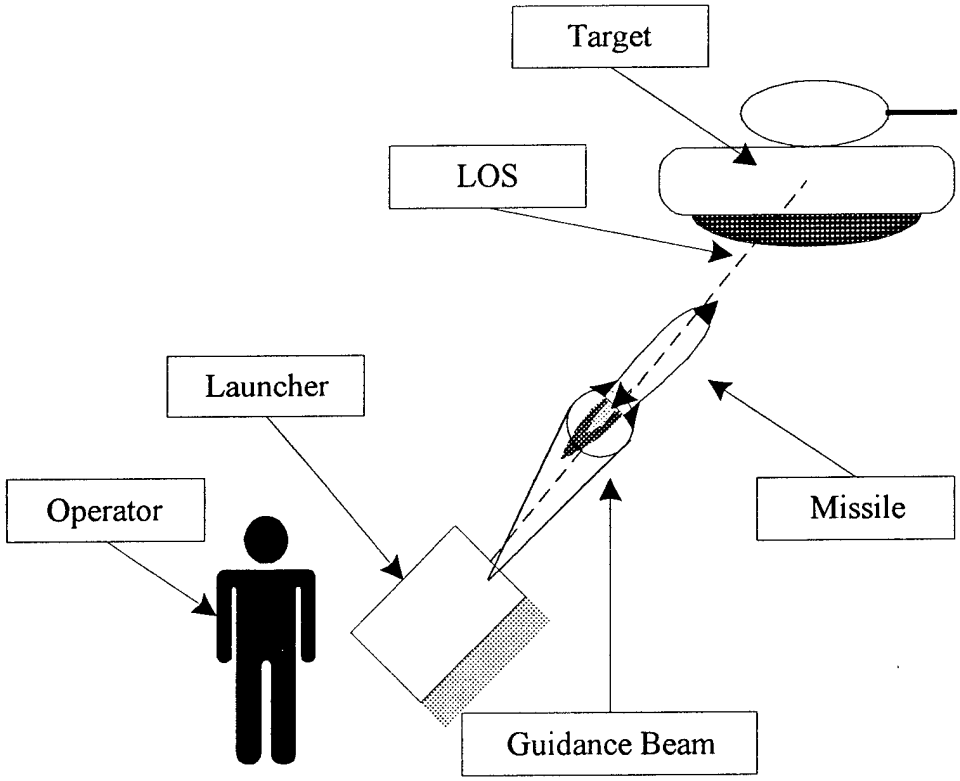


Figure 2: Beam Command Guidance

The velocities of these projectiles has a form that is dominated by the propulsion implementation. Most of these projectiles use a boost-coast type of propulsion where the projectile is boosted to a reasonable speed and it then coasts to the target. Thus the velocities of these projectiles has a form similar to those described in the previous chapter for rockets with drag and the additional complication of the imposed shape of the trajectory. Because of the complexity of calculating the trajectory, two common analytical approaches are used: representing the velocity curve as a mean averaged over the entire trajectory; or simply providing the analyst with a range-time-of-flight curve. The former approach has the disadvantage of overestimating the time-of-flight to range for short ranges and underestimating the time-of-flight to range for longer ranges. Except for Air Defense applications, size considerations for the projectile tend to limit these average velocities to $\sim$ Mach 1 (about 330 m sec^{-1}), so for ranges less than 5 km, times of flight are no more than 15 sec. For Air Defense applications, higher speeds are necessary to permit reasonable pursuit (see below.) Obviously, drag minimization is a concern for these projectiles.

The basic idea of this type of guidance is to provide greater accuracy, and it succeeds very well. Except for the very ends of the trajectories where the projectile is running out of either guidance or velocity, CEPs of radii 0.25-1 m are the norm, making the probability of hit for these projectiles essentially constant over their trajectories.

31.4 Semi-Active Guidance

This type of guidance lies between command and terminal homing guidance in type. It actually combines the characteristics of both. The system can be compared to bistatic sensors described previously, although the transmitter and receiver do not initially have to be widely apart..

The basic implementation is to have an illumination source that tracks the target and projects a beam of light onto the target. The illuminator is usually either a laser or a radar. Laser illuminators are usually designed to produce a small solid angle beam so that the spot on the target is smaller than the target, thus eliminating false targets from background reflections. This is usually not possible with radars but is less important because of the lack of background reflectors. The reflected illumination is then sensed by a seeker in the front of the projectile and used by the projectile to guide itself into the target. This is sketched in figure 3.

The nature of the reflection usually causes different implementations for the different illuminators. Most targets, and especially ground targets (which are dirty!), are diffuse reflectors so that the illuminator and the launcher do not have to be closely located. The only real restrictions are that both the projectile and the operator must be on the same side of the target (dictated by the geometry of propagation), and for safety reasons, that the projectile not be able to see the illumination beam (otherwise it will home on the illuminator to the detriment of it, the operator, and the mission - not a good idea for several military reasons.)

For radar illuminators, the reflected signal has a great deal of variation with respect to the angle from the target. To assure that adequate reflected light is reaching the seeker, it is common to closely locate illuminator and launcher.

The common weakness of this form of guidance is the decrease in energy density over range, compounded by the reflectivity of the target (cross section for radar illumination.) Since the energy must travel first to the target and then to the projectile, the effective range of the system tends to be characterized in terms of the sum of the illuminator-target range and the initial (lock-on) projectile-target range. Operation is further complicated by the need for two LOS if illuminator and projectile are separately located. Even with these restrictions, these seekers are usually easier to build because of the illuminated target signature. They also have the military advantage of allowing a tactically small illumination system unburdened by projectiles that are carried on a robust vehicle that can either stand off in range or make use of terrain masking except during launch. This separation implements a partial Fire-and-Forget (see below) capability since the launcher is not concerned with the guidance after launch.

The accuracy of these systems is good. For laser illumination, accuracy is primarily limited

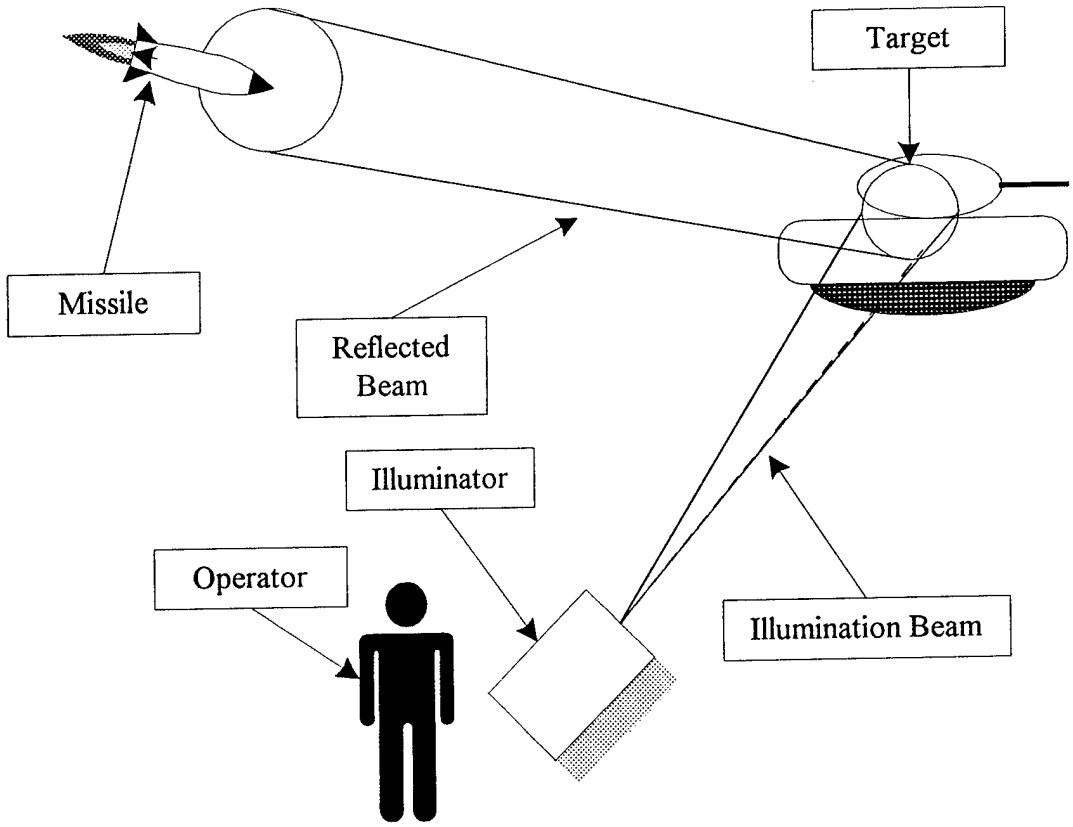


Figure 3: Semi-Active Guidance

by the illumination spot size which we already want to be smaller than the target but want to be large enough to not be hampered by local reflection specularity. For radar systems, the entire target is illuminated, but since the kill mechanism is blast or blast-fragment warhead at closest approach (see discussion of simple pursuit below,) actually hitting the target is not crucial. For laser illumination, target size alone dictates CEPs of radius 0.5-1 m.

The mechanics of these projectiles depends on the illumination form. For radar systems, it is desirable that the trajectory deviate only slightly from the beam LOS to assure adequate reflected energy is reaching the seeker. Because this implementation is most often used in Air Defense applications, gravity becomes a significant factor because of the need to constantly increase altitude. For laser systems, and assuming good diffuse reflectivity, approximately ballistic trajectories can be flown. Nonetheless, the most common analytic descriptions are range-time-of-flight (RTOF) curves for Air Defense systems, and either average velocity or RTOF curves for laser illumination.

31.5 Terminal Homing

The other primary category of guidance is terminal homing. In this category, the projectile carries a seeker that senses a target signature and guides the projectile to the target. These seekers may be either active (usually radar or mm wave) or passive (usually visual, infrared, or mm wave.) Obviously active seekers generate their own target signature while passive seekers use a naturally occurring target signature. This is why most passive seekers operate in the infrared, making use of the natural heat signature of a powered, human environment. Active seekers looking at ground targets must contend with the problem of false targets while passive seekers must deal with both naturally occurring false targets and background clutter that may mask the target. In general, seekers fall into two categories: non-imaging and imaging.

31.5.1 Non-Imaging Seekers

In general, non-imaging seekers sense (hopefully) only one predominant illumination source in their FOV, and this is assumed to be the target. The seeker generates guidance commands that guide the projectile in the direction of that illumination source. Obviously, this system can be very accurate so long as the illumination source has an angular subtense that is a small part of the FOV. Since the target does not fill the FOV until the projectile is very close to it, a common technique is to discontinue track once the target has some critical angular subtense and fly the projectile inertially.

The common application of non-imaging seekers is Air Defense and these seekers tend to operate in the infrared, classically in the 3-5 μ band, utilizing the hot exhaust of the aircraft exhaust as a signature. Some seekers now operate in the 8-14 μ band for engagement of helicopters and unmanned aerial vehicles (UAVs) whose signature has a lower temperature and thus is peaked at longer wavelength. Because of the lower signature energies produced, these implementations tend to be somewhat shorter in range.

31.5.2 Imaging Seekers

As a rule, imaging seekers operate either in the visual or the infrared to take advantage of human experience in their design and to provide a close reliability between operator and projectile that what is seen by the operator is actually what is being engaged. In effect, these seekers are either TVs or FLIRs which do not have a displayed image, but process the focal plane signals directly. Occasionally, systems may implement imaging at mm wavelengths although this is primarily limited to large projectiles. Additionally, these systems are usually passive although the recent interest in imaging laser radar seekers is a notable exception.

In these systems, the target's image is tracked by the seeker and the projectile is guided towards the target. Because the target image has increasing resolvability with decreasing range, exceeding smart tracking algorithms amy allow more effective aim point selection, thus somewhat offsetting the problem with non-imaging seekers when the target fills the FOV.

31.5.3 Accuracy

The accuracy of these systems is inherently no better than those described above, and frequently worse, primarily because of the problem of the target filling the FOV. This is not critical for Air Defense applications, which are the primary application of non-imaging seekers, because of the lethality mechanism employed. This problem is somewhat mitigated in imaging seekers and they are increasing used against ground targets, helicopters, and UAVs. Because the accuracy of these seekers is essentially range independent (after initial boost of the projectile,) they effectively have a constant probability of hit from minimum effective range (post boost) to maximum effective range (kinematic limit.)

31.5.4 Mechanics

The mechanics of these projectiles is somewhat a matter of design and target nature. This is most aptly demonstrated by considering a simple case in the mechanics of pursuit, that is, the trajectory described by a projectile homing on a target. We restrict ourselves to a simple two-dimensional problem with neither gravity nor drag.[3] The geometry of the problem is sketched in figure 4.

The target has a trajectory given by

$$\begin{aligned} x_t(t) &= x_t, \\ y_t(t) &= v_t t, \end{aligned} \tag{6}$$

where: x = cross range, and y = down range in the figure. Note that the cross range is kept constant. (This is sometimes known as a Fat, Dumb, and Happy trajectory.) A trajectory with varying y would not greatly add to our example. The projectile always steers toward the target with constant velocity v_p. As a result, the tangent from the projectile trajectory curve to the target trajectory (at any given instant of time,) is

$$\frac{dy_p}{dx_p} = \frac{y_t - y_p}{x_t - x_p}, \tag{7}$$

where we have dropped the time dependence for brevity. In equal time increments, the target and projectile each move according to their individual (but constant) speeds, so we may write

$$\frac{dx_p^2 + dy_p^2}{v_p^2} = \frac{dx_t^2 + dy_t^2}{v_t^2}. \tag{8}$$

If we now define $k \equiv \frac{v_p}{v_t}$, then we may rewrite equation 8 as

$$1 + \left(\frac{dy_p}{dx_p}\right)^2 = k^2 \left[\left(\frac{dx_t}{dx_p}\right)^2 + \left(\frac{dy_t}{dx_p}\right)^2\right] \tag{9}$$

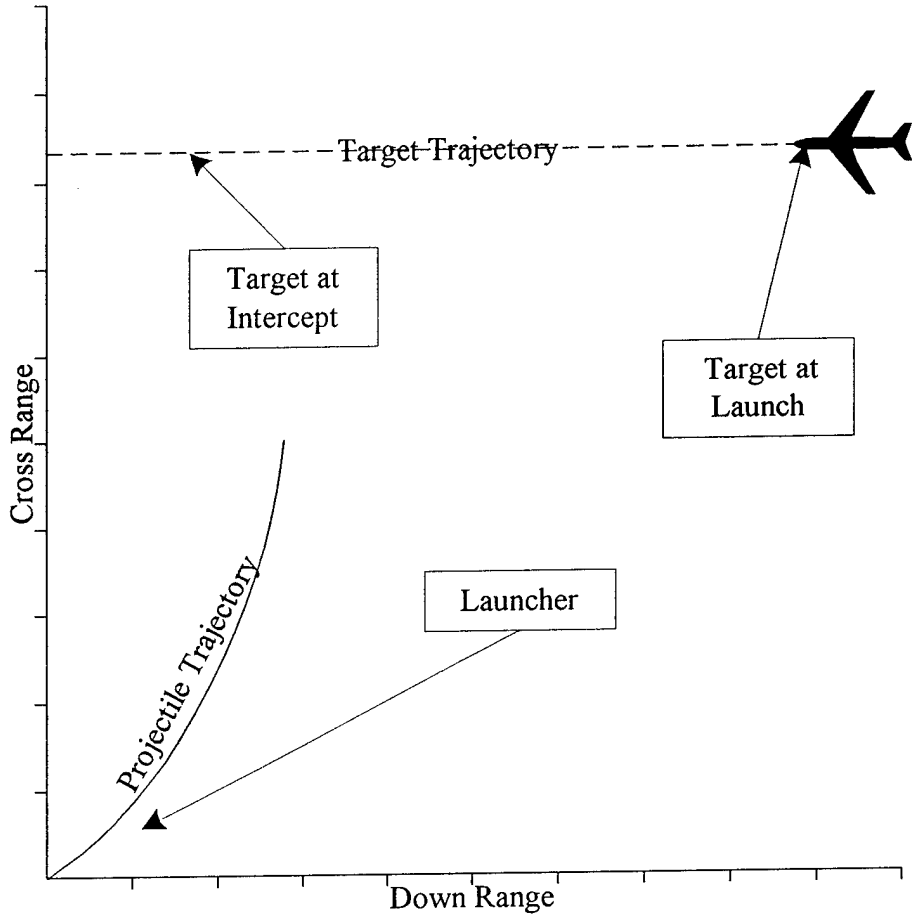


Figure 4: Pursuit Trajectory Geometry

$$= k^2 \left(\frac{dy_t}{dx_p} \right)^2,$$

since x_t is a constant.

Now, we rewrite equation 7 as

$$y_t = y_p + (x_t - x_p) \frac{dy_p}{dx_p}, \tag{10}$$

and differentiate this wrt x_p, yielding

$$\frac{dy_t}{dx_p} = \frac{dy_p}{dx_p} + (x_t - x_p) \frac{d^2 y_p}{dx_p^2} - \frac{dy_p}{dx_p} \tag{11}$$

$$= (x_t - x_p)\frac{d^2y_p}{dx_p^2},$$

again since x_t is a constant. After combining equations 9 and 11, with a bit of algebra, we get

$$\sqrt{1+\left(\frac{dy_p}{dx_p}\right)^2} = k(x_t - x_p)\frac{d^2y_p}{dx_p^2}. \tag{12}$$

Since this differential equation does not contain y_p explicitly, but only its derivatives, we may define the variable $p \equiv \frac{dy_p}{dx_p}$, which changes equation 12 to

$$\sqrt{1+p^2} = k(x_t - x_p)\frac{dp}{dx_p}. \tag{13}$$

This equation may be rearranged and integrated to yield an equation for p as a function of x_p. Using the definition of p, this equation can then be integrated again to get

$$y_p = \frac{kx_t}{k^2-1} + \frac{kx_t}{2(k^2-1)}\left[(k-1)\left(1-\frac{x_p}{x_t}\right)^{1+\frac{1}{k}} - (k+1)\left(1-\frac{x_p}{x_t}\right)^{1-\frac{1}{k}}\right]. \tag{14}$$

This is essentially a state space trajectory, which becomes undefined if the speeds of projectile and target are the same. This is why it is generally desirable to design the projectile with as high a speed as possible; and also why it is possible for aircraft to outrun Air Defense missiles, if they can prolong the engagement long enough, the missile runs out of speed due to drag and gravity. Representative curves for $k = 3, 6, 9$ are shown in figure 5. The improvement with increasing speed is readily apparent.

We need to make several comments about this equation. First, it is an exceedingly simple treatment of pursuit, neglecting both drag and gravity. Additionally, Air Defense missiles normally do not fly this type of trajectory. They incorporate two advances. First, they steer to where they predict the target will be when the missile reaches the target - this reduces the flight time. Second, particularly if they use IR seekers, they steer ahead of the apparent target. Since they home on the exhaust and the exhaust is behind the aircraft, this corrects for a miss error that would occur if they didn't. Finally, we have not considered the more general case of the projectile flying out to meet the target.

This brings us back to our original comment about design. If k is very large as would be the case for a projectile to attack a ground target, (since it has a speed of less than 100 km $hr^{-1} \sim$ 28 m sec^{-1}), there is so much excess speed that it doesn't matter from a pursuit standpoint what the trajectory is, since they are essentially straight lines. Thus, it is not uncommon to force some shape on these trajectories for a variety of reasons such as safety and avoidance of countermeasures. For Air Defense applications, this is not always the case, and they tend to fly modified pursuit trajectories (as noted above.)

As a result of all this, common analytical practice is to express the trajectories of these weapons as RTOF curves. A special case of this is the drawing of intercept contours for Air Defense systems.

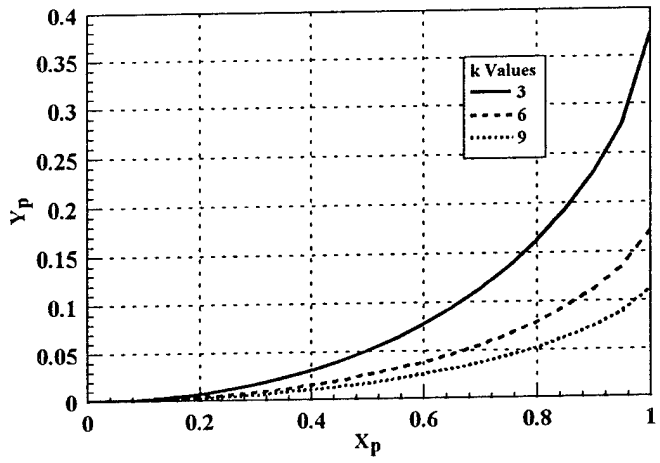


Figure 5: Representative Pursuit Curves

31.6 Warhead Guidance

An increasing common technique is to configure large projectiles (usually rockets) with payloads that are comprised of a number of guided warheads. The intent is to permit a long range projectile to deliver and attack several targets at range. These guided warheads are often called "Smart Munitions" and they have a wide variety of operational practices and design forced on them by the needs of operating without the benefit of human direction. We cannot cover the spectrum of these weapons here (although we shall explore them in greater detail later because of their special attritional features,) so the reader is referred to the literature for greater detail.[4] These systems usually are unpowered, incorporate an infrared or mm wave (or both) seeker(s), and have either HEAT or EFP warheads. These systems are commonly dispensed from a rocket at altitude but in the general target area (as dictated by rocket accuracy.) The systems then fall or glide until they find a target. (They are generally unpowered.) After finding a target, they either engage it immediately (EFP warhead) or glide into it. Since there are several of these in each payload, there is some risk of multiple detections of the same target.

These multiple detections have a random component, but since it is common to design the

FOV of the seekers to be large, all of the warheads often see all of the target area. Thus there is a systematic effect in selecting targets since the seekers are identical.

31.7 Time Lines

The attrition time lines associated with weapons vary enormously in form and geometry. In general, the more dense the weapon system, the shorter its time line. For small arms, there are relatively short time lines after target detection, although because the projectiles are inaccurate at long range, a large number of firings (and thereby long total times,) may be necessary. In general, the time lines of guided weapons are longer, often dominated by communication or organizational times. Thus, despite their higher accuracy and lethality, their lethality rates may be smaller than for unguided weapons.

31.8 Weather and Weapons

The most prevalent effect of weather on weapon performance is wind increasing the inaccuracy of projectiles. While steady wind can be measured and corrected for (to some extent), turbulent wind cannot. This effect is most pronounced for rockets which burn outside the launcher, especially as they are emerging from the launcher.

Weather has another effect on guided weapons. We have already discussed the effect of weather on sensors and how this reduces the effective range of the sensors. These degradations carry over to guided weapons. Thus, radar seekers are usually only degraded by snow or heavy rain while infrared seekers are degraded by fog and heavy rain.

From a military standpoint, heavy ground fog is a terrible weather situation since it degrades both target acquisition, and target engagement by guided weapons. While it has little effect on mobility, fog's degradation can actually be worse than rain's.

31.9 References

[1] Lee, R. G., T. K. Garland-Collins, D. E. Johnson, E. Archer, C. Sparkes, G. M. Moss, and A. W. Mowat, **Guided Weapons**, Brassey's Defense Publishers, London, 1988.

[2] Fowler, Bruce W., "Environmental Effects on Combat Performance: A Lanchester Approach", in L. G. Callahan, Jr., and Lawrence Low, Proceedings of the Workshop on Modeling, Simulation, and Gaming of Warfare, Callaway Gardens, GA, 2-5 December 1984, pp. 145-175.

[3] Davis, Harold T., **Introduction to Nonlinear Differential and Integral Equations**, Dover

Publications, Inc., New York, pp. 113-119.

[4] Heaston, R. J., and C. W. Smoots, "Introduction to Precision Guided Munitions", IIT Research Institute, Chicago, IL, GACIAC HB-82-01, May 1983.

Chapter 32

The Statistical Mechanics of Processes

32.1 Introduction

With this chapter, we turn aside from our direct path to the connection between fundamental principles and attrition to consider the basic mechanics of processes. Rest assured that this bytrip is necessary. It lays the foundation for what we shall do in the next chapter when we finally deal with the question of calculation of attrition rate constants (functions.)

Before we embark on discussions of processes, we must be a bit less cavalier about how we define systems. To this point, we have blithely talked about systems of sensors, weapons, even differential equations. We may describe a *system* as one or more components that perform one or more observable functions. Of necessity, the components must interact with each other in some manner.

A totally independent system is one which interacts with its external environment in at most two ways: initiation of system function (and this interaction is not necessary,) and production of a continuous, discrete, or repeated product. In general, the level of independence of a system is determined by the extent of its interaction with the external environment. This is the basis of the usual distinction of a system from a collection of components - that collection which maximizes internal component interaction and minimizes interaction between the collection and the external environment (that which is not part of the collection.)

From a physical standpoint, the only totally independent system that we can recognize is the universe as a whole. This is interesting in some contexts, but not particularly to us. For this reason, we tend to relax from the requirement of total independence.

Given this relaxation, we may consider a single atom to be a system consisting of electrons and a nucleus. (Or for that matter, the nucleus is a system comprised of elementary particles.) The atom interacts with its external environment in prescribed manners described by the laws of physics. Similarly, a crystalline solid is a system comprised of nuclei, bound electrons, and possibly, conduction electrons. The products of this system are occupation of space (and time), containment of heat, etc. Within the model that mass came into existence with the birth of the

universe, we may view the initiation of the system either with that birth, or with the formation of the system from its components.

The systems that we shall be interested in will frequently be material, but equally often organizational and/or human. They may be systems of organizational or individual behavior (like the psychophysics of detection.)

In general, systems may be comprised of subsystems just as the universe is comprised of galaxies which are... The usual distinction of subsystems is that they have decreasing independence wrt the external environment and the rest of the system. Given the lack of total independence, much of this distinction is somewhat subjective.

Why this interest in systems and subsystems? In general, systems, by approximating or restricting their external interactions, are amenable to aggregation. Aggregation may be described as the simplification of the description of either a system's internal functionality or its interaction with its external environment. This permits simplification. Physicists are familiar with aggregation by the differences among quantum mechanics, statistical mechanics, and thermodynamics. By its nature however, aggregation may not (usually isn't) reversible.

32.2 Process

One criterion of a system is that is executes (or participates in) one or more processes. If the system consists of subsystems, each subsystem executes (participates in) one or more processes. Just as systems may be divisible into subsystems, processes may be divisible into subprocesses. We usually assume that if a system can be described by a diagram relating its initiation, products, and internal and external interactions, then this diagram is a basis for relating the system's processes as well. Generally, this approach is limited by the complexity of the interactions. Controlling this complexity is inherent to the definition of the system.

As an example, detection is a process. It is initiated by some human motivation to search for targets; it is terminated by finding a target or by stopping (in the context of the chapter on Search and Detection.) The product of the process may be seen as either a detection or a stopping (with or without detection.) The process may be immediately re-entered or continued if multiple targets are sought.

In general, we shall diagram systems and their processes using boxes and arrows indicating flow. Processes are usually characterized by a pdf of process completion (product completion), so that processes will, in general, possess temporal moments. They may also have pdfs associated with the products (e.g., probability of detection.) Even when the processes are deterministic, we will ascribe them Dirac delta function pdfs for consistency of treatment.

There is one major distinction that we shall have to consider: the subsystems of a system, denoted by boxes and connected by arrows, may give rise to many more processes than subsystems. In particular, the transitions from one process to another may themselves be processes with durations, initiations (ending of parent process), and products (starting another process.) In

general, we shall try to be explicit when transitions between processes are themselves processes. For clarity, we may explicitly show these transitions as processes.

The interest that we have in systems and their processes in calculating attrition rate constants (functions) lies in their interactions and connectivity. Given an idealized collection of processes, and that they are simply and cleanly connected, which is a primary condition of relative independence, then there are two ways that simple processes can be connected: in series and in parallel. Of the parallel connections, there are two ways that they may interact:. We shall treat each of these in turn.

32.3 Laplace Transforms

Before proceeding with our discussion of processes, it is useful to spend a moment describing Laplace Transforms (LT). We shall follow the same general format as in Optics II.

Assume that $f(t)$ that is well behaved and defined for $t \geq 0$. Define the *Laplace Transform* of $f(t)$ as

$$\begin{aligned} L[f(t)] &= \int_0^\infty e^{-st} f(t)\, dt, \\ &\equiv f(s). \end{aligned} \tag{1}$$

The *Inverse Laplace Transform* is

$$f(t) = \frac{1}{2\pi i} \lim_{\beta \to \infty} \int_{\gamma - i\beta}^{\gamma + i\beta} e^{st} f(s)\, ds, \tag{2}$$

which, unlike the Inverse Fourier Transform, is not symmetric and is considerably more difficult to perform than the transform itself. (For this reason, a common practice with LT is to invert it by wimp inspection of tables of transforms rather than by macho integration.) As before, $i \equiv \sqrt{-1}$.

The LT also obeys the convolution theorem,

$$L\left[\int_0^t f_1(t - t') f_2(t')\, dt'\right] = f_1(s) f_2(s). \tag{3}$$

Derivatives have the transform

$$L\left[\frac{d^n f(t)}{dt^n}\right] = s^n f(s) - \sum_{j=0}^{n-1} s^j \left.\frac{d^{n-j-1} f(t)}{dt^{n-j-1}}\right|_{t=0}, \tag{4}$$

and repeated integrals have the transform

$$L\left[\int_0^t dt_n .. \int_0^{t_2} dt_1 f(t_1)\right] = s^{-n} f(s). \tag{5}$$

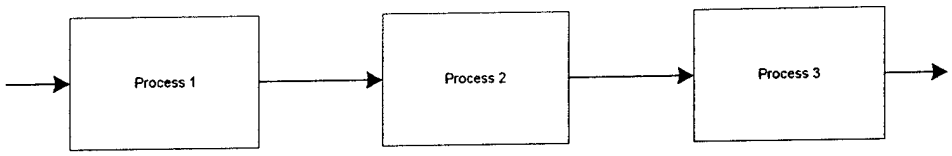


Figure 1: Serial Process Geometry

Also useful are the moments,

$$L\left[t^n f\left(t\right)\right] = \left(-1\right)^n \frac{d^n f\left(s\right)}{ds^n}. \tag{6}$$

While multi-dimensional LT are possible, they seldom arise so we shall not treat with them here.

32.4 Serial Processes Mechanics

The basic geometry of serial connectivity of processes (or just serial processes,) is sketched in figure 1. Three processes are shown. For simplicity, we shall assume that the transitions between processes are not processes and are instantaneous. Thus, the situation we have is that Process 1 must conclude for Process 2 to begin, and Process 2 must conclude for Process 3 to begin. We further assume that each process has associated with it a pdf of completion of the process associated with it. We designate these pdfs by $p_i\left(t\right)$ where the subscript indicates process number. The mathematics of this type of connectivity of processes is described by Renewal Theory. [2]

By now, it should be obvious the cumulative probability that process i has completed by time t is just

$$P_i\left(t\right) = \int_0^t p_i\left(t\right) dt, \tag{7}$$

and the moments are given by

$$\left\langle t^n \right\rangle = \int_0^t t^n p_i\left(t\right) dt. \tag{8}$$

This allows us to compute the expected time to completion and its standard deviation (e.g.,) for each of the processes. The problem we would like to address is how to calculate the expected time to completion for the system of processes.

To address this, we start by considering only the first two processes. Consider that if we want to know the probability that processes one and two have occurred by time t, then we are interested in all possible products of the probability that process one has completed in time t_1

and the probability that process two has completed in time t_2, such that $t_1 + t_2 = t$. We may write this as a probability integral by using the probability of either process as a differential,

$$P_{12}(t) = \int_0^t P_1(t - t')\, dP_2(t'), \tag{9}$$

which is just the summation we wanted. By using the definition of the pdf, we may rewrite this using

$$\begin{aligned} dP_2(t') &= \frac{dP_2(t')}{dt'} dt' \\ &= p_2(t')\, dt', \end{aligned} \tag{10}$$

which gives

$$P_{12}(t) = \int_0^t P_1(t - t')\, p_2(t')\, dt'. \tag{11}$$

If we further apply the definition, then we may differentiate this wrt t, and get

$$p_{12}(t) = \int_0^t p_1(t - t')\, p_2(t')\, dt'. \tag{12}$$

From this, we may infer that

$$\begin{aligned} p_{123}(t) &= \int_0^t p_{12}(t - t")\, p_3(t")\, dt", \\ &= \int_0^t p_3(t")\, dt" \int_0^{t"} p_1(t" - t')\, p_2(t')\, dt' \end{aligned} \tag{13}$$

This iterated form seems simple, but can become especially complicated unless the pdfs are all identical so that we may use the iterated integral trick. A more useful technique is to make use of the LT convolution property, equation 3, to write this as

$$L\left[p_{\{i\}}(t)\right] = \prod_{\{i\}} p_i(s). \tag{14}$$

Of course, the sharp reader will now immediately declaim - "Yeah, but now you've got to calculate the inverse transform!" Actually, we don't because of the properties of the moments. To demonstrate this, we note that

$$\begin{aligned} \int_0^\infty t^n f(t)\, dt &= (-1) \frac{d^n}{ds^n} \int_0^\infty e^{-st} f(t)\, dt \Big|_{s=0} \\ &= (-1) \frac{d^n}{ds^n} f(s) \Big|_{s=0}, \end{aligned} \tag{15}$$

so that all we need to compute moments of the pdf is derivatives of the LT of the pdf. Since we are primarily concerned with getting the expected time, this is relatively easy. At this point, we specifically restrict the set $\{i\}$ to consist of N members to clarify the mathematical notation.

Since

$$p_{\{i\}}(s) = \prod_{i=1}^{N} p_i(s), \tag{16}$$

we may calculate the expected time to complete the system of processes as

$$\begin{aligned} \langle t \rangle_{\{i\}} &= -\frac{d}{ds} \prod_{i=1}^{N} p_i(s) \Big|_{s=0} \\ &= -\sum_{i=1}^{N} \prod_{\substack{j=1 \\ j \neq i}}^{N} p_j(s) \frac{d}{ds} p_i(s) \Big|_{s=0} . \end{aligned} \tag{17}$$

Since $p_i(s) \to 1$, $s \to 0$, this reduces to

$$\begin{aligned} \langle t \rangle_{\{i\}} &= -\sum_{i=1}^{N} \frac{d}{ds} p_i(s) \Big|_{s=0} \\ &= \sum_{i=1}^{N} \langle t \rangle_i . \end{aligned} \tag{18}$$

Thus, the expected time to completion for a system of serial processes is the sum of the expectation times of the individual processes. (I know this is an intuitive result, but isn't it nice to know for sure?)

32.5 Parallel Processes Mechanics

We now turn to the question of parallel processes. That is, processes that (usually) initiate simultaneously and execute in parallel. This is shown figuratively in figure 2. We assume that each of these processes has a pdf defined in the same manner as we used above for serial processes.

Given that all of these processes initiate simultaneously, and each has some duration of execution which is nominally (and usually actually) stochastic, what is the behavior of the system of processes as a whole? We may identify two cases of general interest with regard to this system of processes:

1. the parallel process system concludes when all processes comprising the system complete their execution; and

2. the parallel process system concludes when any process comprising the system completes its execution.

In other words, the first case is concluded when the last process concludes, and the second case is concluded when the first process concludes. We shall refer to the first case as "AND" and the second as "OR" in analogy with electrical circuitry.

32.5.1 AND Parallel Process Mechanics

The AND process has the cumulative probability of conclusion at time t,

$$P_{AND}(t) = \prod_{\{i\}} P_i(t), \tag{19}$$

which is just the product of the cumulative probabilities of conclusion for each constitutive process. We may calculate the pdf of the AND process by differentiating this equation wrt t. To clarify notation, we continue our convention that $\{i\}$ consists of N members. This allows us to write the pdf as

$$p_{AND}(t) = \sum_{i=1}^{N} p_i(t) \prod_{\substack{j=1 \\ j \neq i}}^{N} P_j(t). \tag{20}$$

We note immediately that this is not likely to be neatly expressible, nor likely to be neatly integrable. In principle, we could write equation 20 as

$$p_{AND}(t) = \frac{d}{dt} \prod_{\{i\}} P_i(t), \tag{21}$$

and write the expected time equation as

$$\begin{aligned} \langle t \rangle_{AND} &= \int_0^\infty t p_{AND}(t)\, dt, \\ &= \int_0^\infty t \frac{d}{dt} \prod_{\{i\}} P_i(t)\, dt. \end{aligned} \tag{22}$$

We may integrate this equation by parts to yield

$$\langle t \rangle_{AND} = t \prod_{\{i\}} P_i(t) \Big|_{t=0}^{t=\infty} - \int_0^\infty \prod_{\{i\}} P_i(t)\, dt. \tag{23}$$

This equation has two problems: first, in general, the upper limit of the first rhs term is infinite since $P_i(t) \rightarrow 1$ as $t \rightarrow \infty$; and second, the integral is still not obviously integrable even if we do know it is bounded. We can postulate that the value of the integral is dominated by the slowest of the processes, so in this sense, the equation is insightful.

In general, the most common processes of interest in attrition are Negative Exponentially Distributed (NED), so it is instructive to examine examples. Let each process have a pdf,

$$p_i(t) = \xi_i e^{-\xi_i t}, \tag{24}$$

so the cumulative probability of completion of each process is

$$P_i(t) = 1 - e^{-\xi_i t}, \tag{25}$$

which has expected time to completion of

$$\langle t \rangle_i = \frac{1}{\xi_i}. \tag{26}$$

Let us now consider a parallel process comprised of two subprocesses. The cumulative probability of completion is explicitly

$$P_{AND}(t) = 1 - e^{-\xi_1 t} - e^{-\xi_2 t} + e^{-(\xi_1+\xi_2)t}, \tag{27}$$

and the pdf is just

$$p_{AND}(t) = \xi_1 e^{-\xi_1 t} + \xi_2 e^{-\xi_2 t} - (\xi_1 + \xi_2) e^{-(\xi_1+\xi_2)t}. \tag{28}$$

The expected time to process completion is then just

$$\langle t \rangle_{AND} = \frac{1}{\xi_1} + \frac{1}{\xi_2} - \frac{1}{\xi_1 + \xi_2}, \tag{29}$$

which is not simply reducible. If we rewrite this in terms of the expected times of the subprocesses,

$$\langle t \rangle_{AND} = \frac{(\langle t \rangle_1 + \langle t \rangle_2)^2 - \langle t \rangle_1 \langle t \rangle_2}{\langle t \rangle_1 + \langle t \rangle_2}, \tag{30}$$

which is equally dismal. The limits of the relationship are somewhat cheerier. For example, for a disparity in the sizes of ξ_1 and ξ_2 (and we don't care which because of the symmetry of the equation.), say $\xi_1 \gg \xi_2$, then

$$\begin{aligned} \langle t \rangle_{AND} &\simeq \frac{1}{\xi_1} + \frac{1}{\xi_2} - \frac{1}{\xi_1} \\ &\simeq \frac{1}{\xi_2} \simeq \langle t \rangle_2 . \end{aligned} \tag{31}$$

In this limit,

$$\langle t \rangle_{AND} \simeq \max \left[\langle t \rangle_1 , \langle t \rangle_2 \right] . \tag{32}$$

Alternately, if $\xi_1 \sim \xi_2$, then

$$\langle t \rangle_{AND} \simeq \frac{3}{4} \left(\langle t \rangle_1 + \langle t \rangle_2 \right) , \tag{33}$$

which may be compared with the serial result!

32.5.2 OR Parallel Process Mechanics

The OR parallel process has a very complicated cumulative probability of completion since it is the sum of the cumulative probabilities of each subprocess completing, less the cumulative probabilities of any combination of subprocesses concluding simultaneously. A more compact way of viewing this is that of the cumulative probability of all the subprocesses not completing. One minus this probability is the cumulative probability of completion,

$$P_{OR}(t) = 1 - \prod_{i=1}^{N} \left(1 - P_i(t) \right) . \tag{34}$$

The pdf is then just

$$p_{OR}(t) = \sum_{i=1}^{N} p_i(t) \prod_{\substack{j=1 \\ j \neq i}}^{N} \left(1 - P_j(t) \right) , \tag{35}$$

which seems, on first inspection, equally daunting. In general, this is true, but if the general form of the subprocess cumulative probabilities is

$$P_i(t) = 1 - f_i(t) , \tag{36}$$

then the cumulative probability of the process is just

$$P_{OR}(t) = 1 - \prod_{i=1}^{N} f_i(t) . \tag{37}$$

If we now return to our NED example, this reduces simply to

$$P_{OR}(t) = 1 - \prod_{i=1}^{N} e^{-\xi_i t} , \tag{38}$$

which reduces simply to

$$P_{OR}(t) = 1 - e^{-\xi_{\{i\}} t} , \tag{39}$$

where:

$$\xi_{\{i\}} = \sum_{i=1}^{N} \xi_i. \tag{40}$$

This gives us a deceptively simple form for the expected time to process completion of

$$\langle t \rangle_{OR} = \frac{1}{\xi_{\{i\}}}. \tag{41}$$

Sadly, this does not translate into a simple form in terms of subprocess completion times. The simplest is for two subprocesses,

$$\langle t \rangle_{OR} = \frac{\langle t \rangle_1 \langle t \rangle_2}{\langle t \rangle_1 + \langle t \rangle_2}, \tag{42}$$

and greater numbers of subprocesses give even more complicated expected times.

32.6 References

[1] Churchill, Ruel V., **Operational Mathematics**, McGraw-Hill Book Company, New York, 1958.

[2] Ross, Sheldon M., **Applied Probability Models with Optimization Applications**, Dover Publications, Inc., New York, 1970.

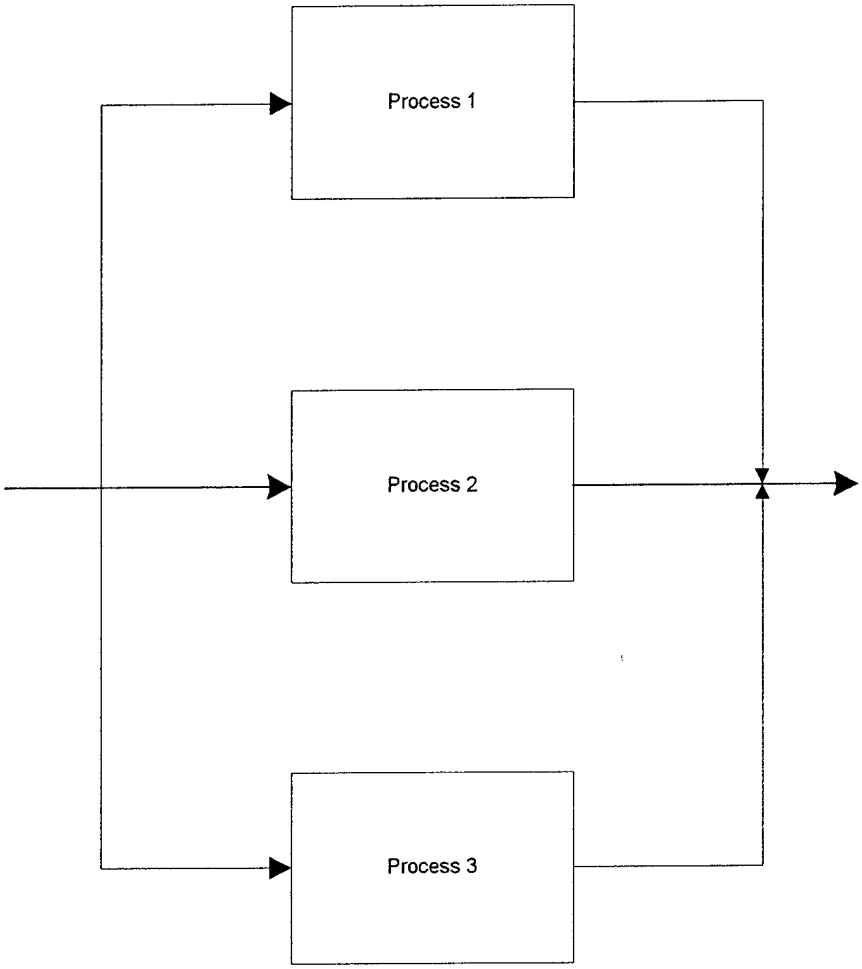


Figure 2: Parallel Process Geometry

This Page Intentionally Left Blank

Chapter 33

Modern Attrition Functionality

33.1 Introduction

It has been a dark and rainy night. With that variation on a cliche sentence, we emerge into the culmination of Part II of the work. Where Part I laid the foundation for basic Lanchester Attrition Theory, Part II has laid the foundation for the principle conjugate to Lanchester Attrition Theory, the theory of attrition rate coefficients (constants and functions,) which is the place where the scientific and technical descriptions of the performance of humans and materiel meet the description of attrition mechanics. In this chapter, we sketch how all of the components we have laboriously built in the preceding chapters of this Part come together to permit the construction of attrition rate coefficients (ARCs). With this construction, we shall relax, even discard, constraints and assumptions that have been held dear, in particular the assumption of constancy of ARCs.

33.2 The Attrition Process

At a top level, the common picture of the Attrition Process is of four subprocesses:

1. Target Acquisition;
2. Target Communication
3. Target Engagement; and
4. Target Assessment.

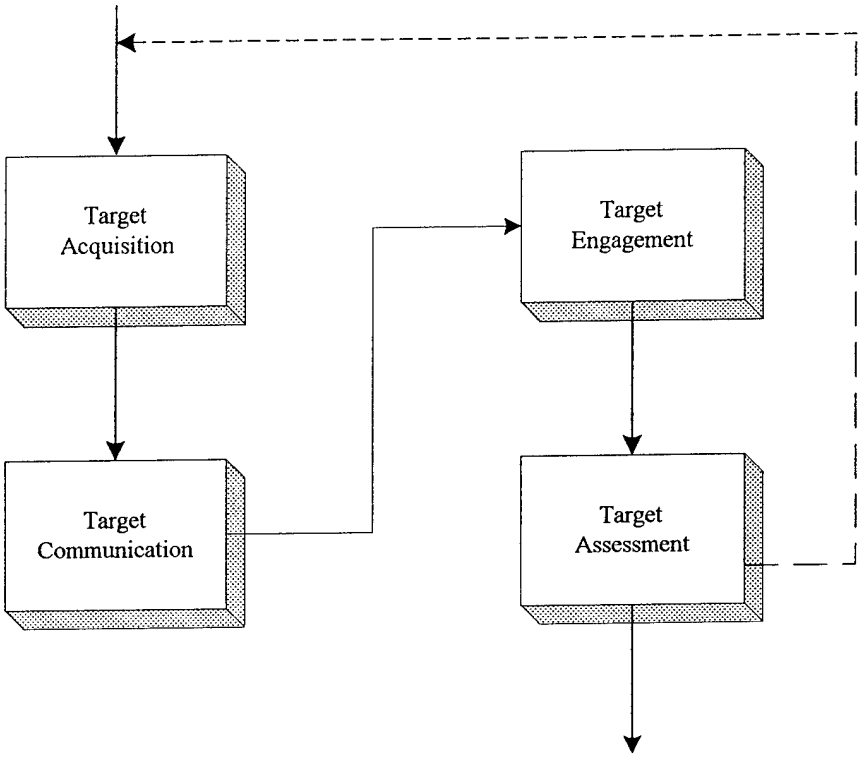


Figure 1: Attrition Process Subprocesses

These are shown as a simple serial process (to indicate the normal order of the subprocesses, but as we shall see, one which doesn't necessarily appertain) in figure 1.[1] The dotted line indicates that the process is commonly repeated.

33.2.1 Target Acquisition

As we promised in an earlier chapter, we now expand the scope of this term while remaining faithful to the original intent. The rationale for this expansion is that the scope we used previously was limited to the search and detection of targets within an engineering context. We expand this now to a broader scope that encompasses both engineering and tactical contexts.

We originally discussed acquisition within the context introduced by Overington: [1] detection; recognition; and identification. Subsequently, we expanded the context to that of Lloyd: [2] search; observation; detection; recognition; and identification. To bring this into a tactical context, we add two processes: orientation: and decision. This gives us:

[1] As always, we tend to err on the side of simplicity while conveying (hopefully) insight. This model of the attrition process is general, open at once to both internal modification and elaboration.

- Orientation (OR) - the recognition of the need for search and the definition of the search space;
- Search (S) - scanning the search space for targets (Chapter XXVII);
- Observation (O) - sensation of the search space images ("looking at");
- Detection (D) - the actual detection of targets (Chapter XXVIII);
- Recognition (R) - the recognition of targets;
- Identification (I) - the identification of targets; and
- Decision (DC) - the decision to note the target's existence for further action.

In keeping with the usual engineering approach, we treat detection, recognition, and identification as cumulative processes. Other than this, we shall treat these processes as occurring essentially simultaneously. In keeping with this, and our inherent acceptance that these processes are stochastic, we write a cumulative probability of Target Acquisition (TA) as

$$\begin{aligned} P_{TA} \simeq\ & P_{OR}P_{S|OR}P_{O|S,OR}P_{D|O,S,OR} \\ & P_{R|D,O,S,OR}P_{I|R,D,O,S,OR}P_{DC|I,R,D,O,S,OR} \end{aligned} \tag{1}$$

Except for the probability of orientation, all of these probabilities are conditional on the preceding probabilities. In normal usage, however, we shall treat these probabilities as independent (except as noted above for detection, recognition, and identification,) and therefore shall drop the conditional subscripts. This treatment is motivated by the differences in the physical nature of the processes and the commonly used (and accepted) mathematical forms. In addition, we note that there is some concern that once combat is joined, less attention is accorded to identification and recognition than would be indicated by field trials. We shall avail ourselves of this with (probably unwarranted and erroneous) frequency to simplify the mathematics. (The consequences of this are Fratricide which we shall discuss in a later chapter.)

Additionally, and as a general rule, we shall simplify our consideration of orientation and decision. Simply put, for our purposes, we shall take orientation as a given of probability one while recognizing that this factor is important under tactical situations such as ambushes and sentrance.[2] Accordingly we shall take this process as not contributing to the duration of Target Acquisition. Further, we shall also take as a given a definition of a search space, usually larger than a single FOV.

[2]Sentrance is a quarter word for the process of standing sentry. We know from psychological tests that sentries tire and become bored, neglecting their duties. ("Sleeping sentries will be shot!") There is some indication that vigilence is NED with a time scale that is long with respect to most attrition processes.

In a similar manner, we shall also take decision as a given. When we consider fire allocation and target prioritization (discussed in a later chapter in Part III,) this may add an additional amount of time to the expected time to Target Acquisition.

We may now consider the time to perform target acquisition. The model we adopt is basically that described in Chapter XXVII. In that chapter, we derived the mean number of glances for detection for random search with detection as

$$\langle i \rangle_{rs} = \frac{1}{p_r p_d}, \tag{2}$$

where: p_r = probability of looking at target (O), and p_d = single glimpse probability of detection. In this model, contributing probabilities are simple denominator multipliers (in keeping with the frequency picture of probability,) and the mean number of glimpses is converted to a mean time by dividing by the glimpse rate g ($\sim 3.3\,\mathrm{sec}^{-1}$). We shall use this model as an estimator of the mean time to target acquisition. As a further limitation, we shall consider only targets that are moving sufficiently slowly that they may be treated as stationary. Truly moving targets must either be treated otherwise or this formalism may be used as an approximation.

The keystone of our model will be a slightly modified version of the RCA [3] or Lloyd model for mean time to look at a target,

$$\langle t \rangle_{search} \simeq \frac{G}{700} \frac{\Omega_s}{\omega} \frac{\omega}{\Omega_t}, \tag{3}$$

where:

G = a congestion factor on [1,10];

Ω_s = the search solid angle, which may be larger than the FOV;

ω = the glance solid angle; and

Ω_t = target solid angle.

It is useful to make some minor modifications in this equation. First, since we know that the search solid angle is normally fixed (at least with respect to azimuth) by sectoring, the quantity $\frac{\omega}{\Omega_s}$ is essentially a parameter and we recognize it as the probability of looking at the target in the total search space, given there is only one target present.[3] We know from search theory that for a few, dispersed targets in a large search space, this probability is increased by the number of targets in the search space. Thus,

$$\langle t \rangle_{search} \rightarrow \frac{G}{700} \frac{\Omega_s}{n_{Es}\omega} \frac{\omega}{\Omega_t}. \tag{4}$$

If there are E enemy elements in an engagement/battle space, and the area of the space is A_E, then the approximate density of enemy elements is

$$\rho_E \simeq \frac{E}{A_E}. \tag{5}$$

[3]This introduces an approximaton. Basically, this neglects any time delay incurred in switching between FOV.

(We know that elements clump in the engagement space - Principle of Concentration, Span of Command/Control, etc., so in practice this approximation must not be used lightly and with gleeful abandon. A battle space of constant density degenerates into Lanchester's ancient combat.) Given this, we may approximate the area of the search space ($\Omega_s \ll 4\pi$) as

$$A_s \simeq \Omega_s R^2, \tag{6}$$

where: R = range of search space. This is nominally the visible range or visibility (or some equivalent for FLIR or radar) or the horizon/mask range, whichever is less.

The number of targets in the search space is then just the product of these two,

$$\begin{aligned} n_{Es} &\simeq \rho_E A_s \\ &= \frac{E\Omega_s R^2}{A_E}. \end{aligned} \tag{7}$$

It is also useful to rewrite the target solid angle explicitly showing range dependence. This is

$$\Omega_t \simeq \frac{a_t}{r^2}, r \leq R, \tag{8}$$

where: a_t = (perceived) target area, and r = range to target. This is approximate since we have clumped all of our targets together at the same range. This violates the assumption that the targets have a density but since the search space is an angle, it is usually acceptable since very few targets are at the shorter ranges. To be more accurate, we should average over the positions of the individual targets to get a mean range, or (better yet?) treat them individually - which violates our desire for simplicity.

If we now substitute these last two equations into the expected time to search, we get

$$\langle t \rangle_{search} \simeq \frac{G}{700} \frac{A_E}{ER^2} \frac{r^2}{a_t}, \tag{9}$$

which we rewrite as

$$\langle t \rangle_{search} \simeq \frac{G}{700} \frac{1}{Ed} \frac{r^2}{R^2}, \tag{10}$$

where: $d \equiv \frac{a_t}{A_E}$, is a quantity that represents the ratio of enemy target size to occupied ground (locally). To coin a term, we shall call this quantity the *dispersivity* since the quantity Ed is just the fraction of occupied ground physically covered by targets and the quantity $\frac{1}{\sqrt{Ed}}$ is approximately the distance (in a mean sense) between targets, normed to the root mean square target dimension. We belabor this point since the selection of what occupied ground means has enormous impact on this value. This quantity also has another enormous advantage. Since it is a ratio and therefore dimensionless, it eases the problem of scaling to linear (rather than area) formations or to volumes (for aircraft.) We note in conclusion that since $A_E \sim km^2$

and $a_t \sim m^2$, then we expect $d \sim 10^{-4} - 10^{-6}$. Also note that this expected time is range dependent and already contains the glance rate. (Thus for non-human search, the time must be scaled accordingly. While mechanized search will probably be scanned or raster search we have already seen that the expected time with uncertain detection are approximately the same for the two techniques.)

We may now proceed to consider the remainder of the terms that comprise the expected time to target acquisition. The derivation immediately above addresses search. The probability of observation, assuming that the observer is neither tired nor bored, and actually "looks at" the target, is determined by the probability that the target can be "looked at". At its simplest, this is nothing more than the probability that LOS exists, $P_{LOS}(r)$, Chapter XXIII.

The probability of detection should be the single glimpse probability of detection, but since we do not have a coherent theoretical way of calculating this, we must develop an approximation. As we recall from Chapter XXIX, Rotman [4] suggested that

$$gp_r p_{\text{det}} \simeq \frac{P_d(\infty)}{3.4}, \tag{11}$$

where:

g = glimpse rate;

p_r = probability of looking at the target;

p_{det} = single glimpse probability of detection; and

$P_d(\infty)$ = infinite time probability of detection.

Since most of the field trials used to develop this empirical relation were confined to single FOV searches, we may approximate the probability of looking at the target as

$$p_r \simeq \frac{\omega}{\Omega_{FOV}}. \tag{12}$$

The glimpse rate is approximately 3.3 sec^{-1}, so we may approximate the single glimpse probability of detection from the infinite time probability of detection as

$$p_{\text{det}} \simeq \frac{P_d(\infty)\,\Omega_{FOV}}{11\omega}. \tag{13}$$

This equation looks dangerous since we may conceive of situations where the single glimpse probability of detection is greater than the infinite time probability of detection. For an eye FOV of about $3°x3°$ this would occur for an FOV $> 100\,\text{deg}^2$. For this reason, a alternate approach based on cumulative probability would be

$$p_{\text{det}} \simeq 1 - (1 - P_d(\infty))^{\frac{1}{m}}, \tag{14}$$

where: m = number of actual glimpses at the target for detection. Note that this is just an inversion of the way we initially approached search with detection in the first place (and is therefore really dependent on something akin to stopping - see Chapter XXVII.) For $P_d(\infty) \ll 1$, as would be the case in Rotman's data, this gives

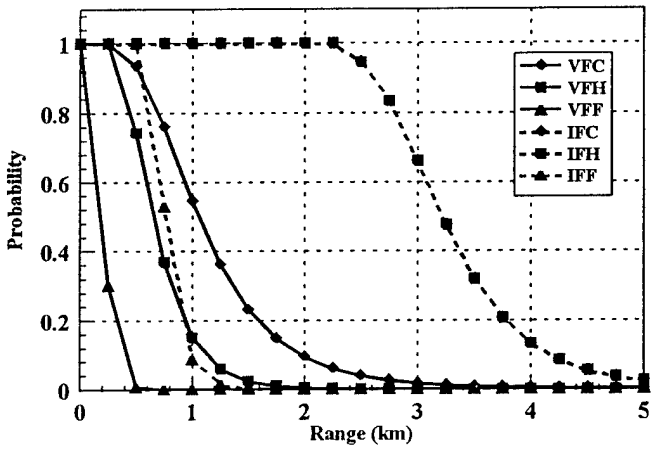


Figure 2: Eye-FLIR Single Glimpse Probability of Detection (Fully Exposed Target)

$$m \simeq \frac{11}{p_r}, \tag{15}$$

and since $\frac{1}{p_r} \simeq$ the mean number of looks to see the target once, we may estimate a single glimpse probability of detection, independent of search, as

$$p_{\text{det}} \simeq 1 - (1 - P_d(\infty))^{\frac{1}{11}}. \tag{16}$$

The value 11 is sufficiently empirical that we may probably risk the neat approximation $11 \simeq 10$, and to make that replacement. If we apply this scheme to the same weather dependent probability of detection curves for eye and FLIR that we presented in Chapter XXVIII, figures 8 and 9, then we may calculate approximate single glimpse probabilities of detection, as shown in figures 2 and 3

Since recognition and identification are cumulative from detection, we may apply the same technique to cumulative detection-recognition, or detection-recognition-identification curves as we have above. (But we cannot apply it independently since they are cumulative.) As this adds little to the derivation, we shall skip this activity and leave it as an exercise for the student.

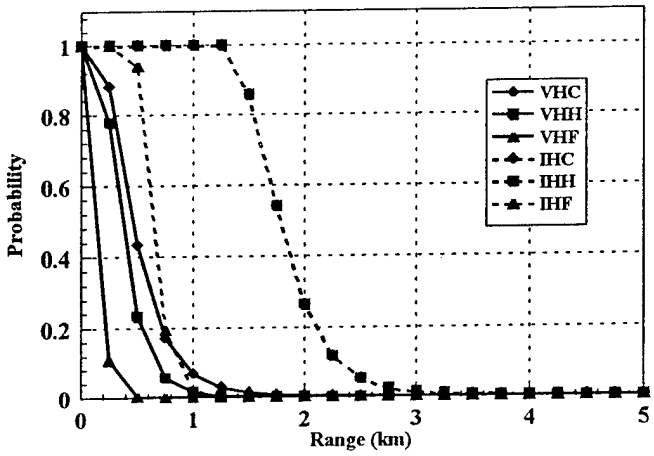


Figure 3: Eye-FLIR Single Glimpse Probability of Detection (Hull Defilade Target)

We are now finally ready to address the problem of expected time to target acquisition. Based on our assumption of the frequency picture of probability, we may use the mean time to search, equation 10, as the basis for mean time for target acquisition. Since we view each inverse probability as the mean number of repetitions of each process, we may estimate the mean time to target acquisition as

$$\langle t \rangle_{TA} \simeq \frac{1}{P_{OR}} \left[\frac{G}{700} \frac{1}{Ed} \frac{r^2}{R^2} \right] \frac{1}{P_O P_D P_R P_I P_{DC}}. \tag{17}$$

Since we have assumed the probabilities of orientation and decision to be one, to represent detection by an approximate single glimpse probability of detection (and to effectively ignore recognition and identification,) and probability of observation as probability of LOS, this reduces to

$$\langle t \rangle_{TA} \simeq \left[\frac{G}{700} \frac{1}{Ed} \frac{r^2}{R^2} \right] \frac{1}{P_{LOS}(r)\, p^{\star}_{\text{det}}(r)}, \tag{18}$$

where we have indicated probability of detection by a $\star$ to indicate that (i) it is single glimpse,

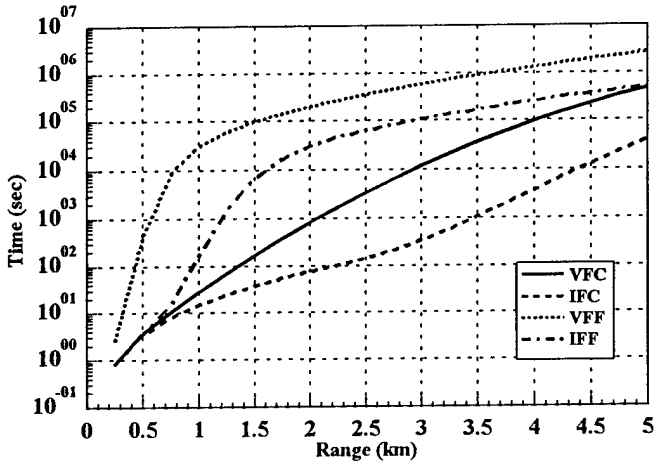


Figure 4: Representative Expected Times to Target Acquisition - 1 Target

and (ii) that this may actually be detection-recognition, or detection-recognition-identification.

For future use, it is convenient at this point to define a specific target acquisition time function,

$$\Phi_{TA}(r) \equiv \frac{700dR^2}{Gr^2} P_{LOS}(r)\, p^{\star}_{\text{det}}(r)\,, \tag{19}$$

so that we may write the expected time in the compact form

$$\langle t \rangle_{TA} \simeq \left[E\Phi_{TA}(r)\right]^{-1}, \tag{20}$$

where we have explicitly separated the force strength (E) and range (and other variable) dependent parts. A plot of the inverse of $\Phi_{TA}(r)$ for selected of the preceding figure data (and other representative parameters,) is shown in figure 4. This is effectively the time for target acquisition of one target and may be approximately compared with figure 1 of Chapter XXIX.

33.2.2 Target Communication

Having found the target and decided that it needs to be engaged, the next Attrition Subprocess is normally Target Communication. This is an enormously variable subprocess in terms of

its expected time to completion, depending both on the weapon system in question and the organization and doctrine for exercising the system. We approach this subprocess by examples.

For a solitary infantryman, target communication is measured in fractions of a second, essentially being the amount of time elapsing from the decision that a target needs engaging and the generation of the nerve impulse to bring the infantryman's weapon to bear. (Note that it does not include the time to bring the weapon to bear - that is part of engagement!) Thus, for this solitary infantryman, the mean time for target communication is effectively zero.

For a unit of infantry, in ranks, and to engage a target in mass using Napoleonic tactics, target acquisition is performed by either an officer or a non-commissioned officer. Having made a decision to engage a target, the officer (e.g.,) must direct the unit to the firing position. This may take considerable time depending on the distance to be traveled.

In modern tanks, the commander usually acquires targets and then hands them off to the gunner for engagement. If the gunner is not currently engaged at the time, target communication is usually a matter of only a few seconds. If the gunner is engaged, then somewhat more time may be required, necessitating that the commander return to target acquisition.

In modern indirect fire, a distinction must be made between direct and general support, and a special consideration must be mentioned about the target acquisition decision subprocess. Another key consideration is whether the searcher has authority to request or authorize fire directly.

Fundamentally, in indirect fire, communication must occur between the observer and the actual weapons' site. Ideally, this is a direct communication over a real time communication link (radio or telephone.) It may however, entail several links of command and organization are intermediary, resulting in additional time. Finally, if the weapons are in direct support, they immediately begin target engagement while if they are in general support, engagement may have to wait for other, previous and equal or higher priority engagements to proceed.

As a result of this, target communication for indirect fire may be the greatest contributor to the mean time to kill. In general however, target communication has much less impact for direct fire systems than for indirect fire systems.

Despite the paradox of negative times, it is common to represent communication processes with normal distributions and until recently, via serial subprocess channels. Thus the mean times to target communication are simply

$$\langle t \rangle_{TC} = \sum_{subprocesses} \langle t \rangle . \tag{21}$$

There is one additional complication that we need to address which complicates the subprocess picture. This is shown in figure 5. For indirect fire, it may be necessary to establish communication directly between weapon site and observer when adjusted fire is necessary. Ranging fire may then occur until rounds are landing in the desired area. (The fire is adjusted.) At this point concerted fire is mounted. The complication arises in that what is commonly thought of as engagement (see below) actually may occur in communication. Where does this go in the

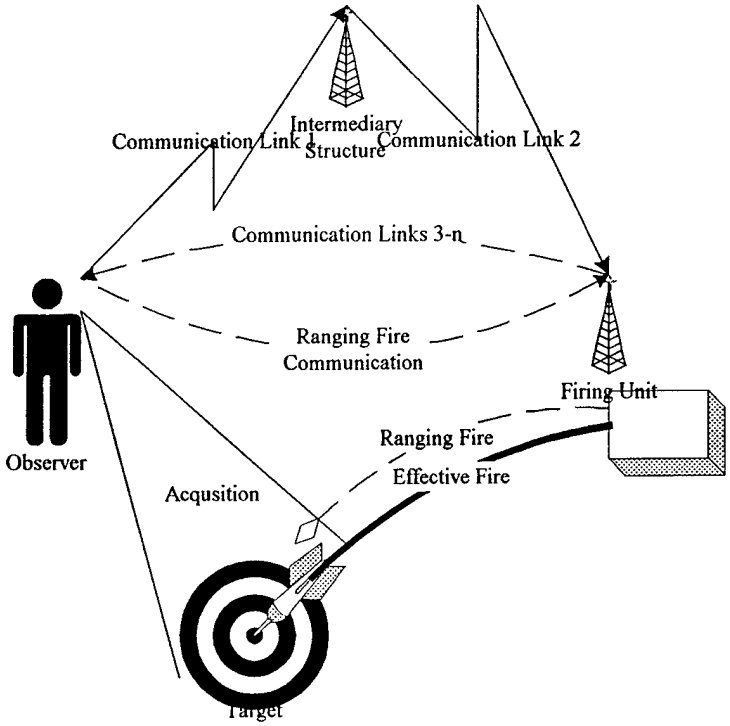


Figure 5: Subprocess Linkages for Adjusted Fire

subprocesses?

The answer is simple. Handle it however you need to. If all you are interested in is the mean time to kill, then so long as all the mean times are combined properly, there is no need to follow this model slavishly. You can conceptually extend the communication subprocess to include the ranging (adjusted) fire; extend the engagement subprocess to include the ranging fire communication; or even insert another subprocess called (e.g.,) Communication-Engagement.

33.2.3 Target Engagement

In general, target engagement is a simpler subprocess in principle than the previous two, but it has a more complicated interface with Target Assessment. Having received the communication to engage a target, the weapon must be loaded (if it is not already,) and trained on the target. It is aimed and fired at the target. At this point, it is usual to determine whether the target has been killed for direct fire systems. For indirect fire systems, some set amount of fire may be brought before this assessment is performed. If the target is killed, the usual effect is to reinitiate target acquisition; if not, engagement is reinitiated.

In engineering level simulations, the actual pointing of the weapon is often an observed

variable so that the "actual" time to train the weapon may be simulated. For our aggregated case, we may assume that the weapon is oriented randomly at the beginning of the engagement with some distribution. A simple calculation of slew time to aim may then be averaged to get the mean time to train. After this, some time is required to aim the weapon and again, this has some distribution. It is possible that a redetection time may be necessary here to account for handoff from observer to operator (if they are different.) If a guided weapon is being used, it may be necessary to track the target for some time prior to launch to "set" the autopilot.

Unless the weapon is a "Fire and Forget" system, and the survivability of the firer necessitates departure (as with an attack helicopter,) it is normal to delay further action until it is determined whether a hit has occurred (of course, reload may occur during this delay.) If a hit occurs, then an additional delay may occur while kill assessment is made. Some armies have a doctrine of multiple shots before kill assessment is made. If the projectile misses, or kill assessment is deferred, engagement is reinitiated. If the engagement space is dense in targets, kill assessment may be deferred until the density is reduced, proceeding directly from engagement back into acquisition.

We shall defer consideration of engagement mean times to completion until the next subsection on assessment.

33.2.4 Target Assessment

We have already discussed much of the consideration of the assessment subprocess in the subsections above. As we have already indicated, the relationship between engagement and assessment is complex and muddled. Under ideal situations, one (or more) projectile(s) is (are) fired at the target. If a hit occurs, kill assessment is performed. Under dense target conditions, kill assessment may be deferred or dropped.

Because of the complexity of this interaction, several possible formulations for mean time to engage and assess are possible. For the idealized situation (which is a general example,) where assessment is performed every time a hit occurs, it is important to perform the separation of hit and kill probabilities (see Chapter XXXI.) In this case, the mean time to completion of engagement and assessment is

$$\langle t \rangle_{TEA} \simeq \langle t \rangle_{train} + \frac{1}{p_{k|h}} \left[\frac{\langle t \rangle_{aim} + \langle t \rangle_{flight} + \langle t \rangle_{hit}}{p_h} + \langle t \rangle_{TA} \right], \tag{22}$$

where:

$\langle t \rangle_{train}$ = expected time to train weapon;
$p_{k|h}$ = probability of kill given a hit;
$\langle t \rangle_{aim}$ = expected time to aim the weapon;
$\langle t \rangle_{flight}$ = expected time to fly the projectile, the inversion of a range-time-of-flight curve;
$\langle t \rangle_{hit}$ = expected time to determine whether a hit has occurred;
p_h = probability of hit, normally range dependent (see Weapons chapters); and
$\langle t \rangle_{TA}$ = expected time to perform kill assessment.

The interpretation of this equation is that training only occurs once, but aiming, flight delay, and hit assessment occur with each shot. The inverse probability of hit represents the number of shots to achieve a hit. Kill assessment occurs whenever there is a hit, and the inverse probability of kill given a hit represents the number of hit necessary to kill. This equation (as do most) ignores depletion of ammunition. If assessment does not occur, or several shots are made before assessment is made, the equation changes appropriately. We note that while this equation is range dependent, it does not depend on force strength for single target engagement systems. For this reason, we may again define a specific Target Engagement (and possibly Assessment) time function,

$$\Phi_{TEA}(r) \equiv \frac{1}{\langle t \rangle_{TEA}}. \tag{23}$$

For indirect fire systems, the situation is similar, but may be very complicated, as we describe below.

33.3 Weapon System Mean Times to Kill

At this point, it is most useful to move to some examples of weapon system mean times to kill.

33.3.1 Infantryman

The problem of the individual or solitary infantryman engaging a single target by aimed (rather than pointed) fire is a useful place to start. By its nature, this is essentially a simple serial system so that subprocess expected times add. In this case, we may effectively ignore the communication time since they are just cognitive and reactive times. Thus the mean time to kill for a multishot weapon (M-1 class of rifle or better,) is

$$\begin{aligned} \langle t \rangle_k &\simeq \langle t \rangle_{TA} + \langle t \rangle_{TEA} \\ &\simeq \frac{1}{\Phi_{TA}(r)} + \frac{1}{\Phi_{TEA}(r)}. \end{aligned} \tag{24}$$

where: $\Phi_{TEA}(r)$ is given by equation 22. Note that if multiple targets are present, but only one is engaged at a time, then the mean time to kill simply becomes

$$\langle t \rangle_k \simeq \frac{1}{E\Phi_{TA}(r)} + \frac{1}{\Phi_{TEA}(r)}, \tag{25}$$

If the infantryman is engaging several targets at once with pointed fire, greater complexity occurs.

Pointed fire is normally used against multiple targets or when the exact location of a single target is not sufficiently well known to permit aimed fire. Pointed fire is characterized by multiple fire, usually in a burst mode, is limited to automatic weapons (or shotguns which we

shall leave as an exercise,) and is scanned either deliberately or through the natural movement of the weapon. (Automatic weapons tend to pull up and to the right.)

The evolution of pointed fire is interesting and worth a brief digression. With the introduction of the crossbow and early guns, fire was a mixture of pointed and aimed fire because of the lack of good sights (and a lack of knowledge of mechanics.) The introduction of guns has steadily liberalized and weakened control on the battlefield that led from massed ranks of infantry to a few skirmishers in front of the massed ranks (Napoleonic Era) to a few snipers in front and around of a lot of skirmishers in front of the massed ranks (American Civil War Era) to a force essentially comprised of a few snipers and the rest skirmishers (Modern Era.) The use of local cover and mask follows this evolution exactly. This evolution was the motivation for (and probably the result of) the development of automatic weapons which have two effects that cannot be realized with aimed, single shot at a time fire: engagement of multiple targets and snipers; and suppressive fire. We must defer consideration of suppressive fire until we consider Phase Aggregation in Part IV.

There are two complications that arise in calculating the expected time to kill a single target in pointed fire - both arise from the presence of multiple targets. If the angular increment per fire is (on average) θ, (which is presumed to be small,) then in a burst of n rounds fired, an angle of $n\theta$ is scanned covering an arc of approximate length $n\theta r$, and an area (above) on the ground of approximately

$$a_s \simeq \frac{n\theta r^2}{2}. \tag{26}$$

The density of targets (enemies) is given by equation 7, so the number of targets (enemies) in the scanned area is

$$\begin{aligned} E_s &\simeq \frac{a_s E}{A_E} \\ &\simeq \frac{n\theta r^2 E}{2A_E}. \end{aligned} \tag{27}$$

If the y standard deviation, σ_y, of the weapon's probability of hit is large compared to the spacing between rounds, θr; that is

$$\sigma_y \gg \theta r, \tag{28}$$

then the composite probability of hit of the burst is approximately uniform (except for tails outside the scanned area) across the scanned burst arc. In this case, the probability of hitting the target is approximately

$$p_h \simeq p_z \left[1 - \left(1 - \frac{w_E}{n\theta r}\right)^{E_s}\right], \tag{29}$$

where: w_E = width of enemy target, and p_z = vertical component of the probability of hit, given approximately by

$$p_z \simeq \frac{1}{\sqrt{2\pi}\sigma_z} \int_{-\frac{h_E}{2}}^{\frac{h_E}{2}} e^{-\frac{z^2}{2\sigma_z^2}} dz, \tag{30}$$

where: h_E = height of enemy, and assuming the fire is approximately pointed at the enemy mid-height. (Of course, since the normal scan is up and right, we should rightly average over point of fire across the arc.) If $h_E >\sim 3\sigma_z$, $p_z \sim 1$.

Regardless of this, if $w_E \ll n\theta r$, equation 29 simplifies to

$$\begin{aligned} p_h &\simeq p_z \frac{w_E E_s}{n\theta r} \\ &= p_z \frac{w_E r E}{2A_E}. \end{aligned} \tag{31}$$

If this quantity is small, then the number of kills per burst is approximately given by the number of bullets which hit and kill,

$$\begin{aligned} \delta E &\simeq n p_{k|h} p_h \\ &= n p_{k|h} p_z \frac{w_E r E}{2A_E}, \end{aligned} \tag{32}$$

assuming no enemy is struck twice.

We may now compute the expected time to kill one enemy. The TA time to completion is given by equation 20, the time to train and fire a burst is $\langle t \rangle_{burst}$, and the assessment time is $\langle t \rangle_{AS}$. If full TA is performed for each burst, then

$$\langle t \rangle_k \simeq \frac{1}{\delta E} \left[\frac{1}{E\Phi_{TA}} + \langle t \rangle_{burst} + \langle t \rangle_{AS} \right], \tag{33}$$

which we note depends on the square of enemy force strength. A more reasonable situation may be the firing of m bursts (perhaps limited by the number of rounds in the weapon so the process is reinitiated by reloading,) then the expected time to kill is

$$\langle t \rangle_k \simeq \frac{1}{m\delta E} \left[\frac{1}{E\Phi_{TA}} + m \left(\langle t \rangle_{burst} + \langle t \rangle_{AS} \right) \right].$$

An interesting variation is the case of the single shot weapon. While we tend to view this as the norm for pre-cartridge small arms: e.g., Napoleonic era warfare; it is also the situation with modern weapons such as LAWs. Since only one shot can occur per acquisition and engagement, the mean time to kill for a LAW is just

$$\langle t \rangle_k \simeq \frac{1}{p_{k|h} p_h} \left[\frac{1}{E\Phi_{TA}(r)} + \frac{1}{\Phi_{TE}(r)} \right], \tag{34}$$

where: $\Phi_{TE}^{-1}(r) \simeq \langle t \rangle_{train} + \langle t \rangle_{aim} + \langle t \rangle_{flight} + \langle t \rangle_{hit} + \langle t \rangle_{AS}$; and the leading probability fraction just represents the number of shots necessary to achieve a kill.

33.3.2 Tank

The tank is the attrition icon of the twentieth century. For a simple system operating with serial subprocesses, the mean time to kill is similar to that for the infantryman, except that target communication, while short, is often included,

$$\begin{aligned} \langle t \rangle_k &\simeq \langle t \rangle_{TA} + \langle t \rangle_{TC} + \langle t \rangle_{TEA} \\ &\simeq \frac{1}{E\Phi_{TA}(r)} + \frac{1}{\Phi_{TC}} + \frac{1}{\Phi_{TEA}(r)}, \end{aligned} \tag{35}$$

where: $\Phi_{TC}^{-1} = \langle t \rangle_{TC}$. A more interesting case is where acquisition-communication and engagement-assessment are performed in parallel. This is an AND system and it depends on whether there are many or few stowed kills on board. This is a somewhat complicated situation, so we depict it in figure 6. The first acquisition-communication subprocesses are performed, and then in all but the last engagement, are performed in parallel with engagement-assessment. For the last engagement, engagement-assessment is performed in serial. Thus, for the tank, the attrition process is modified to a mixture of serial and parallel subprocesses.

If the tank carries N stowed kills, then the expected time to kill N targets is

$$\begin{aligned} \langle t \rangle_{Nk} &= (\langle t \rangle_{TE} + \langle t \rangle_{TC}) \\ &\quad + (N-1) \left[(\langle t \rangle_{TE} + \langle t \rangle_{TC}) + \langle t \rangle_{TEA} - \frac{(\langle t \rangle_{TE} + \langle t \rangle_{TC}) \langle t \rangle_{TEA}}{\langle t \rangle_{TE} + \langle t \rangle_{TC} + \langle t \rangle_{TEA}} \right] \\ &\quad + \langle t \rangle_{TEA} \cdot \end{aligned} \tag{36}$$

If we divide this by N to get the expected time to kill a single target, and substitute explicitly for the times, we get

$$\begin{aligned} \langle t \rangle_k &\simeq \frac{1}{N} \left[\frac{E\Phi_{TA}(r) + \Phi_{TC}}{E\Phi_{TA}(r)\,\Phi_{TC}} + \frac{1}{\Phi_{TEA}(r)} \right] \\ &\quad + \left(1 - \frac{1}{N}\right) \left[\frac{E\Phi_{TA}(r) + \Phi_{TC}}{E\Phi_{TA}(r)\,\Phi_{TC}} + \frac{1}{\Phi_{TEA}(r)} \right. \\ &\quad \left. + \frac{E\Phi_{TA}(r)\,\Phi_{TC}}{E\Phi_{TA}(r) + \Phi_{TC} + E\Phi_{TA}(r)\,\Phi_{TC}\Phi_{TEA}(r)} \right] \end{aligned} \tag{37}$$

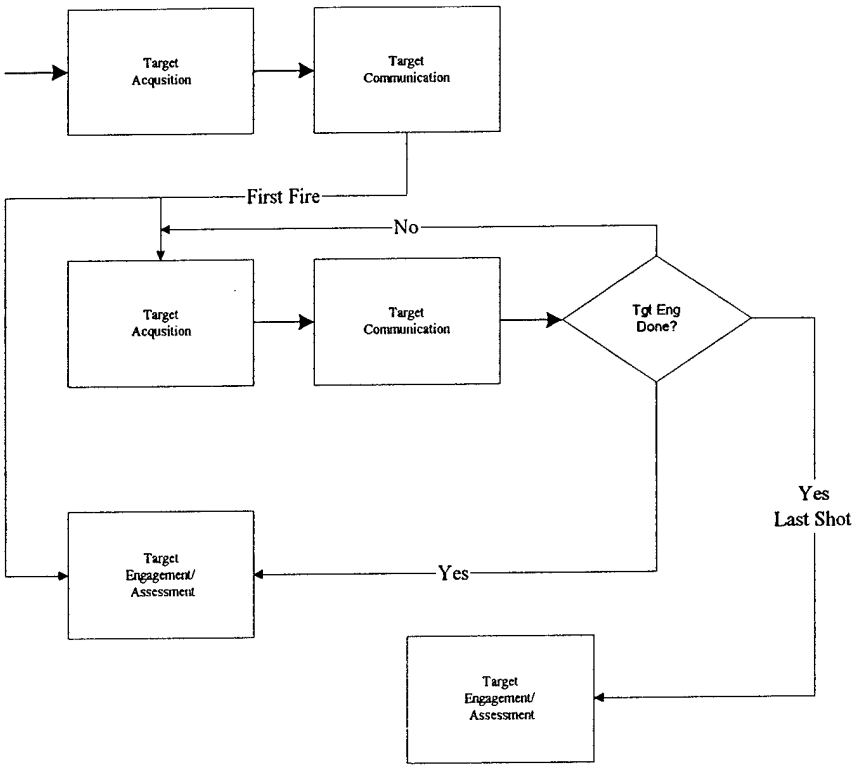


Figure 6: Parallel Tank Attrition Processes

$$= \frac{E\Phi_{TA}(r) + \Phi_{TC}}{E\Phi_{TA}(r)\,\Phi_{TC}} + \frac{1}{\Phi_{TEA}(r)} + \left(1 - \frac{1}{N}\right) \frac{E\Phi_{TA}(r)\,\Phi_{TC}}{E\Phi_{TA}(r) + \Phi_{TC} + E\Phi_{TA}(r)\,\Phi_{TC}\Phi_{TEA}(r)}$$

A useful approximation is to replace $\Phi_{TEA}(r)$ with the inverse rate of fire divided by the probabilities of hit and kill. If we also make the approximation that communication time is small compared to acquisition and engagement times, then this may be modified somewhat to

$$\langle t \rangle_k \simeq \frac{1}{E\Phi_{TA}(r)} + \frac{1}{\Phi'_{TE}} + \left(1 - \frac{1}{N}\right) \frac{1}{E\Phi_{TA}(r) + \Phi'_{TE}}, \tag{38}$$

which may be reduced somewhat to

$$\langle t \rangle_k \simeq \frac{\Phi'^2_{TE} N + 3\Phi'_{TE} E\Phi_{TA}(r)\,N + E^2\Phi^2_{TA}(r)\,N - E\Phi_{TA}(r)\,\Phi'_{TE}}{\Phi'_{TE} E\Phi_{TA}(r)\,N\,(\Phi'_{TE} + E\Phi_{TA}(r))}. \tag{39}$$

This has two limits, a very nonlethal one, $N \to 1$,

$$\langle t \rangle_k \simeq \frac{\Phi'_{TE} + E\Phi_{TA}(r)}{\Phi'_{TE} E\Phi_{TA}(r)}, \tag{40}$$

and a very lethal one, $N \to \infty$,

$$\langle t \rangle_k \simeq \frac{\Phi'^2_{TE} + 3\Phi'_{TE} E\Phi_{TA}(r) + E^2\Phi^2_{TA}(r)}{\Phi'_{TE} E\Phi_{TA}(r)\left(\Phi'_{TE} + E\Phi_{TA}(r)\right)}, \tag{41}$$

which is essentially the same as we would get for a serial process. We note in passing that considering these functions graphically is difficult and misleading since they are functions of force strength which under dynamic attrition conditions will change continuously. We could also apply the limiting cases identified in the preceding chapter.

33.3.3 Artillery

We have already indicated the problem of dividing communication and engagement in indirect fire attrition in the preceding section. For our purposes here, we shall make a decision to include ranging fire in the Target Communication Subprocess. In this case, the time to fire an indirect fire mission is simply

$$\langle t \rangle_{if} \simeq \langle t \rangle_{TA} + \langle t \rangle_{TC'} + \langle t \rangle_{TE} + \langle t \rangle_{TAS}, \tag{42}$$

where we have assumed that target assessment is performed before initiating another mission, and the modified nature of target communication is indicated by a prime. This form implicitly assumes that the desired effect is achieved. The more general form where N target engagements are required has the form

$$\langle t \rangle_{if} \simeq \langle t \rangle_{TA} + N\left(\langle t \rangle_{TC'} + \langle t \rangle_{TE} + \langle t \rangle_{TAS}\right). \tag{43}$$

The value of N can be estimated from the nature of the mission to be performed. We shall comment further on this below.

Let us now assume that n_r shots are fired during ranging, and n_e shots are fired during the engagement subprocess. If we designate the lethal area of a shot as a_L, then the probability of hitting a given spot more than once is just $\frac{a_L}{A_E}$,which we presume is small. (Note that this is a local definition of A_E. It also assumes that the standard deviations of shot hit are large.) The probability of multiple hits is then Poisson, and we may compute an adjusted probability of kill given a shot of

$$\begin{aligned} p^{\star}_{k|s}(n) &\simeq \sum_{m=0}^{\infty} \frac{1}{m!}\left[1 - \left(1 - p_{k|s}\right)^m\right]\left(\frac{a_L n}{A_E}\right)^m e^{-\frac{a_L n}{A_E}} \\ &= 1 - \exp\left[-\frac{a_L n p_{k|s}}{A_E}\right], \end{aligned} \tag{44}$$

where: $p_{k|s}$ = the single fire probability of kill given a shot inside the lethal area, and n = the number of shots. Of course, this assumes that the number of shots is sufficiently large that we may use the infinite limit form of the equation which, as we know, is analytically summable.

It then follows that the number of kills from the mission is approximately

$$\delta E \simeq (n_r + n_e)\, p^{\star}_{k|s}\, (n_r + n_e)\, \frac{a_L}{A_E} E, \tag{45}$$

where the last terms indicate the number of targets in the lethal areas. If we now combine this with the time to execute the mission, the expected time to kill a target is just

$$\langle t \rangle_k \simeq \frac{\langle t \rangle_{if}}{\delta E}. \tag{46}$$

It is appropriate to make one more caveat - this formalism ignores the overlap of subprocesses for separate missions.

33.3.4 Guided Weapons

The advantage of guided weapons is their vastly improved accuracy. This modifies the form of the mean time to kill by ways that we have already seen.

The oldest form of guided weapon is the direct fire missile, such as the ATGM. The missions of these systems tend to be single shot missions akin to what we have already seen for the LAW. This gives a mean time to kill a single target of

$$\langle t \rangle_k \simeq \frac{\langle t \rangle_{TA} + \langle t \rangle_{TC} + \langle t \rangle_{TEA}}{p_h p_{k|h}}, \tag{47}$$

where $\langle t \rangle_{TC}$ may be negligible depending on the mechanical and vehicle implementation of the system. We have already mentioned the situation with respect to air launched but ground acquired systems such as HELLFIRE. With modification for the additional time required (and the requisite range dependence,) the same mean time to kill equation is applicable.

A somewhat more complicated situation is presented by Lock-On After Launch systems such as the Fiber Optics Guided Missile. In this case, the mean time for a single mission is approximately

$$\langle t \rangle_{mission} \simeq \langle t \rangle_{TA} + \langle t \rangle_{TC} + \langle t \rangle_{Launch} + \langle t \rangle_{flight}. \tag{48}$$

In this case, TA refers to any target acquisition (usually of a collection of targets) by an observer away from the weapon system. This system enjoys two advantages in that it may fire from complete defilade and thus cannot be counter-engaged by its targets (except by general indirect fire,) and that it may conduct essentially simultaneous missions. If the system conducts N missions, then the mean time for N missions is approximately

$$\langle t \rangle_{Nmission} \simeq \langle t \rangle_{TA} + \langle t \rangle_{TC} + N \langle t \rangle_{Launch} + \langle t \rangle_{flight}. \tag{49}$$

Since the system operates by acquiring a specific target in the terminal phase of the trajectory, The time available for acquiring this specific target is approximately[4]

$$t_{specific} \simeq \langle t \rangle_{Launch} + \frac{\langle t \rangle_{flight}}{N}. \tag{50}$$

The single mission probability of kill is then just

$$p_{k|mission} \simeq P_d\left(t_{specific}\right) p_h p_{k|h}, \tag{51}$$

and the mean time to kill a single target is just

$$\langle t \rangle_k \simeq \frac{\langle t \rangle_{Nmission}}{N p_{k|mission}}, \tag{52}$$

which is just the fraction of the N mission mean time for one mission divided by the probability a mission will result in a kill - in other words, the duration of an "effective" mission times the number of missions for a kill.

33.4 Attrition Rate Coefficients

We now come to the culmination of the chapter, to the actual formulation of the ARCs. I must warn the reader that this will be rather disappointing because we have already laid the basis, and performed much of the work. The one piece remaining to bring into play is Bonder's Equation which we described in Chapter XXII back at the beginning of Part II. With the addition of this equation, all that is left is a bit of algebra, an occasional approximation, and cataloging of the ARCs for the examples we presented above.

Using the notation of this chapter, the mean time to kill one target is related to the attrition rate coefficient by Bonder's Equation,

$$\alpha = \frac{1}{\langle t \rangle_k}, \tag{53}$$

which is a slight revision of equation 4 of Chapter XXII. In this section, the relaxation of the constancy of the ARCs will be evident, and indeed central to our discussion in the next chapter where we make the leap to Attrition Differential Equations.

Before proceeding, we will make the assignment that the rate functions Φ that we introduced in this chapter can be defined for any single or combination of attrition subprocesses. This seeming complication of notation is made to simplify the notation in this section, which would otherwise become rather complicated in some cases. As a general rule, we shall explicitly separate the dependence on force strengths since this is exactly the dependence that shall consume

[4]This is a simplification for this system. The actual situation is considerably more complicated and we shall discuss this in Part IV.

our attention in the next chapter. Despite this, we shall suppress dependency notation to the maximum extent possible, again for the sake of simple notation.

33.4.1 Infantryman

In keeping with our scheme of examples from the preceding section, we begin by considering the Infantryman. The ARC for simple solitary aimed fire is just

$$\alpha \simeq \frac{E\Phi_{TA}(r)\,\Phi_{TEA}(r)}{E\Phi_{TA}(r)+\Phi_{TEA}(r)}, \tag{54}$$

which as we may see has a complicated dependence on enemy force strength E and range r. Parametric dependence on other parameters is hidden from sight. We note that if $E\Phi_{TA}(r) \gg \Phi_{TEA}(r)$, that is, if $\langle t\rangle_{TA} \ll \langle t\rangle_{TEA}$, then this reduces to

$$\alpha \simeq \Phi_{TEA}(r), \tag{55}$$

which indicates that target engagement and assessment are the critical time subprocesses. This may be recognized as essentially a Quadratic Lanchester ARC except for the range dependence. If the opposite situation holds, where target acquisition is the critical time subprocess, then the limit is

$$\alpha \simeq E\Phi_{TA}(r), \tag{56}$$

which is essentially a Linear Lanchester ARC.

The ARC for pointed fire is considerably more complicated,

$$\alpha \simeq E\Phi_{killed}\left[\frac{E\Phi_{TA}(r)}{E\Phi_{TA}(r)\left(\langle t\rangle_{burst}+\langle t\rangle_{AS}\right)+\frac{1}{m}}\right], \tag{57}$$

where:

$$\Phi_{killed} \equiv \frac{np_{k|h}p_z w_E r}{2A_E}, \tag{58}$$

and $m =$ the number of bursts fired. This is complicated but essentially proportional to enemy force strength squared, which is a behavior we have not seen in the classical theory. As above, this has two possible reductions. If target acquisition is the lengthy part, this reduces to

$$\alpha \simeq E^2\Phi_{killed}\Phi_{TA}(r)\,m, \tag{59}$$

while if assessment is the time consuming part,

$$\alpha \simeq \frac{E\Phi_{killed}}{\langle t\rangle_{burst}+\langle t\rangle_{AS}}. \tag{60}$$

The ARC for a LAW has essentially the same form as aimed fire,

$$\alpha \simeq \frac{E\Phi_{TA}(r)\,\Phi_{TEA}(r)\,p_{k|h}p_h}{E\Phi_{TA}(r)+\Phi_{TE}(r)}, \tag{61}$$

except for the inclusion of the probability of kill given a shot in the numerator. Note that the $\Phi_{TE}(r)$ is a purely time function as defined above.

33.4.2 Tank

The serial form of the ARC for a tank has the same form as for aimed fire above, except for the addition of communication time. This gives

$$\alpha \simeq \frac{E\Phi_{TA}\Phi_{TC}\Phi_{TEA}}{E\Phi_{TA}\Phi_{TC}+E\Phi_{TA}\Phi_{TEA}+\Phi_{TC}\Phi_{TEA}}, \tag{62}$$

which has limiting properties similar to what we have described above. A more interesting formalism is the parallel AND form which, if we ignore communication time as small (and therefore only important only at short ranges.) The nonlethal limit is

$$\alpha \simeq \frac{\Phi'_{TE}E\Phi_{TA}(r)}{\Phi'_{TE}+E\Phi_{TA}(r)} \tag{63}$$

and the lethal limit is

$$\alpha \simeq \frac{\Phi'_{TE}E\Phi_{TA}(r)\left(\Phi'_{TE}+E\Phi_{TA}(r)\right)}{\Phi'^2_{TE}+3\Phi'_{TE}E\Phi_{TA}(r)+E^2\Phi^2_{TA}(r)}. \tag{64}$$

We shall address these in greater detail in the next chapter.

33.4.3 Artillery

The ARC for artillery fire follows directly from its mean time to kill as calculated in the preceding section,

$$\alpha \simeq \frac{E\Phi_{kill}}{1+NE\Phi_{TA}\left(\langle t\rangle_{TC'}+\langle t\rangle_{TE}+\langle t\rangle_{TAS}\right)}. \tag{65}$$

This ARC also has two limiting forms, one Linear Lanchester and the other Quadratic Lanchester, as we would expect, but for different reasons. If $NE\Phi_{TA}\left(\langle t\rangle_{TC'}+\langle t\rangle_{TE}+\langle t\rangle_{TAS}\right) \gg 1$, then this reduces to

$$\alpha \simeq \frac{\Phi_{kill}}{N\Phi_{TA}\left(\langle t\rangle_{TC'}+\langle t\rangle_{TE}+\langle t\rangle_{TAS}\right)}, \tag{66}$$

which is clearly Quadratic Lanchester and generally says that the target density is low (target acquisition is slow.) If the opposite is true, then

$$\alpha \simeq E\Phi_{kill}, \tag{67}$$

which is clearly Linear Lanchester and generally says that target density is high.

33.4.4 Guided Weapons

The ARC for an ATGM has a form similar to that for artillery,

$$\alpha \simeq \frac{E\Phi_{TA}p_{k|h}p_k}{1+E\Phi_{TA}\left(\langle t\rangle_{TC}+\langle t\rangle_{TEA}\right)}, \tag{68}$$

which has the same limiting methodology. For the long ranges that we are primarily interested in for modern combat, this reduces to a Quadratic Lanchester ARC.

The ARC for a Lock-On has a similar form

$$\alpha \simeq \frac{NE\Phi_{TA}p_{k|h}p_kP_d\left(\langle t\rangle_{Launch}+\frac{\langle t\rangle_{flight}}{N}\right)}{1+E\Phi_{TA}\left(\langle t\rangle_{TC}+N\langle t\rangle_{Launch}+\langle t\rangle_{flight}\right)}, \tag{69}$$

which has the same types of limits as artillery. Note that the target acquisition is performed in the target area. This is the ARC for systems such as the Fiber Optics Guided Missile.

33.5 References

[1] Overington, J., **Vision and Acquisition**, Pentech Press, London, 1976.

[2] Lloyd, J. M., **Thermal Imaging Systems**, Plenum Press, New York, 1975.

[3] Radio Corporation of America, **Electro-Optics Handbook**, Solid State Division, Electro-Optics and Devices, Lancaster, PA, 1974.

[4] Rotman, S. R., E. S. Gordon, and M. L. Kowalczyk, " Modeling human search and target acquisition performance: III Target detection in the presence of obscurants", Optical Engineering 28 No. 6, June 1991, pp. 824-829.

This Page Intentionally Left Blank

Chapter 34

Modern Attrition Differential Equations

34.1 Introduction

We come at last to the final chapter of Part II. In the preceding chapter, we concluded our construction of the conjugate theory of Lanchester Attrition Theory, sketching the underlying foundation of physics, engineering, and psychology that supports the construction of Attrition Rate Coefficients (ARCs). There remain two tasks before we may close out the drama of Part II: closure with the dynamics of the Attrition Differential Equations (ADEs) that comprised the bulk of Part I; and opening the entry into Part III that will consider the dynamics of heterogeneous combat. The basic token of both these tasks is, of course, attrition differential equations.

This part of the work has laid a basis, albeit sketchily, for the calculation of ARCs, culminating in the preceding chapter. While interesting and insightful, the ARCs are not end products in and of themselves. Their ultimate utility is providing insight into the dynamics of the attritional aspects of military combat. These ARCs then are the empowerment of calculations using attrition differential equations constructed from them. These calculations then produce force strengths, LERs and FERs, and other parametrics of analysis to support weapon development, doctrine, force and organization structuring, and other planning, management, and leadership activities. To this end, we devote this chapter to consideration of the use of the ARCs in the framework of ADEs.

34.2 Review of Classical Lanchester ADEs

Our starting point for this consideration is a review of the Lanchester ADEs that comprised the basic theme of Part I.[1] This section will strike the immersed reader as a diversion, but I feel it is useful since we have spent an enormous amount of time and effort not only describing the fundamentally different basis of the conjugate theory, but also the basics of these ADEs as well. To these folk, I council a modicum of attention and a bit of patience. To those folk who have

inserted themselves at this point or who may not have all of my sterling words and equations etched on their memories in the mental equivalent of the Mosaic Tablets, I council preparedness for a fast trip.

As we recall, there are two fundamental pairs of ADEs which comprise the contribution of Lanchester: the Linear and the Quadratic ADEs. We shall review each of these in serial. For convenience in our subsequent discussions, we will adopt a slight deviation in our notation for Force Strengths. In Part I, we (fairly) consistently designated force strength for two sides by the capital letters A and B for Red (Amber) and Blue, respectively. These force strengths were assumed to be functions of time, have initial values at time zero indicated by zero subscripts, and represented the number of units not killed (attritted) in each force at a given time. We shall retain this convention, but to facilitate our discussion below, we introduce another notation that makes it considerably easier for us to discuss single ADEs rather than pairs. To this end, we introduce Force Strength functions E and F standing for Enemy and Friendly, respectively. It is a simple matter to associate A and B with E and F (or visa versa).

34.2.1 Linear Lanchester ADEs

As we know, Linear Lanchester ADEs are not linear differential equations. Their form is

$$\begin{aligned} \frac{dE}{dt} &= -\alpha EF, \\ \frac{dF}{dt} &= -\beta FE, \end{aligned} \tag{1}$$

where: α, β are the attrition rate constants. In keeping with our convention in this Part, the right hand sides of these ADEs, less the opposing force's Force Strength will be referred to as the ARC. The continued rationale for this will become evident as we continue. This pair of ADEs has a state solution

$$\alpha\left[F\left(t\right)-F_0\right]=\beta\left[E\left(t\right)-E_0\right], \tag{2}$$

which is the basis of the Linear designation; and has explicit time solutions of the form

$$E\left(t\right)=\frac{E_0\Delta_1}{\beta E_0-\alpha F_0 e^{-\Delta_1 t}}, \tag{3}$$

where:

$$\Delta_1 \equiv \beta E_0 - \alpha F_0. \tag{4}$$

The Friendly solution can be readily formed by symmetry. This material has, of course, been exhaustively described in Chapter III. As we know from the discussion of assumptions in that chapter, the primary common interpretation of these equations is representation of indirect or area fire combat.

34.2.2 Quadratic Lanchester ADEs

The Quadratic Lanchester ADEs have the form

$$\begin{aligned} \frac{dE}{dt} &= -\alpha F, \\ \frac{dF}{dt} &= -\beta E, \end{aligned} \tag{5}$$

which are linear differential equations. This linearity will be a matter of some fundamental consideration in the Part III. The state solution of these ADEs reduces to

$$\alpha \left[F^2(t) - F_0^2 \right] = \beta \left[E^2(t) - E_0^2 \right], \tag{6}$$

which is the basis of the Quadratic designation; and has explicit time solutions of the form

$$\begin{aligned} E(t) &= E_0 \cosh(\gamma t) - F_0 \delta \sinh(\gamma t), \\ F(t) &= F_0 \cosh(\gamma t) - \frac{E_0}{\delta} \sinh(\gamma t), \end{aligned} \tag{7}$$

where:

$$\begin{aligned} \gamma &\equiv \sqrt{\alpha\beta}, \\ \delta &\equiv \sqrt{\frac{\alpha}{\beta}}. \end{aligned} \tag{8}$$

The primary common interpretation of these equations if representation of direct or point fire combat.

34.2.3 Osipov's ADEs

Osipov, of course, was the Russian contemporary of Lanchester, who independently developed the basics of attrition theory.[2] The general form of Osipov's ADEs is

$$\begin{aligned} \frac{dE}{dt} &= -\alpha E^{2-n} F, \\ \frac{dF}{dt} &= -\beta F^{2-n} E, \end{aligned} \tag{9}$$

where: n = attrition order; and the state solution is

$$\alpha \left[F^n(t) - F_0^n \right] = \beta \left[E^n(t) - E_0^n \right]. \tag{10}$$

The explicit time solutions of these ADEs are rather complicated and we will not reproduce them here. The reader is referred to Chapter VIII.

Osipov's ADEs are more general in form, subsuming both of the Lanchester ADEs, and have a more varied application to attrition problems, in particular situations where only the edges of combat formations inflict or experience attrition.

34.2.4 Mixed Lanchester ADEs

We also note, in passing, the association of a pair of ADEs comprised of one Linear ADE and one Quadratic ADE to describe ambush or guerrilla warfare.

34.3 Modern ADEs

As we noted in both the preceding chapter and Chapter XXII, Bonder-Farrell Theory [3] describes the form of an ADE as

$$\frac{dE}{dt} = -f\left(E, F, r, t, \Xi\right) F, \tag{11}$$

where: E and F are the enemy and friendly force strengths, respectively, r and t are range and time, respectively, and Ξ indicates a catch-all of any other variables or parameters. The function f is the ARC as described in the preceding chapter. We note in passing that if f is not a function of either E or F, then the ADE is Quadratic in a Lanchester sense. More importantly, it is linear in the differential equation sense! If f is a function of E, but not of F, then the ADE is Linear in a Lanchester sense.

If we examine the ARC examples in the preceding chapter, we find a repeated form,

$$f = \frac{\varepsilon\alpha E}{\alpha + \varepsilon E}, \tag{12}$$

which reduces to Linear Lanchester if $\alpha \gg \varepsilon E$, or to Quadratic Lanchester if $\alpha \ll \varepsilon E$. This ARC is one of two new basic forms of archetype ADE (akin to the Quadratic and Linear Lanchester ADEs) that we shall consider here. We refer to these as Archetypes since their solutions are insightful even though they do not depend on range or other variables other than force strengths. A convenient name for this Archetype is the Averaged Lanchester ADE.

34.3.1 Example ADEs

In this subsection, we shall now review the different types of ADE that result from the ARCs that we developed in the preceding chapter.

34.3.1.1 Infantryman

The ADE for solitary aimed fire is of the form

$$\frac{dE}{dt} = -\frac{\varepsilon\alpha EF}{\alpha + \varepsilon E}, \tag{13}$$

which is of the Averaged Lanchester type. The ADE for pointed fire is of the form

$$\frac{dE}{dt} = -\frac{\alpha\varepsilon E^2 F}{\alpha\kappa E + \mu}, \tag{14}$$

which reduces to Linear Lanchester if $\alpha\kappa E \ll \mu$, and defines a new archetype of the form

$$\frac{dE}{dt} = -\omega E^2 F, \tag{15}$$

if $\alpha\kappa E \gg \mu$, $\omega \equiv \frac{\varepsilon}{\kappa}$, which we note is an Osipov-Lanchester ADE of attrition order zero. Since we already have a logarithmic equation (see Part I,) we shall designate this as the Zero Lanchester-Osipov ADE. Since the attrition order is integer, we would expect a high possibility of analytical solution for this equation, and this is indeed the case as we shall see below. The ADE for the LAW is of the Averaged Lanchester type.

34.3.1.2 Tanks

We may identify the general form for the tank serial attrition process as Averaged Lanchester. There are two limiting forms for the parallel AND attrition process. The lethal limit is exceedingly complicated, having the form

$$\frac{dE}{dt} = -\frac{\varepsilon\alpha E\left(\alpha + \varepsilon E\right) F}{\alpha^2 + 3\alpha\varepsilon E + \varepsilon^2 E^2}, \tag{16}$$

which reduces to Linear Lanchester if $\alpha \gg \varepsilon E$, and to Quadratic Lanchester if $\alpha \ll \varepsilon E$. The Nonlethal limit is of the Averaged Lanchester type.

34.3.1.3 Artillery

The ADE for artillery fire is also of the Averaged Lanchester type. Because of the dominating importance of the communication time, we shall not spend a great deal of time considering artillery fire further at this time.

34.3.1.4 Guided Weapons

The ADE for many guided weapons are also of the Average Lanchester type, although this is not always the case. We shall defer detailed consideration of guided weapons until later.

34.4 Zero Lanchester-Osipov ADE

As we indicated above, the limit of slow target acquisition for pointed fire leads to a new archetype ADE, the Zero Lanchester-Osipov ADE. The pair of ADEs has the form

$$\begin{aligned} \frac{dE}{dt} &= -\omega E^2 F, \\ \frac{dF}{dt} &= -\lambda F^2 E, \end{aligned} \tag{17}$$

if the equations occur symmetrically. (We note in passing that this pair of differential equations may have application in conditions where target acquisition is very difficult such as modern combat in buildings (Military Operations in Urban Terrain (MOUT) or Built-up Areas (MOBA)) or in fog or darkness with only visual sensors.)

This pair of ADEs has a state solution

$$\left(\frac{E(t)}{E_0}\right)^{\lambda} = \left(\frac{F(t)}{F_0}\right)^{\omega}, \tag{18}$$

which indicates that engagements described by this archetype cannot be won in a conclusion sense; all such are draws implicitly. (This is consistent with what we have observed in previous consideration of logarithmic attrition in Part I.) If we rearrange this equation as

$$F(t) = F_0 \left(\frac{E(t)}{E_0}\right)^{\frac{\lambda}{\omega}} \tag{19}$$

then we may apply the Method of Normal Forms to solve the ADEs. This gives us an Enemy ADE of the form

$$\frac{dE}{dt} = -\omega E^2 F_0 \left(\frac{E}{E_0}\right)^{\frac{\lambda}{\omega}}. \tag{20}$$

This is a simple first order differential equation that can be solved directly by integration, yielding

$$E(t) = \left[E_0^{\frac{\lambda-\omega}{\omega}} - (\lambda-\omega) F_0 E_0^{-\frac{\lambda}{\omega}} t\right]^{\frac{\omega}{\lambda-\omega}}, \tag{21}$$

which clearly does not hold if $\lambda = \omega$. In this case,

$$E(t) = \frac{E_0}{\sqrt{1 + 2\omega E_0 F_0 t}}. \tag{22}$$

The Friendly solutions can easily be constructed by symmetry.

34.5 Averaged Lanchester ADEs

As we have indicated, the most prevalent of the forms arising in our examples of ARCs calculated using Bonder-Farrell Theory is the archetype that we have designated as the Averaged Lanchester ADEs,

$$\begin{aligned} \frac{dE}{dt} &= -\frac{\alpha\varepsilon EF}{\alpha+\varepsilon E}, \\ \frac{dF}{dt} &= -\frac{\beta\eta FE}{\beta+\eta F}. \end{aligned} \tag{23}$$

These equation admit of a state solution,

$$\beta\left[\frac{\alpha}{\varepsilon}E+\frac{E^2}{2}-\frac{\alpha}{\varepsilon}E_0-\frac{E_0^2}{2}\right]=\alpha\left[\frac{\beta}{\eta}F+\frac{F^2}{2}-\frac{\beta}{\eta}F_0-\frac{F_0^2}{2}\right], \tag{24}$$

which we note is quadratic. This is clearly not going to help us very much in solving these equations since eliminating F (we have dropped the t dependence for brevity,) will leave us with a rather nasty radical. An easier approach may be to use the Method of Frobenius.

Let the explicit time solution for the Enemy Force Strength be

$$E(t)=\sum_{i=0}^{\infty}e_i t^i, \tag{25}$$

and the explicit time solution for the Friendly Force Strength similar but with coefficients f_i. Obviously, the first two coefficients in each series are defined by the initial conditions. Now rewrite equations 23 as

$$(\alpha+\varepsilon E)\frac{dE}{dt}=-\alpha\varepsilon EF, \tag{26}$$

and similarly for the Friendly Force Strength. We may now substitute series for the force strengths, giving

$$\begin{aligned} \left(\alpha+\varepsilon\sum_{i=0}^{\infty}e_i t^i\right)\frac{d}{dt}\sum_{j=0}^{\infty}e_j t^j &= -\alpha\varepsilon\sum_{i=0}^{\infty}e_i t^i\sum_{j=0}^{\infty}f_j t^j, \\ \left(\alpha+\varepsilon\sum_{i=0}^{\infty}e_i t^i\right)\sum_{j=0}^{\infty}e_{j+1}(j+1)t^j &= \end{aligned} \tag{27}$$

which we may rewrite by combining the products of series

$$\alpha\sum_{j=0}^{\infty}e_{j+1}(j+1)t^j+\varepsilon\sum_{k=0}^{\infty}t^k\sum_{i=0}^{k}e_{k-i}e_{i+1}=-\alpha\varepsilon\sum_{k=0}^{\infty}t^k\sum_{i=0}^{k}e_{k-i}f_k. \tag{28}$$

If we assume linear independence, then we may remove the infinite series above, and reduce this to

$$\alpha e_{k+1}(k+1)+\varepsilon\sum_{i=0}^{k}e_{k-i}e_{i+1}=-\alpha\varepsilon\sum_{i=0}^{k}e_{k-i}f_k. \tag{29}$$

This equation allows us to bootstrap calculation of the series coefficients in the same manner as we used in solving Osipov's ADEs. Specifically, this gives

$$e_{k+1} = \frac{-\varepsilon}{\alpha(k+1)+\varepsilon E_0}\left[\alpha\sum_{i=0}^{k} e_{k-i}f_k + \sum_{i=0}^{k-1} e_{k-i}e_{i+1}\right]. \tag{30}$$

This is a solution of the Averaged Lanchester ADE. The Friendly complement can be constructed by symmetry.

Despite its neatness, this solution is not particularly useful. Luckily, this pair of ADEs is relatively stable so that simple straightforward open integration approximations are usually valid and we may regress to our old friend - the spreadsheet simulation. This allows us to calculate numbers but does not give us a compact insight into the behavior of these equations since we do not have a simple analytical solution. Filling this gap is part of the value of the conjugate theory of ARCs and we shall explore this in the next section.

34.6 Real World Limits

The ARCs that we established in the preceding chapter are generally considerably more complex in form than the classic archetypes of Lanchester Attrition Theory, the Linear and Quadratic forms, although in many cases, these complex ARCs reduce in some limits to approximations that are archetypes. The question is, what kind of behavior can we expect in the real world? Can we parametrize those situations when the simplifying limits hold, and those situations when the general form holds, recognizing that when the general form holds, analytic solution and the insight that it provides becomes problematic.

To get some insight into this question, we now examine two cases: aimed infantry fire; and lethal, parallel tank fire.

34.6.1 Infantry

As we noted above, the ADE for aimed infantry fire is Averaged Lanchester. Unfortunately, this ARC is dependent on both range and the strength of the enemy force. Without knowing that force strength in detail, how can we draw conclusions about the ARC? To answer this, we examine the general form of the ARC,

$$f = \frac{\varepsilon\alpha E}{\alpha+\varepsilon E}. \tag{31}$$

Since $E > 1$ almost always (except in terminal situation where we know from Part I that these equations don't hold,) then we can compare the values of α and ε, or alternately, the times that are their inverses. Since these times do not depend on enemy force strength, but only range and other parameters, these can be calculated with some confidence.

Let us make some assumptions and assignments. Assume that the infantryman in question has an enemy who is 0.5m wide and 1.9m high, and has an inherent contrast of 0.2. (No British

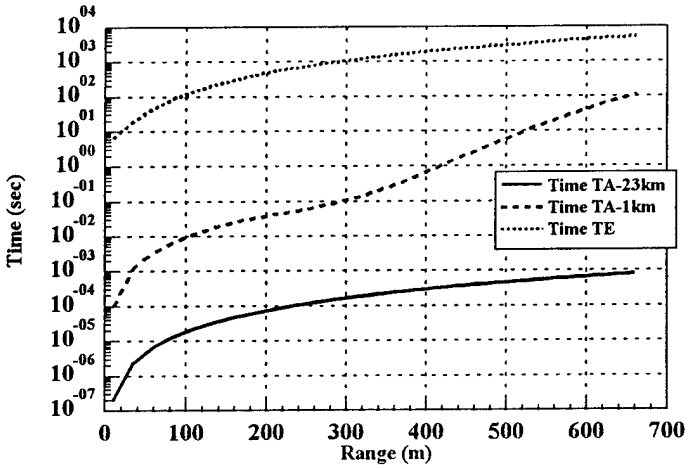


Figure 1: Times of Infantry Attrition

Red Coats here!) Further assume that the only imaging sensor the infantryman has is his eyes and that the field of search is human peripheral vision. Then we can calculate the expected target acquisition time using the equations we developed earlier. Take two limiting cases, a clear day (visibility $\sim$ 23 km,) and a foggy day (visibility $\sim$ 1 km). We plot these values in figure 1.

Now retaining this enemy size, assume that the accuracy of an infantry rifle is about 6 milliradians, that the effective rate of fire is about ten rounds per minute (ammunition discipline after all,) and the probability of kill given a hit is essentially one.[4] Calculate the expected time to kill and plot this value in figure 1

If we examine this figure, we immediately note that the Target Engagement time is much, much greater than either Target Acquisition time. This means that $\alpha \ll \varepsilon$, and we may immediately identify that over all ranges of interest, this ARC is effectively α, which means that the Lanchester Quadratic ADE is a natural limit without any consideration of how large the enemy force strength is.

Before continuing, let us examine the sensitivity of these calculations. Prone targets will be considerably harder to hit and harder to see, so we would expect the effect of considering smaller presented area targets to approximately scale, leaving the ratios about as they are here.

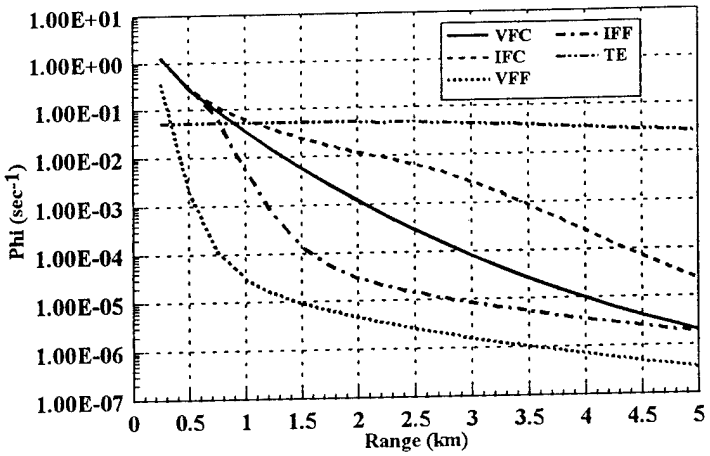


Figure 2: Tank Attrition Subprocess Coefficients

The accuracy used here is a bit high and may be exceeding high given the insight of history. Further, the rates of fire are low, but we rationalize this on the basis of ammunition conservation. Finally, the probability of kill given a hit is idealistic. This all adds up to an optimistic estimate of the expected time to kill. Thus, the conclusion we draw here is probably valid in general. This doesn't directly address the pointed fire case, but the situation is even more pronounced there so that situation is probably not very interesting, especially after we continue our discussion below.

34.6.2 Tanks

In a similar manner, we may examine the same situation for tanks. If we assume a nominal tank to have a basic load of about 40 rounds, and that these rounds have an average probability of kill given a hit of 0.8, then the number of stowed kills (ignoring hit probability,) of about 32. This definitely places us in the realm of lethal fire for the parallel cases. If we again consider a tank of dimensions 2.67m x 2.67m with inherent contrast of 0.2 in the visual band and 2° K in the far infrared, and that the effective rate of fire is four rounds per minute, then we may calculate the same type of times as before. This time we display the coefficients themselves in figure 2.

This is a very complicated situation. For ranges less than between 0,25 km (VFF) and 1 km (VFC), depending on what the sensor is and what the weather conditions are,[1] the rate for target acquisition is greater than the rate of target engagement while for ranges greater than these the rate of target engagement is greater than the rate of target acquisition. Note that the rate of target engagement changes relatively slowly, only about a half an order of magnitude (3 dB as the Radar guys would say,) while the rate of target acquisition changes considerably more.

We may identify three regions of ARC behavior. For short ranges, which depend on the sensor and weather, $\varepsilon E \gg \alpha$, and the ARC can be approximately Quadratic Lanchester. For long ranges, which again depend on sensor and weather (and E), $\varepsilon E \ll \alpha$, and the ARC can be approximately Linear Lanchester. Of the three regions, this is the largest. For intermediary ranges, $\varepsilon E \sim \alpha$, and the ARC must be Averaged Lanchester. This gives us two apparent choices in solving the ADE: we may solve three ADE, one Linear, the next Averaged, and the third Linear, in range, or we may solve the Averaged throughout range. Regardless, we are essentially forced to make a numeric calculation and we lose the goal of having an ADE that reduces to Quadratic.

Given that we must make a numeric calculation, we may examine an approximation. Since the ARC depends on the enemy force strength, and the formalism is only valid when we are adequately removed from conclusion, it seems plausible to examine an approximation where we replace the enemy force strength in the ARC with an averaged force strength of some sort. The obvious starting point is to use the initial enemy force strength so that equation 31 becomes

$$f \simeq \frac{\varepsilon\alpha E_0}{\alpha + \varepsilon E_0}. \tag{32}$$

We may make comparison numerical calculations between an "exact" ARC, equation 31, and an approximate ARC, equation 32. We show these in figure 3. The ADEs are symmetric so we drop the distinction between Enemy and Friendly and use our normal designation Red and Blue. The ADEs are transformed to their range form,

$$\begin{aligned} \frac{dA}{dr} &= \frac{dA}{dt}\frac{dt}{dr} \\ &= -\frac{\varepsilon\alpha AB}{\alpha + \varepsilon A}\left(-\frac{1}{v}\right), \end{aligned} \tag{33}$$

and we arbitrarily open the engagement at 5 km. The Initial Force Strengths are 100 and 50 units respectively, and the Red force has visual sensors while the Blue force has InfraRed sensors. The weather is clear. All rate data is the same as in figure 2.

The message here is pretty clear. First, the ARCs and the ADEs are both highly nonlinear (in the mathematical sense) and decidedly not Quadratic Lanchester. Second, under many circumstances, we may replace the enemy force strength in the ARC by an approximate value. (This

[1]Recall that the V stands for Visual and I stands for InfraRed. The middle F stands for fully exposed, and the tailing C and F stand for clear and fog, respectively.

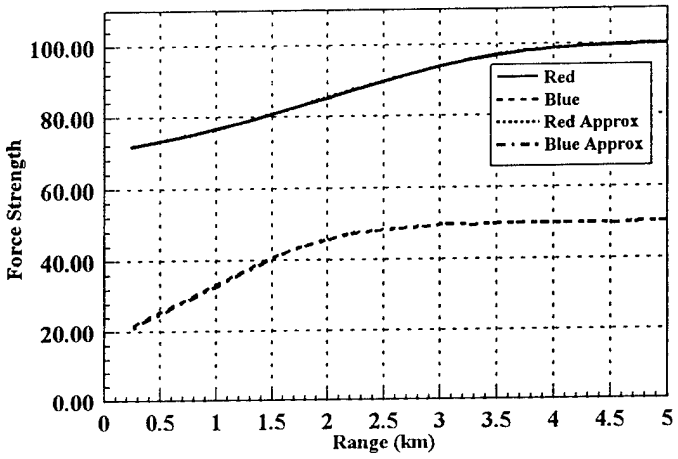


Figure 3: Numeric Averaged Lanchester Calculations

is just a statement of the mean value theorem of the calculus.) As a result, the ARC becomes essentially Quadratic Lanchester albeit still range (and other parameter) dependent. This bodes well for an approximation of the ARC under many circumstances as Quadratic Lanchester.

34.7 Range Averaging

Having established the validity of using the initial force strength in Tank ARC in the preceding section, reducing the ADEs to range dependent Quadratic Lanchester, we are now in a position to examine the form of these ARCs. The two approximate ARCs used in the preceding section are shown in figure 4.These two ARCs are labeled Blue, Red ARC which indicate the firer..

We note the behavior of these ARCs: For some distance at short range, they are essentially constant, then they decay fairly rapidly. If anything, they remind us of Fermi-Dirac distributions for fermions (e.g., electrons,) and that suggests a possible approximation for them - the equivalent of the zero temperature distribution which is a step function. These are shown (approximately)

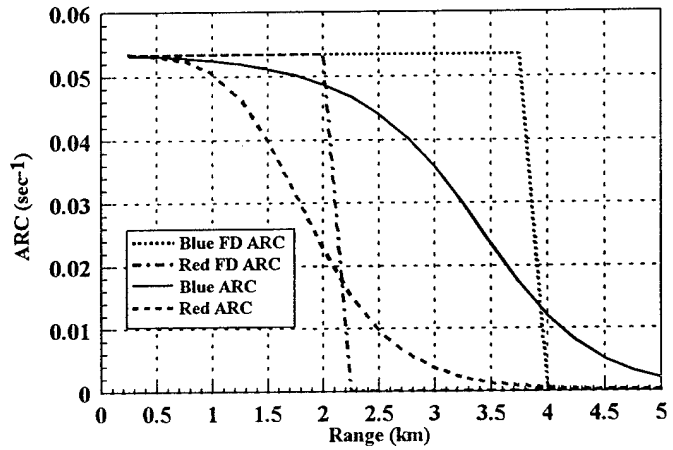


Figure 4: Approximate ARCs

and labeled as Blue, Red FD ARCs.[2] These also serve to show the presence of a killing zone here. For a range band between approximately 2 and 4 km., the Blue force has a significant attrition advantage over the Red force. For shorter and longer ranges than this band, the two forces have essentially equivalent attrition capabilities. From a tactical standpoint, Blue wants to conduct the engagement in this range band. Red on the other hand, wants to conduct the engagement at shorter or loner range. Thus, if Blue is attacking, they want to rapidly close to 4 km and try to conclude the engagement by the time they have closed to 2 km. If they are defending, they again want to open the engagement at long range, but if the engagement has not been concluded by the time Red has closed to 2 km., they may want to withdraw. Alternately, if Red is attacking, they must rapidly close to 2 km., if possible, and if they are defending, their tactic is to lie concealed until Blue has closed to 2 km. Red's defensive tactic is thus an ambush.

How well do these FD ARCs further approximate the range dependent ARCs? We may examine this by repeating the calculations of the preceding section using these FD ARCs. These are shown in figure 5. The calculations using the initial force strengths, introduced in the

[2]The student of solid state physics will probably be wondering if there are things like "dog bone" orbits. The answer is yes. They arise when we consider spatially distributed forces.

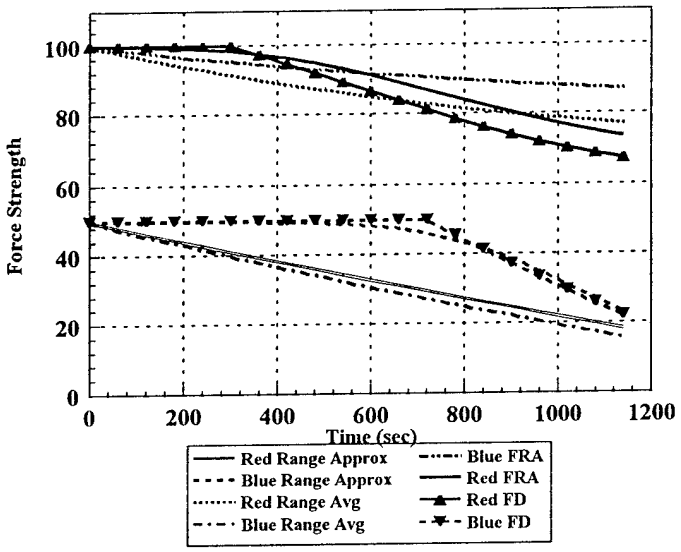


Figure 5: Approximate Attrition Calculation

preceding section, are labeled as Red, Blue Range Approx, while the calculations using the FD ARCs are labeled Red, Blue FD. Relatively good agreement can be seen.

This indicates the possibility of a further approximation. Since the FD ARCs are step functions, they are nothing more than range limited constant ARCs. These would be simple Quadratic Lanchester ARCs except for the fact that the range limits for the two ARCs are different. Is there a way to replace these with simple constant ARCs with the same range limits? This has enormous promise. First, it would allow us to approximate the ARCs as simple Quadratic Lanchester ARCs and thus return to the simplicity of the formalisms we developed in Part I, and second, the tactical situation would dictate the duration of the engagement from simple consideration of range closure given by

$$t_{eng} \simeq \frac{r_0 - r_f}{v}, \tag{34}$$

where:

r_0 = the opening range of the engagement,

r_f = the final or closing range of the engagement, and

v = the closing speed of the engagement. Obviously, if $v = 0$, then the engagement can go

to infinite duration. Regardless, this decoupling would permit simpler computation as well as bounding of the duration of the engagement from consideration of the tactical situation.

The questions now are how to further approximate the ARCs, and how good are the approximations? To address these questions, it is useful to examine the form of the range dependent ADEs. The Red force ADE is

$$\frac{dA(r,t)}{dt} = -f(r)\,B(r,t), \tag{35}$$

which we have written as the time form. We want to represent the ADE as

$$\frac{dA(t)}{dt} = -\alpha B(t). \tag{36}$$

Since we are dealing with the force strengths as spatial points, that is, they have no spatial distribrution, we may integrate these equations wrt time as

$$A(r,t) = A(r_0,0) + \int_0^{t_{eng}} f(r_0 - vt)\,B(r_0 - vt, t)\,dt, \tag{37}$$

and

$$A(t) = A(0) + \alpha \int_0^{t_{eng}} B(t)\,dt, \tag{38}$$

respectively. Since we want the Red force strengths to be (approximately) equal everywhere, we may combine these two equations as

$$\alpha \int_0^{t_{eng}} B(t)\,dt \simeq \int_0^{t_{eng}} f(r_0 - vt)\,B(r_0 - vt, t)\,dt. \tag{39}$$

This suggests several approaches to approximation. Since we also want the Blue force strengths to be (approximately equally) everywhere, an obvious approach seems to be to use the mean value theorem of the calculus to eliminate the integrals, giving

$$\alpha \simeq f(r^*), \tag{40}$$

where r^* is some value of r on $[r_f, r_0]$. The problem with this method is selection of r^*. While its existence is indicated by the calculus, we are at a bit of a loss of a method for estimating it. Thus, we turn to other approaches.

The next simplest approach we may take is to treat the Blue force strength as approximately constant. If we had the actual Blue force strength trajectory, we could, of course, perform these integrals simply, although probably numerically. Since we want to generally replace the range dependent ARCs with approximate constant ARCs in a general manner, we must proceed as if we are ignorant of the actual force strength trajectories.

If we treat the force strength as constant, then we may rewrite equation 40, changing to range form, as

$$\alpha \simeq \frac{\int_{r_f}^{r_0} f(r')\,dr'}{r_0 - r_f}, \tag{41}$$

which we shall refer to as Range Averaging (RA). Calculations using these constant ARCs are shown in figure 5, label as Red, Blue Range Average. We note immediately the loss to the range structure of the force strength trajectories, as we would expect and some degree of inaccuracy compared to the more exact range dependent calculations.

Another approach is to use some idealized form of the force strength trajectory which we shall designate as Force Range Averaging (FRA). This necessitates integrating equation 39, and a useful numerical approximation can be the trapezoid rule. In this case,

$$\alpha \simeq \frac{f(r_0)\,B(r_0,0) + f(r_f)\,B(r_f,t_{eng})}{B(r_0,0) + B(r_f,t_{eng})}. \tag{42}$$

Obviously, this method is very sensitive to the selection of the final force strength value. For the calculations presented as Red, Blue FRA in figure 5, we selected a value of the final force strength as one-half the initial, so that

$$\alpha \simeq \frac{2f(r_0) + f(r_f)}{3}. \tag{43}$$

As we may see, these calculations are comparable to the RA ARCs in accuracy. This accuracy may be improved by iterative recalculation of the ARCs and the force strengths.

34.8 Conclusion

In conclusion, we come full circle. In this Part II of the work, we have established the connection between the technical aspects of the subprocesses of attrition and ARCs. These ARCs result in ADEs that are mathematically non-linear and greatly different from the simple archetypes of the Lanchester ADEs of Part I. Happily, the nonlinearity of these ARCs can be removed by approximation, reducing the ADEs to simple range dependent quadratic ADEs (in a Lanchester sense.) Further, it is desirable, as we shall establish in Part III, to be able to reduce these ADEs further to constant ARC Quadratic ADEs - simple Lanchester archetypes. This can be done, albeit at the cost of greater inaccuracy although this inaccuracy can be reduced somewhat by iterative calculation. This is the completion of the circle. We started with simple constant ARC ADE archetypes in Part I, and we have returned to these simple archetypes (at the cost of some inaccuracy) while retaining the technical basis for calculating the ARCs from first principles. This provides us with entry to Part III where we consider heterogeneous force strengths and their aggregation.

34.9 References

[1] Lanchester, Frederick W., **Aircraft in Warfare: The Dawn of the Fourth Arm**, Constable and Company, LTD., London 1916.

[2] Osipov, M., "The Influence of the Numerical Strength of Engaged Forces on their Casualties", Robert L. Helmbold, and Allan S. Rehm, trans., U. S. Army Concepts Analysis Agency, Bethesda, MD, Research Paper CAA-Rp-91-2, September 1991.

[3] Bonder, Seth, and Robert Farrell, eds., "Development of Analytical Methods of Battalion Force Activities", Systems Research Laboratory, Department of Industrial Engineering, The University of Michigan, Ann Arbor, Michigan, September 1970 (AD 71 4677) and references therein.

[4] Fowler, Bruce W., "Environmental Effects on Combat Performance: A Lanchester Approach", in L. G. Callahan, Jr., and Lawrence Low, eds., **Proceedings of the Workshop on Modeling, Simulation, and Gaming of Warfare**, Callaway Gardens, GA, 2-5 December 1984.

This Page Intentionally Left Blank


DEFENSE TECHNICAL INFORMATION CENTER
50
YEARS
EXCELLENCE
IN INFORMATION
SERVICE


DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA