THEATER-LEVEL STOCHASTIC AIR-TO-AIR ENGAGMENT MODELING VIA EVENT OCCURRENCE NETWORKS USING PIECEWISE POLYNOMIAL APPROXIMATION

Dissertation

D.R. Denhard, Major, USAF

AFIT/DS/ENS/01-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
01 Sep 2001	Final	01 May 1998 - 01 Sep 2001

Title and Subtitle	**Contract Number**
Theater-Level Stochastic Air-to-Air Engagement Modeling via Event Occurrence Networks Using Piecewise Polynomial Approximation	**Grant Number**
	Program Element Number

Author(s)	**Project Number**
Major D. R. Denhard, USAF	**Task Number**
	Work Unit Number

Performing Organization Name(s) and Address(es)	**Performing Organization Report Number**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 P Street, Building 640 WPAFB, OH 45433-7765	AFIT/DS/ENS/01-01

Sponsoring/Monitoring Agency Name(s) and Address(es)	**Sponsor/Monitor's Acronym(s)**
Air Force Studies and Analyses Agency 1570 Air Force Pentagon Washington DC 20330-1570 ATTN: Maj raig Knierem	**Sponsor/Monitor's Report Number(s)**

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes

Abstract
This dissertation investigates a stochastic network formulation termed an event occurrence network (EON). EONs are graphical representations of the superposition of several terminating counting processes. An EON arc represents the occurrence of an event from a group of (sequential) events before the occurrence of events from other event groupings. Events between groups occur independently, but events within a group occur sequentially. A set of arcs leaving a node is a set of competing events, which are probabilistically resolved by order relations. An important EON metric is the probability of being at a particular node or set of nodes at time t. Such a probability is formulated as an integral expression (generally a multiple integral expression) involving event probability density functions. This integral expression involves several stochastic operators: subtraction; multiplication; convolution, and integration. For the EON probability metric, simulation is generally computationally costly to obtain accurate estimates for large EONs, transient nodes, or "rare" states. Instead, using research with probabilistic activity networks, a numerical approximation technique using piecewise polynomial functions is developed. The dissertation's application area is air-to-air combat modeling.

Subject Terms		
Air Combat, Event Occurrence Networks, Probabilistic Networks, Activity Networks, Piecewise Polynomial Approximation, Simulation		
Report Classification unclassified		**Classification of this page** unclassified
Classification of Abstract unclassified		**Limitation of Abstract** UU
Number of Pages 379		

THEATER-LEVEL STOCHASTIC AIR-TO-AIR ENGAGEMENT MODELING VIA
EVENT OCCURRENCE NETWORKS USING PIECEWISE POLYNOMIAL
APPROXIMATION

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

D.R. Denhard, B.S., M.S.

Major, USAF

September 2001

AFIT/DS/ENS/01-01

THEATER-LEVEL STOCHASTIC AIR-TO-AIR ENGAGEMENT MODELING VIA
EVENT OCCURRENCE NETWORKS USING PIECEWISE POLYNOMIAL
APPROXIMATION

D.R. Denhard, B.S., M.S.
Major, USAF

Approved:

<u>Date</u>

_____
Jack M. Kloeber, LTC(retired), USA (Chairman)   _____

_____
Dennis W. Quinn, PE, Professor   _____

_____
Dennis C. Dietz, LtCol(retired), USAF (Member)   _____

_____
Raymond R. Hill, LtCol, USAF (Member)   _____

_____
E. Price Smith, LtCol, USAF (Member)   _____

Accepted:

_____   _____
Robert A. Calico, Jr.   Date
Dean, Graduate School of Engineering and Management

Dedicated in loving memory to my father
whose grin I can sense from above

Acknowledgements

Many people deserve a great deal of thanks in bringing this dissertation to a conclusion. First and foremost, I would like to express my sincere thanks to my beautiful wife. Without her, this dissertation would never have been completed. Thanks, babe.

To my committee chairman Dr. Dennis Dietz (Lt Col, USAF Retired) and Dr. Jack Kloeber (LTC, USA Retired) thanks for your guidance and patience. Two of the smartest men I have ever met, your critiques of the research were always focused, productive, and timely.

To the many committee members I have had over the years: Col Jack Jackson, Lt Col Edward Pohl, Lt Col E. Price Smith, and Maj Paul Murdock, thanks for your guidance and participation. A special thanks to Lt Col Ray Hill whose guidance over the years and attention to detail piloted me toward graduation.

D.R. Denhard

Table of Contents

List of Figures

xi

List of Tables

AFIT/DS/ENS/01-01

Abstract

This dissertation investigates a stochastic network formulation termed an event occurrence network (EON). EONs are graphical representations of the superposition of several terminating counting processes. An EON arc represents the occurrence of an event from a group of (sequential) events before the occurrence of events from other event groupings. Events between groups occur independently, but events within a group occur sequentially. A set of arcs leaving a node is a set of competing events, which are probabilistically resolved by order relations. EONs differ from other stochastic networks discussed in the literature such as Activity Networks, Petri Nets, Task Precedence Graphs, and Belief Networks and are strongly influenced by the research of Ancker, Gafarian, Kress and several associates to model m versus n stochastic ground combat.

An important metric for an EON is the probability of being at a particular node or set of nodes at time t. Such a probability is formulated as an integral expression (generally a multiple integral expression) involving event probability density functions. This integral expression involves several stochastic operators:

- Subtraction,

- Multiplication,

- Convolution, and

- Integration.

In the literature, simulation-based methods are the dominant approximation technique for obtaining metrics from stochastic networks due to their logical simplicity and efficient implementation. However, for the EON probability metric, simulation is generally computationally costly to obtain accurate estimates for large EONs, transient nodes, or "rare" states. Instead, using research by Fergueson, Shortell, and

Lawrence with probabilistic activity networks, a numerical approximation technique using piecewise polynomial functions is developed.

The dissertation's application area is air-to-air engagement modeling. The dissertation contains a comprehensive overview of analytical and simulation based theater-level air combat models including an in-depth review of the US Air Force's THUNDER air-to-air engagement submodel. Additionally, the dissertation provides an overview of air-to-air engagements and the combat environment in which these engagements take place. Factors discussed include: strategy; missions; air-to-air spatial layout; aircraft; detection and identification; weapons; command, control, communications, and information (C3I); electronic warfare; and tactics. The dissertation concludes with a large EON air-to-air engagement example.

THEATER-LEVEL

STOCHASTIC

AIR-TO-AIR ENGAGEMENT

MODELING

VIA

EVENT OCCURRENCE NETWORKS

USING PIECEWISE POLYNOMIAL APPROXIMATION

I. Introduction

Since the introduction of the aircraft into war[1], the battle for control of the air has become an essential component of modern warfare. In fact, since the German attack on Poland in 1939, no nation has won a war; no major offensive has succeeded; and no defense has sustained itself against an opponent who has had control of the air [193: page 10]. The Gulf War provided a vivid example of the devastating effects that can occur when one side (Iraq) cedes sufficient control of the air to its opponent (Coalition Forces) [93].

To have air superiority means having sufficient control of the air to conduct air, sea, and land operations attacks on an opponent without serious air opposition as well as the ability to prevent effective air operations from an opponent [193: page 10] [101: page 7,8]. The natural consequences of two or more opponents battling to achieve air superiority are combat engagements. Air combat engagements may be divided into three categories: air-to-air (e.g., aircraft versus aircraft); air-to-

---

[1]World War I was the first war in which aircraft were used with the first missions being reconnaisance [33: page xvii].

surface (e.g., aircraft versus ground target) and surface-to-air (e.g., surface-to-air missile (SAM) versus aircraft). This dissertation addresses the former, air-to-air engagements.

1.1 Air-to-Air Engagement

Since the classical, stylistic dogfights of World War I, air-to-air engagements have fascinated the military air community. Although surface-to-air engagements usually account for more aircraft kills than air-to-air engagements, surface-to-air engagements lack the glamorous, mono-a-mono, competitive descriptions that are generally associated with air combat[2] [101: page 7,8] [185: pages 56, 72]. Historical accounts of air-to-air engagements paint these engagements as "little slice[s] out of time"[3] which are unrepeatable; the aircraft, the pilots, the weapons, the tactics, and the conditions all come together in an instant of time to determine a result, and if the engagement were repeated, the outcome could easily be reversed [173: pages 10, 98].

Ten years after Wilbur and Orville Wright's first flight in a powered aircraft, unarmed reconnaissance aircraft quickly became established as a vital part of warfare. The usefulness of reconnaissance aircraft led to the development of "predator" aircraft whose sole purpose was to destroy these reconnaissance aircraft. The first reported firing of an air gun occurred on 20 Aug 1910 when Lieutenant (USA) Jacob Earl Fickel fired his rifle at a target from his single seat Curtiss biplane. Although Fickel was exploring the potential of the aircraft as an attack platform, the concept was of obvious equal importance as an air-to-air means to attack other aircraft [122: pages 8 and 9].

Fickel's test led to the development of a light weight, low recoil machine gun (developed by Samuel Neal McClean with later improvement by Colonel (USA) Issac

---

[2] In reality, air combat is far from a glamorous, gentleman's competition. Historically, in the majority of aircraft kills (over 80%), downed pilots were not even aware that they had been fired upon.

[3] Available combat time for most combat aircraft is on the order of five minutes or less.

N. Lewis) which proved ideal for air operation. Attached to the front right-hand seat of a Wright Model B flyer, the first trials of the gun were made on 2 June 1912. On 22 August 1914, 2nd Lieutenant (RAF) L.A. Strange with a Lewis machine gun attached to his Farman biplane took-off in pursuit of a German Albatros which was flying a reconnaissance mission over Maubeuge and is credited with the first use of a machine gun in combat. Until this date, aircrews persisted with pistols and rifles despite the difficulty aircrews faced in putting their aircraft into a position where they could get a clear shot at the opponent. It was as late as 5 October 1914, more than a decade after the Wright Brothers' first flight, that the first aircraft was recorded as shot down in an air-to-air engagement. A French rear-engined Voisin, piloted by Sergeant Joseph Franz shot down a German Avaitax. By the end of 1915, air combat had become an integral part of warfare. In today's time-compressed air combat world, life or death, win, lose or draw generally takes place in under five minutes [122: page 9] [80: page 11] [185: page 56] [192: pages 7,8].

1.2 Military Models

Before discussing air-to-air engagement models, some background information and terminology pertaining to military models used for combat analysis may be helpful to the reader. Several structures exist for categorizing military models[4] [106: page 3]. Figure 1 depicts one such structure for categorizing military models based on *scope*. Scope refers to the intended purpose of the model. Military models range in scope from *engineering* models (e.g., a model of aircraft aerodynamic performance) to *engagement* models (e.g., a model of a one versus one air-to-air engagement) to *mission* models (e.g., a model of a flight group of aircraft from take-off to attack on target and back) to the *campaign* or *theater-level* models "which represent a set of missions, operations, or battles in the pursuance of a military campaign objective" [99: page 5]. The theater-level model, in contrast to models with lesser scope,

---

[4]Reference [106] lists several military modeling taxonomies.

```
        INCREASING                    HIGHER
       AGGREGATION                  RESOLUTION

                        THEATER-
                          LEVEL

                         MISSION

                       ENGAGEMENT

                       ENGINEERING
```
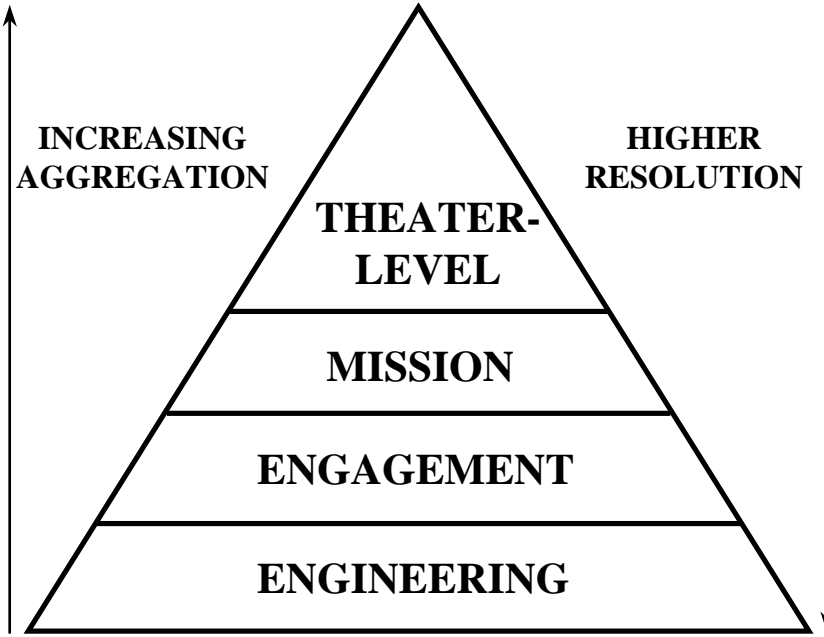
Figure 1. Military Model Categorization (Based on Model Scope)

shows the overall interactions of combined forces (air, land, and naval) in military operations.

An important model quality is the level of detail or *resolution* of aspects used in the model [6: page 7]. Figure 2 quantifies different aspects of model resolution [70: page 1004]. These aspects include:

- Entity resolution. This aspect of resolution refers to the level of object modeled. Higher entity resolution might mean modeling individual aircraft rather than only modeling to the level of an aircraft flight.

- Attribute resolution. This aspect of resolution refers to the detail of an object's attributes. Higher attribute resolution might mean specifying the types of air-to-air munitions carried by an aircraft rather than merely assigning a composite air-to-air munition to the aircraft.

4

- Logical-Dependency or Constraint resolution. This aspect of resolution refers to the level at which constraints are enforced on the attributes of an object and their interrelationships. Higher constraint resolution might mean requiring an air-to-air munition to have a captive-carry reliability rather than merely assuming that the air-to-air munition will always fire [50: page 3].

- Process resolution. This aspect of resolution refers to the level at which various effects cause changes in the attribute of an object. Higher process resolution might mean computing air-to-air attrition at the aircraft level, rather than computing air-to-air attrition at the flight level and then spreading the attrition equally across the flight's aircraft.

- Spatial and Temporal resolution. These aspects of resolution refer to the scales used for space and time. Higher spatial resolution might mean tracking aircraft position in three dimensional space rather than a two dimensional space. Higher temporal resolution might mean tracking time advancement in steps of seconds rather than in steps of minutes.

The terms "high" and "low" resolution are used in a relative context here and are not inherent properties of the aspects of a model. These terms only have meaning when comparing two or more models. For example, model A's air-to-air attrition process might be considered of high resolution when compared to model B's attrition process. However, model A's air-to-air attrition process might also be considered of low resolution when compared to model C's attrition process.

Military models found lower on the pyramid structure in Figure 1 tend to operate at a higher resolution (in terms of the models aspects given above) than models found higher on the pyramid structure[5]. This inverse relationship between model scope and resolution is a direct result of current limitations in model complexity, size, run time performance, and cost. As the scope of a model increases, "the sheer

---

[5]However, the relative resolution between two models can be ambiguous; "one model may have higher resolution than another in some respects, but lower resolution in others" [47: page 4].

RESOLUTION

ENTITY

LOGICAL
DEPENDENCY

PROCESS

TEMPORAL

SPATIAL

ATTRIBUTE
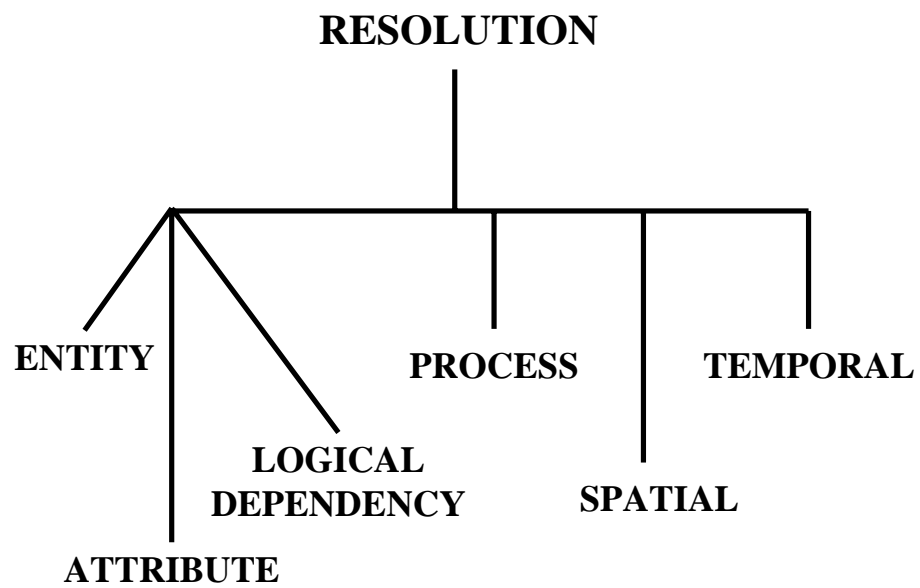
Figure 2. Aspects of Model Resolution

number of combatants and weapon systems makes it impossible to maintain individual item resolution" without sacrificing an increase in model complexity, size, run time performance, and / or cost [109: page 1-1].

Closely associated with the concept of resolution is the concept of *aggregation* and *disaggregation*. In general, aggregation refers to the process of assembling or combining into a mass and disaggregation refers to the antithetical process of disassembling or separating into particulars. Although usually thought of in terms of entity and process aspects of resolution, aggregation and disaggregation apply to other aspects of resolution as well[6]. For example, in the case of temporal resolution, we can aggregate a time step of seconds into minutes and disaggregate minutes into seconds. Military models found higher on the pyramid structure in Figure 1 tend to be more aggregated (in terms of the models aspects given above) than models found lower on the pyramid structure[7].

1.3 Air-to-Air Engagement Models

Today's air-to-air engagement models consist mainly of relatively large, complex simulations and may be roughly divided into two main groups according to their intended purpose: analysis or training. The air-to-air engagement models are used as separate, standalone models or as a section of larger models depending on the scope of the problem being investigated. The level of detail in these models varies depending on the model's scope and purpose.

---

[6]In the literature I have reviewed, aggregation / disaggregation has been used in reference to entity and process resolution exclusively. For example, in the case of entity resolution, aggregating aircraft entities into a flight entity or disaggregating a flight entity into aircraft entities; in the case of process resolution, aggregating specific types of air-to-air munitions on an aircraft into a single or composite munition or disaggregating a composite air-to-air munition into specific types of munitions.

[7]However, as the relative resolution between two models can be ambiguous, so also can the level of aggregation. For example, a campaign model can model air entities to the level of individual aircraft as can an engagement air model, but the air-to-air attrition process used in the campaign model might be more aggregate than the process used in the engagement model.

The Air Force's primary theater-level model for analysis is THUNDER. Operational since 1986, THUNDER is a "computer simulation of conventional air, land, and naval warfare" [29: Volume I, page 1]. The air war model uses a discrete event, time-stepped stochastic simulation whereas the ground war model[8] uses a discrete event, time-stepped *deterministic* simulation. The naval warfare model is limited in scope and includes tasks such as carrier aircraft operations that can be modeled using the air war model. The air-to-air engagement algorithms used in THUNDER are based on research performed by George S. Fishman and Louis R. Moore while at the University of North Carolina [74]. These algorithms have not changed substantially since THUNDER was first introduced in 1986.

In contrast to THUNDER, the Air Force primary model for analyzing air-to-air combat at the mission / engagement level is BRAWLER [49]. Operational since the early 1980's, BRAWLER is the Air Force's highest[9] resolution, analytical, air-to-air engagement model. BRAWLER is a discrete event, time-stepped stochastic simulation that models both with-in visual range and beyond visual range engagements. Multiple flights of aircraft are explicitly modeled through all phases of air-to-air combat. BRAWLER features detailed, data-driven models of most of the avionics and weapons found on modern fighter aircraft. Emphasis has been placed on the command and control aspects of air-to-air engagements including cooperative tactics, surprise, confusion, and situational awareness. BRAWLER is currently used to "tune parameters" in THUNDER's air-to-air engagement model [100: page 60].

In addition to THUNDER and BRAWLER, other air-to-air engagement models exist in the defense community. In terms of analysis models, several theater-level

---

[8]The ground war model is based on the Army's Concept Evaluation Model (CEM) which uses the Attrition Calibration (ATCAL) process and the U.S. Army Concepts and Analysis Agency's Combat Sample Generator (COSAGE) model to attrite ground combatants.

[9]"Highest" as explained in the military models section is a subjective term. In fact, certain aspects of BRAWLER might be of lower resolution than other air-to-air engagement models. However, looking over all aspects, BRAWLER has higher resolution than other models in a majority of cases. Hence the use of the term.

models consider air-to-air engagements such as TACWAR and the Theater Level Campaign (TLC) model[10]. Each of these models approaches the representation of air-to-air engagements differently. For example, TACWAR does not simulate individual flights of aircraft and uses a general attrition algorithm based on "firepower" scores to adjudicate air combat [109: page 1-4, 4-3]. In contrast, TLC uses an air-to-air engagement process based on explicitly modeled flights and the concept of "rounds" of air-to-air combat [100: page 74]. Additionally, mission / engagement level models exist to analyze specific problems such as force ratios. Examples may be found in [14], [72], and [103]. In terms of training models, theater level wargames are used to train command and staff personnel. Several of these models consider air combat such as the Air Force's Air Warfare Simulation (AWSIM) and the Army's Corp Battle Simulation (CBS). Although these training models are not designed for analysis work, they contain air-to-air engagement algorithms that may also have application for analysis models.

The most significant difference between air-to-air engagement models of "high" and "low" resolution is their treatment of the spatial and temporal aspects of modeling. Specifically, higher resolution air-to-air engagement models tend to describe spatial and temporal aspects in greater detail than lower resolution air-to-air engagement models. For example, BRAWLER models the spatial domain in three dimensions (longitude, latitude, altitude) down to units of feet and the temporal domain in seconds [49: page 2.2.1 and page 3.2.1 - 14]. In contrast, THUNDER models the spatial domain in two dimensions (longitude, latitude) down to units of meters with only a few discrete altitude settings and the temporal domain in minutes [29: Volume II, page 35] [103: page 24]. *In air-to-air engagement models, the spatial and temporal aspects of a model directly determine the level of attributes, con-*

---

[10]TACWAR is a deterministic, theater-level model of ground and air combat first developed in 1974 by the Institute of Defense Analysis. TLC is a research prototype, stochastic campaign level model of ground and air combat developed in the early 1990's by RAND. TLC serves as a test bed for investigating issues aimed at improving the next generation of theater-level models.

9

straint, and process aspects that can be effectively modeled. For example, even though both BRAWLER and THUNDER model air-to-air engagements down to the aircraft entity level, the spatial and temporal settings of THUNDER limit THUNDER from explicitly modeling the process of few on few air combat.[11]

1.4 Probabilistic Networks

A probabilistic network is a directed graph $G = (N, A)$ linked to an underlying probabilistic (stochastic) model in which probabilistic events (activities) are represented by arcs or nodes in G. Each probabilistic event is represented by a random variable or a combination of random variables. Probabilistic networks have a wide range of applicability in the literature. These networks are used to formulate and solve problems in such diverse areas as program management [65]; computer and communication systems [32] [186]; and artificial intelligence [150]. The appeal and power of probabilistic networks in formulating and solving problems lies in the graphical representation of the underlying probabilistic models. This graphical representation serves as a visual aid in conveying the structure of the probabilistic model and in formulating possible solution techniques.

This dissertation introduces a new probabilistic network termed an *event occurrence network* (EON) to model air-to-air engagements. An EON is a graphical representation of the superposition of several terminating counting processes. EONs are motivated by the research of Ancker, Gafarian, Kress and several associates [83] [120] [84] [102] [147] [198] in modeling m versus n stochastic ground combat. Event occurrence networks (EONs) are not limited to modeling combat situations. These networks can be applied whenever sets of random events occur.

---

[11]THUNDER can not model aircraft movement in an air-to-air engagement.

1.5 Problem Description

Little research has been devoted to obtaining complexity measures for probabilistic networks with continuous random variables. This situation is a direct result of the inability to compute exact analytical solutions for many probabilistic networks and has led to the use of approximation methods. Simulation-based methods have been the dominant approximation technique due to their logical simplicity and efficient implementation. When a simulation is used to solve a probabilistic network, the main complexity measure is the number of simulation runs and corresponding random variable draws necessary to establish a certain statistical confidence in the output.

However, for some probabilistic networks, simulation can become a computationally costly undertaking to obtain accurate estimates of the output measures. As the number of random variables in the network increases so does the number of replications necessary in order to properly characterize the output measure(s). As an alternative to simulation, numerical approximation and reduction methods have been proposed for some probabilistic networks. Network reduction involves the repeated application of operators such as multiplication (minimum, maximum, k out of n), convolution, and conditional integration to reduce a network to a smaller network through the aggregation of the underlying probabilistic model of the network. Numerical approximation involves the use of functions such as piecewise polynomial functions, expolynomial functions, exponential polynomials functions, etc. to approximate the density or distribution functions of the random variables. For probabilistic networks with several random variables, numerical approximation and reduction techniques can be several orders of magnitude faster than simulation-based methods without significant losses in accuracy.

The overall objective of this research is two-fold. The first goal is to develop a network formulation capable of expressing stochastic event occurrences and their interactions. The concept of event occurrence networks (EONs) will be researched.

EONs are graphical representations of the superposition of several terminating counting processes. The second objective of this research is to further the investigation of the piecewise polynomial functions to approximate density functions and distributions. This investigation includes the stochastic operations of subtraction, multiplication, convolution, and integration. The research will expand the work of Martin [134], Fergueson and Shorttell [73] and Lawrence [128] [127]. The polynomial approximation method developed by this research will be used to solve two analytical air-to-air engagement EONs.

1.6   Sequence of Presentation

The remainder of this dissertation is divided in six chapters. Chapter II provides a literature review of probabilistic networks, their complexity and output measures, and current solution methods. Specifically, the review looks at four different networks application areas: activity networks; stochastic Petri nets; task precedence graphs; and belief networks. Chapter III provides a review of air-to-air engagement literature including analytic and simulated models used for theater-level analysis. Chapter IV introduces event occurrence networks (EONs) and gives several examples of air-to-air engagement EONs. Chapter V formally defines a piecewise polynomial function and provides a literature review summarizing the research into approximating stochastic operators with piecewise polynomial functions. Chapter VI details the techniques/algorithms for solving event occurrence networks using piecewise polynomial approximation. The chapter shows two air-to-air engagement EON examples and compares the piecewise polynomial approximation technique with simulation. Finally, Chapter VII provides a summary of the research and recommendations for further research.

II. Probabilistic Network Literature

Probabilistic (stochastic) networks have a wide range of applicability in the literature. These networks are used to formulate and solve problems in such diverse areas as program management [65]; computer and communication systems [32] [186]; and artificial intelligence [150]. The appeal and power of probabilistic networks in formulating and solving problems lies in the graphical representation of the underlying probabilistic models. This graphical representation serves as a visual aid in conveying the structure of the probabilistic model and in formulating possible solution techniques. Before proceeding with a formal definition of a probabilistic network, a review of graph theoretic terms will be undertaken in order to set the framework for such a definition.

2.1 Graph Theoretic Underpinnings of Probabilistic Networks

The following graph theoretic terminology may be found in any standard graph theory text. Texts used for this review included [18] [182] [25]. A graph G is an ordered pair of disjoint sets (N, A) such that A is a subset of the set of unordered pairs of N. For our purposes, the sets N and A are finite. The set N is termed the set of nodes (or vertices) and the set A termed the set of arcs (or edges). An arc $\{x, y\}$ that joins the nodes x and y is denoted by xy or (x, y). If $xy \in A$, then nodes x and y are adjacent (neighboring) nodes of G and incident with xy. Two arcs are adjacent if they have exactly one common end node. The *order of* G is the number of nodes and is denoted by $|N|$. The *size of* G is the number of arcs and is denoted by $|A|$. $G(n, m)$ denotes an arbitrary graph of order n and size m. The size of a graph of order n is at least 0 and at most $\binom{n}{2}$. A graph $G' = (N', A')$ is a *subgraph* of $G = (N, A)$ if $N' \subset N$ and $A' \subset A$.

A graph G is a *r-partite* graph with node classes $N_1, N_2, \ldots, N_r$ if $N(G) = N_1 \cup N_2 \cup \ldots \cup N_r$ and $N_i \cap N_j = \emptyset$ whenever $1 \leq i < j \leq r$, and no arc joins two nodes in the same class. By definition, a graph G does not contain a loop (i.e., an arc joining a node to itself) nor multiple arcs (i.e., two or more arcs joining the same two nodes). In a multigraph H multiple arcs and loops are allowed. If the arcs of a graph G or multigraph H are identified with ordered pairs of nodes then the graph or multigraph is considered to be *directed*. Otherwise, G or H is undirected. An order pair (a, b) is said to be an arc directed from node a to node b and is denoted as ab or (a, b).

A directed path is a directed graph P of the form $N(P) = \{x_0, x_1, \ldots, x_l\}$ and $A(P) = \{x_0 x_1, x_1 x_2, \ldots, x_{l-1} x_l\}$ where the elements of $\{x_0, x_1, \ldots, x_{l-1}\}$ are distinct. The nodes x_0 and x_l are the end nodes of P and $l = \mid A(P) \mid$ is the length of P. A directed path in which $x_0 = x_l$ and $l \geq 1$ is termed a cycle. A directed graph G which contains no cycles is *acyclic*. Otherwise, it is cyclic. A directed acyclic graph (dag) D is *homeomorphic* from another dag D' if D can be obtained from D' by repeatedly inserting nodes of in-degree and out-degree one in the middle of arcs of D' where the in-degree and out-degree of a node x is the number of arcs incident into and out of x respectively. A dag D is *transitive* if for any two nodes x and y such that there is a path from x to y, either $x = y$ or xy is an arc. The transitive dag of G, $G_T = \{N, A_T\}$, is referred to as the transitive closure of G.

For an arc $x_i x_j \in A$, node x_i is an immediate *predecessor* (parent) of node x_j and node x_j is an immediate *successor* (child) of node x_i. For all $x_i \in N$, $\pi(x_i)$ denotes the set of immediate predecessors of x_i and $\theta(x_i)$ denotes the set of immediate successors of x_i. A directed graph G can be described by listing either $\pi(x_i)$ or $\theta(x_i)$ for all $x_i \in N(G)$. If $\pi(x_i) = \emptyset$, then node x_i is referred to as a source node. If $\theta(x_i) = \emptyset$, then node x_i is referred to as a sink node. Additionally, if there is a directed path from node x_i to node x_j and $x_i \neq x_j$ then node x_j is a descendant of

node x_i. In contrast, if no directed path exists from node x_i to node x_j, then node x_j is a non-descendant of node x_i.

Clearly, the power of graphs as analytical tools lies in their ability to visually represent structural relations. When using directed graphs, the concept of precedence relations is introduced. Precedence relations are specified by listing $\pi(x_i)$ for all $x_i \in N(G)$. Having introduced the graph theory terminology and the concept of precedence relations, we can now proceed with a definition of a probabilistic network. A probabilistic network is a directed graph $G = (N, A)$ linked to an underlying probabilistic (stochastic) model in which probabilistic events (activities) are represented by arcs or nodes in G. Each probabilistic event is represented by a random variable or a combination of random variables. Since G is directed, G allows for the representation of event precedence. Two possible modes of graphical representation exist:

1. Activity-on-Arc (AoA) representation and

2. Activity-on-Node (AoN) representation.

If $G = (N, A)$ is an AoA representation, the sets of arcs, A, denotes the events (activities) and the set of nodes, N, denotes the precedence relations among events. In contrast, if $G = (N, A)$ is an AoN representation, N denotes the events (activities) and A denotes the precedence relations among events. Elmaghraby, et al. [62] state that a AoN graph is unique due to the one-to-one correspondence of both activities and precedence, but an AoA graph is not.

2.2 Examples of Probabilistic Networks

2.2.1 Activity Networks. An activity network is a directed, acyclic graph that is used to represent activities and their interrelationships (i.e., precedence relationships). Figure 3 shows the structure of three different categories of activity networks. Activity networks have their genesis in solving program/project man-

agement planning and control problems and first appeared under the names of the critical path method (CPM) and program evaluation and review technique (PERT) networks in 1959 [62]. Probabilistic or stochastic activity networks are activity networks in which some or all of the activity durations are assumed to be random variables. These activity networks appear under the names PERT networks, probabilistic activity networks (PANs), and generalized activity networks (GANs) in the literature. As will be seen shortly, the name probabilistic activity network has a specific and limited meaning in the literature. As a result, the term stochastic activity network will be used in place of probabilistic activity network to denote an activity network in which some or all of the activity durations are assumed to be random variables. Elmaghraby [65], Adlakha and Kulkarni [1], Lawrence [128] provide overviews of research on stochastic activity networks.

Activity networks separately define the terms event and activity. An event is a "well-defined occurrence in time" and an activity is "any undertaking that consumes time and resources" [65]. The separation of the event and activity terms clarifies the concept that an event may depend upon several activities, with a specified number of these activities having to be accomplished in order for the event to be accomplished[1]. If the principal resource of an activity network is time, precedence relationships among events determine which activities must be completed before other activities can be started and thus determine the order of occurrence of the activity network's events.

Elmaghraby [65] makes two remarks about precedence in activity networks:

1. Precedence is a binary relationship; if an activity a precedes another activity b (denoted $a \prec b$), then a must be completed before b is started.

2. Precedence is a transitive relationship; if $a \prec b$ and $b \prec c$, $a \prec c$.

---

[1]This concept is implied in our definition of probabilistic network.
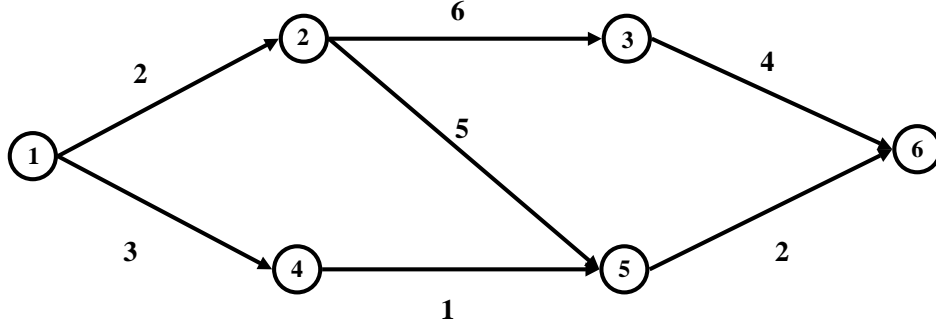
Deterministic Activity Network (DAN)
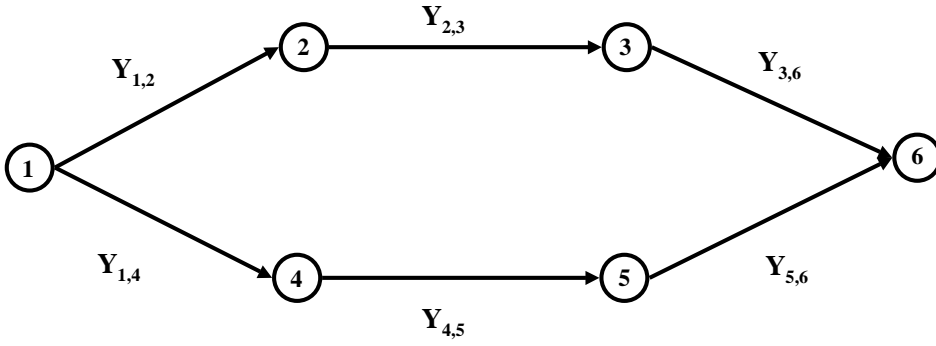
Probabilistic Activity Network (PAN)
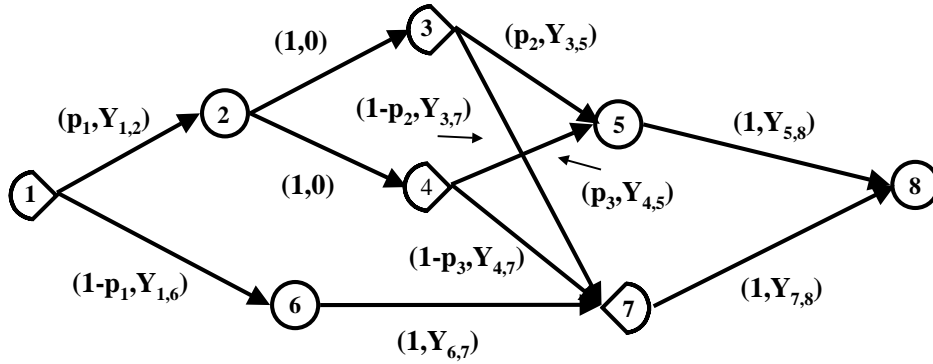
Generalized Activity Network (GAN)

Figure 3. Three Categories of Activities Networks: DAN, PAN, and GAN

Since precedence is a transitive relationship, looping (cycling) is not permitted in activity networks. In other words, if activity a precedes activity b and activity b precedes activity c then activity c cannot precede activity a. Hence, activity networks are directed, acyclic graphs.

Activity networks can be modeled using either an AoA or AoN graphical representation. As will be seen in section 2.3, algorithms exist for converting a AoN graphical representation into a AoA representation and vice versa. Activity networks presented in the literature generally use the AoA representation. When discussing activity networks, an AoA representation will be assumed for the remainder of this dissertation unless otherwise noted. Elmaghraby [65] divided activity networks into three categories:

1. Deterministic Activity Networks (DANs);

2. Probabilistic Activity Networks (PANs); and

3. Generalized Activity Networks (GANs).

Figure 3 illustrates the three activity network categories. DANs are activity networks in which activity (arc) durations are assumed to be constant. In contrast, PANs are activity networks in which activity (arc) durations are assumed to be random variables. Both DANs and PANs require that event (node) realizations occur only when each of the event's successor activities (arcs) are accomplished. This requirement corresponds to the AND logical operator[2]. For this reason, DANs and PANS are also referred to as all-AND networks. In an all-AND activity network such as a DAN or PAN, all activities in the network must be accomplished in order for the activity network to be completed.

In order to relax the precedence relations among activities, all-AND activity networks were extended along two lines. First in DANs, strict precedence among

---

[2] All successor activities (arc) of an event (node) are started when all of the event's (node's) predecessor activities (arcs) are realized.

activities (i.e., completing all preceding activities before starting a successor activity) was relaxed which led to DANs with generalized precedence relations (GPRs) [64]. Second, in DANs and PANs, the non-realization of activities was permitted (i.e., every activity in the project does not have to completed) which led to generalized activity networks (GANs) [59] [60]. In contrast to DANs and PANs, GANs expand the node (event) logical operator set. In a GAN, a node (event) is constructed from the combination of an input (receiver) and output (emitter) side. Three input sides are possible:

1. AND - a node (event) is realized when all arcs (activities) leading into the node are realized;

2. INCLUSIVE-OR - a node is realized when any one arc or combination of arcs leading into the node is realized; and

3. EXCLUSIVE-OR - a node is realized when one and only one arc leading into the node is realized.

Two output sides are possible:

1. DETERMINISTIC (must follow) - a node from which all emanating arcs are undertaken and

2. PROBABILISTIC (may follow)- a node from which an emanating arc may be realized with probability ≤ 1.

Since each node has a receiver and emitter, six node types are possible in a GAN. Figure 4 shows the possible combinations. Additionally, a GAN can accommodate activity (arc) durations that are constant and/or random variables. Clearly, a DAN is a GAN in which all nodes (events) are of the AND-DETERMINISTIC type and all activities (arc) durations are constant. Similarly, a PAN is a GAN in which all nodes (events) are also of the AND-DETERMINISTIC type, but whose activity (arc) durations are random variables.
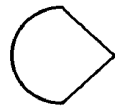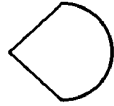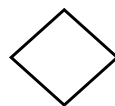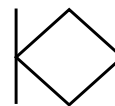
Output Input	Deterministic	Probabilistic
AND		
Inclusive-OR		
Exclusive-OR		

Figure 4. Six Nodes Types in a Generalized Activity Network

Since Elmaghraby's original work [59] [60] on GANs in the mid-1960's, very little work has appeared in the literature modifying the structure of activity networks. Instead this work has proceeded in the commercial realm due to the wide spread adoption of simulation as the preferred analytic tool for solving GANs. (Solution methods for probabilistic networks will be discussed later in section 2.4). One of the main pioneers in the use of simulation as an analytic tool for solving GANs is Pritsker. Following the introduction of the general evaluation and review technique (GERT) [157] [159] to solve all EXCLUSIVE-OR activity networks, Pritsker developed a family of GERT simulation (GERTS) tools to solve more general networking problems [156]. Through the use of simulation, event (node) logical operators and activity (arc) types have greatly expanded. Today, a simulation of an activity network using a general purpose simulation environment such as VISUAL SLAM can have over twenty different node types and five arc types [158].

In Figure 3, all three activity network category examples are shown with a single start node termed a *source* node and a single termination node termed a *sink* node. Referred to as a "two terminal" network [13], this is the standard form for
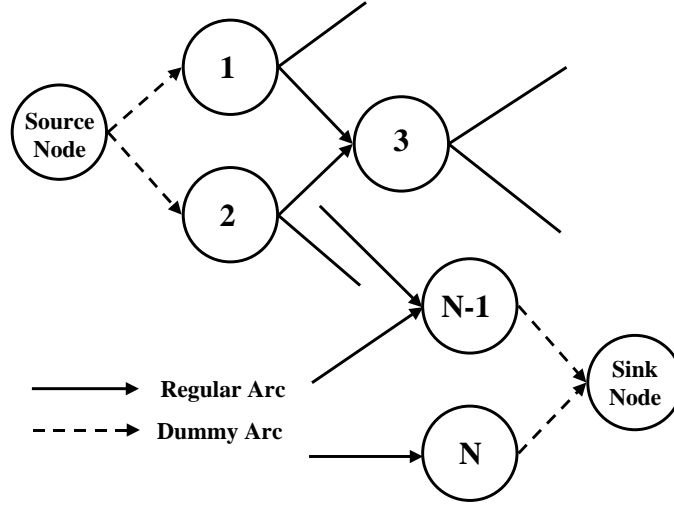
Figure 5. Conversion of a Multiple Source and Sink Activity Network Into a Two Terminal Activity Network

activity networks found in the literature. An activity network may have multiple sources and sinks, but it usually converted into a two terminal network through the use of dummy nodes and arcs (activities of zero duration). As shown in Figure 5, this conversion is accomplished by inserting a dummy (source) node before the source nodes and connecting the dummy node to the source nodes via arcs of zero duration. Similarly, the sink nodes are connected to a single dummy (sink) node through dummy arcs of zero duration.

In the literature, the three primary metrics of interest in analysis of activity networks are:

1. Network (project) completion time;

2. Shortest path; and

3. Activity and path criticalities.

Clearly for stochastic activity networks, network completion time is expressed in the form of a distribution. Adlakha and Kulkarni [1], Lawrence [128] provide a comprehensive review of research on the completion time of stochastic activity networks.

Analogous to network completion time, the shortest path in an activity network is a useful metric for managers. Alexopoulos [3] provides a review of research on the shortest path metric. In addition to network completion time and shortest path, activity and paths criticalities provide measures of the activities and paths respectively that are most likely to compete for the longest duration in a network. These criticalities help managers identify those tasks which can become bottlenecks in a project/program. Bowman [21] provides a review of research on activity and path criticalities.

2.2.2 Petri Nets.   A Petri net is a directed, bipartite graph. Figure 6 shows an example of a Petri net. Petri nets were first used to model computer and communication systems performance, but the use of Petri nets has expanded to include other systems where the modeling of concurrency (parallelism) and synchronization (control of parallelism) are important. Murata [141], and Ciardo, et al. [32] provide overviews of research on Petri Nets. A Petri Net is generally defined by a 5-tuple $(\mathcal{P}, \mathcal{T}, \mathcal{I}, \mathcal{O}, \mathcal{M})$ where $\mathcal{P}$ is the set of places, $\mathcal{T}$ is the set of transitions, $\mathcal{I} \subset \mathcal{P} \times \mathcal{T}$ is the input function, $\mathcal{O} \subset \mathcal{T} \times \mathcal{P}$ is the output function, and $\mathcal{M} \subset \mathcal{P} \times Z^+$ is a marking (Z^+ denotes the set of all nonnegative integers) [117].

As shown in Figure 6, the two disjoint sets of nodes of the bipartite graph are drawn as circles which correspond to places and bars which correspond to transitions. Places contain tokens (possibly multiple tokens) which are drawn as dots. A place is input to a transition if there is an arc from the place to transition. Similarly, a place is the output of a transition if there is an arc from the transition to the place. A marking of a Petri net is given by a vector μ which contains as its entries the number of tokens in each place. Tokens move based upon the enabling and firing of transitions. A transition is enabled in a marking if all of its input places contain at least as many tokens as the multiplicity of the corresponding input arc from that place. A transition fires by removing tokens from each of its input places and depositing tokens to the output places according to the output arc multiplicity. Each
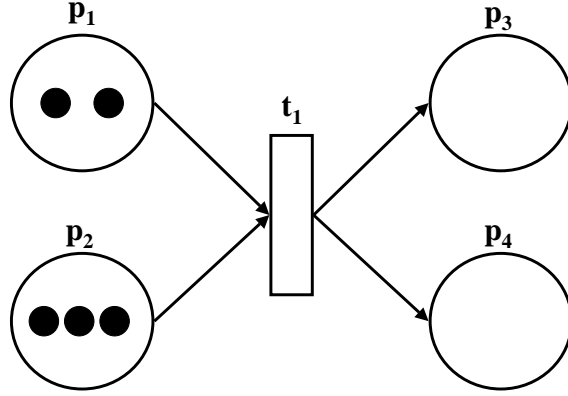
Figure 6. Petri Net Example

firing of a transition results in a new marking of the Petri net. Each marking defines a state of the system. If the number of tokens in the Petri net is bounded, then there are a finite number of markings. The reachability set of a Petri net is defined as the set of all markings that are reachable from the initial marking. The reachability tree is a graphical representation of the reachability set where nodes represent the various reachable states (i.e., markings) and arcs represent possible paths from one state (marking) to another state. A reachability graph is similar to a state space transition diagram. Petri nets have been extended for increased ease of use and enhanced modeling power. For instance, inhibitor arcs can be allowed that prevent firing of a transition when there is a token in any one of its inhibitor places [32] [133]. As an illustration of a firing, Figure 7 shows the result of transition t_1 firing from figure 6. Initial marking $\mu^0 = (2, 3, 0, 0)$ becomes marking $\mu^1 = (1, 0, 1, 3)$ after t_1 fires.

A Stochastic Petri net (SPN) is a Petri net where the transition firings are:

1. immediate (i.e., zero firing time) or

2. timed (i.e, probabilistic firing time distribution).
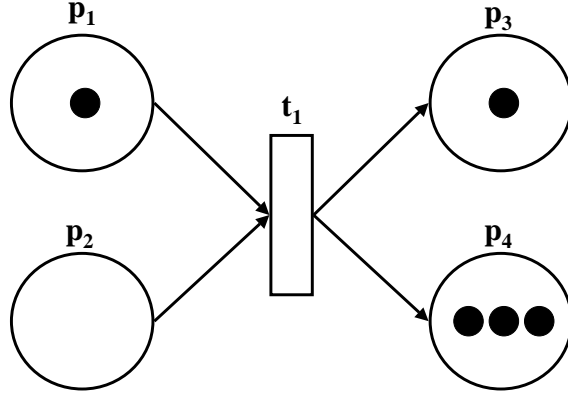
Figure 7. Petri Net Example, Transition Firing

The firing of immediate transitions is given priority over the firing of time transitions. Additionally, each immediate transition is given a weight (probability) which determines its firing probability in case the transition is in conflict with another immediate transition. As a consequence of the two types of transition firings, the reachability set a stochastic Petri net can be divided into vanishing markings (at least one immediate transition is enabled) and tangible markings (otherwise). The tangible markings of a stochastic Petri net correspond to the states of an underlying stochastic process termed the marking process [32] [133].

Based on the above discussion of Petri nets, the reader of this dissertation might have the feeling that there is an underlying correspondence between Petri nets and activity networks. In fact, there is correspondence between the two network types albeit an informal one. The parallelism between Petri nets and activity networks has been addressed formally by several authors in the literature. Quichaud and Chretienne [161] were the first to suggest a link between Petri net theory and activity networks. They derived performance evaluation results for the transient behavior of bipolar synchronization schemes and also proposed an extension of the standard

GERT analysis to networks containing AND nodes (recall the standard GERT analysis was restricted to all EXCLUSIVE-OR activity networks). Elmaghraby [61] elaborated on this link citing the following similarities and differences. Similarities between Petri nets and activity networks include:

- Modeling of an entity moving through a network (i.e., Petri nets use tokens while activity networks use transactions);

- Modeling of activities and relationships among activities including such topics as concurrency, synchronization, etc.; and

- Exploring similar analysis issues (e.g., deadlock (Petri net) vs feasibility (activity network) or reachability (Petri net) vs realizability (activity network)).

Differences between Petri nets and activity networks include:

- Formalism: from their inception, Petri nets addressed issues of logical concurrency (such as reachability, coverability, liveness, boundedness, conversation, etc.) while activity networks focused on issues of timing, economic value, and resource allocation;

- Concept of State: the concept of state is prominent in Petri nets and lacking in activity networks; and

- Resource Representation: it is much harder to represent resources in Petri nets than activity networks.

In conclusion, Elmaghraby stated that activity networks have more potent and general capabilities than Petri nets to model and analyze systems; however he noted that Petri nets retain a logical structure, especially when representing actions that are conditional upon the realization of other events. He proposed research into a "paradigm" that blended the best features of both Petri nets and activity networks.

Elmaghraby, et al. [66] further discussed the blending of concepts from Petri net and activity network theory. The authors proposed a new paradigm PETAN

25

(PETri/Activity Nets) as an approach to the analysis and design of software production processes. The PETAN paradigm combined the theory of generalized activity networks (GANs), available simulation software (SLAM II[3]) for modeling (GANs), and the theory of Petri nets. Elmaghraby, et al. stated that GANs possess unmatched analytical power as a result of being derived from the theory of stochastic processes and software such as SLAM II provided a means to simulate the GANs and hence the underlying stochastic process. Unfortunately, as the authors stated, modeling through either activity networks or stochastic process software is often non-intuitive and demanding in time and resources. However, the authors proposed that the theory of Petri nets provided a highly intuitive modeling process. In the example provided by the authors, a SLAM II simulation model of a software production process GAN is given and solved and a Petri Net model of the GAN is created to intuitively explain the events of the process.

Other researchers have discussed the link between Petri net and activity network theory. Cubaud [45] showed that the computation of transition firing dates in a stochastic event graph (a stochastic Petri net for which each place has at most one input transition and at most one output transition) is equivalent to the computation of the completion time of a PERT network. Lee and Murata [129] introduced a beta-distributed stochastic Petri Net (BSPN) which integrated the PERT activity networks and Petri nets to solve uncertainty and concurrency problems in large software projects.

2.2.3 Task Precedence (Series-Parallel) Graphs. A task precedence graph is a series-parallel directed, acyclic graph. Figure 8 shows an example of a task precedence graph. Task precedence graphs are used to model and evaluate parallel applications such as parallel programs. Trogemann and Gente [186] provide an overview of research on task precedence graphs. The nodes of a task precedence

---

[3]SLAM II is a predecessor of VISUAL SLAM.
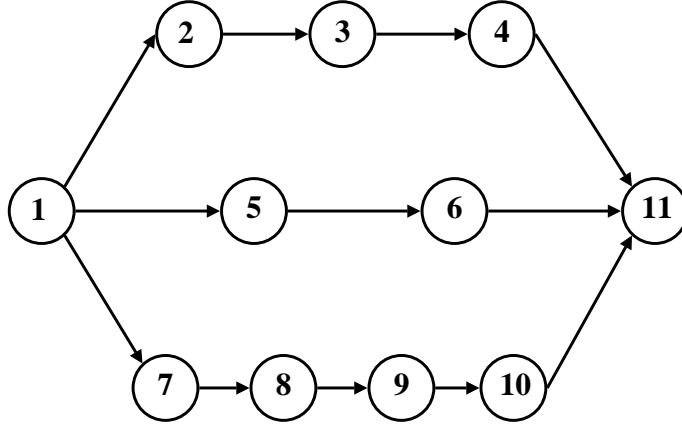
Figure 8. Task Precedence Graph Example

graph such as shown in Figure 8 represent the subtasks of the parallel applications and the arcs represent the precedence relationships between the subtasks. Task precedence graphs are subsets of AoN activity networks. As will be shown in section 2.3, the structure of these precedence graphs and series-parallel directed graphs in general enable efficient solution (in polynomial time). Solutions to series-parallel directed graphs will be discussed in greater detail in section 2.4.

2.2.4 Belief Networks. Informally, a belief network [150] is a directed, acyclic graph linked to a joint probability distribution with certain independence properties. Figure 9 shows two simple examples of a belief network. Belief networks provide graphical representations of causal dependence. These networks are used to model dependencies in such diverse areas as medical diagnosis, natural language understanding, circuit fault diagnosis, pattern recognition, machine vision, financial auditing, map learning, sensor validation, and forecasting [114].

Belief networks have also been called Bayesian networks, causal networks [150], probabilistic influence diagrams [171], or knowledge maps [95]. Pearl [150] and Jen-

27
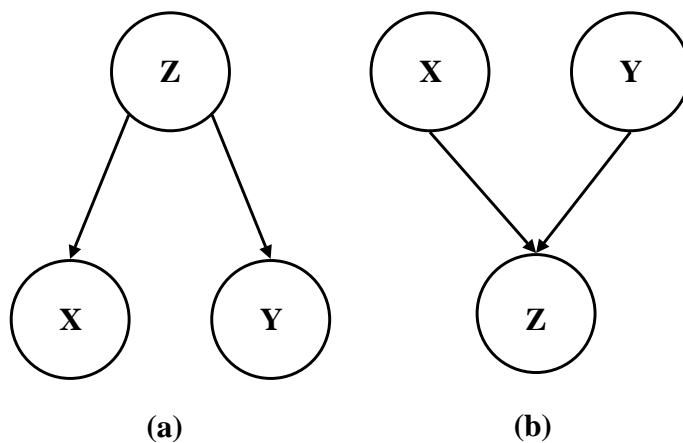
Figure 9. Belief Network Example

zarli [114] provide overviews of research in this area. The nodes of a belief network represent random (termed chance) variables whose distributions are quantified by presence or absence of predecessor arcs into the nodes. Specifically, the chance variable of a node with predecessor arcs is represented by a conditional distribution given its immediate node predecessors (termed parents). Additionally, the chance variable of a node without predecessor arcs is represented by a marginal distribution. The joint distribution for all the chance variables in the belief network is then obtained by multiplying all the conditional and marginal distributions of the nodes in the network [150].

For example, consider the two belief networks in Figure 9. X, Y, and Z represent chance (random) variables in both networks. In belief network (a), the joint probability distribution of X, Y, and Z denoted $f(x, y, z)$ is specified as:

$$f(x, y, z) = f(z)f(x \mid z)f(y \mid z)$$

where $f(z)$ is the marginal probability distribution of Z and $f(x \mid z)$ and $f(y \mid z)$ are the conditional probability distributions of X and Y respectively given $Z = z$. In this case, node X and node Y both have Z as a predecessor. In belief network (b), $f(x, y, z)$ is specified as:

$$f(x, y, z) = f(x)f(y)f(z \mid x, y)$$

where $f(x)$ and $f(y)$ are the marginal probability distributions of X and Y respectively and $f(z \mid x, y)$ is the conditional distribution of Z given $X = x$ and $Y = y$. In this case, node X and node Y are both predecessors of node Z.

Jenzarli [114] gave the following formal definition of a belief network: a belief network is pair $B = (D, P)$, where $D = (N, A)$ is a directed, acyclic graph with variables as nodes and P is a joint probability distribution for the variables in N that can be factored into conditional probabilities, one for each variable given its immediate predecessors. More precisely, if $X_1, \ldots, X_n$ are the random variables (represented as network nodes) and $\pi(x_i)$ denotes the set of immediate successors of node x_i then the joint distribution for the variables in N can be represented as follows [150]:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \pi(x_i)) \tag{1}$$

where $P(x_i \mid \pi(x_i) = \emptyset)) = P(x_i)$.

Shimony [174] stated that a belief network saves many orders of magnitude in the size of representation. Equation 1 shows where the savings in size of representation comes from. For example, representing the full distribution of n binary (and thus discrete) random variables directly requires 2^n probabilities. Using a belief network, one only needs [150]:

$$\sum_{i=1}^{n} 2^{\pi(x_i)}$$

probabilities. Since in most belief networks appearing in the literature, the maximum in-degree of a typical network is much less than n, a considerable savings in size of representation is achieved.

Pearl [150] stated the following important characteristics of directed acyclic graphs and belief networks:

1. Criterion of "d-separation" : if N_1, N_2, and N_3 are three disjoint subsets of nodes in a directed, acyclic graph D, then N_3 is said to be d-separate N_1 from N_2, if along every path between a node in N_1 and a node in N_2 there is a node x_w satisfying one of the following conditions: (a) x_w has "converging arrows" and neither x_w or its descendants are in N_3, or (b) x_w does not have "converging arrows" and x_w is in N_3. For example in Figure 9, belief network (a), nodes X and Y are d-separated by node Z. D-separation permits determination of which set of variables are independent of each other given a third set in a belief network.

2. Given a directed, acyclic graph D and a joint probability distribution P, a necessary and sufficient condition for D to be a belief network of P is that each variable X be conditionally independent (with respect to probability distribution) of its non-descendants variables (represented by non-descendant nodes in D) given its immediate predecessor variables (represented by predecessor nodes in D).

Jenzarli [114] showed a correspondence between activity and belief networks by demonstrating how to model the AoN representation of a probabilistic activity network (PAN) as a belief network. Jenzarli terms these networks PERT belief networks. In PERT belief networks, the nodes represent completion times and the arcs represent probabilistic dependencies between nodes. The completion times of node i: $i = 1, \ldots, n$ denoted C_i, is given by:

$$C_i = max(C_j \mid C_j \in \pi(C_i)) + D_i$$

where D_i is the duration of activity i, $i = 1, 2, \ldots, n$. The joint probability of $C_1, \ldots, C_n$ is determined from equation 1 where C_n is the network completion time.

2.3 Complexity of Probabilistic Networks

Loosely defined, the complexity of a probabilistic network refers to the difficulty in analyzing a network. As will be shown later in this section, the complexity of a probabilistic network is directly related to three factors and their interactions: the topology of the network; the underlying probabilistic model(s); and the solution method(s) used. Solution methods for probabilistic networks will be discussed in more detail in section 2.4; however, in this section, some solution methods will be briefly introduced in order to show how these methods impact on the issue of complexity in probabilistic networks. Network complexity measures are quantitative descriptions of the complexity of a network and are used to

- Predict analysis processing time requirements; and

- Compare two or more proposed algorithms by ensuring that algorithms are evaluated at several points in the range of complexity [115] [63].

As an example of complexity of probabilistic networks and their measures, consider the probabilistic activity networks (PANs) introduced in section 2.2.1.

Figure 10(a) and (b) shows two different PANs each with four nodes and five arcs. As stated in section 2.2.1, a primary metric of interest in analysis of activity networks is network (project) completion time. For PANs, network completion time is expressed in the form of a distribution. Let X_{ij} denote the random variable associated with the duration of arc (activity) ij with density function f_{ij} and distribution function F_{ij}. The network completion time, denoted T_n, of a PAN with n nodes is:

$$T_n = \max_{\pi_k}(T(\pi_k)) \tag{2}$$
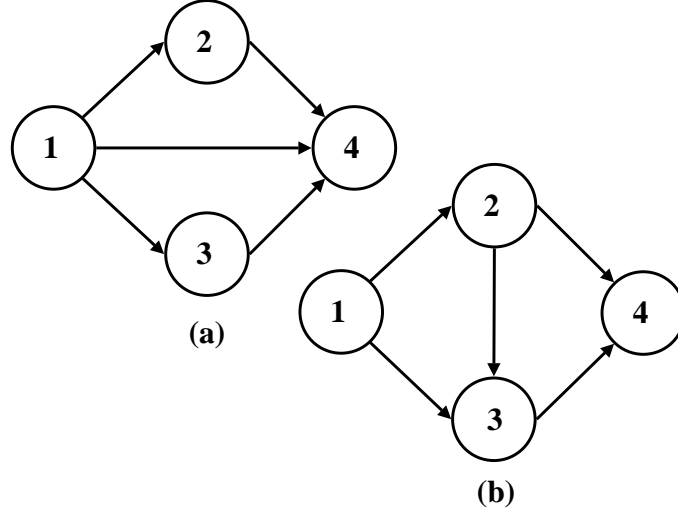
Figure 10. Complexity Example - PAN

where $\pi_k, k = 1, 2, \ldots, r$ are the paths that lead from node 1 to node n and $T(\pi_k)$ is the completion time of path π_k [65]. For Figure 10(a) and (b), the network completion time is:

$$T_4 = \max(T(\pi_1), T(\pi_2), T(\pi_3))$$

where in part (a):

$$T(\pi_1) \quad = \quad X_{12} + X_{24} \tag{3}$$

$$T(\pi_2) \quad = \quad X_{14} \tag{4}$$

$$T(\pi_3) \quad = \quad X_{13} + X_{34} \tag{5}$$

and in part (b):

$$T(\pi_1) \quad = \quad X_{12} + X_{24} \tag{6}$$

$$T(\pi_2) \quad = \quad X_{12} + X_{23} + X_{34} \tag{7}$$

$$T(\pi_3) \quad = \quad X_{13} + X_{34} \tag{8}$$

From equation 2 and equations 3 thru 5, the network completion time for Figure 10(a) is:

$$T_4 = \max(X_{12} + X_{24}, \ X_{14}, \ X_{13} + X_{34}) \qquad (9)$$

and from equation 2 and equations 6 thru 8, the network completion time for Figure 10(b) is:

$$T_4 = \max(X_{12} + X_{24}, X_{12} + X_{23} + X_{34}, X_{13} + X_{34}) \qquad (10)$$

In order to determine the completion times of the PANs, three basic mathematical operators are needed:

1. Convolution;

2. Maximum (multiplication); and

3. Conditional Integration [134] [65] [63].

Graphically, these operators enable a PAN to be 'reduced' to a two terminal network (source and sink) with a single arc between the nodes. The convolution operator is referred to as a 'series reduction' operator. Two arcs are said to be in *series* if the arcs are separated by a single node with only one arc incident into the node and the other arc emanating from the node. With this series condition, the two arcs can be combined into a single arc with the convolution operator since the distribution function of the duration of the new arc can be obtained by convoluting the distribution functions of two arcs in series. Graphically, a series reduction at node v is possible when arc (u, v) is the unique arc into v and (v, w) is the unique arc out of v. The series reduction then eliminates node v and replaces arcs (u, v) and (v, w) with arc (u, w).

Similarly, the maximum operator is a 'parallel reduction' operator. Two arcs are said to be in *parallel* if the arcs have the same starting and ending nodes. With this parallel condition, the two (or more) arcs can be combined into a single arc

with the maximum operator since the distribution function of the duration of the new arc can be obtained by multiplying the distribution functions of the two (or more) arcs in parallel. Graphically, a parallel reduction at nodes v and w replaces two (or more) arcs joining the nodes with a single arc (u, v). As an example of the convolution and maximum operators consider the PAN in Figure 10(a). Network reduction may proceed as follows:

1. Obtain $Y_1 = X_{12} + X_{24}$, a convolution operation;

2. Obtain $Z_1 = \max(Y_1, X_{14})$, a maximum operation;

3. Obtain $Y_2 = X_{13} + X_{34}$, a convolution operation; and finally,

4. Obtain $T_4 = \max(Z_1, Y_2)$, a maximum operation.

Consequently two convolution and two maximum operations are needed to reduce the network in Figure 10(a) and determine the network completion time. Analytically, the solution to equation 9 is:

$$F_{T_4}(t) = (\int_0^t F_{X_{24}}(t - x_{12})dF_{X_{12}}) * F_{X_{14}} * (\int_0^t F_{X_{34}}(t - x_{13})dF_{X_{13}})$$

Note the order in which the series and parallel reductions are applied does not affect the final outcome, since the reduction obeys the Church-Rosser property [190].

The PAN in Figure 10(b) can not be completely reduced using only the convolution and maximum operators since the three paths π_1, π_2, and π_3 are dependent. Specifically, arc X_{12} appears in paths π_1 and π_2 and arc X_{34} appears in paths π_2 and π_3. This dependency can be removed by conditioning on the values of the common arcs that are elements of more than one path. The PAN in figure 10(a) may be reduced through two maximum and two conditional integration (conditioning on X_{12} and X_{34}) operations. Analytically, the solution to equation 10 is:

$$F_{T_4}(t) = \int_0^t \int_0^t F_{X_{24}}(t - x_{12})F_{X_{23}}(t - x_{12} - x_{34})F_{X_{13}}(t - x_{34})dF_{X_{12}}dF_{X_{34}}$$

If one assumes that the conditional integration is the most computational time consuming operator, followed by the convolution operator, and then the multiplication operator, then the computing requirements of the PAN in Figure 10(b) will be greater than the requirements of the PAN in Figure 10(a). In other words, informally, it appears that the PAN in Figure 10(b) is more complex than the PAN in Figure 10(a) if one uses computing requirements as a measure of network complexity.

Formally, Elmaghraby and Herroelon [63] define network complexity in an activity network context 'as the difficulty in analysis and synthesis of a given network'. The authors argued that the measurement of network complexity cannot be meaningfully accomplished unless the use of the measure (i.e., objective of the analysis) is known beforehand. For PANs, Elmaghraby and Herrolen give the following measure of network complexity (MNC):

$$MNC = g(v, m, q) \qquad (11)$$

where

- g is a (monotone increasing) calibration function which is determined empirically, and whose form and value depend on the computer hardware and software and the programmer's skill;

- v is the number of convolutions;

- m is the number of maximums; and

- q is the number of conditional integrations.

For example in Figure 10(a), $v = 2$, $m = 2$, and $q = 0$ and in in Figure 10(a), $v = 0$, $m = 2$, and $q = 2$ resulting in $g(2, 2, 0)$ and $g(0, 2, 2)$ respectively.

As shown in equation 2, the number of maximum operators in a PAN is equal to the number of independent paths π in parallel. In the absence of path independence, the number of maximum operators is directly related to the number of

35

conditioned arcs. Since at the time of their writing no systematic procedure existed for determining the minimum number of arcs to be conditioned on, Elmaghraby and Herrolen concluded that the number of maximum and conditional integration operators must be determined conjointly. In addition, the authors argued that the measure of complexity may be confounded by the algorithm (solution method) employed. For example, the MNC in equation 11 measures the complexity of a PAN assuming that the analytical procedure used is that of convoluting activities in series and maximizing activities in parallel, and conditioning on common activities.

The graph in Figure 10(b) is referred in the literature as the '*interdictive* (forbidden)' graph [56] [190] [55] [54] [13]. Duffin [56] showed that if a two-terminal directed, acyclic graph (st-dag) does not contain a subgraph homeomorphic from the interdictive graph, then the graph is series-parallel[4]. Specifically, Duffin proved that a st-dag is series-parallel if and only if the graph does not contain a subgraph homeomorphic from the interdictive graph. Series-parallel st-dags may be either edge (arc) series-parallel or vertex (node) series-parallel [190]. The PAN in figure 10(a) which has AoA graphical representation is an example of an edge series-parallel st-dag while the task precedence graph defined in section 2.2.3 is an example of a vertex series-parallel st-dag. Since the literature that is of relevance to this dissertation uses edge series-parallel terminology, the term series-parallel will be used to mean edge series-parallel unless otherwise noted. A st-dag is series-parallel if it can be obtained iteratively in the following way:

- A single arc is two-terminal series-parallel (with the tail being the source and the head being the sink);

- If G_1 and G_2 are two terminal series-parallel, so is the graph obtained by identifying the sources and sinks, respectively (parallel composition); and

---

[4]Series-parallel graphs may also be defined by recursive definitions; definitions that specify graph characteristics; and constructive definitions. See [164].

- If G_1 and G_2 are two terminal series-parallel, so is the graph obtained by identifying the sink of G_1 with the source of G_2 (series composition) [13].

As the name implies, series-parallel st-dags are *completely reducible* (separable) to a single arc through series and parallel reductions. In the case of PANs, the series and parallel reduction translates to a convolution and maximum operation respectively in the underlying probabilistic model. In contrast, st-dags with embedded interdictive graphs are *irreducible* (non-separable). Series-parallel reductions may still be performed, but eventually the st-dag will be in an irreducible form. For PANs, Dodin [55] proved that a PAN is not completely reducible if and only if it contains the interdictive graph.

Bein, et al. [13] showed how to eliminate embedded interdictive graphs in a st-dag by applying successive node reductions. A node reduction is a generalization of a series reduction and can occur whenever a node v has in-degree or out-degree 1. If v has in-degree 1 and (u, v) is the arc into v and $(v, w_1), \ldots, (v, w_k)$ are the arcs out of v, then the node reduction is accomplished by deleting node v and replacing the $k + 1$ arcs (u, v) and $(v, w_1), \ldots, (v, w_k)$ with the k arcs $(u, w_1), \ldots, (v, w_k)$. The case where v has out-degree 1 is symmetric. A node reduction at v on graph G is denoted by $G \circ v$.

Based on the number of node reductions required to reduce a st-dag with embedded interdictive graphs, Bein, et al. [13] introduced a st-dag complexity measure that describes how nearly series-parallel a st-dag is. This measure termed the reduction complexity of G, denoted $\mu(G)$, is the minimum number of node reductions sufficient (along with series and parallel reductions) to reduce G to a single arc. Specifically, $\mu(G)$ is smallest c for which there exists a sequence $v_1, \ldots, v_c$ such that $[\ldots[[[G] \circ v_1] \ldots \circ v_c]$ is a single arc where $[G]$ denotes the graph that results when all possible series and parallel reductions have been applied to G.

Bein, et al. [13] proved that there exists a polynomial-time algorithm to minimize node reductions. The algorithm is obtained by showing that $\mu(G)$ is equal to

the size of a minimum node cover of its complexity (transitive) graph $C(G)$. The complexity graph, $C(G)$, of a st-dag G is defined as follows: The arc (v, w) is in $C(G)$ if and only if there exits paths $\pi_1(v, w)$, $\pi_2(v, w)$, $\pi(1, w)$, and $\pi(v, n)$ such that $\pi_1(v, w) \cap \pi(v, n) = \{v\}$ and $\pi(1, w) \cap \pi_2(v, w) = \{w\}$. Furthermore, either $\pi_1(v, w) = \pi_2(v, w)$ or $\pi_1(v, w) \cap \pi_2(v, w) = \{v, w\}$, and the arcs common to $\pi(1, w)$, and $\pi(v, n)$, if any, form a single path. The definition implies that neither nodes 1 nor n appears in $C(G)$. Constructing $C(G)$ is equivalent to computing the transitive closure of G. Since $C(G)$ is a transitive dag, the minimum node cover for $C(G)$ denoted N' can then be computed by finding the maximum matching in a bipartite graph - (the complement of a minimum node cover is a maximum independent set, which in a transitive dag corresponds to a Dilworth chain decomposition [76]).

Bein, et al. [13] proposed the following reduction sequence for a st-dag G

- Step 1: Construct $C(G)$

- Step 2: Compute N'

- Step 3: While $N' \neq 0$ do:

    - Perform all series and parallel reductions possible in G and let $G = [G]$

    - Find $v \in N'$ such that the in-degree or out-degree of node v is 1

    - Let $G = G \circ v$

    - Let $V' = V' - \{v\}$

The authors showed that the first two steps in the reduction sequence occur in a time bound of $O(n^{2.5})$. This time estimate is based on an upper bound of $O(n^{2.37})$ for computing the transitive closure of a graph of n vertices (i.e., step 1) [40] and a time bound of $O(n^2)$ for computing the maximum matching in a bipartite graph [76].

Reyck and Herroelen [162] showed an example of constructing a $C(G)$ and computing N' for a st-dag containing 9 nodes and 14 arcs. The authors note that although the reduction complexity index (CI) introduced by Bein, et al. [13] was

defined only for AoA networks, the CI can be used for a AoN network by transferring a AoN network into a AoA network representation. Reyck and Herroelen mentioned the work of Kawburowski, et al. [116] who develop a polynomial time algorithm which generates an AoA network with minimal number of nodes and CI from a given AoN network.

Although Reych and Herroelen [162] applied the reduction CI to the multiple resource constrained problem (RCPSP) in which it is assumed that an activity is subject to technological precedence constraints and cannot be interrupted once begun and the discrete time/cost trade-off problem (DTCTP) in activity networks of the CPM type, using a single nonrenewable resource, *the reduction CI has not been applied to stochastic networks to our knowledge.* Instead, Lawrence [128] defined two interdependent complexity measures for PANs:

- The first complexity measure is the 2-tuple $(c, \aleph_c)$, where c is the number of *terminating* cross-connections in the network and $\aleph_c$ is the cardinality of these cross-connections; and

- The second complexity measure is the 2-tuple $(c', \aleph'_c)$, where c is the number of *starting* cross-connections in the network and $\aleph'_c$ is the cardinality of these cross-connections.

Lawrence defined a terminating cross-connection as the ending node, x_j, for paths $\pi_i, i \geq 2$ emerging from a common node x_i that is predecessor to node x_j, but not an immediate predecessor. x_j is referred to as the terminal node for the cross-connection. Similarly, a starting cross-connection is the starting node, x_i for paths $\pi_i, i \geq 2$ that end at nodes $x_j, j \geq 2^5$. The presence of a terminating or starting cross-connection indicates a dependency among paths in the PAN (i.e., the PAN does not have a series-parallel topology).

---

[5]Note the number of paths and ending nodes are the same.

Lawrence [128] further showed that terminating and starting cross connections may be determined from the adjacency matrix, M, of the PAN. The adjacency matrix $M = [m_{ij}]$ of a graph G or network is the $n \times n$ matrix with m_{ij} defined as follows:

$$m_{ij} = \begin{cases} 1, & \text{if } (x_i, x_j) \in A \\ 0, & \text{otherwise} \end{cases}$$

The number of terminating cross-connections c is the number of columns in the adjacency matrix M whose column sums are ≥ 2, excluding the sink node column. Conversely, $\aleph_c$ is the maximum of the c column sums. Similarly, the number of starting cross-connections c' is the number of rows in the adjacency matrix M whose row sums are ≥ 2, excluding the source node and $\aleph'_c$ is the maximum of the c' row sums.

As a final remark about the complexity of PANs, Hagstrom [91] obtained the following results about PANs with *discrete, finite range* activity durations:

1. Computing a value of the distribution function of network completion time is P-complete;

2. Computing the mean of the distribution function of network completion time is at least as hard; and

3. Neither of the problems in (1) and (2) can be computed in polynomial in the number of points in the range of the project duration unless P = NP.

The complexity of the problem relies mostly in the enumeration of the paths of the network and manipulation of path durations, which are dependent random variables.

As can be seen by the above discussion, very little research has been devoted to obtaining complexity measures for activity networks with continuous random activity durations (i.e., stochastic activity networks). This observation extends to other probabilistic networks such as stochastic Petri nets and belief networks. I believe the lack of research is a direct result of the necessity to use approximation

techniques for solving probabilistic networks. For example consider the convolution and conditional integration operators mentioned above; both these operators require an integral operation[6]. For many continuous distributions such as the Gamma (with non-integer shape parameter), Normal, Lognormal, and Beta, it is not analytically possible to compute the integral operator exactly (i.e., these distribution functions do not have a closed-form). This inability to compute exact analytical solutions for many probabilistic networks has led to the use of approximation methods. Simulation-based methods have been the dominate approximation technique due to their logical simplicity and efficient implementation. When a simulation is used to solve a probabilistic network, the main complexity measure is the number of simulation runs and corresponding random variable draws necessary to establish a certain statistical confidence in the output.

Soo and Jung-Mo [180] studied the complexity of Petri nets. The authors obtained fifteen structural and dynamic complexity measures and tested these measures on 75 randomly selected "practical" Petri nets. Soo and Jung-Mo observed that adopting a maximum firing rule for the firing of transition in a Petri net reduces the number of nodes (places and transition) in Petri net sample by four on average. The maximum firing rule is defined as follows:

- Let M_k be the kth marking,

- T_k be the set of transitions that are fireable from M_k, and

- $I(t)$ be the set of input places for $t \in T_k$.

If $t_1, \ldots, t_n \in T_k$ and $I(t_1) \cap \ldots \cap I(t_k) = \emptyset$, then $t_1, \ldots, t_n$ are fired in parallel. Soo and Jung-Mo further observed that the number of markings (states) of Petri nets grows exponentially with respect to the increasing size (node and arcs) of the nets.

The exponential behavior of markings of Petri nets extends to stochastic Petri nets. As with Petri nets, research into the complexity of stochastic Petri nets is fo-

---

[6]In the case of the conditional integration operator, multiple integral operations may be required.

cused on reducing the size of the underlying net. For example, German [87] presented a modeling paradigm whose aim is to manage the complexity of net specification. In the proposed framework termed SPNL (Stochastic Petri Net Language), a stochastic Petri net is decomposed into submodels (termed processes) which interact via ports, reward, and results. Ports are arcs that cross process boundaries. Rewards and results measure the internal state and action of the processes. Additionally, Haddad, et al. [90] introduced firing transitions with phase distribution firing time in (bounded) generalized stochastic Petri nets. The authors stated that the transitions produce increases in both the space and time computational complexity of the tangible marking steady-state probabilities. Haddad, et al. solved this problem by a structural decomposition of the underlying Petri net. This decomposition led to a reduction in the computational resources needed to determine the steady-state probabilities. As a final example, Herman, et al. [98] developed a procedure to construct large generalized stochastic Petri nets from smaller components using hierarchical composition. Additionally, the authors provided a method to obtain performance indices of large generalized stochastic Petri nets by stepwise compositional reduction.

Since belief networks are Bayesian in structure, one can update the information in the network using instantiations of some or all the random variables in the network. These instantiations are referred to as evidence. Without evidence one is solving for the joint *prior* distribution of the random variables in the belief network. With evidence one is solving for the joint *posterior* distribution of the belief network. For belief networks with discrete random variables, Kim and Pearl [119] and Pearl [150] showed that finding the posterior distribution[7] or the maximum a-posteriori probability (MAP)[8] of all the variables in the network given the evidence have polynomial-time algorithms in the special case of singly connected networks. A singly connected belief network is a network in which no more than one path exists

---

[7]Also called probabilistic inference.
[8]Also called the most probable explanation (MPE).

between any two nodes. A PERT belief network [114] is an example of a singly connected network. MAP is the distribution that maximizes $P(A \mid E)$ where A is the instantiation (or value assignment) that assign values to all of random variables in the network and E is the evidence. For multiple connected belief networks with binary (discrete) random variables, Cooper [39] and Shimony [174] showed respectively that finding the posterior distribution and MAP of all the variables in the network given the evidence is NP-hard.

For belief networks with continuous random variables, simulation-based approximations for finding the prior and/or posteriori (given evidence) distributions of all the random variables in the network include straight simulation [149] [148], forward simulation [97] [82] [172], and randomized approximation schemes such as Gibbs sampling [149] [15] [31] [105] [199] [89] . Forward simulation is useful in solving belief networks without evidence (i.e., prior distributions) while randomized approximation schemes such as Gibbs sampling are useful in solving belief networks with evidence (i.e., posterior distributions).

2.4   Solution Methods for Probabilistic Networks

Solution methods for output measure(s) in probabilistic networks may be categorized as *exact* or *approximate*. For reasons that were alluded to in sections 2.2 and 2.3 and that will be expanded in greater detail in this section, approximation methods (specifically simulation) dominate the solution methods currently in use. In the review that follows, I decided not to segregate the solution methods by network example opting instead to show the similarities between the solution methods of the various network examples. However when describing a solution method, for historical perspective, I will note the network example for which the solution method was developed. Additionally, based on the research application (air-to-air engagement modeling), the majority of the review is focused on solution methods for probabilistic networks with continuous random variables.

Exact solutions of output measure(s) for many typical probabilistic networks with continuous random variables cannot be achieved. As mentioned in section 2.3, it is not analytically possible to compute a network reduction operation exactly for some distributions (for example, the series reduction operation (convolution) for two Beta distributions). However, there exist classes of functions for which reduction operators yield exact results (for which the classes are closed under the operators). Trogemann and Gente [186] stated that the following items should be taken into account when selecting a class of function to represent random activity durations if the total completion time of a task precedence graph is to be computed in exact form:

1. The class of functions must be closed with respect to multiplication, differentiation, integration, and convolution;

2. The class of functions must be capable of representing empirical distributions to any required degree of accuracy; and

3. The execution time of reduction operations must not require excessive computer time or memory.

These items have applicability to the reduction of non series-parallel graphs as well.

Developed for reducing directed, acyclic networks, Martin [134] used bounded domain, *density* functions represented by the class of piecewise polynomials. A piecewise polynomial is a function, $g(t)$ of the form:

$$
g(t) = \begin{cases}
p_0(t) & \text{if } r_0 \leq t \leq r_1 \\
p_1(t) & \text{if } r_1 \leq t \leq r_2 \\
\vdots & \vdots \\
p_n(t) & \text{if } r_n \leq t \leq r_{n+1}
\end{cases}
\tag{12}
$$

where p_i, $i = 0, 1, \ldots, n$ are real polynomials. In order for $g(t)$ in equation 12 to represent a density function:

1. $g(t) \geq 0$ and

2. $\int_{r_0}^{r_{n+1}} g(x)dx = 1$.

Martin showed that the class of piecewise polynomials is closed under series (convolution) and parallel (maximum) reduction operations.

Developed for series-parallel directed graphs, Cubaud [46] [45] used bounded distributions represented by the class of power functions (PFs). A PF is a function, $PF(t)$ of the form:

$$PF(t) = (\frac{t}{b})^a \quad \text{for } t \in [0, b] \tag{13}$$

where $b \in R^+$, $a = \frac{m}{(b-m)}$, and $m < b$. As equation 13 shows, the class of power functions is a subset of the class of polynomials. Cubaud showed that computing a moment of the maximum of independent, power distributed random variables can be performed in time proportional to the number of variables.

Developed for reducing task precedence graphs, Sahner [164] and Sahner and Trivedi [165] showed that distributions represented by the class of exponential polynomials (EPs) were closed under the operations of multiplication, convolution, differentiation, and integration. An EP is a function, $F(t)$ of the form:

$$F(t) = \sum_{i=1}^{n} a_i t^{k_i} e^{\lambda_i t} \tag{14}$$

where $k_i \in N$, $a_i \in C$, and $\lambda_i \in C$. Not every EP defined in 14 is a valid distribution. For an EP to be a distribution of a nonnegative random variable, it must satisfy the following properties:

1. $F(t)$ is real-valued;

2. $0 \leq F(t) \leq 1 \ \forall t$; and

3. $F(t)$ is monotone nondecreasing and right continuous.

These properties imply that if $\lambda_i \in R$, then $\lambda_i \leq 0$, and if $\lambda_i = 0$, then $k_i = 0$. Additionally, in order for $F(t)$ to be real-valued, complex numbers must occur in conjugate pairs.

The class of EP distributions includes all of the Coxian phase-type distributions [42] [143]. Coxian distributions are defined to be those distributions whose Laplace transforms $L(s)$ are fractions whose numerators and denominators are polynomial in s, such that the degree of the numerator is less than the degree of the denominator. Sahner [164] showed that the class of EP is exactly equivalent to the class of Coxian distributions. A Phase-type distribution has a rational Laplace transform and is thus a proper subset of the Coxian distributions. The class of EP includes exponential, hyperexponential, Erlang, and mixtures of Erlang distributions. EP distributions may also be augmented with a mass at zero and/or infinity to represent, respectively, the probability that an event takes no time or never finishes. As a final note, Sahner gives analytic expressions for *other parallel reduction (multiplicative) operations such as minimum, k out of n, and probabilistic (one path is chosen)*. The minimum operator is equivalent to INCLUSIVE-OR receiver node operation in generalized activity networks.

Trogemann and Gente [186] observed that the class of EPs has the disadvantage that many parameters are required to represent distributions with small variation and large mean value (i.e., coefficients of variation much less than 1). For example, an Erlang distribution with mean 10 and variance 1 contains 100 stages and has distribution function $1 - \exp^{-0.1t} \sum_{i=0}^{99} \frac{0.1^i t^i}{i!}$. A 99 term EP would be necessary to represent this Erlang distribution. In place of EPs, Trogemann and Gente recommended the class of θ-exponential polynomials ($\theta - EPs$) for task precedence graphs. θ-EPs are derived from EPs by translation. More precisely, θ-EPs ($H(t)$)

are functions of the following form:

$$H(t) = \sum_{i=1}^{n} H_i(t) \quad \text{with} \quad H_i(t) = \begin{cases} a_i(t - \theta_i)^{k_i} e^{\lambda_i(t-\theta_i)} & \text{for } t \geq \theta_i \\ 0 & \text{for } t < \theta_i \end{cases} \quad (15)$$

where $k_i \in N_0$, $a_i \in R$, and $\lambda_i \in R_0^+$. The parameter θ_i is called the deterministic part, a_i the coefficient, θ_i the rate, and k_i the stage. For the example above, a θ-EP can model an Erlang distribution (mean 10 and variance 1) with only four terms (if $\theta = 8$).

Figure 11 shows the phase diagram of θ-EPs. The parameters c_i where:

$$c_i = \frac{a_i k_i!}{\lambda_i^{k_i+1}}$$

can be interpreted as branching probabilities if $c_i \in [0, 1]$. Note every branch consists of a deterministic part followed by an Erlang distribution. The deterministic part can be used to guarantee minimum execution times. Trogemann and Gente showed that the class of θ-EPs is closed under the operations of multiplication, convolution, integration and differentiation.

In order to model event distributions with bounded domains, Trogemann and Gente truncate the right hand end of θ-EP class to form a new class of truncated θ-EPs. More precisely, truncated θ-EPs (H_w) are functions of the following form:

$$H(t) = \sum_{i=1}^{n} H_i(t) \quad \text{with} \quad H_i(t) = \begin{cases} 0 & \text{for } t < \theta_i \\ a_i(t - \theta_i)^{k_i} e^{\lambda_i(t-\theta_i)} & \text{for } \theta_i \leq t \leq w \\ \frac{1}{n\xi} & \text{for } t > w \end{cases} \quad (16)$$

where $\xi = \frac{1}{\sum_{i=1}^{n} H_i(w)}$, $k_i \in N_0$, $a_i \in R$, and $\lambda_i \in R_0^+$ and $\forall i : w > \theta_i$. Both θ-EPs and truncated θ-EPs can approximate empirical distributions to any required degree of accuracy.
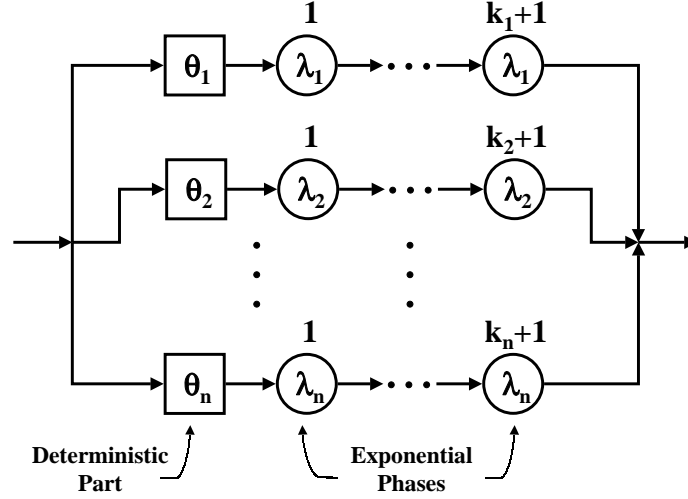
Figure 11. Representation of θ-Exponential Polynomials by Phase Diagram

Kulkarni and Adlakha [121] and Sahner [165] stated that if event distributions in a probabilistic network are exponential, then the network can be transformed into a continuous-time Markov chain. The main drawback of this transformation is that the number of states in the Markov chain increase exponentially with the number of nodes in the graph. Additionally, the practical difficulties of generating the states and transitions of the Markov chain and solving a possibly large system of ordinary differential equations must be contended with. Additionally, Sahner [164] gave an algorithm for transforming any acyclic Markov chain into a series-parallel directed, acyclic graph (dag). The author noted it is more efficient to analyze a chain directly than to translate the chain into a series-parallel dag and reduce the graph.

In stochastic Petri nets (SPNs), one is generally interested in solving for the stationary and/or transient probability of being in a specific marking or state of the reachability set. Exact solution methods vary depending on the subclass of SPNs being analyzed. In other words, the structure and features of a SPN generally depend on the stochastic process (probabilistic model) used to analyze the reachability set. Various subclasses of SPNs are obtained by imposing restrictions on the firing distri-

butions or on the effect of a transition firings on other enable transactions. Ciardo, et al. [32] proposed a hierarchy of SPN classes where modeling power is reduced in exchange for an increasingly efficient solution. Included in their hierarchy are:

- Markov SPNs;

- Semi-Markov SPNs;

- Semi-Regenerative SPNs; and

- Generalized Semi-Markov SPNs.

The descriptions identify the underlying stochastic process used to solve for the stationary (and/or transient probabilities) of a specific marking of a SPN. For example, Markov SPNs are SPNs in which the probabilities of being in a specific marking can be solved by a Markov process such as continuous-time Markov chain. For exact solutions, the main obstacle is often the memory to store the reachability set. Ciardo, et al. gave as a rule of thumb 10^5 to 10^6 markings.

Ciardo, et al. [32] stated that the following firing time distributions are important in practical SPN applications: constant, geometric, discrete, exponential, uniform, polynomial, and expolynomial. The polynomial distribution mentioned by the authors is the same piecewise function defined by Martin [134] in equation 12. The expolynomial distribution is piecewise defined in θ by expressions of the form:

$$\sum_{i=0}^{n} \sum_{j=0}^{m} a_{ij} \theta^i \exp(-\lambda_{ij} \theta) \tag{17}$$

where $a_{ij} \in R$ and $\lambda_{ij} \in [0, +\infty)$. Polynomial and exponential distributions are special cases of the expolynomial distribution shown in expression 17. Haddad, et al. [90] introduced transitions with Coxian firing time distributions in SPN and gave a decomposition method which allowed the derivation of a tensor expression of the generator of the underlying Markov process. This tensor expression reduced the

increased complexity introduced by the Coxian transitions, keeping it of the same order of magnitude as the complexity of the reachability set.

Pritsker's and Happ's [157] and Pritsker's and Whitehouse's [159] introduction of the general evaluation and review technique (GERT) to solve all EXCLUSIVE-OR node generalized activity networks with random activity durations motivated an effort in the late 1960s and early 1970s to find exact solution methods for output measure(s) of stochastic activity networks. Such networks termed (original)[9] GERT networks are signal flow graph (SFG) representations of semi-Markov processes (SMPs)[10] with one or more absorbing states. In the SMP of GERT networks, a state represents an event; the transition probability p_{ij} represents the probability that activity ij will be realized; and the holding time represents the duration of the activity.

In a GERT network, an activity (arc) is represented by the pair of parameters (p, T) where p is the probability of the activity being realized and T is a transform of the time (duration) distribution of the activity. Typical transforms used in GERT networks include the moment generating function, z- or Laplace transforms, and the characteristic function. A transform allows an additive operation such as time to be turned into a multiplicative operation which then can be solved by the three basic types of element combinations in flowgraph theory. The three basic types of element combinations in the theory of flowgraphs are:

- Elements in series;

- Elements in parallel; and

- Loops.

---

[9] The acronym GERT has been expanded to cover several simulation models in which the nodes are not of the EXCLUSIVE-OR type.

[10] A SMP is a stochastic process that makes its transition from any state i to a state j (which may be i) according to the transition-probability matrix of a Markov process, but whose time between transitions is a random variable that may depend on both i and j.

Before Series Combination

After Series Combination
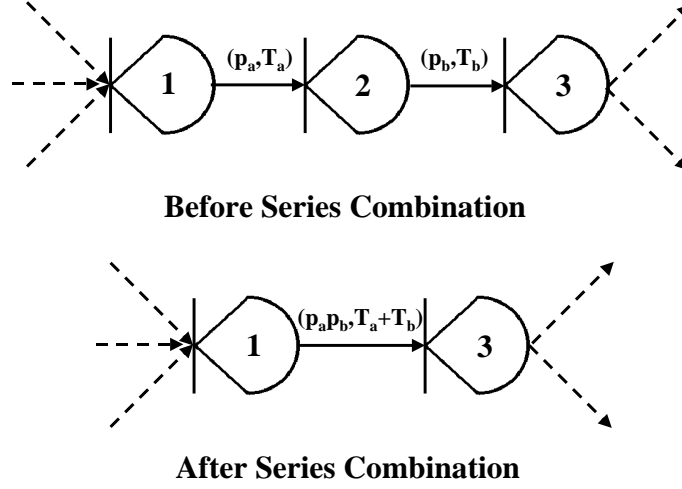
Figure 12. Basic Series Combination in a GERT Network

For example, suppose the time t of an activity duration is specified by its moment generating function (MGF), $M_t(s)$, where:

$$M_t(s) = \begin{cases} \int_t \exp^{st} f(t) dt & \text{if } t \text{ is a continuous random variable} \\ \sum_t \exp^{st} f(t) & \text{if } t \text{ is a discrete random variable} \end{cases}$$

and $f(t)$ is the activity duration density function. For the series system shown in Figure 12, $p_{13} = p_a p_b$ and $t_{13} = t_a + t_b$. However, the MGF of t_{13} is the product of the MGF of t_a and the MGF of t_b. More precisely, $M_{t_{13}} = M_{t_a} M_{t_b}$.

The primary outputs of the original GERT procedure were (a) the probability of realizing a node (and its associated event) and (b) the transform (e.g. MGF) of the time to realize a node, given that the node is realized [157]. In order to obtain the distribution of time to realize a node, one has to invert the transform back into a distribution. This problem is referred to in the literature as the inverse transform problem and is not easily solved for complex GERT networks. Whitehouse [196] proposed four methods to solve this problem:

1. A inversion table look-up operation;

2. An inversion formula;

3. Use the first n moments from the MGF transform to approximate the distribution function; and

4. Assume the form of the distribution and estimate the parameters of the distribution using the moments of the MGF.

Each method has disadvantages. Whitehouse stated that the first and second methods do not appear practical due to the complexity of the transform derived in typically sized GERT networks. The third and four methods are approximations and thus due not lead to exact solution.

Additionally, in theory, AND and INCLUSIVE-OR nodes can be transformed into EXCLUSIVE-OR nodes at the expense of enlarging the number of nodes and arcs in the network. Whitehouse [196] and Elmaghraby [65] provide examples of such expansions. Quichaud and Chretienne [161] extended GERT analysis to networks containing both EXCLUSIVE-OR and AND nodes by searching for analyzable EXCLUSIVE-OR and AND subnetworks and reducing the subnetworks. However, by the late 1960s, simulation (approximation) was seen as a more fruitful solution method for GERT type networks than transform methods. By 1968, Pritsker developed a GERT simulation method to study stochastic problems in inventory, reliability, queuing, and maintenance problems. Other exact solution attempts besides GERT to solve activity networks with random activity durations proved either infeasible or met with limited success. Lawrence [128] and Adlakha and Kulkarni [1] provide an overview of these various attempts. By the late 1970s, exact solution methods for activity networks with random activity durations were abandoned.

When an event distribution of a probabilistic network can not be exactly represented by the functions described above, then the distribution can be approximated by the same functions. For example, piecewise polynomial functions could be fit-

ted to the different event distributions. Unfortunately an exact solution for specific output measure(s) in the "true" sense can no longer be achieved. Many approximation procedures have been developed to solve output measure(s) for probabilistic networks. These procedures generally fall under two categories:

- Simulation and

- Numerical Approximation and Reduction Methods.

Simulation (discrete-event) is an effective approach for approximating the output measure(s) in probabilistic networks, especially if the network has random variables with mixed distributions. At the beginning of simulation experiment, an analyst determines the number of replications (simulations) to be made. Each replication produces a 'sample' value of the output measure(s). These sample values build a 'picture' of the distribution of the output measure(s). The picture is often presented as a frequency histogram of the sample values or set of statistics such as the mean and variance. The accuracy of the picture is a function of the number of replications conducted. Effective employment of simulation requires the user to exercise the simulation through sufficient replications to insure that conditions (e.g., steady state) have been attained with respect to the characterization of the output measure(s).

Two general, accepted types of simulations are found in the literature with regard to output analysis:

1. Terminating - a simulation for which the occurrence of a natural event E specifies the length of each simulation run and

2. Non-Terminating - a simulation for which there is no natural event E to specific the length of the simulation run.

In non-termination simulations one must address steady-state and transient conditions of the output measure(s). Steady-state conditions with respect to output measure(s) are defined as the limits of the measure(s) as the length of the simulation

goes to infinity. In practice, a 'large', finite simulation run may provide 'good' estimates. Transient conditions with respect to output measure(s) refers to fluctuations in the distribution of the output measures. Law and Kelton [126] provide a detailed comparison of the output analysis of terminating and non-terminating simulations. Simulation of the output measures of probabilistic networks may fall into either simulation type. Generally, probabilistic networks without cycles will be terminating (e.g., probabilistic activity networks (PANs)) while networks with cycles will be non-terminating (e.g, ergodic stochastic Petri nets (SPNs)).

For PANs, Van Slyke [191] was the first to propose simulation as a solution method for the network completion time. Burt and Garman [24] and Garman [85] then proposed a conditional simulation approach based on common arcs. Signal et al. [175] developed a conditional simulation procedure similar to that of Burt and Garman. The approach conditions on the arcs in a uniformly directed cut set and then evaluates the integral formulas numerically using simulation. Both Burt and Garman and Sigal, et al. solutions are time consuming and are not practical on large PANs [128]. Following the introduction of the general evaluation and review technique (GERT), Pritsker developed a family of GERT simulation (GERTS) tools for solving generalized activity networks (GAN) and other general networking and stochastic problems [196] [156]. Today, simulation of PANs and GANs can be performed on a PC-based commercial simulation package such as VISUAL SLAM [158]. Variance reduction techniques such as antithetic variates and control variates can be used to reduce the number of runs necessary to obtain output of a required confidence level. Elmaghraby [65] discusses implementation of such reduction techniques in a PAN context.

For stochastic Petri Nets, Ciardo, et al. [32] recommended simulation in cases where a numerical solution is impossible or impractical such as the analysis of Markov SPNs with excessively large reachability sets, transient analysis of semi-Markov SPNs or semi-regenerative SPNs and analysis of a general SPN or a semi-Markov SPN

with non-Markovian subordinated processes. As mentioned in section 2.3, for belief networks with continuous random variables, simulation-based approximations for finding the prior and/or posteriori (given evidence) distributions of all the random variables in the network include straight simulation [149] [148], forward simulation [97] [82] [172], and randomized approximation schemes such as Gibbs sampling [149] [15] [31] [105] [199] [89] . Forward simulation is useful in solving belief networks without evidence (i.e., prior distributions) while randomized approximation schemes such as Gibbs sampling is useful in solving belief networks with evidence (i.e., posterior distributions).

For large probabilistic networks, simulation can become a computationally costly undertaking. As the number of random variables in the network increases, so does the number of replications necessary in order to properly characterize the output measure(s). The objective of an efficient simulation experiment is to accumulate enough sample information on the output measure(s) so that the distribution of the output measure(s) is invariant with increases in the number of replications. Infrequent (rare) events can have a large impact on the number of replications needed to obtain the distribution of output measures. As an example, consider the GAN mentioned by Lawrence [128]. This GAN has a node with two arcs (activities) emanating from it: one that has a high (mean) activity duration when the activity is taken, but a low probability of being taken. The other activity is just the opposite; this activity has a low (mean) activity duration, but a high probability of being taken. The expected number of replications before the first activity is taken will be quite large, and during most simulation experiments the second activity will be repeatedly taken before the first activity is taken for the first time. A output measure, such as mean network completion time, will be initially estimated too low by the corresponding sample statistic, then too high when the statistic spikes the first time the first activity is taken. A large number of replications will be required before the distribution of the mean network completion time is invariant to further

simulation runs. If the simulation experiment is terminated before this invariance is achieved, then an inaccurate statistic will be reported for the mean completion time. Although the above example is pathological, the problem of determining the number of simulation runs to make for a large probabilistic network remains a hard task. If a simulation experiment is stopped to soon for cost and/or time constraints, inaccurate estimates of output measure(s) will result.

As an alternative to simulation, numerical approximation and reduction methods have been proposed for some probabilistic networks. Network reduction involves the repeated application of operators such as multiplication (for example, minimum, maximum, k out of n), convolution, and conditional integration to reduce a network to smaller network through aggregation of the underlying probabilistic model of the network. For probabilistic activity networks (PANs), several numerical approximation and reduction methods have been demonstrated. Reduction processes based on arc (node)[11] duplication has been developed by Martin [134] and Dodin [55] to approximate network completion time. This reduction process transforms an arbitrary PAN (G) into a series-parallel network (G'). The duplications of arcs (nodes) is equivalent to the removal of dependencies in G. Dodin [54] showed network completion time T of both networks are related by:

$$T_G \leq T_{G'} \leq \bar{T}_G$$

where $\bar{T}_G$ is the bound obtained by assuming the path durations are independent. Dodin [51] [53] also developed an approximation of the network complete time of PANs based on the method of sequential approximation. Sequential approximation constructs distributions through the nodes of the network in sequential order.

Dodin implemented the sequential approximation [53] and arc (node) duplication [54] reduction processes using discretization of continuous activity durations.

---

[11]In AoA networks arcs are duplicated. In AoN networks nodes are duplicated.

Later, Lawrence [128] implemented the arc (node) and sequential approximation reduction processes using polygonal approximation of continuous activity durations[12]. Lawrence's implementation method expanded on the work of Martin [134] and Fergeson and Shortell [73][13].

As noted earlier, Martin used piecewise polynomial functions (reference equation 12) to approximate the density functions of random activity durations in series-parallel PANs. Fergeson and Shortell [73] observed that Martin's approach created a problem of 'expanding coefficients' or (as Lawrence [128] [127] terms the problem) of 'exploding coefficients'. To illustrate this problem, consider two arcs in series. Suppose each arc has a random activity duration whose density can be specified by piecewise polynomial functions. If the two arcs are reduced to a single arc through a series reduction, then the random variables representing the activity durations of these arcs are summed (i.e., convoluted). If the density functions of these arcs are represented by polynomials with terms as high as degree m and n respectively, then the density function of the resulting arc can have terms as high as degree $m + n + 1$. A similar result holds for parallel reduction (maximum operator). Thus the series-parallel reduction operations of convolution and maximum increase the number of degree terms in the polynomial functions required to represent the density function of the new arc. The increasing number of terms impacts the amount of computer storage required, because each new (i.e. higher) degree term requires an additional dimension of storage in any array holding the coefficients of the polynomial expression[14]. For larger networks with many series-parallel reduction operations, the computer storage requirements for coefficients quickly explode.

Sahner [171] also noted the 'exploding cofficients' problem in reduction operations when dealing with exponential polynomials (EPs). For example, if a task

---

[12]Lawrence also implemented Dodin's K most critical paths approximation [52] as part of his research.

[13]Lawrence was Fergeson and Shortell's thesis advisor.

[14]A polynomial is completely characterized by its coefficients and hence can be stored in array form.

precedence graph has n nodes with EP distributions (each with a different parameter) and the nodes are to be combined with a maximum operator, then the resulting exact EP will have 2^n terms. Sahner stated that the only way to avoid such an explosion is to use some symbolic approximation. Sahner recommended as further research a study of techniques to reduce the propagation of roundoff and truncation error, due to the finite accuracy of the representation of numbers in a digital computer. Trogemann and Gente [186] also noted that the accuracy of analysis is impacted by rounding errors. They stated that these errors can lead to completely false results for task precedence graphs that contain a large number of nodes making it practically impossible to determine the exact distribution of the completion time for large graphs. As a result, Trogemann and Gente derived approximation formulas using extreme value theory for large task precedence graphs. The authors prove that the limiting distributions from the class of θ-EPs and truncated θ-EPs for parallel (maximum) reduction are Gumbel (extreme value type I) and Weibull (extreme value type III) distributions respectively.

In addition to the problem of 'exploding coefficients', Fergeson and Shortell [73] observed that Martin's approach created a problem of 'proliferating classes'. To illustrate this problem, consider two arcs in parallel. If the density functions of these arcs are represented by c_i polynomials of bounded domain in R, then the partition covered by the polynomials consists of $c_i + 1$ points or c_i classes (cells). If the two arcs are combined in parallel, then the resulting piecewise polynomial function obtained by the maximum operator may be defined on as many as $c_1 + c_2 + 1$ classes. If the two arcs are instead combined in series, then the resulting piecewise polynomial function obtained by the convolution operator may be defined on as many as $c_1 c_2 + c_1 + c_2$ classes. Each of these new classes must be held in computer storage, creating a class storage explosion problem for larger networks.

Fergeson and Shortell [73] solved the 'exploding coefficient' and 'proliferating class' problems by modeling activity duration densities as polynomials of fixed degree,

piecewise-defined on a fixed number of classes. This is accomplished by approximating the activity duration densities of all activities in the network to polynomial densities of the same fixed degree, piecewise defined on the same fixed number of classes over their domains prior to network reduction. Then, when an intermediate or final product is formed by a series or parallel reduction operation, the density of that product is immediately transformed into polynomials of the determined degree, piecewise-defined on the determined number of fixed classes over its domain.

Fergeson and Shortell [73] recommended using simple first degree polynomials fitted piecewise using least square approximation (i.e, least squares regression) over 10 (equally spaced) classes or subintervals. This equates to ten linear regressions over the bound domain of the density function. Fergeson and Shortell further recommended using $10 + I(classwidth * 3)$ regression fitting points positioned uniformly across each class. Additionally, in order to control error build-up from the polygonal approximation, after the ten sets of regression coefficients have been computed, the coefficients are normalized so that the probability under the approximated density function is one.

Fergeson and Shortell [73] termed their method the Polygonal Approximation and Reduction Technique (PART). Fergeson and Shortell [73] examined PANs with uniform, normal, and exponential distributions and showed that the PART compared favorably with simulation (QGERT program) in terms of solution accuracy and had a faster computer processing time and small computer memory requirements than the simulation solution. Fergeson and Shortell [73] noted that the PART experiences it greatest (approximating) error at the peak of a distribution (e.g., approximating the normal density function or approximating the convolution of two normal or exponential distributions). The authors recommended as further research adopting a multiple linear regression approach instead of the simple linear regression approach or increasing the number of classes upwards from ten. Additionally, at class boundaries, two approximations of a density function exist, corresponding to the approximate

line segments of the class which terminates and the class which begins at that class boundary.

As mentioned earlier, Lawrence [128] [127] adopted PART for use with Dodin's arc (node) and sequential approximation reduction processes. Lawrence chose to use Fergeson and Shortell's results in total and justified this choice with the following reasoning:

- The process of fitting a "new" (reduced) piecewise polynomial function to the (higher term) product derived from the convolution or maximization of two "old" piecewise polynomial functions is "better viewed as a data fitting problem, rather than an interpolation problem". This observation is based on the fact that the "old" polynomial functions constitute an approximation of the underlying density functions and thus differ from the true values of densities at possible points of interpolation, although actual errors are not known. Hence, a "new" polynomial function is being fitted to approximate data rather than being interpolated from a given function. Lawrence chose to use the customary least-square approximation to measure the accuracy of the data fitting.

- When approximating the integral or derivative of a continuous function, the higher the degree of the interpolating polynomials, the more accurate the numerically approximated integral or derivative , but the more computationally intensive the construction and subsequent differentiation or integration of the polynomials. Since the accuracy of the approximated integral or derivative can also be increased by refining the partition (increasing the number of classes by adding additional interior points), Lawrence adopts the strategy of Conte and deBoor [37] of approximating the function with polynomial of 'low' degree and then refining the partition until the approximated derivative or integral has the desired accuracy. Lawrence's 'low' degree choice is a first degree or simple linear regression.

Unfortunately, Lawrence did not provide further analysis on 'refining the partition', leaving this topic for further research. Additionally, the initial task of fitting a piecewise polynomial function to a density function is an interpolation problem. Lawrence incorrectly treats this task as a data fitting problem. However, Lawrence [128] did expand PART to include approximations of the triangular, gamma, and beta distributions. Perhaps more importantly, Lawrence showed that PART using Dodin's reduction algorithms were "orders of magnitude faster than simulation-approximations without significant losses in accuracy when the simulation results are taken as 'true' ".

For series-parallel networks, Cubaud [45] showed that the class of power functions can approximate any arbitrary distribution by matching the first two moments of the distribution with the two parameters of the power function. Sahner [171] only partially addressed the issue of fitting distribution data to the class of EP functions. Specifically, Sahner mentioned the work of Augustin and Buscher [10] in the fitting of Coxian phase-type distributions to a given mean and coefficient of variation, and the work of Bux and Herzog [28] in fitting a phase-type distribution to a given mean, variance, and a number of sample values. Sahner noted that Bux and Herzog's method allowed more freedom in the choice of phases than Augustin and Buscher's method. Trogemann and Gente [186] briefly reference [86] as a guide to using θ-EP and truncated θ-EP functions to approximate arbitrary distributions to any degree of accuracy. Unfortunately, neither Sahner nor Trogemann and Gente present an example of these distribution approximating methods with their EP formulations.

Another numerical approximation and reduction approach was developed by Mehrotra, et al. [138] for approximating the network completion time in PANs. The approach is based on an estimation of the number of common activities across critical paths. More precisely, let P denote the set of K critical paths of the network such that V represents the sum of the 'common' activities across the K paths and U_i represents the sum of the 'non-common' activites in the i-th critical path, $i = 1, \ldots, K$.

The completion time of the network T is $\max_{1 \leq i \leq K}(U_i) + V$. Since all critical paths in the PAN will generally not have the same common activities, Mehrotra, et al. proposed using:

1. The average number of common activities;

2. The minimum number of common activities; or

3. The maximum number of common activities.

Additionally, Mehrotra, et al. showed their approximation to have a significant computational time advantage (using a CPU metric) over simulation without sacrificing accuracy when computing the first two moments of the network completion time. For stated purposes of computational ease, Mehrotra, et al. used activity durations with independent and identically distributed (iid) N(4,1) and iid Exp(4) distributions to obtain these results. The authors noted that their approximation can be used with non-iid distributed activities as long as the common activity durations of the network (i.e., the U_is) are identically distributed.

Somarajan and Lau [179] developed an approach to approximating the network completion time of a PAN network based on the Schmeiser-Deutsch distribution [167]. Specifically, the first four moments of network paths are calculated and a Schmeiser-Deutsch distribution is fitted to each path. The network completion time is then computed as the maximum of n random variables represented by each path length. Somarajan and Lau's approach assumes the independence of paths through the network. The first four moments of a network path are calculated from the first four moments of the activity durations in the path. More precisely, if $y = x_1+, \ldots, +x_n$ where x_i are independent random variables whose first four (central) moments are known and $\mu(x_i) = E(x_i) =$ expected value of x_i and $\mu_m(x_i) = (x_i - \mu(x_i))^m =$ the mth central moment of x_i, then:

$$\mu(y) \;=\; \sum_{i=1}^{n} \mu(x_i) \tag{18}$$

62

$$\mu_2(y) \quad = \quad \sum_{i=1}^{n} \mu_2(x_i) \tag{19}$$

$$\mu_3(y) \quad = \quad \sum_{i=1}^{n} \mu_3(x_i) \tag{20}$$

$$\mu_4(y) \quad = \quad \sum_{i=1}^{n} \mu_4(x_i) + 6 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mu_2(x_i)\mu_2(x_j) \tag{21}$$

Equations 18 thru 21 are exact and valid for all distribution forms [92]. The Schmeiser-Deutsch distribution covers all distribution forms with squared skewness β_1 and squared kurtosis β_2 and is of closed form.

2.5   Summary

This chapter provided an overview of probabilistic networks including their graph theoretic underpinnings, complexity, and solution methods for output measures. Specifically, four types of probabilistic networks were investigated; activity networks, Petri nets, task precedence graphs, and belief networks and their structures were compared and contrasted. A probabilistic network was defined as a directed graph $G = (N, A)$ linked to a underlying probabilistic (stochastic) model in which probabilistic events (activities) are represented by arcs or nodes in G. Each probabilistic event is represented by a random variable or a combination of random variables.

Little research has been devoted to obtaining complexity measures for probabilistic networks with continuous random variables. This observation is a direct result of the inability to compute exact analytical solutions for many probabilistic networks and has led to the use of approximation methods. Simulation-based methods have been the dominate approximation technique due to their logical simplicity and efficient implementation. When a simulation is used to solve a probabilistic network, the main complexity measure is the number of simulation runs and corresponding random variable draws necessary to establish a certain statistical confidence in the output.

However, for some probabilistic networks, simulation can become a computationally costly undertaking to obtain accurate estimates of the output measures. As the number of random variables in the network increases, so does the number of replications necessary in order to properly characterize the output measure(s). As an alternative to simulation, numerical approximation and reduction methods have been proposed for some probabilistic networks. Network reduction involves the repeated application of operators such as multiplication (for example - minimum, maximum, k out of n), convolution, and conditional integration to reduce a network to a smaller network through the aggregation of the underlying stochastic process of the network. Numerical approximation involves the use of functions such as piecewise polynomial functions, expolynomial functions, exponential polynomials functions, etc. to approximate the density or distribution functions of the random variables. For probabilistic networks with several random variables, numerical approximation and reduction techniques can be several orders of magnitude faster than simulation-based methods without significant losses in accuracy.

III. Air-to-Air Engagement Literature

3.1 Introduction

In this chapter, a review of air-to-air engagement literature is undertaken. The review in this section deals exclusively with air-to-air engagement modeling. However, in order to accurately model an air-to-air engagement, a modeler should have an adequate understanding of the important components and variables in an air-to-air engagement. To this end, Appendix A provides a brief overview of the complex environment of air-to-air engagements and factors that determine success in this environment. Readers of this dissertation unfamiliar with the air-to-air engagement environment should read the material in Appendix A.

In general, there is a lack of published, public-domain documentation on combat modeling. What little combat modeling documentation that does exist is usually devoted to ground combat modeling (for example, the articles in [140]). To date, I am only able to find three journal references that specifically discuss air-to-air engagement modeling [14] [72] [103]. I believe the lack of documentation on combat modeling (including air-to-air engagement models) is the result of three related factors. First, there does not exist a journal dedicated to combat modeling[1]. Secondly, there is a perception, either real or imagined, that access to information concerning combat models must be kept restricted. Finally, today's combat models consist of large, complex simulations which are not easily explained in the page restrictions of journal articles. Congruently, these simulations are created by civil contractors who are paid to build and maintain the simulations, not to write about them. The result of this lack of information on combat modeling is that combat modeling insights are constantly having to be relearned. Hence, one of the goals of this research is to pro-

---

[1]Articles on combat modeling can be found in journals such as Naval Research Logistics Quarterly and Military Operations Research. Neither of these journals are dedicated solely to combat modeling.

vide a summary of the lessons learned and status of air-to-air engagement modeling at the theater-level.

In section 1.3, it was stated that air-to-air models may be divided into two main groups according to their intended purpose: analysis or training. This dissertation's focus is on air-to-air engagement models used for analysis. Air-to-air engagement models for analysis may be further divided into two subgroups depending on the modeling method employed: analytic or simulated.

3.2 Analytic Models

Analytic models, particularly simple ones, help clarify important relations that are difficult to perceive in a more complex model. Though generally more abstract than simulation models, analytic models are characterized by *transparency* or the degree to which cause and effect relationships in a model are apparent. An analytic model is usually too simple and restricted to directly solve an operational problem, but because of its transparency, its insights provide valuable guidance for simulation model investigations [184].

Analytic combat models may be traced back to the seminal works of Fisk [75], Osipov [145], Lanchester [124][2]. These authors were the first to express the dynamics of combat in mathematical terms. Specifically, they considered combat as a set of coupled differential equations relating the changes over time in the force level of friendly combatants to the force level of the opponent's and/or friendly combatants and vice versa. For example, Lanchester's square law of combat hypothesized that under "modern conditions" that the change over time in the force level of a side's combatants would be proportional to the force level of the opponent's combatants.

---

[2]Helmbold and Relm [96] provide the English translation of the five part series of articles that Osipov published in the Russian journal Voenniy Sbornik and an overview and discussion of the interrelationships between of the works of Fisk, Osipov, and Lanchester.

In mathematical terms, we have

$$\frac{dx}{dt} = -\alpha y \qquad \text{with } x(0) = x_0 \tag{22}$$

$$\frac{dy}{dt} = -\beta x \qquad \text{with } y(0) = y_0 \tag{23}$$

where

- $x(t)$ - the number of x combatants at time t;

- $y(t)$ - the number of y combatants at time t;

- α - the rate at which individual combatants in force y kill opposing combatants in force x; and

- β - the rate at which individual combatants in force x kill opposing combatants in force y.

The state solution for equations 22 and 23 is

$$\beta(x_0^2 - x^2) = \alpha(y_0^2 - y^2) \tag{24}$$

showing the square law relationship.

Lanchester's research started the exploration of using differential equation formulations (both deterministic and stochastic) to model combat attrition (referred to as Lanchester-Type Models of Warfare). Taylor [183] [184] and Fowler [79] provide comprehensive reviews of Lanchester-Type Models of Warfare. Several deterministic Lanchester-type, air-to-air engagement models have been proposed by Frick [81] and Latchaw [125]. The US Air Force's THUNDER theater-level combat model [29] uses a set of deterministic Lanchester equations (in difference form) to allocate intercepting aircraft to target escort aircraft and target non-escort aircraft (e.g. bombers). More precisely, let

- $x(t)$ represent the number of intercepting aircraft in the engagement at time t;

- $y_1(t)$ and $y_2(t)$ represent respectively the number of target escort and non-escort aircraft in engagement at time t;

- a represent the percentage of interceptors allocated to engaging non-escort aircraft ($1 - a$ are allocated to engaging escort aircraft).

The resulting Lanchester equations for attrition in the air-to-air engagement are:

$$\frac{dx}{dt} = -(\alpha y_1 + \beta y_2) \quad \text{with } y_1(0) = y_1^0 \text{ and } y_2(0) = y_2^0 \tag{25}$$

$$\frac{dy_1}{dt} = -(1 - a)\gamma x \quad \text{with } x(0) = x^0 \tag{26}$$

$$\frac{dy_2}{dt} = -a\gamma x \tag{27}$$

where

- α - the rate at which individual escort aircraft in y_1 kill interceptor aircraft in x;

- β - the rate at which individual non-escort aircraft in y_2 kill interceptor aircraft in x;

- γ - the rate at which individual interceptor aircraft in x kill escort and non-escort aircraft respectively in y_1 and y_2.

The state solution for equations 25 thru 27 is:

$$\beta(z_0^2 - z^2) = \theta^2(x_0^2 - x^2) \tag{28}$$

where

- $z(t) = \alpha y_1(t) + \beta y_2(t)$ and

- $\theta = \sqrt{(1 - a)\alpha\gamma + a\beta\gamma}$

Equation 28 is used to determine the winning side (either x or y) through an engagement of total attrition. If side x wins, then $y_1 = 0$ and $y_2 = 0$, since the battle

68

is fought until one side is annihilated ($y_1 = 0$ and $y_2 = 0$ implies $z = 0$). Thus the number of survivors of the engagement on side x are $\sqrt{\theta^2 x_0^2 - \beta z_0^2}$. If side y wins, then $x = 0$. Unfortunately, the number of survivors of the engagement on side y can not be obtained directly from equation 28 since the expression for z involves two unknowns y_1 and y_2 (i.e., one equation and two unknown variables).

However, the survivors on side y can be obtained directly from the force level equations. Specifically, solving equations 25 thru 27 yields force level equations

$$
\begin{aligned}
x(t) &= x_0 \cosh(\theta t) + \frac{z_0}{\theta} \sinh(\theta t) & (29) \\
y_1(t) &= y_1^0 + (1-a)\gamma[\frac{z_0}{\theta^2}[\cosh(\theta t) - 1] - \frac{x_0}{\theta}\sinh(\theta t)] & (30) \\
y_2(t) &= y_1^0 + a\gamma[\frac{z_0}{\theta^2}[\cosh(\theta t) - 1] - \frac{x_0}{\theta}\sinh(\theta t)] & (31)
\end{aligned}
$$

Letting $x(t) = 0$ and solving for t in equation 29 yields

$$
t_y = \frac{1}{2\theta} \ln \frac{1 + \frac{\theta x_0}{z_0}}{1 - \frac{\theta x_0}{z_0}} \tag{32}
$$

The survivors on side y are then determined by equations 30 and 31 using $t = t_y$ from equation 32.

A measure of effectiveness (MOE) determines the allocation a and $1-a$ of interceptors to non-escort aircraft and escort aircraft. Specifically, the winning side and survivors on that side are calculated for $a = 0, .05, \ldots, 1$ (i.e, step i, $i = 1, 2, \ldots, 21$). The percentage of non-escort aircraft lost (A_I) and of interceptors lost (B_I) in allocation step i is determined. The following measure of effectiveness is calculated for each allocation step:

$$
MOE_i = \frac{100 * A_i}{1 + (100 * V * [B_i - C_0])^2} \tag{33}
$$

where

- V is the measure of the value of interceptor losses; and

- C_0 is the percentage of interceptor losses when $a = 0$

The largest value of MOE_i in equation 33 determines the allocation a and $1 - a$ of interceptors to non-escort aircraft and escort aircraft.

Due to the general recognition that combat is a stochastic series of events [7] and the limiting assumption of large force numbers, the use of deterministic Lanchester equations as analytic tools have been limited to large force-on-force engagements. In their place, stochastic analytic models have been pursued. These models are able to analyze small or intermediate numbers of combatants, the engagement size appropriate for modern air combat (reference Appendix A).

Whitehouse [196] used an EXLUSIVE-OR generalized activity network (GAN) or GERT network to model a simple one versus one air-to-air engagement. Figure 13 shows the network for the air-to-air engagement. Node 1 represents the event that the interceptor engages the target aircraft. Three possible outcome activities exist:

1. The interceptor approaches the target aircraft, fires a missile, and kills the target aircraft. This outcome activity is represented by arc $(1, 3)$ with probability p_1 and a random occurrence time of $t_{(1,3)}$ with moment generating function (MGF) $M_{(1,3)}(s)$. This arc terminates at node 3 with the event being a shoot down of the target aircraft.

2. The interceptor approaches the target aircraft, fires a missile, but misses the target aircraft. This outcome activity is represented by arc $(1, 2)$ with probability p_2 and a random occurrence time of $t_{(1,2)}$ with MGF $M_{(1,2)}(s)$. This arc terminates at node 2 with the event being a missed interceptor missile shot.

3. The interceptor approaches the target aircraft. The target aircraft fires a missile first and shoots down the interceptor. This outcome activity is represented by arc $(1, 5)$ with probability p_3 and a random occurrence time of $t_{(1,5)}$ with
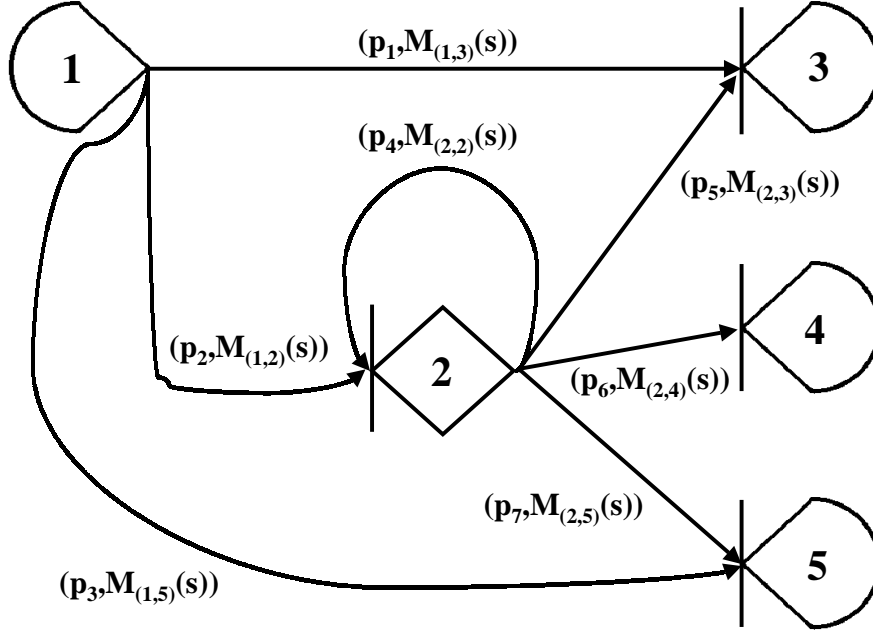
Figure 13. GERT Network of Simple Air Duel with Unlimited Passes

MGF $M_{(1,5)}(s)$. This arc terminates at node 5 with the event being a shoot down of the interceptor.

If the interceptor misses the target aircraft, then four possible outcome activities exist:

1. The interceptor approaches the target aircraft, fires a missile, but misses the target aircraft again. This outcome activity is represented by the loop $(2,2)$ with probability p_4 and a random occurrence time of $t_{(2,2)}$ with MGF $M_{(2,2)}(s)$. This arc terminates at node 2 with the event being a missed interceptor missile shot.

2. The interceptor approaches the target aircraft again, fires a missile, and kills the target aircraft. This outcome activity is represented by arc $(2,3)$ with probability p_5 and a random occurrence time of $t_{(2,3)}$ with MGF $M_{(2,3)}(s)$.

71

This arc terminates at node 3 with the event being a shoot down of the target aircraft.

3. The interceptor terminates the engagement. This outcome activity is represented by arc $(2, 4)$ with probability p_6 and a random occurrence time of $t_{(2,4)}$ with MGF $M_{(2,4)}(s)$. This arc terminates at node 4 with the engagement being terminated.

4. The interceptor approaches the target aircraft again. The target aircraft fires a missile and shoots down the interceptor. This outcome activity is represented by arc $(2, 5)$ with probability p_7 and a random occurrence time of $t_{(2,5)}$ with MGF $M_{(2,5)}(s)$. This arc terminates at node 5 with the event being a shoot down of the interceptor.

The unconditional moment generating function of the time to shoot down the target aircraft is

$$W_3 = p_1 M_{(1,3)}(s) + \frac{p_2 M_{(1,2)}(s) p_5 M_{(2,5)}(s)}{1 - p_4 M_{(2,2)}(s)} \tag{34}$$

Similar expressions can be derived for the unconditional moment generating functions for the time to terminate the engagement and to shoot down the interceptor. Whitehouse [196] stated that more complex air-to-air engagements can be modeled in a similar manner. A significant difficulty of the GERT models as mentioned in section 2.4 occurs when trying to obtain the distribution of time to realize a node. In order to obtain the distribution of time to realize a node, one has to invert a transform such as given in equation 34 back into a distribution. This problem is not easily solved for complex GERT networks.

Feigin, et al. [72] used a discrete space, continuous time Markov chain to analyze a m-on-n air-to-air engagement. Specifically, at time $t = 0$, an air-to-air engagement starts between opposing flight group of m homogeneous[3] blue aircraft

---

[3]In terms of aircraft performance and weapons characteristics.

72

and a flight of n homogeneous red aircraft. At any time t, $t > 0$, an individual aircraft is in one of four roles (or states):

1. <u>Pursuer</u> - involved in a duel with an opponent aircraft and in the process of acquiring a firing position against the opponent aircraft;

2. <u>Evader</u> - involved in a duel with an opponent aircraft and trying to minimize the opponent's firing opportunity;

3. <u>Free</u>- not engaged in a duel with an opponent aircraft; seeking an opponent aircraft to attack;

4. <u>Downed</u> - aircraft has been shot out of the sky by an opponent aircraft.

At $t = 0$, all aircraft are assumed to be in the free role.

A change in an aircraft role is associated with one of the following four events:

1. A pursuer downs an evader with rate λp where λ is the average rate at which the pursuer reaches a firing position and fires and p is the kill probability of a single weapon release.

2. A pursuer is acquired by a free aircraft from the evader's flight group with average rate θ. In this event, the free aircraft becomes a pursuer, the pursuer assumes the role of an evader, and the (original) evader becomes a free aircraft.

3. Two free aircraft from the opposing flight group engage in a duel as the result of one of the aircraft acquiring the other with average rate θ (i.e., one aircraft becomes a pursuer while the other aircraft becomes an evader).

4. An evader disengages a pursuer with average rate α.

The objective of each flight group (blue and red) is to destroy as many aircraft in the opposing flight group as possible and each aircraft has an unlimited number of firing opportunities. For a given m blue and n red aircraft there are $k = k(m, n)$ ways based on the four events above that the $m + n$ aircraft could be assigned the

m	n	k
1	1	5
2	2	19
4	2	39
3	3	49
6	3	109
5	5	195
6	6	335
8	8	824
10	10	1715

Table 1.   Number of the States, k, of the Markovian Model Given m Blue Aircraft and n Red Aircraft

roles of pursuer, evader, free, or downed. Feigin, et al. use k states for the Markov chain. Table 1 shows the possible states for a given value of m and n.

Feigin, et al. [72] considered the loss ratio number:

$$\rho(t) = \frac{N_B(0) - N_B(t)}{N_R(0) - N_R(t)} \tag{35}$$

and the attrition ratio:

$$\tau(t) = \frac{1 - \frac{N_B(t)}{N_B(0)}}{1 - \frac{N_R(t)}{N_R(0)}} \tag{36}$$

as output measures where $N_B(t)$ and $N_R(t)$ refer to the number of aircraft on the blue and red sides respectively at time t. The loss and attrition ratios in equations 35 and 36 are the main metrics used to assess the degree of air superiority at the theater-level.

Feigin, et al. [72] stated that their model "should be regarded as a prototype of a large family of more complete Markov models". The model can easily be expanded to include more detailed aircraft roles and events. For example, two aircraft from the same side (lead and wingman) could be allowed to engage a single opponent aircraft rather than the current single pursuer restriction. Also states could be added to take into account the number of missiles remaining. The payment for a

more detailed model is a larger state space and hence more computational time to solve the resulting Markov chain. Additionally, one may question the validity of the Markovian (memoryless) property in modeling air-to-air engagements (specifically, does the conditional distribution of the future state $S(t+s)$ given the present state $S(s)$ and the past states $S(u), 0 \leq u < s$ depend only on the present state $X(s)$ and not on the past states $S(u)$). Depending on structure (i.e., roles and events) used in the model, this property will not always hold. For example, consider an air-to-air engagement in which a flight of four aircraft loses two of its aircraft. It is unlikely that the flight will function in the same manner (aggressiveness, coordination, etc.) given the loss of the two aircraft as they would without an aircraft loss. Finally, the fact that a homogeneous Markov process is used to model the engagement reflects the assumption that there is no effect of time on the transition probability rates. In other words, as the air-to-air engagement progresses no account is taken of factors such as fatigue or learning on the part of the pilots.

Hong, et al. [103] furnish three non-Markovian air-to-air engagement models. The first model illustrates a one versus one air-to-air engagement using the "interfiring" distributions[4] of missiles (one radar-guided and one infrared) on each aircraft to determine engagement outcomes. The second model is also a one versus one air-to-air engagement but, unlike the first model, it uses the "interkilling" distributions[5] of missiles to determine engagement outcomes. Additionally, the second model increases the number of missiles on each aircraft to four (two radar-guided and two infrared) and divides the engagement into three distinct phases: beyond visual range (BVR) engagement, visual detection, and within visual range (WVR) engagement. Finally, the third model expands the state space of the second model to include two aircraft on one of the sides, making the air-to-air engagement a two versus one

---

[4]The interfiring distribution describes the distribution of time to a missile *launch*.

[5]In contrast to the interfiring distribution, the interkilling distribution represents the distribution of time to a missile *kill*. In other words, with the interkilling distribution, we are interested only in the missile launches that kill (shoot down) an opponent aircraft.

encounter. All three models assume that the engagement continues until one side is killed or all weapons are exhausted.

The three models were developed under the assumption that an air-to-air engagement can be viewed as the superposition of several independent, terminating renewal processes. The technique of superpositioning independent, terminating renewal processes was first proposed by Ancker [5] for modeling one versus one stochastic ground combat, termed by Ancker a "stochastic duel". This technique has been used in follow-on research by Ancker, Gafarian, Kress and several associates [83] [120] [84] [102] [147] [198] to model m versus n stochastic ground combat[6]. In the m versus n stochastic combat model, each combatant completes an activity that leads to a combat event (e.g., firing a weapon or killing a target). The occurrence of a combat event is modeled as a separate renewal process for each combatant with the interarrival times of the renewal process representing individual combat events for the combatant. By assuming that each combatant completes a combat event independently of the other combatants (e.g., assuming each combatant is shooting at a passive, fixed target), the renewal processes are made independent.

In order to calculate the probability of being in a particular state, the technique of supplementary variables [41] is used. More precisely, at time t let the supplementary variable Y_n denote the time since the nth combat event. Then the first order

---

[6]Ancker [5] studied the one-on-one stochastic combat model. The next attempt to solve a stochastic combat model for a larger size was the two-on-one stochastic model by Gafarian and Ancker [83]. They developed the general solution for the state probabilities for homogeneous two-on-one stochastic combat and from these derived the winning probabilities. Kress [120] solved the many-on-one stochastic combat problem. Next, Gafarian and Manion [84] solved the homogeneous two-on-two combat model. Hong [102] followed by developing a solution for the state probabilities of heterogeneous two-on-one and homogeneous three- on-two combats. Later Parkhideh and Gafarian [147] provided the general solution to many-on- many heterogeneous stochastic combat of which many-on-many homogeneous combat is a subset. Unfortunately, Parkhideh and Gafarian's solution is a method of exhaustive enumeration and has a strong exponential computation time which makes it practically impossible to use for battles beyond 4 on 4. As a result, Yang and Gafarian [198] developed a fast approximation for the many-on-many homogeneous case. Unfortunately, as can be seen by the references, Yang's article was published before Parkhideh's. However, Parkhideh [146] completed his research before Yang [197]. Undoubtedly, Parkhideh's research led to Yang's research.
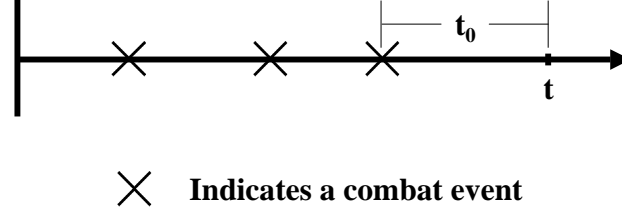
Figure 14. Backward Recurrence Time Example

(instantaneous) probability that a $n + 1$th combat event will occur in the interval $(t, t + dt)$ is given by:

$$\frac{f(y_n)}{F^c(y_n)} \tag{37}$$

where $f(\cdot)$ and $F^c(\cdot)$ are the density and complementary cumulative distribution functions of the combat event random variable (e.g., interfiring time or interkilling time). In renewal theory, Y_n is referred to as the *backward recurrence time*. Figure 14 illustrates the concept of backward recurrence time. In Figure 14 a particular combatant (at time t) has had three combat events and it has been t_0 time units since the last combat event. With the use of the technique of supplementary variables, the non-Markovian combat models are transformed into semi-Markovian models.

A major difficulty of the Hong, et al. [103] model is the (exponential) explosion of the state space with an increase in the number of combat events. This explosion is directly related to the requirement to model possible combinations of events. For example, a one-on-one air-to-air engagement model with interfiring distributions where each aircraft has one radar-guided and one infrared missile (i.e., four combat events) has 21 possible states. In order to reduce the state space size, Hong, et al.

recommend using interkilling distributions rather than interfiring distributions. This reduces the number of combat events, hence reducing the state space of the model.

3.3 Simulation Models

As stated in section 1.3, today's air-to-air engagement models consist mainly of relatively large, complex simulations. To many, these simulated models are considered more credible than analytic models because these models generally contain more detail than analytic models [184]. In simulated models, the processes and activities of combat are "acted out". In this section, THUNDER's air-to-air engagement submodel will be discussed. Although there are other theater-level air-to-air engagement simulation models, THUNDER is the predominate model used by the US Air Force for theater-level analysis [29]. Appendix B details the air-to-air engagement submodel used in THUNDER for those readers who are not familiar with this theater-level model.

As can be seen from Appendices A and B, air environment factors in a theater level air-to-air engagement model such as those contained in THUNDER are treated at a very aggregated level when compared to an engagement level model such as BRAWLER. THUNDER input parameters (such as probability of engagement (ENG), relative range advantage (RRA), probability of kill (PK), degree of aircraft command and control (ACC), etc.) attempt to account for factors that influence the outcome of air-to-air engagements such as tactics, sensor performance, range advantage, weapons usage, etc. in a general manner. In contrast, the air environment in an engagement level model such as BRAWLER is treated in great detail. Sensors and weapons are explicitly modeled and their performance compared to flight tests and laboratory experiments. Tactics are modeled explicitly and are based upon intelligence estimates of threat weapon system capabilities, threat cultural biases, and pilot proficiency. Blue tactics are taken from USAF air manuals

such as Multi-Command Manual (MCM) 3-1, actual tactics development flight tests, and discussions with experienced pilots.

As a result of the aggregated treatment of air environment factors in the THUNDER "high resolution" air-to-air engagement submodel, the submodel has been criticized in certain areas for not accurately reflecting the actual air environment. Major criticisms include:

- The single shot probability of kill, $SSPK$, calculation considers only firing platform / weapon load versus opponent platform combinations and is approximately the same regardless of whether the firing weapon is employed as the primary munition or employed with different tactics, or the firing platform encounters the same opponent platform armed with a different weapons load. [9] [176] For example, consider an F-15C armed with an AIM-120 as its primary weapon and the AIM-9 as its secondary weapon being engaged by a threat aircraft armed with a beyond visual range (BVR) missile less capable than the AIM-120 but with longer range than the AIM-9. In the majority of the engagements, the AIM-120 is likely to dominate all outcomes, negating the threat missile range advantage over the AIM-9. In the situation that the F-15C is not armed with AIM-120s or has expended them during previous engagements, this threat BVR range advantage is not reflected in THUNDER.

- The number of weapons fired per engagement by an aircraft is the same for all air-to-air engagements regardless of type of opponent aircraft faced [9].

- A multiple weapon salvo is always modeled as a SHOOT-SHOOT firing doctrine while the more typical USAF SHOOT-LOOK-SHOOT firing doctrine is not modeled [9] [176].

- When an air-to-air engagement occurs, weapons are always fired. The model does not allow for aircraft to disengage before weapons release [9].

Although the criticisms of the THUNDER "high resolution" air-to-air engagement submodel discussed above do have merit, the model appears to produce acceptable results in terms of bottom-line theater-level air superiority metrics such as aircraft loss and exchange ratios. In section 1.3, it was mentioned that BRAWLER is used to calibrate data used in THUNDER. An example of this calibration process is given in reference [9]. In this calibration process, individual air-to-air engagement outcomes from THUNDER were examined from an input data perspective and compared to the BRAWLER inputs and results. Specifically, four THUNDER input variables were considered ENG, RRA, PK, and L. The result of the calibration process were favorable with BRAWLER and THUNDER air-to-air engagement outcomes matching within intervals attributed to the inheritant differences of the models. With as many input data factors as the THUNDER "high resolution" air-to-air engagement submodel allows one could change other factors and lead to the similar outcomes in terms of theater-level metrics.

3.4   Summary

This chapter provided a review of air-to-air engagement literature. Today's air-to-air engagement models consist mainly of relatively large, complex simulations and may be roughly divided into two main groups according to their intended purpose: analysis or training. This dissertation's focus is on air-to-air engagement models used for analysis. The US Air Force predominate simulation model used for analysis is THUNDER. Appendix B details the air-to-air engagement submodel used in THUNDER.

On the other hand, analytic models (particularly simple ones) help clarify important relations that are difficult to perceive in a more complex model. Though generally more abstract than simulated models, analytic models are characterized by transparency or the degree to which cause and effect relationships in a model

are apparent. Analytic air-to-air engagement models discussed include Lanchester models [124] [29] and probabilistic models [196] [72] [103].

IV. Event Occurrence Networks

4.1 Introduction

In this chapter, a new network formulation termed an *event occurrence network* (EON) is introduced to model the interaction of groups of event sequences. An EON is a graphical representation of the superposition of several terminating counting processes. EONs differ from the probabilistic networks previously discussion in chapter II and are motivated by the research of Ancker, Gafarian, Kress and several associates [83] [120] [84] [102] [147] [198] in modeling m versus n stochastic ground combat. An analysis method termed "bucket" analysis is introduced to accommodate a node (state) explosion problem as the number of the events in an EON increases. The general integral solution to the probability of being at any node (state) in an EON at time t is discussed. To illustrate the capabilities of EONs, two simple air-to-air engagement examples are given.

4.2 Description

An event occurrence network (EON) is a probabilistic network in which an arc represents the occurrence of an event from a group of (sequential) events before the occurrence of events from other event groupings. Events between groups occur independently, but events within a group occur sequentially. A set of arcs leaving a node is a set of competing events, which are probabilistically resolved by order relations. More precisely, consider groups of sequential events:

$$G_{1_1}, G_{1_2}, \ldots, G_{1_{n_1}}$$
$$G_{2_1}, G_{2_2}, \ldots, G_{2_{n_2}}$$
$$\vdots \qquad\qquad \vdots$$
$$G_{s_1}, G_{s_2}, \ldots, G_{s_{n_s}}$$

where $G_g, g = 1, \ldots, s$ represent s independent groups of events and G_{g_e} represents event e in group g. Let continuous random variable E_{g_e} represent the occurrence (completion) time t of event G_{g_e} with

- density function $f_{E_{g_e}}(t)$;

- cumulative distribution function $F_{E_{g_e}}(t)$; and

- complementary cumulative distribution function $F^c_{E_{g_e}}(t)$.

The complementary cumulative distribution function $F^c_{E_{g_e}}(t)$ represents the non-occurrence of event G_{g_e} by time t and is equivalent to $1 - F_{E_{g_e}}(t)$. Additionally, only one event can occur in any infinitesimally small time interval dt; *two or more events cannot occur at the same instant in time.*

An EON is the graphical representation of the superposition of several terminating counting processes. A counting process $\{N(t), t \geq 0\}$ is a stochastic process[1] in which $N(t)$ represents the total number of events that have occurred up to time t [163]. Random variable E_{g_e} (defined previously above) represents the interarrival time of the eth event for counting process $N_g(t), g = 1, \ldots, s$. If E_{g_e} are identically distributed for all e, then counting process $N_g(t)$ is a renewal process. Cox and Smith [43] have discussed the superposition of renewal processes.

As stated in the introduction to this chapter, EONs differ from the probabilistic networks previously discussion in chapter II. For example, consider generalized activity networks (GANs). Two parameters are associated with the arc of a GAN:

- the (discrete) probability p that an arc is taken given the node from which it emanated is realized and

- the time t required to accomplish the activity which the arc represents.

---

[1] A stochastic process $\{X(t), t \in T\}$ is a collection of random variables (where t is an index and $X(t)$ is a random variable).
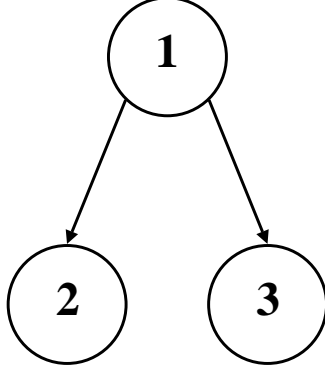
Figure 15. Simple, Two Event Event Occurrence Network

For our purposes, assuming T is a continuous random variable, a GAN arc occurs with probability $P(t)$ and has length given by the distribution of T. On the other hand, an EON arc occurs with (continuous) probability $p(t)$ which is determined by stochastic order relations among the event random variables and has length given by a distribution conditioned on the ordering of the event random variables. For example, consider an EON composed of two event groups with one event in each group (events G_{1_1} and G_{2_1}). Figure 15 graphically depicts the EON for this model which has three states. The probability that arc 12 is chosen is equivalent to the probability that G_{1_1} occurs before G_{2_1} ($P(t) = P(E_{1_1} < E_{2_1})$) with the length given by the conditional distribution $f_{E_{1_1}|E_{1_1}<E_{2_1}}(t)$.

4.3 Air-to-Air Engagement Examples

As an example of an EON, consider the first air-to-air engagement model discussed by Hong, et al. [103]. As discussed in section 3.2, this model described a one versus one air-to-air engagement in which each aircraft could fire one radar-guided and one infrared missile each. A constraint of the model was that the radar-guided

missile must always be fired before the infrared missile. Let G_{g_1} and G_{g_2} represent the events that side g $(g = 1, 2)$ fires a radar-guided and infrared missile respectively. Further let random variables E_{g_1} and E_{g_2} represent the occurrence times to fire the radar-guided and infrared missile respectively for side g^2. Figure 16 shows the EON for this model.

The EON shown in Figure 16 has 9 nodes. Node 1 represents a "non-event" for this model (i.e., the non-occurrence of the first event); neither aircraft has fired a missile. Nodes 2 and 3 represent possible occurrences (outcomes) for the first event; one of the aircraft has fired a radar missile. Node 2 represents the event that the aircraft on side 1 has fired the radar missile (i.e., event G_{1_1}) while Node 3 represents the event that the aircraft on side 2 has fired the radar missile (i.e., event G_{2_1}). Nodes 4, 5, and 6 represent possible occurrences for the second event. Node 4 represents the event that the aircraft on side 1 has fired an infrared missile given that the first event was a radar missile fired by the same aircraft. Node 5 represents the event of both aircraft firing radar missiles. Two separate paths lead to this node, depending on which aircraft fired the first radar missile. Finally, Node 6 represents the event that the aircraft on side 2 has fired an infrared missile given that the first event was a radar missile fired by the same aircraft. In a similar manner, Nodes 7 and 8 represent possible occurrences for the third event and node 9 represents the occurrence of the fourth and last event.

Note the "flow-down" structure of the graph in Figure 16. Each set of nodes on the same horizontal line corresponds to the possible occurrences of the uth event $(u = 1, 2, 3, 4)$. Six paths lead from node 1 to node 9. These paths consists of the following sequences of events:

---

^{2}Recall from section 3.2 that E_{g_1} and E_{g_2} are referred by Hong, et al. as interfiring times.
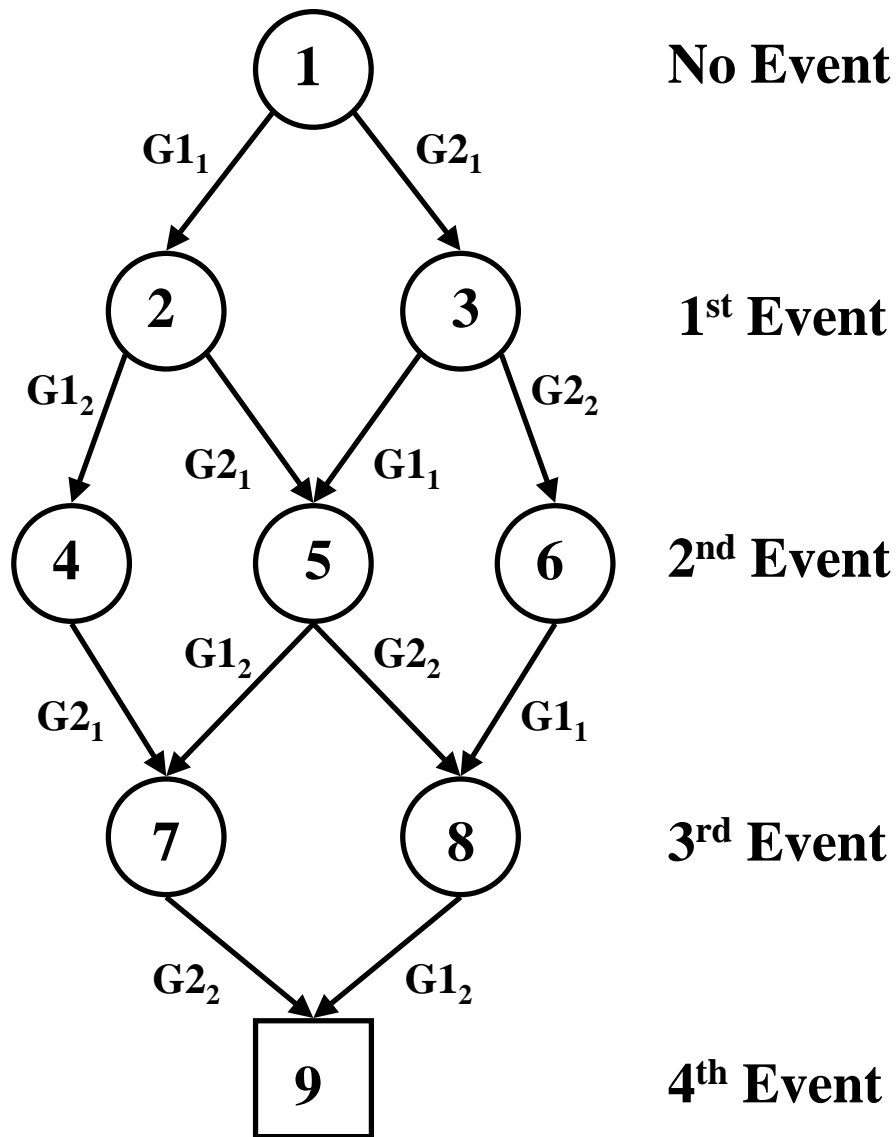
Figure 16.   Event Occurrence Network for One vs One Air-to-Air Engagement Example

Number	Node Path	Event Sequence
1	$1 \to 2 \to 4 \to 7 \to 9$	$(G_{1_1}, G_{1_2}, G_{2_1}, G_{2_2})$
2	$1 \to 2 \to 5 \to 7 \to 9$	$(G_{1_1}, G_{2_1}, G_{1_2}, G_{2_2})$
3	$1 \to 2 \to 5 \to 8 \to 9$	$(G_{1_1}, G_{2_1}, G_{2_2}, G_{1_2})$
4	$1 \to 3 \to 6 \to 8 \to 9$	$(G_{2_1}, G_{2_2}, G_{1_1}, G_{1_2})$
5	$1 \to 3 \to 5 \to 8 \to 9$	$(G_{2_1}, G_{1_1}, G_{2_2}, G_{1_2})$
6	$1 \to 3 \to 5 \to 7 \to 9$	$(G_{2_1}, G_{1_1}, G_{1_2}, G_{2_2}).$

The air-to-air engagement EON shown in Figure 16 terminates with probability one at node 9 as $t \to \infty$ with both missiles being expended by each aircraft. Clearly, this engagement is not a realistic example since the possibility of a missile kill has not been included (this is the actual engagement model discussed by Hong, et al. [103]). Suppose missile kills are allowed in the EON. More precisely, let p_{g_1} and p_{g_2} represent the respective probabilities that the radar and infrared missile from side g shoots down (kills) the opponent aircraft given that the missile is fired. (Additionally, let $q_{g_1} = 1 - p_{g_1}$ and $q_{g_2} = 1 - p_{g_2}$ represent the respective probabilities that the radar and infrared missile from side g misses the opponent aircraft given that the missile is fired). The EON for this one versus one air-to-air engagement is shown in Figure 17.

For the EON shown in Figure 17, let $G_{g_1^{(1)}}$ and $G_{g_2^{(1)}}$ represent the events that side g ($g = 1, 2$) fires a radar-guided and infrared missile respectively that *shoot down* the side $\overline{g}$ and let $G_{g_1^{(0)}}$ and $G_{g_2^{(0)}}$ represent the events that side g fires a radar-guided and infrared missile respectively that *misses* side $\overline{g}$. The EON shown in Figure 17 has 21 nodes. Two different node types are present in the graphical representation but the graph is not bipartite since arcs join two nodes from the same class (e.g., arc $(1,3)$). A square node represents the occurrence of an event that precludes the occurrence of future events. Termed a "*concluding*" event, this type of node completely describes an absorbing state in the underlying probabilistic model. For example, node 2 describes the concluding event $G_{1_1^{(1)}}$ that the aircraft on side 1 fired
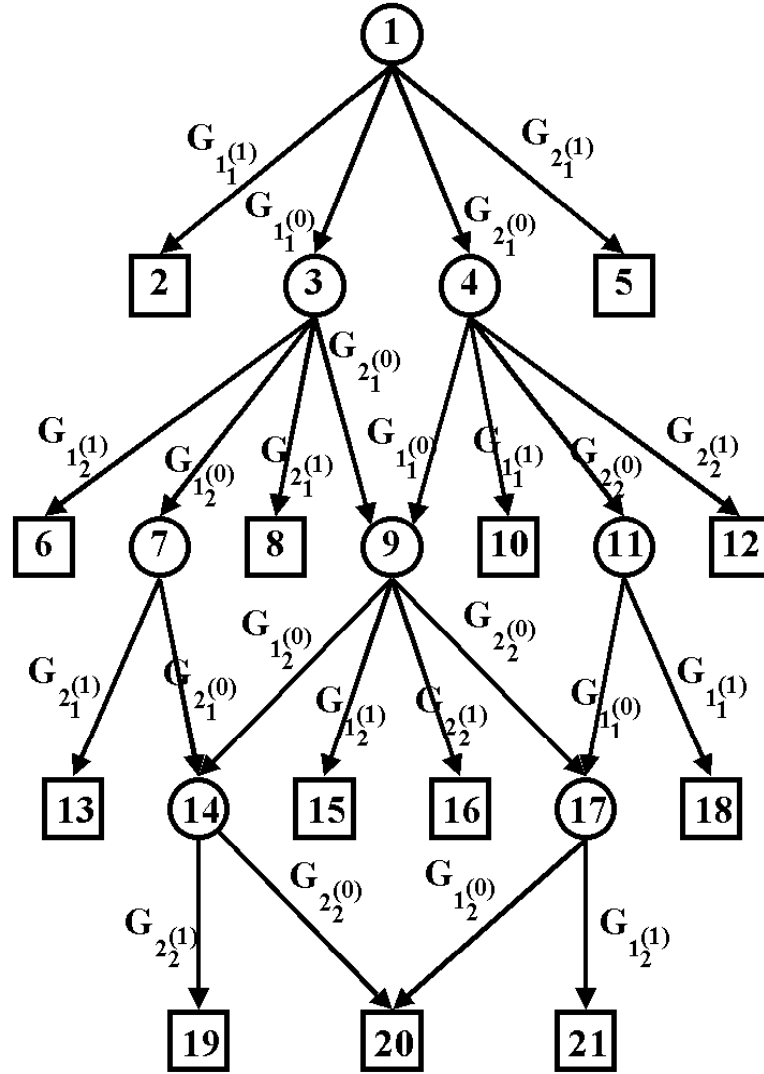
Figure 17. Event Occurrence Network for One vs One Air-to-Air Engagement Example with Missile Kills

a radar missile and shot down the aircraft on side 2. A circled node represents the occurrence of event that does not preclude the occurrence of future events. Termed an "*intermediate*" event, this type of node completely describes a transient state in the underlying probabilistic model. For example, node 3 represents the intermediate event $G_{1_1^{(0)}}$ that the aircraft on side 1 fired a radar missile missing the aircraft on side 2.

The EON shown in Figure 17 has 13 concluding events and 8 intermediate events[3]. Note that some of the square nodes share the same concluding event. For example, nodes 2, 10, and 18 all share the same concluding event $G_{1_1^{(1)}}$; i.e., the aircraft on side 1 fires a radar missile and shoots down the aircraft on side 2. However, the sequence of events leading up to the concluding event for each of these nodes differs. For node 2, the sequence of events is $(G_{1_1^{(1)}})$. For node 10, the sequence of events is $(G_{2_1^{(0)}}, G_{1_1^{(1)}})$. Finally, for node 18, the sequence of events is $(G_{2_1^{(0)}}, G_{2_2^{(0)}}, G_{1_1^{(1)}})$.

4.4   Underlying Probabilistic Model

Suppose one is interested in the probability of being at a particular node or set of nodes at time t in an EON. For example, what is the probability that the aircraft on side 1 shoots down the aircraft on side 2 with a radar missile at time t (the probability of being at node 2, 10, or 18) for the EON in Figure 17. *The probabilistic model generated by the EON structure is a non-Markovian probabilistic model.* In order to calculate the probability of being at a particular node at time t, we need to know the history of all completed events. The states of the underlying probabilistic model of EON correspond on a one-to-one basis with the nodes of the EON; *absorbing states correspond to event sequences that end with a concluding event and transient states correspond to event sequences that end with an intermediate event.* Square nodes are absorbing states and circled nodes are transient states. Appendix C details the probability of being at a particular node in Figure 17 at time t.

---

[3]Node 1 is a "non-event", but will also be classified as an intermediate event.

Table 2 gives $P_a(t)$ the probability of being at a particular node a (state in the underlying probabilistic model) at time t, $t = 1, 2, \ldots, 7$ for the EON in Figure 17 assuming that missile launch times are *exponentially distributed*. Specifically, if continuous random variable $E_{g_e(\kappa)}$ represents the completion time of event $G_{g_e(\kappa)}$ where side $g = 1, 2$ fires a missile of type e, $e = 1, 2$ (where 1 - radar-guided and 2 - infrared) at side $\bar{g}$ that either misses ($\kappa = 0$) or shoots down ($\kappa = 1$) side $\bar{g}$ where $\bar{g}$ represents the opponent side of side g, then

- $E_{1_1(\kappa)} \sim exp(\lambda_{1_1} t)$ with $\lambda_{1_1} = 1.25$ (i.e., side 1 fires a radar missile);

- $E_{2_1(\kappa)} \sim exp(\lambda_{2_1} t)$ with $\lambda_{2_1} = 1.15$ (i.e., side 2 fires a radar missile);

- $E_{1_2(\kappa)} \sim exp(\lambda_{1_2} t)$ with $\lambda_{1_2} = 1.85$ (i.e., side 1 fires an infrared missile); and

- $E_{2_2(\kappa)} \sim exp(\lambda_{2_2} t)$ with $\lambda_{2_2} = 1.65$ (i.e., side 2 fires an infrared missile).

Additionally, the probabilities that the radar and infrared missiles on side 1 and 2 shoot down the other side when fired are $p_{1_1} = p_{2_1} = 0.65$ and $p_{1_2} = 0.85$ - $p_{2_2} = 0.75$ respectively. Exact (analytic) expressions for the probabilities in Table 2 were derived and are summarized in Appendix C.

From Table 2, as $t \to 7$ minutes, the probability of being at a particular node reaches a steady-state value. Note that the probability of being at an intermediate node (transient state) approaches zero as $t \to 7$. The following observations can be made about the steady-state probability of being at a particular node for this air-to-air engagement example:

- First strike radar missile kills occur with a probability of .65 ($P_2(7) + P_5(7)$).

- Slightly over 90% of the air-to-air engagements have ended within two events (missile firings); $P_2(7) + P_5(7) + P_6(7) + P_8(7) + P_{10}(7) + P_{12}(7) = .9096$.

- Only slightly more than .4% of the air-to-air engagements end without a shoot down; $P_{20}(t) = .0046$.

90

Time t	0	1	2	3	4	5	6	7
$P_1(t)$	1.0000	0.0907	0.0082	0.0007	0.0001	0.0000	0.0000	0.0000
$P_2(t)$	0.0000	0.3078	0.3358	0.3383	0.3385	0.3385	0.3385	0.3385
$P_3(t)$	0.0000	0.0298	0.0042	0.0005	0.0000	0.0000	0.0000	0.0000
$P_4(t)$	0.0000	0.0287	0.0042	0.0005	0.0000	0.0000	0.0000	0.0000
$P_5(t)$	0.0000	0.2832	0.3089	0.3112	0.3114	0.3115	0.3115	0.3115
$P_6(t)$	0.0000	0.0712	0.0926	0.0952	0.0955	0.0955	0.0956	0.0956
$P_7(t)$	0.0000	0.0074	0.0042	0.0016	0.0005	0.0002	0.0001	0.0000
$P_8(t)$	0.0000	0.0339	0.0440	0.0453	0.0454	0.0454	0.0454	0.0454
$P_9(t)$	0.0000	0.0095	0.0021	0.0003	0.0000	0.0000	0.0000	0.0000
$P_{10}(t)$	0.0000	0.0347	0.0454	0.0468	0.0470	0.0470	0.0470	0.0470
$P_{11}(t)$	0.0000	0.0099	0.0054	0.0019	0.0006	0.0002	0.0001	0.0000
$P_{12}(t)$	0.0000	0.0528	0.0692	0.0713	0.0715	0.0716	0.0716	0.0716
$P_{13}(t)$	0.0000	0.0034	0.0079	0.0099	0.0106	0.0109	0.0109	0.0109
$P_{14}(t)$	0.0000	0.0023	0.0021	0.0010	0.0004	0.0001	0.0000	0.0000
$P_{15}(t)$	0.0000	0.0123	0.0207	0.0222	0.0223	0.0224	0.0224	0.0224
$P_{16}(t)$	0.0000	0.0097	0.0163	0.0174	0.0176	0.0176	0.0176	0.0176
$P_{17}(t)$	0.0000	0.0033	0.0028	0.0011	0.0004	0.0001	0.0000	0.0000
$P_{18}(t)$	0.0000	0.0050	0.0115	0.0142	0.0151	0.0154	0.0155	0.0155
$P_{19}(t)$	0.0000	0.0012	0.0043	0.0062	0.0070	0.0072	0.0073	0.0074
$P_{20}(t)$	0.0000	0.0008	0.0028	0.0039	0.0044	0.0045	0.0046	0.0046
$P_{21}(t)$	0.0000	0.0022	0.0075	0.0105	0.0116	0.0119	0.0120	0.0121
Total	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 2. Probability of Being at a Node for the One vs One Air-to-Air Engagement Event Occurrence Network Example with Missile Kills

- In answer to the example question posed above, the probability that the aircraft on side 1 shoots down the aircraft on side 2 with a radar missile is .401 ($P_2(7) + P_{10}(7) + P_{18}(7)$).

As demonstrated in Appendix C, the probability of being at any node (state) of the EON in Figure 17 at time t can be calculated exactly when the event probability density functions are all of exponential form (due to closure of the exponential distribution under the operation of integration). However, for many event probability density functions (e.g., gamma with non-integer shape parameter) or combinations of event probability density functions (e.g. exponential and normal), the integral expressions given in Appendix C have no closed form solution. In such cases, approximations of the integral expressions are needed. One solution technique discussed in the literature is polynomial approximation. Discussion of this technique is deferred until Chapter V.

4.5 Node (State) Explosion Problem

A drawback of the EON approach to modeling the interaction of groups of event sequences is a node (state) *explosion* problem as the number of the events in the network is increased. For example, Table 3 shows the number of nodes (states) for s group EONs with s events in each group ($s = 1, \ldots, 6$). This explosion problem is directly linked to the requirement to explicitly track event sequences. Since the underlying probabilistic model of an EON is non-Markovian, past event occurrences are important in determining the probability of being at a current node and the time associated with reaching the node.

One approach to minimizing the network explosion problem is network aggregation. This approach is used by Ancker, Kress, Gafarian, and associates [83] [120] [84] [102] [147] [198] in modeling m versus n stochastic ground combat. Recall from chapter I that aggregation refers to the process of assembling or combining in mass. Aggregation in EONs is process aggregation. This approach attempts to combine

s	Number of Nodes
2	21
3	208
4	2625
5	40176
6	722701

Table 3. Number of Nodes for s Group Event Occurrence Networks with s Events in Each Group

sets of events into fewer, reduced (aggregated) sets of events; i.e., reduce the number of states in the probabilistic model. The cost paid for aggregation is a loss of information (detail) in the model.

For example, the one versus one air-to-air engagement EON shown in Figure 17 may be aggregated into the EON shown in Figure 18. This EON only has 4 nodes (3 concluding and 1 intermediate) associated with two killing events. More precisely, let G_g describe the event that the aircraft from side g shoots down the aircraft from side $\bar{g}$, $g = 1, 2$. Node 1 describe the non-event that a kill on either side has not yet taken place and both sides still have missiles to launch. Nodes 2 and 4 describe the concluding event (i.e., single event sequence) that the aircraft on side 1 and 2 respectively shoots down the aircraft on the other side. Finally, node 3 describes the concluding event that both sides launch all missiles without achieving a kill; engagement ends without a shoot down.

For the aggregated EON and the associated probabilistic model in the Figure 18, the following nodes and associated states have been combined from the unaggregated EON in Figure 17:

- Node 1_{agg} = Nodes $1, 3, 4, 7, 9, 11, 14,$ and 17;

- Node 2_{agg} = Nodes $2, 6, 10, 15, 18,$ and 21;

- Node 3_{agg} = Nodes $5, 8, 12, 13, 16$ and 19; and

- Node 4_{agg} = Node 20

93
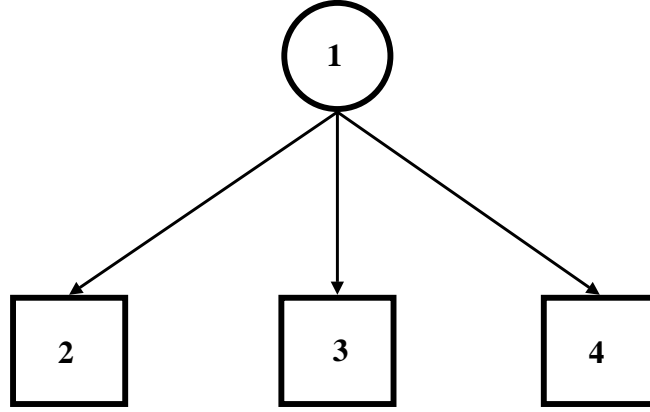
Figure 18.   Aggregated Event Occurrence Network for One vs One Air-to-Air Engagement Example with Missile Kills

where a_{agg} describes the new (aggregated) nodes. Table 4 gives $P_a^{agg}(t)$ the probability of being at a particular node a (state in the underlying probabilistic model) at time t, $t = 1, \ldots, 7$ for the aggregated EON in Figure 18 assuming the same parameters as used to calculate Table 2 for the unaggregated EON.

The probability of being a particular node at time t as shown in Table 4 may be calculated using two different methods:

Time t	0	1	2	3	4	5	6	7
$P_1^{agg}(t)$	1.0000	0.1817	0.0332	0.0075	0.0020	0.0006	0.0001	0.0000
$P_2^{agg}(t)$	0.0000	0.4333	0.5134	0.5272	0.5301	0.5308	0.5309	0.5310
$P_3^{agg}(t)$	0.0000	0.0008	0.0028	0.0039	0.0044	0.0045	0.0046	0.0046
$P_4^{agg}(t)$	0.0000	0.3842	0.4506	0.4614	0.4636	0.4641	0.46.43	0.4644
Total	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4.   Probability of Being at a Node for the One vs One Air-to-Air Engagement Aggregated Event Occurrence Network Example with Missile Kills

1. By summing the probabilities of being at the appropriate original nodes in the unaggregated EON. For example, $P_1^{agg} = P_1(t) + P_3(t) + P_4(t) + P_7(t) + P_9(t) + P_{13}(t) + P_{14}(t) + P_{17}(t)$.

2. By defining new distribution(s) which model the occurrence of the new aggregated event(s) from the unaggregated event distributions. For example, one aggregated event for the EON in Figure 18 is the shooting down of an aircraft. The density function for such an event can be defined using the density functions of the unaggregated events:

$$f_{g_{agg}}(t) = p_{1_1} f_{E_{g_1^{(\kappa)}}}(t) + q_{1_1} p_{1_2} [f_{E_{g_1^{(\kappa)}}}(t) \oplus f_{E_{g_2^{(\kappa)}}}(t)] + q_{1_1} q_{2_1} \delta(t - \infty) \quad (38)$$

for $g = 1, 2$ where $\oplus$ denotes the convolution of the density functions and δ is the Dirac Delta function[4].

The density function in equation 38 is a finite mixture of density functions. Note that detail has been lost with the aggregated EON; the distribution of time for the aircraft on side g to shoot down the aircraft on side $\bar{g}$ is known, but not which missile type shoots down the aircraft.

Another approach to the node (state) explosion problem is a method loosely termed by this researcher as 'lq bucket" analysis (based on the analogy of a fluid flowing down through a system of buckets). In this method, *intermediate nodes of an EON may be transformed into concluding nodes*, effectively reducing the number of nodes in follow-on event tiers. This approach allows an EON to be "truncated" to the appropriate event level. For example, consider the EON in Figure 17 once again. The steady-state findings of Table 2 showed that slightly more than 90% of the air-

---

[4]In equation 38, the Dirac Delta function is defined such that:

$$\delta(s) = \begin{cases} 1 & s \geq 0 \\ 0 & s < 0 \end{cases}$$

to-air engagements end within two events (missile firings); $P_2(7) + P_3(7) + P_6(7) + P_8(7) + P_{10}(7) + P_{12}(7) = .9096$. Rather than calculate the probability of being at the remaining nodes $7, 9, 11, 13$, thru 21, one can change nodes 7, 9, and 11 to concluding nodes and calculate the probability of being at these three nodes instead of calculating the probability of the other 9 remaining nodes $(13-21)$ in the network.

The method of "bucket" analysis eliminates rare (as defined by the analyst) event sequences. In terms of the fluid analogy, the underlying probabilistic model of an EON can be viewed as a construct for the "flow of probability". Clearly, the probability of being at node 1 in any EON attime $t = 0$ is one; no events have occurred by $t = 0$. Probability then "flows" to other nodes in the network based on the occurrence of event sequences. Absorbing states can be viewed as buckets capturing fluid (probability) from higher nodes in the network and transient states as leaky buckets capturing fluid from higher nodes, but feeding this fluid to lower nodes in the EON. By changing an intermediate node in an EON to a concluding node, one has changed a transient state to an absorbing state in the underlying probabilistic model.

Additionally, through the use of "bucket" analysis, one does not have to solve for the probability of being in a transient state directly. By viewing transient states as absorbing states, one can immediately calculate the probability of being in a particular transient state. For example, consider calculating $P_3(t)$ for the EON in Figure 17. Using conservation of probability:

$$P_3(t) = P_{\lfloor 3 \rfloor}(t) - [P_6(t) + P_{\lfloor 7 \rfloor}(t) + P_8(t) + P_{\lfloor 9b_1 \rfloor}(t)]$$

where $P_{\lfloor a \rfloor}(t)$ is the probability of being in transient state a *when state a is viewed as an absorbing state* and $9b_1$ is event sequence $(G_{1_1^{(0)}}, G_{2_1^{(0)}})$. Although one still must calculate the probabilities associated with viewing a transient state as an absorbing state, this computation is generally simplified due to the state's relationship with

an associated absorbing state(s). For example, truncated state 7 is associated with absorbing state 6 in Figure 17. Equation 39 below gives the probability of being absorbing state 6.

$$
\begin{aligned}
P_6(t) \;=\; & \frac{q_{1_1} p_{1_2} \lambda_{1_1} \lambda_{1_2}}{(\lambda_{2_1} + \lambda_{1_2})(\lambda_{1_1} + \lambda_{2_1})} \left[1 + \frac{(\lambda_{1_1} + \lambda_{2_1})}{(\lambda_{1_2} - \lambda_{1_1})} \exp(-(\lambda_{2_1} + \lambda_{1_2})t) \right. \\
& \left. - \frac{(\lambda_{2_1} + \lambda_{1_2})}{(\lambda_{1_2} - \lambda_{1_1})} \exp(-(\lambda_{1_1} + \lambda_{2_1})t) \right]
\end{aligned}
\tag{39}
$$

The probability of being in truncated state 7 is the same as equation 39 except p_{1_2} is replaced with q_{1_2}.

4.6   Integral Solution

In order to solve for the probability of being at a particular node or set of nodes at time t in a general s group event occurrence network (EON), one of two solution approaches may be used: simulation or integral approximation. In order to use the integral approximation solution technique, the probability of being at any node in the network must be expressed in integral form. In this section, the integral solution to a general s group EON is presented.

4.6.1   State 1 Integral Expression.   As shown in Appendix C, the probability of being at node (in state) 1 at time t, $P_1(t)$, for the air-to-air example in Figure 17 is:

$$
P_1(t) = P(E_{1_1^{(\kappa)}} > t, E_{2_1^{(\kappa)}} > t).
$$

$P_1(t)$ is equivalent to the probability of the first event not occurring by time t and can be written in the form:

$$
P_1(t) = F^c_{E_{1_1^{(\kappa)}}}(t) * F^c_{E_{2_1^{(\kappa)}}}(t)
\tag{40}
$$

97

due to the independence of events between event groups. For a s group EON, the expression in equation 40 is generalized as:

$$P_1(t) = \prod_{i=1}^{s} F^c_{E_{g_1^{(\kappa)}}}(t).$$

(41)

Hence, in order to determine the probability of being in state 1 at time t of an event occurrence network, one needs to determine the complementary cumulative distribution functions of the time to first event completions in each event grouping. Additionally, the expression in equation 41 is equivalent to

$$P_1(t) = 1 - F_{min}(t)$$

(42)

where $F_{min}(t)$ is the minimum of first event occurrence distributions.

4.6.2   Event Tier 1 Integral Expressions.   For the first event tier (those nodes (states) where the first event is completed in the EON probabilistic model), $2s$ nodes are possible for a s group EON. These $2s$ nodes consist of s pairs of concluding and intermediate nodes. A concluding node (absorbing state) indicates the completion of the first event from event group g with "success" (discrete) probability p_{g_1} (e.g., $G_{g_1^{(1)}}$). The corresponding intermediate node (transient state) indicates a first event completion with "failure" probability $1 - p_{g_1}$ (e.g. $G_{g_1^{(0)}}$). For example, consider the air-to-air example in Figure 17. Since there are two event groups in this example, four states are associated with the first event tier (two absorbing and two transient states). States 2 and 3 are the corresponding absorbing and transient states associated with the first event completion from event group one (G_{1_1}) while states 3 and 4 are associated with the first event completion from event group two (G_{2_1}).

98

The probability of being in (absorbing) state 2 at time t, $P_2(t)$, for the air-to-air example in Figure 17 as shown in Appendix C is:

$$
\begin{aligned}
P_2(t) &= P(E_{2_1^{(\kappa)}} > E_{1_1^{(1)}}) \\
&= \int_0^t F^c_{E_{2_1^{(\kappa)}}}(e_{1_1}) f_{E_{1_1^{(1)}}}(e_{1_1}) de_{1_1}.
\end{aligned}
\tag{43}
$$

For a r group EON, equation 43 may be generalized to:

$$
\begin{aligned}
P_{2g}(t) &= P(E_{1_1^{(\kappa)}} > E_{g_1^{(1)}}, \ldots, E_{g-1_1^{(\kappa)}} > E_{g_1^{(1)}}, \\
&\qquad E_{g+1_1^{(\kappa)}} > E_{g_1^{(1)}}, \ldots, E_{s_1^{(\kappa)}} > E_{g_1^{(1)}}) \\
&= \prod_{\substack{j=1 \\ j\neq g}}^{s} P(E_{j_1^{(\kappa)}} > E_{g_1^{(1)}}) \\
&= \int_0^t \prod_{\substack{j=1 \\ j\neq g}}^{s} F^c_{E_{j_1^{(\kappa)}}}(e_{g_1}) f_{E_{g_1^{(1)}}}(e_{g_1}) de_{g_1}
\end{aligned}
\tag{44}
$$

for any *first tier* absorbing state $2g$, $g = 1, \ldots, s$.

The probability of being in (transient) state 3 at time t, $P_3(t)$, for the air-to-air example in Figure 17 as shown in Appendix C is:

$$
\begin{aligned}
P_3(t) &= P(E_{1_2^{(\kappa)}} > t - E_{1_1^{(0)}}, E_{2_1^{(\kappa)}} > t) \\
&= P(E_{2_1^{(\kappa)}} > t)\, P(E_{1_2^{(\kappa)}} > t - E_{1_1^{(0)}}) \\
&= F^c_{E_{2_1^{(\kappa)}}}(t) * \left[\int_0^t F^c_{E_{1_2^{(\kappa)}}}(t - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} \right].
\end{aligned}
\tag{45}
$$

For a s group EON, equation 45 may be generalized to:

$$
\begin{aligned}
P_{2g+1}(t) &= P(E_{1_1^{(\kappa)}} > t, \ldots, E_{g-1_1^{(\kappa)}} > t, E_{g_2^{(\kappa)}} > t - E_{g_1^{(0)}}, \\
&\qquad E_{g+1_1^{(\kappa)}} > t, \ldots, E_{s_1^{(\kappa)}} > t)
\end{aligned}
$$

99

$$= \prod_{\substack{j=1 \\ j \neq g}}^{s} P(E_{j_1^{(\kappa)}} > t) \, P(E_{g_2^{(\kappa)}} > t - E_{g_1^{(0)}})$$

$$= \prod_{\substack{j=1 \\ j \neq g}}^{s} F_{E_{j_1^{(\kappa)}}}^c (t) * \left[\int_0^t F_{E_{g_2^{(\kappa)}}}^c (t - e_{g_1}) f_{E_{g_1^{(0)}}}(e_{g_1}) de_{g_1} \right] \tag{46}$$

for any first tier transient state $2g + 1$, $g = 1, \ldots, s$. Comparing equations 41 and 46, $P_{2g+1}(t)$ is similar in form to $P_1(t)$ except that $F_{E_{g_1^{(\kappa)}}}^c (t)$ has been replaced by:

$$\int_0^t F_{E_{g_2^{(\kappa)}}}^c (t - e_{g_1}) f_{E_{g_1^{(0)}}}(e_{g_1}) de_{g_1} \tag{47}$$

to reflect the fact that the first event from group g has been completed with probability $1 - p_{g_1}$, but the second event from group g has not yet occurred. If group g does not have a second event, the integral in equation 47 is replaced with $F_{E_{g_1^{(0)}}}(t)$ and equation 46 simplifies to:

$$P_{2g+1}(t) = \prod_{\substack{j=1 \\ j \neq g}}^{s} F_{E_{g_1^{(\kappa)}}}^c (t) * F_{E_{g_1^{(0)}}}(t) \tag{48}$$

4.6.3   Event Tier u Integral Expressions.   The number of event tiers in an EON is the sum of the number of events in each event group in the network. Specifically, the number of event tiers, n_T, in a s group EON is:

$$n_T = \sum_{g=1}^{s} n_g \tag{49}$$

where n_g is the number of events in event group g. For example, $n_T = 36$ for a 6 group EON with 6 events in each group.

Section 4.6.2 showed that $2s$ possible nodes exist in the first event tier for a s group EON . For event tier u, $2 \leq u \leq n_T$, the number of nodes in the network depends on two interrelated factors,

1. The number of intermediate nodes in the previous event tier $u-1$ and

2. The number of event groups that still have events that can occur.

As an illustration consider the second event tier. For the second event tier, between $2(s-1)$ and $2s^2$ nodes are possible for a s group EON. Specifically for *each* intermediate node in the first event tier, either $s-1$ or s second tier (follow-on) nodes can occur representing:

- $s-1$ events from each of the remaining $s-1$ event groups that did have a (first) event occur in the first event tier and

- a (possible) second event from gth event group (but, only if $n_g > 1$).

Since an event occurrence can be a success (concluding node) or a failure (intermediate node), for *each* intermediate node in the first event tier between $2(s-1)$ and $2s$ nodes are possible. Considering there are s intermediate nodes in the first event tier, the total number of nodes for the second event tier can range from $2s(s-1)$ to $2s^2$ nodes.

The air-to-air EON example in Figure 17 contains two event groups with two events each, thus eight nodes ($2s^2$ nodes) are predicted for the second event tier (four concluding and four intermediate nodes). However, there are only seven nodes (four concluding and three intermediate nodes) present in Figure 17 for the second event tier. The difference is a result of combining in the EON graph, event sequences that have the same set of completed events. In the air-to-air example in Figure 17, node 9 accounts for the two event sequences $(G_{1_1^{(0)}}, G_{2_1^{(0)}})$ and $(G_{1_1^{(0)}}, G_{2_1^{(0)}})$ that can be generated by the event set $\left\{ G_{1_1^{(0)}}, G_{2_1^{(0)}} \right\}$.

Node combining in an EON graph only happens for sets of event occurrences that are all failures (that is, when combining intermediate nodes (transient states)). Each concluding node (absorbing state) in an EON describes a unique set of events. In other words, no two concluding nodes describe the same set of completed events.

At the second event tier, $s(s-1)$ intermediate nodes may be combined into $\frac{s(s-1)}{2}$ nodes.

The probability of being in any concluding or intermediate node (absorbing or transient state) in event tier u, denoted *absorb* and *trans* respectively, at time t can be expressed as a *single*, multi-dimensional integral. Specifically for event tier u, the probability of being in absorbing state *absorb* is:

$$
\begin{aligned}
P_{absorb}(t) &= P(E_{1:T^{(\iota_1)}} \, \alpha_1 \, E_{g:T^{(1)}} \mid E_{1:T-1^{(0)}} < E_{g:T^{(1)}}, \ldots, \\
&\quad E_{g-1:T^{(\iota_{g-1})}} \, \alpha_{g-1} \, E_{g:T^{(1)}} \mid E_{g-1:T-1^{(0)}} < E_{g:T^{(1)}}, \\
&\quad E_{g+1:T^{(\iota_{g+1})}} \, \alpha_{g+1} \, E_{g:T^{(1)}} \mid E_{g+1:T-1^{(0)}} < E_{g:T^{(1)}}, \ldots, \\
&\quad E_{s:T^{(\iota_s)}} \, \alpha_s \, E_{g:T^{(1)}}) \mid E_{s:T-1^{(0)}} < E_{g:T^{(1)}}) \\
&= \int_0^t \prod_{\substack{j=1 \\ j \neq g}}^{s} P(E_{j:T^{(\iota_j)}} \, \alpha_j \, e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T}) f_{E_{g:T^{(1)}}}(e_{g:T}) \, de_{g:T} \quad (50)
\end{aligned}
$$

where $E_{j:T^{(\iota_j)}}$ $(j = 1, \ldots, g-1, g+1, \ldots, s)$ represents the time of occurrence $(\iota_j = 0)$ / non-occurrence $(\iota_j = \kappa)$ of the first T_j events in event group j and α_j $(j = 1, \ldots, g-1, g+1, \ldots, s)$ represents the probabilistic ordering of events $E_{j:T^{(\iota_j)}}$ and $E_{g:T^{(1)}}$ $(\alpha_j : \leq \text{ or } \alpha_j : >)$.

For example, consider concluding node (absorbing state) 72 in the 3 group by 2 event EON shown in Figure 19. This state occurs in event tier 5 as a result of concluding event $G_{1_2^{(1)}}$. The four intermediate events leading to $G_{1_2^{(1)}}$ are events $G_{1_1^{(0)}}$, $G_{2_1^{(0)}}$, $G_{2_2^{(0)}}$, and $G_{3_1^{(0)}}$. From equation 50, the probability of being in state 72 at time t, $P_{72}(t)$, is:

$$
\begin{aligned}
P_{72}(t) &= \int_0^t P(E_{2:2^{(0)}} \leq e_{1:2} \mid E_{2:1^{(0)}} < e_{1:2}) \\
&\quad P(E_{3:2^{(0)}} > e_{1:2} \mid E_{3:1^{(0)}} < e_{1:2}) f_{E_{1:2^{(1)}}}(e_{1:2}) \, de_{1:2}
\end{aligned}
$$
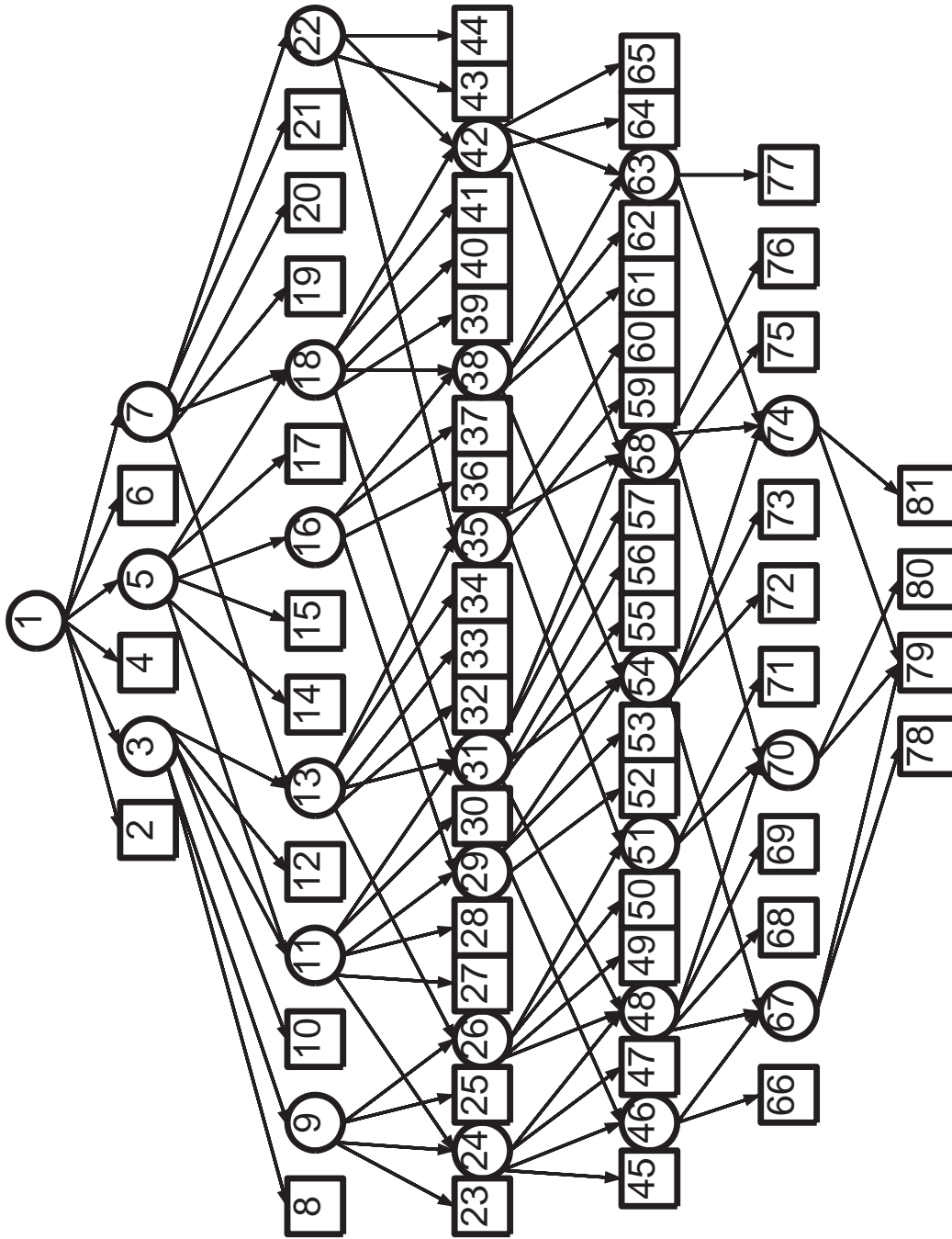
Figure 19. Event Occurrence Network with 3 Event Groups with 2 Events in Each Group

In the above expression for random variable $E_{2:2^{(\iota_2)}}$, $\alpha_2 :\leq$ and $\iota_2 = 0$ reflecting that intermediate events $G_{2_1^{(0)}}$ and $G_{2_2^{(0)}}$ (the two events in event group 2) have occurred sometime before concluding event $G_{1_2^{(1)}}$. Similarly for $E_{3:2^{(\iota_3)}}$, $\alpha_2 :>$ and $\iota_3 = k$ reflecting that intermediate event $G_{3_1^{(0)}}$ occurred before $G_{1_2^{(1)}}$, but $G_{1_2^{(1)}}$ occurred before event $G_{3_2^{(\kappa)}}$. Substituting expressions for $P(E_{2:2^{(0)}} \leq e_{1:2} \mid E_{2:1^{(0)}} < e_{1:2})$, $P(E_{3:2^{(0)}} > e_{1:2} \mid E_{3:1^{(0)}} < e_{1:2})$, and $f_{E_{1:2^{(1)}}}(e_{1:2})de_{1:2}$:

$$
\begin{aligned}
P_{72}(t) \;=\; & \int_0^t \left[\int_0^{e_{1T}} \int_0^{e_{2T}} f_{E_{2_2^{(0)}}}(e_{2T} - e_{2_1}) f_{E_{2_1^{(0)}}}(e_{2_1}) de_{2_1} de_{2T} \right] \\
& \left[\int_{e_{1T}}^{\infty} \int_0^{e_{1T}} f_{E_{3_2^{(\kappa)}}}(e_{3T} - e_{3_1}) f_{E_{3_1^{(0)}}}(e_{3_1}) de_{3_1} de_{3T} \right] \\
& \int_0^{e_{1T}} f_{E_{1_2^{(1)}}}(e_{1T} - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} de_{1T} \qquad (51)
\end{aligned}
$$

which is a multiple integral with 6 dimensions.

For event tier u, the probability of being in intermediate node (transient state) *trans* is

$$
\begin{aligned}
P_{trans}(t) \;=\; & P(E_{1:T^{(\iota_1)}} \, \alpha_1 \, t \mid E_{1:T-1^{(0)}} < t, \; \ldots \; , E_{s:T^{(\iota_s)}} \, \alpha_s \, t \mid E_{s:T-1^{(0)}} < t) \\
\;=\; & \prod_{j=1}^{s} P(E_{j:T^{(\iota_j)}} \, \alpha_j \, t \mid E_{j:T-1^{(0)}} < t) \qquad (52)
\end{aligned}
$$

where $E_{j:T^{(\iota_j)}}$ $(j = 1, \ldots, s)$ represents the time of occurrence $(\iota_j = 0)$ / non-occurrence $(\iota_j = \kappa)$ of the first T_j events in event group j and α_j $(j = 1, \ldots, s)$ represents the probabilistic ordering of events $E_{j:T^{(\iota_j)}}$ $(\alpha_j :\leq$ or $\alpha_j :>)$.

As an illustration, consider intermediate node (transient state) 67 in the 3 group by 2 event EON shown in Figure 19. This state occurs in event tier 5 as a result of intermediate events $G_{1_1^{(0)}}$, $G_{1_2^{(0)}}$, $G_{2_1^{(0)}}$, $G_{2_2^{(0)}}$, and $G_{3_1^{(0)}}$. From equation 52,

104

the probability of being in state 67 at time t, $P_{67}(t)$,is:

$$P_{67}(t) = P(E_{1:2^{(0)}} \leq t)P(E_{2:2^{(0)}} \leq t)P(E_{3:2^{(\kappa)}} > t \mid E_{3:1^{(0)}} < t)$$

In the above expression for random variable $E_{1:2^{(\iota_1)}}$, $\alpha_1 :\leq$ and $\iota_1 = 0$ reflecting that all events in event group 1 have taken place by time t. Similarly for $E_{2:2^{(\iota_2)}}$, $\alpha_2 :\leq$ and $\iota_2 = 0$. For $E_{3:2^{(\iota_3)}}$, $\alpha_3 :>$ and $\iota_3 = \kappa$ reflecting that event $G_{3_1^{(0)}}$ occurred before time t, but event $G_{3_2^{(\kappa)}}$ has yet to occur by t. Substituting expressions for $P(E_{1:2^{(0)}} \leq t)$, $P(E_{2:2^{(0)}} \leq t)$, and $P(E_{3:2^{(\kappa)}} > t \mid E_{3:1^{(0)}} < t)$:

$$P_{67}(t) = \left[\int_0^t \int_0^{e_{1T}} f_{E_{1_2^{(0)}}}(e_{1T} - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} de_{1T} \right]$$
$$\left[\int_0^t \int_0^{e_{2T}} f_{E_{2_2^{(0)}}}(e_{2T} - e_{2_1}) f_{E_{2_1^{(0)}}}(e_{2_1}) de_{2_1} de_{2T} \right]$$
$$\left[\int_t^\infty \int_0^t f_{E_{3_2^{(\kappa)}}}(e_{3T} - e_{3_1}) f_{E_{3_1^{(0)}}}(e_{3_1}) de_{3_1} de_{3T} \right]$$

which is also a multiple integral in 6 dimensions.

For the last (n_T th) event tier of an EON, all nodes describe concluding events (absorbing states) in the network. The probability of being at any of the nodes in this event tier at time t, except one, is calculated using equation 50. The exception is the case where all events end in failure. The probability of being in this absorbing state at time t is calculated using equation 52 where $\alpha_j :\leq$ and $\iota_j = 0$, $\forall j$, $j = 1, \ldots, s$.

4.7 Bucket Analysis

In section 4.5, the concept of bucket analysis was introduced. This method reduces the number of nodes (states) in an EON by transforming *intermediate nodes into concluding nodes*; thereby effectively reducing the number of nodes in follow-on event tiers. As a further illustration, consider the 3 group EON in Figure 19 where each group has 2 events. This network has 81 states. Suppose as $t \rightarrow \infty$, $P_{[7]}(t) < 0.005$ (the (bucket) probability that event $G_{3_1^{(0)}}$ is completed before either

event $G_{1_1^{(0)}}$ or $G_{2_1^{(0)}}$ is "rare"). Since the node paths proceeding from node 7 will not contribute significant (as defined by the analyst) probability to nodes in follow-on event tiers, node 7 can be transformed from an intermediate node to a concluding node. Figure 20 shows this transformation. By truncating node 7, the number of nodes in the network are reduced from 81 to 75 (Nodes 19, 20, 21, 22, 43, and 44 are eliminated from Figure 19).

In section 4.6, single, multi-dimensional integral expressions were derived for the probability of being at a given concluding or intermediate node (absorbing or transient state) in an EON at time t. Although many possible paths (events sequences) can lead to a given state, *the probability of being in that state at time t is independent of the path actually taken.* In other words, individual paths to a given state do not have to be tracked explicitly and the probability "delivered" by these paths calculated in order to determine the probability of being in the state. For example consider the probability of being in transient state 9 at time t, $P_9(t)$, for the air-to-air example in Figure 17. Using equation 52,

$$P_9(t) = \left[\int_t^\infty \int_0^t f_{E_{1_2^{(\kappa)}}}(e_{1_T} - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} de_{1_T} \right]$$
$$\left[\int_t^\infty \int_0^t f_{E_{2_2^{(\kappa)}}}(e_{2_T} - e_{2_1}) f_{E_{2_1^{(0)}}}(e_{2_1}) de_{2_1} de_{2_T} \right] \qquad (53)$$

Two event sequences lead to state 9: $(G_{1_1^{(0)}}, G_{2_1^{(0)}})$ and $(G_{2_1^{(0)}}, G_{1_1^{(0)}})$. The integral in equation 53 can be decomposed into the following individual path contributions,

$$P_9(t) = \int_t^\infty \int_0^t \left[\int_t^\infty \int_{e_{1_1}}^t f_{E_{2_2^{(\kappa)}}}(e_{2_T} - e_{2_1}) f_{E_{2_1^{(0)}}}(e_{2_1}) de_{2_1} de_{2_T} \right]$$
$$f_{E_{1_2^{(\kappa)}}}(e_{1_T} - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} de_{1_T} +$$
$$\int_t^\infty \int_0^t \left[\int_t^\infty \int_0^{e_{1_1}} f_{E_{2_2^{(\kappa)}}}(e_{2_T} - e_{2_1}) f_{E_{2_1^{(0)}}}(e_{2_1}) de_{2_1} de_{2_T} \right]$$
$$f_{E_{1_2^{(\kappa)}}}(e_{1_T} - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} de_{1_T}. \qquad (54)$$
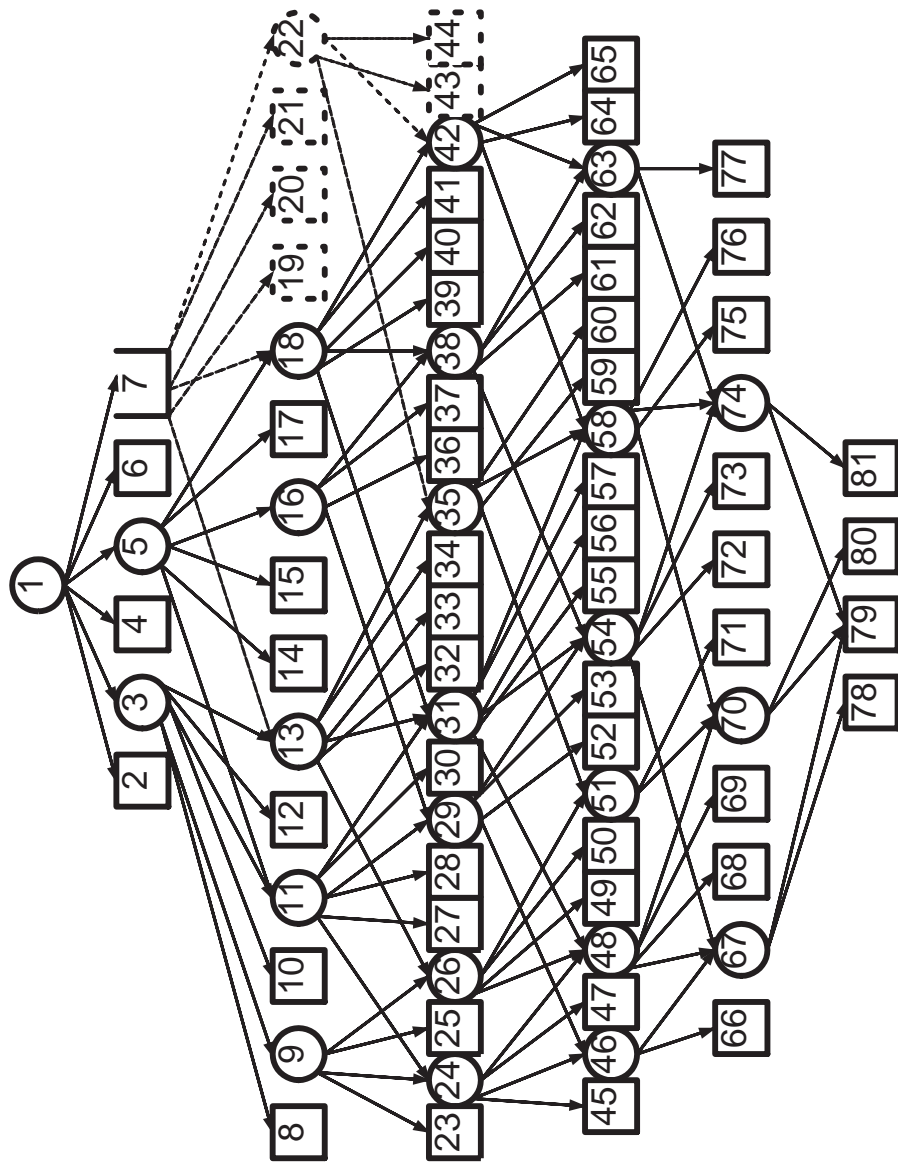
Figure 20.   Revised Event Occurrence Network with 3 Event Groups with 2 Events
in Each Group with Node 7 Truncated

The first integral in the equation 54 represents the probability of being in state 9 at time t due to the event sequence $(G_{1_1^{(0)}}, G_{2_1^{(0)}})$ while the second integral represents the probability of being in state 9 due to sequence $(G_{2_1^{(0)}}, G_{1_1^{(0)}})$.

In general, the single integral expression representing the probability of being in any state of an EON at time t can be decomposed into separate integrals representing the individual path contributions of possible event sequences. This decomposition can be generated by conditioning on the random variables within one event group and decomposing the integral expressions of other event groups based on the conditioned random variables. For the integral expressions in equation 54, the decomposition was generated by first conditioning on the random variables in event group one and then decomposing the integral expression in event group two to reflect the cases that event $G_{1_1^{(0)}}$ is completed *before* event $G_{2_1^{(0)}}$ (first integral) and $G_{1_1^{(0)}}$ is completed *after* event $G_{2_1^{(0)}}$ (second integral).

When nodes are truncated, not only are other nodes eliminated but event sequence(s) for remaining nodes may be eliminated as well. As an illustration consider the truncation of node 7 in the 3 group EON shown in Figure 19. By truncating node 7, paths from node 7 to nodes 13 and 18 have been eliminated. Additionally, the network paths from node 7 through node 22 to nodes 35 and 42 have been eliminated as well. *If a node truncation removes event sequence(s) terminating at a given state then equations 50 and 52 are no longer exact solutions for the probability for being in that state at time t.* For example, consider the probability of being at node 72 at time t, $P_{72}(t)$. Twelve event sequences terminate at node 72. Table 5 shows these sequences and the associated node paths. If node 7 is truncated, the three event sequences with asterisks (5, 6, and 12) are removed (since the paths associated with these sequences contain node 7). Equation 51 no longer exactly represents $P_{72}(t)$ when node 7 is truncated – by eliminating event sequences 5, 6, and 12, the probability associated with sequences has also been eliminated. Since equation 51 contains the probability from the three eliminated event sequences, the equation can

Number	Event Sequence	Node Path
1	$(G_{1_1}^{(0)},G_{2_1}^{(0)},G_{3_1}^{(0)},G_{2_2}^{(0)},G_{1_2}^{(1)})$	$1 \to 3 \to 11 \to 31 \to 54 \to 72$
2	$(G_{1_1}^{(0)},G_{3_1}^{(0)},G_{2_1}^{(0)},G_{2_2}^{(0)},G_{1_2}^{(1)})$	$1 \to 3 \to 13 \to 31 \to 54 \to 72$
3	$(G_{2_1}^{(0)},G_{1_1}^{(0)},G_{3_1}^{(0)},G_{2_2}^{(0)},G_{1_2}^{(1)})$	$1 \to 5 \to 11 \to 31 \to 54 \to 72$
4	$(G_{2_1}^{(0)},G_{3_1}^{(0)},G_{1_1}^{(0)},G_{2_2}^{(0)},G_{1_2}^{(1)})$	$1 \to 5 \to 18 \to 31 \to 54 \to 72$
5*	$(G_{3_1}^{(0)},G_{1_1}^{(0)},G_{2_1}^{(0)},G_{2_2}^{(0)},G_{1_2}^{(1)})$	$1 \to 7 \to 13 \to 31 \to 54 \to 72$
6*	$(G_{3_1}^{(0)},G_{2_1}^{(0)},G_{1_1}^{(0)},G_{2_2}^{(0)},G_{1_2}^{(1)})$	$1 \to 7 \to 18 \to 31 \to 54 \to 72$
7	$(G_{1_1}^{(0)},G_{2_1}^{(0)},G_{2_2}^{(0)},G_{3_1}^{(0)},G_{1_2}^{(1)})$	$1 \to 3 \to 11 \to 29 \to 54 \to 72$
8	$(G_{2_1}^{(0)},G_{1_1}^{(0)},G_{2_2}^{(0)},G_{3_1}^{(0)},G_{1_2}^{(1)})$	$1 \to 5 \to 11 \to 29 \to 54 \to 72$
9	$(G_{2_1}^{(0)},G_{2_2}^{(0)},G_{1_1}^{(0)},G_{3_1}^{(0)},G_{1_2}^{(1)})$	$1 \to 5 \to 16 \to 29 \to 54 \to 72$
10	$(G_{2_1}^{(0)},G_{3_1}^{(0)},G_{2_2}^{(0)},G_{1_1}^{(0)},G_{1_2}^{(1)})$	$1 \to 5 \to 18 \to 38 \to 54 \to 72$
11	$(G_{2_1}^{(0)},G_{2_2}^{(0)},G_{3_1}^{(0)},G_{1_1}^{(0)},G_{1_2}^{(1)})$	$1 \to 5 \to 16 \to 38 \to 54 \to 72$
12*	$(G_{3_1}^{(0)},G_{2_1}^{(0)},G_{2_2}^{(0)},G_{1_1}^{(0)},G_{1_2}^{(1)})$	$1 \to 7 \to 18 \to 38 \to 54 \to 72$

Table 5.   Event Sequences and Associated Paths for Node 72 for the 3 Group Event Occurrence Network with 2 Events in Each Group

only serve as an upper bound on the actual value of $P_{72}(t)$. However, the rational behind truncating a node was that the "flow of probability" through the node was relatively small. By truncating node 7, the analyst has made a calculated decision that "not much" probability flows through the node and hence the event sequences associated with that node. Thus, equation 51 should be an accurate approximation to $P_{72}(t)$.

An exact solution for the probability of being at any state at time t in the network under node truncation could be determined and is straight forward.   The following steps are necessary:

1. For the given state, determine the event sequence(s) that have been removed by the node truncation.

2. Determine the probability integral for these event sequences.

3. Subtract this calculated integral from the appropriate, non-truncated integral (equation 50 or 52)

As an illustration, consider the three event sequences eliminated at node 72 by the truncation of node 7. The probability of each of these sequences at time t is:

$$
\begin{aligned}
P_{path\,5}(t) &= \int_0^t \left[\int_{e_{1_1}}^{e_{1_T}} \int_{e_{1_1}}^{e_{2_T}} f_{E_{2_2^{(0)}}}(e_{2_T} - e_{2_1}) f_{E_{2_1^{(0)}}}(e_{2_1}) de_{2_1} de_{2_T} \right] \\
&\quad \left[\int_{e_{1_T}}^{\infty} \int_0^{e_{1_1}} f_{E_{3_2^{(\kappa)}}}(e_{3_T} - e_{3_1}) f_{E_{3_1^{(0)}}}(e_{3_1}) de_{3_1} de_{3_T} \right] \\
&\quad \int_0^{e_{1_T}} f_{E_{1_2^{(1)}}}(e_{1_T} - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} de_{1_T}
\end{aligned} \tag{55}
$$

$$
\begin{aligned}
P_{path\,6}(t) &= \int_0^t \left[\int_{e_{1_T}}^{\infty} \int_0^{e_{1_1}} \left[\int_{e_{1_1}}^{e_{1_T}} \int_{e_{3_1}}^{e_{1_1}} f_{E_{2_2^{(0)}}}(e_{2_T} - e_{2_1}) f_{E_{2_1^{(0)}}}(e_{2_1}) de_{2_1} de_{2_T} \right] \right. \\
&\quad \left. f_{E_{3_2^{(\kappa)}}}(e_{3_T} - e_{3_1}) f_{E_{3_1^{(0)}}}(e_{3_1}) de_{3_1} de_{3_T} \right] \\
&\quad \int_0^{e_{1_T}} f_{E_{1_2^{(1)}}}(e_{1_T} - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} de_{1_T}
\end{aligned} \tag{56}
$$

$$
\begin{aligned}
P_{path\,12}(t) &= \int_0^t \left[\int_{e_{1_T}}^{\infty} \int_0^{e_{1_1}} \left[\int_{e_{3_1}}^{e_{1_1}} \int_{e_{3_1}}^{e_{2_T}} f_{E_{2_2^{(0)}}}(e_{2_T} - e_{2_1}) f_{E_{2_1^{(0)}}}(e_{2_1}) de_{2_1} de_{2_T} \right] \right. \\
&\quad \left. f_{E_{3_2^{(\kappa)}}}(e_{3_T} - e_{3_1}) f_{E_{3_1^{(0)}}}(e_{3_1}) de_{3_1} de_{3_T} \right] \\
&\quad \int_0^{e_{1_T}} f_{E_{1_2^{(1)}}}(e_{1_T} - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} de_{1_T}.
\end{aligned} \tag{57}
$$

Hence the probability of being in any of the three event sequences at time t is:

$$
\begin{aligned}
P_{path\,5+6+12}(t) &= P_{path\,5}(t) + P_{path\,6}(t) + P_{path\,12}(t) \\
&= \int_0^t \left[\int_{e_{1_T}}^{\infty} \int_0^{e_{1_1}} \left[\int_{e_{3_1}}^{e_{1_1}} \int_{e_{3_1}}^{e_{2_T}} f_{E_{2_2^{(0)}}}(e_{2_T} - e_{2_1}) f_{E_{2_1^{(0)}}}(e_{2_1}) de_{2_1} de_{2_T} \right] \right. \\
&\quad \left. f_{E_{3_2^{(\kappa)}}}(e_{3_T} - e_{3_1}) f_{E_{3_1^{(0)}}}(e_{3_1}) de_{3_1} de_{3_T} \right] \\
&\quad \int_0^{e_{1_T}} f_{E_{1_2^{(1)}}}(e_{1_T} - e_{1_1}) f_{E_{1_1^{(0)}}}(e_{1_1}) de_{1_1} de_{1_T}.
\end{aligned} \tag{58}
$$

Subtracting the integral in equation 58 from the integral in equation 51 yields:

$$
P_{72-truncated\,node\,7}(t) = P_{72}(t) - P_{path\,5+6+12}(t).
$$

4.8 Summary

In this chapter, a new network formulation termed an *event occurrence network* (EON) was introduced to model the interaction of groups of event sequences. An EON is a probabilistic network in which the arcs represent the occurrence of an event from a group of sequential events before events from other groups. Events between groups occur independently, but events within a group occur sequentially. A set of arcs leaving a node is a set of competing events, which are probabilistically resolved by the minimum type operators. EONs differ from the probabilistic networks previously discussion in chapter II and are motivated by the research of Ancker, Gafarian, Kress and several associates in modeling m versus n stochastic ground combat.. An analysis method termed "bucket" analysis is introduced to accommodate a node (state) explosion problem as the number of the events in a EON is increased. The general integral solution to an EON was discussed. To illustrate the capabilities of EONs, two simple air-to-air engagement examples are given.

V. Piecewise Polynomial Approximation

5.1 Introduction

As shown in chapter IV, an important metric for an Event Occurrence Network (EON) is the probability of being at a particular node or set of nodes at time t. Such a probability is formulated as an integral expression (possibly a multiple integral expression) involving event probability density functions (reference section 4.6). This integral expression involves three probabilistic operators:

- Multiplication,

- Convolution, and

- Conditional Integration.

For many event probability density functions (e.g., gamma with non-integer shape parameter) or combinations of event probability density functions (e.g. exponential and normal), these integral expressions have no closed form solution. In such cases, an approximation of the integral expression is needed to obtain a solution. Finding a "good" approximation is a nontrivial problem.

Appendix D, section D.1 provides a review of the theoretical construct of a norm as a measure of the accuracy of fit of an approximating function g to a function f. Specifically, section D.1 summarizes the approximation error function, e ($e = f - g$), for the widely used L_p norm. Unfortunately, such measurement theory can not be extended to EON probability integral expressions. In this case the approximation error function denoted e_I is the difference between the integral function denoted I and the approximation integral function denoted I_a, that is $e_I = I - I_a$. When there is no closed formed expression for I, the evaluation of the error function e_I can not be undertaken. Instead, some form of empirical test must be performed in order to measure the accuracy of the approximation.

5.2 Piecewise Polynomial Functions

An example of a piecewise polynomial function was given in section 2.4 (equation 12) in conjunction with Martin's solution [134] to reducing directed, acyclic networks with continuous random variable activity durations (i.e., probabilistic activity networks). Informally, a piecewise polynomial function is a function, $\rho(x)$, defined over a closed interval obtained by *dividing an interval into subintervals and constructing a different polynomial over each subinterval.*

More precisely, let the interval $[a, b]$ with knot sequence $a = x_1 < x_2 < \ldots < x_n = b$ or $\Delta = \{x_i\}_1^n$ be partitioned into $n - 1$ subintervals I_k such that

$$I_k = \begin{cases} [x_k, x_{k+1}) & \text{if } k = 1, 2, \ldots, n - 2 \\ [x_{n-1}, x_n] & \text{if } k = n - 1 \end{cases} \tag{59}$$

and let P^m be the linear space[1] of polynomials of degree m (i.e., $P^m = \{p(x) : p(x) = \sum_{j=0}^m c_j x^j, \ c_j \in R\}$). The *space of piecewise polynomial functions* is defined to be the linear space $(P_m, \mathbf{v}, \Delta)$ of functions $\rho(x)$ such that:

- There exist polynomials, $p_k \in P^m$ ($k = 1, 2, \ldots, n - 1$) such that $\rho(x) = p_k(x)$ for $x \in I_k$, $k = 1, 2, \ldots n - 1$; and

- $\frac{d^l p_i}{dx}(x_{i+1}) = \frac{d^l p_{i+1}}{dx}(x_{i+1})$ for $i = 1, 2, \ldots, n - 2$ where $l = -1, 0, \ldots, m - v_i$.

where the multiplicity vector $\mathbf{v} = (v_1, v_2, \ldots, v_{n-2})$ is defined such that $v \in Z$ and $1 \leq v_i \leq m + 1$.[2] [168]

---

[1]A linear space over a scalar field F is a nonempty set X that satisfies the following conditions:

1. There exists a mapping of $X \times X$ into X called addition and written $x_1 + x_2$, and

2. There exists a mapping of $F \times X$ into X called scalar multiplication and written αx.

Addition and scalar multiplication must satisfy several conditions. Typical conditions are mentioned in reference [142:page 161].

[2]The definition is a modified version of polynomial spline definition found in reference [168:page 108].

The vector $\mathbf{v}$ specifies the continuity and derivative (smoothing) constraints between the piecewise segments at knots. If $v_i = m + 1$ (i.e., $l = -1$), the two polynomial pieces p_i and p_{i+1} in the intervals adjoining the knot x_{i+1} are unrelated to each other (e.g., may be a jump discontinuity at x_{i+1}). If $v_i < m + 1$ (i.e., $l \geq 0$) then the two polynomial pieces have their first $m - v_i$ derivatives continuous across the knot x_{i+1}.

In most applications, the setting for the approximation problems is the closed interval $[a, b]$ as defined above. However, every piecewise polynomial has a natural extension to the real line $(-\infty, \infty)$. To accomplish this extension, two additional polynomials $p_0, p_n \in P^m$ must be defined such that:

$$
\rho(x) = \begin{cases} p_o(x) & x < a \\ p_k(x) & x \in I_k \\ p_n(x) & x > b \end{cases}
$$

In fact, this is the form of the piecewise polynomial function $g(x)$ given in section 2.4, equation 12. [168]

5.3   Piecewise Polynomial Approximation and Interpolation

In section D.2, polynomial interpolation and approximation is introduced. The main drawback of polynomials for approximation purposes is that the class is relatively inflexible. Polynomial approximation does an adequate job on sufficiently small intervals but, when larger intervals of approximation are required severe oscillations often appear - particularly if m (the degree of the polynomial) is more than 3 or 4. This observation suggests that in order to achieve a class of approximating functions with greater flexibility, work should be performed with polynomials of relatively low degree and the interval of interest should be divided into small subintervals. The definition of piecewise polynomial function shown in section 5.2 provides the foundation for this type of approximation.

Schumaker [168] proved that the finite, linear space $(P_m, \mathbf{v}, \Delta)$ has dimension $m + 1 + K$ where $K = \sum_{i=1}^{n-2} v_i$. A basis for $(P_m, \mathbf{v}, \Delta)$ can be defined with elements that have support over $m + 1$ subintervals (i.e., $m + 2$ knots). In order to define this basis, an extended partition denoted by $\tilde{\Delta}$ must be associated with $(P_m, \mathbf{v}, \Delta)$. The partition $\tilde{\Delta}$ is defined as:

$$\tilde{\Delta} = \{y_1 \leq y_2 \leq \ldots y_{2(m+1)+K}\} = \{y_i\}_{i=1}^{2(m+1)+K}$$

such that

$$y_1 \leq \ldots \leq y_{m+1} \leq a = x_1, \quad x_n = b \leq y_{m+2+K} \leq \ldots \leq y_{2(m+1)+K}$$

and

$$y_{m+2} \leq \ldots \leq y_{m+1+K} = \overbrace{x_2, \ldots, x_2}^{v_1}, \ldots, \overbrace{x_{n-1}, \ldots, x_{n-1}}^{v_{n-2}}.$$

The points $\{y_i\}_{m+2}^{m+1+K}$ in $\tilde{\Delta}$ are uniquely determined. The first and last $m + 1$ points in $\tilde{\Delta}$ can be chosen arbitrary. One popular choice is $y_1 = \ldots = y_{m+1} = a$ and $y_{m+2+K} = \ldots = y_{2(m+1)+K} = b$.

A basis for $(P_m, \mathbf{v}, \Delta)$ associated with the extended partition $\tilde{\Delta}$ is given by the set $\{B_i^m\}_{i=1}^{m+1+K}$ where:

$$B_i^m(x) = (-1)^{m+1}(y_{i+m+1} - y_i)[y_i, \ldots, y_{i+m+1}](x - y)_+^m, \quad a \leq x \leq b \qquad (60)$$

with $[\tau_i, \ldots, \tau_{i+r}]f$ representing the rth divided difference[3] of a function f (in this case, $f = (x - y)^m_+$) and $(x - y)^m_+ = (x - y)^m (x - y)^0_+$, $m > 0$ with

$$(x - y)^0_+ = \begin{cases} 0 & x < y \\ 1 & x \geq y \end{cases}.$$

The basis elements, B_i^m, in equation 60 are referred to in the literature as *B-splines* and have the following properties:

1. $B_i^m(x) = 0$ for $x \notin [y_i, y_{i+m+1}]$;

2. $B_i^m(x) > 0$ for $x \in (y_i, y_{i+m+1})$;

3. $\sum_{i=1}^{m+1+K} B_i^m(x) = 1$ for all $x \in [a, b]$; and

4. $\sum_{i=j-m}^{j} B_i^m(x) = 1$ for all $x \in [y_j, y_{j+1}]$.

Properties 1 and 2 verify that B-splines have positive support over $m + 1$ intervals. The term $(-1)^{m+1}$ in equation 60 provides the positive support attribute in Property 2. Properties 3 and 4 above show that B-splines form a partition of unity. The factor $(y_{i+m+1} - y_i)$ in equation 60 is a normalization factor designed to produce this partition of unity. [48] [168]

A point of clarification may be helpful to the reader at this time. The definition of the space of piecewise polynomial functions, $(P_m, \mathbf{v}, \Delta)$, given in the beginning of this section enabled *the smoothness constraints of the piecewise function $\rho(x)$ to be built into the bases of $\rho(x)$*. Such an approach reduces the approximating problem to *solely fulfilling the appropriate set of interpolation and/or approximation constraints.*

---

[3]Given any points $\tau_i, \ldots, \tau_{i+r}$ and a function f, the rth difference over the points $\tau_i, \ldots, \tau_{i+r}$ is:

$$[\tau_i, \ldots, \tau_{i+r}]f = \frac{[\tau_{i+1}, \ldots, \tau_{i+r}]f - [\tau_i, \ldots, \tau_{i+r-1}]f}{\tau_{i+r} - \tau_i}$$

where

$$[\tau_i, \tau_{i+1}]f = \frac{f(\tau_{i+1}) - f(\tau_i)}{\tau_{i+1} - \tau_i}.$$

For polynomial approximation, one popular criterion in the literature is the "least squares" criterion. Extending this criterion to piecewise polynomial approximation, the L_2 norm discussed in section D.2 (equations 129 and 130) is used to construct a unique approximating polynomial function h_k in each subinterval k, $k = 1, \ldots, n-1$. Specifically, consider the semi-normed linear space $((P_m, \mathbf{v}, \Delta), \|e\|_{2,\mathbf{x},w})$ defined on $R^{r_1+,\ldots,+r_{n-1}}$. Necessary conditions to minimize the total error function,

$$e_T = \sum_{k=1}^{n-1} f - h_k$$

for the weighted L_2 norm defined by equation 130 are:

$$
\begin{aligned}
&\min \quad \sum_{k=1}^{n-1} \|e\|_{2,\mathbf{R^r k},w_k} \\
&s.t. \quad \frac{d^l p_i}{dx}(x_{i+1}) = \frac{d^l p_{i+1}}{dx}(x_{i+1}) \quad i = 1, \ldots, n-2; \quad l = -1, 0, \ldots, m - v_i
\end{aligned}
\tag{61}
$$

Consider approximating functions $h_k(x)$ that have the form:

$$h_k(x) = \sum_{j=0}^{m} c_j^{(k)} \varphi_j^{(k)}(x) \qquad k = 1, \ldots, n-1 \tag{62}$$

Here h_k is a linear combination of functions $\varphi_0^{(1)}, \varphi_1^{(1)}, \ldots, \varphi_m^{(1)}, \ldots, \varphi_0^{(n-1)}, \varphi_1^{(n-1)}, \ldots, \varphi_m^{(n-1)}$. In this case, equation 61 may be written as:

$$\min E(\mathbf{c}, \lambda) = \mathbf{y}^{\mathbf{T}}\mathbf{W}\mathbf{y} - 2\mathbf{y}^{\mathbf{T}}\mathbf{W}\mathbf{A}\mathbf{c} + \mathbf{c}\mathbf{A}^{\mathbf{T}}\mathbf{W}\mathbf{A}\mathbf{c} + \lambda^{\mathbf{T}}\mathbf{R}\mathbf{c} \tag{63}$$

where the above vectors and matrices are defined as follows:

1.

$$
\mathbf{y} = \begin{bmatrix} \mathbf{y_1} \\ \vdots \\ \mathbf{y_{n-1}} \end{bmatrix} \quad \text{where } \mathbf{y_k} = \begin{bmatrix} f(x_{k1}) \\ \vdots \\ f(x_{kr_k}) \end{bmatrix}, \, k = 1, \ldots, n-1
$$

is a column vector of r_k observations on $f(x)$ in the kth interval. By definition, an observation $z_k \in R^{r_k}$ is defined such that:

$$z_k \in [x_k, x_{k+1}) \quad \text{for} \quad k = 1, \ldots, n - 2$$
$$z_k \in [x_k, x_{k+1}] \quad \text{for} \quad k = n - 1;$$

2. $\mathbf{W} = Diag\,[\mathbf{W_1}, \ldots, \mathbf{W_{n-1}}]$ is an r x r block diagonal matrix where $\mathbf{W_k}$ is an r_k x r_k matrix of the form:

$$\mathbf{W_k} = \begin{bmatrix} \mathbf{W}_{(\sum_{i=1}^{k-1} \mathbf{r_i}) + 1} & & & \\ & \ddots & & \mathbf{0} \\ \mathbf{0} & & \ddots & \\ & & & \mathbf{W}_{\sum_{i=1}^{k} \mathbf{r_i}} \end{bmatrix}, \quad k = 1, \ldots, n - 1;$$

3. $\mathbf{A} = Diag\,[\mathbf{A_1}, \ldots, \mathbf{A_{n-1}}]$ is an r x $(m+1)(n-1)$ block diagonal matrix where $\mathbf{A_k}$ is an r_k x $(m+1)$ matrix of the form:

$$\mathbf{A_k} = \begin{bmatrix} \varphi_0^{(k)}(x_{k1}) & \ldots & \varphi_m^{(k)}(x_{k1}) \\ \vdots & & \vdots \\ \varphi_0^{(k)}(x_{kr_k}) & \ldots & \varphi_m^{(k)}(x_{kr_k}) \end{bmatrix}, \quad k = 1, \ldots, n - 1;$$

4.

$$\mathbf{c} = \begin{bmatrix} \mathbf{c_1} \\ \vdots \\ \mathbf{c_{n-1}} \end{bmatrix} \quad \text{where} \quad \mathbf{c_k} = \begin{bmatrix} c_0^{(k)} \\ \vdots \\ c_m^{(k)} \end{bmatrix}, \quad k = 1, \ldots, n - 1$$

is a column vector of m approximating function coefficients;

5.

$$\mathbf{R} = \begin{bmatrix} \mathbf{R_1} & -\mathbf{R_{1'}} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{R_2} & -\mathbf{R_{2'}} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{R_{n-2}} & -\mathbf{R_{n-2'}} \end{bmatrix}$$

is a $\sum_{i=1}^{n-2}(m - v_i + 1)$ x $2(m+1)(n-2)$ matrix where $\mathbf{R_i}$ is an $(m - v_i + 1)$ x $(m+1)$ matrix of the form:

$$\mathbf{R_i} = \begin{bmatrix} \varphi_0^i(x_{i+1}) & \ldots & \varphi_m^i(x_{i+1}) \\ \frac{d}{dx}\varphi_0^i(x_{i+1}) & \ldots & \frac{d}{dx}\varphi_m^i(x_{i+1}) \\ \vdots & & \vdots \\ \frac{d^{m-v_i}}{dx}\varphi_0^i(x_{i+1}) & \ldots & \frac{d^{m-v_i}}{dx}\varphi_m^i(x_{i+1}) \end{bmatrix} \text{ and }$$

$\mathbf{R_{i'}}$ is an $(m - v_i + 1)$ x $(m+1)$ matrix of the form:

$$\mathbf{R_{i'}} = \begin{bmatrix} \varphi_0^{i+1}(x_{i+1}) & \ldots & \varphi_m^{i+1}(x_{i+1}) \\ \frac{d}{dx}\varphi_0^{i+1}(x_{i+1}) & \ldots & \frac{d}{dx}\varphi_m^{i+1}(x_{i+1}) \\ \vdots & & \vdots \\ \frac{d^{m-v_i}}{dx}\varphi_0^{i+1}(x_{i+1}) & \ldots & \frac{d^{m-v_i}}{dx}\varphi_m^{i+1}(x_{i+1}) \end{bmatrix}.$$

In equation 63 the modified error function, $E(\mathbf{c}, \lambda)$, is constructed as the sum

$$e_T + \lambda^{\mathbf{T}}\mathbf{Rc}$$

where λ represents a r column vector of Lagrange multipliers. Necessary conditions to minimize $E(\mathbf{c}, \lambda)$ are:

$$\frac{\partial E}{\partial \mathbf{c}} = \mathbf{0} \quad \frac{\partial E}{\partial \lambda} = \mathbf{0} \ .$$

Hence,

$$\frac{\partial E}{\partial \mathbf{c}} = -2\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{y} + 2\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A}\mathbf{c} + \mathbf{R}^\mathbf{T}\lambda = \mathbf{0} \qquad (64)$$

$$\frac{\partial E}{\partial \lambda} = \mathbf{R}\mathbf{c} = \mathbf{0} \qquad (65)$$

Premultiplying both sides of equation 64 by $\mathbf{R}(\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}$:

$$-2\mathbf{R}(\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{y} + 2\mathbf{R}\mathbf{c} + \mathbf{R}(\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}\mathbf{R}^\mathbf{T}\lambda = \mathbf{0}$$

Using equation 65 and solving for λ:

$$\lambda = 2[\mathbf{R}(\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}\mathbf{R}^\mathbf{T}]^{-1}(\mathbf{R}(\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{y})$$

Using this expression for λ and solving for $\mathbf{c}$:

$$\begin{aligned}
\mathbf{c} &= (\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{y} + (\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}\mathbf{2}[\mathbf{R}(\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}\mathbf{R}^\mathbf{T}]^{-1} \\
&\quad (\mathbf{R}(\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{y}) \qquad (66)
\end{aligned}$$

The expression $(\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{y}$ in equation 66 above represents the unconstrained contribution to the coefficient vector - when the multiplicity vector $\mathbf{v} = (m-1,\ldots,m-1)$ - developed in Appendix D, equation 137.

As discussed in Appendix D for large m, the matrix $\mathbf{A}^\mathbf{T}\mathbf{W}\mathbf{A}$ can be ill-conditioned making its inverse difficult to solve numerically. One way of reducing the effect of ill-conditioning is to introduce orthogonality into the matrix. For equation 66, this can be accomplished by requiring orthogonality between functions $\varphi_j^{(k)}(x)$ within each piecewise segment $h_k(x)$. Specifically, for approximating func-

tions h_k of the form in equation 62, the orthogonality requirements are:

$$\langle \varphi_i^{(k)}, \varphi_j^{(k)} \rangle = 0 \quad \text{for } i \neq j \text{ and}$$

$$\langle \varphi_i^{(k)}, \varphi_i^{(k)} \rangle = 0 \quad \text{for all } i$$

for all $k = 1, \ldots, n - 1$. With these requirements, the matrix $\mathbf{A^T W A}$ reduces to a diagonal matrix $\mathbf{B}$ of the form $Diag[\mathbf{B_1}, \ldots, \mathbf{B_{n-1}}]$ where:

$$\mathbf{B_k} = \begin{bmatrix} \langle \varphi_0^{(\mathbf{k})}, \varphi_0^{(\mathbf{k})} \rangle & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \langle \varphi_\mathbf{m}^{(\mathbf{k})}, \varphi_\mathbf{m}^{(\mathbf{k})} \rangle \end{bmatrix} \quad \varphi_\mathbf{j}^{(\mathbf{k})} = \begin{bmatrix} \varphi_j^{(k)}(x_{k1}) \\ \vdots \\ \varphi_j^{(k)}(x_{kr_i}) \end{bmatrix}, \quad k = 1, \ldots, n - 1.$$

Hence requiring orthogonality between functions $\varphi_j^{(k)}(x)$ within each piecewise segment $h_k(x)$ reduces equation 66 to:

$$\mathbf{c = B^{-1}A^T W y + B^{-1}[2[RB^{-1}R^T]^{-1}(RB^{-1}A^T W y)]} \qquad (67)$$

When the approximating function $h(x)$ is of the form in equation 62, a total of $(m + 1)(n - 1)$ values ($c_j^{(k)}$ coefficients) must be stored to represent $h(x)$. For example, suppose $h(x)$ is a piecewise monic polynomial approximation -

$$\begin{aligned} \varphi_0^{(1)} =, \ldots, = \varphi_0^{(n-1)} &= 1 \\ \varphi_1^{(1)} =, \ldots, = \varphi_1^{(n-1)} &= x \\ \vdots \qquad\qquad &\quad \vdots \\ \varphi_m^{(1)} =, \ldots, = \varphi_m^{(n-1)} &= x^m \end{aligned}$$

Recall that the space of piecewise polynomial functions $(P_m, \mathbf{v}, \Delta)$ has dimension $m + 1 + K$ - the continuity restrictions transform the $(m + 1)(n - 1)$ dimensioned vector space $R^{(m+1)(n-1)}$ into the $m + 1 + K$ dimensioned vector space $R^{\sum_{i=1}^{n-2}(m-m_i)}$. A basis for $(P_m, \mathbf{v}, \Delta)$ was introduced in equation 60. This B-spline representation

may be stored in $m + 1 + K$ values. Hence, in general, the piecewise approximating function representation in equation 62 is bulkier than the B-Spline representation. However, when the approximation has to be evaluated at a large number of points as with the L_2 norm case, it is advantageous to use the approximating function representation in equation 62. For example, to evaluate the approximating function at any point in the domain requires only $m(m + 1)$ operations for the piecewise approximation function in equation 62 as compared to the $3m(m + 1)/2$ operations required by the most efficient B-spline expansions.

As discussed in Appendix D, an interpolating polynomial is a polynomial approximation function in which the polynomial approximation function matches exactly a finite number of points from the original function. Interpolating polynomials provide adequate approximations over small intervals, but when applied to larger intervals severe oscillations often appear - particularly with $m \geq 3$. To avoid these problems, an alternative interpolation approach divides the interval of approximation into smaller intervals and uses a different (interpolating) polynomial over each interval. Referred to in the literature as "interpolation with piecewise polynomial functions", an interpolating piecewise polynomial function is a piecewise polynomial function $h(x)$ in which the knot sequence is formed from the n interpolation points. [71] [168]

One set of interpolating piecewise polynomial functions that has received much discussion in the literature are *spline* (interpolating) functions. Let Δ be a partion of the interval $[a, b]$ as in definition 59 and let m be a positive integer, the space of spline (interpolating) functions, S_m, is a subspace of the space of piecewise polynomials $(P_m, \mathbf{v}, \Delta)$ such that:

- $f(x_k) = p_k(x_k)$ for $k = 1, \ldots, n - 1$ and $f(x_n) = p_{n-1}(x_n)$;

- $\mathbf{v} = (1, \ldots, 1)$; and

- $\exists\, m - 1$ boundary conditions.

As an example of boundary conditions, consider cubic interpolating splines $(m = 3)$. With cubic splines, one of the following sets of boundary conditions is generally specified:

- $\frac{d^2 p_1}{dx}(x_1) = \frac{d^2 p_{n-1}}{dx}(x_n) = 0$ (natural or free boundary); or

- $\frac{dp_1}{dx}(x_1) = \frac{df}{dx}(x_1)$ and $\frac{dp_{n-1}}{dx}(x_n) = \frac{df}{dx}(x_n)$ (clamped boundary).

Earlier it was shown that B-splines form a basis for $(P_m, \mathbf{v}, \Delta)$. Since S_m is a subspace of $(P_m, \mathbf{v}, \Delta)$, B-splines also form a basis for S_m. Hence every spline function $\rho(x) \in S_m$ has a unique representation as:

$$\rho(x) = \sum_{i=1}^{m+n-1} c_i B_i^m(x) \text{ for } x \in [a, b] \tag{68}$$

Given the set of n interpolating points and the representation of $\rho(x)$ in equation 68, the following system of equations for an approximating spline function, $h(x)$, is delineated:

$$h(x_k) = \sum_{i=1}^{m+n-1} c_i B_i^m(x_k) = f(x_k), \quad k = 1, \ldots, n \tag{69}$$

The system of equations in 69 can be represented in matrix form as:

$$\mathbf{Bc} = \mathbf{y}$$

where the above vectors and matrices are defined as follows:

$$\mathbf{B} = \begin{bmatrix} B_1^m(x_1) & \ldots & B_{m+n+1}^m(x_1) \\ \vdots & & \vdots \\ B_1^m(x_n) & \ldots & B_{m+n+1}^m(x_n) \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_{m+n+1} \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}$$

A unique solution for $h(x)$ requires an additional $m - 1$ equations. As mentioned above, these equations are formed from $m - 1$ boundary assumptions. Additionally, the matrix $\mathbf{B}$ is a banded matrix as each B-spline, B_i^m, has non-zero support over the interval $[x_i, x_{i+m+1}]$.

As a final note, piecewise polynomial approximation and spline interpolation can also be viewed from a regression prospectus. "Piecewise" (polynomial) regression literature is divided into two main areas: piecewise regression and spline regression. [170] Specifically, consider the relationship between y and x:

$$y = f(x; \Theta) + \varepsilon \tag{70}$$

obtained by piecing together different functions (curves) over different intervals, i.e.,

$$f(x; \Theta) = \begin{cases} f_1(x; \Theta_1) & x < \alpha_1 \\ f_2(x; \Theta_2) & \alpha_1 \leq x < \alpha_2 \\ \vdots & \vdots \\ f_n(x; \Theta_n) & \alpha_{n-1} \leq x \end{cases} \tag{71}$$

In piecewise regression, the n functions are referred to as phase models or regimes, and the interval points α as change points or join points. In spline regression, the individual phase models ($f_i(x; \Theta_i)$, $i = 1, 2, \ldots, n$) are polynomials and more stringent conditions are imposed on these models at the changepoints (α). Poirier [152] developed a piecewise regression scheme using cubic splines. Buse and Lim [26] showed that cubic spline regression scheme developed by Poirier is a special case of restricted least-squares estimation.

5.4   *Past Use in Density Function Estimation and Stochastic Operations*

Cleroux and McConaloque [34] first applied cubic spline interpolation to convolution of random variables which arise in reliability problems. The authors used splines in their application because maintaining the derivative of the probability density functions across spline segments that make up the interpolation was considered important.

Clerox and McConalogue [34] [135] developed an algorithm using the Lobatto formula (a variation of Gaussian quadature) and a spline method to approximate the n-fold convolution of distribution functions. In order for the algorithm to be applicable, the probability density function has to be analytic (i.e., continuous first derivative) on the interval $[0, \infty)$. The algorithm was tested on the exponential, the truncated normal, gamma, and Weibull probability distributions. Its application to the gamma and Weibull distributions is restricted to the case of the shape parameter $\alpha > 1$.

McConalogue [136] generalized the algorithm to a subclass of singular probability density functions having finite singularities at the origin and infinity. This work made it possible to approximate the n-fold convolution of the gamma and Weibull distributions with shape parameters α as small as $\frac{1}{2}$. Unfortunately, implementation of this generalized algorithm is not straight-forward.

Mahjooli [132] used piecewise polynomial interpolation to approximate the n-fold convolution of probability density functions which are analytic over a bounded interval. Specifically, Mahjooli developed two approximation schemes using cubic spline functions in the form of B-Splines over a uniform partition of the bounded interval. In the first scheme, all six non-symmetric B-splines are discarded and only the symmetric normalized B-splines are considered. In the second scheme only four non-symmetric B-splines are eliminated and the two non-symmetric B-splines with largest support are included.

Santoro [166] used normalized B-splines to represent any probability density function with finite bounds. Normalized B splines, denoted M_i^m, are based on a scheme introduced by Curry and Schoenberg that produces a related basis for $(P_m, \mathbf{v}, \Delta)$ where:

$$M_i^m(x) = \frac{m}{y_{i+m+1} - y_i} B_i^m(x) \quad a \leq x \leq b \tag{72}$$

The basis elements, M_i^m, in equation 72 have the additional property [48] [166] that:

$$\int_a^b M_i^m(x)dx = \int_{y_i}^{y_{i+m+1}} M_i^m(x)dx = 1.$$

In contrast to the above interpolation schemes, Fergeson and Shortell [73] recommended using simple first degree polynomials that were fitted piecewise using least squares regression over ten equally space subintervals to approximate maximum and convolution stochastic operators. This equates to ten linear regressions over the bounded domain of the density function. Fergeson and Shortell recommended using $10 + I(classwidth * 3)$ regression fitting points positioned uniformly across each class. Here *classwidth* refers to the length of a subinterval. Additionally, in order to control error build-up from the polygonal approximation, after the ten sets of regression coefficients have been computed, the coefficients are normalized so that the probability under the approximated density function is one.

Fergeson an Shortell [73] termed their method the Polynomial Approximtion and Reduction Technique (PART). The authors used this technique with uniform, normal, and exponential distribution functions. Fergeson and Shortell noted that the PART experiences it greatest (approximating) error at the peak of a distribution (e.g., approximating the normal density function or approximating the convolution of two normal or exponential distributions). The authors recommended as further research adopting a multiple linear regression approach instead of the simple linear regression approach or increasing the number of classes upwards from ten. Additionally, at class boundaries, two approximations of a density function exist, corresponding to the approximate line segments of the class which terminates and the class which begins at that class boundary.

Lawrence [128] [127] adopted PART for use with Dodin's arc (node) and sequential approximation reduction processes. Lawrence also choose a first degree or simple linear regression. Lawrence expanded Fergeson and Shortell's technique to

include approximation of the maximum and convolution operators for the triangular, gamma, and beta distributions. Lawrence documented that the PART algorithms were orders of magnitude faster than simulation-approximations without significant losses in accuracy when simulation results are taken as "true".

In contrast to Lawrence, Badinelli [11] used orthogonal polynomials to piecewise approximate the probability density functions of continuous random variables and the convolution of random variables for inventory models. Badinelli looked at five classical orthogonal polynomial approximations:

1. Tchebyshev polynomials of the first type;

2. Tchebyshev polynomials of the second type;

3. Legendre polynomials;

4. Laguerre polynomials; and

5. Hermite polynomials.

Badinelli [11] recommended that, if the density function of a random variable is decreasing or increasing over a wide range of its domain and/or changing shape, then a more efficient method of approximation was to divide the domain into segments and perform a separate approximation over each segment (i.e., piecewise approximation) rather than fitting a single approximation over the entire domain. As he stated, in this manner, "such a piecewise approximation method can achieve a very close approximation with three or four polynomial approximations of order 3 or 4 instead of a single polynomial approximation of order 7 or 8". Badinelli [11] commented that with orthogonal polynomials, unlike interpolation methods such as splines and Lagrange polynomials, one can incorporate into the approximation as much or as little data as needed independently of choosing the number of segments in the piecewise approximation, the order of the polynomial approximation on each segment, and the boundaries of each segment.

For bounded domains, Badinelli [11] set upper and lower bounds of the piece-wise approximation of the density functions of the random variables based on the bounded values of the random variable. For unbounded domains, Badinelli set the bounds of the piecewise approximation at "points beyond which the probability function had negligble magnitude". Badinelli states that in most applications such points exist and can be conservatively estimated using the Tchebyshev theorem for upper bounds on tail probabilities and then tighted as density function values are computed and found to be too small to be included in the approximation. Once the upper and lower bounds of the approximation have been established, the interval between them is divided into N adjoining segments (Badinelli does not give a criterion for selecting N.)

Badinelli [11] choose two measures for accuracy of fit:

- the average of the absolute value of the difference between true complementary cumulative distribution function, F_c and the fitted complementary cumulative distribution function Φ expressed as a fraction of F_c and

- the average of the absolute value of the difference between true complementary cumulative distribution function of F, G and the fitted complementary cumulative distribution function of F, α expressed as a fraction of G.

Specifically, the measures are:

$$\overline{\mid error_{F_c} \mid} = \int_{x_l}^{x_u} \frac{\mid F_c(t) - \Phi(t) \mid}{F_c(t)} dt \tag{73}$$

$$\overline{\mid error_G \mid} = \int_{x_l}^{x_u} \frac{\mid G(t) - \alpha(t) \mid}{G(t)} dt \tag{74}$$

where

- $F_c(x) = \int_x^\infty f(t)dt$;

- $G(x) = \int_x^\infty F(t)dt = \int_x^\infty \int_t^\infty f(t)dtds$;

128

- f is the density function of the continuous random variable;

- $\Phi(a) = \int_a^\infty \phi(\tau)d\tau$;

- $\alpha(a) = \int_a^\infty \Phi(\tau)d\tau = \int_x^\infty \int_\sigma^\infty \phi(\tau)d\tau d\sigma$; and

- ϕ is the piecewise orthogonal polynomial that approximates f.

Badinelli [11] stated that no theoretical results on the optimal choice of the five classical, orthogonal polynomial types to minimize equations 73 and 74 exist in the literature. The only theoretical result is for the accuracy of fit of ϕ to f (reference section D.1). As a result, Badinelli used empirical results for equations 73 and 74 to indicate the best choice of polynomial type and the best order of the approximation when applied to a given density function. In most of the empirical tests, Badinelli used a normal density function since routines for computing integrated values of the normal, accurate to the fifth decimal place, were available. Additionally, Badinelli assumed that a procedure yielding an accurate approximation to various segments of the normal density function would perform well on other density function such as the gamma and beta functions which were also tested.

Badinelli drew the following conclusions from the empirical tests:

- For normal, gamma, and beta density functions, the *Tchebyshev polynomials of the second type* obtained the best approximation in terms of the accuracy of fit criteria.

- For every polynomial type there is an optimal power of approximation. When the order of the polynomial is too small, the approximation cannot take on all of the complexity in shape of the density function. When the order of the polynomial is too high, the approximation is volatile over the interval.

- Increasing the number of points used in the approximation procedure enhances the accuracy of the approximation in a *given segment*. Badinelli suggested 50 data points as typically adequate based on his empirical tests.

129

- Error terms from an orthogonal polynomial approximation show symmetry. This has the advantageous effect of canceling positive and negative errors in the fitted density function under integration, making *"the fit of the integral more accurate than the fit of the integrand [i.e., density function] "*.

- The accuracy of the approximation of some of the five orthogonal polynomial types deteriorates at the extreme values of a segment since the weighting functions decrease as one approaches these values. However, the errors at the "fringes" of a segment or domain can be effectively truncated by fitting the polynomial to a wider interval than the segment or domain over which the approximation is to be used.

- Accuracy of the convolution on normal density functions did not significantly deteriorate as the number of convolutions increased. Badinelli used convolutions from 2 to 10.

Badinelli took the number of data points and the order of the polynomial approximation as proxy measures of CPU time and stated that he kept them at the smallest values possible while achieving desired accuracy. However, he did not state these measures quantitatively.

Kennedy [118] used non-classical, orthogonal polynomials to approximate the density functions of a random variable based on sample moments of the density function (the actual underlying density function is assumed to be unknown). Kennedy's method first established approximate density functions based on lower order moments. These density functions were then refined by orthogonal polynomial series expansions based on higher order moments. Kennedy's method relied on the important result that a probability distribution with a bounded domain can be uniquely determined from its moments. [181] Specifically, the second, third and fourth moments about the mean provide information about the spread, skewness, and peakedness of the distribution function while higher order moments contain an increasing amount of information about the tails of the distribution.

Kennedy [118] mentioned four methods for approximating the density function from sample moment data:

1. Pearson Distributions;

2. Series Expansion;

3. Transformation to Normality; and

4. Principle of Maximum Entropy.

Kennedy stated that classical orthogonal polynomials form the basis for the majority of series expansions. As mentioned in section D.2, Kennedy observed that classical orthogonal polynomials are extremely useful since they are inexpensive to compute and possess the minimum least squares property. However, Kennedy also noted that there are some limitations to use of classical orthogonal polynomials to approximate functions:

1. To obtain a very high degree of accuracy an excessive polynomial order may be required and

2. Convergence is not guaranteed to be uniform unless particular types of polynomials are used (e.g. polynomials obtained after the application of the Bernstein operator)[4] [153]

Kennedy [118] stated that these potential limitations may be overcome by choosing a weighting function that resembles the specific function being approximated. In section D.2, it was mentioned that non-classical orthogonal polynomials may be generated using the Gram-Schmidt orthogonalization process *with any positive weighting function*. For the approximating density functions, Kennedy recommended selecting weighting functions of similar shape to the density functions. Since the density functions are not know but the sample moments of the density function

---

[4]In other words, the nth order polynomial approximation to a function may not necessarily be better than the $n-1$th order approximation.

are known, Kennedy proposed using the extreme value, Pearson type I, and Johnson transform distribution as the weighting functions for the polynomial approximation based on first three or four sample moments. These weighting functions are then used in an orthogonal series expansion with the higher order moments. In this manner, Kennedy stated an orthogonal polynomial series expansion can be obtained which is stable at high order (e.g. for 6th to 10th) allowing accurate approximation of tail probabilities.

Kennedy [118] applied his method to sum, difference, product, and quotient operations on random variables. Kennedy showed that the moments of the output of these random variable operations can be expressed exactly in terms of the moments of the input variables. The density function of the output of the operations can then be approximated using the non-classical orthogonal polynomial method described above.

5.5 Discussion

The material presented in this chapter has been focused on providing the necessary notation and background to rationally discuss the piecewise polynomial approximation technique(s) to use for solving the multiple integrals present in Event Occurrence Network (EON) formulations. The author has organized this discussion in two focus areas:

- Approximation power of least squares versus interpolation schemes for stochastic operations and

- Advantages/disadvantages of continuity requirements between piecewise polynomial segments.

As can be seen in literature review of piecewise polynomial approximation techniques (section 5.4), the research trend has moved from interpolation approaches towards least squares approaches. The research of the approximation power of least

squares versus interpolation schemes for stochastic operations is scant. Unser and Daubechies [187] investigated the approximation power as a function of the sampling step h (i.e., spacing between sample points) for (discrete) convolution-based signal processing algorithms. The authors'analysis summarized that, for larger values of h, the least squares approximation behaved like an interpolation with twice the order. For small values of h, both the least squares and interpolation approximation exhibit general $O(h^L)$ behavior of the error[5], but the least squares error is usually smaller (by a known proportion) than the interpolation error.

In terms of the flexibility in approximation, the least squares schemes have a decisive advantage over interpolation schemes. With least squares techniques, as Badinelli [11] points out, one can incorporate as much or as little data as needed independently of choosing the number of segments in the piecewise approximation or the order of the polynomial approximation on each segment. With the least squares techniques, the numerical integration in EON formulations can be accomplished with as few points and as small a polynomial order as necessary on each piecewise segment to achieve desired accuracy.

In terms of the second focus area, continuity requirements between piecewise polynomial segments, the literature is again split between least squares and inter- polation schemes. The interpolations schemes generally involve the use of spline functions, which by definition are a subspace of $C^{m-1}[a,b]$, while the least squares schemes generally do not specify continuity requirements. However, it was shown in section 5.3 that continuity requirements can also be easily incorporated into least squares schemes. The literature is quite uninstructive on when to impose/not im- pose continuity requirements. For example, Cleroux and McConaloque [34] consider continuity across segments to be important in approximating the convolution of random variables while Badinelli [11] considers continuity not to be important for

---

[5]An approximation procedure has an Lth order of approximation if it reproduce all polynomials of degree m = L-1.

approximating convolutions. In fact, neither of the authors state the rationale for their recommendations.

It is readily apparent that continuity requirements add additional computational burden to piecewise polynomial approximations. Specifically, each continuity requirement between two segments adds an additional constraint (equation) to the approximation problem. However, the *most convincing arguments against specifying continuity requirements between piecewise polynomial segments are the results achieved without their use*. Neither Fergueson and Shortell, Lawerence, or Badnelli maintained continuity between segments and achieved adequate results for stochastic operators.

Based on the above discussion, this dissertation pursues least square piecewise polynomial approaches *without continuity requirements across segments*. Specifically, the two approaches taken are a generalized version of the Fergueson/Shortell/Lawrence PART algorithm and a variant of the PART algorithm using Badinelli's research on piecewise orthogonal polynomial functions.

Fergueson/Shortell/Lawrence used simple piecewise polynomial segments of degree one in the PART algorithm. This approach approximated maximum and convolution operators well for random variables having linear probability density functions such as the uniform or triangular distributions. However, this approach did not work as well for nonlinear probability density functions such as the normal or exponential distributions. Fergueson and Shortell [73] recommended investigation of piecewise polynomials segments of degree greater than one, but Lawrence [128] [127] did not pursue this recommendation in his research. A generalized PART algorithm uses a least squares approximation with piecewise polynomial functions of degee $m \geq 1$. Using the notation in equation 66, the coefficients of the piecewise polynomial functions are:

$$\mathbf{c} = (\mathbf{A^T A})^{-1} \mathbf{A^T y}. \tag{75}$$

Note that the PART algorithm eliminates the weight matrix $\mathbf{W}$ and continuity component (involving the matrix $\mathbf{R}$) in equation 66. As discussed previously, the main problem with this approach is that for large m, the matrix $\mathbf{A^T A}$ can be ill-conditioned, making its inverse difficult to solve numerically. In fact, in implementing the generalized PART algorithm in the EON solution, ill-conditioning of the matrix becomes a severe problem in the tails of distributions. This problem forced a switch of inversion engines from LU decomposition to singular value decomposition (SVD).

To eliminate potential conditioning problems with inverting $\mathbf{A^T A}$, orthogonal polynomials may be used. With this approach, $\mathbf{A^T A}$ reduces to a diagonal matrix $\mathbf{B}$ of the form $Diag[\mathbf{B_1}, \ldots, \mathbf{B_{n-1}}]$ where:

$$\mathbf{B_k} = \begin{bmatrix} \langle \varphi_{\mathbf{0}}^{\mathbf{(k)}}, \varphi_{\mathbf{0}}^{\mathbf{(k)}} \rangle & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \langle \varphi_{\mathbf{m}}^{\mathbf{(k)}}, \varphi_{\mathbf{m}}^{\mathbf{(k)}} \rangle \end{bmatrix} \quad \varphi_{\mathbf{j}}^{\mathbf{(k)}} = \begin{bmatrix} \varphi_j^{(k)}(x_{i1}) \\ \vdots \\ \varphi_j^{(k)}(x_{ir_i}) \end{bmatrix}, \quad k = 1, \ldots, n-1.$$

The coefficients of the piecewise orthogonal polynomials using the PART algorithm are:

$$\mathbf{c} = (\mathbf{B})^{-1} \mathbf{A^T y}. \tag{76}$$

This approach differs from Badinelli's in several respects. First, and most significantly, Badnelli did not use a reduction step with his algorithm - he was left with the problems of 'exploding coefficients' and 'proliferating classes'. Secondly, Badinelli used polynomials satisfying a *continuous* orthogonality relation which allows the relation $\langle \varphi_i^{(k)}, \varphi_j^{(k)} \rangle = 0$ ($i \neq j$) to hold for any point x in the subinterval $[x_k, x_{k+1}]$. For discrete applications such as the PART algorithm, the above orthogonality condition holds only when x is one of the zeros of $\varphi_j^{(k)}$. To overcome this restriction on x a different orthogonality scheme must be used. Hayes [94] [169] suggested using the three-term recurrence relationship:

$$\varphi_{j+1}^{(k)}(x) = 2(x - a_{j+1}^{(k)})\varphi_j^{(k)}(x) - b_j^{(k)}\varphi_{j-1}^{(k)}(x)$$

$$, k = 1, \ldots, n - 1; j = 2, \ldots, m \quad (77)$$

beginning with initial polynomials

$$\varphi_0^{(k)}(x) = 1 \quad \text{and} \quad \varphi_1^{(k)}(x) = 2(x - a_1)$$

where $x \in [-1, 1]$ and a_{j+1} and b_j are chosen to the make the (discrete) orthogonal relation hold

$$a_{j+1} \quad = \quad \frac{\langle \varphi_{\mathbf{j}}^{\mathbf{k}}, \mathbf{x}\varphi_{\mathbf{j}}^{\mathbf{k}} \rangle}{\langle \varphi_{\mathbf{j}}^{\mathbf{k}}, \varphi_{\mathbf{j}}^{\mathbf{k}} \rangle} \quad (78)$$

$$b_j \quad = \quad \frac{\langle \varphi_{\mathbf{j}}^{\mathbf{k}}, \varphi_{\mathbf{j}}^{\mathbf{k}} \rangle}{\langle \varphi_{\mathbf{j-1}}^{\mathbf{k}}, \varphi_{\mathbf{j-1}}^{\mathbf{k}} \rangle} \quad (79)$$

with $b_0^k = 0$ and $a_1^k = \bar{\mathbf{x}}$. Hayes's method is similar to the Gram-Schmidt orthogonalization process discussed in Appendix D, section D.2, with the factor unity instead of the factor 2 in equation 77. The orthogonal polynomials generated by Hayes's method may be expressed in terms of classical orthogonal polynomials such as Tchebyshev polynomials via transformation [169]. However, such a transformation increases computing time.

Lastly, Badnelli used weight functions and expanded intervals in his approximation. These two techniques provide greater power of approximation but must be used with apriori knowledge. They are not well suited to a generalized algorithm. Based on Badnelli's research, the dissertation employs orthogonal polynomials generated using Hayes's method in the PART algorithm. At this time, it needs to be highlighted that although the use of piecewise orthogonal polynomials eliminates conditioning problems with the matrix $\mathbf{A}^{\mathbf{T}}\mathbf{A}$, it introduces the problem of maintaining the orthogonality of the polynomials during the reduction process.

5.6 Summary

In this chapter, a formal definition of a piecewise polynomial function was given. This definition was followed by an overview of piecewise polynomial approximation techniques of least squares and interpolation. Next, a literature review summarizing the research into approximating stochastic operators with piecewise polynomial functions was given. This literature review lead to a discussion of the most appropriate technique(s) for solving the multiple integrals present in Event Occurrence Network formulations. Two approaches were forwarded for further study, a generalized version of the PART algorithm and a variant of the PART algorithm using piecewise orthogonal polynomial functions.

VI. EON Solution Using Piecewise Polynomial Approximation

6.1 Introduction

In section 4.6.3, equations 50 and 52 delineate respectively the probability of being in any absorbing and transient state of an Event Occurrence Network (EON) at time t. These multiple integral probability expressions consist of several stochastic operations. In equation 50,

$$P_{absorb}(t) = \int_0^t \prod_{\substack{j=1 \\ j \neq g}}^{s} P(E_{j:T^{(\iota_j)}} \, \alpha_j \, e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T}) f_{E_{g:T^{(1)}}}(e_{g:T}) de_{g:T},$$

the integral expression $\int_0^t de_{g:T}$ is a conditional integration operation with multiplication operation $[\prod_{\substack{j=1 \\ j \neq g}}^{s} P(E_{j:T^{(\iota_j)}} \, \alpha_j \, e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})][f_{E_{g:T^{(1)}}}(e_{g:T})]$. The subexpression $\prod_{\substack{j=1 \\ j \neq g}}^{s} P(E_{j:T^{(\iota_j)}} \, \alpha_j \, e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$ consists of $s-1$ multiplication operations. Additionally, the subexpression $P(E_{j:T^{(\iota_j)}} \, \alpha_j \, e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$ consists of the following stochastic operations:

- if $P(E_{j_1^{(\kappa)}} > e_{g:T})$ - the first event in event group j has not occurred before the first T_g events in event group g, so an additional integration operation is required;

- if $P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$ - the first T_j events in event group j have not occurred before the first T_g events in event group g given that the first $T-1_j$ events in event group j have occurred before the first T_g events in event group g, so $T_j - 1$ convolution operations, two integrations, and a subtraction operation are required; or

- if $P(E_{j_{n_j}^{(0)}} < e_{g:T})$ - all events in event group j have occurred before the first T_g events in event group g, so $n_j - 1$ convolution operations and an integration operation are required.

Finally, the subexpression $f_{E_{g:T^{(1)}}}(e_{g:T})$ consists of $T-1$ convolution operations. Similarly for equation 52,

$$P_{trans}(t) = \prod_{j=1}^{s} P(E_{j:T^{(\iota_j)}} \, \alpha_j \, t \mid E_{j:T-1^{(0)}} < t),$$

the expression consists of s multiplication operations. Additionally, the subexpression $P(E_{j:T^{(\iota_j)}} \, \alpha_j \, t \mid E_{j:T-1^{(0)}} < t))$ consists of stochastic operations similar to those for the absorbing state subexpression $P(E_{j:T^{(\iota_j)}} \, \alpha_j \, e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$ above.

In this chapter, the techniques/algorithms for approximating these stochastic operations and solving EONs are given. The FORTRAN95 computer program that solves for the probability of being at any node of an EON at time t is over $8,000$ lines long - much too large to include in this dissertation. Although the algorithms used to produce the EON solution are delineated in this chapter, readers desiring the algorithms at the code level should contact the author. The chapter contains two air-to-air engagement EON examples. One of the networks is the simple air-to-air engagement example created by Hong, et al. [103] shown in Figure 17 and the other is a more complex engagement involving over 60 event distributions. Additionally, Appendix E contains interval data and polynomial coefficients for all the examples given in this chapter.

6.2   *Probability Density Function Approximation*

The first step in solving for the probability of being at a node a of an EON at time t requires inputting each event occurrence density function, $f_{E_{ge}}(t)$, as a piecewise polynomial function. Specifically, let the interval $[a, b]$ be partitioned into $n-1$ subintervals I_k such that

$$I_k = \begin{cases} [x_k, x_{k+1}) & \text{if } k = 1, 2, \ldots, n-2 \\ [x_{n-1}, x_n] & \text{if } k = n-1 \end{cases}$$

and a piecewise polynomial function of degree m, $h_k(t)$:

$$h_k(t) = \sum_{j=0}^{m} c_j^{(k)} \varphi_j^{(k)}(t), \quad c_j^k \in R$$

approximate the event occurrence density function in each subinterval using the L_2 (least squares) norm.

Assuming no continuity between segments,

$$\mathbf{c} = (\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{y} \tag{80}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y_1} \\ \vdots \\ \mathbf{y_{n-1}} \end{bmatrix} \quad \text{and} \quad \mathbf{y_k} = \begin{bmatrix} f(x_{k1}) \\ \vdots \\ f(x_{kr_k}) \end{bmatrix}, \quad k = 1, \ldots, n-1$$

is a column vector of r_k observations on $f(x)$ in the kth interval. By definition, an observation $z_k \in R^{r_k}$ is defined such that:

$$z_k \in [x_k, x_{k+1}) \quad \text{for} \quad k = 1, \ldots, n-2$$
$$z_k \in [x_k, x_{k+1}] \quad \text{for} \quad k = n-1;$$

and $\mathbf{A} = Diag\,[\mathbf{A_1}, \ldots, \mathbf{A_{n-1}}]$ is an r x $(m+1)(n-1)$ block diagonal matrix where $\mathbf{A_k}$ is an r_k x $(m+1)$ matrix of the form:

$$\mathbf{A_k} = \begin{bmatrix} \varphi_0^{(k)}(x_{k1}) & \cdots & \varphi_m^{(k)}(x_{k1}) \\ \vdots & & \vdots \\ \varphi_0^{(k)}(x_{kr_k}) & \cdots & \varphi_m^{(k)}(x_{kr_k}) \end{bmatrix}, \quad k = 1, \ldots, n-1;$$

.

When using monic polynomials, as discussed in section 5.5, the key operation in equation 80 is the inversion of the matrix $\mathbf{A}^{\mathbf{T}}\mathbf{A}$. Two solution techniques are

available for this inversion: LU decomposition or Singular Value Decomposition (SVD). Unlike LU decomposition, SVD enables a clear diagnosis of ill-conditioning situations allowing "zeroing" out of singularities in the matrix $\mathbf{A^T A}$. However, SVD requires that a threshold of singularity be set apriori. [154] [155]

When using orthogonal polynomials, the inversion of the matrix $\mathbf{A^T A}$ in equation 80 is a trivial task. The more arduous effort is the calculation of the orthogonal polynomials $(\varphi_j^{(k)})$ for each interval I_k. Hayes' [94] [169] three-term recurrence relationship shown in section 5.5, equation 77 is used to calculate the orthogonal polynomials.

As an example of fitting monic and orthogonal piecewise polynomial functions to a probability density function, consider the completion time function of event G_{1_1} in the simple air-to-air engagement example created by Hong, et al. [103] shown in Figure 17 - here, $E_{1_1^{(\kappa)}} \sim exp(\lambda_{1_1} t)$ with $\lambda_{1_1} = 1.25$. The left side of Figure 21 shows graphs of density function $f_{E_{1_1^{(\kappa)}}}(t)$ with the linear, quadratic, and cubic polynomial approximation techniques respectively. For each graph, two monic (LUD, SVD) and one orthogonal polynomial least squares technique were used. The density function was truncated at 99% of its total area - the interval of approximation is $[0, 3.6841361488]$. In order to maintain the integrity of the original exponential density function, the area of the piecewise polynomial approximation is brought back to the area of original exponential density function (one) by adding, in this case, .01 to the last subinterval. The implementation adds the factor $(\frac{.01}{classwidth})$ to coefficient $c_0^{(k)}$. Here *classwidth* refers to the length of a subinterval, which in this case is 0.36541361488. Additionally, each piecewise polynomial approximation was determined by the using ten subintervals and eleven points in each interval based on Fergeson and Shortell's recommendation of using $10 + I(classwidth * 3)$ fitting points. *Inspecting the approximations visually, all three piecewise polynomial techniques appear to fit the density function adequately for the linear, quadratic, and cubic polynomial cases.*

Figure 21.   Graph of $f(t)$ and Approximation Error for $E_{1_1^{(\kappa)}}$ in the Simple Air-to-Air Engagement Event Occurrence Network $(E_{1_1^{(\kappa)}} \sim exp(1.25))$

Polynomial Approximation	Error Bound
Linear - Monic (LUD)	$[-0.027, 0.019]$
Linear - Monic (SVD)	$[-0.027, 0.019]$
Linear - Orthogonal	$[-0.027, 0.016]$
Quadratic - Monic (LUD)	$[-0.027, 0.0006]$
Quadratic - Monic (SVD)	$[-0.027, 0.0006]$
Quadratic - Orthogonal	$[-0.027, 0.0006]$
Cubic - Monic (LUD)	$[-0.027, 0.000016]$
Cubic - Monic (SVD)	$[-0.027, 0.000016]$
Cubic - Orthogonal	$[-0.027, 0.000014]$

Table 6. Error Bounds for Piecewise Polynomial Approximations of $f_{E_1^{(\kappa)}}(t)$

In order to quantitatively determine the adequacy of the fit, the right side of Figure 21 shows graphs of the (pointwise) error for the three approximation techniques. All error graphs appear similar in shape having a damped oscillatory nature. Table 6 shows the error bounds for these graphs. Note that *an additional order of precision is gained with each increase in the order of the polynomial approximation* except for the last subinterval. Recall in the last subinterval, an area of .01 was added to the approximation in order to maintain the integrity of the original exponential density function. This addition adversely affects the approximation of the density function in the last subinterval, but is necessary to maintain the accuracy of other stochastic operators such as multiplication and convolution. Additionally, the two monic piecewise polynomial approximations (LU Decomposition, SVD) and the orthogonal piecewise polynomial approximation show *similar accuracy for each increase in the order of the polynomial approximation.*

As an additional example (from Lawrence's research [128]), consider fitting monic and orthogonal piecewise polynomial functions to a truncated normal density function $N(5, 1.67)$ defined on $[0, 10]$. Figure 22 shows the graphs of this truncated normal density function with the linear, quadratic, and cubic polynomial approximation techniques respectively and the graphs of the approximation error for each of these approximation techniques. For the linear case, the "Lawrence" graphs refer

Polynomial Approximation	Error Bound
Linear - Monic (LUD)	$[-0.0059, 0.0036]$
Linear - Monic (SVD)	$[-0.0059, 0.0036]$
Linear - Orthogonal	$[-0.0059, 0.0036]$
Linear - Lawrence	$[-0.0057, 0.0037]$
Quadratic - Monic (LUD)	$[-0.00043, 0.00042]$
Quadratic - Monic (SVD)	$[-0.00043, 0.00042]$
Quadratic - Orthogonal	$[-0.00043, 0.00042]$
Cubic - Monic (LUD)	$[-0.000025, 0.000025]$
Cubic - Monic (SVD)	$[-0.000025, 0.000025]$
Cubic - Orthogonal	$[-0.000025, 0.000025]$

Table 7.   Error Bounds for Piecewise Polynomial Approximations of a truncated $N(5, 1.67)$ density function

to the coefficients generated using simple regression by Lawrence [128]. It should be noted that Lawrence did not use a matrix representation for this regression.

Inspecting the approximations visually, all three piecewise polynomial techniques appear to fit the density function adequately for the linear, quadratic, and cubic polynomial cases. All error graphs appear similar in shape having a damped oscillatory nature. Table 7 shows the error bounds for these graphs. As with the exponential example, an additional order of precision is gained with each increase in the order of the polynomial approximation; however, the accuracy between monic and orthogonal techniques is similar for a given polynomial approximation order.

6.3   Integration Stochastic Operator Approximation

The integration of a polynomial expression is a straightforward task. The integration of a piecewise polynomial approximation is slightly more complex, but nevertheless is also undemanding. Consider the integration of the piecewise polyno-

144
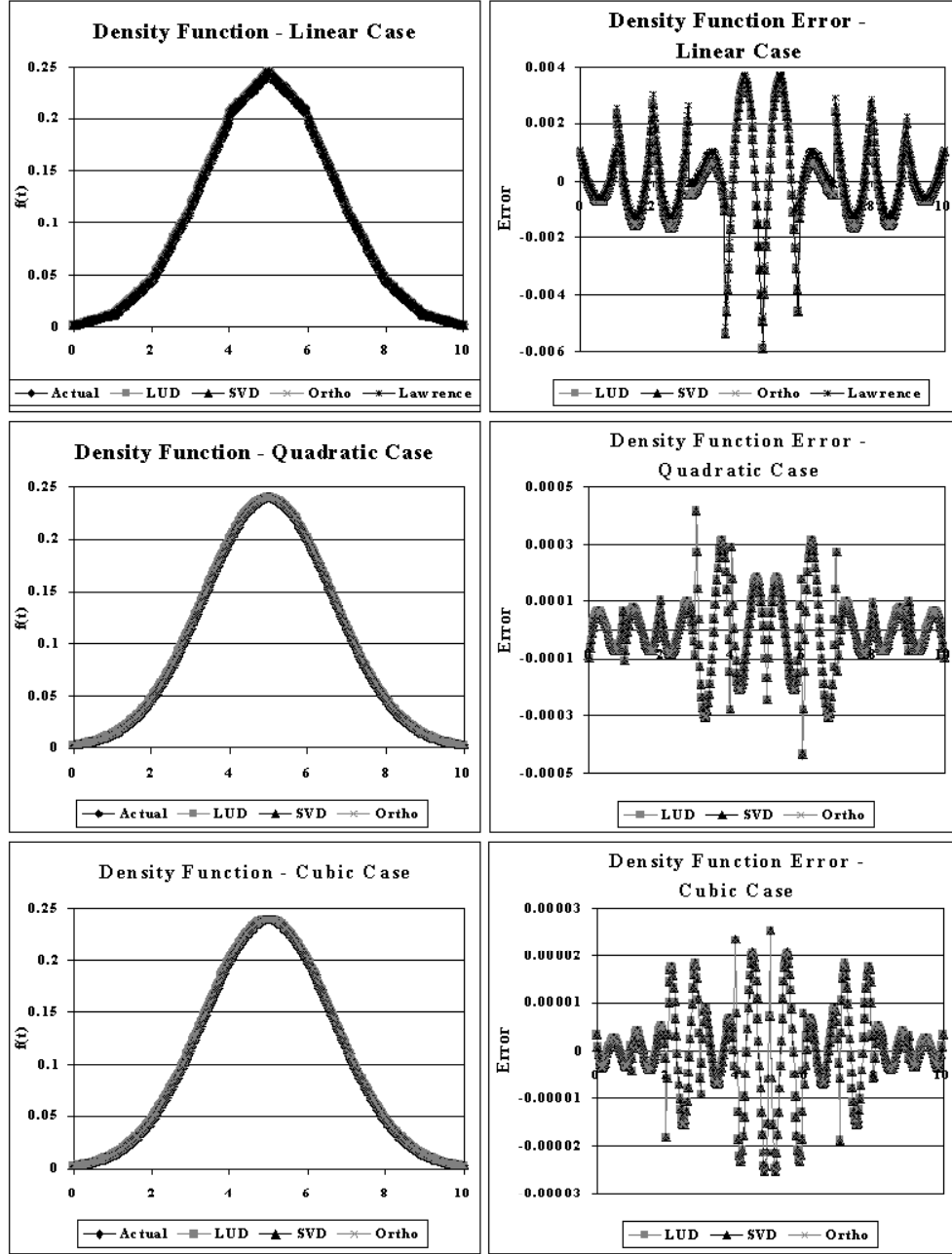
Figure 22. Graph of $f(t)$ and Approximation Error for Truncated Normal Density Function $N(5, 1.67)$ defined on $[0, 10]$

mial function h(x) over the interval $[d_1, d_2]$ where:

$$h(x) = \begin{cases} 0 & x < x_1 = a \\ h_k(x) & x \in I_k, \ k = 1, 2, \ldots, n-1 \\ 0 & x > x_n = b \end{cases}$$

Specifically, let the function H represent this integration, that is:

$$H = \int_{d_1}^{d_2} h(x) dx.$$

The coverage of the interval $[d_1, d_2]$ determines the form of H. Six cases exist:

1. $d_1 < x_1, \ d_2 \leq x_1$;

2. $d_1 < x_1, \ x_1 < d_2 \leq x_n$;

3. $d_1 < x_1, \ x_n < d_2$;

4. $x_1 \leq d_1 < x_n, \ x_1 < d_2 \leq x_n$;

5. $x_1 \leq d_1 < x_n, \ x_n < d_2$; and

6. $x_n < d_1, \ x_n < d_2$.

Cases 1 and 6 are trivial with $H = 0$ over the interval of approximation $[a, b]$. Case 2 involves integration over the interval $[a, d_2]$ while case 3 integrates over $[a, b]$. Case 4 uses $[d_1, d_2]$. Finally, case 5 uses the integration interval $[d_1, b]$.

As an example of calculating H, consider calculating $P(E_{j_1^{(\kappa)}} > e_{g:T})$ which is the complementary cumulative distribution function $F^c_{E_{j_1}}(e_{g:T})$. Recall $F^c_{E_{j_1}}(e_{g:T})$ represents the non-occurrence of event G_{j_1} by time $e_{g:T}$ in an EON and is equivalent to $1 - F_{E_{j_1}}(e_{g:T})$. Specifically,

$$\begin{aligned} F^c_{E_{j_1}}(e_{g:T}) &= 1 - F_{E_{j_1}}(e_{g:T}) \\ &= 1 - \int_0^{e_{g:T}} f_{E_{j_1}}(e_{g:T}) de_{g:T} \end{aligned}$$

$$= 1 - \int_0^{e_{g:T}} \sum_{k=1}^{l|e_{g:T} \in I_l} h_k(e_{g:T}) de_{g:T}$$

$$= 1 - \int_0^{e_{g:T}} \sum_{k=1}^{l|e_{g:T} \in I_l} \sum_{j=0}^{m} c_j^{(k)} \varphi_j^{(k)}(e_{g:T}) de_{g:T}$$

$$= 1 - (\sum_{k=1}^{l-1} \sum_{j=0}^{m} c_j^{(k)} \int_{x_k}^{x_{k+1}} \varphi_j^{(k)}(e_{g:T}) de_{g:T} +$$

$$\sum_{j=0}^{m} c_j^{(l)} \int_{x_l}^{t} \varphi_j^{(k)}(e_{g:T}) de_{g:T}) \quad \text{for all } e_{g:T} \mid e_{g:T} \leq x_n \qquad (81)$$

For $e_{g:T} > x_n$, $F_{E_{j_1}}^c(e_{g:T}) = 0$.

The left side of Figure 23 shows the graph of the complementary cumulative distribution function for the first event from event group 1, $F_{E_{1_1}}^c(t)$ in the air-to-air engagement example created by Hong, et al. [103] with the linear, quadratic, and cubic polynomials approximation techniques respectively. Again these approximations used ten subintervals and eleven least square points in each subinterval. Inspecting the approximations visually, all three piecewise polynomial fits appear to fit the complementary cumulative distribution function adequately.

On the right side of Figure 23, all error graphs appear similar in shape having a damped oscillatory nature. Table 8 shows the error bounds for these graphs. As an aside, Badnelli's [11] observation that "the fit of the integral [is] more accurate than the fit of the integrand [i.e., density function]" is validated for the approximations of $F_{1_1^{(\kappa)}}^c(t)$ and $f_{1_1^{(\kappa)}}(t)$. The left side of Figure 24 displays the relative error of each of polynomial approximation techniques for the $\exp(1.25)$ density function ($f_{1_1^{(\kappa)}}(t)$) while the right side of figure shows the relative error for the $\exp(1.25)$ complementary cumulative distribution function ($F_{1_1^{(\kappa)}}^c(t)$). For each case (linear, quadratic, and cubic), the complementary cumulative distribution function approximation gains an order of magnitude of accuracy over the density function approximation.

As was the case with the density function approximations, the two monic piecewise polynomial approximations (LU Decomposition, SVD) and the orthogonal
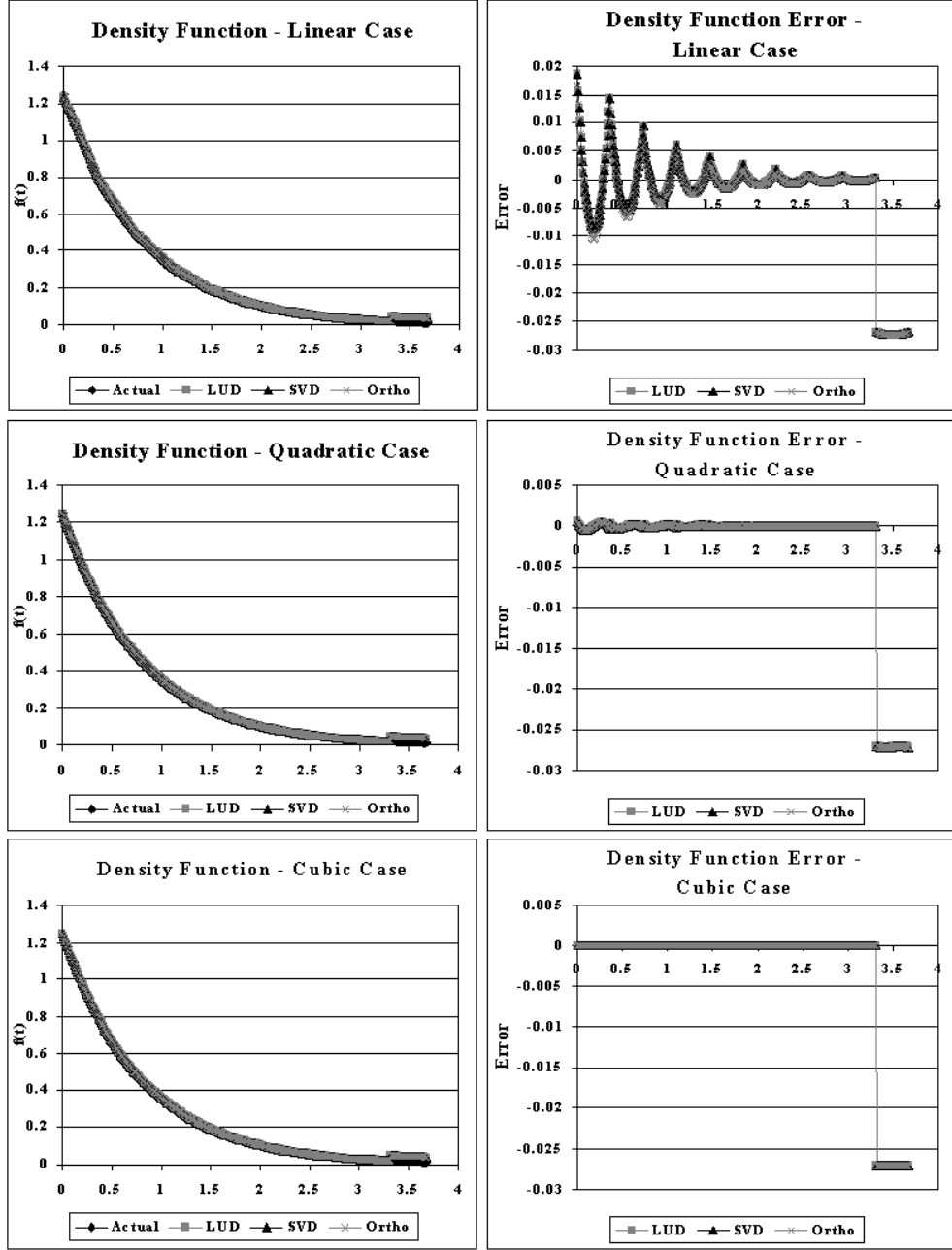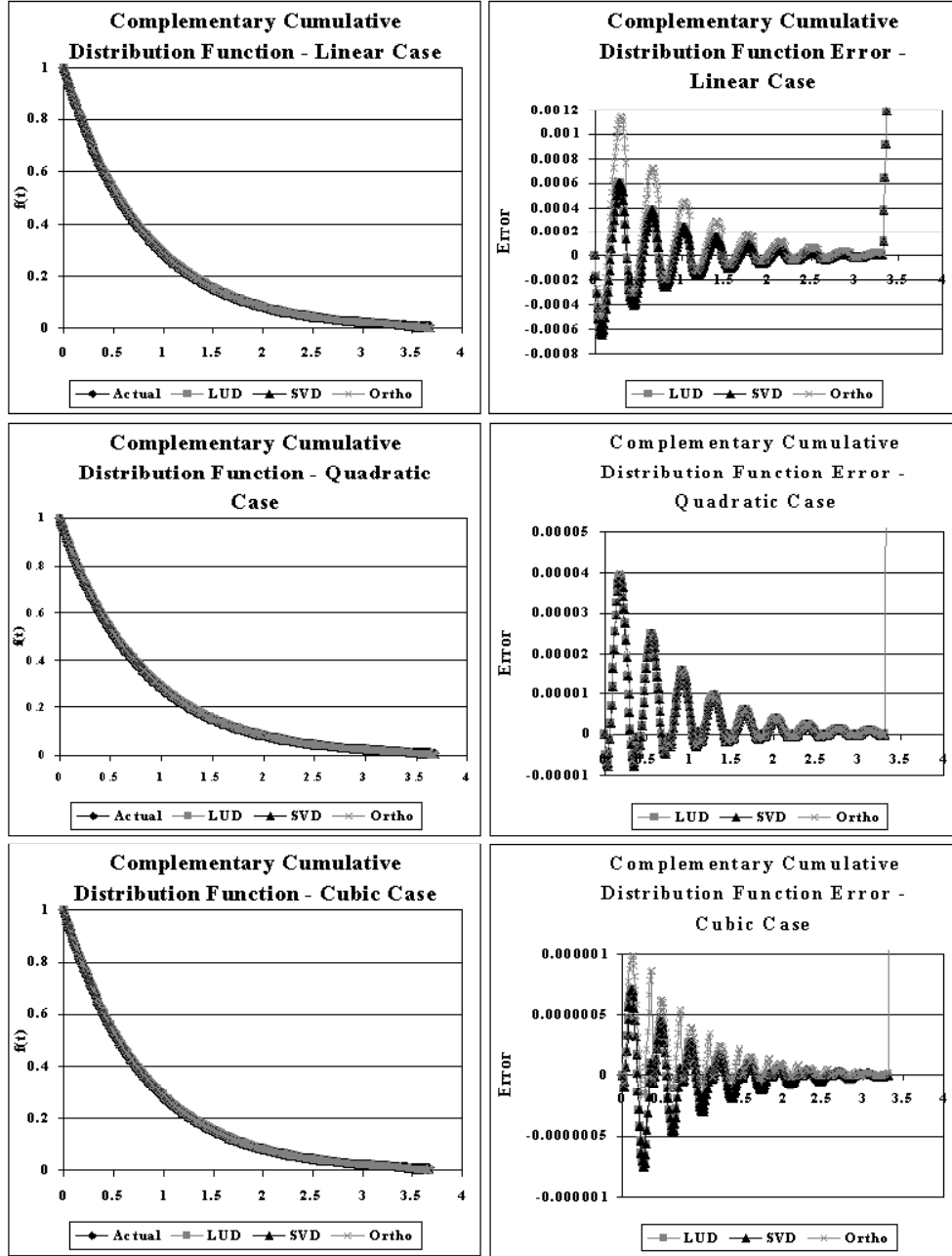
Figure 23.   Graph of $F^c_{1_1(\kappa)}(t)$ and Approximation Error in the Simple Air-to-Air Engagement Event Occurrence Network $(E_{1_1(\kappa)} \sim exp(1.25))$
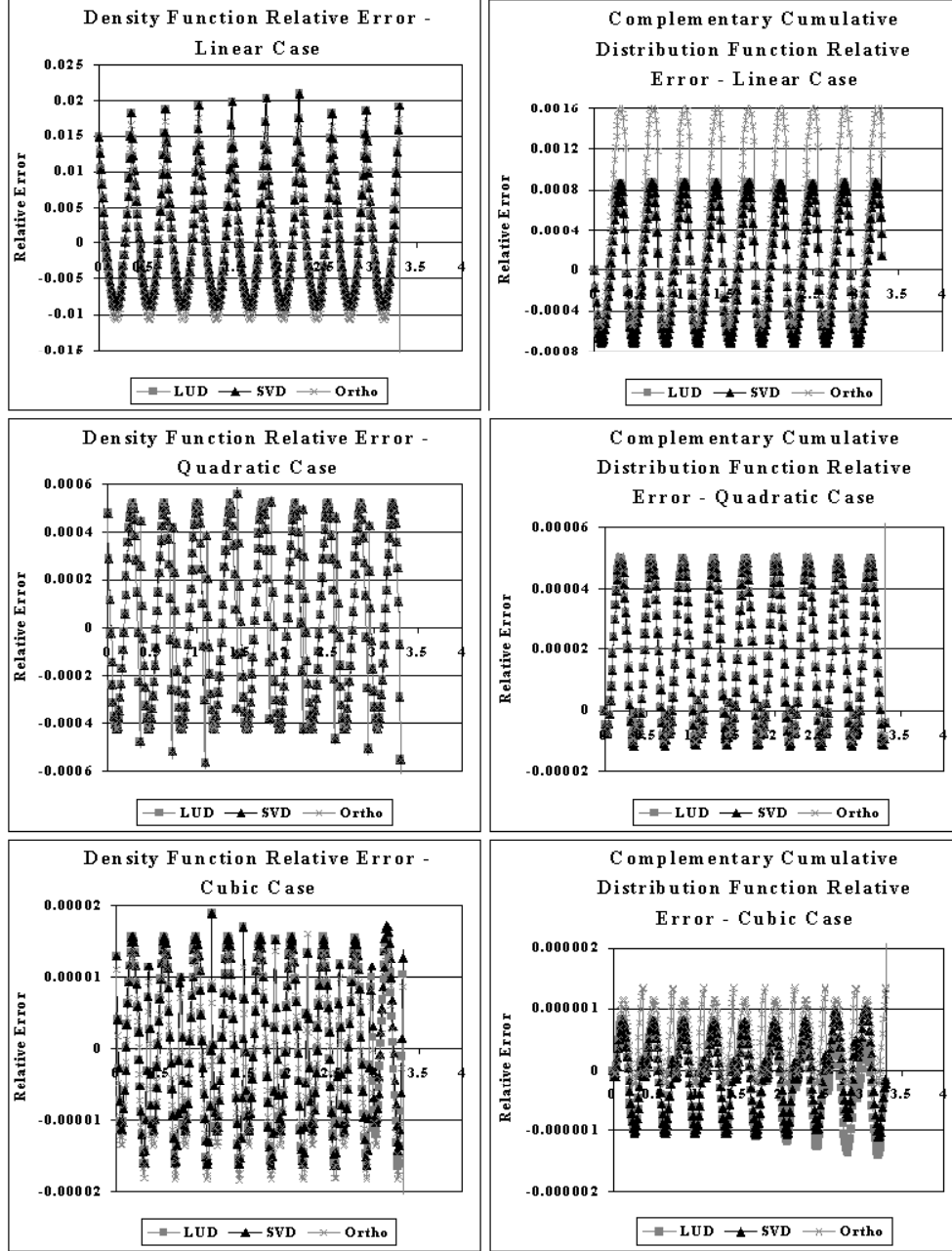
Figure 24. Graph of the Relative Approximation Error of $f_{1_1^{(\kappa)}}(t)$ and $F^c_{1_1^{(\kappa)}}(t)$ in the Simple Air-to-Air Engagement Event Occurrence Network

Polynomial Approximation   **Error Bound**

Linear - Monic (LUD)	$[-0.00064, 0.0099]$
Linear - Monic (SVD)	$[-0.00064, 0.0099]$
Linear - Orthogonal	$[-0.00049, 0.0099]$
Quadratic - Monic (LUD)	$[-0.0000078, 0.0099]$
Quadratic - Monic (SVD)	$[-0.0000078, 0.0099]$
Quadratic - Orthogonal	$[-0.0000077, 0.0099]$
Cubic - Monic (LUD)	$[-0.00000075, 0.0099]$
Cubic - Monic (SVD)	$[-0.00000075, 0.0099]$
Cubic - Orthogonal	$[-0.00000018, 0.0099]$

Table 8.   Approximation Error Bounds for Piecewise Polynomial Approximations of $F^c_{E_{1_1^{(\kappa)}}}(t)$

piecewise polynomial approximation show similar accuracy for each increase in the order of the polynomial approximation. However, the orthogonal polynomial technique requires, for each degree m, the storage of $\sum_{i=1}^{m+1} i$ coefficients where as the monic polynomial technique requires storage of $m + 1$ coefficients for each piecewise segment. For this reason, further development of the orthogonal polynomial approximation technique for EON solutions is not pursued. One further note, since the SVD technique allows a clear diagnosis of ill-conditioning situations and a fix to such situations, monic piecewise polynomial results will be shown using SVD inversion only for the remainder of this chapter.

6.4 Approximation of $P(E_{j:T^{(\iota_j)}} \alpha_j e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$

Recall from section 6.1, the expression $P(E_{j:T^{(\iota_j)}} \alpha_j e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$ consists of the following stochastic operations:

- $P(E_{j_1^{(\kappa)}} > e_{g:T})$;

- $P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$; or

- $P(E_{j_{n_j}^{(0)}} < e_{g:T})$.

The solution to $P(E_{j_1^{(\kappa)}} > e_{g:T})$ was given in equation 81 in the last section. Similarly, the solution to $P(E_{j_{n_j}^{(\kappa)}} < e_{g:T})$ is a byproduct of equation 81 as $P(E_{j_{n_j}^{(\kappa)}} < e_{g:T})$ is the distribution function $F_{E_{j_{n_j}}}(e_{g:T})$:

$$
\begin{aligned}
F_{E_{j_{n_j}}}(e_{g:T}) \ =\ & \sum_{k=1}^{l-1}\sum_{j=0}^{m} c_j^{(k)} \int_{x_k}^{x_{k+1}} \varphi_j^{(k)}(e_{g:T})de_{g:T}\ + \\
& \sum_{j=0}^{m} c_j^{(l)} \int_{x_l}^{t} \varphi_j^{(k)}(e_{g:T})de_{g:T} \quad \text{for all } e_{g:T} \mid e_{g:T} \le x_n \qquad (82)
\end{aligned}
$$

For $e_{g:T} > x_n$, $F_{E_{j_{n_j}}}(e_{g:T}) = 1$.

The solution to $P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$ can be derived from the expression $P(E_{j:T^{(\kappa)}} > e_{g:T})$. Conditioning on the completion of event $G_{j:T-1}$:

$$
\begin{aligned}
P(E_{j:T^{(\kappa)}} > e_{g:T}) \ =\ & P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T}) + P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(\kappa)}} > e_{g:T}) \\
=\ & P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T}) + P(E_{j:T-1^{(\kappa)}} > e_{g:T}) \qquad (83)
\end{aligned}
$$

From equation 83,

$$
\begin{aligned}
P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T}) \ =\ & P(E_{j:T^{(\kappa)}} > e_{g:T}) - P(E_{j:T-1^{(\kappa)}} > e_{g:T}) \\
=\ & F_{E_{j:T}}^c(e_{g:T}) - F_{E_{j:T-1}}^c(e_{g:T}) \qquad (84)
\end{aligned}
$$

6.5   *Multiplication Stochastic Operator Approximation*

Of all stochastic operations necessary to solve for the probability of being in a particular absorbing and transient state of an EON at time t, the multiplication operation is the most prevalent. In equation 50 alone, fourteen different multiplication operations are possible -

1. $P(E_{j_1^{(\kappa)}} > e_{g:T})f_{E_{g:T(1)}}(e_{g:T})$ - $F_{E_{j_1^{(\kappa)}}}^c(e_{g:T})f_{E_{g:T(1)}}(e_{g:T})$;

2. $P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})f_{E_{g:T(1)}}(e_{g:T})$ -
$[F_{E_{j:T^{(\kappa)}}}^c(e_{g:T}) - F_{E_{j:T-1^{(\kappa)}}}^c(e_{g:T})]f_{E_{g:T(1)}}(e_{g:T})$;

151

3. $P(E_{j_{n_j}^{(0)}} < e_{g:T}) f_{E_{g:T^{(1)}}}(e_{g:T})$ - $F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) f_{E_{g:T^{(1)}}}(e_{g:T})$;

4. $P(E_{i_1^{(\kappa)}} > e_{g:T}) P(E_{j_1^{(\kappa)}} > e_{g:T})$ - $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) F^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$;

5. $P(E_{i:T^{(\kappa)}} > e_{g:T} \mid E_{i:T-1^{(0)}} < e_{g:T}) P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$ -
$[F^c_{E_{i:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{i:T-1^{(\kappa)}}}(e_{g:T})][F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]$;

6. $P(E_{i_{n_i}^{(0)}} < e_{g:T}) P(E_{j_{n_j}^{(0)}} < e_{g:T})$ - $F_{E_{i_{n_i}^{(0)}}}(e_{g:T}) F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$;

7. $P(E_{i_1^{(\kappa)}} > e_{g:T}) P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$ -
$F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]$;

8. $P(E_{i_1^{(\kappa)}} > e_{g:T}) P(E_{j_{n_j}^{(0)}} < e_{g:T})$ - $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$;

9. $P(E_{i_{n_i}^{(0)}} < e_{g:T}) P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})$ -
$F_{E_{i_{n_i}^{(0)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]$;

10. $[P(E_{i_1^{(\kappa)}} > e_{g:T}) P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})] f_{E_{g:T^{(1)}}}(e_{g:T})$ -
$[F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]] f_{E_{g:T^{(1)}}}(e_{g:T})$;

11. $[P(E_{i_1^{(\kappa)}} > e_{g:T}) P(E_{j_{n_j}^{(0)}} < e_{g:T})] f_{E_{g:T^{(1)}}}(e_{g:T})$ -
$[F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) F_{E_{j_{n_j}^{(0)}}}(e_{g:T})] f_{E_{g:T^{(1)}}}(e_{g:T})$;

12. $[P(E_{i_{n_i}^{(0)}} < e_{g:T}) P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T})] f_{E_{g:T^{(1)}}}(e_{g:T})$ -
$[F_{E_{i_{n_i}^{(0)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]] f_{E_{g:T^{(1)}}}(e_{g:T})$;

13. $P(E_{i_1^{(\kappa)}} > e_{g:T}) P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T}) P(E_{k_{n_k}^{(0)}} < e_{g:T})$ -
$F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})] F_{E_{k_{n_k}^{(0)}}}(e_{g:T})$; and

14. $[P(E_{i_1^{(\kappa)}} > e_{g:T}) P(E_{j:T^{(\kappa)}} > e_{g:T} \mid E_{j:T-1^{(0)}} < e_{g:T}) P(E_{k_{n_k}^{(0)}} < e_{g:T})] f_{E_{g:T^{(1)}}}(e_{g:T})$ -
$[F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})] F_{E_{k_{n_k}^{(0)}}}(e_{g:T})] f_{E_{g:T^{(1)}}}(e_{g:T})$.

Multiplication operations $(10) - (14)$ are required to extend the solution to EONs whose number of event groups s is three or greater. Concurrently, equation 52 requires the multiplication operations in $(4) - (9)$ and (13) above.

6.5.1 $F^c_{E_{j_1^{(\kappa)}}}(e_{g:T}) f_{E_{g:T^{(1)}}}(e_{g:T})$. Let the complementary cumulative distribution function $F^c_{E_{j_1^{(\kappa)}}}$ and density function $f_{E_{g:T^{(1)}}}(e_{g:T})$ be defined as follows:

$$F^c_{E_{j_1^{(\kappa)}}}(e_{g:T}) = \begin{cases} 1 & \text{if } 0 \le e_{g:T} < d1_1 \\ F^c_{E_{j_1^{(\kappa)}}}(e_{g:T}) & \text{if } d1_1 \le e_{g:T} \le d1_2 \\ 0 & \text{if } e_{g:T} > d1_2 \end{cases}$$

and

$$f_{E_{g:T^{(1)}}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \le e_{g:T} < d2_1 \\ f_{E_{g:T^{(1)}}}(e_{g:T}) & \text{if } d2_1 \le e_{g:T} \le d2_2 \\ 0 & \text{if } e_{g:T} > d2_2. \end{cases}$$

In order to calculate the product $F^c_{E_{j_1^{(\kappa)}}}(e_{g:T}) f_{E_{g:T^{(1)}}}(e_{g:T})$, one first must determine which function reaches a value of zero first - the product is zero after the end of one of the boundary points $d1_2$ or $d2_2$. Additionally, if the intervals $[d1_1, d1_2]$ and $[d2_1, d2_2]$ are disjoint or overlap at only one (boundary) point, the domain of the product is $[d2_1, d2_2]$ (the interval of the density function) with two possible product cases:

- if $d1_2 \le d2_1$, the product is 0 since $F^c_{E_{j_1^{(\kappa)}}}(e_{g:T}) = 0$ over the interval $[d2_1, d2_2]$ and

- if $d2_2 \le d1_1$, the product is $f_{E_{g:T^{(1)}}}(e_{g:T})$ since $F^c_{E_{j_1^{(\kappa)}}}(e_{g:T}) = 1$ over this interval.

Assuming overlap between the two functions, two cases are possible for the domain of the product -

1. if $d1_2 < d2_2$, then the interval of the product is $[d2_1, d1_2]$ and

2. if $d1_2 \ge d2_2$, then the interval of the product is $[d2_1, d2_2]$.

Let $F^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$ and $f_{E_{g:T^{(1)}}}(e_{g:T})$ be approximated by the piecewise polynomial functions $H^c_{E_{j_1^{(\kappa)}}}(e_{gT})$ and $h_{E_{g:T^{(1)}}}(e_{gT})$ over the subinterval sets I_{k_1} and I_{k_2} respectively.

If I_{k_1} is partitioned into $n_1 - 1$ subintervals and I_{k_2} is partitioned into $n_2 - 1$ subintervals, then the piecewise polynomial approximation of $F^c_{E_{j_1^{(\kappa)}}}(e_{g:T})f_{E_{g:T^{(1)}}}(e_{g:T})$, $H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})h_{E_{g:T^{(1)}}}(e_{g:T})$, is partitioned into potentially $n_1 + n_2 - 1$ subintervals.

Hence the subinterval set of $H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})h_{E_{g:T^{(1)}}}(e_{g:T})$, I_k, is composed of a set of boundary points $\{x_{k_1}\} \cup \{x_{k_2}\}$ where $k_1 = 1, \ldots, n_1$ and $k_2 = 1, \ldots, n_2$ - all possible subinterval points of I_{k_1} and I_{k_2}. Let $H^c_{E_{j_1^{(\kappa)}}}(e_{g_T})h_{E_{g:T^{(1)}}}(e_{g_T})_k$ represent the form of the product $H^c_{E_{j_1^{(\kappa)}}}(e_{g_T})h_{E_{g:T^{(1)}}}(e_{g_T})$ in the kth subinterval of I_k -

$$
\begin{aligned}
H^c_{E_{j_1^{(\kappa)}}}(e_{g_T})h_{E_{g:T^{(1)}}}(e_{g_T})_k &= [\, H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})h_{E_{g:T^{(1)}}}(e_{g:T})\,]_k \\
&= \mathbf{c}^{(k_1)}\big(\varphi^{(k_1)}(e_{g:T})\big)^T \mathbf{c}^{(k_2)}\big(\varphi^{(k_2)}(e_{g:T})\big)^T. \quad (85)
\end{aligned}
$$

where

- $\mathbf{c}^{(k_1)}$ is the $m+2$-dimension polynomial coefficient vector and $\varphi^{(k_1)}$ is the $m+2$-dimension polynomial basis function vector for the k_1st interval of I_{k_1} and

- $\mathbf{c}^{(k_2)}$ is the $m+1$-dimension polynomial coefficient vector and $\varphi^{(k_2)}$ is the $m+1$-dimension polynomial basis function vector for the k_2nd interval of I_{k_2}.

Assuming a monic piecewise polynomial approximation, the above expression may be written as:

$$
H^c_{E_{j_1^{(\kappa)}}}(e_{g_T})h_{E_{g:T^{(1)}}}(e_{g_T})_k = \mathbf{c}^{(k_1)}\mathbf{C}^{(k_2)}\mathbf{x}^T, \quad k \in I_k \quad (86)
$$

where

$$
\mathbf{C}^{(k_2)} = \begin{bmatrix}
c_0^{(k_1)} & \ldots & c_m^{(k_1)} & 0 & \ldots & 0 \\
0 & c_0^{(k_1)} & \ldots & c_m^{(k_1)} & \ddots & \vdots \\
\vdots & \ddots & \ddots & & \ddots & 0 \\
0 & \ldots & 0 & c_0^{(k_1)} & \ldots & c_m^{(k_1)}
\end{bmatrix} \quad \text{and}
$$

$$
\mathbf{x} = \big(1, e_{g:T}, \ldots, (e_{g:T})^{2m+1}\big)
$$

154

Note that the order of each piecewise polynomial segment in equation 86 is $2m + 1$. In order to manage a potential problem of "exploding coefficients" once the "expanded" piecewise polynomial approximation in equation 86 is calculated, the approximation is "reduced" by fitting a piecewise polynomial approximation of degree m to the "expanded" approximation. In order to control error build-up from the "reduced" approximation, the "reduced" approximation is then normalized to the area of the "expanded" approximation. This normalization is accomplished by dividing each coefficient in each piecewise approximation by:

$$\frac{\text{"expanded" approximation area}}{\text{"reduced" approximation area}}.$$

Additionally, each piecewise (reduced) approximation is checked to verify that it does not yield values less than zero - if it does, the piecewise approximation is increased by a factor c_{0+}^k necessary to guarantee values greater than zero. Non-zero values generally appear in the tails of a product function where values of the function approach zero.

As a final note, if $d1_1 > d2_1$, then $\mathbf{c}^{(k_1)} = \mathbf{1}$ until $d1_1 \geq x_k, x_k \in I_k$.

6.5.2 $[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]f_{E_{g:T(1)}}(e_{g:T})$. Let the function $F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ be defined as follows:

$$F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T}) & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 0 & \text{if } e_{g:T} > d1_2 \end{cases}$$

and $f_{E_{g:T(1)}}(e_{g:T})$ be defined as in section 6.5.1.

If the intervals $[d1_1, d1_2]$ and $[d2_1, d2_2]$ are disjoint or overlap at only one (boundary) point, the product $[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]f_{E_{g:T(1)}}(e_{g:T})$ is 0 with two possible domain cases:

- if $d1_2 \leq d2_1$, the domain of the product is $[d1_1, d2_2]$ and

- if $d2_2 \leq d1_1$, the domain of the product is $[d2_1, d1_2]$.

Assuming overlap between the two functions, four cases are possible for the domain of the product -

1. if $d1_1 > d2_1$ and $d1_2 \leq d2_2$, then the interval of the product is $[d1_1, d1_2]$;

2. if $d1_1 \leq d2_1$ and $d1_2 \leq d2_2$, then the interval of the product is $[d2_1, d1_2]$;

3. if $d1_1 > d2_1$ and $d1_2 > d2_2$, then the interval of the product is $[d1_1, d2_2]$; and

4. if $d1_1 \leq d2_1$ and $d1_2 > d2_2$, then the interval of the product is $[d2_1, d2_2]$,

Equations 85 and 86 in section 6.5.1 are still valid for approximating $[H^c_{E_{j:T(\kappa)}}(e_{g:T}) - H^c_{E_{j:T-1(\kappa)}}(e_{g:T})]h_{E_{g:T(1)}}(e_{g:T})_k$ - replacing $H^c_{E_{j_1(\kappa)}}(e_{g:T})$ with $H^c_{E_{j:T(\kappa)}}(e_{g:T}) - H^c_{E_{j:T-1(\kappa)}}(e_{g:T})$. $H^c_{E_{j:T(\kappa)}}(e_{g:T}) - H^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ is the piecewise polynomial approximation of $F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ over the subinterval I_{k_1}.

6.5.3 $F_{E_{j_{n_j}^{(0)}}}(e_{g:T})f_{E_{g:T(1)}}(e_{g:T})$. Let the distribution function $F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ be defined as follows:

$$
F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 1 & \text{if } e_{g:T} > d1_2 \end{cases}
$$

and $f_{E_{g:T(1)}}(e_{g:T})$ be defined as in section 6.5.1.

If the intervals $[d1_1, d1_2]$ and $[d2_1, d2_2]$ are disjoint or overlap at only one (boundary) point, the domain of the product is $[d2_1, d2_2]$ (the interval of the density function) with two possible product cases:

- if $d1_2 \leq d2_1$, the product is $f_{E_{g:T(1)}}(e_{g:T})$ since $F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) = 1$ over the interval $[d2_1, d2_2]$ and

- if $d2_2 \leq d1_1$, the product is 0 since $F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) = 0$ over this interval.

Assuming overlap between the two functions, two cases are possible for the domain of the product -

1. if $d1_1 > d2_1$, then the interval of the product is $[d1_1, d2_2]$;

2. if $d1_1 \leq d2_1$, then the interval of the product is $[d2_1, d2_2]$;

Equations 85 and 86 in section 6.5.1 are still valid for approximating $H_{E_{j_{n_j}^{(0)}}}(e_{g:T}) h_{E_{g:T^{(1)}}}(e_{g:T})_k$ - replacing $H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$ with $H_{E_{j_{n_j}^{(0)}}}(e_{g:T})$. $H_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ is the piecewise polynomial approximation of $F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ over the subinterval I_{k_1}. As a final note, if $d1_2 < d2_2$, then $\mathbf{c}^{(k_1)} = \mathbf{1}$ in equations 85 and 86 for all $x_k \geq d1_2, x_k \in I_k$.

6.5.4 $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) F^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$.   Let the complementary cumulative distribution functions $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})$ and $F^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$ be defined as follows:

$$
F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) = \begin{cases} 1 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 0 & \text{if } e_{g:T} > d1_2 \end{cases}
$$

and

$$
F^c_{E_{j_2^{(\kappa)}}}(e_{g:T}) = \begin{cases} 1 & \text{if } 0 \leq e_{g:T} < d2_1 \\ F^c_{E_{j_2^{(\kappa)}}}(e_{g:T}) & \text{if } d2_1 \leq e_{g:T} \leq d2_2 \\ 0 & \text{if } e_{g:T} > d2_2. \end{cases}
$$

In order to calculate the product $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) F^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$, one first must determine which distribution reaches a value of zero first - the product is zero after the end of one of boundary points $d1_2$ or $d2_2$. Additionally, if the intervals $[d1_1, d1_2]$ and $[d2_1, d2_2]$ are disjoint or overlap at only one (boundary) point, the product of two complementary cumulative distribution functions is simply the complementary distribution function of the lower interval.

Assuming overlap between the two complementary cumulative distribution functions, four cases are possible for the domain of the product -

1. If $d1_1 \leq d2_1$ and $d1_2 \leq d2_2$, then the interval of the product is $[d1_1, d1_2]$;

2. If $d1_1 \leq d2_1$ and $d2_2 < d1_2$, then the interval of the product is $[d1_1, d2_2]$;

3. If $d2_1 < d1_1$ and $d1_2 \leq d2_2$, then the interval of the product is $[d2_1, d1_2]$; and

4. If $d2_1 < d1_1$ and $d2_2 < d1_2$, then the interval of the product is $[d2_1, d2_2]$.

Let $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})$ and $F^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$ be approximated by the piecewise polynomial functions $H^c_{E_{i_1^{(\kappa)}}}(e_{g:T})$ and $H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$ over the subinterval sets I_{k_1} and I_{k_2} respectively. If I_{k_1} is partitioned into $n_1 - 1$ subintervals and I_{k_2} is partitioned into $n_2 - 1$ subintervals, then the piecewise polynomial approximation of $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})F^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$, $H^c_{E_{i_1^{(\kappa)}}}(e_{g:T})H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$, is partitioned into potentially $n_1 + n_2 - 1$ subintervals.

Hence the subinterval set of $H^c_{E_{i_1^{(\kappa)}}}(e_{g:T})H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$, I_k, is composed of a set of boundary points $\{x_{k_1}\} \cup \{x_{k_2}\}$ where $k_1 = 1, \ldots, n_1$ and $k_2 = 1, \ldots, n_2$ - all possible subinterval points of I_{k_1} and I_{k_2}. Let $H^c_{E_{i_1^{(\kappa)}}}(e_{g:T})H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})_k$ represent the form of the $H^c_{E_{i_1^{(\kappa)}}}(e_{g:T})H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})$ in the kth subinterval of I_k -

$$
\begin{aligned}
H^c_{E_{i_1^{(\kappa)}}}(e_{g:T})H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})_k &= [\, H^c_{E_{i_1^{(\kappa)}}}(e_{g:T})H^c_{E_{j_1^{(\kappa)}}}(e_{g:T}) \,]_k \\
&= \mathbf{c}^{(k_1)}\big(\varphi^{(k_1)}(e_{g:T})\big)^T \mathbf{c}^{(k_2)}\big(\varphi^{(k_2)}(e_{g:T})\big)^T. \qquad (87)
\end{aligned}
$$

where

- $\mathbf{c}^{(k_1)}$ is the $m+2$-dimension polynomial coefficient vector and $\varphi^{(k_1)}$ is the $m+2$-dimension polynomial basis function vector for the k_1st interval of I_{k_1} and

- $\mathbf{c}^{(k_2)}$ is the $m+2$-dimension polynomial coefficient vector and $\varphi^{(k_2)}$ is the $m+2$-dimension polynomial basis function vector for the k_2nd interval of I_{k_2}.

Assuming a monic piecewise polynomial approximation, the above expression may be written as:

$$H^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) H^c_{E_{j_1^{(\kappa)}}}(e_{g:T})_k = \mathbf{c}^{(k_1)} \mathbf{C}^{(k_2)} \mathbf{x} \tag{88}$$

where

$$\mathbf{C}^{(k_2)} = \begin{bmatrix} c_0^{(k_1)} & \dots & c_{m+1}^{(k_1)} & 0 & \dots & 0 \\ 0 & c_0^{(k_1)} & \dots & c_{m+1}^{(k_1)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \dots & 0 & c_0^{(k_1)} & \dots & c_{m+1}^{(k_1)} \end{bmatrix} \text{ and}$$

$$\mathbf{x} = \left(1, e_{g:T}, \dots, (e_{g:T})^{2(m+1)}\right)$$

Note that the order of each piecewise polynomial segment in equation 88 is $2(m+1)$. As in section 6.5.1, in order to manage a potential problem of "exploding coefficients" once the "expanded" piecewise polynomial approximation in equation 88 is calculated, the approximation is "reduced" by fitting a piecewise polynomial approximation of degree $m + 1$ to the "expanded" approximation. For this operation, an area normalization routine as outlined in section 6.5.1 was *not* used as test cases indicated no significant improvement in accuracy when employed. However, each piecewise approximation is checked to verify that it does not yield values less than zero - if it does the piecewise approximation is increased by a factor c_{0+} necessary to guarantee values greater than zero. As stated in section 6.5.1, non-zero values can appear in the tails of the probability functions where values of the function approach zero.

As an example of the product of two complementary cumulative distribution functions consider the probability of being in state 1 at time t, $P_1(t)$, for the simple air-to-air engagement example created by Hong, et al. [103] shown in Figure 17.

Approximation	Error Bound
Linear	$[-0.003, 0.002]$
Quadratic	$[-0.00013, 0.00016]$
Cubic	$[-0.000007, 0.00014]$

Table 9. Error Bounds for Piecewise Polynomial Approximations of $P_1(t)$

Recall from equation 40 in section 4.6 that:

$$P_1(t) = F^c_{E^{(\kappa)}_{1_1}}(t) * F^c_{E^{(\kappa)}_{2_1}}(t)$$

The left side of Figure 25 shows the graph of $P_1(t)$ and the three piecewise polynomial approximations using the linear, quadratic, and cubic monic polynomials techniques respectively. The three piecewise polynomial approximations were again determined by the least squares criterion (SVD inversion engine) using ten subintervals. Inspecting the approximations visually, all three piecewise polynomial fits appear to fit $P_1(t)$ adequately.

In order to quantitatively determine the adequacy of the fit, the right side of Figure 25 shows graphs of the (pointwise) error for the three approximations. Table 9 shows the error bounds for these graphs. The error graphs exhibit damped oscillatory behavior that moves towards zero except for the "hump" between $[3.25, 5.25]$. As expected this is caused by truncation of the exponential functions and subsequent area renormalization. In order to fully realize the accuracy gains with the cubic approximation, truncating the event occurrence density functions at greater than 99% of the total areas would be recommended.

6.5.5 $\left[F^c_{E_{i:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{i:T-1^{(\kappa)}}}(e_{g:T})\right]$

$\left[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})\right]$.   Let the functions $F^c_{E_{i:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{i:T-1^{(\kappa)}}}(e_{g:T})$
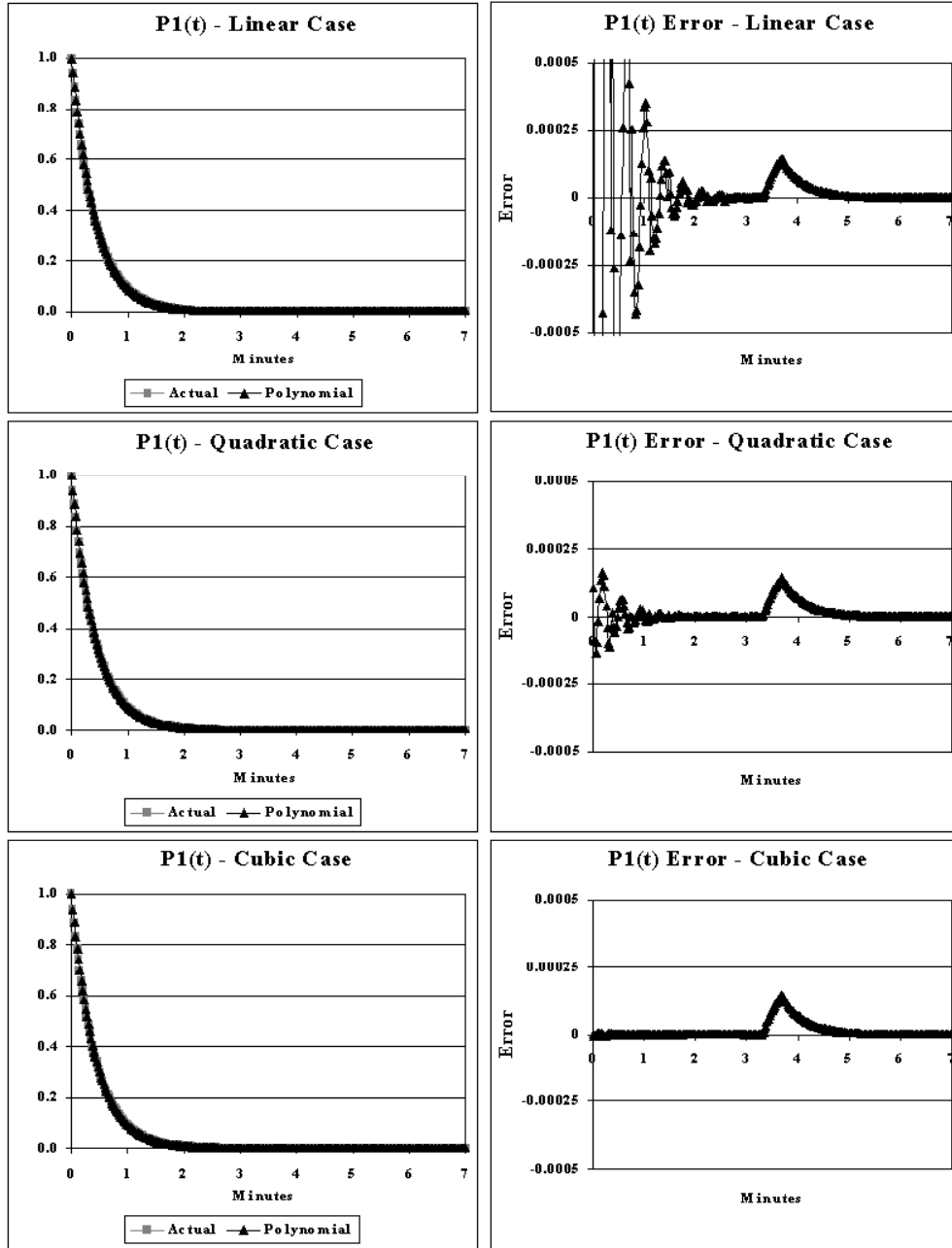
Figure 25. Graph of $P_1(t)$ and Error for the Simple Air-to-Air Engagement Event Occurrence Network

and $F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ be defined as follows:

$$F^c_{E_{i:T(\kappa)}}(e_{g:T}) - F^c_{E_{i:T-1(\kappa)}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F^c_{E_{i:T(\kappa)}}(e_{g:T}) - F^c_{E_{i:T-1(\kappa)}}(e_{g:T}) & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 0 & \text{if } e_{g:T} > d1_2 \end{cases}$$

and

$$F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T}) & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 0 & \text{if } e_{g:T} > d1_2. \end{cases}$$

The domain of this product is identical to the domain of the product $[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]f_{E_{g:T(1)}}(e_{g:T})$ detailed in section 6.5.2. If the intervals $[d1_1, d1_2]$ and $[d2_1, d2_2]$ are disjoint or overlap at only one (boundary) point, the product $[F^c_{E_{i:T(\kappa)}}(e_{g:T}) - F^c_{E_{i:T-1(\kappa)}}(e_{g:T})][F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]$ is 0 with two possible domain cases:

- if $d1_2 \leq d2_1$, the domain of the product is $[d1_1, d2_2]$ and

- if $d2_2 \leq d1_1$, the domain of the product is $[d2_1, d1_2]$.

Assuming overlap between the two functions, four cases are possible for the domain of the product -

1. if $d1_1 > d2_1$ and $d1_2 \leq d2_2$, then the interval of the product is $[d1_1, d1_2]$;

2. if $d1_1 \leq d2_1$ and $d1_2 \leq d2_2$, then the interval of the product is $[d2_1, d1_2]$;

3. if $d1_1 > d2_1$ and $d1_2 > d2_2$, then the interval of the product is $[d1_1, d2_2]$; and

4. if $d1_1 \leq d2_1$ and $d1_2 > d2_2$, then the interval of the product is $[d2_1, d2_2]$.

Contrary to the domain of the product, the product $[F^c_{E_{i:T(\kappa)}}(e_{g:T}) - F^c_{E_{i:T-1(\kappa)}}(e_{g:T})][F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]$ is determined using

equations 87 and 88 in section 6.5.4. The approximating function is $\big[H^c_{E_{i:T(\kappa)}}(e_{g:T}) - H^c_{E_{i:T-1(\kappa)}}(e_{g:T})\big]\big[H^c_{E_{j:T(\kappa)}}(e_{g:T}) - H^c_{E_{j:T-1(\kappa)}}(e_{g:T})\big]_k$ where $H^c_{E_{i:T(\kappa)}}(e_{g:T}) - H^c_{E_{i:T-1(\kappa)}}(e_{g:T})$ and $H^c_{E_{j:T(\kappa)}}(e_{g:T}) - H^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ are the piecewise polynomial approximations of $F^c_{E_{i:T(\kappa)}}(e_{g:T}) - F^c_{E_{i:T-1(\kappa)}}(e_{g:T})$ and $F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ respectively over the subintervals I_{k_1} and I_{k_2}.

6.5.6 $F_{E_{i_{n_i}^{(0)}}}(e_{g:T})F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$.   Let the distribution functions $F_{E_{i_{n_i}^{(0)}}}(e_{g:T})$ and $F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ be defined as follows:

$$
F_{E_{i_{n_i}^{(0)}}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F_{E_{i_{n_i}^{(0)}}}(e_{g:T}) & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 1 & \text{if } e_{g:T} > d1_2 \end{cases}
$$

and

$$
F_{E_{i_{n_i}^{(0)}}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F_{E_{i_{n_i}^{(0)}}}(e_{g:T}) & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 1 & \text{if } e_{g:T} > d1_2. \end{cases}
$$

If the intervals $[d1_1, d1_2]$ and $[d2_1, d2_2]$ are disjoint or overlap at only one (boundary) point, the domain of the product and the product itself have two possible outcomes:

- if $d1_2 \leq d2_1$, the domain of the product is $[d2_1, d2_2]$ with product $F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ as a result of $F_{E_{i_{n_i}^{(0)}}}(e_{g:T}) = 1$ over $[d2_1, d2_2]$ and
- if $d2_2 \leq d1_1$, the domain of the product is $[d1_1, d1_2]$ with product $F_{E_{i_{n_i}^{(0)}}}(e_{g:T})$ as a result of $F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) = 1$ over $[d1_1, d1_2]$.

Assuming overlap between the two distribution functions, four cases are possible for the domain of the product -

1. if $d1_1 > d2_1$ and $d1_2 \leq d2_2$, then the interval of the product is $[d1_1, d2_2]$;

163

2. if $d1_1 \leq d2_1$ and $d1_2 \leq d2_2$, then the interval of the product is $[d2_1, d2_2]$;

3. if $d1_1 > d2_1$ and $d1_2 > d2_2$, then the interval of the product is $[d1_1, d1_2]$; and

4. if $d1_1 \leq d2_1$ and $d1_2 > d2_2$, then the interval of the product is $[d2_1, d1_2]$.

Equations 87 and 88 in section 6.5.4 are valid for approximating $H_{E_{i_{n_i}^{(0)}}}(e_{g:T}) H_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ where $H_{E_{i_{n_i}^{(0)}}}(e_{g:T})$ and $H_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ are the piecewise polynomial approximations of $F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ and $F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ respectively over the subintervals I_{k_1} and I_{k_2}. As final notes, if $d1_2 < d2_2$, then $\mathbf{c}^{(k_1)} = \mathbf{1}$ in equations 87 and 88 for all $x_k \geq d1_2, x_k \in I_k$. If $d2_2 < d2_1$, then $\mathbf{c}^{(k_2)} = \mathbf{1}$ in equations 87 and 88 for all $x_k \geq d2_2, x_k \in I_k$

6.5.7 $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]$. This multiplication operator is similar to the $F^c_{E_{j_1^{(\kappa)}}}(e_{g:T}) f_{E_{g:T^{(1)}}}(e_{g:T})$ operator in section 6.5.1. In fact, the no-overlap and overlap domains are identical. The products for the no-overlap cases are:

- if $d1_2 \leq d2_1$, the product is 0 since $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) = 0$ over the domain $[d2_1, d2_2]$ and

- if $d2_2 \leq d1_1$, the product is $F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})$ since $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) = 1$ over this domain.

The product approximation $H^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[H^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - H^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]_k$ for the overlap domains $[d2_1, d1_2]$ and $[d2_1, d2_2]$ is determined from equations 87 and 88 in section 6.5.4. As a final note, if $d1_1 > d2_1$, then $\mathbf{c}^{(k_1)} = \mathbf{1}$ until $d1_1 \geq x_k, x_k \in I_k$.

6.5.8 $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$. Let the complementary cumulative distribution function $F^c_{E_{i_1^{(\kappa)}}}$ and distribution function $F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ be defined as follows:

$$F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) = \begin{cases} 1 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 0 & \text{if } e_{g:T} > d1_2 \end{cases}$$

164

and

$$F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \le e_{g:T} < d1_1 \\ F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) & \text{if } d1_1 \le e_{g:T} \le d1_2 \\ 1 & \text{if } e_{g:T} > d1_2. \end{cases}$$

If the intervals $[d1_1, d1_2]$ and $[d2_1, d2_2]$ are disjoint or overlap at only one (boundary) point, the two separate domain and product cases are possible:

- if $d1_2 \le d2_1$, the product is 0 over the domain $[d1_1, d2_2]$ and

- if $d2_2 \le d1_1$, the product is

$$F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \le e_{g:T} < d2_1 \\ F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) & \text{if } d2_1 \le e_{g:T} < d2_2 \\ 1 & \text{if } d2_2 \le e_{g:T} < d1_1 \\ F^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) & \text{if } d1_1 \le e_{g:T} \le d1_2 \\ 0 & \text{if } e_{g:T} > d1_2 \end{cases}$$

over the domain $[d2_1, d1_2]$.

The product approximation $H^c_{E_{i_1^{(\kappa)}}}(e_{g:T}) H_{E_{j_{n_j}^{(0)}}}(e_{g:T})_k$ for the overlap domain $[d2_1, d1_2]$ is determined from equations 87 and 88 in section 6.5.4. As final notes, if $d1_1 > d2_1$, then $\mathbf{c}^{(k_1)} = \mathbf{1}$ until $d1_1 \ge x_k, x_k \in I_k$ and if $d2_2 < d1_2$, then $\mathbf{c}^{(k_2)} = \mathbf{1}$ for all $x_k \ge d2_2, x_k \in I_k$.

6.5.9 $F_{E_{i_{n_i}^{(0)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]$. This multiplication operator is similar to the $F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) f_{E_{g:T^{(1)}}}(e_{g:T})$ operator in section 6.5.3. In fact, the no-overlap and overlap domains are identical. The products for the no-overlap cases are:

- if $d1_2 \le d2_1$, the product is $[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]$ since $F_{E_{i_{n_i}^{(0)}}}(e_{g:T}) = 1$ over the domain $[d2_1, d2_2]$ and

- if $d2_2 \leq d1_1$, the product is 0 since $F_{E_{i_{n_i}^{(0)}}}(e_{g:T}) = 0$ over this domain.

The product approximation $H_{E_{i_{n_i}^{(0)}}}(e_{g:T})[H^c_{E_{j:T(\kappa)}}(e_{g:T}) - H^c_{E_{j:T-1(\kappa)}}(e_{g:T})]_k$ for the overlap domains $[d1_1, d2_2]$ and $[d2_1, d2_2]$ is determined from equations 87 and 88 in section 6.5.4. As a final note, if $d1_2 < d2_2$, then $\mathbf{c}^{(k_1)} = \mathbf{1}$ for all $x_k \geq d1_2, x_k \in I_k$.

6.5.10 $[F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]]$

$f_{E_{g:T(1)}}(e_{g:T})$. The product $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]$ was discussed in section 6.5.7. The form of this product is as follows:

$$F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})] =$$

$$\begin{cases} 0 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})] & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 0 & \text{if } e_{g:T} > d1_2. \end{cases}$$

The domain of the product $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]$ is similar in form to the domain of the product $F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ in section 6.5.2. As a result, the domain logic and product equations from section 6.5.2 are applicable to the calculation of $[F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]]f_{E_{g:T(1)}}(e_{g:T})$.

6.5.11 $[F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})F_{E_{j_{n_j}^{(0)}}}(e_{g:T})]f_{E_{g:T(1)}}(e_{g:T})$. The product

$F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ was discussed in section 6.5.8. The form of this product is as follows:

$$F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) = \begin{cases} 0 & \text{if } 0 \leq e_{g:T} < d1_1 \\ F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})F_{E_{j_{n_j}^{(0)}}}(e_{g:T}) & \text{if } d1_1 \leq e_{g:T} \leq d1_2 \\ 0 & \text{if } e_{g:T} > d1_2. \end{cases}$$

The domain of the product $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})F_{E_{j_{n_j}^{(0)}}}(e_{g:T})$ is similar in form to the domain of the product $F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ in section 6.5.2. As a result, the domain logic and product equations from section 6.5.2 are applicable to the calculation of $[F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})F_{E_{j_{n_j}^{(0)}}}(e_{g:T})]f_{E_{g:T(1)}}(e_{g:T})$.

6.5.12 $[F_{E_{i_{n_i}^{(0)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]]$
$f_{E_{g:T(1)}}(e_{g:T})$. The product $F_{E_{i_{n_i}^{(0)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]$ was discussed in section 6.5.9. The form of this product is as follows:

$$F_{E_{i_{n_i}^{(0)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})] =$$

$$\begin{cases} 0 & \text{if } 0 \le e_{g:T} < d1_1 \\ F_{E_{i_{n_i}^{(0)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})] & \text{if } d1_1 \le e_{g:T} \le d1_2 \\ 0 & \text{if } e_{g:T} > d1_2. \end{cases}$$

The domain of the product $F_{E_{i_{n_i}^{(0)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]$ is similar in form to the domain of the product $F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ in section 6.5.2. As a result, the domain logic and product equations from section 6.5.2 are applicable to the calculation of $F_{E_{i_{n_i}^{(0)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]]f_{E_{g:T(1)}}(e_{g:T})$.

6.5.13 $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]F_{E_{k_{n_k}^{(0)}}}(e_{g:T})$. The product $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]$ was discussed in sections 6.5.7 and 6.5.10. It has been shown that the domain of this product is similar in form to the domain of the product $F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})$ in section 6.5.2. As a result, the domain logic and product equations from section 6.5.9 are applicable to the calculation of $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]F_{E_{k_{n_k}^{(0)}}}(e_{g:T})$.

6.5.14 $[F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T(\kappa)}}(e_{g:T}) - F^c_{E_{j:T-1(\kappa)}}(e_{g:T})]F_{E_{k_{n_k}^{(0)}}}(e_{g:T})]$
$f_{E_{g:T(1)}}(e_{g:T})$. This last EON multiplication stochastic operator is calculated using two previous operators:

1. the product $F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]F_{E_{k_{n_k}^{(0)}}}(e_{g:T})$ in section 6.5.13 and

2. the product $[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]f_{E_{g:T^{(1)}}}(e_{g:T})$ in section 6.5.2. Here replacing the product $[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]$ with the product $[F^c_{E_{i_1^{(\kappa)}}}(e_{g:T})[F^c_{E_{j:T^{(\kappa)}}}(e_{g:T}) - F^c_{E_{j:T-1^{(\kappa)}}}(e_{g:T})]F_{E_{k_{n_k}^{(0)}}}(e_{g:T})]$.

6.6   Convolution Stochastic Operator Approximation

Equations 50 and 52 require calculation of convolution(s) of event occurrence density functions. Specifically, consider the (convoluted) density function of the occurrence time of the first T events in event group g - i.e., $f_{E_{g:T}}(t)$. This density function may be expressed as a convolution of the density function of the occurrence time of the first $T - 1$ events in event group g and the event occurrence density function of the Tth event in event group g; i.e.,

$$f_{E_{g:T}}(t) = (f_{E_{g:T-1}} \oplus f_{E_{g_T}})(t) \tag{89}$$

The first step in solving the convolution in equation 89 is to determine the applicable, approximation partition for $f_{E_{g:T}}(t)$. Specifically, let $f_{E_{g:T-1}}(t)$ and $f_{E_{g_T}}(t)$ be approximated by the piecewise polynomial functions $h_{E_{g:T-1}}$ and $h_{E_{g_T}}$ and over the subinterval sets I_{k_1} and I_{k_2} respectively. If I_{k_1} is partitioned into $n_1 - 1$ subintervals and I_{k_2} is partitioned into $n_2 - 1$ subintervals, then the piecewise polynomial approximation of $f_{E_{g:T}}(t)$, $h_{E_{g:T}}(t)$, is partitioned into potentially $n_1 n_2 - 1$ subintervals.

The subinterval set of $h_{E_{g:T}}(t)$, I_k, is composed of boundary points $x_{k_1} + x_{k_2}$, where $k_1 = 1, \ldots, n_1$ and $k_2 = 1, \ldots, n_2$ - all possible combinations of I_{k_1} and I_{k_2}. The domain of the convolution approximation $h_{E_{g:T}}(t)$ is:

$$[\min_{k_1, k_2}(x_{k_1} + x_{k_2}), \max_{k_1, k_2}(x_{k_1} + x_{k_2})].$$

Let $h_{E_{g:T}}(t)_k$ represent the form of the $h_{E_{g:T}}(t)$ in the kth subinterval of I_k -

$$h_{E_{g:T}}(t)_k = (h_{E_{g:T-1}} \oplus h_{E_{g_T}})(t)_k$$

$$= \sum_{J_{k_1}, J_{k_2}} \int_{lower}^{upper} h_{E_{g_T}}(t-x)_{k_2} h_{E_{g:T-1}}(x)_{k_1} dx. \tag{90}$$

Here $J_{k_1} \subset I_{k_1}$ and $J_{k_2} \subset I_{k_2}$ such that J_{k_1} and J_{k_2} are valid sets of subintervals contributing to the kth subinterval of I_k, i.e., $[x_k, x_{k+1}]$. The conditions for the k_1st subinterval of I_{k_1} and k_2nd subinterval of I_{k_2} contributing to the kth subinterval of I_k are:

$$x_k \geq x_{k_1} + x_{k_2} \tag{91}$$

$$x_{k+1} \leq x_{k_1+1} + x_{k_2+1} \tag{92}$$

Conditions 91 and 92 may be written for any $x \in [x_k, x_{k+1}]$ as:

$$x_{k_1} \leq x \leq x_{k_1+1} \quad \text{and} \quad x_{k_2} \leq t - x \leq x_{k_2+1}$$

$$x_{k_1} \leq x \leq x_{k_1+1} \quad \text{and} \quad t - x_{k_2+1} \leq x \leq t - x_{k_2}.$$

The above expression simplifies to:

$$\max(x_{k_1}, t - x_{k_2+1}) \leq x \leq \min(x_{k_1+1}, t - x_{k_2}). \tag{93}$$

From equation 93, the limits of integration in equation 90 are:

$$upper = \min(x_{k_1+1}, t - x_{k_2}) \tag{94}$$

$$lower = \max(x_{k_1}, t - x_{k_2+1}) \tag{95}$$

for each subinterval combination $[x_{k_1}, x_{k_1+1}] \in J_{k_1}$ and $[x_{k_2}, x_{k_2+1}] \in J_{k_2}$ contributing to $[x_k, x_{k+1}] \in I_k$.

In equation 90, the subexpression $\int_{lower}^{upper} h_{E_{g_T}}(t-x)_{k_2} h_{E_{g:T-1}}(x)_{k_1} dx$ can be written in the following matrix form:

$$\int_{lower}^{upper} \mathbf{c}^{(k_2)} (\varphi^{(k_2)}(t-x))^T \mathbf{c}^{(k_1)} (\varphi^{(k_1)}(x))^T dx$$

where

- $\mathbf{c}^{(k_1)}$ is the $m+1$-dimension polynomial coefficient vector and $\varphi^{(k_1)}$ is the $m+1$-dimension polynomial basis function vector for the k_1st interval of I_{k_1} and

- $\mathbf{c}^{(k_2)}$ is the $m+1$-dimension polynomial coefficient vector and $\varphi^{(k_2)}$ is the $m+1$-dimension polynomial basis function vector for the k_2nd interval of I_{k_2}.

Assuming a monic piecewise polynomial representation, the above expression can be written as:

$$
\begin{aligned}
\int_{lower}^{upper} h_{E_{g_T}}(t-x)_{k_2} h_{E_{g:T-1}}(x)_{k_1} dx &= \int_{lower}^{upper} \mathbf{c}^{(k_1)} (\mathbf{x}^T \mathbf{x})(\mathbf{C}^{(k_2)})^T \mathbf{t}^T dx \\
&= (\mathbf{c}^{(k_1)} \mathbf{X} (\mathbf{C}^{(k_2)})^T) \mathbf{t}^T \big|_{lower}^{upper} \qquad (96)
\end{aligned}
$$

where

$$
\mathbf{X} = \begin{bmatrix}
x & \frac{x^2}{2} & \cdots & \frac{x^{m+1}}{m+1} \\
\frac{x^2}{2} & \frac{x^3}{3} & \cdots & \frac{x^{m+2}}{m+2} \\
\vdots & \vdots & & \vdots \\
\frac{x^{m+1}}{m+1} & \frac{x^{m+2}}{m+2} & \cdots & \frac{x^{2m+1}}{2m+1}
\end{bmatrix} ;
$$

$\mathbf{C}^{(k_2)}$ is an upper triangular matrix of the form

$$
\mathbf{C}^{(k_2)} = \begin{bmatrix}
\binom{0}{0} c_0^{(k_2)} & -\binom{1}{1} c_1^{(k_2)} & \cdots & (-1)^{m-1} \binom{m-1}{m-1} c_{m-1}^{(k_2)} & (-1)^m \binom{m}{m} c_m^{(k_2)} \\
\binom{1}{0} c_1^{(k_2)} & -\binom{2}{1} c_2^{(k_2)} & \cdots & (-1)^{m-1} \binom{m}{m-1} m c_m^{(k_2)} & \\
\vdots & \vdots & \ddots & & \\
\vdots & -\binom{m}{1} c_m^{(k_2)} & & \mathbf{0} & \\
\binom{m}{0} c_m^{(k_2)} & & & &
\end{bmatrix} ;
$$

and $t = (1, t, \ldots, t^m)$. When either the *upper* or *lower* limits of integration are a constant, $d = x_{k_1+1}$ or x_{k_1} respectively, the evaluation at that limit is obtained by simply substituting d into the matrix $\mathbf{X}$ in equation 96. In this case, the convolution segment $h_{E_{g:T}}(t)_k$ maintains the same polynomial degree m as each of the convoluted piecewise segments $h_{E_{g:T-1}}(x)_{k_1}$ and $h_{E_{g_T}}(t-x)_{k_2}$. However, when an *upper* or *lower* limit is of the form $t - d$, $t - d = t - x_{k_2}$ or $t - x_{k_2+1}$, the degree of convolution increases to $2m + 1$.

When an *upper* or *lower* limit is of the form $x = t - d$, the matrix expression $(\mathbf{c}^{(k_1)}\mathbf{X}(\mathbf{C}^{(k_2)})^T)\mathbf{t}^T$ in equation equation 96 may be reformulated as:

$$\mathbf{c}^{(k_1)}(\mathbf{c}^{(k_2)}\mathbf{X'T})^T \tag{97}$$

where $\mathbf{X'}$ is a $m + 1$ x $2(m + 1)$ matrix of the form:

$$\mathbf{X'} = \left[\begin{array}{cccc|c} (-1)^{(m+1)}\frac{d^{m+1}}{m+1} & (-1)^{(m)}\frac{d^m}{m} & \cdots & -d \\ (-1)^{(m+1)}\frac{d^{m+2}}{m+2} & (-1)^{(m)}\frac{d^{m+1}}{m+1} & \cdots & -\frac{d^2}{2} & \mathbf{I} \\ \vdots & \vdots & & \vdots \\ (-1)^{(m+1)}\frac{d^{2m+1}}{2m+1} & (-1)^{(m)}\frac{d^{2m}}{2m} & \cdots & -\frac{d^{m+1}}{m+1} \end{array} \right]$$

with $\mathbf{I}$ being a $(m+1)$ x $(m+1)$ identity matrix and $\mathbf{T}$ is an $2(m+1)$ x $m+1$ matrix of the form:

$$\mathbf{T} = \left[\begin{array}{ccccc}
 & & & & 1 \\
 & \mathbf{0} & & 1 & \binom{m}{1}t \\
 & & \ddots & \binom{m-1}{1}t & \vdots \\
 & 1 & \ddots & \vdots & \vdots \\
1 & \binom{1}{1}t & \cdots & \binom{m-1}{m-1}t^{m-1} & \binom{m}{m}t^m \\
\hline
\dfrac{t}{\binom{0}{0}} & \dfrac{t^2}{2\binom{1}{1}} & \cdots & \cdots & \dfrac{t^{m+1}}{m+1\binom{m}{m}} \\[2ex]
\dfrac{t^2}{2\binom{1}{0}} & \dfrac{t^3}{3\binom{2}{1}} & \cdots & \cdots & \dfrac{t^{m+2}}{m+2\binom{m+1}{m}} \\[2ex]
\vdots & \vdots & & & \vdots \\[1ex]
\dfrac{t^{m+1}}{m+1\binom{m}{0}} & \dfrac{t^{m+2}}{m+2\binom{m+1}{1}} & \cdots & \cdots & \dfrac{t^{2m+1}}{2m+1\binom{2m}{m}}
\end{array}\right]$$

with $\mathbf{0}$ being an upper triangular matrix of zeros.

As an example of the above formulation, consider the convolution of two truncated normal density functions from Lawrence [128]. Suppose:

$$f_{E_{g_T}} \sim N(7, 2) \qquad \text{truncated on } [1, 13]$$
$$f_{E_{g:T-1}} \sim N(13, 1) \quad \text{truncated on } [10, 16].$$

The final form of this example convolution is the truncated normal density function:

$$(f_{E_{g:T-1}} \oplus f_{E_{g_T}}) \sim N(20, \sqrt{5}) \text{ truncated on } [11, 29].$$

Figure 26 shows the graph of this convolution with the linear, quadratic, and cubic monic piecewise polynomial techniques respectively using Single Value Decomposition matrix inversion. Inspecting the graphs visually, the monic piecewise

Polynomial Approximation	Error Bound
Linear	$[-0.000377, 0.0005174]$
Quadratic	$[-0.000800, 0.00055261]$
Cubic	$[-0.000808, 0.000551]$

Table 10.   Error Bounds for Piecewise Polynomial (Expanded) Approximations of the Truncated Normal Convolution Example

polynomial techniques appear to fit the truncated normal density function convolution adequately.

As with the results shown for multiplication stochastic operation, the approximations shown in Figure 26 consist of "expanded" piecewise polynomial fits that are "reduced" in degree in order to manage the problem of "exploding coefficients". Figure 27 shows the convolution error associated with the linear, quadratic, and cubic piecewise polynomial approximations for the "expanded" representation. Note that the cubic piecewise polynomial approximation starts to oscillate over the region $[24, 29]$. This is a direct result of the relatively high order of the "expanded" convolution; in this case the order is $m = 7$. In general, from equation 97, the order of the "expanded" convolution is $2m + 1$. This oscillation problem makes *applying PART above $m > 3$ problematic* for EON solutions. Additionally as Figure 27 and Table 10 show, the linear piecewise polynomial (expanded) approximation has less error over regions of the convolution then the quadratic and cubic (expanded) approximations (e.g. the peak of the distribution). This result is unique to this example and is not a general trend.

In order to quantitatively determine the adequacy of the "reduced" representation in Figure 26, Figure 28 shows graphs of the (pointwise) error for the three approximation techniques. All error graphs appear similar in shape having an oscillatory nature. Table 11 shows the error bounds for these graphs. As with the stochastic multiplication operation, a key concern with the convolution operation is the potential for error buildup - i.e., approximating two density functions and then
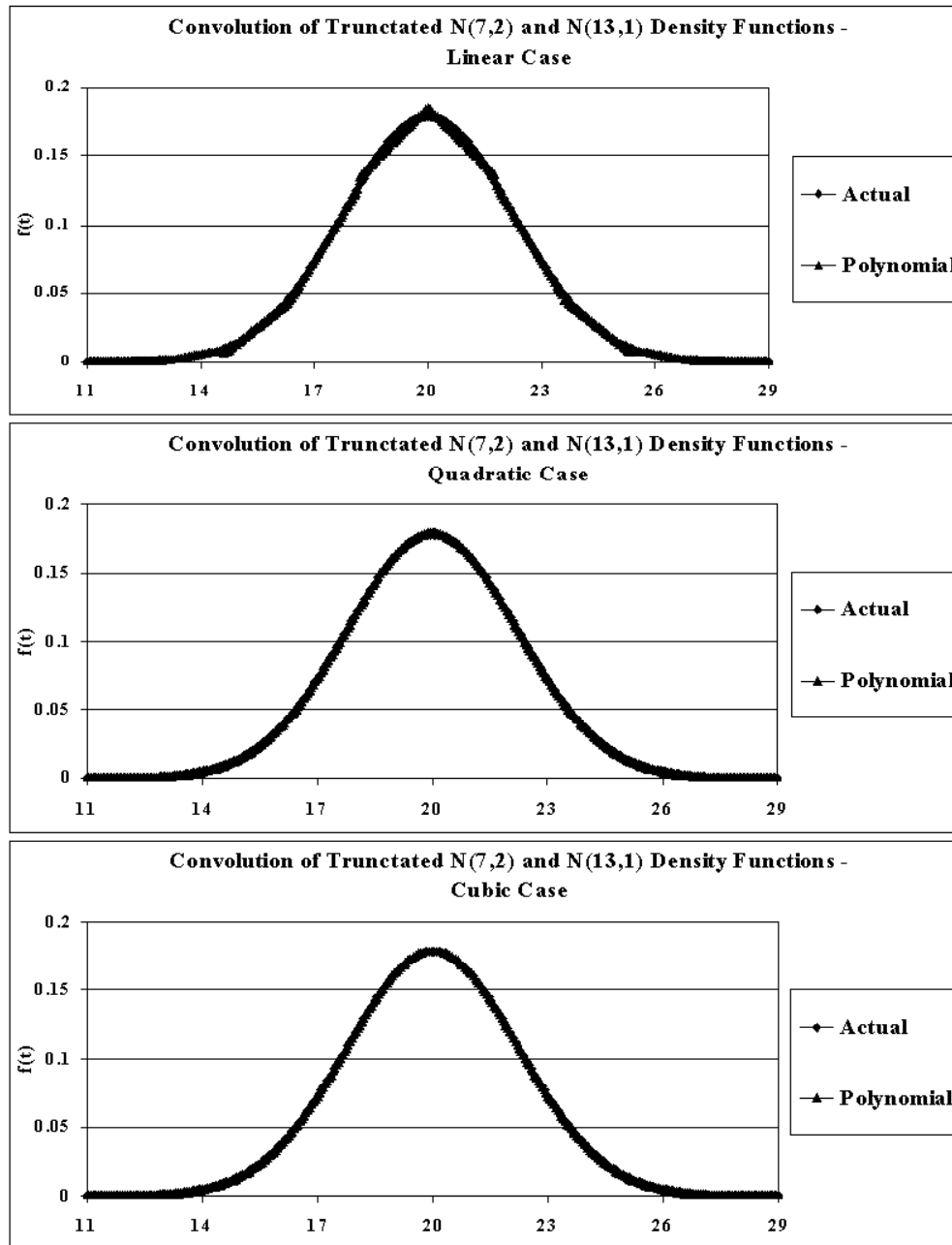
Figure 26. Graph of the Convolution of Two Truncated Normal Density Functions
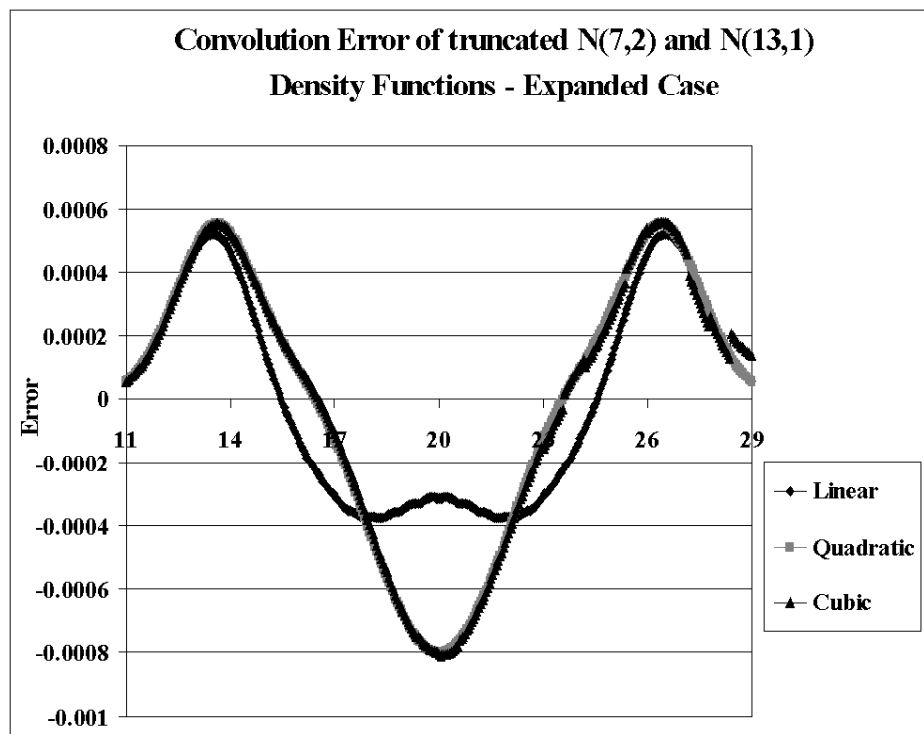
Figure 27. Graph of the Convolution Error of Two Truncated Normal Density Functions - Expanded Case

Polynomial Approximation	Error Bound
Linear	$[-0.00711, 0.00426]$
Quadratic	$[-0.00136, 0.000640]$
Cubic	$[-0.000845, 0.000555]$

Table 11.   Error Bounds for Piecewise Polynomial (Reduced) Approximations of the Truncated Normal Convolution Example

using those approximations to approximate the convolution of the density functions. Note from Tables 7 and 11 that the order of the "reduced" convolution error is similar in magnitude to the error in approximating the truncated normal density function in Figure 22. Although an additional order of precision is not gained with each increase in the order of the polynomial approximation (unlike the truncated normal density function approximation), the least squares convolution approximations have the advantageous effect of canceling positive and negative errors in the individual density function approximations.

As a final observation on this example, Figure 29 shows the pointwise error of Lawrence's single degree approximation and the author's. Note that the author's approximation has less pointwise error on average than Lawrence's. This is due to a formation error in Lawrence's solution. The correct representation on [128], page 115 for $c = bf_i$ or bf_{i+1} should be:

$$[f_0^i g_0^j c + \frac{1}{2}(f_1^i g_0^j - f_0^i g_1^j)c^2 - \frac{1}{3}f_1^i g_1^j c^3] + (f_0^i g_1^j c + \frac{1}{2}f_1^i g_1^j c^2)t.$$

Of interest to the reader and a potential area of further investigation for speeding up the convolution algorithm is the way in which algorithm identifies the valid sets of density function subintervals J_{k_1} and J_{k_2} contributing to the kth subinterval I_k of the convolution. Martin [134] choose to index his convolution algorithm by convolution subinterval I_k while Lawrence [128] indexed by density function subinterval I_{k_1}. Specifically, Martin choose I_k and then searched for the first subintervals

176
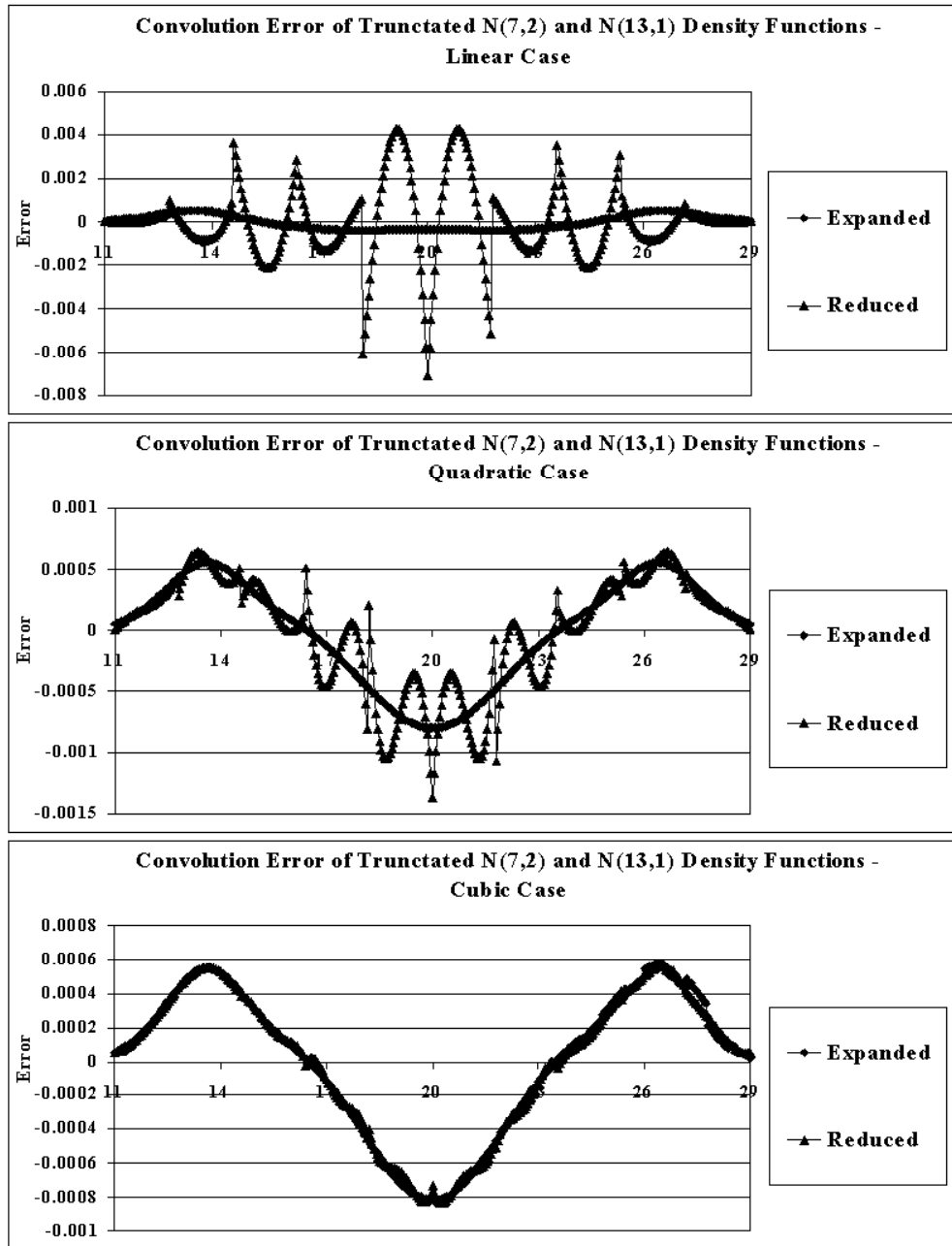
Figure 28. Graph of the Convolution Error of Two Truncated Normal Density
Functions - Reduced Case
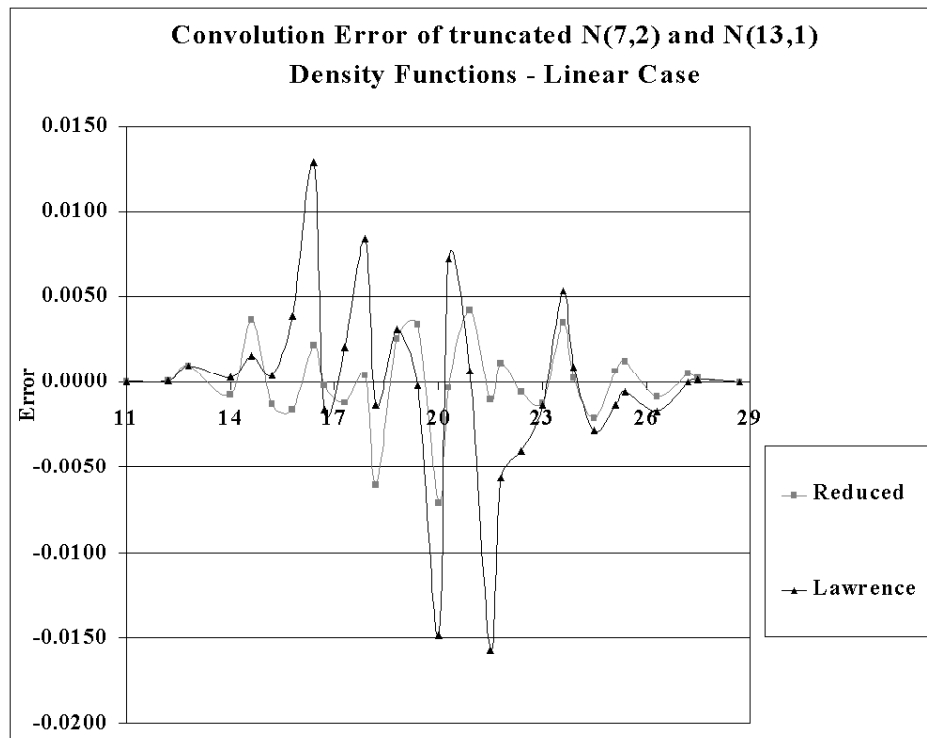
Figure 29.   Graph of the Convolution Error of Two Truncated Normal Density
Functions using Lawrence's algorithm

of I_{k_1} and I_{k_2} to form I_k. Lawrence first found the subinterval J_{k_1} of the density function $f_{E_{g_T}}$ that contributes to the I_k subinterval and then the corresponding subintervals J_{k_2} of density function $f_{E_{g:T-1}}$ that contribute to I_k. The author has chosen Lawrence's indexing framework based on his implementation with a reduction routine with the understanding that it has not been validated to be the most efficient.

As another example of the convolution stochastic operation, consider the convolution of two shifted exponential density functions. Specifically, let

$$f_{E_{g_T}} = \tfrac{1}{2}\exp(-\tfrac{1}{2}(t-1)), \qquad 1 \le t < \infty$$
$$f_{E_{g:T-1}} = \tfrac{1}{5}\exp(-\tfrac{1}{5}(t-20)), \qquad 20 \le t < \infty.$$

The final form of this convolution is the density function:

$$(f_{E_{g:T-1}} \oplus f_{E_{g_T}}) =$$

$$\begin{cases} -\tfrac{1}{3}\exp(-0.2t+4.5)[\exp(-0.3(t-20))-\exp(-0.3)], & [21,30.21) \\ -\tfrac{1}{3}\exp(-0.2t+4.5)[\exp(-0.3(10.21))-\exp(-0.3)], & [30.21,44.03) \\ -\tfrac{1}{3}\exp(-0.2t+4.5)[\exp(-0.3(10.21))-\exp(-0.3(t-43.03))], & [43.03,53.23). \end{cases}$$

The above formulation is based on truncating each of the exponential density functions at 99% of their total area; i.e., $f_{E_{g_T}}$ is truncated on $[1, 10.21]$ and $f_{E_{g:T-1}}$ is truncated on $[20, 43.03]$. Figure 30 shows the graph of this convolution with the linear, quadratic, and cubic monic piecewise polynomial techniques respectively using SVD matrix inversion. Inspecting the graphs visually, the monic piecewise polynomial techniques appear to fit the convolution adequately for the quadratic and cubic cases; however the linear approximation has difficulty approximating the shape of the convolution in the interval $[21, 24]$. This is a direct result of each "reduced" approximation using only ten subintervals. The first "reduced" subinterval is $[21, 24.22]$ which is required to span five "expanded" approximation subintervals.
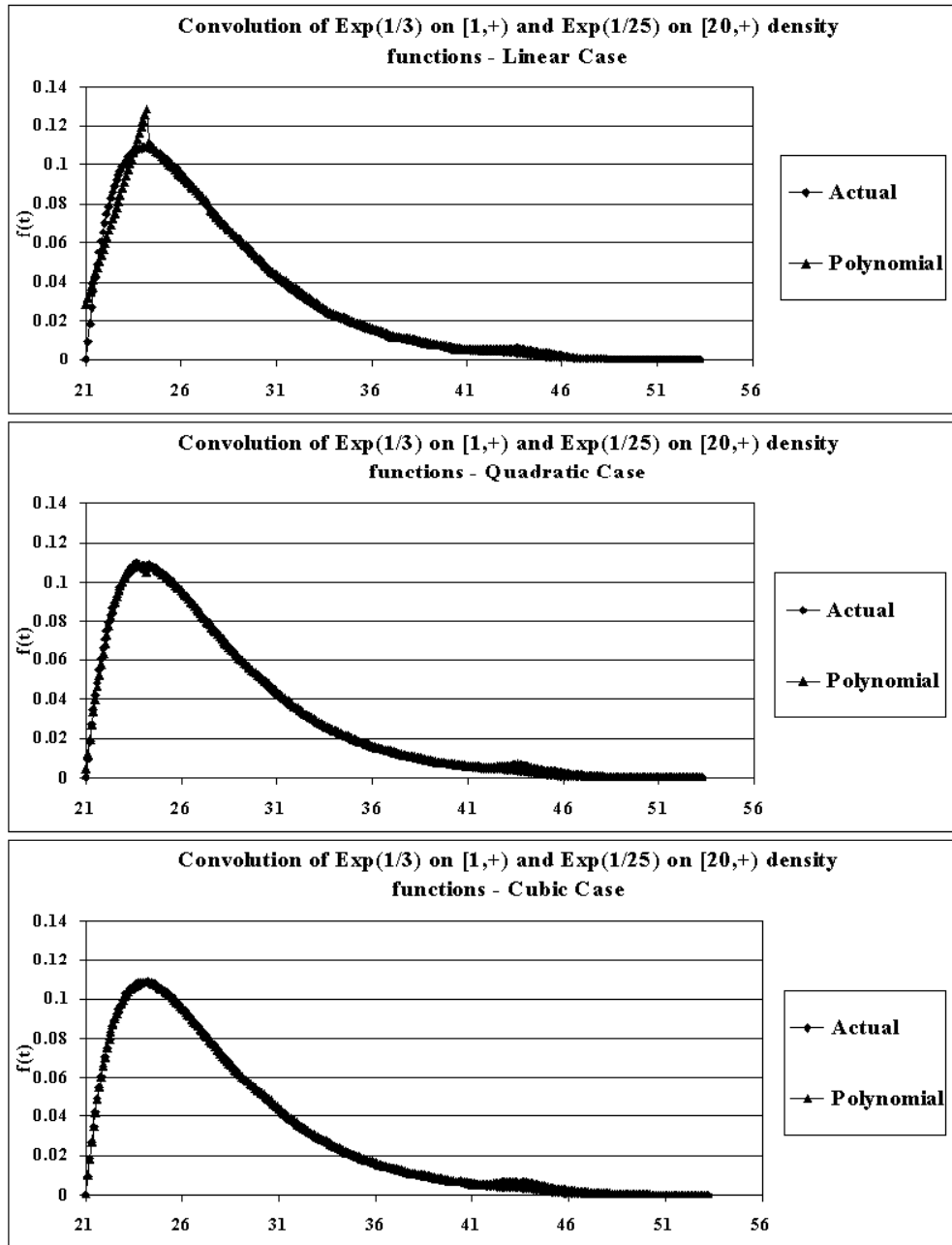
Figure 30. Graph of the Convolution of Two Truncated Exponential Density Functions
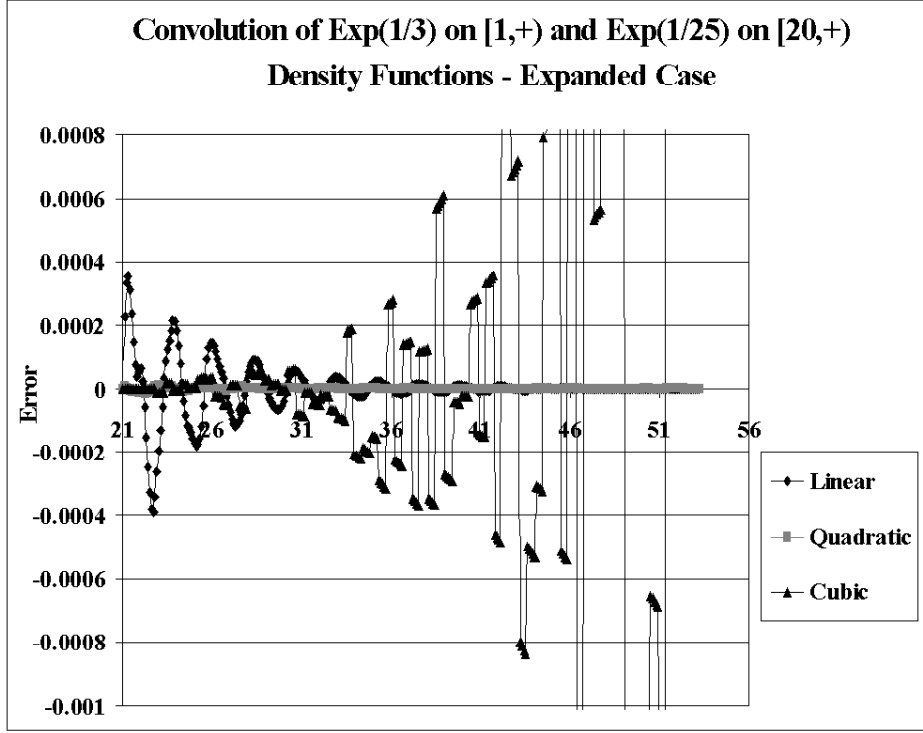
Figure 31.   Graph of the Convolution Error of Two Truncated Exponential Density
Functions - Expanded Case

Figure 31 shows the convolution error associated with the linear, quadratic, and
cubic piecewise polynomial approximations for the "expanded" representation. Note
that the cubic piecewise polynomial approximations significantly increase in error
over the region $[34, 53.23]$. As shown with the truncated normal convolution example,
this is a direct result of the relatively high order of the "expanded" convolution.
Coefficients c_0^k and c_7^k for the cubic "expanded" approximations are on the order of
$10^{-2} - 10^1$ and $10^{-12} - 10^{-8}$ respectively. Additionally as Figure 31 and Table 12
show, the cubic piecewise polynomial (expanded) approximation has an equivalent
error bound over the convolution region as the linear (expanded) approximation.

In order to quantitatively determine the adequacy of the "reduced" represen-
tation in Figure 30, Figure 32 shows graphs of the (pointwise) error for the three
approximation techniques. All error graphs appear similar in shape having a damped

Polynomial Approximation	Error Bound
Linear	$[-0.000308, 0.000352]$
Quadratic	$[-0.000308, 0.00000887]$
Cubic - Monic (SVD)	$[-0.00362, 0.00338]$

Table 12. Error Bounds for Piecewise Polynomial (Expanded) Approximations of the Truncated Exponential Convolution Example

Polynomial Approximation	Error Bound
Linear	$[-0.00287, 0.00140]$
Quadratic	$[-0.00453, 0.00333]$
Cubic	$[-0.000305, 0.000358]$

Table 13. Error Bounds for Piecewise Polynomial (Reduced) Approximations of the Truncated Exponential Convolution Example

oscillatory nature. Table 13 shows the error bounds for these graphs. From Tables 6 and 13, note that the order of the "reduced" convolution error is similar in magnitude to the error in approximating the exponential density function in Figure 21. Although an additional order of precision is not gained with each increase in the order of the polynomial approximation (unlike the exponential normal density function approximation), the least squares convolution approximations have the advantageous effect of canceling positive and negative errors in the individual density function approximations. Additionally, note that the "reduced", cubic representation has corrected the error associated with the "expanded", cubic representation - again this is a result of canceling positive and negative errors in the "expanded" representation.

6.7 Simple Air-to-Air Engagement Example

In section 4.3, a simple one versus one air-to-air engagement was detailed [103]. In this engagement, each aircraft could fire one radar-guided and one infrared missile at an opponent with the firing constraint that the radar-guided missile must always be fired before the infrared missile. Recall that random variable $E_{g_e^{(\kappa)}}$ represented
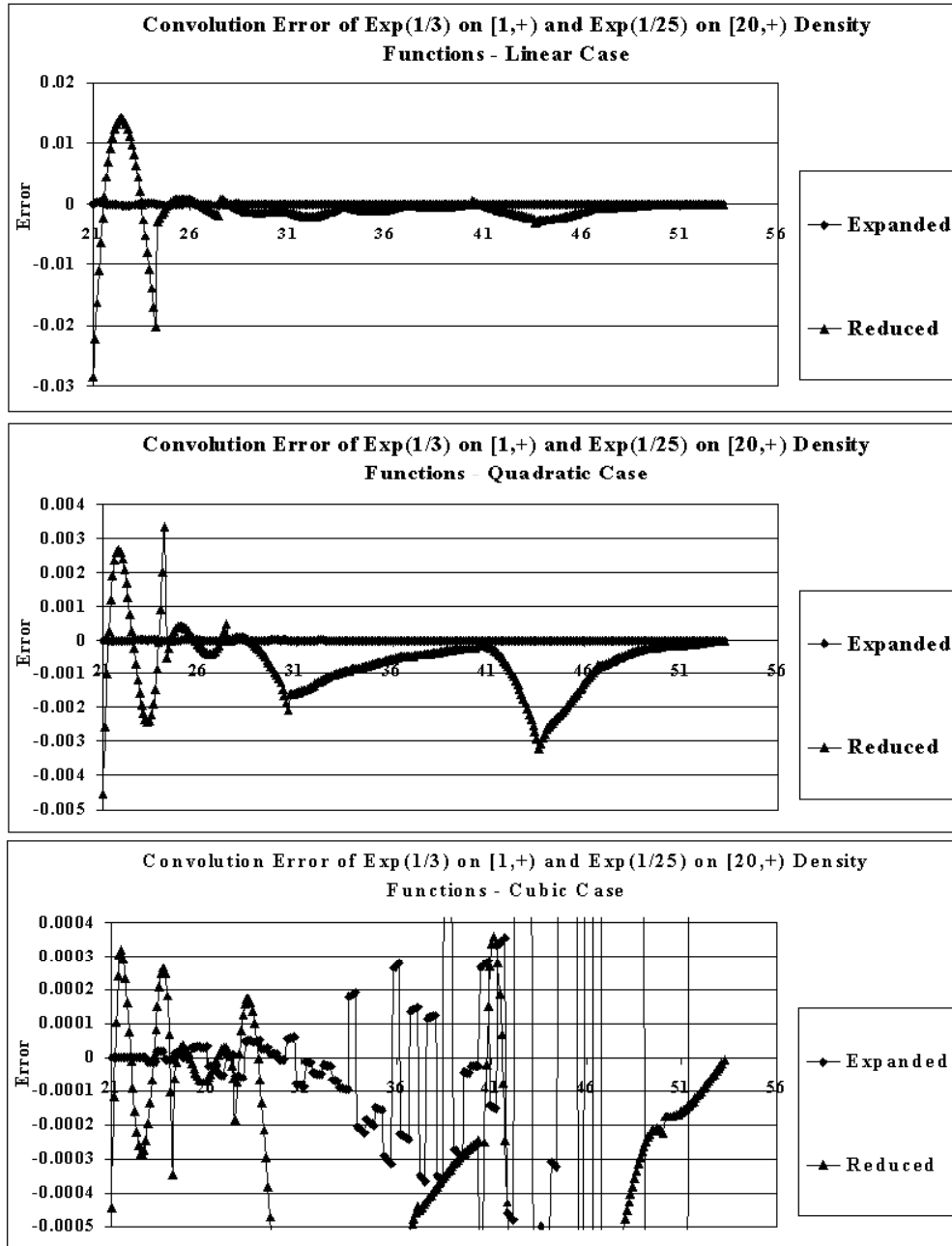
Figure 32. Graph of the Convolution Error of Two Truncated Exponential Density
Functions - Reduced Case

the completion time of event $G_{g_{e}^{(\kappa)}}$ where side $g = 1, 2$ fires a missile of type $e, e = 1, 2$ (where 1 - radar-guided and 2 - infrared) at side $\bar{g}$ that either misses ($\kappa = 0$) or shoots down ($\kappa = 1$) side $\bar{g}$ where $\bar{g}$ represents the opponent side of side g. The EON for this one versus one air-to-air engagement is shown in Chapter IV, Figure 17.

For this EON, specific event completion functions are:

- $E_{1_1^{(\kappa)}} \sim exp(\lambda_{1_1} t)$ with $\lambda_{1_1} = 1.25$ (i.e., side 1 fires a radar missile);

- $E_{2_1^{(\kappa)}} \sim exp(\lambda_{2_1} t)$ with $\lambda_{2_1} = 1.15$ (i.e., side 2 fires a radar missile);

- $E_{1_2^{(\kappa)}} \sim exp(\lambda_{1_2} t)$ with $\lambda_{1_2} = 1.85$ (i.e., side 1 fires an infrared missile); and

- $E_{2_2^{(\kappa)}} \sim exp(\lambda_{2_2} t)$ with $\lambda_{2_2} = 1.65$ (i.e., side 2 fires an infrared missile).

Additionally, the probabilities that the radar and infrared missiles on side 1 and 2 shoot down the other side when fired are $p_{1_1} = p_{2_1} = 0.65$ and $p_{1_2} = 0.85$, $p_{2_2} = 0.75$ respectively.

Implementing a generalized EON solution technique using the core concepts of the Polynomial Approximation and Reduction Technique, the probability of being in any state in the EON in Figure 17 was calculated. These calculations consisted of:

- fitting linear, quadratic, and cubic monic piecewise polynomials (using SVD matrix inversion) to the above density functions;

- determining the integral formation for each state probability; and

- solving the integral expression using the stochastic operations outlined in earlier sections of this chapter.

Each of the exponential density functions was truncated at 99% of its total area with each piecewise polynomial approximation determined using ten subintervals and $10 + I(classwith * 3)$ fitting points. In order to judge the accuracy of the EON solution technique, two comparisons where made:

Approximation	Error Bound
Linear	$[-0.003, 0.002]$
Quadratic	$[-0.00013, 0.00016]$
Cubic	$[-0.000007, 0.00014]$
Simulation	$[-0.007, 0.009]$

Table 14.   Error Bounds for Piecewise Polynomial Approximations and Simulation of $P_1(t)$

- one comparison with the exact (analytic) expressions for the probabilities as summarized in Appendix C and

- one comparison with a simulation of the state probabilities.

For this example, the simulation was run for $500,000$ iterations with a time increment (histogram) width of .025 minutes.

Consider the probability of being in transient state 1 at time t - $P_1(t)$. The left side of Figure 33 shows the graph of $P_1(t)$, the simulation of $P_1(t)$, and the monic, piecewise polynomial approximations (linear, quadratic, and cubic) respectively. Inspecting the approximations and simulation visually, all three piecewise polynomial fits and the simulation appear to fit $P_1(t)$ adequately.

In order to quantitatively determine the adequacy of the fit, the right side of Figure 33 shows graphs of the (pointwise) error. Table 14 shows the error bounds for these graphs. The error graphs exhibit damped oscillatory behavior that moves towards zero. In general, *the piecewise polynomial approximations provide greater accuracy than the simulation results.*

For event tier one, Figures 34 and 35 show the graphs of $P_2(t)$ (absorbing state) and $P_4(t)$ (transient state), the simulation of $P_2(t)$ and $P_4(t)$, and the monic piecewise polynomial approximations as well as the pointwise error for approximation techniques. As was the case for $P_1(t)$, all three piecewise polynomial fits and the simulation appear to fit the actual solution to $P_2(t)$ and $P_4(t)$ adequately,
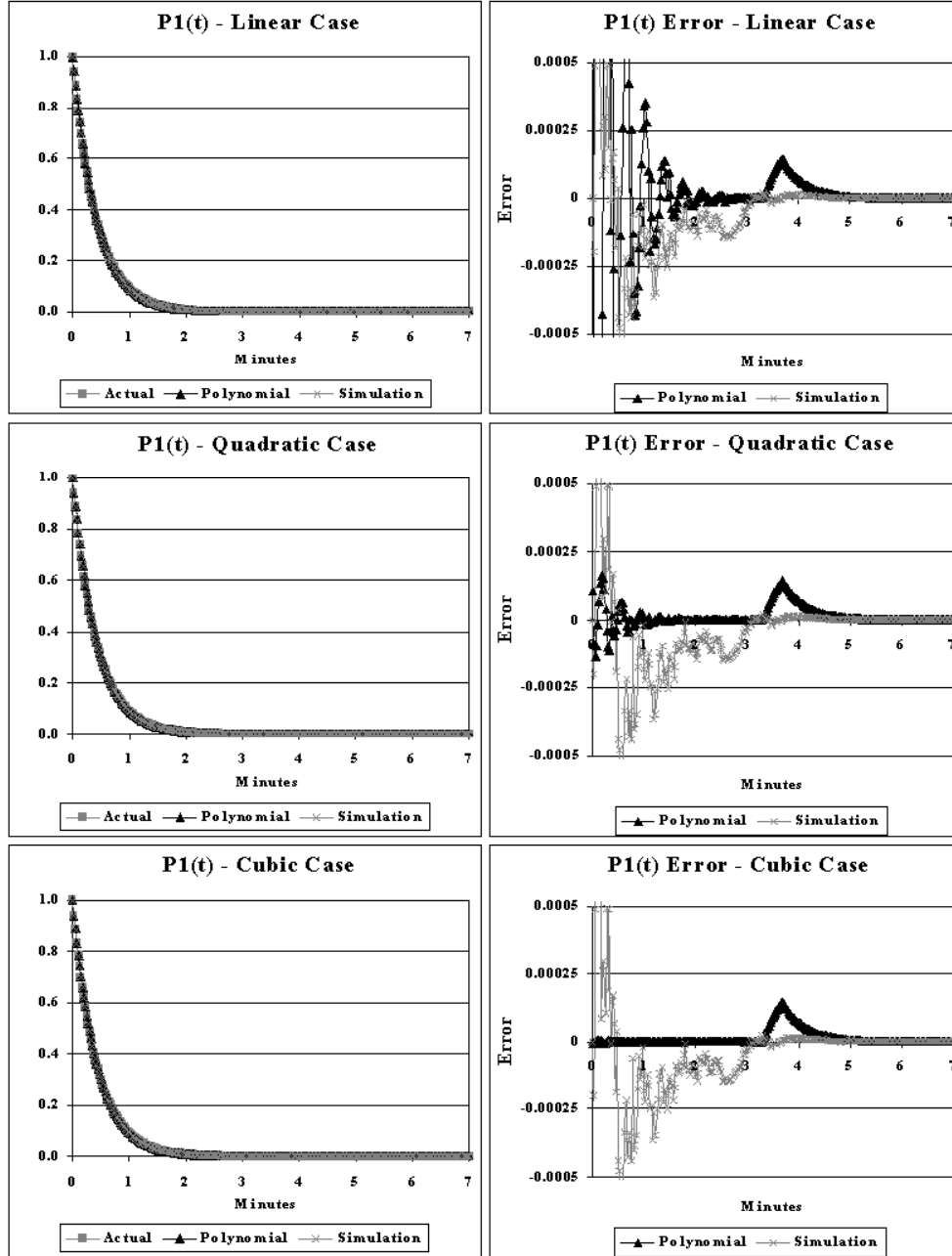
Figure 33.   Graph of $P_1(t)$ and Error for the Simple Air-to-Air Engagement Event
Occurrence Network
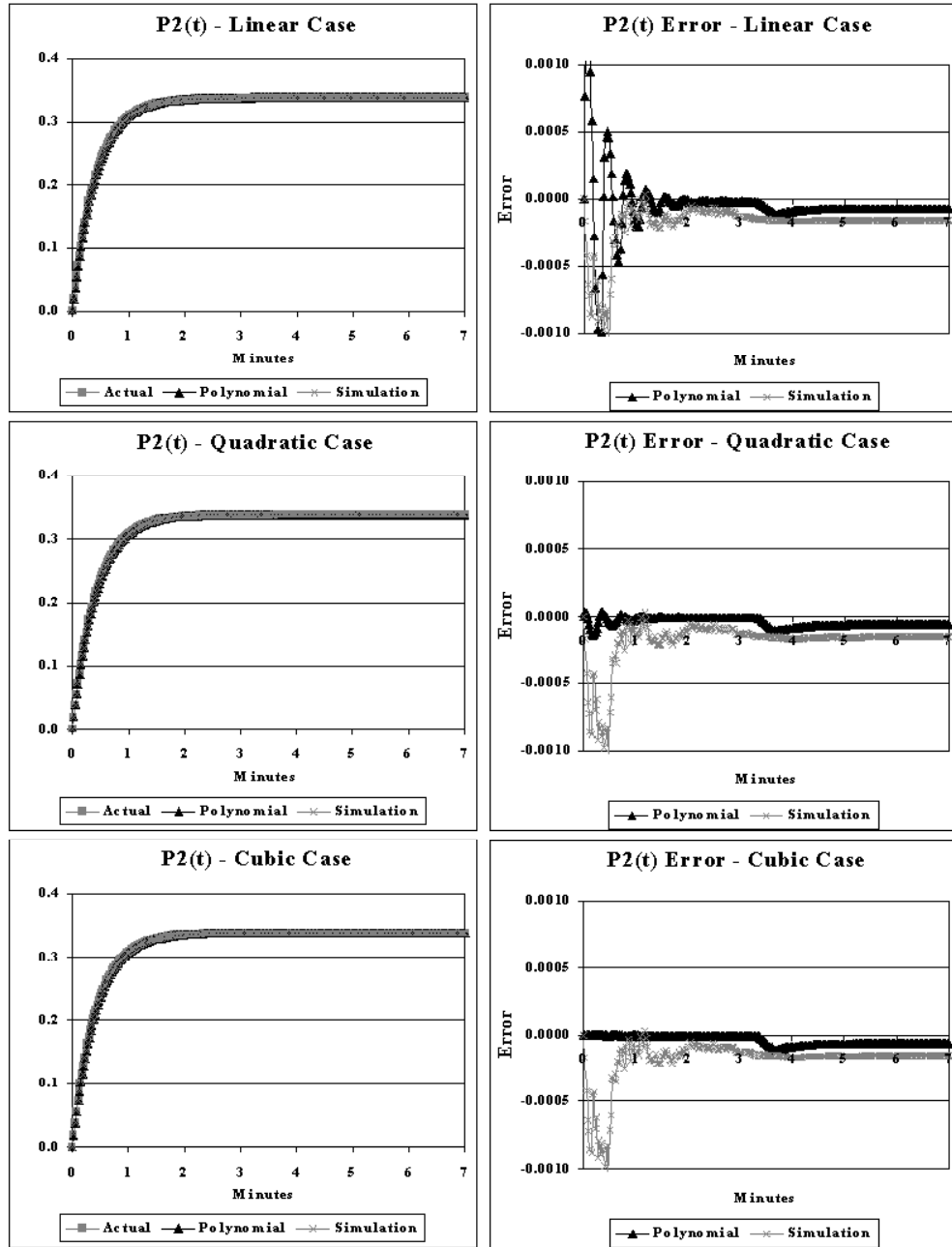
Figure 34. Graph of $P_2(t)$ and Error for the Simple Air-to-Air Engagement Event Occurrence Network
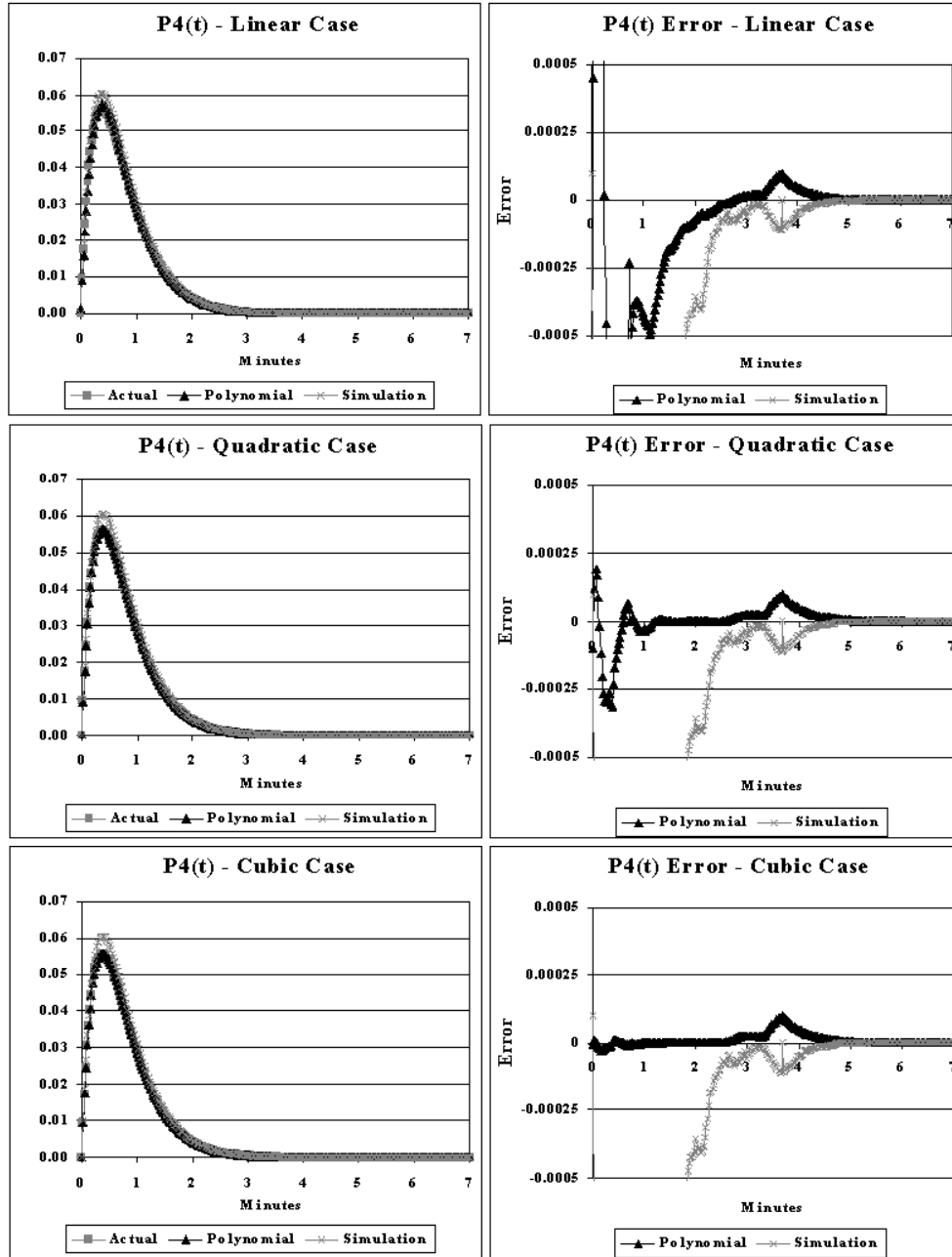
Figure 35. Graph of $P_4(t)$ and Error for the Simple Air-to-Air Engagement Event Occurrence Network

Approximation	Error Bound
Linear	$[-0.0012, 0.0013]$
Quadratic	$[-0.00015, 0.00003]$
Cubic	$[-0.00012, 0.000001]$
Simulation	[-0.001,0.00003]

Table 15. Error Bounds for Piecewise Polynomial Approximations and Simulation of $P_2(t)$

Approximation	Error Bound
Linear	$[-0.002, 0.0026]$
Quadratic	$[-0.00032, 0.00019]$
Cubic	$[-0.000026, 0.000098]$
Simulation	[-0.0045,0.000098]

Table 16. Error Bounds for Piecewise Polynomial Approximations and Simulation of $P_4(t)$

The right side of Figures 34 and 35 shows graphs of the (pointwise) error. Tables 15 and 16 show the error bounds for these graphs. In general for $P_2(t)$, the simulation results provide greater accuracy then the linear piecewise polynomial approximation while the quadratic and cubic piecewise polynomial approximations provide greater accuracy than the simulation results. For $P_4(t)$, the piecewise polynomial approximations provide greater accuracy than the simulation results.

Figures 36, 37, 38, and 39 show the pointwise error for various absorbing and transient states in event tiers two, three, and four. In general, the following observations hold:

- For absorbing states, the simulation results provide greater accuracy then the linear piecewise polynomial approximation while the quadratic and cubic piecewise polynomial approximations provide greater accuracy than the simulation results;

- For transient states, the piecewise polynomial approximations provide greater accuracy than the simulation results;

- The cubic piecewise polynomial approximation provides the best accuracy of all polynomial approximations; and

- The greatest percentage increase in accuracy occurs when moving from the linear piecewise polynomial approximation to the quadratic piecewise polynomial approximation.

Exceptions include:

- The approximations for transient state $P_9(t)$ where the simulation results provide greater accuracy then the linear piecewise polynomial approximation;

- The approximation for absorbing state $P_{16}(t)$ where the simulation results provide greater accuracy then both the linear and quadratic piecewise polynomial approximations; and

- The approximation for absorbing state $P_{20}(t)$ where the piecewise polynomial approximations provide greater accuracy than the simulation results.

As $t \rightarrow 7$ minutes, the probability of being at a particular node in the air-to-air engagement EON reaches a steady-state value. The following observations can be made about the steady-state probability of being at a particular node for this air-to-air engagement example using the cubic piecewise polynomial approximation:

- First strike radar missile kills occur with a probability of .65000000 ($P_2(7) + P_5(7)$) (versus .64999997 for the exact solution).

- Slightly over 90% of the air-to-air engagements have ended within two events (missile firings); $P_2(7) + P_5(7) + P_6(7) + P_8(7) + P_{10}(7) + P_{12}(7) = .909523$ versus .909525 for the exact solution.

- Only slightly more than .4% of the air-to-air engagements end without a shoot down; $P_{20}(t) = .0045877$ (versus .0045868 for the exact solution).

190
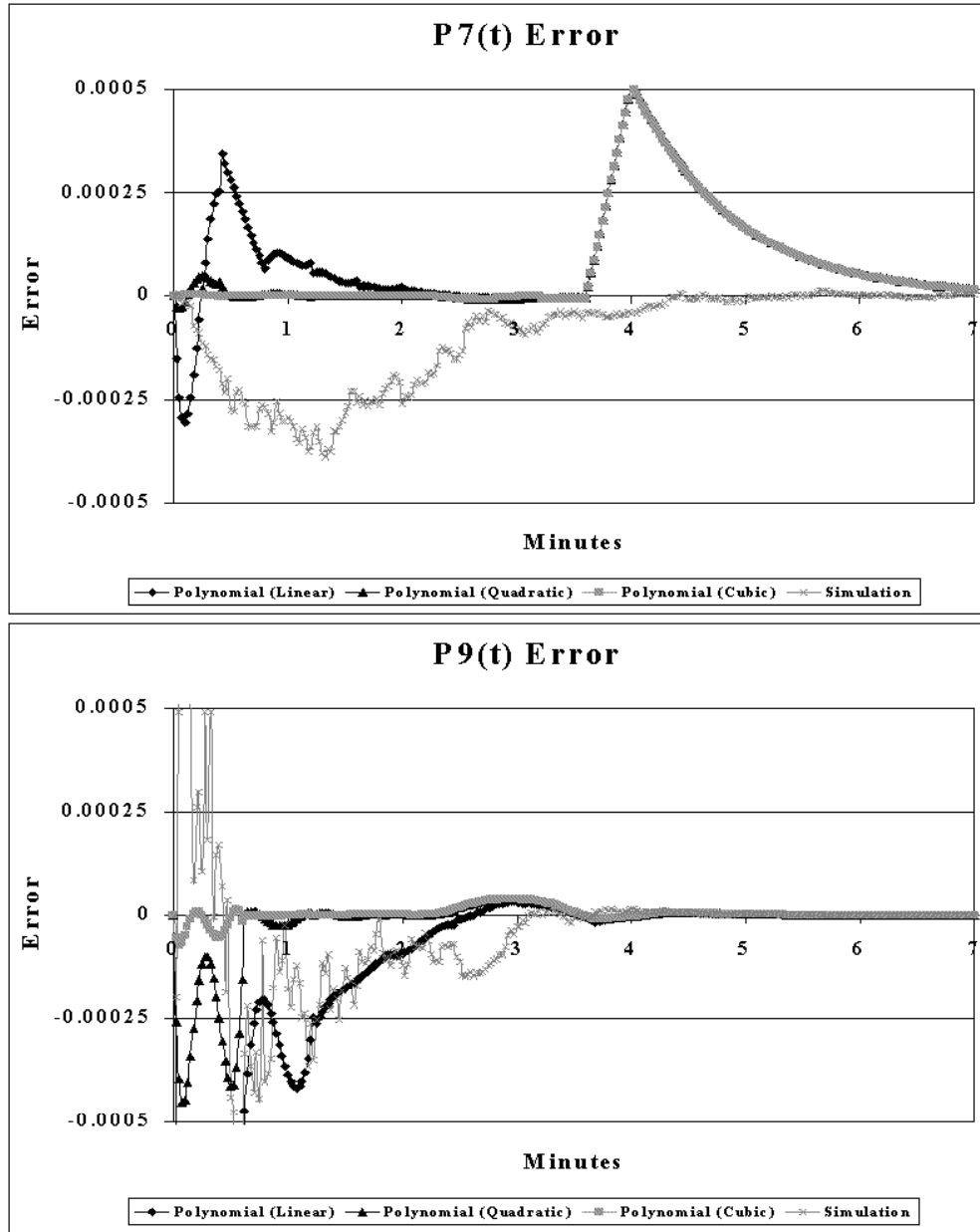
Figure 36.   Graph of $P_7(t)$ and $P_9(t)$ Error for the Simple Air-to-Air Engagement Event Occurrence Network
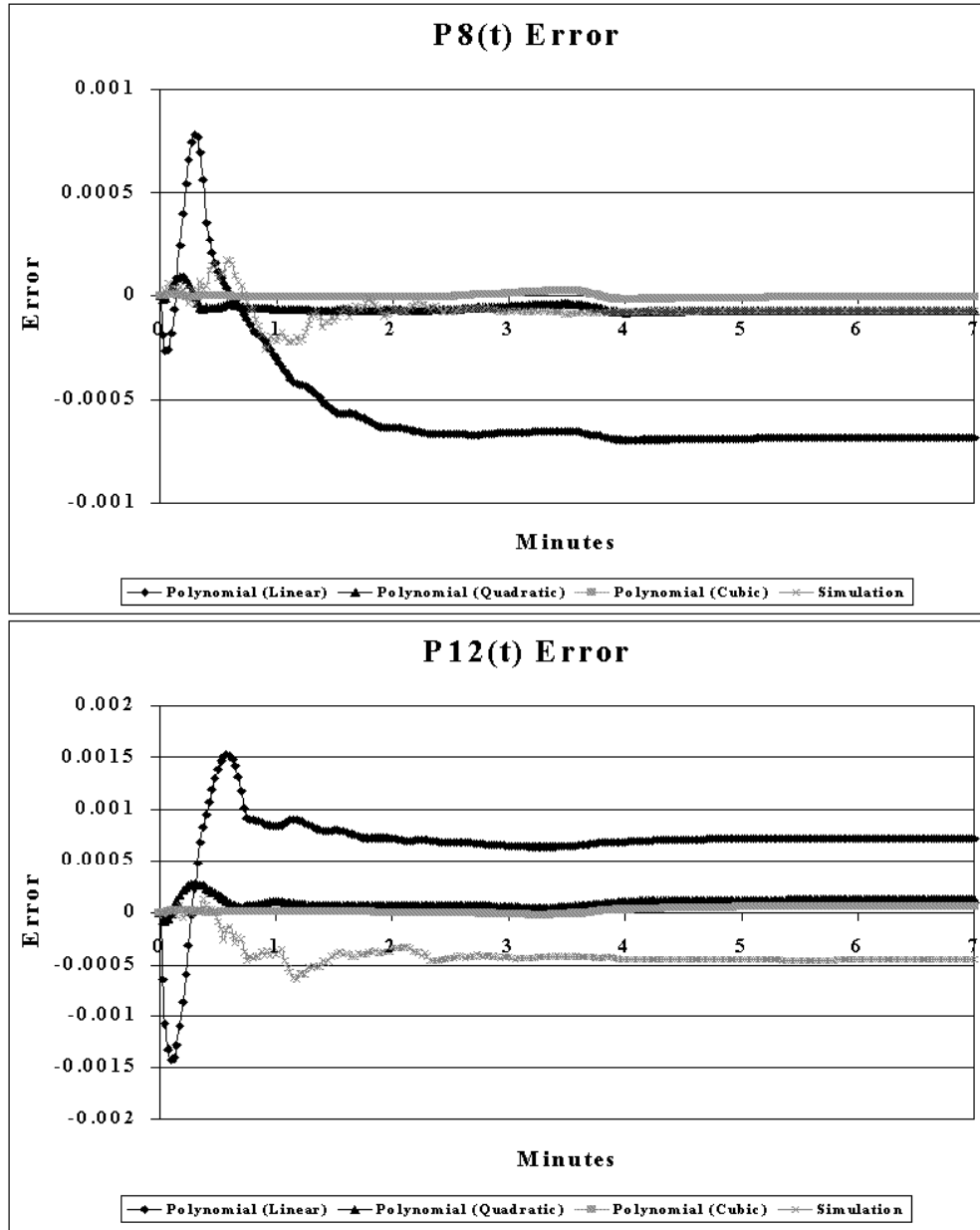
Figure 37. Graph of $P_8(t)$ and $P_{12}(t)$ Error for the Simple Air-to-Air Engagement Event Occurrence Network
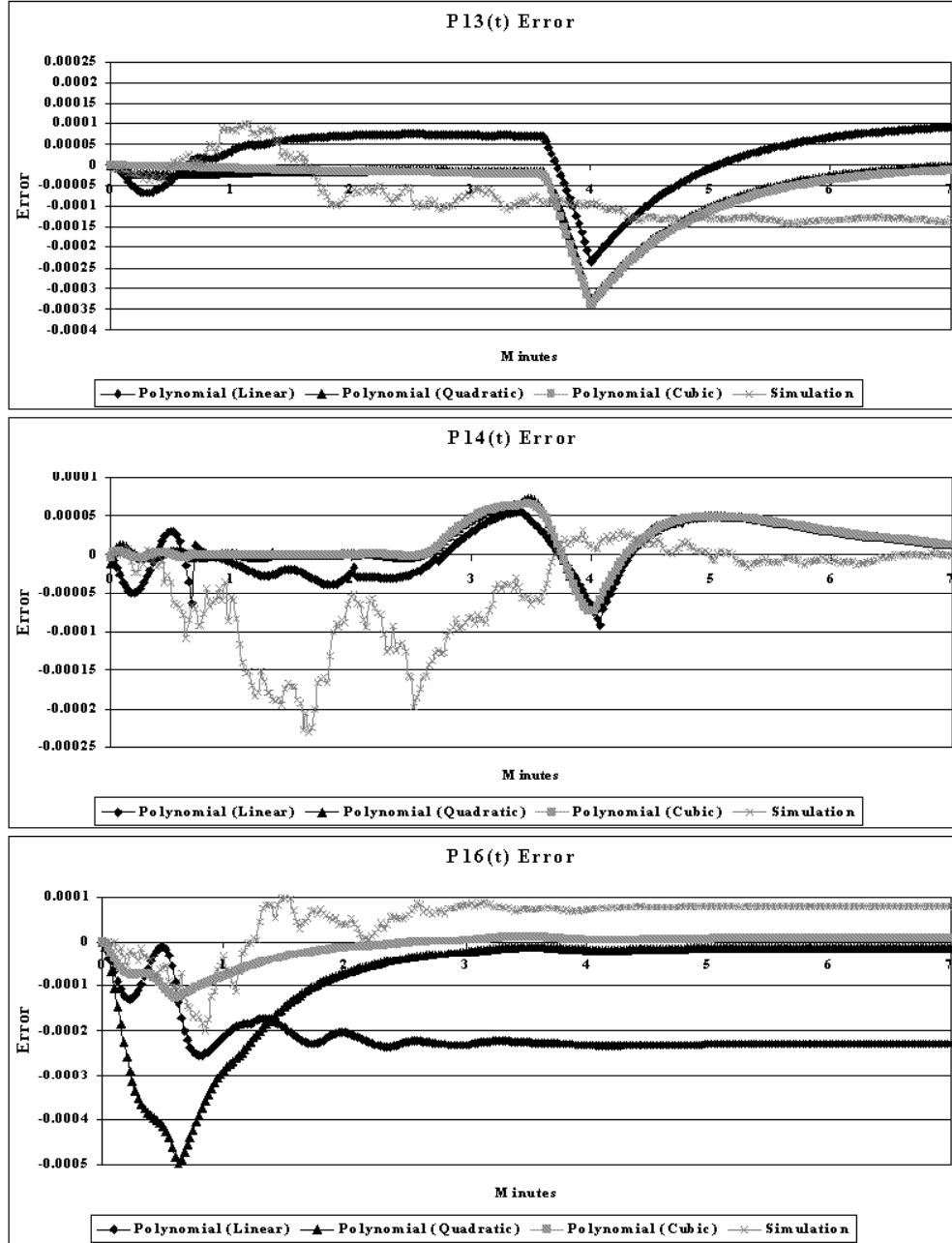
Figure 38.   Graph of $P_{13}(t)$, $P_{14}(t)$, and $P_{16}(t)$ Error for the Simple Air-to-Air Engagement Event Occurrence Network

193
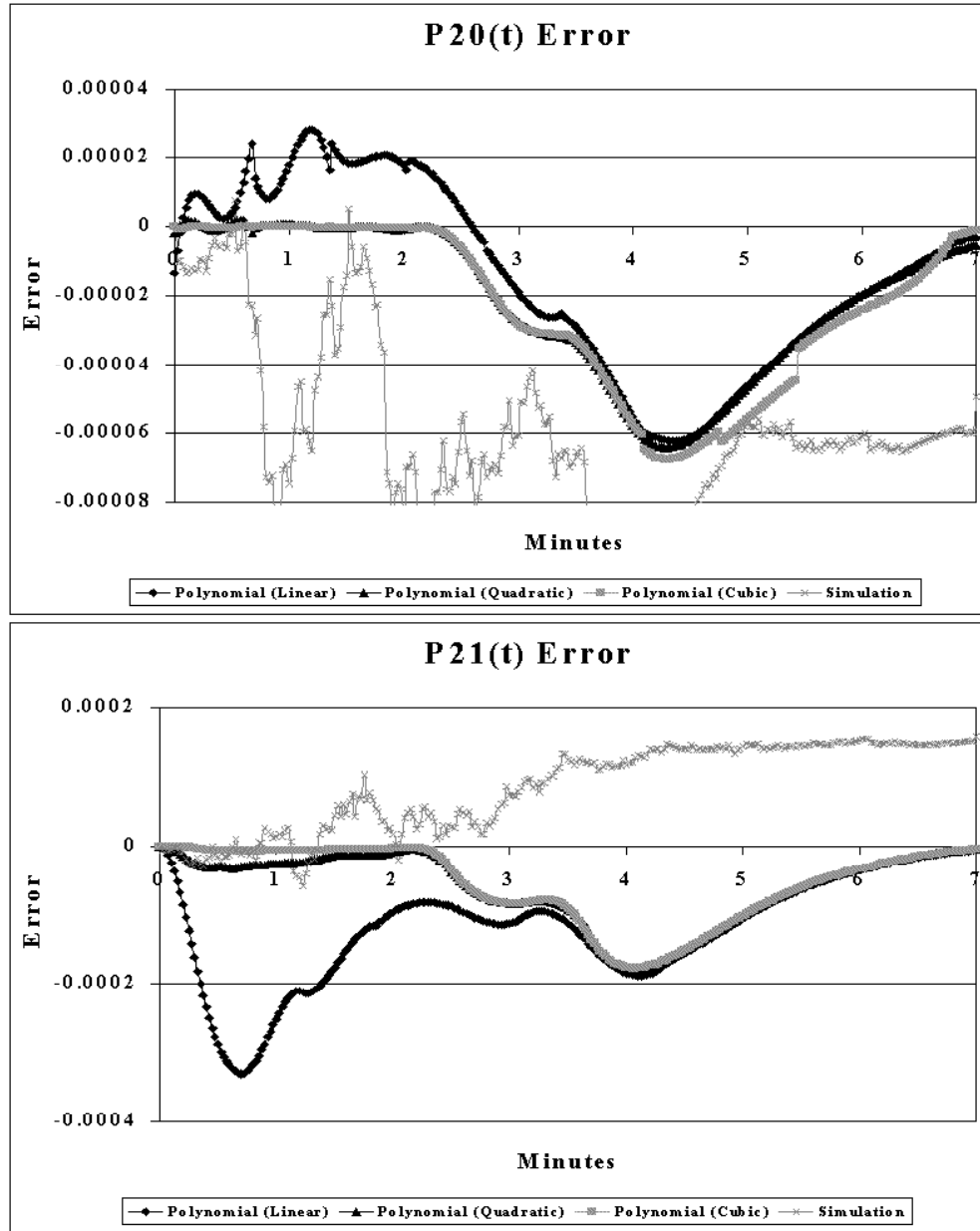
Figure 39. Graph of $P_{20}(t)$ and $P_{21}(t)$ Error for the Simple Air-to-Air Engagement Event Occurrence Network

194

- In answer to the example question posed above, the probability that the aircraft on side 1 shoots down the aircraft on side 2 with a radar missile is .42561 $(P_2(7) + P_{10}(7) + P_{18}(7))$ (versus .42561 for the exact solution - an agreement to five decimal places).

6.8   Large EON Air-to-Air Engagement

The large air-to-air example consists of a blue offensive counter air (OCA) mission. For a description of the terminology used in this section refer to Appendix A. As shown in figure 40, four blue aircraft are sweeping red airspace prior to a four-ship blue strike mission. Vectored by a blue airborne early warning aircraft, the four blue sweep aircraft are attempting to engage two red defensive combat air patrols (CAPs) - each CAP consists of a two-ship flight of red aircraft. Unbeknownst to the blue aircraft, two red surface-to-air missile (SAM) batteries wait to engage aircraft behind the red CAPs. The blue aircraft are loaded out with six radar missiles and two infrared missiles while the red aircraft are loaded out with four radar missiles and two infrared missiles. Each red SAM battery has eight SAMs.

The EON for this example consists of 10 event groups as follows:

- Event Groups 1-4: each blue aircraft with each of these event groups having *six* events

- Event Groups 5-8: each red aircraft with each of these event groups having *six* events; and

- Event Groups 9-10: each SAM battery with each of these event groups having *four* events

for a total of 56 events in this air-to-air example. Each blue aircraft is employing shoot-shoot tactics on its first shot opportunity firing two radar missiles. The next blue aircraft shot is a single radar missile followed by a single infrared missile if necessary. Additionally if needed, each blue aircraft will repeat the above sequence.
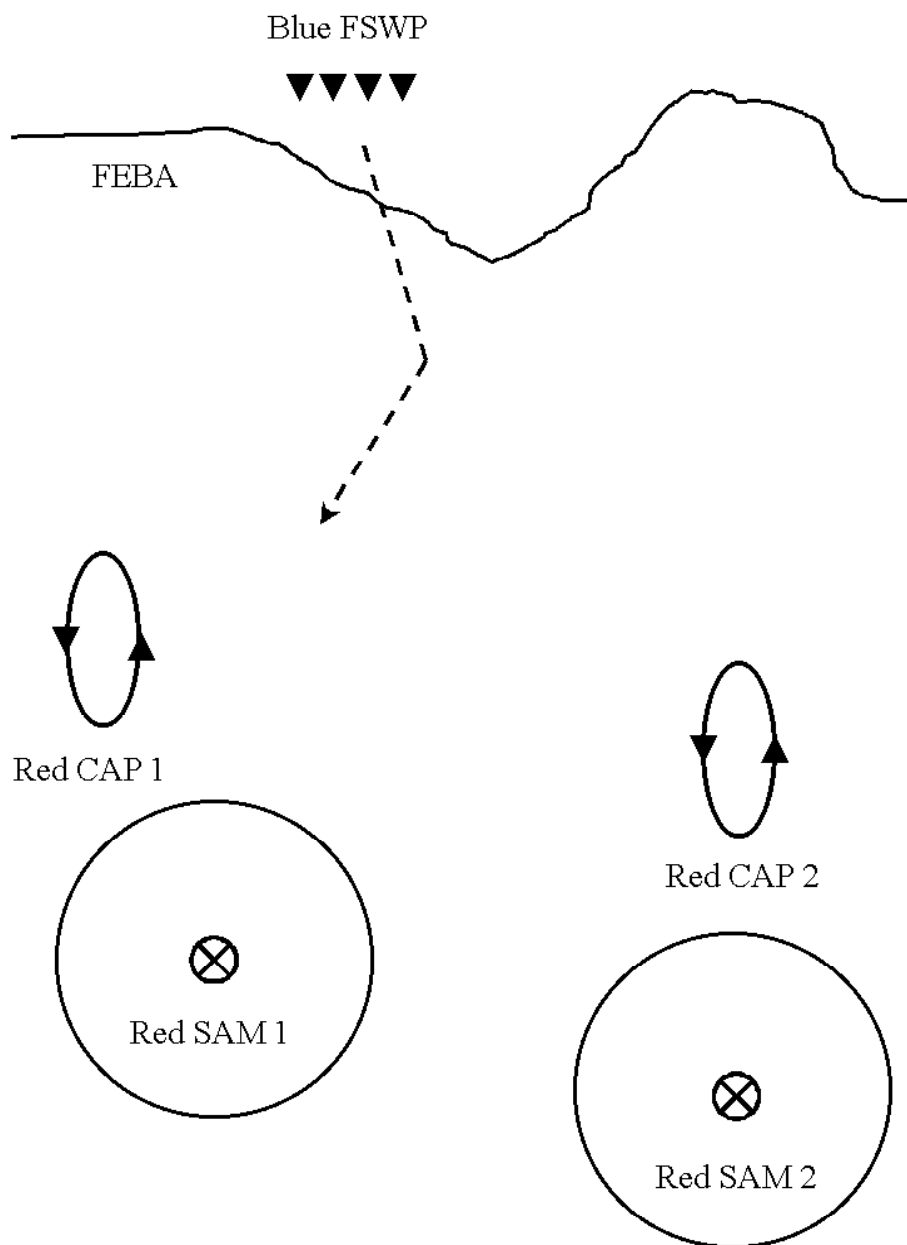
Figure 40. Graphical Depiction of Blue Offensive Counter Air (OCA) mission

The above represents the case where the blue flight engages the red flight through the merge (firing four missiles each); blows through the merge; turns to reengage; and engages the red aircraft again (firing an additional four missiles each). In contrast, the red aircraft initially employ a shoot-look-shoot strategy employing two radar missiles then a follow-on infrared missile shot. This sequence is repeated after the blue aircraft reengage. Finally, each red SAM battery employs shoot-shoot tactics firing two missiles per SAM event.

The event distribution and probability of success for these events are:

- Event Group 1

  - Event 1 - $E_{1_1^{(\kappa)}} \sim exp(.4)$, $p_{1_1} = .5$ - that is the first blue aircraft fires two radar missiles with an exponential distribution of occurrence ($\lambda = .4$), once fired the missiles have a probability of success of .5;

  - Event 2 - $E_{1_2^{(\kappa)}} \sim exp(1)$, $p_{1_2} = .4$;

  - Event 3 - $E_{1_3^{(\kappa)}} \sim$ (truncated) $N(1, .2)$ over $[0, 2]$, $p_{1_3} = .6$;

  - Event 4 - $E_{1_4^{(\kappa)}} \sim exp(.5)$, $p_{1_4} = .5$;

  - Event 5 - $E_{1_5^{(\kappa)}} \sim exp(1)$, $p_{1_5} = .4$; and

  - Event 6 - $E_{1_6^{(\kappa)}} \sim$ (truncated) $N(1, .2)$ over $[0, 2]$, $p_{1_6} = .6$.

- Event Group 2

  - Event 1 - $E_{2_1^{(\kappa)}} \sim exp(.41)$, $p_{2_1} = .5$;

  - Event 2 - $E_{2_2^{(\kappa)}} \sim exp(1.01)$, $p_{2_2} = .4$;

  - Event 3 - $E_{2_3^{(\kappa)}} \sim$ (truncated) $N(1, .2)$ over $[0, 2.1]$, $p_{2_3} = .6$;

  - Event 4 - $E_{2_4^{(\kappa)}} \sim exp(.51)$, $p_{2_4} = .5$;

  - Event 5 - $E_{2_5^{(\kappa)}} \sim exp(1.01)$, $p_{2_5} = .4$; and

  - Event 6 - $E_{2_6^{(\kappa)}} \sim$ (truncated) $N(1, .2)$ over $[0, 2.1]$, $p_{2_6} = .6$.

- Event Group 3

    - Event 1 - $E_{3_1^{(\kappa)}} \sim exp(.42)$, $p_{3_1} = .5$;

    - Event 2 - $E_{3_2^{(\kappa)}} \sim exp(1.02)$, $p_{3_2} = .4$;

    - Event 3 - $E_{3_3^{(\kappa)}} \sim$ (truncated) $N(1, .2)$ over $[0, 2.2]$, $p_{3_3} = .6$;

    - Event 4 - $E_{3_4^{(\kappa)}} \sim exp(.52)$, $p_{3_4} = .5$;

    - Event 5 - $E_{3_5^{(\kappa)}} \sim exp(1.02)$, $p_{3_5} = .4$; and

    - Event 6 - $E_{3_6^{(\kappa)}} \sim$ (truncated) $N(1, .2)$ over $[0, 2.2]$, $p_{3_6} = .6$.

- Event Group 4

    - Event 1 - $E_{4_1^{(\kappa)}} \sim exp(.43)$, $p_{4_1} = .5$;

    - Event 2 - $E_{4_2^{(\kappa)}} \sim exp(1.03)$, $p_{4_2} = .4$;

    - Event 3 - $E_{4_3^{(\kappa)}} \sim$ (truncated) $N(1, .2)$ over $[0, 2.3]$, $p_{4_3} = .6$;

    - Event 4 - $E_{4_4^{(\kappa)}} \sim exp(.53)$, $p_{4_4} = .5$;

    - Event 5 - $E_{4_5^{(\kappa)}} \sim exp(1.03)$, $p_{4_5} = .4$; and

    - Event 6 - $E_{4_6^{(\kappa)}} \sim$ (truncated) $N(1, .2)$ over $[0, 2.3]$, $p_{4_6} = .6$.

- Event Group 5

    - Event 1 - $E_{5_1^{(\kappa)}} \sim$ (truncated) $N(4, 3)$ over $[0, 8]$, $p_{5_1} = .4$;

    - Event 1 - $E_{5_2^{(\kappa)}} \sim$ (truncated) $N(1, .5)$ over $[0, 2]$, $p_{5_2} = .5$;

    - Event 1 - $E_{5_3^{(\kappa)}} \sim$ (truncated) $N(1, .1)$ over $[0, 2]$, $p_{5_3} = .9$;

    - Event 1 - $E_{5_4^{(\kappa)}} \sim$ (truncated) $N(6, 1)$ over $[0, 6]$, $p_{5_4} = .4$;

    - Event 1 - $E_{5_5^{(\kappa)}} \sim$ (truncated) $N(1, .5)$ over $[0, 2]$, $p_{5_5} = .5$; and

    - Event 1 - $E_{5_6^{(\kappa)}} \sim$ (truncated) $N(1, .1)$ over $[0, 2]$, $p_{5_6} = .9$.

- Event Group 6

    - Event 1 - $E_{6_1^{(\kappa)}} \sim$ (truncated) $N(4, 3)$ over $[0, 8.1]$, $p_{6_1} = .4$;

- Event 1 - $E_{6_2^{(\kappa)}} \sim$ (truncated) $N(1,.5)$ over $[0,2.1]$, $\ p_{6_2} = .5$;

- Event 1 - $E_{6_3^{(\kappa)}} \sim$ (truncated) $N(1,.1)$ over $[0,2.1]$, $\ p_{6_3} = .9$;

- Event 1 - $E_{6_4^{(\kappa)}} \sim$ (truncated) $N(6,1)$ over $[0,6.1]$, $\ p_{6_4} = .4$;

- Event 1 - $E_{6_5^{(\kappa)}} \sim$ (truncated) $N(1,.5)$ over $[0,2.1]$, $\ p_{6_5} = .5$; and

- Event 1 - $E_{6_6^{(\kappa)}} \sim$ (truncated) $N(1,.1)$ over $[0,2.1]$, $\ p_{6_6} = .9$.

- Event Group 7

  - Event 1 - $E_{7_1^{(\kappa)}} \sim$ (truncated) $N(4,3)$ over $[0,8.2]$, $\ p_{7_1} = .4$;

  - Event 1 - $E_{7_2^{(\kappa)}} \sim$ (truncated) $N(1,.5)$ over $[0,2.2]$, $\ p_{7_2} = .5$;

  - Event 1 - $E_{7_3^{(\kappa)}} \sim$ (truncated) $N(1,.1)$ over $[0,2.2]$, $\ p_{7_3} = .9$;

  - Event 1 - $E_{7_4^{(\kappa)}} \sim$ (truncated) $N(6,1)$ over $[0,6.2]$, $\ p_{7_4} = .4$;

  - Event 1 - $E_{7_5^{(\kappa)}} \sim$ (truncated) $N(1,.5)$ over $[0,2.2]$, $\ p_{7_5} = .5$; and

  - Event 1 - $E_{7_6^{(\kappa)}} \sim$ (truncated) $N(1,.1)$ over $[0,2.2]$, $\ p_{7_6} = .9$.

- Event Group 8

  - Event 1 - $E_{8_1^{(\kappa)}} \sim$ (truncated) $N(4,3)$ over $[0,8.3]$, $\ p_{8_1} = .4$;

  - Event 1 - $E_{8_2^{(\kappa)}} \sim$ (truncated) $N(1,.5)$ over $[0,2.3]$, $\ p_{8_2} = .5$;

  - Event 1 - $E_{8_3^{(\kappa)}} \sim$ (truncated) $N(1,.1)$ over $[0,2.3]$, $\ p_{8_3} = .9$;

  - Event 1 - $E_{8_4^{(\kappa)}} \sim$ (truncated) $N(6,1)$ over $[0,6.3]$, $\ p_{8_4} = .4$;

  - Event 1 - $E_{8_5^{(\kappa)}} \sim$ (truncated) $N(1,.5)$ over $[0,2.3]$, $\ p_{8_5} = .5$; and

  - Event 1 - $E_{8_6^{(\kappa)}} \sim$ (truncated) $N(1,.1)$ over $[0,2.3]$, $\ p_{8_6} = .9$.

- Event Group 9

  - Event 1 - $E_{9_1^{(\kappa)}} \sim exp(.2)$, $\ p_{9_1} = .75$;

  - Event 2 - $E_{9_2^{(\kappa)}} \sim exp(1.2)$, $\ p_{9_2} = .5$;

  - Event 3 - $E_{9_3^{(\kappa)}} \sim exp(1.2)$, $\ p_{9_3} = .2$; and

Event Tier	Number of States in the Tier	Probability (cumulative)
1	20	0.54586
2	155	0.77278
3	770	0.88813
4	2705	0.94507
5	5290	0.97529
6	5993	0.99055
7	5495	0.99801
8	2399	1.00059
9	20	1.00061

Table 17.   Number of States and Cumulative Probability for Each Event Tier in the Large Air-to-Air example

- Event 4 - $E_{9_4^{(\kappa)}} \sim exp(1.2)$, $p_{9_4} = .1$.

• Event Group 10

- Event 1 - $E_{10_1^{(\kappa)}} \sim exp(.21)$, $p_{10_1} = .75$;

- Event 2 - $E_{10_2^{(\kappa)}} \sim exp(1.21)$, $p_{10_2} = .5$;

- Event 3 - $E_{10_3^{(\kappa)}} \sim exp(1.21)$, $p_{10_3} = .2$; and

- Event 4 - $E_{10_4^{(\kappa)}} \sim exp(1.21)$, $p_{10_4} = .1$.

The above event occurrence network (EON) has 56 event tiers. This EON was first run with the bucket probability code engaged - truncating transient nodes if less than 0.00001 of probability was estimated to pass through the nodes and using the cubic monic piecewise polynomial approximation. With the bucket code engaged, the EON fully truncated at event tier 9 with 22848 total states. Table 17 shows the total number of states for each tier as well as the steady state probability of being in all absorbing states in that tier and all preceding tiers - i.e., the cumulative probability.

For event tier 1, the total steady state probability of being in an absorbing state in this event tier is 0.54586 (refer to Table 17). This probability decomposes into the following steady state probabilities:

Event Tier	Number of States in the Tier
1	20
2	155
3	770
4	2915
5	9418
6	24967
7	61074
8	136608
9	283300
10	550223

Table 18.   Number of States and Cumulative Probability for Each Event Tier in the Large Air-to-Air example

- the steady state probability of the blue aircraft flight's first missile shots ending the engagement is .39116;

- the steady state probability of the red aircraft flight's first missile shots ending the engagement is .00978; and

- the steady state probability of the red SAM batteries first missile shots ending the engagement is .14492.

Before presenting the probability estimates over time, the EON was solved using simulation for comparison. *Since the network has a mixture of exponential and truncated normal distributions, no closed form analytic solution can be derived* as was the case for the one versus one example given in the last section. The initial simulation approach was to run the network generation program to determine the number of states in the EON. However, the computer employed ran out of memory during event tier 11 or after more $1,069,451$ states where created. The computer was a PC having a Pentium III processor chip running at 666MHz, 384MB of RAM, and 512MB of virtual memory. Table 18 shows the number of states for each event tier through event tier 10.

In order to produce simulation results, a truncated version of the air-to-air EON was run. Comparing tables 17 and 18, transient node truncation does not begin with the piecewise polynomial approximation until event tier 4. By taking the above air-to-air engagement EON and using only the first *two* events in each of the ten event groups, the simulation code could be executed. This new (reduced) EON will allow exact comparisons for states generated in event tiers 1 and 2 of the original air-to-air engagement EON.

For this example, the simulation was run for two million iterations with a time increment (histogram) width of 0.05 minutes. Figure 41 shows the approximation and simulation graphs of the probability of being in a representative absorbing and transient state from event tier 1 at time t. The absorbing state shown is where the first blue aircraft kills a red aircraft before another blue aircraft, red aircraft, and SAM battery fires a shot. The transient state shown is where the first red SAM battery fires a two missile salvo at the blue aircraft flight and misses before either a blue or red aircraft fires a missile. The piecewise polynomial approximation and simulation results for the absorbing and transient state are close with a maximum absolute error of 0.00273.

Figure 42 shows the approximation and simulation graphs of the probability of being in a representative absorbing and transient state from event tier 2 at time t. The absorbing state shown is where the third red aircraft fires a missile unsuccessfully at the blue aircraft then the second red SAM battery kills a blue aircraft. The transient state shown is where the second blue aircraft fires a two missile salvo unsuccessfully at a red aircraft followed by the fourth blue aircraft firing a two missile salvo unsuccessfully. For this tier, the piecewise polynomial approximation and simulation results are showing more difference graphically. In fact, the absorbing state did not show up in any of the simulation runs. This is a general trend as the events have less probability of occurring in lower tiers - *the simulation technique has a difficult time picking up these "rare states" with the needed accuracy.*

202
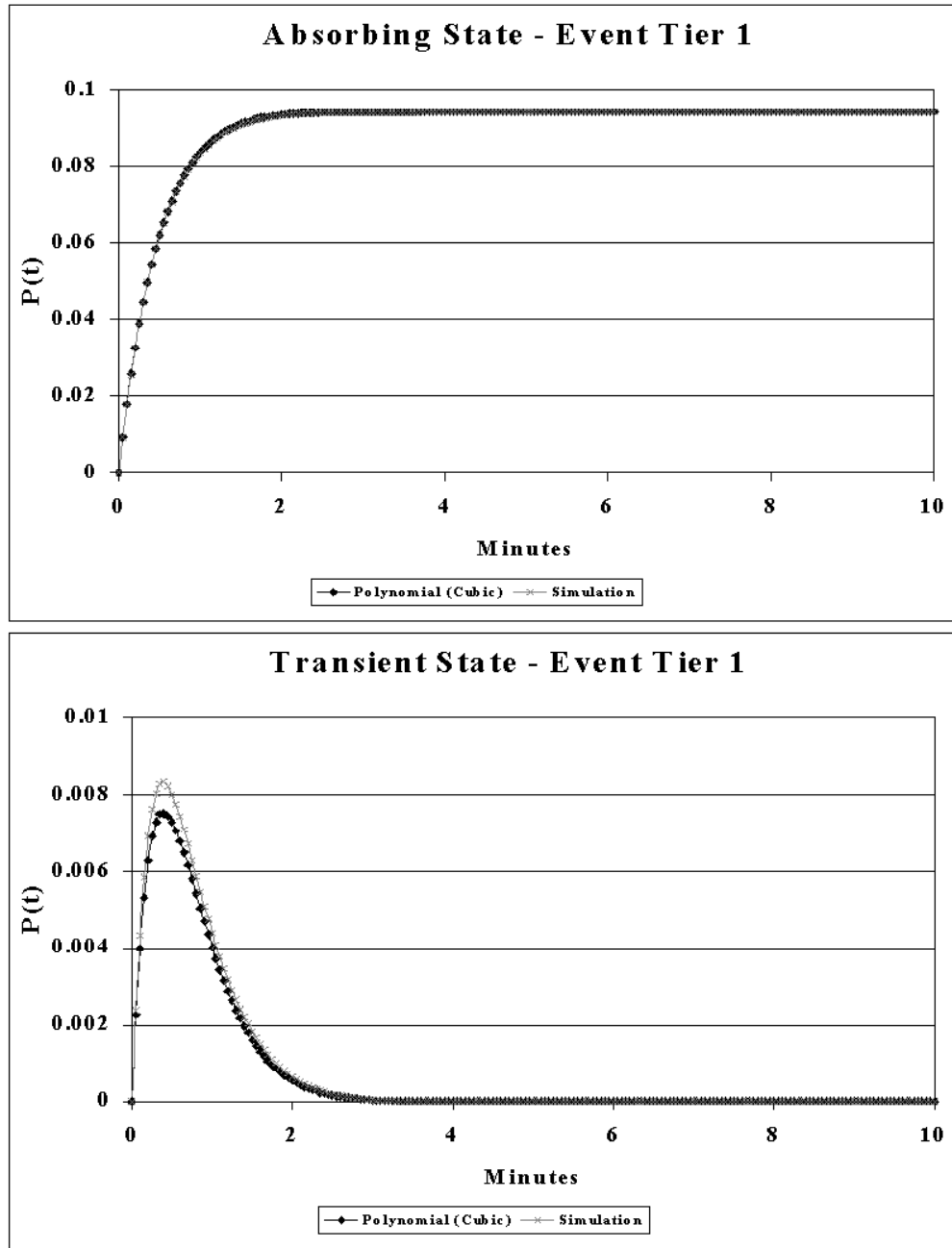
Figure 41.   Probability of Being in a Representative Absorbing and Transient State
at a time t for Event Tier 1 for the Large air-to-air engagement
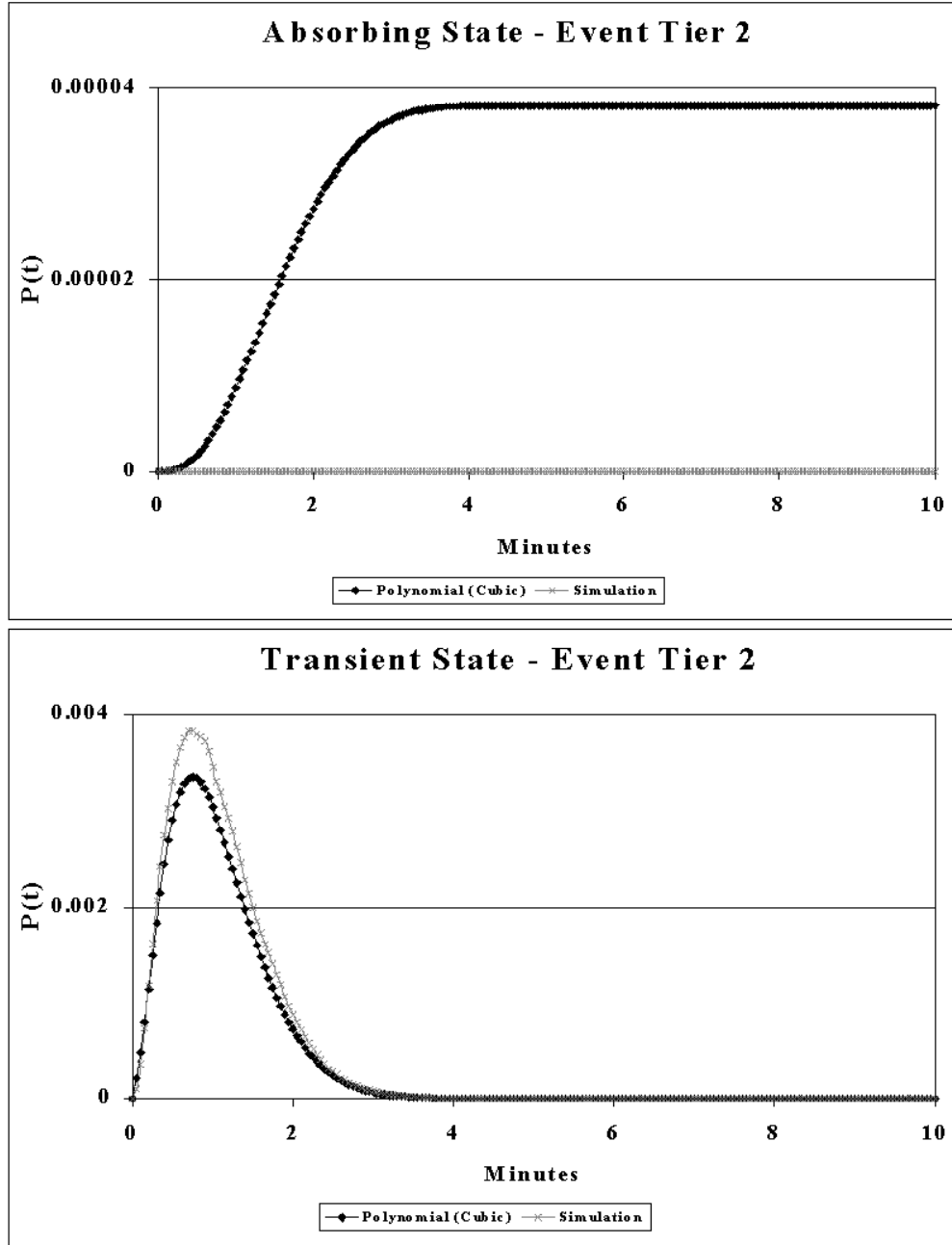
Figure 42. Probability of Being in a Representative Absorbing and Transient State
at a time t for Event Tier 2 for the Large air-to-air engagement
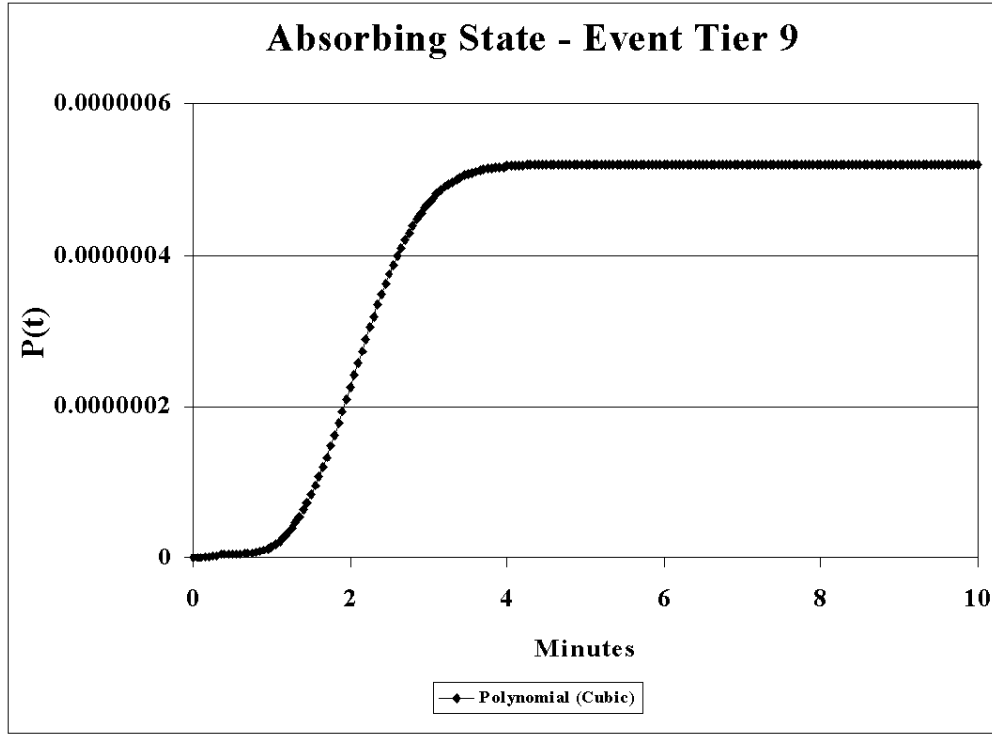
Figure 43. Probability of Being in a Representative Absorbing State at a time t for Event Tier 9 for the Large air-to-air engagement

As a final example, Figure 43 shows the approximation graph of the probability of all blue aircraft firing three missile each and then the first SAM battery killing a blue aircraft by time t.

As can be seen by the above example, the current event occurrence network formulation has a critical limitation - *the network concludes after only one successful event.* For illustration, if a blue aircraft shoots down a red aircraft the engagement is terminated. Clearly, this is not representative of real world air combat as the surviving red aircraft may not get the chance to disengage nor may all blue aircraft survive. This inherent network limitation can be offset to some degree by setting several of the success probabilities (p_{g_e}) to zero thus allowing the engagement to continue for additional entity kills. In fact when viewing such an EON formulation, Ancker and Gafarian's general renewal (GR) model is shown to be an EON subset.

For example, consider two versus two ground combat with the following rules [198] [147]:

- Two sides A and B, conduct a continuous time engagement with initial force size 2 on the A side and 2 on the B side;

- At the beginning of the battle, each combatant on side A picks a target on side B at random and fires. Each combatant fires until he is killed or makes a kill, at which time he immediately shifts to a new target picked at random and resumes firing;

- Each combatant on side B picks a target on side A at random and fires. Each combatant fires until he is killed or makes a kill, at which time he immediately shifts to a new target picked at random and resumes firing;

- All the combatants on one side are visible and within the weapon range of all the combatants on the other side;

- The ammunition supply is unlimited;

- Every combatant fires independently;

- Combat ends when one side is annihilated;

- Combatants have different interkilling time distributions; and

- If a combatant's target is killed by himself, he starts afresh the killing process on his next target, whereas, if his target is killed by another combatant from his side, his remaining time to kill is carried over to his next target.

The Ancher and Gafarian's GR model for this example is shown in Figure 44. Here combat starts with two combatants on each side state $(2, 2)$ and concludes when one of four absorbing states is reached - state $(2, 0)$, state $(1, 0)$, state $(0, 2)$, or state $(0, 1)$. The identical network can be generated via an EON formulation by setting $p_{1_1} = p_{2_1} = 0$ and $p_{1_2} = p_{2_2} = 1$ and creating the following event occurrences:

- Event Group 1 - $E_{1_1} = K_{A_1}(t)$ and $E_{1_2} = K_{A_2}(t)$

206

- Event Group 2 - $E_{2_1} = K_{B_1}(t)$ and $E_{2_2} = K_{B_2}(t)$

where $K_{i_j}(t)$ is the distribution of time to the jth kill event on side i. Figure 45 shows this EON using the state notation from Figure 17.

In order to handle the possibility of multiple successful events, the current EON formulation can be easily expanded to include: *multiple absorbing and transient states* or *reactive decision making*.

To illustrate the concept of multiple absorbing and transient states consider Figure 46. The top portion of Figure 46 shows the typical EON formulation developed in this dissertation. Here event G_{g_e} has just occurred with probability of success p_{g_e} (State 2) and probability of failure $1 - p_{g_e}$ (State 3). The bottom portion of Figure 46 shows an expanded EON formulation with two absorbing and transient states. Here event G_{g_e} has just occurred with probabilities of success $p_{g_e}^{(a)}$ and $p_{g_e}^{(b)}$ (States 2a and 2b) and probabilities of failure $q_{g_e}^{(c)}$ and $q_{g_e}^{(d)}$ (States 3c and States 3d). In order to maintain conservation of probability it is required that:

$$p_{g_e}^{(a)} + p_{g_e}^{(b)} + q_{g_e}^{(c)} + q_{g_e}^{(d)} = 1$$

To show the flexibility of this expanded EON, consider event G_{g_e} to be the occurrence of the first event in the large air-to-air engagement above where the first blue aircraft fires a missile salvo. States 2a could represent the occurrence where blue destroys a red aircraft and red ends the engagement while State 2b shows the occurrence where blue destroys a red aircraft and blue ends the engagement. Concurrently, States 3c could represent the occurrence where blue destroys a red aircraft and the engagement continues while state 3d shows the occurrence where blue fires but does not destroy a red aircraft.

To illustrate reactive decision making, consider the following EON formulation. Suppose two event groups exist, each with two events. However, condition the set of events in group two dependent on the outcome of the first event in the network.
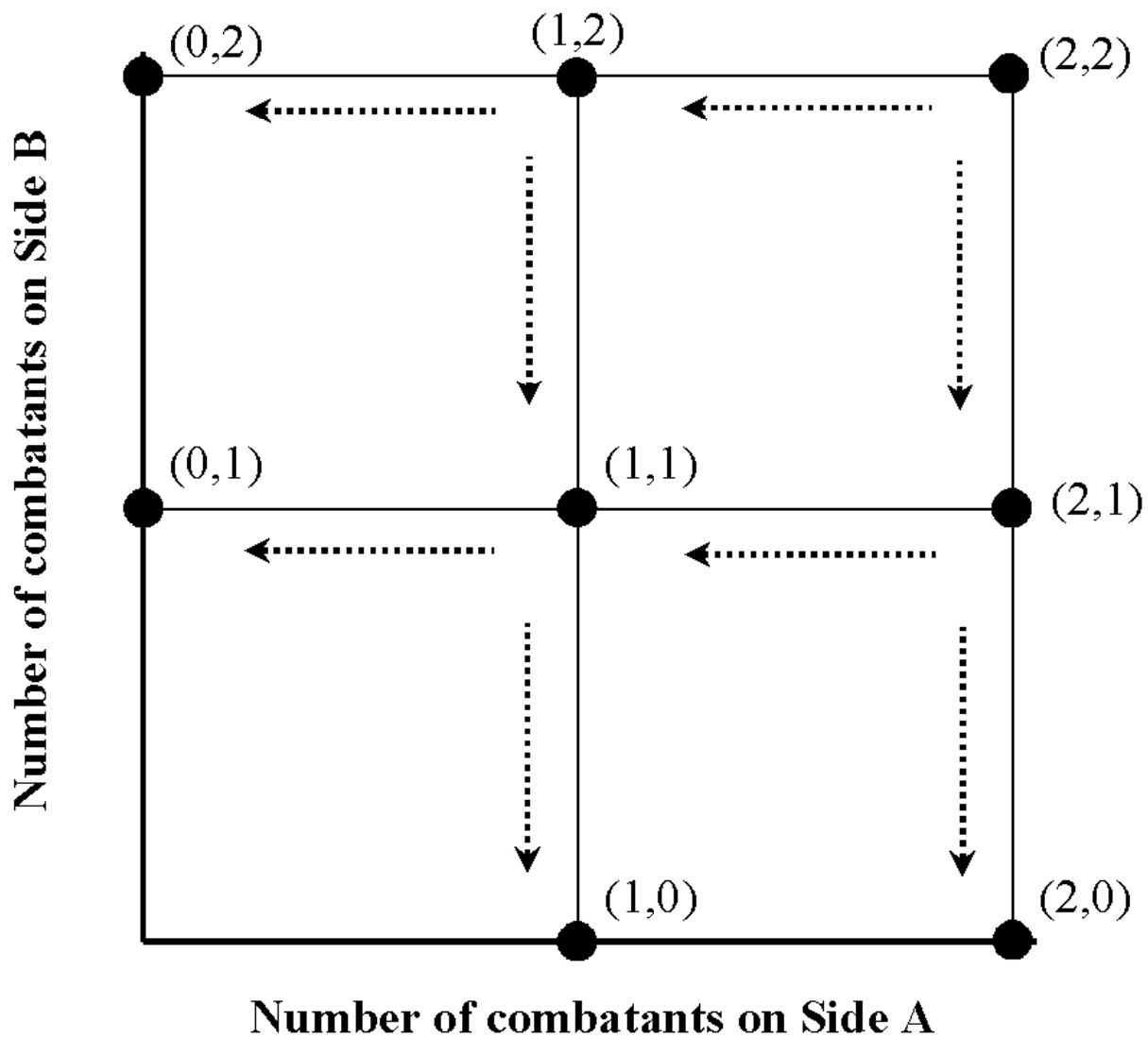
Figure 44. Ancker and Gafarian's stochastic kill network for two versus two ground combat
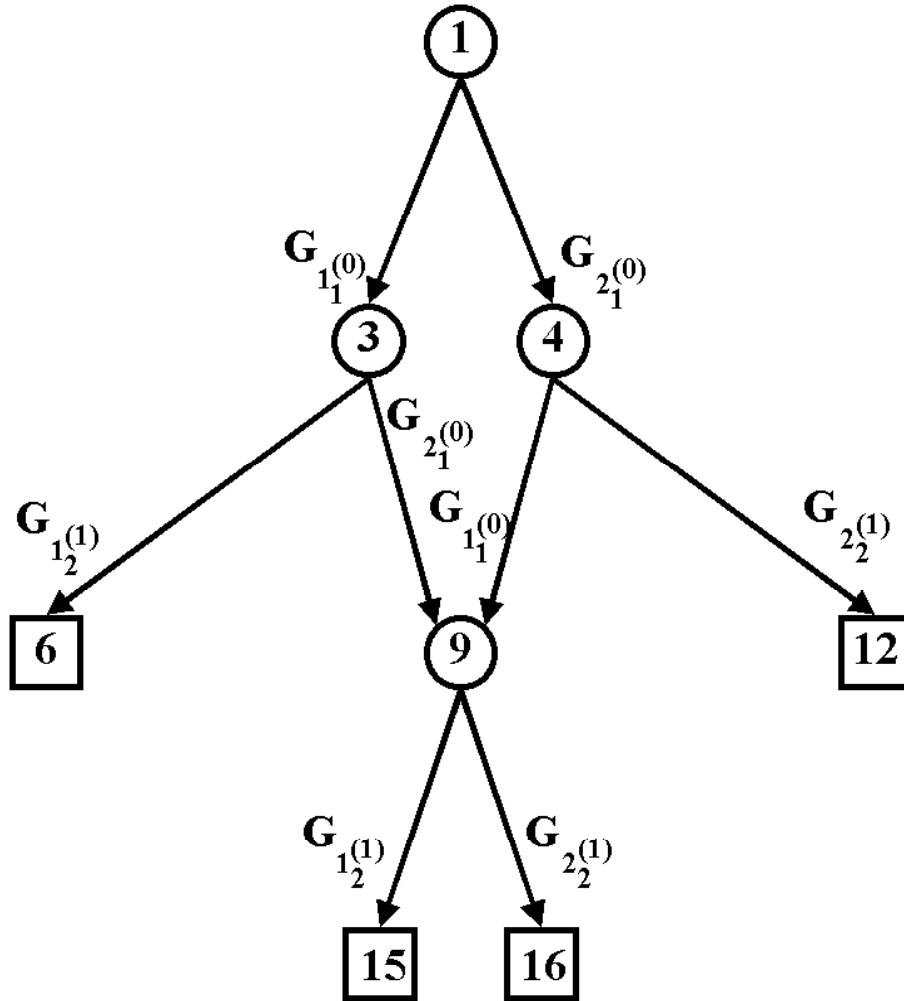
Figure 45. Event Occurrence Network formulation of Ancker and Gafarian' s stochastic kill network for two versus two ground combat
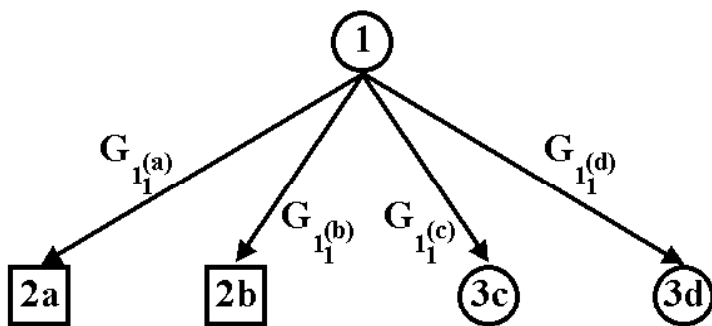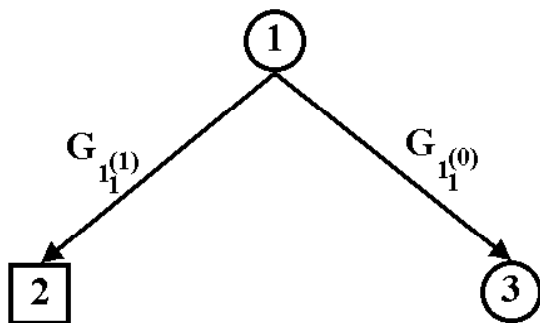
Figure 46. Event Occurrence Network formulation with Multiple Absorbing and Transient States

If event G_{1_1} occurs first then the set of events in group two is $\{G_{2_1}, G_{2_2}\}$ else the set of events in group two is $\{G'_{2_1}, G'_{2_2}\}$. To show the flexibility of this expanded EON, consider event G_{1_1} to be a missile firing from side 1 in an air engagement between two aircraft. The set of events on side 2 are now dependent on whether side 1 fires first. If side 1 does fire first, then side 2 could have a set of defensive events like defeat side 1's missile shot (G_{2_1}) and disengage (G_{2_2}). If side 1 does not fire first, then side 2 could have a set of offensive events like fire a radar missile at side 1 (G'_{2_1}) and fire an infrared missile at side 1 (G'_{2_2}).

Derivation of the probability (integral solution) for being in any state at time t for these expanded event occurrence network formulations is left for further research.

6.9 Summary

In this chapter, the techniques/algorithms for solving event occurrence networks using piecewise polynomial approximation were given. The chapter showed two air-to-air engagement EON examples. One of the networks is the simple air-to-air engagement example created by Hong, et al. [103] shown in Figure 17 and the other is a more complex engagement involving over 50 event distributions shown in Figure 40. For the simple air-to-air engagement in which an exact analytic solution could be derived, the quadratic and cubic approximations generally showed greater accuracy than the simulation solution. For the complex air-to-air engagement (for which an exact analytic solution can not be calculated), the piecewise polynomial and simulation solution showed similar results for probable events. However as the events had less probability of occurring in lower tiers, the simulation technique had a difficult time picking up these "rare states" with the needed accuracy.

VII. Summary and Recommendations

7.1 Research Goals and Summary

This goal of this research was to improve on the air-to-air combat algorithm used in USAF's campaign model THUNDER. Improvement centered on creating an air-to-air combat algorithm that was event-driven versus scripted. Concurrent with this research effort, Emergent Information Technologies (the developer of THUNDER) began development on an improved air-to-air adjudication algorithm [22], [160] for incorporation into the USAF's next generation campaign model STORM. The work in this dissertation had a direct influence on the creation of the new event-based air-to-air adjudicator.

In order to meet the above research goal, two major research objectives had to be accomplished. The first objective required developing a network formulation capable of expressing stochastic event occurrences and their interactions. Fulfilling this objective led to the author's concept of event occurrence networks (EONs). EONs are graphical representations of the superposition of several terminating counting processes. An EON arc represents the occurrence of an event from a group of (sequential) events before the occurrence of events from other event groupings. Events between groups occur independently, but events within a group occur sequentially. A set of arcs leaving a node is a set of competing events, which are probabilistically resolved by order relations. EONs differ from probabilistic networks previous discussed in the literature such as Activity Networks, Perti Nets, Task Precedence Graphs, and Belief Networks and are strongly influenced by the research of Ancker, Gafarian, Kress and several associates [83] [120] [84] [102] [147] [198] to model m versus n stochastic ground combat.

The second research objective required developing an Event Occurrence Network (EON) solution technique. An important metric for an EON is the probability of being at a particular node or set of nodes at time t. Such a probability is formu-

212

lated as an integral expression (generally a multiple integral expression) involving event probability density functions. This integral expression involves several stochastic operators:

- Subtraction,

- Multiplication,

- Convolution, and

- Integration.

In the literature, simulation-based methods are the dominant approximation technique for obtaining metrics from probabilistic networks due to their logical simplicity and efficient implementation. When a simulation is used to solve a probabilistic network, the main complexity measure is the number of simulation runs and corresponding random variable draws necessary to establish a certain statistical confidence in the output.

However, for some probabilistic networks (large networks or transient analysis of smaller networks), simulation can become a computationally costly undertaking to obtain accurate estimates of the output measures. As the number of random variables in the network increases, so does the number of replications necessary in order to properly characterize the output measure(s). As an alternative to simulation, numerical approximation has been proposed for some probabilistic networks. Numerical approximation involves the use of functions such as piecewise polynomial functions, expolynomial functions, exponential polynomials functions, etc. to approximate the density or distribution functions of the random variables. Two potentially fatal problems in using numerical approximation to determine probabilistic network output metrics are 'exploding coefficients' and 'proliferating classes'. A numerical approximation reduction technique developed by Fergueson and Short-tell [73] and Lawrence [128] [127] using single order piecewise polynomial approximation solves these problems by taking the product of a stochastic operation and

reducing its degree and class structure to the degree and class structure of the original functions in the stochastic operation. The research has expanded and operationalized the technique to the general order m polynomial case as well as formulated reduction techniques for subtraction, multiple multiplication applications, and integration stochastic operators.

This research effort has yielded five major contributions to the literature. These contributions are labeled with either an *applied* or *theoretical* designator to indicate the nature of the contribution. Contributions in this dissertation include:

1. Introduction of event occurrence networks (EONs) as a probabilistic network framework in which to investigate the interaction of independent groups of events (refer to Chapter IV). Integral expressions were developed to determine the probability of being in a particular node in the network. From the prospective of stochastic processes, this solution is equivalent to determining the probability of being in a given state at time t of a probabilistic model composed of the superposition of several terminating counting processes (Theoretical).

2. Adaptation and expansion Fergeson/Shortell/Lawrence's (least squares) piecewise polynomial approximation and reduction technique (PART) to solve for the probability of being in any node of an EON at time t. Fergeson/Shortell/ Lawrence developed PART for serial and parallel network operations in order to approximate the network completion distribution of a probabilistic activity network. Here, the series operation involved the stochastic operation of convolution (of two density functions) and the parallel operation involved the stochastic operation of multiplication (maximum of two density functions). Adapting PART concept to EONs, the set of stochastic operations was expanded to include:

   (a) Subtraction operation;

(b) Fourteen multiplication operations;

(c) Convolution operation; and

(d) Three integration operations

All the above operations were developed for general, piecewise polynomials of degree m. These operations are now available for use by programmers (Theoretical).

3. In-depth literature review (Chapter II) of probabilistic networks, their complexity and output measures, and current solution methods. Specifically, the review looked at four different network application areas: activity networks; stochastic Petri nets; task precedence graphs; and belief networks. The application areas were compared and contrasted in an effort to determine similarities among the network types. (Applied)

4. Literature review (Appendix A) of the operational air-to-air engagement environment. Specifically, the review provided an overview of air-to-air engagements and the combat environment in which these engagements take place. Factors discussed include: strategy; missions; air-to-air spatial layout; aircraft; detection and identifications; weapons; command, control, communications, and information (C3I); electronic warfare; and tactics. This review can be used by researchers and modelers to model air-to-air engagements at the theater-level. (Applied)

5. Literature review of analytical and simulated theater-level air combat models including an in-depth review of the US Air Force's THUNDER air-to-air engagement submodel (Appendix B). Several researchers have benefited from an advanced copy of this review. (Applied)

7.2 Recommendations

Event occurrence networks (EONs) are not limited to modeling combat situations. These networks can be applied whenever sets of independent random events occur. For example, one promising application area is large scale simulations. Here EONs and "bucket" analysis could be used to eliminate rare events and rare event sequences. By eliminating rare events and event sequences, the amount of computer simulation time needed for analysis might be significantly reduced as well as, perhaps more importantly, an understanding of what caused the event or event sequence to be rare to begin with.

Based on this research, the following is recommended for further study:

- Expand the set of input event distributions. Currently, only the exponential, truncated right exponential, truncated left normal, and truncated two-side normal distributions have been included. This task is straight-forward as new distributions can be added with ease to the EONSOLVE code through subroutine POLY_DENSITY_Fit.

- Develop the probability (integral solution) for being in any state at time t for two expanded event occurrence network formulations: multiple absorbing and transient states and reactive decision making. These formulations would led to the modeling of a greater set of combat situations.

- Develop a piecewise polynomial algorithm to solve for the completion time of general activity networks using the work of Bein,et al. [13], Agrawal, et al. [2], Lawrence [128], and this dissertation. The solution would have a direct commercial application.

216

Appendix A.   Air-to-Air Engagement Environment

The following summary is intended to illuminate what the experts say are the important components and variables in air-to-air engagements and to distill this information where possible to aid in the modeling of these engagements. The material below has been taken from unclassified, air-to-air engagement literature. This summary is not intended to be an exhaustive literature survey, but to provide an overview of the complex environment of air-to-air engagements and the factors that determine success in this environment.

An air-to-air engagement begins when opponent aircraft are detected and ends when the opponent aircraft are either killed or neutralized or the attack is terminated. The time from beginning of the engagement to end may be a matter of seconds. [188: page v] The outcome of an air-to-air engagement depends on many interwoven factors. None of these factors should be looked at in isolation - the *interaction effects* are much more important in determining the results of an engagement. The following factors of the air-to-air engagement environment are discussed: strategy; missions; air-to-air spatial layout; aircraft; detection and identification; weapons; command, control, communications, and information (C3I); electronic warfare; and tactics and maneuvers.

A.1   Strategy

As with land and sea engagements, when two combatants meet in the air environment, they have several engagement choices. For instance, both combatants could continuously engage each other until one combatant is destroyed; i.e., a war of attrition. Another possibility is for one combatant to move in, strike a blow, and retire before the other combatant can respond. Yet another approach involves doing something unexpected, like using a decoy combatant, pulling a third combatant into the engagement, or striking indirectly at the combatant by destroying his base

of operation or an air space controller. [193: page 114] The engagement choice(s) exercised by a combatant depend on the air strategy employed by that combatant.

An important component of an effective air strategy is an analysis of the nature of the opponent in the air environment. Many ways exist to categorize an opponent. For example, he may be rational, irrational, fanatic, rigid, flexible, independent, innovative, determined, or doctrinaire. To the extent an opponent can be assigned to a category, his air-to-air engagement plans can be anticipated, and his reactions in the air environment can be predicted. Related to analyzing the nature of the opponent in the air environment, is knowing the nature of your own side in the air environment. As you are analyzing the strengths and weaknesses of your opponent in the air environment, so is your opponent, if he is competent, analyzing yours. Your opponent will try to exploit your weaknesses and avoid your strengths. By knowing the nature of your side in the air environment, you can predict opponent air strategy. [193: page 128]

Air strategy evolves into rules of engagement (ROE) and general instructions for air-to-air engagements. It has a significant effect on how air-to-air engagements will be fought. For example, consider these three ROE

1. "BVR" ROE: Aircraft have permission to fire at any unknown;

2. "ID" ROE: In order for aircraft to fire at a target, the target must be first confirmed as hostile by <u>any</u> electronic or visual means; or

3. "Visual" ROE: In order for aircraft to fire at a target, the target must be confirmed as a hostile by visual means or through a hostile act.

Clearly, aircraft with BVR ROE have advantage over aircraft with Visual ROE in terms of first shoot capability; BVR ROE aircraft do not have to identify an aircraft visually (typically within 3 nautical miles (nm) range) whereas the Visual ROE aircraft do. Hence, the BVR ROE aircraft can launch weapons (if the aircraft is equipped with such weapons) at a longer range than the Visual ROE aircraft.

218

As section A.9 shows first shoot advantage can be a decisive factor in air-to-air engagements. [49:page 3.2.1 - 20]

A.2   *Missions*

Air-to-air engagements are a byproduct of the interactions between opponents executing separate air missions. There are many ways to classify an air mission, I will use the mission classification scheme used in United States Air Force THUNDER simulation. [29] THUNDER classifies twenty-two different air missions. [29: Volume I, pages 23-24]. Alphabetically (by acronym), the air missions are:

- Airborne Refueling (AAR)
- Airborne Early Warning (AEW)
- Air-to-Air Escort (AIRESC)
- Battlefield Air Interdiction (BAI)
- Barrier Combat Air Patrol (BARCAP)
- Close Air Support (CAS)
- Close-in Non-Lethal Air Defense Jamming (CJAM)
- Close-in Lethal Air Defense Suppression (CSUP)
- Defensive Counter-Air (DCA)
- Lethal Direct Air Defense Suppression (CSUP)
- Escort Non-Lethal Air Defense Jamming (EJAM)
- Escort Lethal Air Defense Suppression (ESUP)
- Fighter Sweep (FSWP)
- High Value Asset Attack (HVAA)
- Long Range Air Interdiction (INT)
- Offensive Counter Air (OCA)

- Over FLOT Defense Counter Air (ODCA)

- Reconnaissance (RECCE)

- Standoff Non-Lethal Air Defense Jamming (SJAM)

- Standoff Reconnaissance (SREC)

- Standoff Lethal Air Defense Suppression (SSUP)

- Strategic Target Interdiction (STI)

An air-to-air engagement occurs when a friendly or own aircraft group intercepts an opponent group of aircraft or vice versa. For illustration purposes, assume that a friendly aircraft group intercepts an opponent aircraft group. An intercept can occur one of two ways:

1. previous airborne friendly aircraft vector to the opponent aircraft, or

2. friendly aircraft launch and vector to the opponent aircraft.

In the THUNDER mission classification scheme, the previous airborne friendly aircraft are executing a BARCAP mission and the launched friendly aircraft are executing either a DCA, ODCA, or HVAA mission. A BARCAP mission involves friendly aircraft patrolling a designated area in friendly territory positioned so as to facilitate interception of opponent aircraft before the opponent aircraft strike a friendly target. A BARCAP formation generally consists of a rotating 'racetrack' orbit (i.e., a circular route with one aircraft or element pointing toward the expected threat axis at all times.) The BARCAP mission is best suited to situations when the direction of the approach or route of the opponent aircraft is known with some degree of certainty. The BARCAP mission provides the flight with certain tactical advantages. These include:

- aircraft are deployed in the air;

- sufficient time to use radar to search for, detect, and monitor all targets along the threat axis; and

- at least one aircraft is in position to monitor the threat axis.

An important consideration for the BARCAP mission is orbiting altitude. Orbiting altitude is usually chosen according to the expected altitude of the opponent, but weapons and fuel considerations also play important roles. [29: Volume II, page 50] [173: pages 325-327] [104: page 11]

DCA, ODCA, and HVAA missions wait on the ground until opponent aircraft are detected and then are 'scrambled' to intercept. The DCA mission involves friendly aircraft(s) sitting on runway alert and waiting to intercept opponent aircraft detected in *friendly territory* while an ODCA mission involves a friendly aircraft sitting on runway alert waiting to intercept an opponent aircraft detected in *opponent territory*. Finally, a HVAA mission involves a friendly aircraft sitting on runway alert waiting to intercept a AEW, AAR, or SREC mission opponent aircraft. DCA, ODCA, and HVAA mission aircraft have the advantage of countering attacks from any direction with equal ease. [29: Volume II, page 49, 50] [173: pages 325, 330]

The primary goal of a BARCAP, DCA, ODCA, or HVAA mission is to protect a friendly target[1]. Although destruction of opponent strike or high value asset (HVA) aircraft is ideal, destruction is not the only way to accomplish the mission goal. Merely threatening opponent strike or HVA aircraft is often sufficient to force aircraft off-course or to defend themselves (e.g., drop bombs in the case of the strike aircraft) leading to a mission abort. In either case, the friendly aircraft have achieved a 'mission kill'. [173: page 331]

Additionally the opponent aircraft group may contain an air escort element. THUNDER classifies escort missions as either FSWP or AIRESC. A FSWP mission involves friendly flights flying over the opponent's territory for the purpose of engaging and destroying opponent aircraft. The FSWP mission denies the opponent

---

[1]In the case of the HVAA mission, protecting the target means preventing a high value asset (HVA) aircraft from performing its mission. For example, preventing a reconnaissance aircraft from flying over a given area.

use of the airspace and makes the airspace safer for friendly strike aircraft. FSWP missions are offensive in nature and flown as a precursor to strike missions such as CAS, BAI, OCA, INT, and STI. FSWP mission objectives are achieved when opponent air combat aircraft in the strike aircraft package's ingress and/or egress route are shot down or driven off. Often the opponent's air defense network will be superior to friendly capabilities in the area, so the unexpected attack should be guarded against. Command, control, communication and information (C3I) are critical elements in the success of the FSWP. The ability of the friendly flight to find, identify, and engage opponent aircraft while maintaining an acceptable chance of survival rest in great measure on relative C3I capabilities. C3I capabilities will be discussed in greater detail in section A.7. [29: Volume II, page 50] [173: pages 317, 320,321] [177: page 40]

Although FSWP generally provides the most effective means of clearing a safe path through opponent territory for a friendly strike package, sometimes the necessary of protecting other types of aircraft such as transports in a hostile environment requires an air escort (AIRESC) mission. In general, there are four types of escort missions:

- *Reception Escort*: meet the escorted aircraft as they return from hostile territory and guard its retreat from pursuing opponent aircraft;

- *Remote Escort*: position ahead of the escorted aircraft taking the form of a sweep;

- *Detached Escort*: position strategically around the escorted aircraft, normally within visual range, to engage opponent aircraft trying to attack the escort aircraft; and

- *Close Escort*: position around the escorted aircraft for terminal defense.

Today's USAF doctrine stresses independent FSWP missions over AIRESC missions. [29: Volume II, page 49, 58] [173: pages 333, 336-338, 340, 344]

The goal of any air mission whether its an air-to-air mission or an air-to-ground mission is to accomplish the mission while maintaining adequate survivability. In many cases, one must weigh the probability of mission success against the chances of survival. Except for very critical missions, where success is absolutely essential, when the probability of survival becomes unacceptable, aircraft will abort their mission. [173: page 296]

A.3 Air-to-Air Engagement Spatial Layout

There are an infinite number of starting conditions for any air-to-air engagement; engagements have unique spatial layouts or setups. Figure 47 shows one such layout. In this figure, two aircraft are in the process of entering an air-to-air engagement[2]. This layout can be easily extended to multiple air combat by picturing each aircraft in Figure 47 as a flight. Each aircraft passes through five combat zones:

1. *Outside Sensor Zone*: In this zone, an aircraft can not detect the presence of an opponent with his own sensor system (e.g. radar). Detection requires the intervention of a third party such as a dedicated airborne or ground-based radar platform (e.g. AWACs).

2. *Beyond Visual Range (BVR) Intercept Zone*: In this zone, an aircraft's sensor system is able to detect the presence of an opponent. The pilot of the aircraft is positioning the aircraft for weapons employment.

3. *BVR Engagement Zone*: Aircraft is within the maximum launch range of its BVR weaponry.

4. *BVR/WVR Transition Engagement Zone*: Aircraft is well within range of its BVR weaponry and is transitioning to a visual contact with the opponent. Short-range weaponry becomes available for launch towards the end of this zone.

---

[2]Or as is the case with the majority of engagements, one of the aircraft is engaging the other undetected.

5. *Within Visual Range (WVR) Engagement Zone*: Pilot of the aircraft is able to visually identify the opponent. [173: page 99]

Each of the combat zones in Figure 47 depends on the unique spatial and environment factors of the engagement. For example, given the RCS of the aircraft on side B and the sensor system capabilities of the aircraft on side A, the pilot on side A is able to theoretically detect the aircraft on side B at 60 nautical miles (nm); transition from the Outside Sensor Zone to BVR Intercept Zone. This detection distance can easily change based on the relative position of the two aircraft; it is harder to detect an aircraft straight on then on the its beam (side). As another example consider the transition to the WVR engagement zone, the pilot on side B is able to detect the aircraft on side A at 5 nm while the pilot on side A is able to detect the aircraft on side B at 4 nm. These visual detection distances can clearly change based on environmental factors such as the position of sun and human factor capabilities such as the eye sight of the pilots. [44: page 64]

As a final comment about the spatial layout of air-to-air engagements, Figure 47 includes two sets of dashed lines that denote the transition from an offensive to neutral posture and the transition from a neutral to defensive posture for a side. For example, consider the offensive/neutral posture line for the aircraft on side B. This line coincides with the maximum launch range of the BVR weapon of the aircraft on side A. The aircraft on side B is no longer in an offensive posture if it is in the weapons envelope of the aircraft on side A. Offensive, neutral, and defensive postures will be discussed in greater detail in section A.9.

A.4   Aircraft

In air-to-air engagements, aircraft performance is important especially when approaching the within visual range (WVR) transition zone shown in Figure 47 (i.e., less than 10 nm). In order to employ a weapon (shoot) at an opponent it is necessary
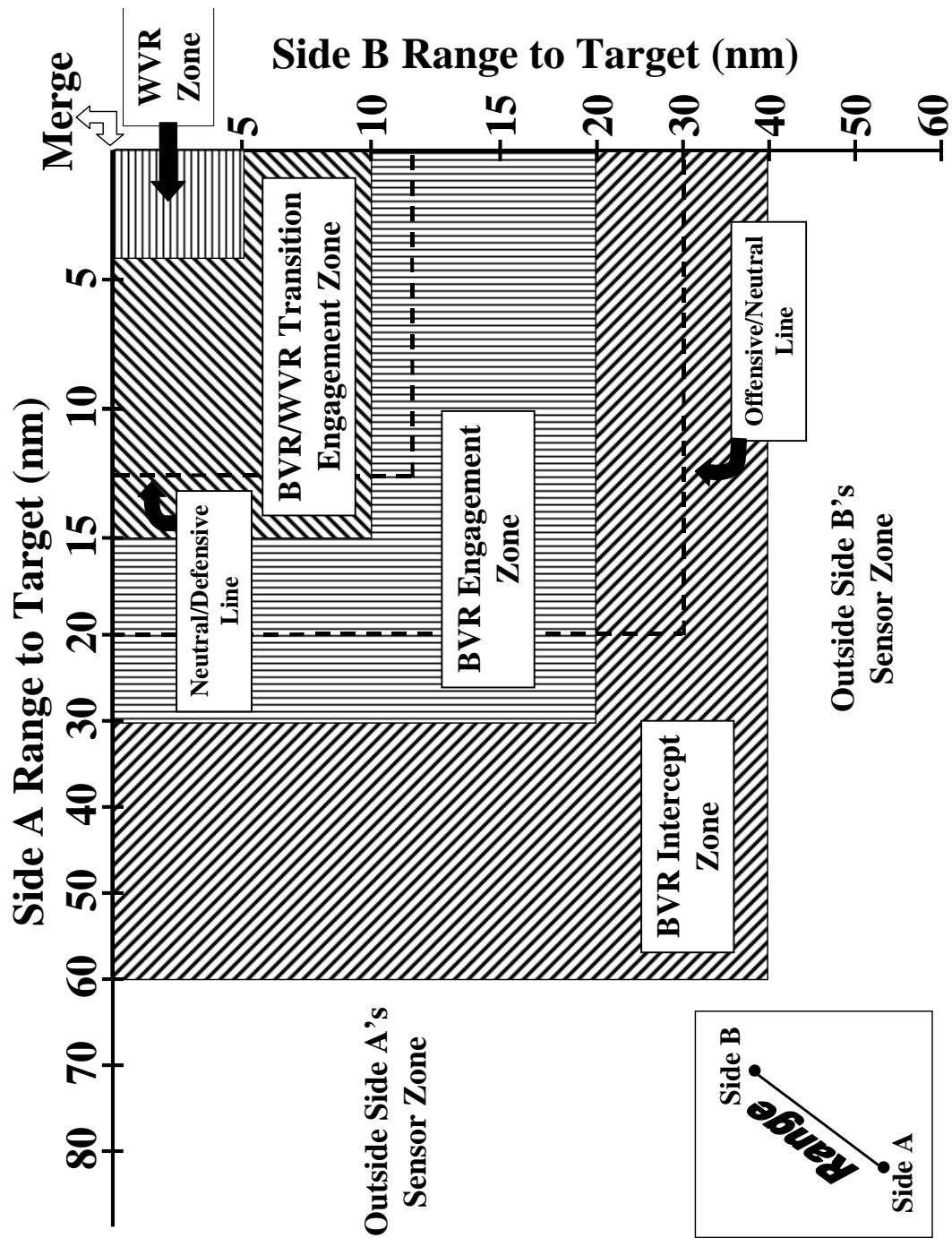
Figure 47. Air-to-Air Engagement Spatial Layout

225

to point the aircraft at the opponent or in the opponents near vicinity, depending on the weapon being employed. The performance measures of most interest are:

1. Turn Performance (instantaneous and sustained) and

2. Energy Performance (climb and acceleration).

Turn performance or maneuverability is the ability of an aircraft to change the direction of its motion in flight while energy performance is the ability to change energy state (i.e., to climb and/or to accelerate). Other aspects of aircraft performance such as takeoff, landing, range, and endurance, although critical, are concerned more with how the air combat aircraft gets to and from the air-to-air engagement area rather than how it performs within that area. [17: page 14] [173: page 139, 387, 395]

Turn performance is classified as either instantaneous or sustained. Instantaneous turn performance refers to the aircraft's turn capabilities at any given moment under existing flight conditions such as altitude and speed. Sustained turn performance refers to the aircraft's turn capabilities for an extended length of time under existing flight conditions. Aircraft turns are measured in three interrelated ways:

- Load Factor (g): indicates the 'centrifugal acceleration' generated by the turn and is usually expressed in units of G[3];

- Turn Radius (R): indicates how tight the aircraft is turning and is generally expressed in feet or nautical miles (nm)[4]; and

- Turn Rate (TR): indicates how fast the aircraft moves around the turn radius and is usually expressed in degrees per second.

Well defined physical relationships exist between the three parameters above. For fairly high load factors:

$$R \sim \frac{V^2}{g} \quad \text{and} \quad TR \sim \frac{g}{V} \tag{98}$$

---

[3]One G unit is equivalent of the nominal acceleration of gravity, $32.2 ft/sec^2$.

[4]If looking down on an aircraft from above, the turn radius is the distance from the center of the turn circle to the aircraft.

where V is the true airspeed (speed in relation to the ground) of the aircraft. As can be seen from equation 98, *turn radius (R) and turn rate (TR) are inversely proportional at a given load factor (G) and airspeed.* The greater the turn rate the smaller the turn radius and vice versa. Equation 98 also shows that turn radius is minimized and turn rate is maximized at a high load factor and low airspeed. [173: pages 388,390] [19: page 36, 37]

Based on the discussion above, in order to obtain a maximum load performance (minimum turn radius (R) and maximum turn rate (TR)), the best course of action would appear to be a high load factor (g) at minimum airspeed (V). However, there is a limit to the load factor that can be obtained at any given airspeed. Figure 48 shows the relationship between load factor and indicated airspeed (speed shown on the aircraft's airspeed indicator). The left side of Figure 48 labeled 'lift limit' indicates the maximum load factor that an aircraft can generate at a given airspeed. The curvature of the boundary reflects the variation of lift capability (L) with the square of airspeed. Lift capability is the aerodynamic force generated by the aircraft perpendicular to its direction of motion and represents the force available to turn the aircraft. Lift capability is mainly produced by an aircraft's wings. The specific lift capability that can be produced by a given wing is dependent on the airspeed and altitude of the aircraft and is roughly proportional to the square of airspeed. The upper and lower boundaries of Figure 48 depict the structural-strength limits of the aircraft. The intersection of the lift limit and structural limit define a crucial aircraft speed in fighter performance known as the corner velocity (V_c). At the corner velocity, an aircraft attains its maximum instantaneous turn performance. [173: pages 388-390] [19:page 37]

To see that an aircraft attains maximum instantaneous turn performance at corner velocity (V_c), a closer look at Figure 48 is necessary. In the case of turn rate (TR), the rapid rise in load factor (g) above aircraft speed (V) leads to improved instantaneous turn performance, culminating at the corner velocity as evidenced by
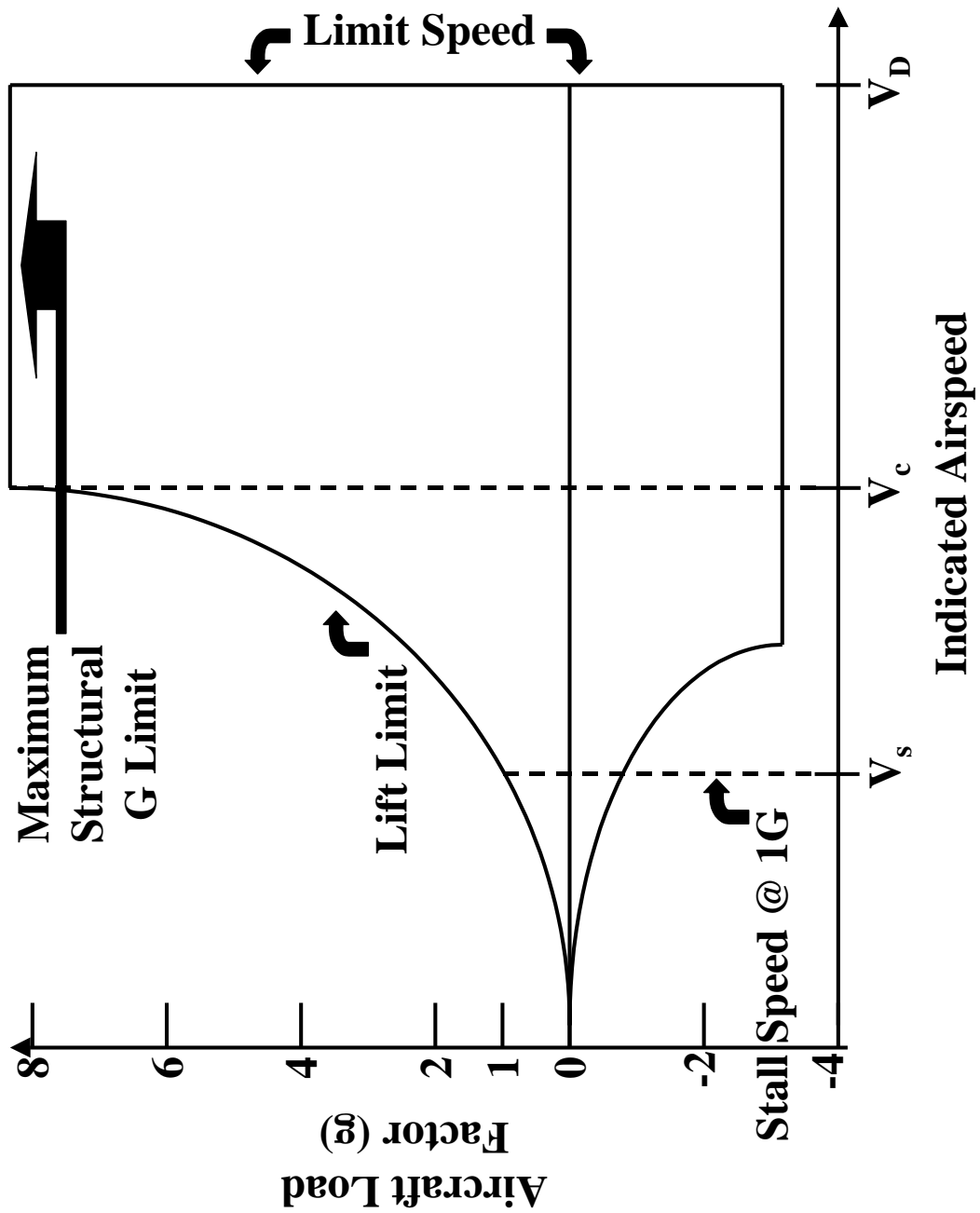
Figure 48. Aircraft Load Factor (g) versus Indicated Airspeed (V)

equation 98. Since the load factor is limited by the aircraft's structure above corner velocity, further increases in airspeed degrade (decrease) the instantaneous turn rate of the aircraft. In the case of turn radius (R), the rapid rise in the load factor above stall speed leads to improved instantaneous turn performance as turn radius decreases as evidenced by equation 98 until after corner velocity is obtained. In most modern fighters, V_c is between 400 to 500 knots of indicated airspeed. *The aircraft with the greatest useable load factor at a given airspeed has superior instantaneous turn performance (i.e., faster turn rate and smaller radius) at that speed.* [173: pages 390, 391]

Sustained turn performance is a little more complex. In Figure 48, the boundaries of the figure are dependent on the aircraft's weight, altitude, configuration, and power setting. No aircraft can sustain corner velocity while pulling a maximum load factor for long. Pulling too great a load factor will cause airspeed to 'bleed off' below corner velocity due to aircraft drag effects. This speed lost can only be regained by reducing the load factor (further reducing turn performance) until the aircraft accelerates back to corner velocity, by reducing altitude (gravity provides needed acceleration), or increasing engine thrust. Since maintaining the capability to stay at corner velocity and altitude in a turn is an important consideration, sustained turn performance depends on the ability of engine thrust to overcome increased drag. A important measure of engine thrust is the thrust-to-weight ratio at a given altitude and airspeed. As the name implies, the thrust-to-weight (T/W) ratio compares the amount of thrust produced by an aircraft's engines to the weight of the aircraft. The higher the ratio, the quicker an aircraft can gain or regain airspeed. For most combat aircraft, the T/W ratio is around 0.7 to 0.9; however, for high performance aircraft such as the F-15 and F-16, the ratio is greater than 1.0 in some altitude and airspeed regions allowing acceleration in the vertical direction. T/W is a critical component in sustained turn performance. [173: pages 407-409] [33: page 1] [17: page 14] [192: page 35]

As stated above, energy performance is the ability to change energy state; i.e., to climb and/or accelerate. The energy state of an aircraft is expressed through a quantity known as specific energy, E_S and is defined mathematically as

$$E_S = \frac{M * V^2}{G} + M * A \tag{99}$$

where M is the mass of the aircraft; V is the true speed of the aircraft; G is the acceleration of gravity (32.2 ft/sec); and A is the altitude of the aircraft. The first term in equation 99 is termed *kinetic energy* and represents the energy associated with the speed of the aircraft. The second term in equation 99 is *potential energy* and represents the energy stored in the altitude of the aircraft. [19: page 36] [173: page 394]

The energy state of an aircraft can be changed through the through the application of power (thrust and drag). The rate of change in specific energy, E_S, is known as specific excess power, P_S and is expressed mathematically as

$$P_S = \frac{T - D}{W} * V \tag{100}$$

where T is total engine thrust; D is total aircraft drag; W is aircraft weight; and V is true airspeed. Equation 100 shows that whenever thrust is greater than drag, the rate of change in specific energy is positive resulting in an increase in energy state. Conversely, if drag exceeds thrust at any time, the energy state will decrease. The specific excess power, P_S of an aircraft under given conditions of weight, configuration, engine thrust, speed, altitude, and load factor determines the available energy performance or energy maneuverability. Higher specific excess power can give an aircraft the ability to out zoom and out maneuver a less energy-efficient opponent [173: page 394] [93: page 279]

Presently, all high performance combat aircraft are subsonic aircraft, with the ability to make short supersonic dashes through the use of afterburners[5]. Although aircraft may enter air-to-air engagements at speeds greater than the speed of sound, results from Southeast Asia and the Middle East suggest that most air-to-air engagements take place at speeds less than the speed of sound. Higher airspeeds, particularly above the speed of sound, typically mean reduced maneuverability. . [33: pages 5, 10] [17: page 14] [185: page 48]

One important area of research in aircraft maneuvering is thrust-vectored engines. Thrust-vector engines enable pivoting of the exhaust nozzles. This pivoting can provide such characteristics as turn without bank and increased load factor with increased pitch. For example at the September 1996 Farnborough (England) Air Show, the Soviet Sukhoi Su-37 was able to perform 'super-cobras', 160° pull-backs from the horizontal and 360° 'kulbits' or somersaults. Both maneuvers are impossible in non-thrust vectoring aircraft. Although many defense analysts criticize, the combat value of "super-maneuvering" aircraft in a short range air-to-air engagement (i.e., the loss of energy or speed in a short range engagement is usually fatal), these maneuvers can causes AWACS and air intercept radars to "hiccup" and temporarily lose track of a target for several seconds. The US will not have an equivalent maneuvering aircraft until the F-22 is operational. [38] [185: pages 16, 93]

Most air combat aircraft generally operate or cruise at an altitude of 20,000 to 25,000 ft and an airspeed of 450 knots. Many air combat aircraft experience a marked changed in performance when passing an altitude of 25,000 to 27,000 ft; combat aircraft are much more maneuverable in the mid-twenties - however, engine

---

[5]Afterburner refers to a series of engine settings where additional fuel is sprayed directly into the exhaust gases of the final combustion chamber. The highest afterburner setting can provide a 65% increase in thrust of today's turbofan combat aircraft engines. Unfortunately using afterburner settings consumes fuel at roughly three to four times the rate of non-afterburner 'dry' engine settings. The highest maximum power setting without afterburner is termed mil (military) power. The USAF's next-generation Advanced Tactical Fighter (ATF) is required to sustain supersonic cruise speeds (speeds above Mach 1.5) at altitude in mil power (i.e., without the use of afterburners).

performance drops off significantly in the upper twenties and turn performance is thus degraded. The lower twenties often offers the right balance of high true airspeed and endurance (low fuel consumption) while providing adequate indicated airspeed for maneuvering. For greater range (endurance) and better maneuverability, most of today's combat aircraft cruise at high subsonic speed on the order of Mach 0.8 to 0.9. In terms of the speeds at an air-to-air engagement, USAF combat aircraft such as the F-15C will pursue the intercept phase of an engagement at about Mach 0.9 accelerating to Mach 1.1 to launch beyond visual range (BVR) weapons. The increase in speed serves two purposes: first, the increased speed adds to BVR missile speed thus increasing BVR missile reach; second , if the BVR missiles miss, then the speed can be converted into a turn or energy advantage. The Eagles will generally stay at this speed (Mach 1.1) unless entering a turning engagement where opponent speed and own ship cornering velocity come into play. [185: pages 48, 72] [30] [192: page 64]

A.5 Detection and Identification

Analogous to a ground battle, in order to destroy an opponent in the air, he must first be found. The range at which an opponent aircraft is detected, the detection range, is a critical factor in air-to-air engagements. Simply put *detection range equates to reaction time.* The greater the detection range, the larger the time to react to the opponent aircraft and more reaction time means more time to set-up specific engagement tactics. In the air environment, aircraft are detected by radar, infrared sensors, electronic countermeasures such as a radar and missile warning receivers, communication intercepts, and visual observation. In all these cases, detection depends on extracting and interpreting information provided by the electromagnetic spectrum. [33:page 15] [49: page 8.2-8].

The electromagnetic spectrum is an ordered array of known electromagnetic (EM) radiations or energies. The spectrum extends from short (in terms of wave-

length) cosmic rays through long electric current. The two principal ways in which EM radiation is described are by its *frequency, f* and *wavelength, λ*. The frequency of a particular EM wave is the number of oscillations in the wave during a one second time period while wavelength is the distance through which the wave propagates in one complete cycle. The wavelength of an EM wave is related to the frequency of the wave by the equation:

$$\lambda = \frac{c}{f}$$

where c is equal to the speed of light (and of all EM radiation) through space approximately $3x10^8$ m/s. [195: pages F-89,90] [12: pages 1-3 thru 1-5]

Table 19 details the electromagnetic spectrum. Frequencies between about 15 Hz to 20 kHz are commonly referred to as audio frequencies, because the human ear can detect oscillations in the air pressure at these frequencies. Radio communications typically takes place at frequencies from 10 kHz to 3 GHz and overlap radar frequencies which take place between 1 GHz and 10^{12} Hz. Infrared radiation occupies a region from 10^{12} to $8x10^{14}$ Hz while visible light occupies a very narrow portion of the spectrum, with frequencies in the order of 10^{15} Hz. [12: page 1-3] [137:page 185] [69:page 157]

By far, the main method for aircraft detection involves extracting and interpreting energy from the radar portion of the electromagnetic spectrum. The term radar is derived from the original name given to the technique *Radio Detection And Ranging* invented by the British during World War II. At its most basic level, a radar system operates by radiating EM energy and detecting the echo returned from reflecting objects (targets). The nature of the echo signal provides information about the target such as range, aspect, elevation, velocity, etc. [12: page 1-5] [178: page 1.1]

Radars can be classified by their use and generally fall into three categories: *early warning, acquisition, or guidance.* Early-warning radars are usually low frequency, long-wavelength sets requiring large antennas. Their size generally precludes

Frequency (Hz)	Wavelength (m)	Type
10^{23}	3×10^{-15}	Cosmic Photons
10^{22}	3×10^{-14}	Gamma Rays
10^{21}	3×10^{-13}	Gamma Rays, X Rays
10^{20}	3×10^{-12}	X Rays
10^{19}	3×10^{-11}	Soft X Rays
10^{18}	3×10^{-10}	Ultraviolet, X-Rays
10^{17}	3×10^{-9}	Ultraviolet
10^{16}	3×10^{-8}	Ultraviolet
10^{15}	3×10^{-7}	Visible Spectrum
10^{14}	3×10^{-6}	Infrared
10^{13}	3×10^{-5}	Infrared
10^{12}	3×10^{-4}	Far-infrared
10^{11}	3×10^{-3}	Microwaves
10^{10}	3×10^{-2}	Microwaves, Radar
10^{9}	3×10^{-1}	Radar
10^{8}	3	Television, FM Radio
10^{7}	3×10^{1}	Short-Wave Radio
10^{6}	3×10^{2}	AM Radio
10^{5}	3×10^{3}	Long Wave Radio
10^{4}	3×10^{4}	Induction Heating
10^{3}	3×10^{5}	
10^{2}	3×10^{6}	Power
10^{1}	3×10^{7}	Power
1	3×10^{8}	
0	∞	Direct Current

Table 19. Electromagnetic Spectrum

their use on air combat aircraft. Early-warning radars are characterized by relatively long range and poor resolution and are used primarily for airborne or ground-based intercept control. An aircraft detected by an early warning radar may appear on a radar display as a 'blip' representing several miles in width and many closely spaced aircraft may appear as a single target. Air combat aircraft are generally equipped with an acquisition radar which has a higher frequency, smaller antenna, shorter range, and better resolution than early warning radars. Acquisition radars usually have the capability to 'track' a target in order to gain more detailed information on its relative position, speed, altitude, etc. Air combat aircraft radars also are generally capable of guiding air-to-air weapons to a target (thus also serving as guidance radars). Given sufficient resolution, a radar can discern something about the target's size and shape (refer to identification techniques explained later in this section). Table 20 shows the general frequency designations (bands) at which radar systems operate. Each frequency band has particular characteristics that make it better for certain applications than for others. [178: pages 1.1, 1.14, 1.15] [173: pages 346, 347]

Aircraft may be detected by either ground-based or airborne radar. USAF air superiority doctrine relies mainly on the use of airborne radar platforms such as the E-3 Sentry Airborne Warning and Control System (AWACS) and E-2 Hawkeye platforms as well as the radars of air combat aircraft. The E-3 Sentry's radar (APY-2) operates in the S or E/F band, meaning that it generates radar waves in the 2-to-4 GigaHertz (GHz) range, with a wavelength of from 15 to 7.5 cm. The APY-2 radar uses the pulse-Doppler principle, relying on precise measurement of the tiny frequency shift in waves reflected from moving objects to distinguish flying aircraft from background ground clutter. This gives the radar the ability to 'look down' and detect low-flying targets, as long as they are moving faster than 80 knots. [33: pages 111, 115]

The APY-2 radar has a 360° scanning capability and can operate in seven modes:

Standard Radar Bands	Frequency	Wavelength
HF	3-30 MHz	3000 cm - 300 cm
VHF	30-300 MHz	300-100 cm
UHF	0.3-1 GHz	100-30 cm
L	1-2 GHz	30-15 cm
S	2-4 GHz	15-7.5 cm
C	4-8 GHz	7.5-3.75 cm
X	8-12 GHz	3.75-2.5 cm
Ku	12-18 GHz	2.5-1.6 cm
K	18-27 GHz	1.6-1.1 cm
Ka	27-40 GHz	1.1-0.75 cm
MM	40-100 GHz	0.75-0.3 cm

NATO Bands	Frequency	Wavelength
A	0-250 Hz	
B	250-500 Hz	120-60 cm
C	0.5-1 GHz	60-30 cm
D	1-2 GHz	30-15 cm
E	2-3 GHz	15-10 cm
F	3-4 GHz	10-7.5 cm
G	4-6 GHz	7.5-5 cm
H	6-8 GHz	5-3.75 cm
I	8-10 GHz	3.75-3 cm
J	10-20 GHz	3-1.5 cm
K	20-40 GHz	1.5-0.75 cm
L	40-60 GHz	0.75-0.5 cm
M	60-100 GHz	0.5-0.3 cm

Table 20. General Frequency Designations (Bands) For Radar Systems

1. *Pulse-Doppler Non-Elevation Scan (PDNES)*: This high pulse repetition frequency (PRF)[6] mode provides look down surveillance to the surface using pulse-Doppler, with narrow Doppler filters and a sharp beam to eliminate ground clutter. The PDNES mode is optimized for detection out to the radar horizon; however this mode does not measure target elevation. The PDNES mode is used when range information is paramount to elevation data. At an altitude of 9114 m (29902 ft), this mode is able to detect targets out to a range of approximately 213 nautical miles (nm).

2. *Pulse-Doppler Elevation Scan (PDES)*: This mode is similar to the PDNES mode but includes a electronic vertical radar beam scan to provide target elevation with a slight decrease in maximum detection range.

3. *Beyond-the-Horizon (BTH)*: This low PRF mode provides long-range surveillance of medium and high altitude aircraft without Doppler[7]. This mode obtains target range and azimuth information only; no target elevation data is provided.

4. *Passive*: In this mode, the radar transmitter is turned off and the receiver left on to obtain electronic countermeasures (ECM) information such as enemy jamming location.

5. *Maritime*: This mode provides a maritime surveillance capability. This involves the use of a very short radar pulse to provide the resolution required to detect moving or anchored surface ships.

6. *Interleaved One*: The PDNES and Maritime modes can be interleaved.

7. *Interleaved Two*: The PDES and BTH modes can be interleaved.

---

[6]PRF refers to the number of radar pulses sent out in a given time period. PRF will be discussed in greater detail in F-15C's radar system (APG-63) description.

[7]Since the radar beam is above the horizon, there is no ground clutter and a Doppler beam is not needed.

The PDNES/Maritime and PDES/BTH interleaved modes are noted as providing the necessary performance for most operational situations. Further flexibility is generated by the APY radar being able to divide the surveillance volume of azimuth scans into to a maximum of 32 sectors, each of which can be configured with its own operating mode. [78] [23: page 186] [113: page 566] [111: page 247]

Most air combat aircraft radar systems operate in the X-band (8 to 12.5 GHz). This frequency band offers the best compromise in terms of radar system design and performance. By far, the most critical factor in radar system design for an air combat aircraft is antenna size. For an air combat aircraft, the most convenient position for the antenna is the aircraft's nose where limited space (due to aerodynamic considerations) constrains allowable antenna size. An X band radar system produces a good compromise antenna size while maintaining adequate system performance. Dropping (frequency) to a higher performance S-band radar system such as on the E-3 Sentry demands a larger antenna while increasing to a K-band radar offers a smaller antenna size, but a radar signal which is adversely affected by meteorological conditions such as rain.[8] USAF air combat aircraft such as the F-15C Eagle, F-15E Strike Eagle, F-16C Falcon, and F-22 Raptor contain sophisticated radar systems. [88: page 31] [185: page 73]

The F-15C Eagle uses the APG-63 radar system[9] while the F-15E Strike Eagle uses the APG-70 radar system. Each system is an all-weather, programmable, multimode, pulse-Doppler radar operating with a number of selectable frequencies in the I-/J-bands (8 to 20 GHz). With a detection range of greater than 100 nm (115 nm for the APG-70) against a large radar cross section aircraft (e.g. a Tu-95 Bear bomber), both the APG-63 and APG-70 radar combine long range with features such as automatic detection and lock-on within 10 nm (11.5 nm for APG-

---

[8]In addition, the X-band generates adequate gain (the output/input power ratio) to produce a finely-controlled beam without unwanted big sidelobes. Sidelobes give away the aircraft's presence on opponent radar warning receivers without contributing to detection capability.

[9]Some versions of the Eagle are equipped with the APG-70 radar system.

70). These systems contain a programmable signal processor which gives the ability to rapidly change the radar (through software rather than a hard wired circuit design) to accommodate new tactics, modes, and weapons. Both the APG-63 and APG-70 can simultaneously track and engage multiple targets. [33: page 18] [111: page 268, 273] [78] [88: page 32] [113: page 147]

Three pulse repetition frequency (PRF) ranges are used in the APG-63 and APG-70 radar systems: low, medium and high. PRF refers to the number of radar pulses transmitted in a given time period. Each of the three PRF ranges has its strengths and weaknesses. Low PRF (typically 100 to 1000 pulses per second) radar modes transmit radar pulses at sufficient intervals so that each pulse has enough time to travel out to the target, be reflected, and return to the radar before the next pulse is transmitted. Low PRF radar transmissions provide excellent 'look-up', long range capability. However, low PRF radar transmissions tend to produce massive ground clutter when searching in a 'look-down' capacity; making targets harder to differentiate. In contrast, high PRF (typically 10,000 to 100,000 pulses per second) radar modes transmit radar pulses sooner than reflected pulses are returned. In high PRF radar modes, the radar is able to measure the Doppler shift in pulses reflected by the target.[10] Thus, high PRF unlike low PRF radar transmissions can acquire and discriminate targets from ground clutter when searching in a 'look-down' capacity. However, high PRF radar transmissions are poor at providing target distance due to the high reflected pulse return rate and are generally useless in beam (side) attacks where there is relatively little Doppler shift.[11] [185: pages 73, 74] [88: pages 31] [57: pages 98-105] [67: page 46]

---

[10] Recall Doppler shift is the tiny shift in frequency of a reflected pulse caused by the relative motion between the target and the radar system. As the target approaches the radar system more reflected pulses per second reach the system, but as the target recedes fewer reflected pulses per second reach the system. Hence, the Doppler shift is providing velocity information only.

[11] In order to improve distance determination, lower frequency-modulation (FM) or two simultaneous, closely-spaced frequencies are transmitted. But this is at the expense of reduced detection range due to the complex processing required of the returning signals. In addition, FM ranging has an error of five percent or so.

Medium PRF (typically 1000 to 10,000 pulses per second) radar modes offers a good compromise to low and high PRF modes at distances less than 40 nm or so. Medium PRF transmissions provide target distances while acquiring and discriminating targets from ground clutter when searching in a 'look-down' capacity. The three PRF ranges used in the APG-63 and APG-70 radar systems effectively filter out ground clutter, giving the radar a 'look-down, shoot down' capability (i.e. the radar can track and engage an aircraft flying at low altitude) while retaining an excellent 'look-up, shoot up' capability. [185: pages 73, 74] [88: pages 31] [57: pages 98-105]. [33: page 18]

The APG-63[12] radar's horizontal or azimuth scan has three selectable arcs, 30°, 60°, or 120°, centered directly in front of the aircraft. The vertical or elevation scan has three selectable 'bars' (a bar is a slice of airspace with a vertical depth of 1.5° per bar) - 2 bar (3°), 4 bar (6°), or 6 bar (9°) - for varying vertical coverage. With an antenna sweep speed of 70°/sec/bar, the largest search pattern (a 120°, 6 bar scan) can take up to fourteen seconds to complete. [33: page 20]

Additionally, the APG-70 radar system has a low probability of intercept mode (LPI) mode designed to defeat the radar warning receivers and electronic support measure (ESM) detectors (reference section A.8) on opponent aircraft, by using techniques like frequency-hopping and power regulation. This means an Eagle or Strike Eagle can conduct an active radar search and RWR/ESM-equipped aircraft will not detect the search. Conventional radars locate aircraft by transmitting high-energy pulses in a narrow frequency band and finding high-energy returns. However, a good RWR/ESM-equipped aircraft can pick up these high-energy pulses at over two times the radar's effective range. On the other hand, LPI radars transmit low-energy pulses over a wide band of frequencies. When the multiple returns are received by the aircraft, the radar's signal processor integrates all the individual pulses back together. Even though, the amount of returning energy is about the

---

[12]Scan performance statistics could not be found in the literature for the APG-70 radar system.

same as a conventional radar's high-energy pulse, most warning systems can not detect the LPI radar since each individual LPI pulse has significantly less energy than a conventional radar pulse. Additionally, the LPI radar frequencies used do not necessarily correlate to the normal frequency search pattern of these warning systems. This gives the Eagle or Strike Eagle a tremendous advantage in any long-range engagement, as the pilot does not have to establish a lock-on when firing a medium-range missile such as the AIM-120 AMRAAM. The first realistic indication that an opponent aircraft will have of an attack by an Eagle or Strike Eagle may be the screams from his radar-warning receiver telling him that the AMRAAM's radar has lit off, locked on, and is in the final stages of intercept. [33: page 21, 22]

The F-16C Falcon uses the APG-68 radar which is an X-band (8-12 GHz), pulse Doppler radar. The radar has range out to 80, can scan a 120° arc horizontally, and has elevation setting of 2, 4, or 6 bar. As with the F-15, the range at which air-to-air targets are detected is increased by using a high PRF mode. Once detected, a medium PRF mode can be employed to get additional range and angle information. The APG-68 radar is capable of tracking ten targets simultaneously. [33: page 92] [111: page 270]

The radar system to be used on the F-22, the APG-77, differs considerably from the radar systems described above. For instance, the antenna of the APG-77 is a fixed, elliptical, active array containing about 1,500 individual, radar Transmit/Receive (T/R) modules. Each T/R module is about the size of an adult's finger and is a complete radar system in its own right. The APG-77 can sweep a 120° multiple-bar search pattern. However, instead of taking fourteen seconds to sweep a 120°, six-bar pattern like the APG-63, the APG-77 will search the equivalent volume almost instantaneously. This is because the active array can form multiple radar beams to rapidly scan an area and does not need motors or mechanical linkages to move the antenna for a scan. The APG-77 will also have the critical LPI search mode. [33: page 22]

Recall it was stated as the beginning of this section that USAF air superiority doctrine relies mainly on the use of airborne radar platforms. This reliance is a direct result of the offensive nature of USAF air superiority doctrine which seeks to obtain control of opponent airspace. Since the ground under an opponent's airspace is by definition controlled by the opponent, airborne radar platforms in conjunction with space-based assets such as satellites provide the most effective means to monitor opponent airspace. In contrast, a majority of the nations in the world including several of the former Soviet Union republics and associated satellite nations do not have the adequate airborne assets (both in number and quality) to conduct a prolonged offensive air campaign. The air superiority doctrine of these nations centers on maintaining an adequate air defense capability. Detection of opponent aircraft in an air defense system relies not only on the use of airborne radar platforms, but also on ground-based radar systems. [193] [131]

Ground-based radar systems are quite diverse in nature. Systems may be mobile or fixed; two-dimensional (provide bearing and range information only) or three dimensional (bearing, range, and height information); and provide high or low altitude coverage or both. To provide an illustrative example of the capabilities of such systems, consider the ground-based radar systems of the People's Republic of China[13]. [67: page 30] [27]

Two dimensional air defense radar systems are generally characterized by relatively long range; however these systems can only provide bearing and range information. Chinese two dimensional radar systems include the 408C, HN-401R, and HN-503 mobile, long range air defense warning radar systems. The 408C ground-based radar system can operate simultaneously in two band regions within the VHF frequency band: 100 to 120 MHz (considered the 'low' band region) and 150 to 180

---

[13]I tried first to find an open, unclassified reference documenting the performance characteristics of the ground-based radar systems of the former Soviet Union since these systems are prevalent in many of regions of the world where conflict can be expected to take place. Unfortunately, I was not able to locate a reference. As a substitute, I choose the Chinese ground-based radar systems where information was readily available.

MHz ('high' band region). The low band radar wave has a horizontal beamwidth of 8° while the high band radar wave has a horizontal beamwidth of 12°. The 408C radar has a pulse length of 10 microseconds (3000 m) and a PRF of 200 Hz. The HN-401R ground-based radar system operates in the UHF band (0.3 to 1 GHz). The radar has a detection range of approximately 162 nm against a 1.5 square meter target. The HN-401R radar has a range resolution of 450 m and an azimuth resolution of 9°. Lastly, the HN-503 ground-based radar system operates in the VHF frequency band (30 to 300 MHz). The radar has a pulse length of 430 microseconds (129 km) and a PRF of 200 Hz. This radar has longest detection range of any of China's air defense radars with a range of approximately 389 nm. The Type HN-503 radar has a range resolution of 450 m and an azimuth resolution of 7°. [27] [111: pages 8,9, 13]

In contrast to two dimensional air defense radar systems, three dimensional air defense radar systems are characterized as generally being of shorter range than their two dimensional counterparts; however three dimensional systems provide altitude information in addition to bearing and range. Chinese three dimensional radar systems include the JY-8, the JY-8A, JY-9, JY-14 air surveillance radars. The JY-8 radar is a mobile, van-mounted air surveillance radar that operates in the G/H bands (4 to 8 GHz). This radar can be employed as the main radar sensor for an automated tactical defense system, or can be used as an independent radar. The JY-8 radar has a horizontal beamwidth of 0.55°, a vertical beamwidth of 0.9°, and a pulse length of $3 - 3.3$ microseconds (900- 990 m). This radar also has three PRFs 365, 500, and 600 Hz. The JY-8 radar has a detection range of approximately 188 nm and can cover altitudes up to 13.5 nm. Additionally, the JY-8 has a range resolution of 1000 m and an azimuth resolution of 0.6° and can simultaneously track up to 36 targets. A variant of the JY-8, the JY-8A mobile air surveillance radar can carry out fire control of missiles and/or antiaircraft artillery guns. This radar has a reduced detection range of approximately 80 nm and a reduced altitude coverage up to 6.5 nm. [111: pages 10, 11] [27]

The Chinese JY-9 radar is a mobile, low altitude, three dimensional air surveillance radar designed for air defense gap filling, airport surveillance, and coastal defense. This radar operates in E-band (2 to 3 GHz) with a 20 microsecond (6000 m) pulse length. The radar rotates at 6/12 rpm providing 360° azimuth coverage and 40° vertical coverage. The JY-9 radar has a horizontal beamwidth of 1.5°, a vertical beamwidth of 40° and a PRF of 850 Hz. This radar has a detection range of approximately 80 nm[14]. Additionally, the JY-9 has a range resolution of 120-240 m and an azimuth resolution of 1.3° and can simultaneously track up to 200 targets. The JY-14 is a fixed, long range three dimensional air surveillance radar operating in the F-band (3 to 4 GHz). The radar rotates at 6 rpm providing 360° azimuth coverage and 20° vertical coverage. The JY-14 had a detection range of approximately 169 nm and can cover altitudes up to 13.4 nm. Additionally, the JY-14 has a range resolution of 240m, an azimuth resolution of 0.9°, and an altitude resolution of 1°. [111: pages 11,12] [27]

In addition to the two / three dimensional ground-based radar systems, the People's Republic of China has two over-the-horizon (OTH) radar systems in North China and the South China Sea. As there name implies OTH radar systems can look past the radar horizon (the line of sight limited by radar by the curvature of the earth). These systems generally use the ionosphere as a "mirror in the sky" to bounce transmitted radar waves off the upper atmosphere and use the same path for reflected returns. Unfortunately, OTH radar warning systems have no close-in capability. The Chinese OTH radar systems operate in HF-band (3 to 30 MHz) have a reported detection range from 378 to 1890 nm with a 60° azimuth coverage. [67: pages 18, 21, 22] [111: page 12] [27]

In order to direct surface-to-air weapons such as AA guns or surface-to-air missiles, an air defense system needs fire control radars that can precisely track the flight of an opponent aircraft. The People's Republic of China has several surface-

---

[14]With a probability of detection of .80 and Swerling I of 2 sq m.

to-air fire control radar systems. In addition to the JY-8A radar mentioned above, other surface-to-air fire control radar systems include the 311 and 702. The mobile, 311 surface-to-air fire control radar system is used with anti-aircraft (AA) guns (usually 37 and 55 mm) and comes in three versions (A, B, and C). All versions of this radar operate in the I/J-band (8 to 20 GHz) and is capable of both search and track functions and is generally used with a computer and optical range finder. The JY-9 radar has a horizontal beamwidth of 2.6°, a vertical beamwidth of 2.4° and two pulse lengths: narrow - 0.3 micro seconds (90 m), PRF of 2500 Hz and broad - 0.9 microseconds (270 m), PRF of 850 Hz. The 311-A has a detection range (on a fighter size target) of approximately 16 nm in search mode with a maximum tracking range of 13.5 nm. The minimum reliable tracking range is about 500 m. The 311-B introduced a identification friend or foe (IFF) systems and increased the frequency coverage and maximum detection range (search mode) to 19 nm. The 311-C has a maximum detection range of 22 nm. The Chinese 702 surface-to-air fire control radar system is a mobile, all weather system consisting of radar, optical unit, and computer and operates with various caliber AA guns and surface-to-air missiles. The radar operates in the G/H/I-bands (4 to 10 GHz) and is used primarily against aircraft attacking from low and medium altitude. The 702 has a maximum detection range of 22 nm. [111: pages 77, 78] [27]

Once detected by the ground-based radars of an air defense system, air defense aircraft will be vectored to intercept opponent aircraft either from defensive ground alert (DCA, OCA or HVAA missions) or combat air patrol (BARCAP mission) positions. Several formidable, potential opponent aircraft present challenges to USAF air superiority operations, two of these include the Russian MIG-29 Fulcrum and the Sukhoi Su-27 Flanker-B. Both the Fulcrum and the Flanker contain sophisticated radar systems. The Su-27SK, Flanker-B (K is the export version of the Su-27) has the NIIP S-27 Myech (N-001) track while scan, coherent pulse-Doppler radar system. The N-001 radar operates in the I/low J-band (8 to 12 GHz). This radar has 'look

down, shoot down' capability and can track up to ten targets simultaneously while engaging two. The N-001 radar has a search range up to 130 nm and tracking range up to 92 nm against large size targets and a search range of 54 nm and tracking rage of 35 nm against MIG-21 size targets in the forward hemisphere. The search range (against MIG-21 size targets) is up to 22 nm in the rear hemisphere. The N-001 radar has three different fields of vision: 60° by 10°; 20° by 5°; and 3° by 3° and a search field of 120° by 75°. [23: page 91] [113: page 439] [36] [111: page 264]

The MIG-29C Fulcrum has the RLS RP (N019 Sapfir-29) coherent pulse-Doppler radar system. The N019 radar operates in the I/low J-band (8 to 12 GHz). This radar has 'look down, shoot down' capability and can track up to ten targets simultaneously while engaging one. The N019 radar has a search range up to 55 nm and tracking range up to 38 nm against large size targets and a search range of 38 nm against MIG-21 size targets in the forward hemisphere. The search range varies from 19 nm (against MIG-21 size targets) to 28 nm (against large size targets) in the rear hemisphere. The N019 radar tracking limits are: +60° and −38° in elevation; and 67° in azimuth. [23: page 78] [113: page 402] [111: page 263]

As stated earlier in this section, a radar system operates by radiating EM energy (in this case, radar waves) and detecting the echo returned from reflecting objects (targets). As is the case with EM energy such as light waves, a radar wave can be deflected, diffused, or absorbed. Radar reflectivity is affected by the physical properties of the reflected object and the aspect of the object presented to radar. The amount of reflected radar radiation provide by an aircraft can be expressed by an area metric known as radar cross section (RCS). RCS is a complex measure that depends on the cross-sectional area of that radiated aircraft (geometric cross section); how well the aircraft reflects radar radiation (material reflectivity), and how much of the reflected radar radiation travels back toward the radar system (directivity). For example of the affects of geometric cross section, take an aircraft flying directly toward a radar wave. The physical dimensions presented to the radar system would

246

be fairly small: the leading edge of the wings and tail plus the cross-section of the fuselage; and engines are probably all that would be seen. If the same aircraft now turns through 90°, the radar system would now see the whole length of the fuselage and the slab section of the wings and tail. Clearly, the 90° case would present a larger radar image or radar cross section because there is more of the target for the radar waves to hit and reflect back. [67: page 17]

Additionally, RCS is defined as the cross sectional area of a perfectly conducting sphere of radius a ($RCS = \pi a^2$) that reflects the same amount of energy as the aircraft. This conducting sphere is the only radar object with a geometric cross-sectional area equal to its RCS[15]. [131: page 20]

Factors that determine the detection range of a specific radar system include maximum range of the radar; aircraft (including munitions) radar signature; and jamming techniques used. Appendix B provides a general discussion of these factors.

In recent years, the concept of 'stealth' or low observable technology has garnered a great deal of discussion. Two stealth techniques exist to reduce a radar system effective detection range:

- External shaping and

- Coating the aircraft with radar-absorbing materials (RAMs).

Both these techniques reduce the effective detection range by reducing the amount of reflected radar radiation provided to the radar system (i.e., RCS of the target aircraft). By carefully shaping the external surfaces of an aircraft, one can reduce the RCS of the aircraft. The F-22 incorporates such shaping in its design. For existing non-stealth aircraft designs such as the F-15 or F-16, RAM coatings can cut the RCS. Both shaping and coating are effective against both the typical short

---

[15]This is true only when the circumference of the sphere is at least an order of magnitude larger than the radar wavelength under consideration

wavelengths of conventional air combat aircraft and the longer wavelengths of air and ground-based defense radars. [33: pages 12, 13, 15] [35: pages 53, 54]

Although radar is the main sensor used to detect aircraft in the air, infrared (IR) sensors are becoming an increasingly important and capable means for aircraft detection. IR sensors are passive; these sensors detect infrared energy produced by an aircraft. The majority of infrared energy produced by an aircraft is absorbed by water vapor and carbon dioxide gas in the atmosphere; however, two aircraft detection windows exist in the infrared band. The first window (termed 'mid-IR') occurs at a frequency of 2 to $5\mathrm{x}10^{14}$ Hz (wavelength of 2 to $5\mathrm{x}10^{-6}$ meters). Infrared energy from the heat of an aircraft's engine parts and exhaust falls into this mid-IR region. The second window (termed 'long IR') occurs at a frequency of 8 to $15\mathrm{x}10^{14}$ Hz wavelength of 8 to $15\mathrm{x}10^{-6}$ meters. Infrared energy caused by solar heating or by air friction on the fuselage of the aircraft falls into this long-IR region. Modern dual-band infrared search and track (IRST) systems can look for aircraft in both IR windows. IRSTs are composed of wide field-of-view sensors that use automated detection and tracking routines to find targets in highly cluttered backgrounds. [33: pages 15, 16, 23] [35: page 54] [131: page 18]

Russian air combat aircraft such as the MIG-29 and Su-27 contain sophisticated IRST sensors. The Su-27SK is equipped with the OLS-27 IRST that provides azimuth and elevation information. This IRST can scan large areas and detect aircraft at ranges out to 10 to 15 nm although 5 to 8 nm is a more reasonable range estimate against a non-afterburning, non-IR stealthly aircraft. The OLS-27 is combined with a laser-range finder to provide distance information. The laser-range finder is capable of tracking a target out to a range of 27 nm in the rearward hemisphere and 8.1 nm in the forward hemisphere. The MIG-29C is equipped with UOMZ OEPS-29 IRST which can detect fighter-sized targets out to 8 nm in the forward hemisphere and 10 nm in the rearward hemisphere. [113: pages 402, 439] [23: pages 77, 91]

On the other hand, the USAF has chosen not to install such systems on current air combat aircraft such as the F-15 and F-16. Apparently, this position is beginning to change due to the fact that at least two USAF aircraft participating in the Gulf War were *probably* shot down by Iraqi MIG-29s using a combination of that aircraft's IRST, helmet-mounted sight, and the R-73 Archer infrared air-to-air missile. [185: page 58]

Clearly, current IRST systems do not provide greater detection ranges than radar systems; however, IRST systems are passive (i.e., do not generate the EM radiation needed for detection) where as radar systems are active. Thus, the use of IRST systems can not be detected whereas the use of radar systems can. Stealth countermeasures to IR sensors only address the mid-IR detection band. Little can be done about the long IR band; some IR emission from solar or friction heating of an aircraft's outer skin will always remain present. Mid-IR stealth countermeasures include cooling engine exhaust where the most infrared energy is generated. [33: page 16] [35: page 54]

Once an aircraft is detected, in order to engage the opponent (especially from beyond visual range (BVR)), under most common rules of engagement one must be certain he is first an opponent. Timely aircraft identification is probably the toughest problem in the air environment. This problem is not unique to air-to-air engagements as it applies to surface-to-air engagements as well. Aircraft can be identified visually or electronically. Two types of electronic identification procedures exist:

- cooperative identification and

- noncooperative identification.

Cooperative identification depends on transponders carried by friendly aircraft. In cooperative identification, an unknown aircraft is interrogated with a coded radio signal from a friendly aircraft's Identification Friend or Foe (IFF) electronic identi-

fication system. Friendly aircraft respond with a coded reply and are classified as friendly aircraft. Opponent aircraft do not reply. Unfortunately, neither do friendly aircraft with inoperative equipment. Despite advances in IFF systems, one basic problem has remained: determining whether the lack of a response from an aircraft means that the aircraft is hostile or whether the aircraft is friendly with a non-responsive transponder. The relatively low reliability of the cooperative identification procedure has led to restrictive rule of engagement (ROE) that require several independent means of verifying that a non-replying aircraft is really, truly an opponent before the aircraft can be fired at. The tragic shoot down of two Army helicopters in Northern Iraq in 1994 by F-15Cs provides a vivid illustration of shortcoming of the cooperative identification procedure. [101: page50] [33: page 21] [173: page 294] [111: page 4]

In contrast, the noncooperative electronic identification procedure relies on various radar methods of identifying aircraft without a response and is able to identify most friendly and opponent aircraft. How this is done is highly classified. Nevertheless, one possible means discussed in open sources is to focus a high-resolution radar beam on a target and with advanced signal processing determine a return pattern that is unique to the aircraft and can be compared to the known parameters of friendly and hostile aircraft. For example, with a high-resolution radar beam you might be able count the number of blades in the aircraft's engine fan or compressor. Knowing the blade count tells you the type of engine and can give you a good idea as to whether the aircraft is hostile. [101: page 50] [111: pages 4,5] [33: page 21] [185: page 16]

Most USAF air combat aircrafts are equipped with cooperative and noncooperative electronic identification capability. For example, the E-3 Sentry is equipped with a highly sophisticated IFF system[16], capable of interrogating virtually any IFF transponder in the world within 200 nm. The system reportedly has some noncoop-

---

[16] APX-103 IFF/Tactical Digital Link (IFF/TADIL-C) system

erative capabilities as well. The F-15C Eagle and F-15E Strike Eagle APG-70 radar contains noncooperative identification capabilities and the F-16C Falcon may have noncooperative identification capabilities as well. [33: page 21, 92, 113]

The F-22 will have an improved capability to conduct noncooperative identifications. The F-22's APG- 77 radar will be able to form incredibly fine beams allowing the signal processor to generate a high-resolution radar image of an aircraft through Inverse Synthetic Aperture Radar (ISAR) mode processing. An ISAR-capable radar uses the Doppler shifts caused by rotational changes in the aircraft's position with respect to the radar antenna to create a 3-D map of the aircraft. With a good 3-D radar image, an integrated aircraft-combat system might conceivably identify the target by comparing the image to a stored database. [33:page 23]

A.6   Weapons

The aircraft gun ushered in the era of air-to-air engagement and is, by far, the most widely used and important air-to-air weapon in history. However, the development and employment of capable air-to-air missiles has expanded the combat domain of air-to-air engagements. Since the June 1967 Arab-Israeli War, all major air-to-air engagements have included the use of air-to-air missiles. [93] The advent of air-to-air missiles enabled aircraft to shoot and kill opponents quicker and from greater distances than guns alone have allowed. Aircraft no longer have to close to within hundreds of feet of an opponent and maneuver into a small cone behind the opponent in order to enable the aircraft to hold guns on the opponent long enough to fire sufficient bullets to destroy the opponent. [173: page 1] [17: page 14]

Air-to-air missiles may be divided into two main categories: radar-guided and non radar-guided (or passively guided). Radar-guided missiles are further divided into two categories: semi-active and active. Active radar-guided air-to-air missiles such as the US Advanced Medium Range Air-to-Air Missile (AMRAAM) carry their own radar where as semi-active radar-guided air-to-air missiles such as the Sparrow

(AIM-7) have a radar receiver, but no transmitter and pick up radar signals transmitted by the aircraft and reflected from the target. Semi-active radar-guided air-to-air missiles require the launching aircraft to lock on (with radar) a single opponent and to remain locked on until the missile impacts in order to provide targeting information for the missile. If radar lock is lost from the launching aircraft, the missile will not track the opponent. In contrast, active radar-guided air-to-air missiles can provide their own targeting information. An aircraft launching an active radar-guided missile can break off after missile launch and leave to fight another opponent or exit the combat area. [93: page 46-51]. [50] [17: page 18] [151: page 19]

In the USAF air-to-air missile inventory, the AMRAAM has replaced the final version of the Sparrow, the AIM-7M. Unfortunately, the Sparrow was plagued by significant technological flaws during its five decades of the service. The AIM-7 was over four times as effective in Desert Storm as it was in Vietnam; however, the missile only achieve a 36% kill rate in Desert Storm (9% in Vietnam). Additionally, almost half of the AIM-7s launched in Desert Storm failed to function properly. [33: page 127]

The current deployed version of the AMRAAM is the AIM-120A. Termed the 'Slammer' the AIM-120A is 7 inches in diameter and 72 inches long with a center (stabilizing) fin of 21 inches and a rear (guidance) fin span of 25 inches. The AIM-120A weights 335 lb. The AIM-120A had a complete air intercept radar system. The radar of the firing aircraft sends the AMRAAM the three dimensional heading and speed of the target aircraft, then the missile flies out to a point where it switches to its own radar[17]. If the target aircraft is within the radar 'cone' of the AIM-120A's seeker, the missile locks up the target aircraft and then proceeds to intercept. Addi-

---

[17]AMRAAM may be fired in a number of different modes. At longer range, the missile may be continually updated with fresh target information from the launching aircraft's radar while in flight until the missile's radar takes over. Alternatively, if the aircraft survivability is an issue, the launching aircraft can let the missile go without mid-course updates and trust the AMRAAM's active radar to work unaided. Finally, for shorter engagement distances, lock-on of the missile's seeker can be accomplished prior to launch.

tionally, the AIM-120A is equipped with a proximity fuse that detects and detonates when a target aircraft is within lethal range. The impetus for the AMRAAM's introduction into service was the gradual implementation of ground-based and airborne sensors that provided a detailed picture of the air environment at ranges impossible a decade ago. Information from assets such as AWACs piped via datalink to the aircraft combined with aircraft sensors allow the identification and targeting of opponent aircraft at very long ranges with a high degree of accuracy. [33: page 130, 131] [38]

As an illustrative example, consider launching an AIM-120A at a target aircraft from an F-16C. In the F-16C, the pilot selects an air-to-air mode for the radar such as BORE (Boresight - i.e., where the radar is sighted down the centerline of the aircraft) or TWS, DOGFIGHT (where the radar is in a mode useful for close-in dogfighting). The pilot then selects the AIM-120A from the stores control panel and chooses either SLAVE or BORE to program the AIM-120's radar to accept commands from the F-16's onboard APG-68 radar.[18] Once a radar contact is established with the target aircraft, the F-16's onboard weapons computer establishes a fire control solution, including elapsed time from AIM-120A launch until the AIM-120A's radar goes active. At this point, the F-16's Head Up Display (HUD) will begin to give the pilot steering cues to bring the F-16 into range to fire. Once the HUD gives the pilot an IN RANGE indication, the pilot press the weapons release switch on the control stick. The AIM-120A is launched and will accept updates from the F-16's radar (if pilot has selected a FIRE AND UPDATE mode) until either pilot breaks radar contact from maneuvering or the AIM-120A hits the target. At this point, the firing aircraft is free to fire another missile, seek another target aircraft, or leave the engagement area. [33: page 132]

---

[18]The SLAVE option locks the seeker onto whatever target the F-16's radar is currently tracking, while BORE points the F-16's radar straight ahead along the line of flight where the first target the radar sees is locked.

Other capable radar-guided missiles include the semi-active Russian Vympel R-27 (NATO code name AA-10 'Alamo')[19]. Future developments include the French MICA, UK FMRAAM, and Russia's first active radar-guided missile the R-77, (NATO code name AA-12, 'AMRAAMski'). [38] [20]

The majority of non radar-guided (passively guided) air-to-air missiles are heat-seeking or infrared (IR) missiles such as the US Sidewinder (AIM-9)[20]. The frequency of the IR portion of the electromagnetic spectrum is just below that of visible light and well above that of radar. These missiles guide on the IR energy produced by the opponent aircraft. All IR missiles are launch and leave or fire and forget; i.e., the missile is autonomous after launch, tracking on the opponent at which the missile was directed. For example, to launch an AIM-9 at a target aircraft in the F-16C all that is required is to select AAM from the stores control panel. The seeker in the AIM-9 then looks for the target aircraft in front of the F-16C.[21] When the seeker is locked on to the target aircraft, an audio 'growl' tone in the pilot's headset alerts the pilot that the AIM-9 is ready to launch. All the pilot of the F-16C has to do to launch the AIM-9 is squeeze the trigger. Once an IR missile is in flight, the firing aircraft is free to fire another missile, seek another target, or leave the engagement area. [33: page 15, 121, 125] [173: page 38] [17: page 20]

The currently deployed, third generation version of the Sidewinder is the AIM-9M. Termed the 'Mike' the AIM-9M is 5 inches in diameter and 113 inches long with a forward canard wingspan of 15 inches and a rear stabilizer wingspan of 24.8 inches. The AIM-9M weights 194 lb. The AIM-9M has an off-boresight capability[22] of 27° and scans in both short- and middle- wavelength infrared light as well as the

---

[19] Also, IR versions of this missile.

[20] "When the first test launches of what would become the Aerial Intercept Missile Nine (AIM-9) were conducted in 1953, the missile's snakelike flight path towards the test targets provided the name it would carry for the next half century of service, the Sidewinder". [33: page 121]

[21] Or if the radar is already locked onto the target aircraft, the seeker head can be slaved to the radar and the seeker will also lock onto the target aircraft.

[22] The ability to lock up a target off the firing aircraft's centerline.

long wavelength (ultraviolet) spectrum. Additionally, the AIM-9M is equipped with a proximity fuse like the AIM-120A. If the missile should miss the target (a rare event due to the accuracy of the guidance system), the warhead will still detonate 'spewing' its fragmentation pattern at the target aircraft. [33: page 123, 124] [20]

Other capable IR missiles include the Russian Vympel R-73 (NATO code name AA-11 'Archer') and Israel Python 4. Many defense experts believe the AA-11 has provided its combat aircraft, the Sukhoi Su-27 'Flanker' and the MiG-29 'Fulcrum' with a significant advantage in short range combat scenarios such as 'bug-out', the moment that an friendly aircraft breaks out of the engagement and 'rundown', the subsequent tail chase over NATO aircraft.[23] The AA-11 has an off-boresight capability of 45° expanded to 60° immediately after launch and a deflector-based thrust-vectoring system that provides for increased agility. Additionally, the AA-11 is the first air-to-air missile to be coupled with an operational helmet mounted sight (HMS). The Python 4 has an off-boresight capability of approximately 60° and maneuvers using aerodynamic control. The AA-11 and Python 4 have the led the US to pursue an upgrade to the AIM-9M, the AIM- 9X. AIM-9X program requirements call for a missile with a 60°+ off-boresight capability using a HMS. [151: page 19] [38] [20] [33: page 134].

One method for visualizing the capabilities and limitations of an air-to-air weapon is to study weapon envelopes. A weapons envelope is the area around the target aircraft where missiles or gun can be effective. The weapons envelope is defined by:

- Angle-off - the difference, measured in degrees, between the firing aircraft's heading and the target aircraft's;

- Range - the distance between the firing aircraft and the target aircraft; and

---

[23]This is based on an operational assessment conducted when German reunification brought a squadron of AA-11 equipped MIG- 29s into NATO service. The assessment indicated that Western aircraft equipped with the AIM-9M would be at a distinct disadvantage in air-to-air engagements at ranges below 2.7 nm. [20]

- Aspect Angle - the difference, measured in degrees, between the target aircraft's tail and the firing aircraft.

The weapons envelope of all-aspect missiles such as the AIM-9M and the AIM-120 look like a doughnut with the outside ring being the maximum range of the missile and the inside ring being the minimum range. Figure 49 shows a typical weapon envelope for an all-aspect missile[24]. The weapon envelope in Figure 49 has an oval shape with more of the envelope area in front of the target aircraft than behind. This disportionality is due to the fact that a missile fired at a high aspect (from in front of the target aircraft) has a greater effective range than a missile fired at low aspect (from behind). When a missile is fired at a target aircraft head on, the fact that the target aircraft is flying towards the missile will help the missile reach its target by reducing the distance the missile needs to travel to reach the target aircraft. This difference may vary by as much as 3-5 to 1 from a rear (0°) aspect missile shot. Although shown in two dimensions in Figure 49, an actual air-to-air weapons envelope is three dimensional; however, the envelope would probably only slightly vary in the third dimensional depending on the firing aircraft's position relative to the target aircraft's plane of symmetry. [19: pages 11, 12, 20, 21] [173: pages 25, 46]

When a target aircraft turns, the weapons envelope shifts. Figure 50 shows a typical weapon envelope for an all-aspect missile for a target aircraft making a 5G turn[25] Generally, the limits of the maximum and minimum ranges in front of the target aircraft move in the direction of the turn while the ranges behind the target aircraft move in on the "belly side" of the turn. [19: pages 20, 21, 23]

In addition to changing with target aircraft maneuvers, the weapons envelope of an air-to-air missile also changes with firing aircraft/target aircraft speeds and altitude. Clearly, in order to get a full picture of the capabilities of a particular air-to-air missile, pilots must consider many weapon envelope charts covering a wide

---

[24]Figure 49 is taken from [19: page 20].
[25]Figure 50 is taken from [19: page 23].

Note: Weapon Zone is taken at a fixed Angle-Off
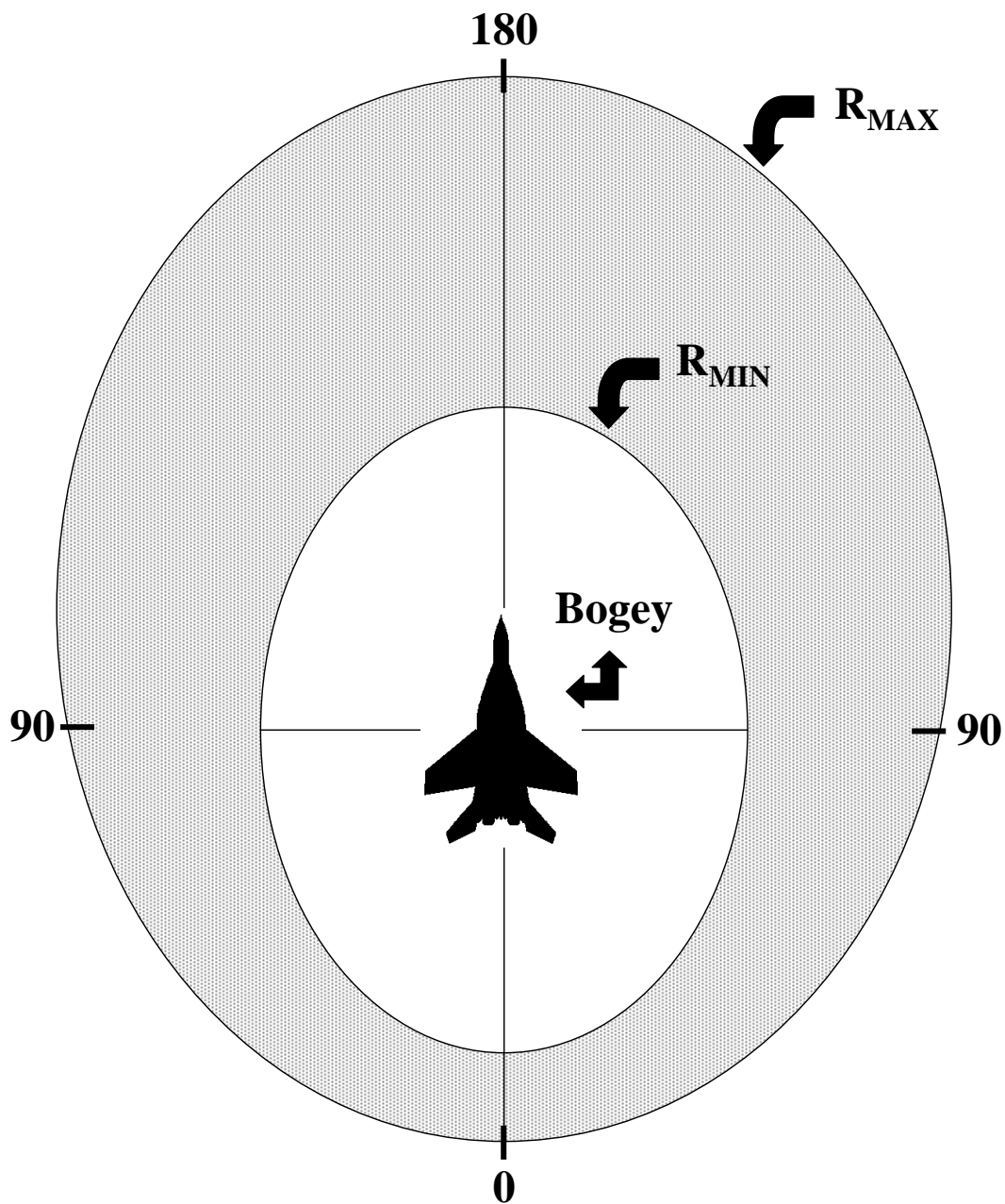
Figure 49. Weapons Envelope of All-Aspect Missile
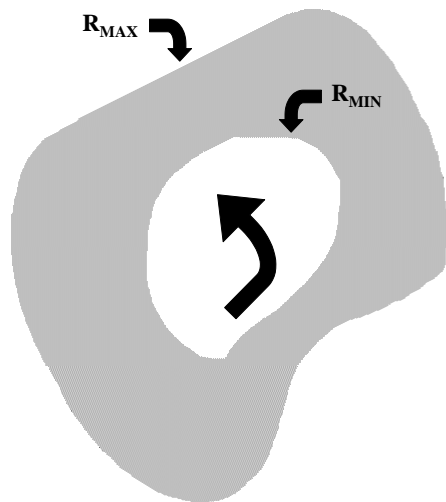
Figure 50. Weapons Envelope of All-Aspect Missile under 5Gs

range of target aircraft maneuvers, firing aircraft/target speeds, and altitudes. Even if pilots could draw each weapon envelope from memory, they still must determine the parameters necessary to recognize which envelope they are in and their position (range and aspect) within the envelope. In order to solve the hard problem of envelope recognition, most combat aircraft are equipped with a tracking radar system and fire-control computer. These systems can accurately assess and display the missile envelope capabilities as well as other limits that might be deemed necessary. In order to make inputs to the radar computer it is necessary to maintain a radar track, rather than just detect the target aircraft. The transition from radar detection to track is referred to target lock. Depending on the sophistication of the radar, this may be an automatic process or a manual process. [173: pages 48, 49]

As shown in Figures 49 and 50, the range of an air-to-air missile can be significantly affected by whether the air-to-air engagement takes place head-on, on the beam, or in a tail-chase configuration. For more capable air-to-air missiles, the best engagement range is usually achieved with the target 30° off nose of the firing aircraft (150° degrees aspect). As mentioned, the range of an air-to-air missile can be significantly affected by both the speed and altitude of the firing and target aircraft. The greater the speed of the firing aircraft, the greater the range of the missile since aircraft speed adds to the missile speed[26]. Also, missile range increases with altitude, since the range of the missile's engine increases with altitude[27]. The range capabilities of air-to-air missiles are generally divided into three groups:

- *Short*: to indicate a missile with a maximum range of 5.4 nm (10 km) or less;

---

[26]Note: As discussed in section A.4, there is a reduction in aircraft maneuverability at speeds above Mach 1.0. Hence, there is a practical limit on aircraft firing speeds if one desires to retain maneuverability. For example, F-15 Eagle generally launch BVR missiles at speeds around Mach 1.1. This firing speed strikes a good balance between increased missile range and adequate maneuverability.

[27]Note: However, missile maneuverability decreases with altitude. Furthermore, missile range is greater when firing from a higher altitude to a target at a lower altitude than firing from a lower altitude to a higher altitude. This is the result of potential energy effects described in section A.4.

- *Medium*: to indicate a missile with a maximum range greater than 5.4 nm (10 km) and less than 54 nm (100 km); and

- *Long*: to indicate a missile with a maximum range greater than 54 nm (100km).

These range groupings are in a constant state of flux as new, higher range missiles enter national inventories. For example, the AIM-9M is consider a short range missile even though it can theoretically fly out to a range of 11 nm against a non-maneuvering target. [112: page 1] [192:pages 64 -68] [173: pages 52-55] [33: page 125]

Typically, IR (non radar-guided) missiles have a shorter range capability than radar-guided missiles. For example, the IR AIM-9M Sidewinder has a maximum engagement range (approaching a head-on target) of 9.6 nm while the radar-guided AIM-120 AMRAAM has a maximum engagement range of approximately 40 nm. In the past, IR missiles were generally only useful at visual ranges; that is, when pilots of opposing aircraft could see each other (distances of several nautical miles or less). This limitation placed IR missiles firmly in the short range missile category. However, more capable IRST sensors (reference section A.5) especially those on Russian aircraft such as the Su-27 and MIG-29 have helped push the launch envelope for IR missiles well past the short range missile category into the medium range missile category. For example, the Russian AA-11 Archer has an engagement range from .16 nm to 16 nm[28] depending of the engagement scenario. Additionally, IR versions of the Russian A-10 Alamo have a maximum engagement range of 38 nm for medium and high altitude and 10.7 nm for low altitude targets using high IR emitting (afterburner) engines[29]. [112] [23] [77]

---

[28]Range data is for newer R-73M2 version. The older R-73M1 version had an engagement range of .16 nm to 10.7 nm.

[29]Referenced distances are for the D version of the AA-10 missile (Russian designation: R-27ET1). The B version of the AA-10 missiles (Russian designation: R-27T1) has a maximum engagement range of 27 nm for medium and high altitude and 10.7 nm for low altitude targets using high IR emitting (afterburner) engines. The AA-10 D version has a longer range solid propellant rocket than the AA-10 B version.

In contrast to IR missiles, radar guided missiles can usually be launched at ranges much greater than a few nautical miles placing them in the medium and sometimes long range missile category. The United States has emphasized fighting air-to-air engagements from longer ranges and has focused development funding on radar guided missiles that can operate in the heart of the medium range missile envelope (25 to 30 nm or more). For example, the AIM-120 AMRAAM can kill an approaching head-on target aircraft out to approximately 40 nm[30] while in a 'tail chase' engagement, which requires the missile to overtake the target, this range drops to approximately 12 nm. The minimum engagement range for the AIM-120 is approximately 1 to 2 nm. Although Russia has emphasized development of shorter range IR missiles, the Russians still have developed a capable semi-active radar guided missiles in the AA-10 Alamo. The Russian AA-10 Alamo has a maximum engagement range of 38 nm for medium and high altitude and 10.7 nm for low altitude targets[31]. [17:page 20]. [33: page 125] [38] [151: page 19] [112] [23] [77]

The speed and operating altitude capability of the above IR and radar guided missiles varies. The AIM-9M Sidewinder has a speed of Mach 2.5 (with a maximum flight time of one minute) and an unreported operating altitude. The AA-11 Archer has a speed of 1350 knots and an operating altitude capability of 65 ft (20 m) to 65,000 ft (20 km). The AA-10 Alamo missiles (IR and radar versions) have an unreported speed with an operating altitude capability of sea level to 70,000 ft (21 km) for the A and B versions with the C and D versions possibly having some limited capability to 90000 ft (27 km). Finally, the AIM-120 AMRAAM had a reported speed of Mach 4 and an operating altitude capability of sea level to 65,000 ft (20 km). [112] [77]

---

[30]The maximum engagement range of the AIM-120 AMRAAM is quite open to debate in the literature. Jane reports a maximum engagement range of 27 nm while Brassey reports a range of 30nm; Clancy a range of 40 nm; and Missile Forecast a range of 54 nm.

[31]Referenced distances are for the C version of the AA-10 missile (Russian designation: R-27ER1). The A version of the AA-10 missiles (Russian designation: R-27R1) has a maximum engagement range of 27 nm for medium and high altitude and 10.7 nm for low altitude targets. The AA-10 D version has a longer range solid propellant rocket than the AA-10 A version.

Typical measures of an air-to-air missile's effectiveness are: the 'no-escape' zone and the single-shot kill probability. The 'no-escape' zone refers to the volume of airspace around the firing aircraft where a target aircraft will be unable to get away from a launched missile, no matter how hard and fast the target aircraft maneuvers. The probability of the missile achieving a single-shot kill, P_{ssk}, is defined as the product of five factors:

$$P_{ssk} = P_e * P_l * P_g * P_f * P_k \qquad (101)$$

where

- P_e - whether an aircraft successfully forces an opponent into an engagement;

- P_l - whether the pilot fires a missile;

- P_g - whether the missile guides successfully;

- P_f - whether the fuse works; and

- P_k - whether the warhead kills the target.

Assuming independent missile launches, the probability of killing a single aircraft can be increased by firing multiple missiles. For example, consider these two missile firing strategies[32]:

- Shoot: In this strategy, an aircraft fires one missile at the target aircraft and waits for the outcome of the missile shot before deciding to firing another missile; and

- Shoot-Shoot: In this strategy, an aircraft fires a second missile at the target before knowing the outcome of the first missile.

The probability of kill is P_{ssk} and $P_{ssk} * P_{ssk} + 2 * (1 - P_{ssk}) * P_{ssk}$ for the 'Shoot' and 'Shoot-Shoot' strategies respectively. If $P_{ssk} = .60$, this results in .60 for the

---

[32]There are higher order cases such as Shoot-Shoot-Shoot; however, the increase in probability of kill is marginal when comparing current missile P_{ssk}s.

'Shoot' strategy and a .84 for the 'Shoot-Shoot' strategy. [33: page 127] [93: page 46,47] [68: page 5]

Radar-guided air-to-air missiles unlike IR air-to-air missiles are generally unaffected by environmental factors such as weather or time of day. This is a direct result of the fact that radar-guided air-to-air missiles provide their own source of illuminating energy required for guidance and homing[33] where as IR missiles depend on the illuminating energy provided by the target aircraft. IR missiles do not function well in clouds or rain because IR energy is absorbed and dissipated by water vapor. Additionally, the greater the environmental background energy or radiation (e.g., sun; reflections off of water, snow, clouds, or terrain.), the stronger the illuminating energy provided by the target aircraft must be in order for an IR missile to see (guide/home on) the target aircraft. Although radar-guided air-to-air missiles are affected by environmental factors such as weather or time of day, they are susceptible to countermeasures (e.g. jamming) that affect the illuminating energy generated by the missile. Missile countermeasures will be discussed in greater detail in section A.8. An additional disadvantage of an active radar-guided missile is its potential for fratricide. Once an active radar-guided missile is separated from the firing aircraft and locked onto a target aircraft, the firing aircraft has no further role in the missile's operation. In a close, crowded engagement, the active radar-guided missile may not be able to discriminate between opponent and friendly aircraft and fratricide could occur [17:page 19] [173:pages 38, 40] [139:page 33]

In contrast to present day air-to-air missiles, today's aircraft guns appear relatively 'low tech'. However, this appearance does not imply that the aircraft gun serves no useful purpose in today's air engagement environment; in fact, quite the opposite is true. Aircraft guns unlike air-to-air missiles can be used against surface targets and have no minimum range requirements. The aircraft gun is all-aspect weapon with a circular weapon envelope around the target aircraft depicting the

---

[33]In the case of semi-active radar missiles, the aircraft provides the illuminating source.

gun's maximum range. Additionally, a gun projectile is not vulnerable to counter-measures once it leaves the barrel. The MIG-29 and SU-27 come equipped with the GSH-301 30 mm cannon. The aircraft gun used on USAF fighters is the M61MA1 Vulcan 20 mm cannon. The version used in the F-15C Eagle holds 940 rounds which yields about 9.4 seconds of firing time. Ammunition used in the Vulcan includes the PGU-28, which has armor piercing, explosive fragmentation, and incendiary capabilities, all in a single round. [19: page 22] [139: page 34] [33: page 68]

A.7 Command, Control, Communications and Information

Command, control, communications, and information (C3I) play a significant role in air-to-air engagements. As an example consider the first night of the air battle in the Gulf War. During the first night of the war, USAF aircraft destroyed the majority of the Iraqi air defense network. Without ground-based command and control, Iraqi interceptors "blasted around" trying to intercept coalition strike aircraft they could not detect. Another example from the Gulf War involves a four ship of F-15Cs patrolling the skies over Baghdad. An E-3 AWACS warned the flight of an Iraqi Mirage F-1, which had just taken off and was closing on the four F-15Cs from astern. With the AWACs warning, one of the F-15Cs broke hard, turned behind the Mirage and destroyed the F-1 before it could inflict damage on the flight. [93: page 172, 175]

At the heart of USAF C3I is the E-3 AWACs. As mentioned in section A.5, the E-3 AWACS flies on the edge of the air battle area, scanning airspace for unidentified aircraft (referred to as 'bogies') and opponent aircraft (referred to as 'bandits'). Typically, an AWAC is organized into separate surveillance and control sections. In the surveillance section, three to five technicians monitor the air traffic in the assigned airspace and pass on information to the control section. Composed of two to five weapons controllers sitting at multi-purpose consoles, the control section guides friendly aircraft to intercept bandits or bogies. Depending on its particular mission,

Band	Frequency (MHz)	Wavelength (m)
HF	$3 - 30$	$100 - 10$
VHF	$30 - 300$	$10 - 1$
UHF	$300 - 1000$	$1 - .3$

Table 21. Communication Bands

an AWACs also may carry senior staff officers, radar technicians, radio operators, a communications technician, and a computer technician. [44: page 63] [33: page 115]

An important aspect of the C3I is the communications equipment available to participants in air-to-air engagements. Voice communication generally takes place in three frequency bands of the electromagnetic spectrum: HF, VHF, and UHF. Table A.7 shows the frequency and wavelengths of these communication bands. The majority of voice communication occurs in the VHF and UHF bands. VHF/UHF communication systems are essentially line-of-sight systems where the range of the equipment is directly proportional to the aircraft's altitude. In order to provide the necessary range for long distance flights, the HF band is used. The F-15C Eagle is equipped with a AN/ARC-164 VHF/UHF auxiliary transceiver and VHF/UHF auxiliary transceiver with cryptographic capability. The basic ARC-164 covers the VHF/UHF band, providing 7000 channels over the range 225 to 400 MHz in 25 KHz increments. In addition, the F-15C is equipped with AN/ARC-190 HF transmitter/receiver. This system provides 280,000 channels over the range 2 to 30 MHz in 100 kHz increments. [88: page 30] [111: page 5] [112: page 320, 321]

In order to provide voice communication capability in a heavy electronic coutnermeasure (ECM) environment, many USAF communication system including the F-15C's AN/ARC-164 are equipped with Have Quick capability. Have Quick provides the user with a jam-resistant VHF/UHF capability by giving the communication systems a frequency-hopping technique. The frequency hopping technique is implemented by storing a pattern of frequencies to be used for a given day and utiliz-

265

ing the pattern according to the time of day. The strength of Have Quick lies in the apparently random manner in which frequencies are jumped. [44: page 63] [113: page 60]

Data communication is another important aspect of any airborne communications system. Airborne communication systems such as the F-15C Eagle's AL/ARC-164 provide for general data transmission. Also, USAF air combat aircraft such as the F-15C and F-16C Falcon are equipped with tactical air navigation (TACAN) radios which allow pilots to determine the separation distances of aircraft within the flight group to a resolution of a tenth of a mile. In addition to the standard communication and TACAN systems, the new Joint Tactical Information Data System (JTIDS) provides for expanded data communication between aircraft. JTIDS terminals allow the 'linking' of any aircraft so equipped to an aerial, local network. The secure (i.e., unjammable and untappable) data link allows the sharing of information from a plane's sensors and other systems with other aircraft, ships, and ground units equipped with JTIDS terminals. JTID terminals are able to transmit a full situational report, including radar contacts, sending aircraft position, altitude, and heading, and even fuel and armament status (counting gun, bomb, and missile rounds onboard). JTIDS terminals are currently available on the E-3 Sentry AWACs, as well as new E-8 Joint-STARS ground surveillance aircraft, and select F-15C aircraft. [113: page 59] [44: page63] [33: page 67]

JTIDS operates on 51 frequencies in the 960 to 1215 MHz frequency band sharing with TACAN, but avoids the IFF transponder frequencies which are also in this band. Future "data-fusion" efforts beyond JTIDS include collecting infrared and ultraviolet scans from satellites or a RC-135 Cobra Ball which can pick up rapid movement such as a rocket launch and electronic intelligence (ELINT) from platforms such as the RC-135V/W Rivet Joints which can plot opponent radars and their status. Clearly, systems like JTIDs are the only "quiet" form of communication in strict emissions control situations. [113: page 51] [185: pages 18, 57]

266

A.8 Electronic Warfare

Electronic warfare (EW) is defined as an action to "determine, exploit, reduce, or prevent" use of the electromagnetic spectrum and is organized into two major categories: EW support measures (ESM) and electronic countermeasures (ECM). ESM involves actions taken to "search for, intercept, locate, record, and analyze" radiated electromagnetic energy while ECM involves actions taken to "prevent or reduce" use of the electromagnetic spectrum. [178: page 9.2] [101: page 36] [144: page 25]

ESM uses passive radar intercept or warning systems which detect radar emissions from aircraft, radar missiles, and surface-to-air missile installations. ESM systems are preset to cover the portion of electromagnetic spectrum used by expected threat radars. The simplest system is the radar warning receiver (RWR) which supplies the relative bearing of the emission on an aircraft display. Most combat aircraft have RWRs which are tuned to provide a warning only when an opponent's fire control radar has established a lock on the aircraft. ESM systems then increase in complexity through tactical ESM to full ELINT (intelligence-gathering) capability. These systems are designed to find and analyze radar emissions and then classify the type of radar producing the emissions. ELINT capable combat aircraft include specialized EW aircraft such as the EF-111A Raven. The range at which a radar emission is detected by a radar warning system is inversely proportional to the squared radar range rather than the quaded radar range need for detection. For this reason, radar warning systems can detect a radiating radar at far greater distances then those of radar's own target detection capability. [178: page 9.3, 9.4] [33: page 24]

With its Block 30/35 update, the E-3 Sentry will be equipped with four passive ESM frequency receivers that are capability of intercepting radar emissions from more than 225 nm. These receivers are also capable of detecting the mode in which the radar is operating (e.g., wide-scan, narrow-scan, etc.). This ESM capability can be used to provide raw data as to the target's height and speed as well as to

classify the radar system and hence the aircraft using the system. For example, search emissions characteristic of an NO-193 Slot Back radar system would indicate the target to be a MIG-29. [185: page 16]

The F-15C Eagle and F-15E Strike Eagle are equipped with the Loral AN/ALR-56C radar warning receiver (RWR). This RWR has a digitally controlled, dual-channel receiver covering the E- and J-bands (2 to 20 GHz) and is capable of sorting and identifying threat aircraft. The AN/ALR-56C is part of the Tactical Electronic Warfare System (TEWS) that provides control of all self-protection systems. The TEWS also includes the Magnavox AN/ALQ-128 electronic warfare system. This equipment probably covers the area beyond the H/I/J band of the AN/ALR-56. The Su-27SK Flanker and MiG-29C Fulcrum have the SPO-15 Beryoza (L-006) 360° RWR. On the Su-27SK, the RWR sensors are mounted on the sides of the air intakes and on the tailboom while on the MiG-29C, the sensors are on the wingroot extensions, wing tips, and port fin. [111: pages 263, 264, 575, 576]

In addition to radar warning systems, ESM also include dedicated missile-warning systems (for both infrared and radar missiles). Historically, 80% of all aircraft shot in an air-to-air engagement never detected the opponent that killed them. Missile-warning systems can provide 360° spherical coverage around the air-craft, enabling a pilot to know when a missile has been fired. [33: page 24]

In contrast to a ESM system's objective to obtain information, the objective of an ECM system is to deny the opponent information (i.e., detection, position, track, and classification). ECM systems may be classified in a number of ways: main purpose, active or passive, etc. The most common type of ECM is radar jamming. Radar jamming denies the opponent the use of specific wavelengths or portions of the electromagnetic spectrum by transmitting signal(s) that degrade the receptive capability of radar systems. Radar jamming may be active (powered signals) or passive (no power required). Two typical examples of radar jamming are noise jamming and deceptive jamming. Noise jamming is an attempt to produce

a strong signal that will overpower a target aircraft's return when it is received by an opponent radar. Since reflected target energy is more sensitive to target range than received noise, noise jamming is effective at longer distances but will eventually become ineffective (i.e., 'burnthrough') at closer distances. Deceptive jamming involves techniques that generate false targets on an opponent aircraft's radar. [178: page 9.4, 9.5] [101: page 36] [173: page 56, 57]

Radar Jamming is deployed/employed in five major ways: standoff, escort, mutual support (cooperative), self-screening, and stand-forward. In the standoff jamming (SOJ) case, the jamming platform remains close, but outside the range of the opponent's weapons. SOJ platforms typically employ high-power jamming in order to disrupt radar systems at the long ranges required to stay out of opponent's weapons range. In the escort jamming case, the jamming platform accompanies the strike package, jamming radars to protect the strike aircraft. In the mutual support case, the jamming capabilities of individual jamming platforms are coordinated as combat elements against several radar systems. In the self-screening case, a jammer is used to protect the carrying aircraft. For example, F-15C carries the internally mounted Northrop ALQ-135(V) radar jammer, which operates automatically, requiring only that pilot turn it on. Finally, in the stand-forward jamming case, the jamming platform is located between the strike package and the opponent radars and jams the radars to protect the strike package.[34] [178: page 9.5] [33: page 68]

In addition to radar jamming, one can jam the data and voice communications of opponent aircraft degrading the command and control. For example, in the 1982 Lebanon air war, the Syrians launched waves of aircraft to intercept the Israelis, but the Israelis jammed the data and voice communications on which the Syrian pilots were dependent. As a result, the Syrian fighters were reduced to uncontrolled

---

[34]The stand-forward jammer is usually within the range of opponent weapons and may suffer a high attrition rate. Thus, only the use of relative low-cost remotely piloted vehicles is practical for this mission.

singles trying to operate against a superior number of well-controlled opponent aircraft. [193: page 49]

ECM is also used to defeat air-to-air missiles. For radar-guided air-to-air missiles, several classes of ECM exist. The first class are devices like gate stealers, that attack the tracking circuitry of missile without angle deception. The next class is expendables. Passive expendables like chaff and blivets and; more advanced forms, like illuminated chaff and active expendables. The next class is high power systems like Cross Eye and Cross Pole that disrupt or break angle track. For IR air-to-air missiles, the conventional ECM, like flares and flash lamps, are designed to defeat reticle (and pseudo imaging) seekers. The AN/ALE-47 chaff/flare dispenser in the F-15 Eagle automatically programs the ejection of chaff/flare cartridges in response to inputs from the ESM systems (e.g. in response to evasive aircraft maneuvers). Full automation takes the guesswork out of timing the good discharge of chaff/flare. [189: page 40] [185: page 161]

In order to counter opponent ECM, many aircraft and missiles are equipped with electronic counter-countermeasures (ECCM) to ensure the effective use of the electromagnetic spectrum despite opponent ECM. For example, many radars are "frequency-agile" over a wide band. If the rate of frequency agility is fast enough, a jammer may not be able to follow the frequency changes and the radar can counter the jamming effects.[35] A key role is also played by those ECCM techniques which cannot be classified as electronic, such as human factors, methods of radar operation, and radar deployment tactics. For example, emission control - the appropriate assignment of operating frequencies to various radars - can impact ECM. [178: page 9.1, 9.5]

---

[35] This type of ECCM is effective against spot jamming where the jamming frequency is centered around a narrow bandwidth. Barrage or broadband jamming radiated across the entire band of the radar spectrum could still jam frequency agile systems.

As a final note, many ESM systems control the deployment and operation of ECM and ECCM; with the link between ESM, ECM, ECCM often automatic. For example, a missile-warning system could automatically deploy expendable counter-measures (chaff and flares) or a radar system may automatically change frequencies when jammed. Automated ESM systems reduce pilot workload and react in real time to threats such as a missile launch. [178: page 9.3] cite[page 24]Cla95 [173: page 57]

A.9 Tactics and Maneuvers

The sole purpose of air combat aircraft is to destroy other aircraft. The aircraft itself is only a weapons platform designed to bring weapons into a firing position against an opponent aircraft. The goal of all air combat aircraft maneuvers and tactics is to position the aircraft into weapon firing parameters while simultaneously frustrating those of the opponent. An aircraft maneuver positions an aircraft in the three dimensional combat arena by actuating its control vectors of roll, pitch, and thrust. An aircraft tactic on the other hand is a series of these maneuvers designed to accomplish a specific objective in the air-to-air engagement. For example, suppose the mission for a particular group of friendly aircraft is to intercept a group of opponent strike aircraft (i.e., a BARCAP mission). The friendly aircraft will first assess the initial geometry of the engagement and select an air tactic to establish contact with the opponent aircraft. After acquiring the opponent aircraft with their sensors, the friendly aircraft will attempt to place the opponent aircraft within the lethal firing envelope of their weapons by a series of appropriate maneuvers. As the friendly aircraft close in on the opponent aircraft, the opponent aircraft can be expected to respond with a series of their own tactics. Hence a sequence of interactive maneuvers and tactics will develop between the two groups of aircrafts. In general, the actual maneuvers and tactics selected in a given air-to-air engagement depend on each pilot's understanding or perception of the combat situation. Tactics (and the corresponding maneuvers) are designed to take advantage of the strengths of

the aircraft and weapons and exploit any weaknesses in the opponent. Each tactic has strong and weak points and can be anticipated and defeated by an opponent if applied too often with the same routine. [173: pages 1, 274] [17: page 10] [107: pages 2.1,2.2]

Perception is perhaps the most important factor in the air environment in determining tactics and maneuvers. The task of the pilots in the flight is to obtain as much tactical information (electronic, visual, communicated) as possible and then to filter and analyze the information. The ability of each member of the flight to gather and to pass critical information to other members of the flight plays a significant role in success in the air environment. An important concept in the employment of the tactics and maneuvers is the organization of the flight into aircraft or ship elements. USAF aircraft typically engage opponent aircraft in two or four ship elements. For fuel economy[36] as well as tactical reasons, a tactical (spread out) formation is preferable to a close formation. In a tactical formation, aircraft are position abreast about 1 to 6.5 nm (depending on air environment conditions) and are usually stacked high or low about 2.5 nm. Aircraft elements coordinate tactics and maneuver and share workloads. Once engaged, flights usually attempt to maintain only two-ship elements. This size has been found to be most effective in maintaining mutual support while minimizing coordination problems. [173: page 291, 309, 310] [185: pages 52, 56] [130]

In the majority of cases, the side taking the "first shot" in an air-to-air engagement seizes the initiative by forcing opponent aircraft to react against the missile(s) to prevent a kill. A missile in the air generally attracts the attention of the pilot in the target aircraft often causing him to forget about the firing aircraft or firing his own weapons. For this reason, first shot can be critical even if not successful. A

---

[36]Two fuel states are generally considered in mission planning: Bingo fuel and Joker fuel. Bingo fuel is the fuel state that allows an air combat aircraft to return to a tanker or home base with an acceptable safety reserve. Joker fuel is the fuel state that the pilot of an air combat aircraft should think about disengaging if involved in an air-to-air engagement.

missile in the air establishes a psychological set between combatants that places the pilot of the target aircraft in a defensive frame of mind. This reaction may make the opponent more vulnerable to other weapons as well. Getting the first shot requires long-range weapons and identification systems. As discussed in section A.6 all-aspect air-to-air missiles can be fired at a target aircraft from any direction. Although all-aspect capable, most of these missiles are better in some situations than in other, with beam aspects often causing the most difficulty (reference Figure 49). One way to increase a missile's effective range is to fire at a significantly higher altitude than the opponent aircraft. This gives the missile a reserve of potential energy that can be converted into kinetic energy. [101: page 38] [17: pages 21,22] [173: pages 51, 121, 122]

As shown in section A.3, a typical air-to-air engagement starts with 40 nm plus range separation between friendly and opponent flights. An airspace controller (AWACS or GCI) may point the (friendly) flight in the right direction, but the flight must eventually find the opponent aircraft on their sensor systems. In order to get into weapons parameters, the friendly flight must 'intercept' (reduce the range between) the opponent flight. The choice of intercept tactic(s) depends on the engagement situation. Many factors contribute to this choice - for example, mission goals, rules of engagement; the geometry of the engagement; number of friendly and opponent aircraft; friend and opponent aircraft and weapon capabilities; capability of the pilots; offensive and defensive potential[37]; and so on. In addition to placing the friendly flight within weapons parameters of the opponent flight, the intercept tactic(s) chosen should keep the opponent flight from getting into weapons parameters. Intercept tactics have names such as single-side offset; trail; sweep; pincer; and drag.[38] The basics of almost all of these intercepts involve front-quarter and

---

[37] The choice between optimizing offensive or defensive potential often depends on the likelihood and results of being attacked by the opponent flight. If the flight is confident they can detect and neutralize an opponent flight attack and either defeat or escape the opponent then optimizing offensive potential is a reasonable choice. Otherwise, a defensive posture should be taken.

[38] Refer to chapter 10 of reference [173] for other tactics.

stern-conversion intercepts. As the name implies, the front-quarter intercept is one in which the intercepting flight approaches the target flight from the target flight's front quarter. The stern conversion converts an initial front quarter intercept into a final rear-hemisphere position for the flight. The bottom line on intercept tactics is that there are many tactics to choose from and choice is highly scenario dependent. [101: page 38] [17: page 22] [173: pages 297, 348, 350, 353, 383, 384] [19: page 110]

During the intercept portion of air-to-air engagements, most of the friendly flight's knowledge of opponent is provided by on-board sensors (mainly radar) and communications. Throughout the course of the intercept, the friendly flight attempts to comprehend the tactical situation in terms of the spatial (three dimensional layout) and temporal relationships between the flight, other friendly flights, and the opponent. The process of developing and updating a spatial layout of the opponent is called radar sorting. Sorting requires the flight to distinguish all potential targets in terms of azimuth, elevation, range, and velocity.[39] The goal of sorting is to decide which opponent aircraft to attack. After deciding which aircraft to attack, the friendly flight must become more specific in controlling the intercept geometry. Until this point general rules of thumb were sufficient (e.g., offset left or right, stay at high altitude[40]. Beyond Visual Range (BVR) weapons employment is the culmination of the intercept (if the rules of engagement allow). The friendly flight compares its tactics to the opponent flight's tactics and flies within the weapons envelope of the missile. [104: pages 1,6,7, 14, 18, 19, 22] [19: pages 110-112]

If all has gone according to plan and BVR weapons are launched the flight may proceed to within visual range (WVR) operations (IR missiles and guns). Conversely,

---

[39]Sorting answers such questions as: How many threat aircraft are out there? What formation are they in? What are they doing? The goal is to sort the opponent before entering opponent BVR missile range.

[40]An altitude offset can prevent or delay acquisition of the friendly flight by the opponent flight.

if radar sorting procedures[41] did not differentiate each individual aircraft or the ROE did not allow a BVR weapons employment (e.g., visual identification required or insufficient identification obtained), the flight may have to continue into closer range of the opponent flight. As a result, the flight may opt to forego BVR weapons employment and proceed directly to short range operations. Information provided by the electronic support measures (ESM), e.g. RWR, helps the flight decide whether to continue offensive actions or to execute defensive actions for survival. If defensive, the flight may be forced to abort the engagement. [104: page 20-22]

By 10 nm or less from the target, the flight has entered the WVR transition zone. At this zone, each pilot in the flight must decide whether his aircraft is in an offensive, defensive, or neutral position and whether he should attack, evade and reengage, or disengage. Factors that determine position include radar and ESM situational awareness; for example, a pilot might consider his aircraft to be in:

- An *offensive* position if he has high radar SA and a blank RWR scope;
- A *neutral* position if he has a radar lock and a single RWR spike; and
- A *defensive* position if low radar SA and many RWR spikes.

If deciding to attack, the pilot must execute short range operations which rely on air combat maneuvering tactics; infrared missiles and guns; and visual contact rather than sensors. On the other hand, an evade and reengage action may be attempted if the pilot elects to pursue a follow-on BVR attack. If deciding to disengage, the pilot should perform defensive maneuvering to separate from the engagement. [104: page 23]

When pressing to WVR (typically 3 nm or less), two basic tactical approaches are available: the 'angles' fight and the 'energy' fight. In the angles approach, the pilot first seeks to gain a position advantage (angles) even at the expense of relative

---

[41]Sorting is difficult within 15 nm of the threat because of rapid changes in threat azimuth and altitude.

energy and then maintain or improve this advantage in order to get within weapon firing parameters. In the energy approach, the pilot first seeks to gain an energy advantage without yielding a decisive position disadvantage. Once an energy advantage has been gained, the pilot converts this advantage into a lethal position disadvantage without giving up the entire energy advantage. Each of these approaches has advantages and disadvantages depending on the weapons involved. [173: page 99]

If friendly aircraft enters a turning engagement (termed 'dogfight') with an opponent aircraft, he allows the other aircraft in the opponent flight an opportunity to shoot. USAF tactics emphasize early shots, causing disruption, keeping airspeed up and avoiding getting drawn into a dogfight type engagement. Pilots must realize their 'escape window' or safe path out of an engagement. Factors that the drive the position of the escape window include:

- range from opponent flight (the greater the range, the more open the escape window);

- energy relative to opponent aircraft (the greater the energy, the more open the escape window); and

- angle-off and aspect with opponent flight (head-on pass typically gives the best chance for an open escape window).

A turning engagement among several aircraft can quickly degenerate into a 'furball' type of fight. A furball is not desired since each pilot is operating independently and giving mutual support only by presence. [17: page 23] [19: page vii, 89, 160] [173: page 267]

In a turning engagement, maneuver capability is paramount. In this close-in environment, basic fighter maneuvers (BFM) describe how aircraft maneuver against each other in one-versus-one combat. BFM describes specific concepts of air combat aircraft turns, turning room and turn circles and is generally grouped in three cate-

gories depending on the position of aircraft relative to the opponent in the air-to-air engagement:

- *Offensive BFM*: Offensive BFM implies a position and/or speed <u>advantage</u> over the opponent. The goal of offensive BFM is to shot down the opponent in the minimum amount of time. Hence, maneuvers center around placing the opponent in weapon envelopes while at the same time denying the opponent weapon launch opportunities. Anytime a weapon shot can be taken it should be done.

- *Defensive BFM*: Defense BFM implies a position and/or speed <u>disadvantage</u> over the opponent. The goal of defensive BFM is to stay alive and separate from the opponent if given the opportunity. Maneuvers center around creating weapon launch problems for the opponent thus buying time and surviving a little longer. By extending the engagement, the opponent may be forced into an error. If the opponent fires a missile, the best way to fight the missile is with aspect - by placing the oncoming missile on the beam (90° aspect). Missiles generally have the worst line-of-sight and rate problems to solve at this aspect. The practice of beaming in combination with electronic countermeasures such as chaff and flare disrupts and frequently defeats the tracking capability of the missile and causes the missile to miss.

- *Neutral BFM*: Neutral BFM implies a null position and/or speed over the opponent and vice versa. Two options exists to the pilot : he can separate or he can stay and attempt to get an offensive position on the opponent. If the pilot choices to engage he must be aware of his position in regard to the escape window.

Time considerations are important in turning engagements. Then though a pilot may be winning the engagement he started with one opponent, he may be losing an engagement with a second opponent he does not see. [19] [173] [44: page 65]

As a final note, most air-to-air kills are against aircraft that have no idea that they are about to be fired upon. The further away, an aircraft can fire a missile at an opponent and still have the missile be effective the better. Of particular concern to USAF planners are opponent air combat aircraft attacking high value assets such as AWACs and tankers using radar quiet intercept (RQI) techniques initially relying solely on passive detection systems such as radar warning receivers and infrared search and track (IRST) systems. [19: page 63] [185: pages 57, 58]

Appendix B. THUNDER's Air-to-Air Engagement Submodel

THUNDER's air-to-air engagement submodel consists of two separate engagement models described by the model documentation as "low resolution" and "high resolution" [29: Volume II, page 66]. In order to describe the "high resolution" air-to-air engagement submodel[1], some background information on THUNDER's aircraft entity modeling structure is appropriate. Three hierarchical aircraft entity classes exist in THUNDER: Flight Group, Flight, and Aircraft. Figure 51 illustrates these entity classes and their hierarchy.

The basic aircraft entity class modeled in THUNDER is the aircraft. An aircraft entity consists of a platform and weapon loads (e.g., a F-15C with four AIM-120s (AMRAAM) and four AIM-9s (Sidewinder)). The next aircraft entity class in the hierarchy is the flight. A flight entity consists of a group of *homogeneous* aircraft entities (e.g., a four ship of F-15Cs each carrying four AIM-120s and four AIM-9s). The last aircraft entity class in the hierarchy is the flight group. A flight group entity consists of a group of homogeneous and/or heterogeneous flight(s) (e.g., two flights of four F-15Cs or a flight of four F-15Cs escorting a flight of two F-16Cs to a ground target).

As noted in Figure 51 when an air-to-air engagement occurs between two aircraft flight groups one is termed the *intercepting* flight group and the other the *target* flight group. *An air-to-air engagement in THUNDER is triggered by an intercepting flight group detecting a target flight group.* Detection occurs in two manners:

1. The target flight group is detected while flying through an air zone patrolled by the intercepting flight group. The intercepting flight group is performing a BARCAP mission. (The BARCAP mission is described in section A.2.) or

---

[1]The description of the THUNDER air-to-air engagement submodel will be limited to the "high resolution" submodel only. The "low resolution" model is trivial in comparison to the "high resolution" model. For further information concerning the "low resolution" model, the reader may refer to [29].
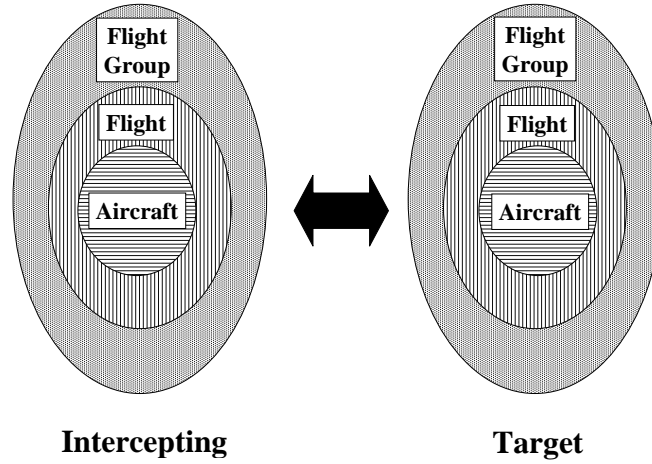
Figure 51. Thunder Aircraft Entity Classes

2. The target flight group is detected by the intercepting flight group after the intercepting flight group has flown to an intercept point to engage the target flight group. The intercepting flight group is performing a DCA, ODCA, or HVAA mission. (DCA, ODCA, or HVAA missions are described in section A.2).

Before explaining the detection process in THUNDER, it might be helpful to understand how flight paths are modeled. Thunder uses a horizontal (x, y), square grid system to model flight paths. Each grid intersection defines a possible point a flight path may pass through. Mission planning modules in THUNDER produce a flight path for every flight group. Flight paths are either direct or non-direct in nature (depending on a departure airfield setting). If a flight group is given a direct flight path, the flight group proceeds on a direct route to the target. In non-direct routing, the flight group proceeds to a point perpendicular to the target and a setback distance on the friendly side of the forward line of troops (FLOT) and then proceeds directly to that target. Non-direct flight paths are intended to minimize the time and distance spent over opponent territory. A few discrete altitude (z) settings may be used for each flight path point. [29: Volume II: page 35, 52, 57, 173]

280

The air zones patrolled by an intercepting flight group[2] performing a BARCAP mission consist of a rectangular patrol area whose size is determined by the *type*, *number*, and *flight size* of aircraft flying the BARCAP mission. Specifically for each aircraft type, input variables width and depth define a rectangular area that can be patrolled by that aircraft type and the input variable flight size defines the standard patrol unit size for that aircraft type. The width of the patrol area is calculated as the product of the width variable and greatest integer value of the ratio of the number of aircraft in the intercepting flight group to the flight size[3]. The depth of the patrol area corresponds to the depth variable. Figure 52[4] shows a patrol area for a four ship of MiG-29s consisting of two flights (i.e, flight size variable for the MiG-29's is two). Each MiG-29 can cover an area of width 65 nm (120 km) and depth 108 nm (200 km). Hence, the patrol area of the four-ship of MiG-29s has a width of 130 nm and a depth of 108 nm [29: Volume II, pages 55, 56].

If the flight path of the target flight group enters the patrol area of the intercepting flight group, the target flight group is subject to detection. The probability of detection at search time t, $P(t)$, is based on a random search algorithm:

$$P(t) = 1 - exp^{\frac{-SWt}{A}} \tag{102}$$

where

- S - the speed of the target flight group's path through the patrol area;

- W - the diameter of the search sweep;

- t - the time the intercepting flight group has spent searching for the target flight group; and

---

[2] In the case of the BARCAP mission, the intercepting flight group is composed of only one aircraft type.

[3] The flight size variable allows for element formations. For example, consider a four-ship of aircraft split as two ship elements. In this scenario, the effective search width would be doubled with each element acting as a separate search unit.

[4] Figure 52 is taken from [29:Volume II, page 55].
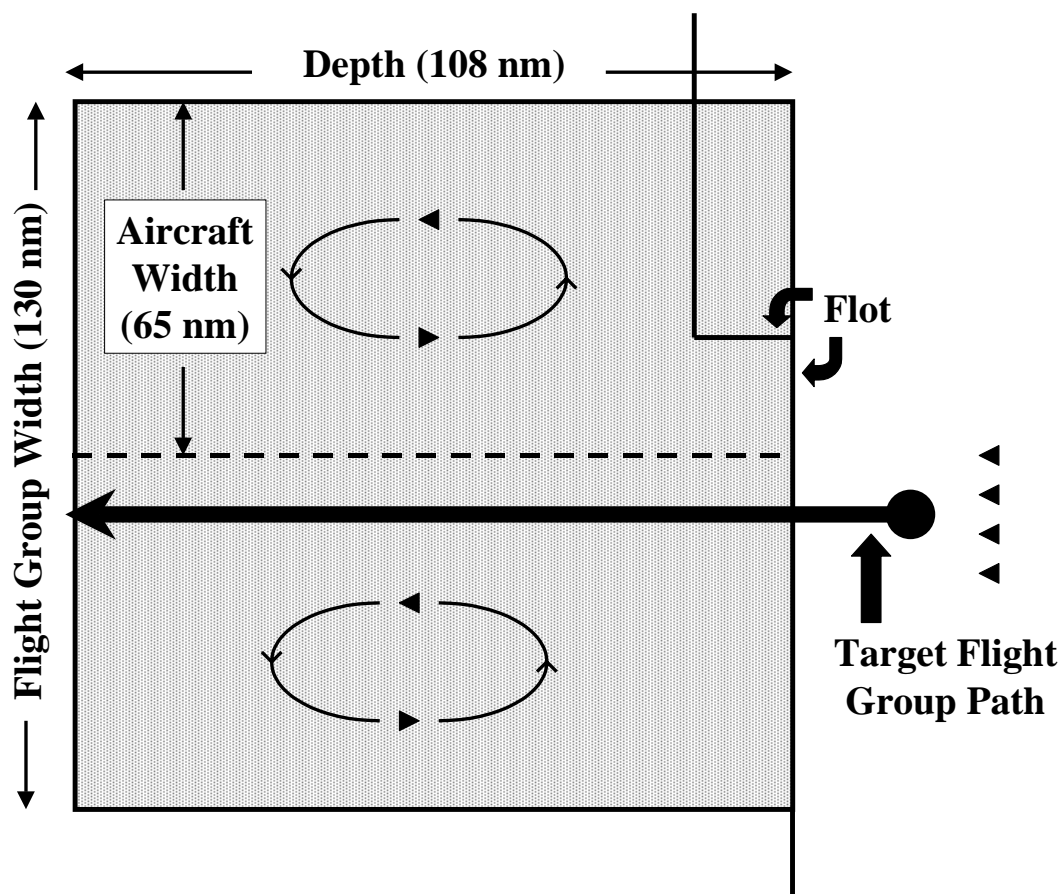
Figure 52. THUNDER BARCAP Patrol Area

- A - the search area the intercepting flight group must cover. [29: page 70, 71]

This search algorithm places a search path of length $L = St$ into the search area, A, in a random fashion. This means that the location and course of the search path at any given time is independent of the location and course at other times not too close to the first. The random search model in equation 102 yields a lower bound on search effectiveness [110: pages 5-10, 5-11].

The variables W^5 and A in equation 102 are dependent on a numerical expression of the *degree of command and control* among aircraft in the intercepting flight group. The concept of command and control will be addressed in greater detail later in this section. For illustration purposes, consider the intercepting flight group's command and control value, $FGCC$ to be a given value between 0 and 1. The search strategies are:

1. If the intercepting flight group was in a "perfect" command and control state ($FGCC = 1$), aircraft in the intercepting flight group would conduct individual, *flight size* searches in *specific strips* of the search area; or

2. If the intercepting flight group was in a "no" command and control state ($FGCC = 0$), aircraft in the intercepting flight group would *combine* the diameter of individual aircraft search sweeps over the *entire* patrol area.

Figure 53 shows the search pattern for the "perfect" command and control state for the four ship of MiG-29s. In this state, each flight element[6] searchs a different subsection of the patrol area defined by the aircraft patrol area parameters width

---

[5]W is defined as the diameter of the search sweep (or circle) which is $2\sqrt{\frac{A_{search}}{\pi}}$ where A_{search} is the search sweep area. For a single aircraft, A_{search} is $sweep * \frac{\pi}{360°} * detect^2$ where $sweep$ is the sweep angle of the radar (in degrees) and $detect$ is the distance at which the aircraft's radar detects the target.

[6]Defined by the flight size variable.

and depth. For this state, denoted 1,

$$W1 = 2 * detect * \sqrt{fsize * \frac{sweep}{360°}} \qquad A1 = \frac{width * depth}{\frac{n}{fsize}}$$

where

- $detect$ - the distance at which an intercepting aircraft's radar detects the target flight group;

- $fsize$ - intercepting aircraft flight (element) size;

- n - the number of intercepting aircraft in the patrol area

- $sweep$ - intercepting aircraft's radar sweep angle;

- $width$ - intercepting aircraft patrol area width; and

- $depth$ - intercepting aircraft patrol area depth.

Figure 54 shows the search pattern for "no" command and control state for the four ship of MiG-29s. In this state each aircraft combines aircraft search sweeps to search the entire aircraft patrol area. For this state, denoted 2,

$$W2 = 2 * detect * \sqrt{n * \frac{sweep}{360°}} \qquad A2 = width * depth$$

[29: page 71].

Letting P1(t) and P2(t) represent the probability of detection at search time t under search state 1 and 2 respectively and weighting the probabilities in search state 1 and 2 by $FGCC$ and $1 - FGCC$ respectively yields:

$$P(t) = FGCC * (1 - exp^{\frac{-S*W1*t}{A1}}) + (1 - FGCC) * (1 - exp^{\frac{-S*W2*t}{A2}}) \qquad (103)$$

Equation 103 is evaluated at

$$t = \min[t_{target}, t_{search}]$$

284
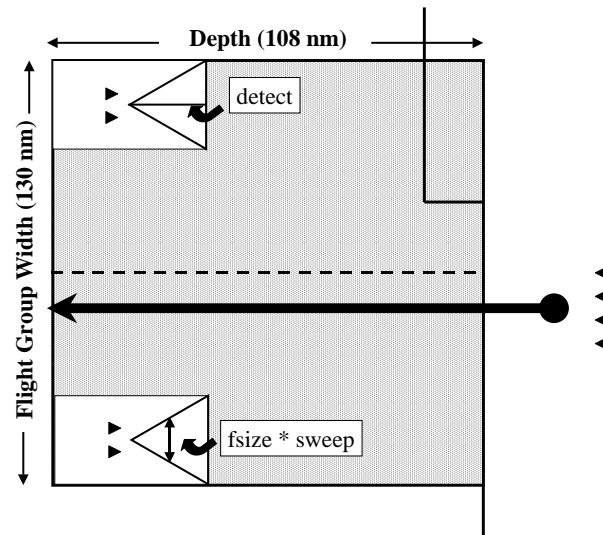
Figure 53. THUNDER BARCAP Search Pattern for Perfect Command and Control State
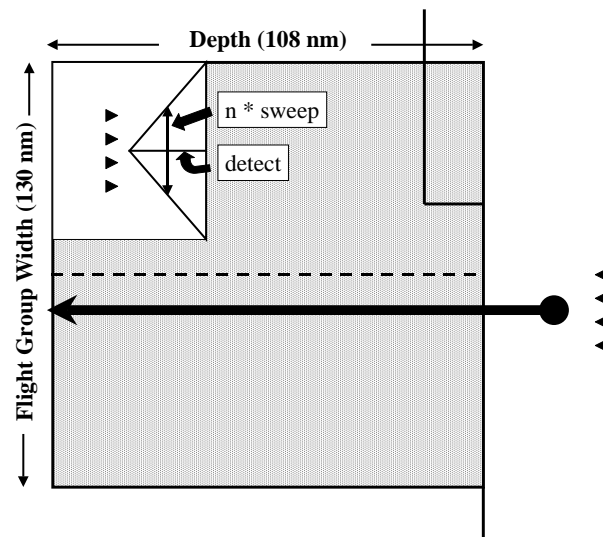
Figure 54. THUNDER BARCAP Search Pattern for No Command and Control State

285

where

- t_{target} - the time the target flight group will be in the patrol area; and

- t_{search} - the time since the intercepting flight group's last air-to-air engagement (i.e., search time). [29: page 71, 72]

Once the value of $P(t)$ is determined, a random draw is made from a $U(0,1)$ to determined if the intercepting flight group detects the target flight group[7]. If the intercepting flight group detects the target flight group an air-to-air engagement takes place else the target flight group passes unhindered through the patrol area.

In order to illustrate equation 103, consider the four-ship of MiG-29s (packaged as two elements) mentioned above covering a patrol area of width of 130 nm and a depth of 108 nm. Assume that a four ship of F-15Cs enters the patrol area traveling at 480 knots parallel to the depth of the patrol area (i.e., $t = \frac{108}{480} = .225$ hrs in the patrol area). Further assume that the MiG-29s have a radar detection range of 30 nm[8] against the F-15Cs; a radar sweep angle of 120o, and a $FGCC$ of 0.63. For this situation,

$$
\begin{aligned}
W1 &= 2*30nm*\sqrt{2*\frac{120°}{360°}} = 48.99nm \\
A1 &= \frac{130nm*108nm}{\frac{4}{2}} = 7020nm^2 \\
P1 &= 1 - exp^{\frac{-S*W1*t}{A1}} = 1 - exp^{-.753} = .529
\end{aligned}
$$

and

$$
\begin{aligned}
W2 &= 2*30nm*\sqrt{4*\frac{120°}{360°}} = 69.28nm \\
A2 &= 130nm*108nm = 14040nm^2 \\
P2 &= 1 - exp^{\frac{-S*W1*t}{A1}} = 1 - exp^{-.533} = .505
\end{aligned}
$$

[7]A detection takes place if the value of $P(t)$ is greater than or equal to the random draw.
[8]The concept of radar detection range will be detailed shortly.

Thus, the detection probability for this example is $P(.225)$ which is .52 $(.63(.529) + .37(.505))$.

Intercepting flight groups performing a DCA, ODCA, or HVAA missions are *first alerted to target flight group by a GCI detection event.* A GCI detection event occurs when target flight group is detected by a ground-based or airborne early warning (AEW) platform. Ground-based detection occurs when the target flight group crosses a ground-based detection line. The location of this line is dependent on several factors including the altitude (high or low) of the target flight group. Specifically, the detection distance, $dist_{GB}$, from the ground-based radar location is computed as

$$dist_{GB} = \min[dist_{max}, (dist_i * RCS^{.25})] \tag{104}$$

where

- $dist_{max}$ - the ground-based radar's maximum detection range;
- $dist_i$ - the ground-based radar's one square meter target detection range depending on the target flight group's altitude, $i = low, high$ and the given radar band; and
- RCS - the target flight group's radar cross section in the ground-based radar band.

Once within the ground-based radar's detection distance, THUNDER considers the target flight group to be detected. An intercepting flight group is then sent to intercept the target flight group at the earliest possible point. [29: Volume II: page 67]

Similarly, airborne early warning (AEW) detection occurs when the target flight group enters the radar detection circle around the AEW platform's orbit point. The AEW detection distance, $dist_{AEW}$, is computed as

$$dist_{AEW} = \min[dist_{max}, (dist_{target} * RCS^{.25}), los] \tag{105}$$

where

- $dist_{max}$ - the AEW platform radar's maximum detection range ;

- $dist_{target}$ - the AEW platform radar's one square meter target dectection range for the given radar band;

- RCS - the target flight group's radar cross section in AEW platform's radar band; and

- los - line of sight between the AEW platform and the target flight group.

los is computed as

$$los = r * [\cos^{-1}(\frac{r}{r + alt_{target}}) + \cos^{-1}(\frac{r}{r + alt_{AEW}})] \qquad (106)$$

where

- r - radius of the earth (6371 km);

- alt_{target} - altitude of the target flight group; and

- alt_{AEW} - altitude of the AEW platform.

Equation 106 assumes a round, 'cue-ball'[9] earth. *Once within the AEW radar's detection distance, THUNDER considers the target flight group to be detected.* An intercepting flight group is then sent to intercept the target flight group at the earliest possible point. [29: Volume II: page 68]

For DCA, ODCA, and HVA missions, once the intercepting flight group arrives at its target point (i.e., the point where the target flight group is reported to be), THUNDER computes the radar detection range, *detect*, at which the intercepting flight group radar can detect the target flight group based upon the intercepting

---

[9]Equation 106 uses the angle from the center of the earth to the AEW platform and the target flight group in place of the tangent of the angle. THUNDER documentation claims this approximation holds since aircraft altitudes are small with respect to the radius of the earth.

flight group's radar and target flight group's radar cross section (RCS) and use of jammers. This radar detection range is the same range used to calculate W in equation 103. Detection occurs, if closest distance between the target flight group's flight path and the target point, $dist_{closest}$ is less than the detection range, $detect$; i.e., if

$$dist_{closest} < detect \qquad (107)$$

[29: Volume II: page 69]

The detection range, $detect$, is calculated using a horizontal (x, y) detection range $hdetect$ and the altitude (z) separation distance, alt, between the intercepting and target flight groups. Based on the Pythagorean relation,

$$detect = \begin{cases} \sqrt{hdetect^2 + alt^2} & \text{if } alt < detect \\ 0 & \text{if } alt \geq detect \end{cases} \qquad (108)$$

The THUNDER horizontal (x, y) detection range algorithm uses the fact that the radar cross section, RCS, of a flight group in a given radar band is proportional to the horizontal radar detection range, $hdetect$, raised to the fourth power. [192: page 85] This is expressed mathematically as:

$$RCS \ \alpha \ hdetect^4 \qquad (109)$$

Each flight group in THUNDER has a radar cross section (RCS). RCS is an area measure of the flight group's ability to reflect search radar energy. The RCS of the flight group is radar band or frequency specific (i.e., the RCS of the flight group varies depending of the band (frequency) of the search radar). For each applicable radar band, the RCS of a flight group is determined by finding the maximum radar cross section of the flights in the flight group. For example, suppose a flight group is

composed of three flights with the following radar cross sections (for a given radar band):

- Flight A: RCS = 3.75 square meters;

- Flight B: RCS = 4.20 square meters;

- Flight C: RCS = 25.00 square meters; and

The RCS of this flight group is 25.00 square meters. The RCS of a flight is calculated as:

$$\min[fsize, n] * RCS$$

where

- $fsize$ - aircraft flight (element) size;

- n - is the number of aircraft in the flight; and

- RCS - aircraft RCS.

THUNDER assumes that the maximum number of aircraft that would be in the resolution cell of the search radar is the flight size specified for particular aircraft - this assumes aircraft in a flight are flying sufficiently close to each other that they do not appear as separate targets. The RCS of an aircraft (for a given radar band) is a function of the aircraft, fuel, and weapons configuration and is calculated as the sum of:

1. basic aircraft RCS[10];

2. aircraft configuration RCS delta (e.g., F-15C vs F-15E);

3. fuel configuration RCS delta (e.g., drop tanks); and

4. weapons configuration RCS delta (e.g., missiles).

---

[10]THUNDER has historically based the basic aircraft RCS on a 20° off-nose aspect at the given radar band (frequency) of the search radar.

[29: Volume II: page 67, 93, 100]

Using the relationship in equation 109, the non-jammed (i.e., assuming no radar jamming), horizontal detection range, $hdetect_{nojam}$, for a flight group of radar cross section, RCS at a given radar band is

$$hdetect_{nojam} = hdetect_{1sq-nojam} * RCS^{.25} \qquad (110)$$

where $hdetect_{1sq-nojam}$ is the non-jammed, horizontal detection distance for a RCS of unit area. [29: Volume II: page 100]

Jamming effects in THUNDER consists of escort (main lobe) and standoff (side lobe)[11]. These jamming effects are assumed to be used only when the search radar is detected. This assumption allows the jammed, horizontal radar detection range, $hdetect_{jam}$ to be related to unjammed, horizontal detection range, $hdetect$. The jammed, horizontal radar detection range is also referred to as the burnthough range; the range at which a jammed search radar can burnthough the jamming. Using the relationship in equation 109, the jammed, horizontal detection range, $hdetect_{jam}$, for a flight group of radar cross section, RCS at a given radar band is

$$hdetect_{jam} = hdetect_{1sq-jam} * RCS^{.25} \qquad (111)$$

where $hdetect_{1sq-jam}$ is the jammed, horizontal detection range for a RCS of unit area. The $hdetect_{1sq-jam}$ value for escort jamming is determined using a lookup table of ranges by jammer type and radar type. The $hdetect_{1sq-jam}$ value for standoff jamming is set in an input curve describing the range as a function of standoff jammer range, jammer type, and radar type. When multiple jammers are present the

---

[11]As the name implies, escort jamming is an electronic countermeasure (ECM) tactic in which the jamming platform accompanies the strike package; jamming radars to protect the strike package. In contrast, in standoff jamming, the jamming platform remains close, but outside the range of opponent air defense systems; jamming these systems to protect the strike package. [178: page 9.6]

jammed, horizontal detection range is taken to be the minimum jammed, horizontal detection range; *jamming effects are not combined.* [29: Volume II: page 100]

The THUNDER "high resolution" air-to-air engagement submodel, hence referred to as the THUNDER air-to-air engagement model, adjudicates combat via three processes:

1. A process that calculates a single shot probability of kill (SSPK) for each aircraft against possible opponent aircraft.

2. A process that aggregates aircraft versus aircraft SSPK calculations to attrition rates for flight versus flight and flight group versus flight group levels.

3. A process that assesses engagement losses using draws from binomial distributions.

In order to illustrate the processes as they are described, let us consider the case where a target flight group composed of a single flight flies through an air zone patrolled by an intercepting flight group composed of a single flight. Specifically, consider an intercepting flight group composed of a flight of four MiG-29s each carrying two AA-10s, two AA-8s, and two GSH-23 and a target flight group composed of a flight of four F-15Cs each carrying four AIM-120s and four AIM-9s flying a FSWP mission (reference section A.2). The intercepting flight group (MiG-29s) will be referred to as side A and the target flight group (F-15Cs) will be referred to side B.

A single shot probability of kill (SSPK) is calculated as follows for each aircraft versus opponent aircraft combination.

$$SSPK = ENG * LCH * PK \tag{112}$$

where

- $SSPK$ - single shot probability of kill of an aircraft against an opponent aircraft;

- ENG - probability of engaging the opponent aircraft;

- LCH - probability of firing a weapon at the opponent aircraft; and

- PK - probability of killing an opponent aircraft given a weapon firing.

Since a flight is composed of homogeneous aircraft in THUNDER, the SSPK calculation can also be viewed as a flight versus flight entity level process. The discussion of the SSPK calculation highlights which entity level (aircraft or flight) the engagement process is being viewed from.

An engagement will always occur in THUNDER if the intercepting flight group detects the target flight group. The probability of an aircraft (or flight) engaging an opponent aircraft (or flight) addresses those air environment variables that affect the ability to engage after initial detection. The probability of engaging the opponent (ENG) at the aircraft (or flight) level is modeled in THUNDER as the product of two factors:

- A general engagement probability, ENG_{AEW}, based on whether or not either flight group is *under airborne or ground-based control* (termed in THUNDER Airborne Early Warning (AEW) control) and

- A modification component, TAC, reflecting the flight group tactics flown by each side.

A flight group is consider to be in AEW control if the controller is in both communication range of the friendly flight group and radar range of the opponent flight group. Four separate AEW control states are possible:

1. Neither side is in AEW control;

2. The own flight group is in AEW control only;

3. The opponent flight group is in AEW control only; and

	$ENG(1)$	$ENG(2)$	$ENG(3)$	$ENG(4)$
Side A (MiG-29s)	.60	.75	.45	.60
Side B (F-15Cs)	.50	.63	.38	.50

Table 22.   THUNDER Example 1: Probability of Engagement Based on AEW Control State

4. Both the own and opponent flight groups are AEW control.

For the current case, let $ENG^i(AEW)$, $i = A, B$ refer to the probability of engagement conditioned on AEW control state, $AEW = 1, \ldots, 4$. Table 22 shows the probabilities for our example. The probability of engagement conditioned on AEW control state is *opponent platform dependent*. For example, in our case, the F-15Cs would have a different set of engagement probability values if they were to engage MiG-23s rather than the MiG-29s. For illustration purposes let us assume that both flight groups are in AEW control, thus from Table 22:

- $ENG^A(AEW) = ENG(4) = .60$ and

- $ENG^B(AEW) = ENG(4) = .50$

As mentioned above, the next contributing factor to a side's ENG is the side's flight group tactics. Flight group tactics are implemented at the flight level. For the intercepting flight group, the flight group can choose to employ one of these three flight group tactics:

1. The intercepting flight group engages the target flight group's escort aircraft only; or

2. The intercepting flight group engages the target flight group's other (non-escort) aircraft only; or

3. The intercepting flight group engages the target flight group's escort aircraft with a percentage $1 - a$ and other (non-escort) aircraft with a percentage a.

Escort aircraft and hence escort flights are those entities that have been given an Air-to-Air Escort (AIRESC) mission (reference section A.2). In THUNDER, a is referred to as the allocation ratio. The allocation ratio is calculated using Lanchester difference equations and an allocation measure of effectiveness. Section 3.2 details this calculation. Since the target flight group (F-15Cs) in our example does not contain a flight with an AIRESC mission (recall the F-15Cs are flying an FSWP mission), the intercepting flight group allocation ratio is $a = 1$.

For the target flight group, the flight group can choose to employ one of these three flight group tactics:

1. All escorts split from the target flight group to engage the intercepting flight group; or

2. All escorts stay with the target flight group; or

3. A percentage m of the escorts stay with the target flight group and a percentage $1-m$ of the escorts split from the target flight group to engage the intercepting flight group.

In THUNDER, m is referred to as the mutual support defender ratio. The mutual support defender ratio is a user inputted variable. For our case, let $m = .50$.

Figure 55 shows the interaction of the intercepting and target flight group tactics. Four modification components, TAC, exist depending on the type of target flight group involved (escort or other):

1. If a flight from the intercepting flight group is engaging an escort flight from the target flight group, $TAC = (1-a)$. The term $1-a$ represents the allocation of aircraft in the intercepting flight to engaging target flight group escort aircraft.

2. If a flight from the intercepting flight group is engaging an other (non-escort) flight from the target flight group, $TAC = a$. The term a represents the allocation of aircraft in the intercepting flight to engaging target flight group other aircraft.

295

	Escort Target Flight	Other (Non-Escort) Target Flight
Intercepting Flight	$ENG_{AEW} * (1 - a)$	$ENG_{AEW} * a$

Table 23. Probability of Engaging A Target Flight, ENG, For An Intercepting Flight in THUNDER

3. If an escort flight from the target flight group is engaging a flight from the intercepting flight group, $TAC = (1 - a) * (1 - m) + m$. The term $(1 - a) * (1 - m)$ represents the portion of the escort flight that engages a portion of an intercepting flight engaging escort aircraft. The term m represents the portion of the escort flight protecting the other (non-escort) aircraft in the target flight group.

4. If an other (non-escort) flight from the target flight group is engaging a flight from the intercepting flight group, $TAC = a * (1 - m) + m$. The term $a * (1 - m)$ represent the portion of the other (non-escort) flight that engages a portion of an intercepting flight engaging the other (non-escort) aircraft. The term m represents the portion of the escort flight protecting the other (non-escort) aircraft in the target flight group.

Table 23 and Table 24 show the probabilities of engaging the opponent (ENG) at the aircraft (or flight) level when for the intercepting and target flights respectively. For our example, with $a = 1$ and $m = .5$:

- $ENG^A = ENG^A(AEW) * a = .60 * 1 = .60$ and
- $ENG^B = ENG^B(AEW) * [a * (1 - m) + m] = .50 * [1 * (1 - .60) + .60] = .50$

In this example, flight group tactics do not affect the engagement probabilities since there are no escort aircraft involved.

The next variable in equation 112, the probability of firing a weapon at the opponent aircraft - LCH, is a combination of several factors:

296

INTERCEPTORS

a

1-a

m   **ESCORTS**   **1-m**
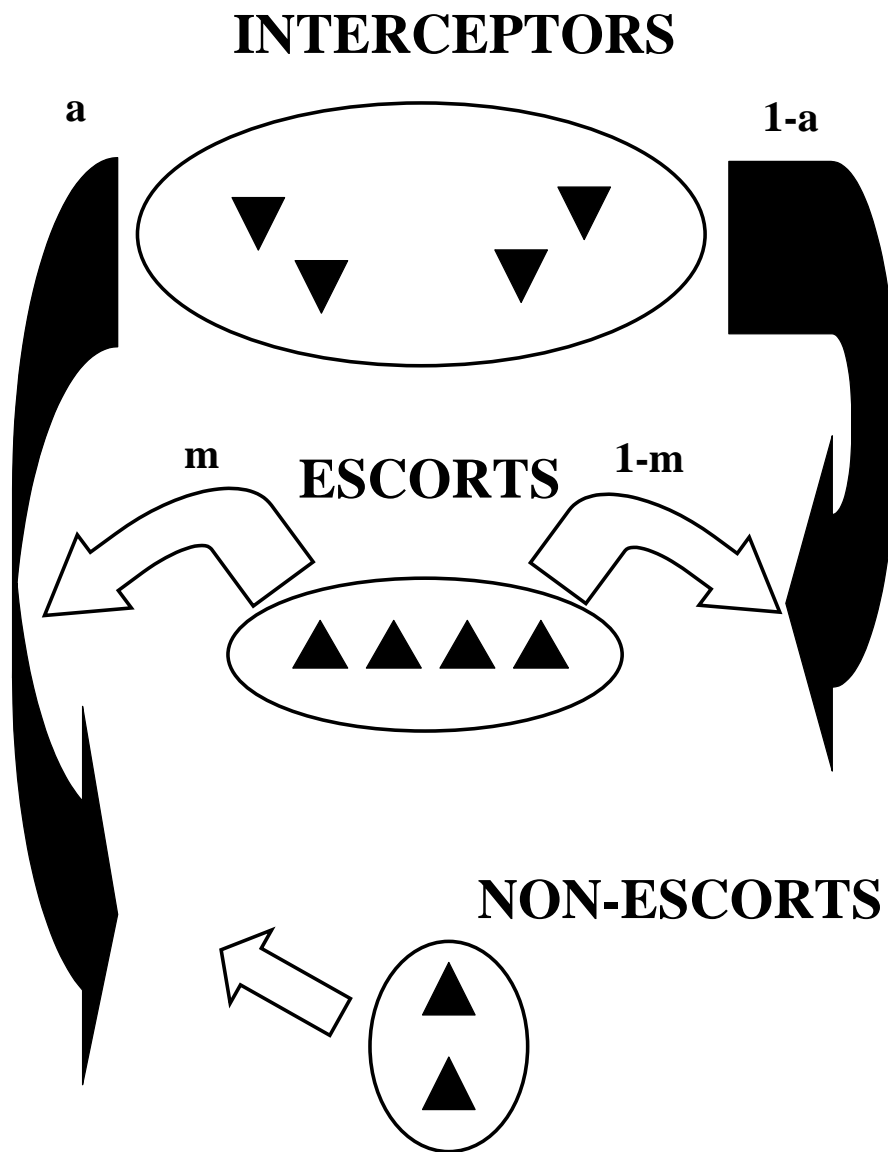
NON-ESCORTS

Figure 55. Interaction of Flight Group Tactics in THUNDER

	Intercepting Flight
Escort Target Flight	$ENG_{AEW} * [(1-a) * (1-m) + m]$
Other (Non-Escort Flight)	$ENG_{AEW} * [a * (1-m) + m]$

Table 24. Probability of Engaging An Intercepting Flight, ENG, For A Target Flight in THUNDER

- Relative Range Advantage - variable indicating the number of weapons (of a specific weapon type) an aircraft can fire before an opposing aircraft can fire a single weapon (of a specific weapon type);

- Opponent's Probability of Engagement; i.e., Opponent's ENG;

- Number of Friendly and Opponent's Aircraft in Each Flight;

- Opponent's Probability of Kill (i.e., Opponent's PK); and

- Opponent's Force Multiplier.

For each aircraft in an air-to-air engagement, THUNDER determines the impact of weapon firings using the concept of a composite or aggregate weapon. Recall each aircraft in THUNDER consists of a platform and weapon loads (e.g,. a F-15C with four AIM-120s (AMRAAM) and four AIM-9s (Sidewinder)). For each air-to-air engagement, THUNDER specifies a user-inputted maximum number of launches for each aircraft. For our example, the maximum number of weapon firings per engagement for the F-15Cs and MiG-29s is set at 2. Additionally, for each aircraft, THUNDER specifies a user-inputted prioritized list of weapons firings. For our example, the prioritized list of weapons firings for the MiG-29s and F-15Cs are:

Firing Sequence	MiG-29	F-15C
1st	AA-10	AIM-120
2nd	AA-8	AIM-9
3rd	AA-10	AIM-120
4th	AA-8	AIM-9
5th	GS-23	AIM-120
6th	GS-23	AIM-9
7th	No Ammo	AIM-120
8th	No Ammo	AIM-9

If the F-15C were to use its maximum number of weapon firings (2) during an engagement the F-15 would fire an AIM-120 and an AIM-9. THUNDER's air-to-air engagement model assumes that *an aircraft will always take its maximum number of weapon firings in an engagement unless the aircraft has an insufficient number of weapons remaining.* For our example, the F-15C and MiG-29 aircraft entities will always fire 2 weapons per engagement unless they have only 1 remaining weapon to fire[12]. An *aggregate weapon* is created from the weapons to be fired in the air-to-air engagement. For our example, the aggregate weapon for the F-15C will be a combination of an AIM-120 and an AIM-9 and for the MiG-29 a combination of an AA-10 and AA-8. The specifics of these combinations will be addressed shortly.

In order to fire an aggregate weapon at an opponent aircraft, one of these three scenarios must exist for the firing aircraft. These scenarios are:

1. The firing aircraft has an overall relative range advantage and can fire its aggregate weapon at the opponent aircraft without threat of being killed by the opponent aircraft's aggregate weapon; or

---

[12]This situation is not possible in our case, since the total number of weapons for each aircraft divided by the maximum number of weapon firings in an engagement are integer-valued (4 for the F-15C and 3 for the MiG-29s).

	AA-10	AA-8	GSH-23
AIM-120	1	1	1
AIM-9	0	0	0

	AIM-120	AIM-9
AA-10	0	1
AA-8	0	0
GSH-23	0	0

Table 25.   Weapon versus Weapon Relative Range Advantage for F-15C and MiG-29 Example

2. The firing aircraft does not have an overall relative range advantage, but the opponent aircraft can not engage the firing aircraft; or

3. The firing aircraft does not have an overall relative range advantage, but the firing aircraft survives an specified number of firings of an aggregate weapon by the opponent aircraft.

Each of these mutually exclusive scenarios, has a probability of occurrence. These probabilities are then added together to determine LCH (the probability of firing a weapon at an opponent aircraft entity). Let these probabilities of occurrence be designated by LCH_i, $i = 1, 2, 3$ respectively. Thus, $LCH = LCH_1 + LCH_2 + LCH_3$.

In order to compute LCH_1, the probability of occurrence of scenario one, the concept of relative range advantage needs to be addressed. THUNDER defines relative range advantage as the number of weapons (of a specific weapon type) an aircraft can fire before an opposing aircraft can fire a single weapon (of a specific weapon type). Table 25 shows the weapon versus weapon relative range advantages for our example. As can be seen from Table 25, an F-15C armed with AIM- 120s has a single relative range advantage against a MiG-29 armed with either AA-10s, AA-8s, or GSH- 23s. In other words, the F-15C can fire an AIM-120 before a MiG-29 armed with either AA-10s, AA-8s, or GSH-23s fire a weapon. Additionally, a MiG-29 armed with AA-10s has a single relative range advantage against an F-15C armed with only AIM-9s.

Scenario one equates to the situation were the firing aircraft can fire a weapon at the opponent aircraft before the opponent aircraft fires a weapon at it. In our

example, there are four possible combinations of first weapons to be fired between a F-15C and a MiG-29. These combinations are:

1. AIM-120 versus AA-10

2. AIM-120 versus AA-8

3. AIM-9 versus AA-10

4. AIM-9 versus AA-8

In weapon combinations 1 and 2, the F-15C has a single relative range advantage over the MiG-29. In weapon combination 3, the MiG-29 has a single relative range advantage over the F-15C. In combination 4, neither aircraft has a relative range advantage. Assuming that weapon combinations are random, the F-15C will have a single relative range advantage over the MiG-29 50% of the time; the MiG-29 will have a single relative range advantage over the F-15C 25% of the time; and neither aircraft will have a relative range advantage over the other 25% of the time. Hence, $LCH_1^A = .50$ and $LCH_1^B = .25$.

LCH_1^A and LCH_1^B may also be derived from the aggregate weapon concept. In our example, assuming the maximum number of weapons firings (2 each), the F-15C will fire an AIM-120 and AIM-9 and the MiG-29 will fire an AA-10 and AA-8 respectively each engagement. The aggregate weapon for the F-15C will be an equal combination of an AIM-120 and an AIM-9 and for the MiG-29 an equal combination of an AA-10 and AA-8. In other words, the aggregate weapon for the F-15C will be have 50% of the characteristics of the AIM-120 and 50% of the characteristics of the AIM-9. THUNDER terms these characteristics fractional launches. Let FL_j^i denote the fractional launches of weapon type j for side $i, i = A, B$. For the MiG-29s $(i = A)$, let $j = 1$ represent the AA-10 and $j = 2$ represent the AA-8 respectively. For the F-15Cs $(i = B)$, let $j = 1$ represent the AIM-120 and $j = 2$ represent the AIM-9 respectively. Hence for our example,

$$FL_1^A = .50$$
$$FL_2^A = .50$$
$$FL_1^B = .50$$
$$FL_2^B = .50$$

Now

$$LCH_1^A = FL_1^B * FL_2^A = .50 * .50 = .25$$
$$LCH_1^B = [FL_1^A * FL_1^B] + [FL_1^A * FL_1^B] = [.50 * .50] + [.50 + .50] = .50$$

LCH_2, the probability of occurrence of scenario two, depends on the occurrence of two independent events:

1. The firing aircraft does not have a relative range advantage on the opponent aircraft. The probability of this event is $1 - LCH_1$.

2. The opponent aircraft (opp) does not engage the firing aircraft. The probability of this event is $1 - ENG_{opp}$.

Hence $LCH_2 = [1 - LCH_1] * [1 - ENG_{opp}]$. For our example,

$$LCH_2^A = [1 - LCH_1^A] * [1 - ENG^B] = [1 - .25] * [1 - .50] = .375$$
$$LCH_2^B = [1 - LCH_1^B] * [1 - ENG^A] = [1 - .50] * [1 - .60] = .2$$

LCH_3, the probability of occurrence of scenario three, depends on the occurrence of three independent events:

1. The firing aircraft does not have a relative range advantage on the opponent aircraft. The probability of this event is $1 - LCH_1$.

2. The opponent aircraft (opp) engages the firing aircraft. The probability of this event is ENG_{opp}.

3. The firing aircraft survives n weapon firings from the opponent aircraft (opp) before his first weapon firing. The probability of this event is $(1 - PK_{opp})^n$ where PK_{opp} is the probability of kill given the firing of an opponent aircraft entity's aggregate weapon.

The probability of kill given the firing of an aggregate weapon is calculated in manner similar to the calculation of LCH_1. *PK values in THUNDER are weapon versus platform dependent.* Let PK_j^i denote the probability of kill of weapon type j for side $i, i = A, B$. Again, for our example, the MiG-29s $(i = A)$, let $j = 1$ represent the AA-10 and $j = 2$ represent the AA-8 respectively. Fpr the F-15Cs $(i = B)$, let $j = 1$ represent the AIM-120 and $j = 2$ represent the AIM-9 respectively. Hence for our example,

$$PK_1^A = .32$$
$$PK_2^A = .16$$
$$PK_1^B = .34$$
$$PK_2^B = .32$$

Now

$$PK^A = [FL_1^A * PK_1^A] + [FL_2^A * FL_2^A] = [.50 * .32] + [.50 * .16] = .24$$
$$PK^B = [FL_1^B * PK_1^B] + [FL_2^B * FL_2^B] = [.50 * .34] + [.50 * .32] = .33$$

The number of weapons firings from the opponent aircraft, n, is determined by considering the product of two factors:

1. A force multiplier variable, FM, that allows particular weapons to act as force multipliers, reducing the attrition of the aircraft entity carrying them.

2. The ratio of number of aircraft entities in both engaging flight entities.

The force multiplier variable, FM, is calculated via the following equation:

$$FM^{opp} = \sum_i \sum_j \min([(RRA_{ij}^{opp}, L_{opp} * FL_j^{opp}] * FL_j^{opp} * FL_i^{fac})$$

303

where

- i represents possible weapon types of the firing aircraft (fac);

- j represents possible weapon types of the opponent aircraft (opp);

- RRA is the relative range advantage of the opponent aircraft and weapon type j against the firing aircraft and weapon type i; and

- L represent the number of weapon firings the opponent aircraft may take.

The first term in the above equation, $\min([(RRA_{ij}^{opp}, L_{opp} * FL_j^{opp}]$, represents the advantage the opponent aircraft has in firing weapon type j against the firing aircraft firing weapon type i. This advantage takes into account the number of weapons of type j that can be fired by the opponent aircraft in an air-to-air engagement. For example, the opponent aircraft with weapon type j may have a two shot advantage ($RRA^{opp} = 2$) over the firing aircraft with weapon type i, but it may have only one weapon of type j to fire ($L_{opp} * FL_j^{opp} = 1$). Hence, the opponent aircraft with weapon type j only has a one shot advantage ($\min[2, 1]$) over the firing aircraft with weapon type i. The second term, $FL_j^{opp} * FL_i^{fac}$, represents the fraction of first weapon firings by the opponent aircraft firing weapon type j and the firing aircraft firing weapon type i. For our example, letting the AIM-120 and AA-10 have a subscript designation of 1 and the AIM-9 and AA-8 have a subscript designation of 2, the force multiplier variables are:

$$
\begin{aligned}
FM^A &= \sum_{i=1}^{2}\sum_{i=1}^{2} \min([(RRA_{ij}^A, L_A * FL_j^A] * FL_j^A * FL_i^B) \\
&= \min[0, 2 * .50] * .50 * .50 + \min[1, 2 * .50] * .50 * .50 \\
&\quad + \min[0, 2 * .50] * .50 * .50 + \min[0, 2 * .50] * .50 * .50 \\
&= 0 * .50 * .50 + 1 * .50 * .50 + 0 * .50 * .50 + 0 * .50 * .50 = .25 \\
FM^B &= \sum_{i=1}^{2}\sum_{i=1}^{2} \min([(RRA_{ij}^B, L_B * FL_j^B] * FL_j^B * FL_i^A) \\
&= \min[1, 2 * .50] * .50 * .50 + \min[1, 2 * .50] * .50 * .50 \\
&\quad + \min[0, 2 * .50] * .50 * .50 + \min[0, 2 * .50] * .50 * .50
\end{aligned}
$$

$$= 1*.50*.50 + 1*.50*.50 + 0*.50*.50 + 0*.50*.50 = .50$$

Letting $x^i, i = A, B$ represent the number of aircraft on side i, for our example, the number of weapons firings from the opponent aircraft, n, is:

$$n^A = FM^A * \frac{x^A}{x^B} = .25 * \frac{4}{4} = .25$$

$$n^B = FM^B * \frac{x^B}{x^A} = .50 * \frac{4}{4} = .50$$

Thus, the probability of the event that the firing aircraft survives n weapon firings from the opponent aircraft before his first weapon firing for our example is:

$$[1 - PK^A]^{n^A} = [1 - .24]^{.25} = .9333$$

$$[1 - PK^B]^{n^B} = [1 - .33]^{.50} = .9047$$

For our example, LCH_3, the probability of occurrence of scenario three is:

$$LCH_3^A = [1 - LCH_1^A] * ENG^B * [1 - PK^B]^{n^B} = [1 - .25] * .25 * .9047 = .3393$$

$$LCH_3^B = [1 - LCH_1^B] * ENG^A * [1 - PK^A]^{n^A} = [1 - .50] * .60 * .9333 = .2800$$

The probability of firing an aggregate weapon, LCH, is the sum of the probabilities of occurrence for the three scenarios given above. Specifically, $LCH = LCH_1 + LCH_2 + LCH_3$. For our example, LCH is:

$$LCH^A = LCH_1^A + LCH_2^A + LCH_3^A = .25 + .3375 + .3393 = .9268$$

$$LCH^B = LCH_1^B + LCH_2^B + LCH_3^B = .50 + .20 + .2800 = .9800$$

For our example, the last variable in equation 112, the probability of killing an opponent given firing a weapon (PK) were shown in the discussion of scenario three

(LCH_3) to be:

$$PK^A = .24$$
$$PK^B = .33$$

Hence for our example, the single shot probability of kill SSPK using equation 112 are:

$$SSPK^A = ENG^A * LCH^A * PK^A = .60 * .9268 * .24 = .1353$$
$$SSPK^B = ENG^B * LCH^B * PK^B = .50 * .9800 * .33 = .1617$$

The second process in the THUNDER air-to-air engagement model aggregates the aircraft versus aircraft SSPK calculations to attrition rates for the flight versus flight and flight group versus flight group levels. The process is conducted in the following manner for each firing flight versus opponent flight combination:

1. The attrition rate of a *single* aircraft in the opponent flight when engaged by the *entire* firing flight firing *one weapon each* is calculated.

2. The attrition rate of the *single* aircraft in the opponent flight when engaged by the *entire* firing flight firing *all available weapons* is calculated.

First, for each firing flight versus opponent flight combination, the attrition rate of a single aircraft in the opponent flight when engaged by the entire firing flight firing *one weapon* each is calculated. This attrition rate calculation determines two discrete probabilities:

- The probability of i aircraft from the firing flight ($i = 1, \ldots, fac$ where fac is the number of aircraft in the firing flight) engaging the opponent aircraft.

- The probability of j out of i engaging firing aircraft kill an opponent aircraft.

	AEW Control	No AEW Control
MiG-29 ACC Value	.63	.38
F-15C ACC Value	.95	.70

Table 26.   Aircraft Degree of Command and Control (ACC) Values for THUNDER Example

The process, used to calculated the probability that i aircraft from the firing flight will engage the opponent aircraft, is based on a numerical expression of the degree of command and control among aircraft in the firing flight. The concept of command and control was introduced with the probability of detection in equation 103. For each aircraft, a user-inputted variable, termed aircraft degree of command and control (ACC, $0 \leq ACC \leq 1$) is provided. An $ACC = 1$ corresponds to "perfect" command and control and an $ACC = 0$ corresponds to "no" command and control [29: Volume II, page 85]. ACC is a subjective measure of the degree of command and control given an aircraft's sensor suite and weapons configuration. ACC values are also dependent on whether or not the firing flight group is under airborne or ground-based control (termed in THUNDER Airborne Early Warning (AEW) control). Table 26 shows the ACC values for the flights in our example. Earlier we assumed that both flight groups were in AEW control, thus

$$ACC^A = .63$$

$$ACC^B = .95$$

A flight's degree of command and control (FCC) value is assumed to be the average of all the ACC values of the aircraft in the flight. Considering that a flight entity by definition is composed of homogenous aircraft entities, $FCC = ACC$.

THUNDER assumes a weighted combination of two firing flight engagement strategies based on the degree of command and control of the flight. These strategies are:

1. If the firing flight was in a "perfect" command and control state ($FCC = 1$), aircraft in the firing flight would divide themselves evenly over the aircraft in the opponent flight; or

2. If the firing flight was in a "no" command and control state ($FCC = 0$), aircraft in the firing flight would divide themselves randomly over the aircraft in the opponent flight.

Letting $S1_i$, $i = 1, \ldots, fac$ represent the probability that i aircraft from the firing flight engage an aircraft from the opponent flight under the first strategy,

$$S1_i = \begin{cases} 1 - \left(\frac{fac}{opp} - m\right) & \text{if } i = m \\ \frac{fac}{opp} - m & \text{if } i = m + 1 \\ 0 & \text{otherwise} \end{cases} \tag{113}$$

where $m = [\frac{fac}{opp}]$ such that $[x]$ is the integer portion of the value x and opp is the number of aircraft in the opponent flight. Under the second strategy $I \sim bin(fac, \frac{1}{opp})$. Letting $S2_i$, $i = 1, \ldots, fac$ represent the probability that i aircraft from the firing flight engage an aircraft from the opponent flight under the second strategy,

$$S2_i = \begin{cases} \binom{fac}{i} * \left(\frac{1}{opp}\right)^i * \left(1 - \frac{1}{opp}\right)^{fac-i} & \text{if } 0 \leq i \leq fac \\ 0 & \text{otherwise} \end{cases} \tag{114}$$

Let $P_i, i = 1, \ldots, fac$ represent the probability that i aircraft from the firing flight engage an aircraft from the opponent flight. Weighting the probabilities in strategy one ($S1_i$) and strategy two ($S2_i$) by FCC and $1 - FCC$ respectively yields,

$$P_i = FCC * S1_i + (1 - FCC) * S2_i \tag{115}$$

For our example, the MiG-29 and F-15C flights both have 4 firing aircraft entities. For the MiG-29 flight,

$$P_1 = FCC * S1_1 + (1 - FCC) * S2_1 = .63 * 1 + (1 - .63) * .4219 = .7861$$

$$P_2 = FCC * S1_2 + (1 - FCC) * S2_2 = .63 * 0 + (1 - .63) * .2109 = .0780$$

$$P_3 = FCC * S1_3 + (1 - FCC) * S2_3 = .63 * 0 + (1 - .63) * .0469 = .0173$$

$$P_4 = FCC * S1_4 + (1 - FCC) * S2_4 = .63 * 0 + (1 - .63) * .0039 = .0014$$

For the F-15C flight,

$$P_1 = FCC * S1_1 + (1 - FCC) * S2_1 = .95 * 1 + (1 - .95) * .4219 = .9711$$

$$P_2 = FCC * S1_2 + (1 - FCC) * S2_2 = .95 * 0 + (1 - .95) * .2109 = .0105$$

$$P_3 = FCC * S1_3 + (1 - FCC) * S2_3 = .95 * 0 + (1 - .95) * .0469 = .0023$$

$$P_4 = FCC * S1_4 + (1 - FCC) * S2_4 = .95 * 0 + (1 - .95) * .0039 = .0002$$

Allowing for the possibility of multiple kill(s) of an opponent aircraft by different firing aircraft, the probability of j out of the i engaging firing aircraft killing an opponent aircraft is:

$$\binom{i}{j} * (1 - SSPK)^{i-j} * SSPK^j$$

Hence, for each firing flight versus opponent flight combination, the attrition rate (probability of kill) of a *single* aircraft in the opponent flight entity when engaged by the *entire* firing flight is firing *one weapon each*:

$$ATTR = \sum_{i=1}^{fac} P_i \sum_{j=1}^{i} \binom{i}{j} * (1 - SSPK)^{i-j} * SSPK^j \qquad (116)$$

or equivalently:

$$ATTR = \sum_{i=1}^{fac} P_i * \left(1 - (1 - SSPK)^i\right) \qquad (117)$$

For our example,

$$
\begin{aligned}
ATTR_A &= .7961 * .1335 + .0780 * (1 - (1 - .1335)^2) + .0173 * (1 - (1 - .1335)^3) \\
&\quad + .0014 * (1 - (1 - .1335)^4) = .1310 \\
ATTR_B &= .9771 * .1617 + .0105 * (1 - (1 - .1617)^2) + .0023 * (1 - (1 - .1617)^3) \\
&\quad + .0002 * (1 - (1 - .1617)^4) = .1622
\end{aligned}
$$

Next, for each firing flight versus opponent flight combination, the attrition rate of a single aircraft in the opponent flight when engaged by the entire firing flight firing *all available weapons* is calculated. This attrition rate calculation is similar to that shown for the firing flight firing one weapon each. Two discrete probabilities are determined by the calculation:

- The probability of that the opponent flight will be engaged i times by the firing flight entity $(i = 1, \ldots, fl)$ where fl is the maximum number of weapon firings of the firing flight.

- The probability that k out of the i engagements by the firing flight kill an opponent aircraft.

THUNDER assumes a weighted combination of two firing flight group engagement strategies based on a numerical expression of the degree of command and control among flights in the firing flight group and on the maximum number of launches any aircraft in the flight group can make (max). The strategies are:

1. If the firing flight group is in a "perfect" command and control state $(FGCC = 1)$, flights in the friendly flight group would divide themselves evenly over the flights in the opponent flight group; or

2. If the firing flight group is in a "no" command and control state ($FGCC = 0$), flights in the friendly flight group would divide themselves randomly over the flights in the opponent flight group.

A flight group's degree of command and control ($FGCC$) value is assumed to be the average of all FCC values of the flights in the flight group.

Letting $S1_j^{FG}$, $j = 1, \ldots, max$ represent the probability that j flights from the firing flight group engage a flight from the opponent flight group entity under the first strategy,

$$S1_j^{FG} = \begin{cases} 1 - \left(\frac{max}{fli} - m\right) & \text{if } j = m \\ \frac{max}{fli} - m & \text{if } j = m + 1 \\ 0 & \text{otherwise} \end{cases} \tag{118}$$

where $m = [\frac{max}{fli}]$ such that $[x]$ is the integer portion of the value x and fli is the number of flights in the opponent flight. Under the second strategy $J \sim bin(max, \frac{1}{fli})$. Letting $S2_j^{FG}$, $j = 1, \ldots, max$ represent the probability that j flights from the firing flight group engage a flight from the opponent flight under the second strategy,

$$S2_j^{FG} = \begin{cases} \begin{pmatrix} fac \\ i \end{pmatrix} * \left(\frac{1}{fli}\right)^j * \left(1 - \frac{1}{fli}\right)^{max-j} & \text{if } 0 \leq j \leq max \\ 0 & \text{otherwise} \end{cases} \tag{119}$$

Let P_j^{FG}, $j = 1, \ldots, max$ represent the probability that j flights from the firing flight group engage a flight from the opponent flight group. Weighting the probabilities in strategy one ($S1_j^{FG}$) and strategy two ($S2_j^{FG}$) by $FGCC$ and $1 - FGCC$ respectively yields,

$$P_j^{FG} = FCC * S1_j + (1 - FCC) * S2_j \tag{120}$$

A note of clarification may be helpful to the reader at this time. THUNDER assumes that an opponent flight in the opponent flight group will be engaged, at most, max

311

times by flights in the firing flight group. When calculating the probability of that the opponent flight will be engaged i times by a *specific* firing flight ($i = 1, \ldots, fl$) where fl is the maximum number of weapon firings of the firing flight, THUNDER uses the appropriate values of P_j^{FG}. For example, consider a firing flight group composed of two flights. Assume flight one and two can fire 2 and 4 weapons per engagement respectively. The most an opponent flight in the opponent flight group can be engaged is 4 times ($max = 4$). The probability that flight one will engage the opponent flight once and twice is P_1^{FG} and P_2^{FG} respectively.

For our example, each flight group has only one flight. Thus, each flight will engage the other with the maximum number of firing opportunities. The MiG-29 and F-15C flights both have two firing opportunities each. Hence for the MiG-29 and F-15C flights, $P_1^{FG} = 0$ and $P_2^{FG} = 1$.

Allowing for the possibility of multiple kill(s) of an opponent aircraft by the different engagements, the probability that k out of the i engagements by the firing flight kill an opponent aircraft is:

$$\begin{pmatrix} i \\ k \end{pmatrix} * (1 - ATTR)^{i-k} * ATTR^k$$

Hence, for each firing flight versus opponent flight combination, the attrition rate (probability of kill) of a *single* aircraft in the opponent flight when engaged by the *entire* firing flight is firing *all available weapons*:

$$ATTR^{FG} = \sum_{i=1}^{fl} P_i^{FG} \sum_{j=1}^{i} \begin{pmatrix} i \\ j \end{pmatrix} * (1 - ATTR)^{i-j} * ATTR^j \tag{121}$$

or equivalently:

$$ATTR^{FG} = \sum_{i=1}^{fl} P_i^{FG} * \left(1 - (1 - ATTR)^i\right) \tag{122}$$

For our example,

$$ATTR_A^{FG} = 1 - (1 - .1622)^2 = .2980$$

$$ATTR_A^{FG} = 1 - (1 - .1310)^2 = .2149$$

In order to ensure that the attrition rates, $ATTR^{FG}$, that a firing flight inflicts on opponent flights are less than or equal to one, THUNDER normalizes the attrition rates for each firing flight by multiplying the attrition rate by the fraction of the number of firings by the firing flight and the number of total firings by the firing flight group or

$$ATTR^{norm} = ATTR^{FG} * \frac{fl}{tl} \tag{123}$$

where fl designates flight firings and tl designates total flight group firings. For our example $fl = tl$, thus $ATTR^{norm} = ATTR^{FG}$.

The last process in the THUNDER air-to-air engagement model is the loss assessment. The loss assessment proceeds as follows:

- For each opponent flight in the opponent flight group, the number of aircraft lost due to the engagement with each firing flight in the firing flight group is calculated using a draw from a binomial distribution: bin(opp, $ATTR^{norm}$). The actual amount of aircraft loss is the minimum of the binomial draw and the number of weapon firings designated for the firing flight[13].

- The total amount of aircraft loss from the opponent flight is the minimum of the sum of actual aircraft losses for each firing flight in the firing flight group and the number of aircraft in the opponent flight[14].

- This assessment is performed for each opponent flight in the opponent flight group.

---

[13]Clearly, actual aircraft losses can not exceed weapon firings.

[14]Clearly, aircraft losses can not exceed the number of aircraft in the opponent flight.

Appendix C. Event Occurrence Network Solution for One Versus

One Air-to-Air Engagement

Figure 17 shows the EON for this one vs one air-to-air engagement from Hong, et al. [103]. Recall for this engagement that each side (aircraft) has two missiles: one radar-guided and one infrared. The radar-guided missile must always be fired before the infrared missile. Let $G_{i_j^{(\kappa)}}$ represent the event where side $i = 1, 2$ fires a missile of type $j, j = 1, 2$ (where 1 - radar-guided and 2 - infrared) at side $\bar{i}$ that either misses ($\kappa = 0$) or shoots down ($\kappa = 1$) side $\bar{i}$ where $\bar{i}$ represents the opponent side of side i. For example, the event $G_{1_1^{(1)}}$ represents the "concluding" event that the aircraft on side 1 fired a radar missile and shot down the aircraft on side 2.

For each node in figure 17 has the following sequence of events occur:

- Node 1 - no events have taken place;
- Node 2* - $(G_{1_1^{(1)}})$;
- Node 3 - $(G_{1_1^{(0)}})$;
- Node 4 - $(G_{2_1^{(0)}})$;
- Node 5* - $(G_{2_1^{(1)}})$;
- Node 6* - $(G_{1_1^{(0)}}, G_{1_2^{(1)}})$;
- Node 7 - $(G_{1_1^{(0)}}, G_{1_2^{(0)}})$;
- Node 8* - $(G_{1_1^{(0)}}, G_{2_1^{(1)}})$;
- Node 9 -

    - $(G_{1_1^{(0)}}, G_{2_1^{(0)}})$ or
    - $(G_{2_1^{(0)}}, G_{1_1^{(0)}})$;
- Node 10* - $(G_{2_1^{(0)}}, G_{1_1^{(1)}})$;

- Node 11 - $(G_{2_1^{(0)}}, G_{2_2^{(0)}})$;

- Node 12* - $(G_{2_1^{(0)}}, G_{2_2^{(1)}})$;

- Node 13* - $(G_{1_1^{(0)}}, G_{1_2^{(0)}}, G_{2_1^{(1)}})$;

- Node 14 -

  - $(G_{1_1^{(0)}}, G_{1_2^{(0)}}, G_{2_1^{(0)}})$ or

  - $(G_{1_1^{(0)}}, G_{2_1^{(0)}}, G_{1_2^{(0)}})$ or

  - $(G_{2_1^{(0)}}, G_{1_1^{(0)}}, G_{1_2^{(0)}})$;

- Node 15* -

  - $(G_{1_1^{(0)}}, G_{2_1^{(0)}}, G_{1_2^{(1)}})$ or

  - $(G_{2_1^{(0)}}, G_{1_1^{(0)}}, G_{1_2^{(1)}})$;

- Node 16* -

  - $(G_{1_1^{(0)}}, G_{2_1^{(0)}}, G_{2_2^{(1)}})$ or

  - $(G_{2_1^{(0)}}, G_{1_1^{(0)}}, G_{2_2^{(1)}})$;

- Node 17 -

  - $(G_{2_1^{(0)}}, G_{2_2^{(0)}}, G_{1_1^{(0)}})$ or

  - $(G_{1_1^{(0)}}, G_{2_1^{(0)}}, G_{2_2^{(0)}})$ or

  - $(G_{2_1^{(0)}}, G_{1_1^{(0)}}, G_{2_2^{(0)}})$;

- Node 18* - $(G_{2_1^{(0)}}, G_{2_2^{(0)}}, G_{1_1^{(1)}})$;

- Node 19* -

  - $(G_{1_1^{(0)}}, G_{1_2^{(0)}}, G_{2_1^{(0)}}, G_{2_2^{(1)}})$ or

  - $(G_{1_1^{(0)}}, G_{2_1^{(0)}}, G_{1_2^{(0)}}, G_{2_2^{(1)}})$ or

  - $(G_{2_1^{(0)}}, G_{1_1^{(0)}}, G_{1_2^{(0)}}, G_{2_2^{(1)}})$;

- Node 20* -

- $(G_{1_1^{(0)}}, G_{1_2^{(0)}}, G_{2_1^{(0)}}, G_{2_2^{(0)}})$ or

- $(G_{1_1^{(0)}}, G_{2_1^{(0)}}, G_{1_2^{(0)}}, G_{2_2^{(0)}})$ or

- $(G_{2_1^{(0)}}, G_{1_1^{(0)}}, G_{1_2^{(0)}}, G_{2_2^{(0)}})$ or

- $(G_{2_1^{(0)}}, G_{2_2^{(0)}}, G_{1_1^{(0)}}, G_{1_2^{(0)}})$ or

- $(G_{1_1^{(0)}}, G_{2_1^{(0)}}, G_{2_2^{(0)}}, G_{1_2^{(0)}})$ or

- $(G_{2_1^{(0)}}, G_{1_1^{(0)}}, G_{2_2^{(0)}}, G_{1_2^{(0)}})$.

- Node 21* -

  - $(G_{2_1^{(0)}}, G_{2_2^{(0)}}, G_{1_1^{(0)}}, G_{1_2^{(1)}})$ or

  - $(G_{1_1^{(0)}}, G_{2_1^{(0)}}, G_{2_2^{(0)}}, G_{1_2^{(1)}})$ or

  - $(G_{2_1^{(0)}}, G_{1_1^{(0)}}, G_{2_2^{(0)}}, G_{1_2^{(1)}})$.

Nodes marked with an $*$ denote a sequence of intermediate events that end with a concluding event. Hence, nodes $1, 5, 6, 8, 10, 12, 13, 15, 16, 18, 19,$ and 21 end with concluding events while node 20 ends the engagement since no events are left to occur (i.e., both aircraft are out of weapons).

Let continuous random variable $E_{i_j^{(\kappa)}}$ represent the occurrence time of event $G_{i_j^{(\kappa)}}$ with

- density function $f_{E_{i_j^{(\kappa)}}}(t)$;

- distribution function $F_{E_{i_j^{(\kappa)}}}(t)$; and

- complementary distribution function $F^c_{E_{i_j^{(\kappa)}}}(t)$.

Also, let p_{i_j} and $q_{i_j} = 1 - p_{i_j}$ represent the probabilities that missile j from side i shoot downs and misses respectively side $\bar{i}$ given that the missile j is fired. For the example shown in chapter IV,

- $E_{1_1^{(\kappa)}} \sim exp(\lambda_{1_1} t)$ with $\lambda_{1_1} = 1.25$ (i.e., side 1 fires a radar missile);

- $E_{2_1^{(\kappa)}} \sim exp(\lambda_{2_1} t)$ with $\lambda_{2_1} = 1.15$ (i.e., side 2 fires a radar missile);

316

- $E_{1_2^{(\kappa)}} \sim exp(\lambda_{1_2}t)$ with $\lambda_{1_2} = 1.85$ (i.e., side 1 fires a heat missile); and

- $E_{2_2^{(\kappa)}} \sim exp(\lambda_{2_2}t)$ with $\lambda_{2_2} = 1.65$ (i.e., side 2 fires a heat missile).

Additionally, the probabilities that the radar and infrared missiles on side 1 and 2 shoot down the other side when fired are $p_{1_1} = p_{2_1} = 0.65$ and $p_{1_2} = 0.85$ - $p_{2_2} = 0.75$ respectively.

Exact expressions for the probabilities in Table 2 were derived and are summarized below:

$$
\begin{aligned}
P_1(t) &= P(E_{1_1^{(\kappa)}} > t, E_{2_1^{(\kappa)}} > t) \\
&= \left[\int_t^\infty f_{E_{1_1^{(\kappa)}}}(x_1)dx_1\right] * \left[\int_t^\infty f_{E_{2_1^{(\kappa)}}}(y_1)dy_1\right] \\
&= \left[\int_t^\infty \lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1\right] * \left[\int_t^\infty \lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1\right] \\
&= \exp(-(\lambda_{1_1}+\lambda_{2_1})t) \\
P_2(t) &= P(E_{2_1^{(\kappa)}} > E_{1_1^{(1)}}) \\
&= \int_0^t F^c_{E_{2_1^{(\kappa)}}}(x_1)f_{E_{1_1^{(1)}}}(x_1)dx_1 \\
&= \int_0^t \exp(-\lambda_{2_1}x_1) * p_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1 \\
&= p_{1_1}\left[\frac{\lambda_{1_1}}{\lambda_{1_1}+\lambda_{2_1}}\right][1-\exp(-(\lambda_{1_1}+\lambda_{2_1})t)] \\
P_3(t) &= P(E_{1_2^{(\kappa)}} > t - E_{1_1^{(0)}}, E_{2_1^{(\kappa)}} > t) \\
&= F^c_{E_{2_1^{(\kappa)}}} * \left[\int_0^t F^c_{E_{1_2^{(\kappa)}}}(t-x_1)f_{E_{1_1^{(0)}}}(x_1)dx_1\right] \\
&= \exp(-\lambda_{2_1}t) * \left[\int_0^t \exp(-\lambda_{1_2}(t-x_1)) * q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1\right] \\
&= q_{1_1}\left[\frac{\lambda_{1_1}}{\lambda_{1_2}-\lambda_{1_1}}\right][\exp(-(\lambda_{1_1}+\lambda_{2_1})t) - \exp(-(\lambda_{2_1}+\lambda_{1_2})t)] \\
P_4(t) &= P(E_{1_1^{(\kappa)}} > t, E_{2_2^{(\kappa)}} > t - E_{2_1^{(0)}}) \\
&= F^c_{E_{1_1^{(\kappa)}}} * \left[\int_0^t F^c_{E_{2_2^{(\kappa)}}}(t-y_1)f_{E_{2_1^{(0)}}}(y_1)dy_1\right] \\
&= \exp(-\lambda_{1_1}t) * \left[\int_0^t \exp(-\lambda_{2_2}(t-y_1)) * q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1\right]
\end{aligned}
$$

317

$$= q_{2_1}\left[\frac{\lambda_{2_1}}{\lambda_{2_2} - \lambda_{2_1}}\right][\exp(-(\lambda_{1_1} + \lambda_{2_1})t) - \exp(-(\lambda_{1_1} + \lambda_{2_2})t)]$$

$$P_5(t) = P(E_{1_1^{(\kappa)}} > E_{2_1^{(1)}})$$

$$= \int_0^t F^c_{E_{1_1^{(\kappa)}}}(y_1)f_{E_{2_1^{(1)}}}(y_1)dy_1$$

$$= \int_0^t \exp(-\lambda_{1_1}y_1) * p_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1$$

$$= p_{2_1}\left[\frac{\lambda_{2_1}}{\lambda_{1_1} + \lambda_{2_1}}\right][1 - \exp(-(\lambda_{1_1} + \lambda_{2_1})t)]$$

$$P_6(t) = P(E_{2_1^{(\kappa)}} > E_{1_1^{(0)}} + E_{1_2^{(1)}})$$

$$= \int_0^t F^c_{E_{2_1^{(\kappa)}}}(x_T)\left[\int_0^{x_T} f_{E_{1_2^{(1)}}}(x_T - x_1)f_{E_{1_1^{(0)}}}(x_1)dx_1\right]dx_T$$

$$= \int_0^t \exp(-\lambda_{2_1}x_T) *$$
$$\left[\int_0^{x_T} p_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T - x_1)) * q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1\right]dx_T$$

$$= p_{1_2}q_{1_1}\frac{\lambda_{1_1}\lambda_{1_2}}{(\lambda_{2_1} + \lambda_{1_2})(\lambda_{1_1} + \lambda_{2_1})}\left[1 + \frac{(\lambda_{1_1} + \lambda_{2_1})}{(\lambda_{1_2} - \lambda_{1_1})}\exp(-(\lambda_{2_1} + \lambda_{1_2})t)\right.$$
$$\left. - \frac{(\lambda_{2_1} + \lambda_{1_2})}{(\lambda_{1_2} - \lambda_{1_1})}\exp(-(\lambda_{1_1} + \lambda_{2_1})t)\right]$$

$$P_7(t) = P(E_{1_1^{(0)}} + E_{1_2^{(0)}} \le t, E_{2_1^{(\kappa)}} > t)$$

$$= F^c_{E_{2_1^{(\kappa)}}}(t)\left[\int_0^t \int_0^{x_T} f_{E_{1_2^{(0)}}}(x_T - x_1)f_{E_{1_1^{(0)}}}(x_1)dx_1 dx_T\right]$$

$$= \exp(-\lambda_{2_1}t) *$$
$$\left[\int_0^t \int_0^{x_T} q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T - x_1)) * q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1 dx_T\right]$$

$$= q_{1_1}q_{1_2}\exp(-\lambda_{2_1}t)$$
$$\left[1 + \frac{\lambda_{1_1}}{\lambda_{1_2} - \lambda_{1_1}}\exp(-\lambda_{1_2}t) - \frac{\lambda_{1_2}}{\lambda_{1_2} - \lambda_{1_1}}\exp(-\lambda_{1_1}t)\right]$$

$$P_8(t) = P(E_{1_1^{(0)}} + E_{1_2^{(\kappa)}} > E_{2_1^{(1)}})$$

$$= \int_0^t f_{E_{2_1^{(1)}}}(y_1)\left[\int_0^{y_1} F^c_{E_{1_2^{(\kappa)}}}(y_1 - x_1)f_{E_{1_1^{(0)}}}(x_1)dx_1\right]dy_1$$

$$= \int_0^t p_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1) *$$
$$\left[\int_0^{y_1} \exp(-\lambda_{1_2}(y_1 - x_1)) * q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1\right]dy_1$$

$$= q_{1_1}p_{2_1}\frac{\lambda_{1_1}\lambda_{2_1}}{(\lambda_{2_1} + \lambda_{1_2})(\lambda_{1_1} + \lambda_{2_1})} *$$

$$\left[1 + \frac{\lambda_{1_1} + \lambda_{2_1}}{\lambda_{1_2} - \lambda_{1_1}} \exp(-(\lambda_{2_1} + \lambda_{1_2})t) - \frac{\lambda_{2_1} + \lambda_{1_2}}{\lambda_{1_2} - \lambda_{1_1}} \exp(-(\lambda_{1_1} + \lambda_{2_1})t)\right]$$

$$\begin{aligned}
P_9(t) &= P(E_{1_1^{(0)}} + E_{1_2^{(\kappa)}} > t, E_{2_1^{(0)}} + E_{2_2^{(\kappa)}} > t, E_{1_1^{(0)}} < E_{2_1^{(0)}}) + \\
&\quad P(E_{1_1^{(0)}} + E_{1_2^{(\kappa)}} > t, E_{2_1^{(0)}} + E_{2_2^{(\kappa)}} > t, E_{2_1^{(0)}} < E_{1_1^{(0)}}) \\
&= \int_0^t F^c_{E_{2_2^{(\kappa)}}}(t - y_1) f_{E_{2_1^{(0)}}}(y_1) \left[\int_0^{y_1} F^c_{E_{1_2^{(\kappa)}}}(t - x_1) f_{E_{1_1^{(0)}}}(x_1) dx_1\right] dy_1 + \\
&\quad \int_0^t F^c_{E_{1_2^{(\kappa)}}}(t - x_1) f_{E_{1_1^{(0)}}}(x_1) \left[\int_0^{x_1} F^c_{E_{2_2^{(\kappa)}}}(t - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1\right] dx_1 \\
&= \left[\int_0^t \exp(-\lambda_{2_2}(t - y_1)) * q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1) * \right. \\
&\quad \left. \left(\int_0^{y_1}\exp(-\lambda_{1_2}(t-x_1)) * q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1\right)dy_1\right] + \\
&\quad \left[\int_0^t \exp(-\lambda_{1_2}(t - x_1)) * q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1) * \right. \\
&\quad \left. \left(\int_0^{x_1}\exp(-\lambda_{2_2}(t-y_1)) * q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1\right)dx_1\right] \\
&= q_{1_1}q_{2_1}\frac{\lambda_{1_1}\lambda_{2_1}}{(\lambda_{1_2} - \lambda_{1_1})(\lambda_{2_2} - \lambda_{2_1})} * [\exp(-(\lambda_{1_1} + \lambda_{2_1})t) + \\
&\quad \exp(-(\lambda_{1_2} + \lambda_{2_2})t) - \exp(-(\lambda_{1_1} + \lambda_{2_2})t) - \exp(-(\lambda_{2_1} + \lambda_{1_2})t)] \\[6pt]
P_{10}(t) &= P(E_{2_1^{(0)}} + E_{2_2^{(\kappa)}} > E_{1_1^{(1)}}) \\
&= \int_0^t f_{E_{1_1^{(1)}}}(x_1)\left[\int_0^{x_1} F^c_{E_{2_2^{(\kappa)}}}(x_1 - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1\right] dx_1 \\
&= \int_0^t p_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1) * \\
&\quad \left[\int_0^{x_1}\exp(-\lambda_{2_2}(x_1 - y_1)) * q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1\right] dx_1 \\
&= q_{2_1}p_{1_1}\frac{\lambda_{1_1}\lambda_{2_1}}{(\lambda_{1_1} + \lambda_{2_2})(\lambda_{1_1} + \lambda_{2_1})} * \\
&\quad \left[1 + \frac{\lambda_{1_1} + \lambda_{2_1}}{\lambda_{2_2} - \lambda_{2_1}}\exp(-(\lambda_{1_1} + \lambda_{2_2})t) - \frac{\lambda_{1_1} + \lambda_{2_2}}{\lambda_{2_2} - \lambda_{2_1}}\exp(-(\lambda_{1_1} + \lambda_{2_1})t)\right] \\[6pt]
P_{11}(t) &= P(E_{1_1^{(\kappa)}} > t, E_{2_1^{(0)}} + E_{2_2^{(0)}} \le t) \\
&= F^c_{E_{1_1^{(\kappa)}}}(t)\left[\int_0^t\int_0^{y_T} f_{E_{2_2^{(0)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1)dx_1dy_T\right] \\
&= \exp(-\lambda_{1_1}t) * \\
&\quad \left[\int_0^t\int_0^{y_2} q_{2_2}\lambda_{2_2}\exp(-\lambda_{2_2}(y_2 - y_1)) * q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1dy_2\right] \\
&= q_{2_1}q_{2_2}\exp(-\lambda_{1_1}t)
\end{aligned}$$

$$\left[1 + \frac{\lambda_{2_1}}{\lambda_{2_2} - \lambda_{2_1}} \exp(-\lambda_{2_2} t) - \frac{\lambda_{2_2}}{\lambda_{2_2} - \lambda_{2_1}} \exp(-\lambda_{2_1} t)\right]$$

$$
\begin{aligned}
P_{12}(t) &= P(E_{1_1^{(\kappa)}} > E_{2_1^{(0)}} + E_{2_2^{(1)}}) \\
&= \int_0^t F^c_{E_{1_1^{(\kappa)}}}(y_T) \left[\int_0^{y_T} f_{E_{2_2^{(1)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1\right] dy_T \\
&= \int_0^t \exp(-\lambda_{1_1} y_T) * \\
&\quad \left[\int_0^{y_T} p_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) dy_1\right] dy_T \\
&= p_{2_2} q_{2_1} \frac{\lambda_{2_1} \lambda_{2_2}}{(\lambda_{1_1} + \lambda_{2_2})(\lambda_{1_1} + \lambda_{2_1})} \left[1 + \frac{(\lambda_{1_1} + \lambda_{2_1})}{(\lambda_{2_2} - \lambda_{2_1})} \exp(-(\lambda_{1_1} + \lambda_{2_2})t)\right. \\
&\quad \left. - \frac{(\lambda_{1_1} + \lambda_{2_2})}{(\lambda_{2_2} - \lambda_{2_1})} \exp(-(\lambda_{1_1} + \lambda_{2_1})t)\right] \\
P_{13}(t) &= P(E_{1_1^{(0)}} + E_{1_2^{(0)}} < E_{2_1^{(1)}}) \\
&= \int_0^t f_{E_{2_1^{(1)}}}(y_1) \left[\int_0^{y_1} \int_0^{x_T} f_{E_{1_2^{(0)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) dx_1 dx_T\right] dy_1 \\
&= \int_0^t p_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) * \\
&\quad \left[\int_0^{y_1} \int_0^{x_T} q_{1_2} \lambda_{1_2} \exp(-\lambda_{1_2}(x_T - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) dx_1 dx_T\right] dy_1 \\
&= \frac{q_{1_1} q_{1_2} p_{2_1}}{(\lambda_{1_2} - \lambda_{1_1})} * \left[(\lambda_{1_2} - \lambda_{1_1}) - \frac{\lambda_{2_1} \lambda_{1_2}}{(\lambda_{1_1} + \lambda_{2_1})} + \frac{\lambda_{1_1} \lambda_{2_1}}{(\lambda_{2_1} + \lambda_{1_2})} - \right. \\
&\quad \exp(-\lambda_{2_1} t) + \frac{\lambda_{2_1} \lambda_{1_2}}{(\lambda_{1_1} + \lambda_{2_1})} \exp(-(\lambda_{1_1} + \lambda_{2_1})t) - \\
&\quad \left. \frac{\lambda_{1_1} \lambda_{2_1}}{(\lambda_{2_1} + \lambda_{1_2})} \exp(-(\lambda_{2_1} + \lambda_{1_2})t)\right] \\
P_{14}(t) &= P(E_{2_1^{(0)}} + E_{2_2^{(\kappa)}} > t, E_{1_1^{(0)}} < E_{2_1^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(0)}} < t) + \\
&\quad P(E_{2_1^{(0)}} + E_{2_2^{(\kappa)}} > t, E_{2_1^{(0)}} < E_{1_1^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(0)}} < t) + \\
&\quad P(E_{2_1^{(0)}} + E_{2_2^{(\kappa)}} > t, E_{1_1^{(0)}} + E_{1_2^{(0)}} < E_{2_1^{(0)}} < t) + \\
&= \int_0^t \int_0^{x_T} F^c_{E_{2_2^{(\kappa)}}}(t - y_1) f_{E_{2_1^{(0)}}}(y_1) \left[\int_0^{y_1} f_{E_{1_2^{(0)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) dx_1\right] dy_1 dx_T + \\
&\quad \int_0^t \int_0^{x_T} f_{E_{1_2^{(0)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) \left[\int_0^{x_1} F^c_{E_{2_2^{(\kappa)}}}(t - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1\right] dx_1 dx_T + \\
&\quad \int_0^t F^c_{E_{2_2^{(\kappa)}}}(t - y_1) f_{E_{2_1^{(0)}}}(y_1) \left[\int_0^{y_1} \int_0^{x_T} f_{E_{1_2^{(0)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) dx_1 dx_T\right] dy_1 \\
&= \int_0^t \int_0^{x_T} \exp(-\lambda_{2_2}(t - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) *
\end{aligned}
$$

$$\left[\int_0^{y_1} q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1\right]dy_1dx_T+$$

$$\int_0^t\int_0^{x_T}q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)*$$

$$\left[\int_0^{x_1}\exp(-\lambda_{2_2}(t-y_1))*q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1\right]dx_1dx_T+$$

$$\int_0^t\exp(-\lambda_{2_2}(t-y_1))*q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)*$$

$$\left[\int_0^{y_1}\int_0^{x_T}q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1dx_T\right]dy_1$$

$$=\frac{q_{1_1}q_{1_2}q_{2_1}\lambda_{2_1}}{(\lambda_{1_2}-\lambda_{1_1})(\lambda_{2_2}-\lambda_{2_1})}\left[-(\lambda_{1_2}-\lambda_{1_1})\exp(-\lambda_{2_2}t)-\right.$$

$$\lambda_{1_2}\exp(-(\lambda_{1_1}+\lambda_{2_1})t)-\lambda_{1_1}\exp(-(\lambda_{1_2}+\lambda_{2_2})t)+$$

$$\lambda_{1_1}\exp(-(\lambda_{2_1}+\lambda_{1_2})t)+\lambda_{1_2}\exp(-(\lambda_{1_1}+\lambda_{2_2})t)+$$

$$\left.(\lambda_{1_2}-\lambda_{1_1})\exp(-\lambda_{2_1}t)\right]$$

$$P_{15}(t)=P(E_{2_1^{(0)}}+E_{2_2^{(\kappa)}}>E_{1_1^{(0)}}+E_{1_2^{(1)}},E_{2_1^{(0)}}>E_{1_1^{(0)}})+$$

$$P(E_{2_1^{(0)}}+E_{2_2^{(\kappa)}}>E_{1_1^{(0)}}+E_{1_2^{(1)}},E_{2_1^{(0)}}<E_{1_1^{(0)}})$$

$$=\int_0^t\int_0^{x_T}F^c_{E_{2_2^{(\kappa)}}}(x_T-y_1)f_{E_{2_1^{(0)}}}(y_1)\left[\int_0^{y_1}f_{E_{1_2^{(1)}}}(x_T-x_1)f_{E_{1_1^{(0)}}}(x_1)dx_1\right]dy_1dx_T+$$

$$\int_0^t\int_0^{x_T}f_{E_{1_2^{(1)}}}(x_T-x_1)f_{E_{1_1^{(0)}}}(x_1)\left[\int_0^{x_1}F^c_{E_{2_2^{(\kappa)}}}(x_T-y_1)f_{E_{2_1^{(0)}}}(y_1)dy_1\right]dx_1dx_T$$

$$=\int_0^t\int_0^{x_T}\exp(-\lambda_{2_2}(x_T-y_1))*q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)*$$

$$\left[\int_0^{y_1}p_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1\right]dy_1dx_T+$$

$$\int_0^t\int_0^{x_T}p_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)*$$

$$\left[\int_0^{x_1}\exp(-\lambda_{2_2}(x_T-y_1))*q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1\right]dx_1dx_T$$

$$=\frac{q_{1_1}p_{1_2}q_{2_1}\lambda_{1_1}\lambda_{1_2}\lambda_{2_1}}{(\lambda_{1_2}-\lambda_{1_1})(\lambda_{2_2}-\lambda_{2_1})}\left[\frac{1}{\lambda_{1_1}+\lambda_{2_1}}(1-\exp(-(\lambda_{1_1}+\lambda_{2_1})t))+\right.$$

$$\frac{1}{\lambda_{1_2}+\lambda_{2_2}}(1-\exp(-(\lambda_{1_2}+\lambda_{2_2})t))-\frac{1}{\lambda_{2_1}+\lambda_{1_2}}*$$

$$\left.(1-\exp(-(\lambda_{2_1}+\lambda_{1_2})t))-\frac{1}{\lambda_{1_1}+\lambda_{2_2}}(1-\exp(-(\lambda_{1_1}+\lambda_{2_2})t))\right]$$

$$P_{16}(t)=P(E_{1_1^{(0)}}+E_{1_2^{(\kappa)}}>E_{2_1^{(0)}}+E_{2_2^{(1)}},E_{2_1^{(0)}}>E_{1_1^{(0)}})+$$

$$P(E_{1_1^{(0)}}+E_{1_2^{(\kappa)}}>E_{2_1^{(0)}}+E_{2_2^{(1)}},E_{2_1^{(0)}}<E_{1_1^{(0)}})$$

$$= \int_0^t \int_0^{y_T} f_{E_{2_2^{(1)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) \left[\int_0^{y_1} F_{E_{1_2^{(\kappa)}}}^c(y_T - x_1) f_{E_{1_1^{(0)}}}(x_1) dx_1 \right] dy_1 dy_T +$$

$$\int_0^t \int_0^{y_T} F_{E_{1_2^{(\kappa)}}}^c(y_T - x_1) f_{E_{1_1^{(0)}}}(x_1) \left[\int_0^{x_1} f_{E_{2_2^{(1)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1 \right] dx_1 dy_T$$

$$= \int_0^t \int_0^{y_T} p_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) *$$

$$\left[\int_0^{y_1} \exp(-\lambda_{1_2}(y_T - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) dx_1 \right] dy_1 dy_T +$$

$$\int_0^t \int_0^{y_T} \exp(-\lambda_{1_2}(y_T - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) *$$

$$\left[\int_0^{x_1} p_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) dy_1 \right] dx_1 dy_T$$

$$= \frac{q_{1_1} q_{2_1} p_{2_2} \lambda_{1_1} \lambda_{2_1} \lambda_{2_2}}{(\lambda_{1_2} - \lambda_{1_1})(\lambda_{2_2} - \lambda_{2_1})} \left[\frac{1}{\lambda_{1_1} + \lambda_{2_1}}(1 - \exp(-(\lambda_{1_1} + \lambda_{2_1})t)) + \right.$$

$$\frac{1}{\lambda_{1_2} + \lambda_{2_2}}(1 - \exp(-(\lambda_{1_2} + \lambda_{2_2})t)) - \frac{1}{\lambda_{1_1} + \lambda_{2_2}} *$$

$$\left. (1 - \exp(-(\lambda_{1_1} + \lambda_{2_2})t)) - \frac{1}{\lambda_{2_1} + \lambda_{1_2}}(1 - \exp(-(\lambda_{2_1} + \lambda_{1_2})t)) \right]$$

$$P_{17}(t) = P(E_{1_1^{(0)}} + E_{1_2^{(\kappa)}} > t, E_{2_1^{(0)}} < E_{1_1^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(0)}} < t) +$$

$$P(E_{1_1^{(0)}} + E_{1_2^{(\kappa)}} > t, E_{1_1^{(0)}} < E_{2_1^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(0)}} < t) +$$

$$P(E_{1_1^{(0)}} + E_{1_2^{(\kappa)}} > t, E_{2_1^{(0)}} + E_{2_2^{(0)}} < E_{1_1^{(0)}} < t) +$$

$$= \int_0^t \int_0^{y_T} F_{E_{1_2^{(\kappa)}}}^c(t - x_1) f_{E_{1_1^{(0)}}}(x_1) \left[\int_0^{x_1} f_{E_{2_2^{(0)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1 \right] dx_1 dy_T +$$

$$\int_0^t \int_0^{y_T} f_{E_{2_2^{(0)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) \left[\int_0^{y_1} F_{E_{1_2^{(\kappa)}}}^c(t - x_1) f_{E_{1_1^{(0)}}}(x_1) dx_1 \right] dy_1 dy_T +$$

$$\int_0^t F_{E_{1_2^{(\kappa)}}}^c(t - x_1) f_{E_{1_1^{(0)}}}(x_1) \left[\int_0^{x_1} \int_0^{y_T} f_{E_{2_2^{(0)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1 dy_T \right] dx_1$$

$$= \int_0^t \int_0^{y_T} \exp(-\lambda_{1_2}(t - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) *$$

$$\left[\int_0^{x_1} q_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) dy_1 \right] dx_1 dy_T +$$

$$\int_0^t \int_0^{y_T} q_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) *$$

$$\left[\int_0^{y_1} \exp(-\lambda_{1_2}(t - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) dx_1 \right] dy_1 dy_T +$$

$$\int_0^t \exp(-\lambda_{1_2}(t - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) *$$

$$\left[\int_0^{x_1} \int_0^{y_T} q_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) dy_1 dy_T \right] dx_1$$

$$
= \frac{q_{1_1} q_{2_1} q_{2_2} \lambda_{1_1}}{(\lambda_{1_2} - \lambda_{1_1})(\lambda_{2_2} - \lambda_{2_1})} \left[-(\lambda_{2_2} - \lambda_{2_1}) \exp(-\lambda_{1_2} t) - \right.
$$
$$
\lambda_{2_2} \exp(-(\lambda_{1_1} + \lambda_{2_1})t) - \lambda_{2_1} \exp(-(\lambda_{1_2} + \lambda_{2_2})t) +
$$
$$
\lambda_{2_1} \exp(-(\lambda_{1_1} + \lambda_{2_2})t) + \lambda_{2_2} \exp(-(\lambda_{2_1} + \lambda_{1_2})t) +
$$
$$
\left. (\lambda_{2_2} - \lambda_{2_1}) \exp(-\lambda_{1_1} t) \right]
$$

$$
P_{18}(t) = P(E_{2_1^{(0)}} + E_{2_2^{(0)}} < E_{1_1^{(1)}})
$$
$$
= \int_0^t f_{E_{1_1^{(1)}}}(x_1) \left[\int_0^{x_1} \int_0^{y_T} f_{E_{2_2^{(0)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1 dy_T \right] dx_1
$$
$$
= \int_0^t p_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) *
$$
$$
\left[\int_0^{x_1} \int_0^{y_T} q_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) dy_1 dy_T \right] dx_1
$$
$$
= \frac{q_{2_1} q_{2_2} p_{1_1}}{(\lambda_{2_2} - \lambda_{2_1})} * \left[(\lambda_{2_2} - \lambda_{2_1}) - \frac{\lambda_{1_1} \lambda_{2_2}}{(\lambda_{1_1} + \lambda_{2_1})} + \frac{\lambda_{1_1} \lambda_{2_1}}{(\lambda_{1_1} + \lambda_{2_2})} - \right.
$$
$$
\exp(-\lambda_{1_1} t) + \frac{\lambda_{1_1} \lambda_{2_2}}{(\lambda_{1_1} + \lambda_{2_1})} \exp(-(\lambda_{1_1} + \lambda_{2_1})t) -
$$
$$
\left. \frac{\lambda_{1_1} \lambda_{2_1}}{(\lambda_{1_1} + \lambda_{2_2})} \exp(-(\lambda_{1_1} + \lambda_{2_2})t) \right]
$$

$$
P_{19}(t) = P(E_{1_1^{(0)}} < E_{2_1^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(1)}} < t) +
$$
$$
P(E_{1_1^{(0)}} + E_{1_2^{(0)}} < E_{2_1^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(1)}} < t) +
$$
$$
P(E_{2_1^{(0)}} < E_{1_1^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(1)}} < t)
$$
$$
= \int_0^t \int_0^{y_T} \int_0^{x_T} f_{E_{2_2^{(1)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) \left[\int_0^{y_1} f_{E_{1_2^{(0)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) dx_1 \right] dy_1 dx_T dy_T +
$$
$$
\int_0^t \int_0^{y_T} f_{E_{2_2^{(1)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) \left[\int_0^{y_1} \int_0^{x_T} f_{E_{1_2^{(0)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) dx_1 dx_T \right] dy_1 dy_T +
$$
$$
\int_0^t \int_0^{y_T} \int_0^{x_T} f_{E_{1_2^{(0)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) \left[\int_0^{x_1} f_{E_{2_2^{(1)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1 \right] dx_1 dx_T dy_T
$$
$$
= \int_0^t \int_0^{y_T} \int_0^{x_T} p_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) *
$$
$$
\left[\int_0^{y_1} q_{1_2} \lambda_{1_2} \exp(-\lambda_{1_2}(x_T - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) dx_1 \right] dy_1 dx_T dy_T
$$
$$
+ \int_0^t \int_0^{y_T} p_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) *
$$
$$
\left[\int_0^{y_1} \int_0^{x_T} q_{1_2} \lambda_{1_2} \exp(-\lambda_{1_2}(x_T - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) dx_1 dx_T \right] dy_1 dy_T
$$

$$+ \int_0^t \int_0^{y_T} \int_0^{x_T} q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T - x_1)) * q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1) *$$

$$\left[\int_0^{x_1} p_{2_2}\lambda_{2_2}\exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1\right]dx_1 dx_T dy_T$$

$$= \frac{q_{1_1}q_{1_2}q_{2_1}p_{2_2}\lambda_{2_1}\lambda_{2_2}}{(\lambda_{1_2} - \lambda_{1_1})(\lambda_{2_2} - \lambda_{2_1})}\left[(\exp(-\lambda_{2_2}t) - 1)\left(\frac{\lambda_{1_2} - \lambda_{1_1}}{\lambda_{2_2}}\right) + \right.$$

$$(\exp(-(\lambda_{1_1} + \lambda_{2_1})t) - 1)\left(\frac{\lambda_{1_2}}{\lambda_{1_1} + \lambda_{2_1}}\right) + (\exp(-(\lambda_{1_2} + \lambda_{2_2})t) - 1) *$$

$$\left(\frac{\lambda_{1_1}}{\lambda_{1_2} + \lambda_{2_2}}\right) + (\exp(-(\lambda_{2_1} + \lambda_{1_2})t) - 1)\left(\frac{-\lambda_{1_1}}{\lambda_{2_1} + \lambda_{1_2}}\right) +$$

$$(\exp(-(\lambda_{1_1} + \lambda_{2_2})t) - 1)\left(\frac{-\lambda_{1_2}}{\lambda_{1_1} + \lambda_{2_2}}\right) + (\exp(-(\lambda_{2_1}t) - 1) *$$

$$\left.\left(\frac{-(\lambda_{1_2} - \lambda_{1_1})}{\lambda_{2_1}}\right)\right]$$

$$P_{20}(t) = P(E_{1_1^{(0)}} < E_{2_1^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(0)}} < t) +$$

$$P(E_{1_1^{(0)}} + E_{1_2^{(0)}} < E_{2_1^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(0)}} < t) +$$

$$P(E_{2_1^{(0)}} < E_{1_1^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(0)}} < t) +$$

$$P(E_{2_1^{(0)}} < E_{1_1^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(0)}} < t) +$$

$$P(E_{2_1^{(0)}} + E_{2_2^{(0)}} < E_{1_1^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(0)}} < t) +$$

$$P(E_{1_1^{(0)}} < E_{2_1^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(0)}} < t)$$

$$= \int_0^t \int_0^{y_T} \int_0^{x_T} f_{E_{2_2^{(0)}}}(y_T - y_1)f_{E_{2_1^{(0)}}}(y_1)\left[\int_0^{y_1} f_{E_{1_2^{(0)}}}(x_T - x_1)f_{E_{1_1^{(0)}}}(x_1)dx_1\right]dy_1 dx_T dy_T +$$

$$\int_0^t \int_0^{y_T} f_{E_{2_2^{(0)}}}(y_T - y_1)f_{E_{2_1^{(0)}}}(y_1)\left[\int_0^{y_1} \int_0^{x_T} f_{E_{1_2^{(0)}}}(x_T - x_1)f_{E_{1_1^{(0)}}}(x_1)dx_1 dx_T\right]dy_1 dy_T +$$

$$\int_0^t \int_0^{y_T} \int_0^{x_T} f_{E_{1_2^{(0)}}}(x_T - x_1)f_{E_{1_1^{(0)}}}(x_1)\left[\int_0^{x_1} f_{E_{2_2^{(0)}}}(y_T - y_1)f_{E_{2_1^{(0)}}}(y_1)dy_1\right]dx_1 dx_T dy_T +$$

$$\int_0^t \int_0^{x_T} \int_0^{y_T} f_{E_{1_2^{(0)}}}(x_T - x_1)f_{E_{1_1^{(0)}}}(x_1)\left[\int_0^{x_1} f_{E_{2_2^{(0)}}}(y_T - y_1)f_{E_{2_1^{(0)}}}(y_1)dy_1\right]dx_1 dy_T dx_T +$$

$$\int_0^t \int_0^{x_T} f_{E_{1_2^{(0)}}}(x_T - x_1)f_{E_{1_1^{(0)}}}(x_1)\left[\int_0^{x_1} \int_0^{y_T} f_{E_{2_2^{(0)}}}(y_T - y_1)f_{E_{2_1^{(0)}}}(y_1)dy_1 dy_T\right]dx_1 dx_T +$$

$$\int_0^t \int_0^{x_T} \int_0^{y_T} f_{E_{2_2^{(0)}}}(y_T - y_1)f_{E_{1_1^{(0)}}}(y_1)\left[\int_0^{y_1} f_{E_{1_2^{(0)}}}(x_T - x_1)f_{E_{1_1^{(0)}}}(x_1)dx_1\right]dy_1 dy_T dx_T$$

$$= \int_0^t \int_0^{y_T} \int_0^{x_T} q_{2_2}\lambda_{2_2}\exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1) *$$

$$\left[\int_0^{y_1} q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1\right]dy_1dx_Tdy_T$$

$$+\int_0^t\int_0^{y_T} q_{2_2}\lambda_{2_2}\exp(-\lambda_{2_2}(y_T-y_1))*q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)*$$

$$\left[\int_0^{y_1}\int_0^{x_T} q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1dx_T\right]dy_1dy_T$$

$$+\int_0^t\int_0^{y_T}\int_0^{x_T} q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)*$$

$$\left[\int_0^{x_1} q_{2_2}\lambda_{2_2}\exp(-\lambda_{2_2}(y_T-y_1))*q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1\right]dx_1dx_Tdy_T+$$

$$\int_0^t\int_0^{x_T}\int_0^{y_T} q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)*$$

$$\left[\int_0^{x_1} q_{2_2}\lambda_{2_2}\exp(-\lambda_{2_2}(y_T-y_1))*q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1\right]dx_1dy_Tdx_T$$

$$+\int_0^t\int_0^{x_T} q_{1_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)*$$

$$\left[\int_0^{x_1}\int_0^{y_T} q_{2_2}\lambda_{2_2}\exp(-\lambda_{2_2}(y_T-y_1))*q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)dy_1dy_T\right]dx_1dx_T$$

$$+\int_0^t\int_0^{x_T}\int_0^{y_T} q_{2_2}\lambda_{2_2}\exp(-\lambda_{2_2}(y_T-y_1))*q_{2_1}\lambda_{2_1}\exp(-\lambda_{2_1}y_1)*$$

$$\left[\int_0^{y_1} q_{2_2}\lambda_{1_2}\exp(-\lambda_{1_2}(x_T-x_1))*q_{1_1}\lambda_{1_1}\exp(-\lambda_{1_1}x_1)dx_1\right])]dy_1dy_Tdx_T$$

$$=\quad \frac{q_{1_1}q_{1_2}q_{2_1}q_{2_2}\lambda_{2_1}\lambda_{2_2}}{(\lambda_{1_2}-\lambda_{1_1})(\lambda_{2_2}-\lambda_{2_1})}\left[(\exp(-\lambda_{2_2}t)-1)\left(\frac{\lambda_{1_2}-\lambda_{1_1}}{\lambda_{2_2}}\right)+\right.$$

$$(\exp(-(\lambda_{1_1}+\lambda_{2_1})t)-1)\left(\frac{\lambda_{1_2}}{\lambda_{1_1}+\lambda_{2_1}}\right)+(\exp(-(\lambda_{1_2}+\lambda_{2_2})t)-1)*$$

$$\left(\frac{\lambda_{1_1}}{\lambda_{1_2}+\lambda_{2_2}}\right)+(\exp(-(\lambda_{2_1}+\lambda_{1_2})t)-1)\left(\frac{-\lambda_{1_1}}{\lambda_{2_1}+\lambda_{1_2}}\right)+$$

$$(\exp(-(\lambda_{1_1}+\lambda_{2_2})t)-1)\left(\frac{-\lambda_{1_2}}{\lambda_{1_1}+\lambda_{2_2}}\right)+(\exp(-(\lambda_{2_1}t)-1)*$$

$$\left.\left(\frac{-(\lambda_{1_2}-\lambda_{1_1})}{\lambda_{2_1}}\right)\right]+$$

$$\frac{q_{1_1}q_{1_2}q_{2_1}q_{2_2}\lambda_{1_1}\lambda_{1_2}}{(\lambda_{1_2}-\lambda_{1_1})(\lambda_{2_2}-\lambda_{2_1})}\left[(\exp(-\lambda_{1_2}t)-1)\left(\frac{\lambda_{2_2}-\lambda_{2_1}}{\lambda_{1_2}}\right)+\right.$$

$$(\exp(-(\lambda_{1_1}+\lambda_{2_1})t)-1)\left(\frac{\lambda_{2_2}}{\lambda_{1_1}+\lambda_{2_1}}\right)+(\exp(-(\lambda_{1_2}+\lambda_{2_2})t)-1)*$$

$$\left(\frac{\lambda_{2_1}}{\lambda_{1_2}+\lambda_{2_2}}\right)+(\exp(-(\lambda_{1_1}+\lambda_{2_2})t)-1)\left(\frac{-\lambda_{2_1}}{\lambda_{1_1}+\lambda_{2_2}}\right)+$$

$$(\exp(-(\lambda_{2_1}+\lambda_{1_2})t)-1)\left(\frac{-\lambda_{2_2}}{\lambda_{2_1}+\lambda_{1_2}}\right)+(\exp(-(\lambda_{1_1}t)-1)*$$

325

$$\left(\frac{-(\lambda_{2_2} - \lambda_{2_1})}{\lambda_{1_1}}\right)\Big]$$

$$
\begin{aligned}
P_{21}(t) &= P(E_{2_1^{(0)}} < E_{1_1^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(1)}} < t) + \\
&\quad P(E_{2_1^{(0)}} + E_{2_2^{(0)}} < E_{1_1^{(0)}} < E_{1_1^{(0)}} + E_{2_2^{(1)}} < t) + \\
&\quad P(E_{1_1^{(0)}} < E_{2_1^{(0)}} < E_{2_1^{(0)}} + E_{2_2^{(0)}} < E_{1_1^{(0)}} + E_{1_2^{(1)}} < t)
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^t \int_0^{x_T} \int_0^{y_T} f_{E_{1_2^{(1)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) \left[\int_0^{x_1} f_{E_{2_2^{(0)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1 \right] dx_1 dy_T dx_T + \\
&\quad \int_0^t \int_0^{x_T} f_{E_{1_2^{(1)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) \left[\int_0^{x_1} \int_0^{y_T} f_{E_{2_2^{(0)}}}(y_T - y_1) f_{E_{2_1^{(0)}}}(y_1) dy_1 dy_T \right] dx_1 dx_T + \\
&\quad \int_0^t \int_0^{x_T} \int_0^{y_T} f_{E_{2_2^{(0)}}}(y_T - y_1) f_{E_{1_1^{(0)}}}(y_1) \left[\int_0^{y_1} f_{E_{1_2^{(1)}}}(x_T - x_1) f_{E_{1_1^{(0)}}}(x_1) dx_1 \right] dy_1 dy_T dx_T
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^t \int_0^{x_T} \int_0^{y_T} p_{1_2} \lambda_{1_2} \exp(-\lambda_{1_2}(x_T - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) * \\
&\quad \left[\int_0^{x_1} q_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) dy_1 \right] dx_1 dy_T dx_T \\
&\quad + \int_0^t \int_0^{x_T} p_{1_2} \lambda_{1_2} \exp(-\lambda_{1_2}(x_T - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) * \\
&\quad \left[\int_0^{x_1} \int_0^{y_T} q_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) dy_1 dy_T \right] dx_1 dx_T \\
&\quad + \int_0^t \int_0^{x_T} \int_0^{y_T} q_{2_2} \lambda_{2_2} \exp(-\lambda_{2_2}(y_T - y_1)) * q_{2_1} \lambda_{2_1} \exp(-\lambda_{2_1} y_1) * \\
&\quad \left[\int_0^{y_1} p_{1_2} \lambda_{1_2} \exp(-\lambda_{1_2}(x_T - x_1)) * q_{1_1} \lambda_{1_1} \exp(-\lambda_{1_1} x_1) dx_1 \right) \right] dy_1 dy_T dx_T
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{q_{1_1} p_{1_2} q_{2_1} q_{2_2} \lambda_{1_1} \lambda_{1_2}}{(\lambda_{1_2} - \lambda_{1_1})(\lambda_{2_2} - \lambda_{2_1})} \left[(\exp(-\lambda_{1_2} t) - 1) \left(\frac{\lambda_{2_2} - \lambda_{2_1}}{\lambda_{1_2}} \right) + \right. \\
&\quad (\exp(-(\lambda_{1_1} + \lambda_{2_1}) t) - 1) \left(\frac{\lambda_{2_2}}{\lambda_{1_1} + \lambda_{2_1}} \right) + (\exp(-(\lambda_{1_2} + \lambda_{2_2}) t) - 1) * \\
&\quad \left(\frac{\lambda_{2_1}}{\lambda_{1_2} + \lambda_{2_2}} \right) + (\exp(-(\lambda_{1_1} + \lambda_{2_2}) t) - 1) \left(\frac{-\lambda_{2_1}}{\lambda_{1_1} + \lambda_{2_2}} \right) + \\
&\quad (\exp(-(\lambda_{2_1} + \lambda_{1_2}) t) - 1) \left(\frac{-\lambda_{2_2}}{\lambda_{2_1} + \lambda_{1_2}} \right) + (\exp(-(\lambda_{1_1} t) - 1) * \\
&\quad \left. \left(\frac{-(\lambda_{2_2} - \lambda_{2_1})}{\lambda_{1_1}} \right) \right]
\end{aligned}
$$

The nodes of EON correspond on a one-on-one basis with states of the underlying stochastic process; absorbing states correspond to event sequences that end with a

concluding event and transient states correspond to event sequences that end with an intermediate event.

Appendix D. Polynomial Approximation

D.1   Norms

One of the most basic problems in the numerical methods literature is the problem of approximating a known function f with a (approximation) function h. In order that h be a "good "approximation to a given function f, we require the error function, e, (i.e., $f - h$) to be "small" in some sense.

More precisely, the criterion of a "small" error function is quantified with a metric[1] known as a *norm*. A norm is a real-valued function, $||\,x\,||$, defined on a linear space X such that:

1. $||\,x\,|| \geq 0$ (Positivity);

2. $||\,x\,|| = 0$ if and only if $x = 0$ (Positive Definiteness);

3. $||\,\alpha x\,|| = |\,\alpha\,| \cdot ||\,x\,||$ for any arbitrary scalar α (Homogeneity); and

4. $||\,x + y\,|| \leq ||\,x\,|| + ||\,y\,||$, (Triangle Inequality). [108] [142]

The pair $(X, ||\cdot||)$ is referred to as a normed linear space. The norm of the error function, $||\,e\,||$ or $||\,f - h\,||$ determines the adequacy of the approximation. The best approximation, h^*, is the approximation such that the condition

$$||\,f - h^*\,|| \leq ||\,f - h\,||$$

---

[1]The framework of metric spaces provides a method of measuring the goodness of an approximation since one of the basic properties of a metric space is a distance function. Specifically, a pair (X, d) is a metric space if X is a nonempty set of objects (called points) and d is a (real-valued) function from $X \times X$ to R (called the metric of the space) satisfying the following four properties for all points x, y, z in X:

1. (Positive) $d(x, y) \geq 0$ and $d(x, x) = 0$,

2. (Strictly Positive) if $d(x, y) = 0$, then $x = y$,

3. (Symmetry) $d(x, y) = d(y, x)$, and

4. (Triangle Inequality) $d(x, y) \leq d(x, z) + d(z, y)$. [153] [8] [142]

for all h in the approximation set A. However, unless A is a compact[2] set, a best approximation may not exist. [153]

Generally, many norms can be defined on a given linear space X. Clearly, the choice of norm plays an important role in the adequacy of the approximation - what may be a good approximation when measured in one norm may be a very poor approximation in another norm. [142] [108] One of the most widely used norms in approximation theory is the L_p norm. For the normed linear space $C(-\infty, \infty)$ of continuous real-valued functions defined on the interval $(-\infty, \infty)$, the L_p norm is defined (using the error function) as:

$$||\,e\,||_p \;\;=\;\; \left[\int_{-\infty}^{\infty} |\,e(x)\,|^p \, dx \right]^{\frac{1}{p}}, \; 1 \leq p < \infty \tag{124}$$

$$||\,e\,||_\infty \;\;=\;\; \sup_{-\infty < x < \infty} |\,e(x)\,| \tag{125}$$

For the n-dimensional normed linear (vector) space R^r of real column vectors $\mathbf{x}$, the L_p norm is defined (using the error function) as:

$$||\,e\,||_{p,\mathbf{x}} \;\;=\;\; \left(\sum_{i=1}^{n} |\,e(x_i)\,|^p \right)^{\frac{1}{p}}, \; 1 \leq p < \infty \tag{126}$$

$$||\,e\,||_{\infty,\mathbf{x}} \;\;=\;\; \max_i |\,e(x_i)\,| \tag{127}$$

The norms defined in equations 126 and 127 are referred to in the literature as *semi-norms*. Semi-norms satisfy properties (1),(3), and (4) of a norm, but not necessarily property (2), because $||\,e\,|| = 0$ may not imply $e(x) = 0$. [194] [108] [58]

D.2 *Polynomial Interpolation and Approximation*

Interpolation refers to the process of approximating a function $f(x)$ with $h(x)$ using a finite number of specified data points. More precisely, given the n paired values $(x_i, f(x_i))$, $i = 1, 2, \ldots, n$, an interpolating (approximating) function $h(x)$ is

---

[2]Closed and bounded. [8]

defined such that:

$$h(x_i) = f(x_i) \qquad i = 1, 2, \ldots, n \tag{128}$$

The criterion in equation 128 minimizes the error function, e, for the L_p (semi-) norm given in equation 126. [108]

An interpolating function $h(x)$ can be used to estimate the value of f at any arbitrary point x. If x lies in the interval $[x_0, x_n]$ then the estimate is termed "interpolation", else the estimate is "extrapolation". Many different types of interpolating functions can be used (e.g., polynomials, rational functions, trigometric functions). For polynomial interpolation, there is a unique polynomial of degree $n - 1$ or less that satisfies the n constraints in equation 128. [58]

As the number of interpolation points increases, the degree of the interpolating polynomial generally must be increased to fit these specified points. However, "high" degree polynomials typically have an oscillatory nature that can lead to inaccurate approximations *within* intervals defined by the interpolation points. Additionally, for functions that change their "shape" over different intervals, a single interpolation polynomial is generally not a good approximation[3]. For these reasons, an alternative interpolation approach uses a different polynomial over each subinterval. This approach is referred to in the literature as "interpolation with piecewise polynomial functions" and will be discussed further in section 5.3. [123] [58] [71]

In many cases it may be unreasonable to require that the approximating function h *exactly fit* the function f at specified points (i.e., $h(x_i) = f(x_i)$). Two cases when an exact fit may not be desired are:

1. If some question exists as to the accuracy of the individual values $f(x_i), i = 1, 2, \ldots, n$ (e.g. $f(x_i) = f_i + \varepsilon_i$ where the errors ε_i are unknown) or

---

[3]These functions change their character too much to be approximated by a polynomial of "low" degree and using a "high" degree polynomial can cause significant approximation error within intervals due to the oscillatory nature of the high degree polynomial.

2. If a "simple" function is desired (e.g. as is our case, f is approximated by a polynomial function h that is easier to integrate).

The first case falls under the area of regression in the literature. Although this dissertation does not view approximation from a regression perspective (i.e., errors in the specified points), results from the regression literature when viewed from the second perspective are applicable.

For the two cases mentioned above, another approximation criterion is required. One popular criterion in the literature is the "least squares" criterion. The least squares criterion is the L_p norm shown in equations 124 and 126 with $p = 2$. The L_2 norm gives equal "weight" to the error function for each point in $C(-\infty, \infty)$ and R^r. For some purposes, it is appropriate to require the approximation to be better (have less error) over some parts of the interval than over other parts. To this end, weighted, generalized versions of the L_p norm defined in equations 124 and 126 are:

$$|| e ||_{p,w} \;=\; \left[\int_{-\infty}^{\infty} w(x) \, | e(x) |^p \, dx \right]^{\frac{1}{p}}, \; 1 \le p < \infty \tag{129}$$

$$|| e ||_{p,\mathbf{x},w} \;=\; \left(\sum_{i=1}^{n} w_i \, | e(x_i) |^p \right)^{\frac{1}{p}}, \; 1 \le p < \infty \tag{130}$$

respectively where $w(x)$ and w_i are *weight functions*. In equation 129, w(x) is defined such that $w(x) \ge 0$ for all x (positive function) and $\int_{-\infty}^{\infty} w(x)dx > 0$. For equation 130, the w_i are positive. If $w(x) = 1$ and $w_i = 1$, the L_p norm in equations 124 and 126 results. [108] [153] [58]

Necessary conditions to minimize the error function, e (i.e., $f - h$), for the weighted L_2 norm defined by equations 129 and 130 are:

$$\frac{\partial \, || e ||_{2,w}}{\partial c_j} = 0 \tag{131}$$

where c_j, $j = 0, \ldots, m$ is the set of parameters for h. For $C[-\infty, \infty]$ and R^r, the solution to equation 131 is respectively:

$$\int_{-\infty}^{\infty} w(x)f(x)h_j(x)dx = \int_{-\infty}^{\infty} w(x)h(x)h_j(x)dx \tag{132}$$

$$\sum_{i=1}^{r} w_i f(x_i)h_j(x_i) = \sum_{i=1}^{r} w_i h(x_i)h_j(x_i) \tag{133}$$

where $h_j = \frac{dh}{dc_j}$, $j = 0, \ldots m$. Equations 132 and 133 are referred to in the literature as *normal equations*. Both these equations contain scalar (inner) product terms. Equation 132 contains the scalar product:

$$(f_1, f_2) = \int_a^b w(x)f_1(x)f_2(x)dx \tag{134}$$

while equation 133 contains the scalar product

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=0}^{n} w_i u_i v_i \tag{135}$$

where f_1 and f_2 are arbitrary functions and $\mathbf{u}$ and $\mathbf{v}$ are arbitrary vectors. [108] [58] [153]

The smallness of $\| e \|_{2,w}$ is strongly depends on the choice of approximating function h. One class of choices are those that have the form:

$$h(x) = \sum_{j=0}^{m} c_j \varphi_j(x). \tag{136}$$

Here, h is a linear combination of basis functions $\varphi_0, \varphi_1, \ldots, \varphi_m$ selected a priori. [4] In this case, equations 132 and 133 reduce to a system of $m + 1$ linear equations for determining the $m + 1$ coefficients (unknowns) c_j - (note each of the $m + 1$ equations contains all $m + 1$ coefficients). For R^r, equation 133 is of the form:

$$(\mathbf{A}^{\mathbf{T}}\mathbf{W}\mathbf{A})\mathbf{c} = \mathbf{A}^{\mathbf{T}}\mathbf{W}\mathbf{y} \tag{137}$$

where

$$\mathbf{A} = \begin{bmatrix} \varphi_0(x_1) & \ldots & \varphi_m(x_1) \\ \vdots & & \vdots \\ \varphi_0(x_r) & \ldots & \varphi_m(x_r) \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} w_1 & & & \\ & \ddots & & \mathbf{0} \\ & \mathbf{0} & \ddots & \\ & & & w_r \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} c_0 \\ \vdots \\ c_m \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_r) \end{bmatrix}$$

When h is a polynomial function p^m of degree m (i.e., $p^m = \sum_{j=0}^{m} c_j x^j$), the set of basis functions ($\{\varphi_0, \varphi_1, \ldots, \varphi_m\}$) in equation 136 are the monic polynomials $1, x, \ldots, x^m$. For this set of basis functions the matrix $\mathbf{A}$ in equation 137 is of the form:

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 & \ldots & x_1^m \\ \vdots & & & \vdots \\ 1 & x_r & \ldots & x_r^m \end{bmatrix}$$

For large m, the matrix $\mathbf{A^T WA}$ can be ill-conditioned, making the linear system in equation 137 difficult to solve numerically. [169]

One way of reducing the effect of ill-conditioning is to introduce *orthogonality* into the linear system. Two functions (or vectors) a and b are orthogonal if $\langle a, b \rangle = 0$ (reference equations 134 and 135). Additionally, a sequence of functions (or vectors) $\phi_0, \phi_1, \phi_2, \ldots$ is an orthogonal system if $\langle \phi_i, \phi_j \rangle = 0$ for $i \neq j$ and $\langle \phi_i, \phi_i \rangle \neq 0$ for all i. If in addition $\|\phi_i\| = 1$ for all i, the sequence is called an *orthonormal* system. Functions (or vectors) in an orthogonal system are linearly independent. [71] [108] [58] [16]

When the approximating function h is composed of a finite orthogonal system of basis functions - i.e.,

$$h = \sum_{j=0}^{m} c_j \phi_j,$$

the best approximation of h is:

$$h^* = \sum_{j=0}^{m} \frac{\langle \phi_j, f \rangle}{\|\phi_j\|^2} \phi_j \tag{138}$$

where $\|\phi_j^2\| = \langle \phi_j, \phi_j \rangle$. In this case, orthogonality reduces equations 132 and 133 from a system of $m+1$ equations in $m+1$ unknowns (c_js) to $m+1$ equations, each of which contains only one unknown (one c_j). For R^r, equation 137 reduces to:

$$\mathbf{Bc} = \mathbf{A^T W y} \tag{139}$$

where

$$\mathbf{B} = \begin{bmatrix} \langle \varphi_\mathbf{0}, \varphi_\mathbf{0} \rangle & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \langle \varphi_\mathbf{m}, \varphi_\mathbf{m} \rangle \end{bmatrix} \qquad \varphi_\mathbf{j} = \begin{bmatrix} \varphi_j(x_\mathbf{1}) \\ \vdots \\ \varphi_j(x_r) \end{bmatrix}$$

The matrix $\mathbf{B}$ is diagonal and does not pose conditioning problems.

Orthogonal basis polynomials on an interval $[a, b]$ may be generated using the following recursive relation based on equation 138:

1. Let ϕ_0 be the constant function $\phi_{0(x)} = 1$, $a \leq x \leq b$.

2. Let ϕ_1 be the linear function $\phi_1(x) = (x - \alpha_0)\phi_0(x)$, $a \leq x \leq b$ where

$$\alpha_j = \frac{\langle \phi_j, x\phi_j \rangle}{\|\phi_j\|^2} \quad j \geq 0.$$

3. Let ϕ_{j+1} be defined by the three-term recurrence relation

$$\phi_{j+1} = (x - \alpha_j)\phi_j - \beta_j \phi_{j-1}(x), \quad a \leq x \leq b \text{ and } j \geq 1$$

where

$$\beta_j = \frac{\|\phi_j\|^2}{\|\phi_{j-1}\|^2}.$$

This process is referred to as the Gram-Schmidt orthogonalization process. [153] [108] [71]

Two major classes of orthogonal polynomials are distinguished in the literature: "classical" and "non-classical". The three systems of classical orthogonal polynomi-

als are: the Jacobi (Tchebysheff, Gegenbauer, and Legendre), Laguerre and Hermite. These classical polynomial systems are distinguished from non-classical orthogonal polynomials by their association with a generalized Rodrigue formula, from which the classical orthogonal polynomials may be derived by successive differentiation of functions of their weighting functions. Additionally, classical orthogonal polynomials also satisfy a Sturm-Liouville second-order differential equation. *The major advantage of the classical systems is that they possess simple recurrence relations which relate any three successive polynomials.* Recurrence formulae also exist for non-classical polynomials, but they are more difficult to obtain. However, non-classical orthogonal polynomials (with any positive weighting function $w(x) \geq 0$) may be generated on any part of the real line using the Gram-Schmidt orthogonalization process detailed above. [153]

Appendix E. Polynomial Coefficients and Interval Data

Table 27 shows the ten subintervals used in the piecewise polynomial approximation for the $exp(1.25)$ density function example shown in Section 6.2, Figure 21. Tables 28 and 29 detail respectively the monic polynomial coefficients $c_j^{(k)}$ for the $exp(1.25)$ density function example using LU Decomposition and Singular Value Decomposition for matrix inversion. Here the polynomial function in subinterval k is of the form $h_k(t) = \sum_{j=0}^{m} c_j^{(k)} t^m$ where m is the degree of the polynomial.

Tables 30 and 31 show the polynomial coefficients $c_j^{(k)}$ and orthogonal polynomials $\varphi_j^{(k)}(t)$ for the $exp(1.25)$ density function example. Here the polynomial function in subinterval k is of the form $h_k(t) = \sum_{j=0}^{m} c_j^{(k)} \varphi_j^{(k)}(t)$ where m is the number of orthogonal polynomials in the approximation.

Table 32 shows the ten subintervals used in the piecewise polynomial approximation for the truncated $N(5, 1.67)$ density function example shown in Section 6.2, Figure 22. Tables 33 and 34 detail respectively the monic polynomial coefficients $c_j^{(k)}$ for this density function example using LU Decomposition and Singular Value Decomposition for matrix inversion. Here the polynomial function in subinterval k is of the form $h_k(t) = \sum_{j=0}^{m} c_j^{(k)} t^m$ where m is the degree of the polynomial.

Tables 35 and 36 show the polynomial coefficients $c_j^{(k)}$ and orthogonal polynomials $\varphi_j^{(k)}(t)$ for the truncated $N(5, 1.67)$ density function example. Here the polynomial function in subinterval k is of the form $h_k(t) = \sum_{j=0}^{m} c_j^{(k)} \varphi_j^{(k)}(t)$ where m is the number of orthogonal polynomials in the approximation.

For the polynomial coefficients and interval data for the other examples in this dissertation please contact the author.

Subinterval Number	Subinterval
1	$[0.0000000000, 0.3684136149)$
2	$[0.3684136149, 0.7368272298)$
3	$[0.7368272298, 1.1052408446)$
4	$[1.1052408446, 1.4736544595)$
5	$[1.4736544595, 1.8420680744)$
6	$[1.8420680744, 2.2104816893)$
7	$[2.2104816893, 2.5788953042)$
8	$[2.5788953042, 2.9473089190)$
9	$[2.9473089190, 3.3157225339)$
10	$[3.3157225339, 3.6841361488]$

Table 27.   Subintervals for Piecewise Polynomial Approximations of the $exp(1.25)$ Density Function

For $m = 1$:

k	c_0^k	c_1^k
1	1.231370510296	-1.246767603180
2	1.066757481160	-0.786657176086
3	0.855939505277	-0.496347122840
4	0.655438832079	-0.313173862567
5	0.486352235305	-0.197599348686
6	0.352800130830	-0.124676760318
7	0.251583355072	-0.078665717609
8	0.177024469408	-0.049634712284
9	0.123232640605	-0.031317386257
10	0.085034368704	-0.019759934869

For $m = 2$:

k	c_0^k	c_1^k	c_2^k
1	1.249399931011	-1.536008080579	0.779142635807
2	1.212093131723	-1.331384096130	0.491605768461
3	1.116364261258	-1.068597316666	0.310182270208
4	0.979340861316	-0.818445095023	0.195711781507
5	0.824933104661	-0.607391633327	0.123485785943
6	0.672262969480	-0.440647563088	0.077914263597
7	0.533271723484	-0.314252667876	0.049160576848
8	0.413730833083	-0.221135103180	0.031018227034
9	0.315106455281	-0.153947394376	0.019571178146
10	0.236280363821	-0.106233008228	0.012348578610

For $m = 3$:

k	c_0^k	c_1^k	c_2^k	c_3^k
1	1.249983709582	-1.561185597064	0.958356925575	-0.324299082099
2	1.243893607460	-1.513905832924	0.830835492964	-0.204618887630
3	1.214363899533	-1.394040914818	0.666914754186	-0.129105791930
4	1.151448435002	-1.222802835655	0.510827946088	-0.081460245257
5	1.057076259740	-1.029952206176	0.379117751479	-0.051397941023
6	0.940474721940	-0.839315325476	0.275050037595	-0.032429916606
7	0.813079105493	-0.665776373274	0.196160096052	-0.020461884658
8	0.685224983674	-0.516529951343	0.138038073257	-0.012910591626
9	0.564645897848	-0.393400399424	0.096099550257	-0.008146041438
10	0.456204297309	-0.294991396665	0.066316116730	-0.005139869187

Table 28. Monic Polynomial Coefficients for $exp(1.25)$ Density Function (Using LU Decomposition)

338

For $m = 1$:

k	c_0^k	c_1^k
1	1.231370510296	−1.246767603180
2	1.066757481160	−0.786657176086
3	0.855939505277	−0.496347122840
4	0.655438832079	−0.313173862567
5	0.486352235305	−0.197599348686
6	0.352800130830	−0.124676760318
7	0.251583355072	−0.078665717609
8	0.177024469408	−0.049634712284
9	0.123232640605	−0.031317386257
10	0.085034368704	−0.019759934869

For $m = 2$:

k	c_0^k	c_1^k	c_2^k
1	1.249399931009	−1.536008080576	0.779142635805
2	1.212093131720	−1.331384096129	0.491605768460
3	1.116364261250	−1.068597316652	0.310182270203
4	0.979340861313	−0.818445095019	0.195711781505
5	0.824933104627	−0.607391633280	0.123485785928
6	0.672262969498	−0.440647563110	0.077914263603
7	0.533271723604	−0.314252667977	0.049160576870
8	0.413730832899	−0.221135103047	0.031018227011
9	0.315106455258	−0.153947394362	0.019571178144
10	0.236280364650	−0.106233008701	0.012348578677

For $m = 3$:

k	c_0^k	c_1^k	c_2^k	c_3^k
1	1.249983715413	−1.561185604346	0.958356930045	−0.324299083610
2	1.243893613291	−1.513905839974	0.830835496827	−0.204618888580
3	1.214363905221	−1.394040920223	0.666914756186	−0.129105792088
4	1.151448437162	−1.222802836744	0.510827944251	−0.081460244733
5	1.057076244215	−1.029952169161	0.379117730024	−0.051397936278
6	0.940474789838	−0.839315398044	0.275050069513	−0.032429921608
7	0.813078716998	−0.665775891235	0.196159897562	−0.020461856677
8	0.685225762575	−0.516530815103	0.138038387451	−0.012910628677
9	0.564642956947	−0.393397607447	0.096098656411	−0.008145946538
10	0.483352604872	−0.294995607165	0.066317320603	−0.005139983779

Table 29. Monic Polynomial Coefficients for $exp(1.25)$ Density Function (Using Singular Value Decomposition)

Polynomial coefficients:

k	$\mathbf{c_0^k}$	$\mathbf{c_1^k}$	$\mathbf{c_2^k}$	$\mathbf{c_3^k}$
1	1.003472081781	-0.115033832125	0.006609509747	-0.000253380787
2	0.633148079980	-0.072581441243	0.004170318718	-0.000159872468
3	0.399489431207	-0.045795793425	0.002631293224	-0.000100872708
4	0.252060790662	-0.028895192208	0.001660233785	-0.000063646376
5	0.159039607124	-0.018231633744	0.001047536700	-0.000040158148
6	0.100347208178	-0.011503383213	0.000660950975	-0.000025338079
7	0.063314807998	-0.007258144124	0.000417031872	-0.000015987247
8	0.039948943121	-0.004579579343	0.000263129322	-0.000010087271
9	0.025206079066	-0.002889519221	0.000166023379	-0.000006364638
10	0.015903960712	-0.001823163374	0.000104753670	-0.000004015815

Orthogonal Polynomials:

k	$\varphi_j^{(k)}(t)$	Polynomial
1	$\varphi_0^{(1)}(t)$	1.000000000000
	$\varphi_1^{(1)}(t)$	$10.857362047581t - 2.000000000000$
	$\varphi_2^{(1)}(t)$	$117.882310632259t^2 - 43.429448190325t + 2.400000000000$
	$\varphi_3^{(1)}(t)$	$1279.890925539874t^3 - 707.293863793552t^2 + 99.366577459464t$ -2.304000000000
2	$\varphi_0^{(2)}(t)$	1.000000000000
	$\varphi_1^{(2)}(t)$	$10.857362047581t - 6.000000000000$
	$\varphi_2^{(2)}(t)$	$117.882310632259t^2 - 130.288344570976t + 34.400000000000$
	$\varphi_3^{(2)}(t)$	$1279.890925539874t^3 - 2121.881591380657t^2 + 1141.673334027268t$ -198.912000000000
3	$\varphi_0^{(3)}(t)$	1.000000000000
	$\varphi_1^{(3)}(t)$	$10.857362047581t - 10.000000000000$
	$\varphi_2^{(3)}(t)$	$117.882310632259t^2 - 217.147240951626t + 98.340000000000$
	$\varphi_3^{(3)}(t)$	$1279.890925539874t^3 - 3536.469318967761t^2 + 3226.286847162877t$ -971.520000000000
4	$\varphi_0^{(4)}(t)$	1.000000000000
	$\varphi_1^{(4)}(t)$	$10.857362047581t - 14.000000000000$
	$\varphi_2^{(4)}(t)$	$117.882310632259t^2 - 304.006137332276t + 194.400000000000$
	$\varphi_3^{(4)}(t)$	$1279.890925539874t^3 - 4951.057046554864t^2 + 6353.207116866290t$ -2704.128000000000

Table 30. Orthogonal Coefficients and Polynomials for $exp(1.25)$ Density Function

Orthogonal Polynomials:

k	$\varphi_j^{(k)}(t)$	Polynomial
5	$\varphi_0^{(5)}(t)$	1.000000000000
	$\varphi_1^{(5)}(t)$	$10.857362047581t - 18.000000000000$
	$\varphi_2^{(5)}(t)$	$117.882310632259t^2 - 390.865033712927t + 322.340000000000$
	$\varphi_3^{(5)}(t)$	$1279.890925539874t^3 - 6365.644774141970t^2 + 10522.434143137510t$ -5780.736000000000
6	$\varphi_0^{(6)}(t)$	1.000000000000
	$\varphi_1^{(6)}(t)$	$10.857362047581t - 22.000000000000$
	$\varphi_2^{(6)}(t)$	$117.882310632259t^2 - 477.723930093577t + 482.340000000000$
	$\varphi_3^{(6)}(t)$	$1279.890925539874t^3 - 7780.232501729072t^2 + 15733.967925976530t$ -10585.344000000000
7	$\varphi_0^{(7)}(t)$	1.000000000000
	$\varphi_1^{(7)}(t)$	$10.857362047581t - 26.000000000000$
	$\varphi_2^{(7)}(t)$	$117.882310632259t^2 - 564.582826474228t + 674.400000000001$
	$\varphi_3^{(7)}(t)$	$1279.890925539877t^3 - 9194.820229316192t^2 + 21987.808465383390t$ -17501.952000000030
8	$\varphi_0^{(8)}(t)$	1.000000000000
	$\varphi_1^{(8)}(t)$	$10.857362047581t - 30.000000000000$
	$\varphi_2^{(8)}(t)$	$117.882310632259t^2 - 651.441722854877t + 898.399999999999$
	$\varphi_3^{(8)}(t)$	$1279.890925539872t^3 - 10609.407956903270t^2 + 29283.955761357930t$ -26914.559999999960
9	$\varphi_0^{(9)}(t)$	1.000000000000
	$\varphi_1^{(9)}(t)$	$10.857362047581t - 34.000000000000$
	$\varphi_2^{(9)}(t)$	$117.882310632259t^2 - 738.300619235527t + 1154.399999999999$
	$\varphi_3^{(9)}(t)$	$1279.890925539872t^3 - 12023.995684490370t^2 + 37622.409813900360t$ -39207.167999999940
10	$\varphi_0^{(10)}(t)$	1.000000000000
	$\varphi_1^{(10)}(t)$	$10.857362047581t - 38.000000000000$
	$\varphi_2^{(10)}(t)$	$117.882310632259t^2 - 825.195156616181t + 1442.400000000005$
	$\varphi_3^{(10)}(t)$	$1279.890925539881t^3 - 13438.583412077560t^2 + 47003.170623010900t$ -54763.776000000280

Table 31.   Orthogonal Coefficients and Polynomials for $exp(1.25)$ Density Function
            - Continued

Subinterval Number	Subinterval
1	$[0, 1)$
2	$[1, 2)$
3	$[2, 3)$
4	$[3, 4)$
5	$[4, 5)$
6	$[5, 6)$
7	$[6, 7)$
8	$[7, 8)$
9	$[8, 9)$
10	$[9, 10]$

Table 32. Subintervals for Piecewise Polynomial Approximations of the truncated $N(5, 1.67)$ Density Function

For $m = 1$:

k	c_0^k	c_1^k
1	0.001662393757	0.010609217972
2	−0.022699126009	0.033821213619
3	−0.094355289232	0.069579051148
4	−0.136217437278	0.084503329005
5	0.045530120186	0.040074139950
6	0.446271519687	−0.040074139950
7	0.708815852768	−0.084503329005
8	0.601435222247	−0.069579051148
9	0.315513010184	−0.033821213619
10	0.107754573472	−0.010609217972

For $m = 2$:

k	c_0^k	c_1^k	c_2^k
1	0.002762128135	0.003409589123	0.007198953768
2	0.011960922828	−0.014479380816	0.016099480781
3	0.010111954602	−0.015314871422	0.016977899031
4	−0.205755216033	0.124556705101	−0.005722679024
5	−0.696989535437	0.371671079632	−0.036844387740
6	−0.664717513262	0.365216675221	−0.036844387745
7	0.467543932595	−0.010103124620	−0.005722679024
8	1.554753143337	−0.324243109162	0.016977899029
9	1.477115192363	−0.307510234710	0.016099480776
10	0.756753396128	−0.147388664476	0.007198953768

For $m = 3$:

k	c_0^k	c_1^k	c_2^k	c_3^k
1	0.002662919130	0.004906744622	0.003302742374	0.002597474414
2	0.003511049833	0.003362380339	0.003890968591	0.002713002821
3	0.058898574555	−0.074973938118	0.041064550598	−0.003211553473
4	0.265129867233	−0.282927356359	0.111253950050	−0.011140631355
5	−0.062994505709	−0.053428941420	0.057893025384	−0.007017586241
6	−1.825567974940	1.000844496164	−0.152634596377	0.007017588312
7	−2.579381365578	1.400038380053	−0.222965085729	0.011140636210
8	0.204060663139	0.217149021754	−0.055282060831	0.003211553787
9	3.139233345190	−0.895082185374	0.085281004672	−0.002713000930
10	2.979480597567	−0.850204425106	0.081227028599	−0.002597476307

Table 33. Monic Polynomial Coefficients for the Truncated $N(5, 1.67)$ Density Function (Using LU Decomposition)

For $m = 1$:

k	c_0^k	c_1^k
1	0.001662393757	0.010609217972
2	−0.022699126009	0.033821213619
3	−0.094355289232	0.069579051148
4	−0.136217437278	0.084503329005
5	0.045530120186	0.040074139950
6	0.446271519687	−0.040074139950
7	0.708815852768	−0.084503329005
8	0.601435222247	−0.069579051148
9	0.315513010184	−0.033821213619
10	0.107754573472	−0.010609217972

For $m = 2$:

k	c_0^k	c_1^k	c_2^k
1	0.002762128135	0.003409589123	0.007198953769
2	0.011960922828	−0.014479380816	0.016099480781
3	0.010111954603	−0.015314871423	0.016977899031
4	−0.205755216008	0.124556705086	−0.005722679022
5	−0.696989535719	0.371671079755	−0.036844387753
6	−0.664717513438	0.365216675289	−0.036844387751
7	0.467543933561	−0.010103124926	−0.005722679001
8	1.554753148513	−0.324243110543	0.016977899121
9	1.477115189344	−0.307510234000	0.016099480734
10	0.756753388662	−0.147388662902	0.007198953685

For $m = 3$:

k	c_0^k	c_1^k	c_2^k	c_3^k
1	0.002662917901	0.004906742358	0.003302740850	0.002597473216
2	0.003511048214	0.003362378786	0.003890966797	0.002713001569
3	0.058898547226	−0.074973903430	0.041064531621	−0.003211551986
4	0.265129757788	−0.282927237334	0.111253902135	−0.011140626549
5	−0.062994383594	−0.053428975805	0.057893011590	−0.007017583946
6	−1.825570848509	1.000846068033	−0.152634895577	0.007017607493
7	−2.579397853407	1.400045905443	−0.222966240924	0.011140695550
8	0.203917763892	0.217206131027	−0.055289666798	0.003211891404
9	3.139909683244	−0.895321217987	0.085309142098	−0.002714104067
10	2.979527776498	−0.850219341831	0.081228600456	−0.002597531464

Table 34.   Monic Polynomial Coefficients for the Truncated $N(5, 1.67)$ Density Function (Using Singular Value Decomposition)

Polynomial coefficients:

k	$\mathbf{c_0^k}$	$\mathbf{c_1^k}$	$\mathbf{c_2^k}$	$\mathbf{c_3^k}$
1	0.006966558300	0.002652135296	0.000449934538	0.000040585530
2	0.028030906145	0.008454764020	0.001006217387	0.000042390661
3	0.079587261245	0.017393653132	0.001061118519	−0.000050180513
4	0.159534036516	0.021124484582	−0.000357667381	−0.000174072331
5	0.225849341552	0.010017895880	−0.002302773863	−0.000109649761
6	0.225849341552	−0.010017895880	−0.002302773863	0.000109649761
7	0.159534036516	−0.021124484582	−0.000357667381	0.000174072331
8	0.079587261245	−0.017393653132	0.001061118519	0.000050180513
9	0.028030906145	−0.008454764020	0.001006217387	−0.000042390661
10	0.006966558300	−0.002652135296	0.000449934538	−0.000040585530

Orthogonal Polynomials:

k	$\varphi_j^{(k)}(t)$	**Polynomial**
1	$\varphi_0^{(1)}(t)$	1.000000000000
	$\varphi_1^{(1)}(t)$	$4.000000000000t - 2.000000000000$
	$\varphi_2^{(1)}(t)$	$16.000000000000t^2 - 16.000000000000t + 2.444444444444$
	$\varphi_3^{(1)}(t)$	$64.000000000000t^3 - 96.000000000000t^2 + 36.888888888889t$ -2.444444444444
2	$\varphi_0^{(2)}(t)$	1.000000000000
	$\varphi_1^{(2)}(t)$	$4.000000000000t - 6.000000000000$
	$\varphi_2^{(2)}(t)$	$16.000000000000t^2 - 48.000000000000t + 34.444444444444$
	$\varphi_3^{(2)}(t)$	$64.000000000000t^3 - 288.000000000000t^2 + 420.888888888888t$ -199.333333333333
3	$\varphi_0^{(3)}(t)$	1.000000000000
	$\varphi_1^{(3)}(t)$	$4.000000000000t - 10.000000000000$
	$\varphi_2^{(3)}(t)$	$16.000000000000t^2 - 80.000000000000t + 98.444444444444$
	$\varphi_3^{(3)}(t)$	$64.000000000000t^3 - 480.000000000000t^2 + 1188.888888888880t$ -972.222222222222
4	$\varphi_0^{(4)}(t)$	1.000000000000
	$\varphi_1^{(4)}(t)$	$4.000000000000t - 14.000000000000$
	$\varphi_2^{(4)}(t)$	$16.000000000000t^2 - 112.000000000000t + 194.444444444444$
	$\varphi_3^{(4)}(t)$	$64.000000000000t^3 - 672.000000000000t^2 + 2340.888888888880t$ -2705.11111111111

Table 35. Orthogonal Coefficients and Polynomials for the Truncated $N(5, 1.67)$ Density Function

Orthogonal Polynomials:

k	$\varphi_{\mathbf{j}}^{(\mathbf{k})}(\mathbf{t})$	**Polynomial**
5	$\varphi_0^{(5)}(t)$	1.000000000000
	$\varphi_1^{(5)}(t)$	$4.000000000000t - 18.000000000000$
	$\varphi_2^{(5)}(t)$	$16.0000000000000t^2 - 144.000000000000t + 322.444444444444$
	$\varphi_3^{(5)}(t)$	$64.000000000000t^3 - 864.000000000000t^2 + 3876.888888888880t$
		-5782.000000000000
6	$\varphi_0^{(6)}(t)$	1.000000000000
	$\varphi_1^{(6)}(t)$	$4.000000000000t - 22.000000000000$
	$\varphi_2^{(6)}(t)$	$16.0000000000000t^2 - 176.000000000000t + 482.444444444444$
	$\varphi_3^{(6)}(t)$	$64.000000000000t^3 - 1056.000000000000t^2 + 5796.888888888880t$
		-10586.888888888800
7	$\varphi_0^{(7)}(t)$	1.000000000000
	$\varphi_1^{(7)}(t)$	$4.000000000000t - 26.000000000000$
	$\varphi_2^{(7)}(t)$	$16.0000000000000t^2 - 208.000000000000t + 674.444444444444$
	$\varphi_3^{(7)}(t)$	$64.000000000000t^3 - 1248.000000000000t^2 + 8100.888888888880t$
		-17503.777777777700
8	$\varphi_0^{(8)}(t)$	1.000000000000
	$\varphi_1^{(8)}(t)$	$4.000000000000t - 30.000000000000$
	$\varphi_2^{(8)}(t)$	$16.0000000000000t^2 - 240.000000000000t + 898.444444444444$
	$\varphi_3^{(8)}(t)$	$64.000000000000t^3 - 1440.000000000000t^2 + 10788.888888888800t$
		-26916.666666666600
9	$\varphi_0^{(9)}(t)$	1.000000000000
	$\varphi_1^{(9)}(t)$	$4.000000000000t - 34.000000000000$
	$\varphi_2^{(9)}(t)$	$16.0000000000000t^2 - 272.000000000000t + 1154.444444444440$
	$\varphi_3^{(9)}(t)$	$64.000000000000t^3 - 1632.000000000000t^2 + 13860.888888888800t$
		-39209.555555555500
10	$\varphi_0^{(10)}(t)$	1.000000000000
	$\varphi_1^{(10)}(t)$	$4.000000000000t - 38.000000000000$
	$\varphi_2^{(10)}(t)$	$16.0000000000000t^2 - 304.000000000000t + 1442.444444444440$
	$\varphi_3^{(10)}(t)$	$64.000000000000t^3 - 1824.000000000000t^2 + 17316.888888888800t$
		-54766.444444444400

Table 36. Orthogonal Coefficients and Polynomials for the Truncated $N(5, 1.67)$ Density Function - Continued

Bibliography

1. Adlakha, V.G. and Kulkarni, V. G. A Classified Biography of Research on Stochastic PERT Networks: (1966-1987). *INFOR*, 27:272–296, 1989.

2. Agrawal, M. K., Elmaghraby, S. E., and Herroelen, W. S. DAGEN: A Generator of Testsets for Project Activity Nets. *European Journal of Operational Research*, 90:376–382, 1996.

3. Alexopoulos, Christos. State Space Partitioning Methods for Stochastic Shortest Path Problems. *Networks*, 30:9–21, 1997.

4. Allen, Myron B. III and Isaacson, Eli L. *Numerical Analysis For Applied Science*. John Wiley & Sons, New York, 1998.

5. Ancker, C. J. Jr. *One-on-One Stochastic Duels*. Military Applications Section of the Operations Research Society of America, Arlington VA, 1982.

6. Anderson, Lowell B., Cushman, John H., Gropman, Alan L., Roske, Vincent P. Jr., and editors. SIMTAX: A Taxonomy for Warfare Simulations. Technical report, Military Operations Research Society, Alexandria VA, 1989.

7. Anker, C. J. A Proposed Foundation for a Theory of Combat. *Naval Logistics Research Quarterly*, 42:311–343, 1995.

8. Apostol, Tom M. *Mathematical Analysis*. Addison-Wesley Publishing Company, Reading MA., 1974.

9. ASC/XR. BRAWLER / THUNDER Calibration Methodology. Internal Report, 1995. Unclassified Version.

10. Augustin, R. and Buscher, K. Characteristics of the COX-Distribution. *ACM Sigmetrics Performance Evaluation Review*, 12(1):22–32, Winter 1982-1983.

11. Badinelli, Ralph D. Approximating Probability Density Functions and Their Convolutions Using Orthogonal Polynomials. *European Journal of Operational Research*, 95:221–230, 1996.

12. Barton, David K., et al., editors. *Radar Evaluation Handbook*. Artech House, Boston MA, 1991.

13. Bein, Wolfgang W., Kamburowski, Jerzy, and Stallmann, Matthias F. M. Optimal Reduction of Two-Terminal Directed Acyclic Graphs. *SIAM Journal on Computing*, 21(6):1112–1129, 1992.

14. Berkowitz, D. and Dresher, M. A Game-Theory Analysis of Tactical Air War. *Operations Research*, 7, 1959.

15. Berzuini, C., Bellazzi, R., and Quaglini, S. Temporal Reasoning With Probabilities. In *Fifth Workshop on Uncertainty and Artificial Intelligence*, pages 14–21, 1989.

16. Bjorck, Ake. *Numerical Methods for Least Squares Problems.* Society for Industrial and Applied Mathematics, Philadelphia, 1996.

17. Bogusch, Roy J. A Model to Investigate Losses and Exchange Ratios in Large Scale Air-to-Air Engagements. Master's thesis, AFIT/GST/OS/83M-1, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 1983.

18. Bollobâs, Bêla. *Graph Theory: An Introductory Course.* Springer-Verlag, New York, 1979.

19. Bonanni, Pete. *The Art of the Kill.* Spectrum Holobyte, Alameda CA, 1993.

20. Bonsignore, Ezio and Bustin, Ian. Air-to-Air Missiles: The Battle Begins. *Military Technology*, May 1994.

21. Bowman, R.A. Efficient Estimation of Arc Criticalities in Stochastic Networks. *Management Science*, 41(1):58–67, 1995.

22. Jim Brady. Proceedings of the Military Operations Research Society Symposium. Technical Report MORS-99, S3I, Montrey CA, 1999.

23. Brassey's. World Aircraft and Systems Directory, 1996/97. Brassey's (UK) Ltd (London), 1996.

24. Burt, John M. and Garman, Mark B. Conditional Monte Carlo: Simulation Technique for Stochastic Network Analysis. *Management Science*, 18(3):207–217, 1971.

25. Busacker, Robert G. and Saaty, Thomas L. *Finite Graphs and Networks.* McGraw-Hill, New York, 1965.

26. Buse, A. and Lim, L. Cubic Splines as a Special Case of Restricted Least Squares. *Journal of the American Statistical Association*, 72(401):64–68, 1977.

27. Bussert, James C. Chinese Modernization Aids Air Defense Radar, Missiles. *Signal*, July 1994.

28. Bux, W. and Herzog, U. The Phase Concept: Approximations of Measured Data and Performance Analysis. In *Computer Performance*, pages 23–38. North-Holland, 1977.

29. CACI Products Company, Arlington, VA. *THUNDER Analyst Manual*, version 6.3 edition, June 1995.

30. Chapa, Capt Mike. F-15C ABC's. F-15C Instructor Pilot and F-16C Aggressor Pilot, November 1997. Personnel Interview.

31. Chavez, R.M. and Cooper, G.F. An Empirical Evaluation of a Randomized Algorithm for Probabilistic Influence. In Henrion, M., Shachter, R. D., Kanal, L. N., and Lemmer, J.F., editors, *Uncertainty in Artificial Intelligence*, chapter 5, pages 197–207. Elsevier, Amsterdam, 1990.

32. Ciardo, Gianfranco, German, Reinhard, and Linderman, Christoph. A Characterization of the Stochastic Process Underlying a Stochastic Petri Net. *IEEE Transactions on Software Engineering*, 20(7):506–515, 1994.

33. Clancy, Tom. *Fighter Wing*. Berkley Books, New York, 1995.

34. Cléroux, Robert and McConaloque, Denis J. A Numerical Algorithm for Recursively-Defined Convolution Integrals Involving Distribution Functions. *Management Science*, 22(10):1138–1146, 1976.

35. Coniglio, Sergio. Air-to-Air Combat in the Stealth Era. *Military Technology*, April 1995.

36. Coniglio, Sergio. The Sukhoi Su-27 Combat Aircraft Family. *Military Technology*, Special Supplement 1995.

37. Conte, S.D. and de Boor, Carl. *Elementary Numerical Analysis: An Algorithmic Approach*. McGraw-Hill Book Company, New York, 1980.

38. Cook, Nick. Russian Missiles Hot Up Western Market. *Jane's Defense Weekly*, pages 27–30, Nov 6 1996.

39. Cooper, Gregory F. The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks. *Artificial Intelligence*, 42:393–405, 1990.

40. Coppersmith, D. and Winograd, S. Matrix Multiplication via Arithmetic Progression. In *19th Annual ACM Syposium on Theory of Computing*, pages 1–6, 1987.

41. Cox, D. R. and Miller, H. D. *The Theory of Stochastic Processes*. John Wiley & Sons, New York, 1965.

42. Cox, D.R. The Use of Complex Probability in the Theory of Stochastic Processes. *Proc. Cambridge Philosophical Soc.*, 51:313–319, 1955.

43. Cox, D.R. and Smith, W.L. The Superposition of Renewal Processes. *Biometrika*, 41:91–99, 1954.

44. Coyne, James P. Fighting in Fours. *Air Force Magazine*, April 1993.

45. Cubaud, Pierre H. Performance Evaluation of non-Markovian Stochastic Event Graphs. In *Sixth International Workshop on Petri Nets and Performance*, pages 94–102, 1995.

46. Cubuad, P. *Modelès et Outils Pour la Prédiction de Performance des Systèmes Informatiques Parallèles*. PhD thesis, Université René Descartes, 1992.

47. Paul K. Davis and Richard Hillestad. Proceedings of Conference on Variable Resolution. Technical Report CF-103, RAND, Santa Monica CA, 1993.

48. deBoor, Carl. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.

49. Decision-Science Applications, Washington, D.C. *The Brawler Combat Simulation User Manual*, reversion 6.2 edition, October 1995.

50. Denhard, David R. Estimation of the Captive-Carry Survival Function for the Advanced Medium Range Air-to-Air Missile (AMRAAM). Master's thesis, AFIT/GOR/ENC/95M-01, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 1995.

51. Dodin, Bajis. On Estimating the Probability Distribution Functions in PERT Type Networks. Technical Report OR Report No. 153 (Revised), OR Program, North Carolina State University, 1980.

52. Dodin, Bajis. Determining the K Most Critical Paths in PERT Networks. *Operations Research*, 32:859–877, 1984.

53. Dodin, Bajis. Approximating the Distribution Functions in Stochastic Networks. *Computers and Operations Research*, 12(3):251–264, 1985.

54. Dodin, Bajis. Bounding the Project Completion Time Distribution in PERT Networks. *Operations Research*, 33(4):862–881, 1985.

55. Dodin, Bajis. Reducibility of Stochastic Networks. *OMEGA*, 13(5):223–232, 1985.

56. Duffin, R. J. Topology of Series-Parallel Networks. *Journal of Mathematical Analysis and Applications.*, 10:303–318, 1965.

57. Editors of Time-Life Books. *Air Combat.* Time-Life Books, Alexandria VA, 1991.

58. Eldén, Lars and Wittmeyer-Koch, Linde. *Numerical Analysis: An Introduction.* Academic Press, Inc., Boston, 1990.

59. Elmaghraby, S. E. An Algebra for the Analysis of Generalized Activity Networks. *Management Science*, 10(3):494–514, 1964.

60. Elmaghraby, S. E. On Generalized Activity Networks. *Journal of Industrial Engineering*, 17(11):621–631, 1966.

61. Elmaghraby, S. E. System Modeling: Petri Nets and Activity Networks in Juxaposition. In *1992 IEEE International Conference of Systems, Man, and Cybernetics*, pages Vol 1: 853–860, 1992.

62. Elmaghraby, S. E. Activity Nets: A Guided Tour Through Some Recent Developments. *European Journal of Operational Research*, 82:383–408, 1995.

63. Elmaghraby, S. E. and Herroelen, Willy S. On the Measurement of Complexity in Activity Networks. *European Journal of Operational Research*, 5:223–234, 1980.

64. Elmaghraby, S. E. and Kamburowski, Jerzy. The Analysis of Activity Networks Under Generalized Precedence Relations (GPRs). *Management Science*, 38(9):1245–1263, 1992.

65. Elmaghraby, S.E. *Activity Networks: Project Planning and Control by Network Methods.* John Wiley & Sons, New York, 1977.

66. Elmaghraby, S.E., Baxter, E.I., and Vouk, M.A. An Approach to the Modeling and Analysis of Software Production Processes. *International Transactions in Operational Research*, 2(1):117–35, 1995.

67. Elsman, M. B. Group Captain. *Air Defence*. Brassey's, UK, 1989.

68. Elwood, Maj Joe and Van Orne, Maj Skip. A Pyrrhic Victory or How Many Missiles Do I Shoot. *USAF Fighter Weapons Review*, Winter 1986.

69. Encyclopedia Americana. Electromagnetic Radiation. *The Encyclopedia Americana International Edition*, 10:157, 1994.

70. Evans, G.W. and Mollaghasemi, M. and Russell, E.C. and Biles, W.E., editor. *Families of Models That Cross Levels of Resolution: Issues for Design, Calibration and Management.* Proceedings of the 1993 Winter Simulation Conference, 1993.

71. Faires, J. Douglas and Burden, Richard L. *Numerical Methods*. PWS Publishing Company, Boston, 1993.

72. Feigin, P.D., Pinkas, O., and Shinar, J. A Simple Markov Model for the Analysis of Multiple Air Combat. *Naval Research Logistics Quarterly*, 31:413–429, 1984.

73. Fergeson, Capt Jerry A. and Shortell, Patallen H. Polygonal Approximation and Reduction of Activity Resource Consumption Distributions in Program Management Networks. Master's thesis, AFIT/LSSR 18-78A, School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, June 1978.

74. Fishman, George S. and Moore, Louis R. TAC THUNDER Technical Development Papers (Restricted Use). Technical report, CACI Inc. - Federal, Arlington VA, 1995.

75. Fiske, B.A. American Naval Policy. In *Proceedings of the U.S. Naval Institute*, pages 1–80. 113, 1905.

76. Ford, L. Jr and Fulkerson, D. *Flows in Networks*. Princeton University Press, Princeton, NJ, 1962.

77. Forecast International. Missile Forecast. Forecast International/DMS, 1996.

78. Forecast International. Radar Forecast. Forecast International (Newton, Conn), 1997.

79. Fowler, Bruce W. De Physica Belli: An Introduction to Lanchestrial Attrition Mechanics. Technical Report SOAR 95-01, Defense Modeling, Simulation and Tactical Information Center (DMSTTIAC), Huntsville AL, 1995.

80. Franks, Norman. *Aircraft Versus Aircraft*. Macmillian Publishing Company, New York, 1986.

81. Frick, Roy Kimble. Interaction of Forces as Discrete Processes With Application to Air Battle Analysis. Technical Report AD 720-237, Aeronautical Systems Division, Wright-Patterson AFB OH, 1970.

82. Fung, R. and Chang, K.-C. Weighing and Integrating Evidence For Stochastic Simulation in Bayesian Networks. In Henrion, M., Shachter, R. D., Kanal, L. N., and Lemmer, J.F., editors, *Uncertainity in Artificial Intelligence*, chapter 5, pages 209–219. Elsevier, Amsterdam, 1990.

83. Gafarian, A.V. and Ancker, C. J. Jr. The Two-on-One Stochastic Duels. *Naval Logistics Research Quarterly*, 31:309–324, 1984.

84. Gafarian, A.V. and Manion, K.R. Some Two-on-Two Stochastic Duels. *Naval Logistics Research Quarterly*, 36:721–764, 1989.

85. Garman, Mark B. More on Conditioned Sampling in the Simulation of Stochastic Networks. *Management Science*, 19(1):90–95, 1971.

86. Gelenbe, E. and Pujolle, G. *Introduction to Queueing Networks*. John Wiley & Sons, Chichester, UK, 1987.

87. German, R. SPNL: Processes as Language-Oriented Building Blocks of Stochastic Petri Nets. In 9_{th} *International Conference on Computer Performance Evaluation, Modeling Techniques, and Tools*, pages 123–134, 1997.

88. Gething, Michael J. *Modern Fighting Aircraft: F-15 Eagle*. Arco Publishing, Inc., New York, 1983.

89. Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. Software for the Gibbs Sampler. Technical report, Medical Research Council, Biostatistics Unit, Cambridge, UK, 1992.

90. Haddad, S., Moreaux, P., and Chiola, G. Efficient Handling of Phase-Type Distributions in Generalized Stochastic Petri Nets. In 18_{th} *International Conference on Application and Theory of Petri Nets*, pages 175–194, 1997.

91. Hagstrom, Jane N. Computational Complexity of PERT Problems. *Networks*, 18:139–147, 1988.

92. Hahn, G.J. and Shapiro, S.S. *Statistical Methods in Engineering*. Wiley, New York, 1968.

93. Hallon, Richard P. *Storm Over Iraq*. Smithsonian Institution Press, Washington, 1992.

94. Hayes, J.G. Numerical methods for Curve and Surface Fitting. *Journal of the Institute of Mathematics Applied*, 10:144–152, 1974.

95. Heckerman, D. Probabilistic Similarity Networks. *Networks*, 20(5):607–636, 1990.

96. Helmbold, Robert L. and Rehm, Allan S. "The Influence of the Numerical Strength of Engaged Forces on Their Casualties" by M. Osipov. *Naval Logistics Research Quarterly*, 42(3):435–490, 1995.

97. Henrion, M. Propagating Uncertainity in Bayesian Networks by Probabilistic Logic Sampling. In Lemmer, J.F. and Kanal, L. N., editors, *Uncertainity in Artificial Intelligence*, chapter 2, pages 149–163. North-Holland, Amsterdam, 1988.

98. Hermanns, H., Herzog, U., Mertsiotakis, V., and Rettelbach, M. Exploiting Stochastic Process Algebra Achievements fof Generalized Stochastic Petri Nets. In *Seventh International Workshop on Petri Nets and Performance Models*, pages 183–192, 1997.

99. Hillestad, Richard, Moore, Louis, and Bennett, Bart. Modeling for Campaign Analysis: Lessons for the Next Generations of Models (Draft Report). Technical Report DRR-1088-AF, RAND, Santa Monica CA, 1995.

100. Hillestad, Richard J. and Moore, Louis. The Theater-Level Campaign Model: A Research Prototype for a New Generation of Combat Analysis Model. Technical Report MR-388-AF/A, RAND, Santa Monica CA, 1996.

101. Holmes, Maj James A. *The Counterair Campaign: A Short Guide to Air Superiority for Joint Force Commanders*. Air University Press, Maxwell AFB AL, 1993.

102. Hong, Y. G. *Some Extensions of Stochastic Square Law Combat Models and Approximations*. PhD thesis, University of Southern California, 1989.

103. Hong, Yoon Gee, Yang, Shin Hyuck, and Kim, Young Hui. Development of Stochastic Aerial Combat Models. Unpublished, 1997.

104. Houck, Michael R., Whitaker, Leslie A., and Kendall, Robert R. An Information Processing Classification of Beyond-Visual-Range Air Intercepts. Technical Report AL/HR-TR-1993-0061 (DTIC AD-A266 927), Human Resources Directoriate: Aircrew Training Research Division, William AFB AZ, 1993.

105. Hrycej, T. Gibbs Sampling in Bayesian Networks. *Artificial Intelligence*, 46:351–363, 1990.

106. Hughes, Wayne P., editor. *Military Modeling (Second Edition)*. Military Operations Research Society, Alexandria VA, 1989.

107. Ilhan, Mustafa. An Analysis of Escort Formations . Master's thesis, AFIT/GOR/ENS/92M-16, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 1992.

108. Isaacson, Eugene and Keller, Herbert Bishop. *Analysis of Numerical Methods*. John Wiley & Sons, New York, 1966.

109. Hartman J. Lecture Notes in Aggregated Combat Modeling. Course Notes, Naval Postgraduate School, 1985.

110. Hartman J. Lecture Notes in High Resolution Combat Modeling. Course Notes, Naval Postgraduate School Edited by LTC Jack Kloeber Jr. and Lt Col Jack Jackson, 1996.

111. Jane's. Radar and Electronic Warfare Systems 1996/97. Jane's Information Group Limited (UK), 1996.

112. Jane's. Air-Launched Weapons. Jane's Information Group Limited (UK), 1997.

113. Jane's. All the World's Aircraft 1997/98. Jane's Information Group Limited (UK), 1997.

114. Jenzarli, Ali. *Modeling Dependence in Project Management.* PhD thesis, University of Kansas, 1995.

115. Kaimann, R.A. Coefficient of Network Complexity. *Management Science,* 21(2):172–177, 1974.

116. Kamburowski, J., Michael, D.J., and Stallman, M. Optimal Construction of Project Activity Networks. In *1992 Annual Meeting of the Decision Science Institute,* pages 1424–1426, 1992.

117. Kant, K. *Introduction to Computer System Performance Evaluation.* McGraw-Hill, Inc., New York, 1992.

118. Kennedy, Christopher Alan. *Groundwater Risk Assessment Using Probability Density Functions Approximated by Non-Classical Orthogonal Polynomials.* PhD thesis, University of Waterloo, Canada, 1996.

119. Kim, J.H. and Pearl, J. A Computation Model For Causal and Diagnostic Reasoning in Influence Systems. In *IJCAI,* 1983.

120. Kress, M. The Many-on-One Stochastic Duel. *Naval Logistics Research Quarterly,* 34:713–720, 1987.

121. Kulkarni, V.G. and Adlakha. Markov and Markov-Regenerative PERT Networks. *Operations Research,* 34(5):769–781, 1986.

122. Laming, Tim and Flack, Jeremy. *80 Years of Air to Air Combat.* Chartwell Books, Inc., London UK, 1996.

123. Lancastar, Peter and Šalkauskas, Kęstutis. *Curve and Surface Fitting: An Introduction.* Academic Press, London UK, 1986.

124. Lanchester, F.W. *Aircraft in Warfare: The Dawn of the Fourth Arm.* Constable and Co., Ltd., London UK, 1916.

125. Latchaw, John H. A Lanchester Model for Air Battles. Master's thesis, AFIT/GSA/SM/72-8, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, February 1972.

126. Law, Averill M. and Kelton, W. David. *Simulation Modeling and Analysis.* McGraw-Hill, Inc., New York, 1991.

127. Lawrence, Frederick P. and Cochran, Jeffery K. A New Computational Approach for Program Management Networks. In 17$_{th}$ *International Conference on Computers and Industrial Engineering*, pages 339–343, 1995.

128. Lawrence, Frederick Peter. *Analytic Solution of Stochastic Project Management Networks by the Method of Polygonal Approximation*. PhD thesis, Arizona State University, 1994.

129. Lee, Gang-Soo and Murata, Tadao. A β-Distributed Stochastic Petri Net Model for Software Project Time/Cost Management. *Journal of Systems Software*, 26:149–165, 1994.

130. Lovejoy, Capt Mark. F-15E Heinkein Talks. F-15E Weapon System Officer and Fighter Weapons Graduate, August 1997. Personnel Interview.

131. Macfadzean, Robert H. M. *Surface-Based Air Defense System Analysis*. Artech House, Norwood MA, 1992.

132. Mahjooli, Hashem. *A Spline-Based Algorithm for Approximation of the n-Fold Convolution of the Probability Density Functions*. PhD thesis, Texas A & M University, 1984.

133. Malhotra, Manish and Trivedi, Kishor S. Dependability Modeling Using Perti Nets. *IEEE Transactions on Reliability*, 44(3):428–440, 1995.

134. Martin, J. J. Distribution of the Time Through A Directed, Acyclic Network. *Operations Research*, 13(1):46–66, 1965.

135. McConaloque, Denis J. Convolution Integrals Involving Probability Distribution Functions (Algorithm 102). *Computer Journal*, 21(3):270–272, 1978.

136. McConaloque, Denis J. Numerical Treatment of Convolution Integrals Involving Distributions with Densities Having Singularities at the Origin. *Communications in Statistics: Simulation and Computations*, B10(3):265–280, 1981.

137. McGraw-Hill Encyclopedia. Electromagnetic Radiation. *McGraw-Hill Encyclopedia of Science and Technology*, 6:185, 1997.

138. Mehrota, Kishan, Chai, John, and Pillutla, Sharma. A Study of Approximating the Moments of the Job Completion Time in PERT Networks. *Journal of Operations Management*, 14:277–289, 1996.

139. Mets, David R. Checking Six Is Not Enough: The Evolution and Future of Air Superiority Armament. Technical Report AU-ARI-CPSS-91-14, CADRE, Maxwell AFB AL, 1992.

140. Military Operations Research Society. *Warfare Modeling*. John Wiley & Sons, Danvers, MA, 1995.

141. Murata, Tadao. Petri Nets: Properties, Analysis and Applications. *Proceedings of the IEEE*, 77(4):541–580, 1989.

142. Naylor, Arch W. and Sell, George R. *Linear Operator Theory in Engineering and Science.* Springer-Verlag, New York, 1982.

143. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models.* The Johns Hopkins University Press, Baltimore MD, 1981.

144. O'Grady, Capt Scott with Jeff Coplin. *Return with Honor.* Doubleday, New York, 1995.

145. Osipov, M. The Influence of the Numerical Strength of Engaged Sides on their Casualties. *Voenniy Sbornik*, (6-10):59–74, 25–36, 310–340, 25–37, 93–96, 1915.

146. Parkhideh, S. *General Solutions to Some Hetergeneous and Homogeneous Stochastic Combat.* PhD thesis, University of Southern California, 1990.

147. Parkhideh, S. and Gafarian, A.V. General Solution to Many-on-Many Heterogeneous Stochastic Combat. *Naval Logistics Research Quarterly*, 43:937–953, 1996.

148. Pearl, J. Addendum to Evidential Reasoning Using Stochastic Simulation of Causal Models. *Artiificial Intelligence*, 33:131, 1987.

149. Pearl, J. Evidential Reasoning Using Stochastic Simulation of Causal Models. *Artiificial Intelligence*, 32:245–257, 1987.

150. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Influence.* Morgan Kaufman, San Mateo CA, 1988.

151. Pierrot, Lane. *A Look at Tomorrow's Tactical Air Forces.* Congress of the United States, Congressional Budget Office, Washington DC, 1997.

152. Poirer, Dale J. Piecewise Regression Using Cubic Splines. *Journal of the American Statistical Association*, 68(343):515–524, 1973.

153. Powell, M.J.D. *Approximation Theory and Methods.* Cambridge University Press, London UK, 1981.

154. Press, William H., Vetterling, William T., Teukolsky, Saul A., and Flannery, Brian P. *Numerical Recipes in Fortran 77, Second Edition.* Cambridge University Press, New York, 1992.

155. Press, William H., Vetterling, William T., Teukolsky, Saul A., and Flannery, Brian P. *Numerical Recipes in Fortran 90.* Cambridge University Press, New York, 1996.

156. Pritsker, A.A.B. *Modeling and Analysis Using Q-GERT Networks.* Halsted Press, New York, 1978.

157. Pritsker, A.A.B and Happ, W.W. GERT: Graphical Evaluation and Review Technique, Part I. Fundamentals. *Journal of Industrial Engineering*, 17(5):267–74, 1966.

158. Pritsker, A.A.B, O'Reilly, Jean J., and LaVal, David K. *Simulation with Visual SLAM and AweSim.* John Wiley & Sons, New York, 1997.

159. Pritsker, A.A.B and Whitehouse, G. E. GERT: Graphical Evaluation and Review Technique, Part II. Applications. *Journal of Industrial Engineering*, 17(5):293–301, 1966.

160. Pugh, Capt David M. A Validation Assessment of the STORM Air-to-Air Prototype Algorithm. Master's thesis, AFIT/GOA/ENS/00M-06, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2000.

161. Quichaud, D. and Chretienne, P. Performance Evaluation of Interpreted Biploar Synchronization Schemes Using GERT. In *Advances in Petri Nets Theory*, pages 343–370. Springer-Verlag, 1998.

162. Reyck, Bert Deand Herroelen, Willy. On the Use of the Complexity Index as a Measure of Complexity in Activity Networks. *European Journal of Operations Research*, 91:347–366, 1996.

163. Ross, Sheldon M. *Stochastic Processes*. John Wiley & Sons, New York, 1983.

164. Sahner, Robin A. *A Hybrid, Combinatorial-Markov Method of Solving Performance and Reliability Models*. PhD thesis, Duke University, 1986.

165. Sahner, Robin A. and Trivedi, Kishor S. Performance and Reliability Analysis Using Directed Acyclic Graphs. *IEEE Transactions on Software Engineering*, 13(10):1105–1114, 1987.

166. Santoro, E. Some Probabilistic Properties of B-Splines and an Application to Dimensional Tolerance. *European Journal of Operational Research*, 24(11):221–230, 1992.

167. Schmeiser, B.W. and Deutsch, S. J. A Versatile Four Parameter Family of Probability Distribution Suitable For Simulation. *AIIE Transactions*, 9(2):176–182, 1977.

168. Schumaker, Larry L. *Spline Functions: Basic Theory*. John Wiley & Sons, New York, 1981.

169. Seber, G.A.F. *Linear Regression Analysis*. John Wiley & Sons, New York, 1977.

170. Seber, G.A.F. and Wild, C.J. *Nonlinear Regression*. John Wiley & Sons, New York, 1989.

171. Shachter, R.D. Evaluating Influence Diagrams . *Operations Research*, 34(6):871–882, 1986.

172. Shachter, R.D. and Peot, M.A. Simulation Approaches to General Probabilistic Influence on Belief Networks. In Henrion, M., Shachter, R. D., Kanal, L. N., and Lemmer, J.F., editors, *Uncertainty in Artificial Intelligence*, chapter 5, pages 221–231. Elsevier, Amsterdam, 1990.

173. Shaw, Robert L. *Fighter Combat: Tactics and Manuevering.* Naval Institute Press, Annapolis MD, 1985.

174. Shimony, Solomon Eyal. Finding MAPs for Belief Networks is NP-hard. *Artifical Intelligence*, 68:399–410, 1994.

175. Sigal, C. E., Pritsker, A.A.B., and Solberg, J.J. The Use of Cutsets in Monte Carlo Analysis of Stochastic Networks. *Mathematics and Computers in Simulation*, 21:376–384, 1979.

176. Sinclair, James D. BRAWLER / THUNDER Calibration Summary. Internal Memo, 1995.

177. Singleton, CAPT Jerry . Say It in Pilot Talk. *Proceedings of the United States Naval Institute*, 122(2):39–41, Feb 1996.

178. Skolnik, Merrill (editor). *Radar Handbook.* McGraw-Hill, Inc., New York, 1990.

179. Somarajan, C. and Lau, Hon-Shiang. Stochastic Activity Networks: A Four-Parameter Approach for Completion Times. *International Journal of Systems Science*, 25(10):1607–1619, 1994.

180. Soo, Lee Gang and Jung-Mo, Yoon. An Empirical Study on the Complexity Metrics of Petri Nets. *Microelectronics and Reliability*, 32(3):323–329, 1990.

181. Stuart, Alan and Ord, J. Keith. *Kendall'a Advanced Theory of Statistics (Volume 1): Distribution Theory.* Edward Arnold, London UK, 1994.

182. Swamy, M. N. S and Thulasiraman, K. *Graphs, Networks, and Algorithms.* John Wiley & Sons, New York, 1981.

183. Taylor, James G. *Force-On-Force Attrition Modeling.* Military Applications Section of the Operations Research Society of America, Arlington VA, 1980.

184. Taylor, James G. *Lanchester Models of Warfare.* Military Applications Section of the Operations Research Society of America, Arlington VA, 1983.

185. Thornborough, Anthony. *Modern Fighter Aircraft: Technology and Tactics.* Patrick Stephens Limited, England, 1995.

186. Trogemann, Georg and Gente, Matthias. Performance Analysis of Parallel Programs Based on Directed Acyclic Graphs. *Acta Informatica*, 34:411–428, 1997.

187. Unser, Michael and Daubechies, Ingrid. On the Approximation Power of Convolution-Based Least Squares Versus Interpolation. *IEEE Transactions on Signal Processing*, 45(7):1697–1711, 1997.

188. USAF Scientific Advisory Board. *New World Vista's Attack Volume: Air and Space Power for the 21st Century: Attack Volume.* USAF Scientific Advisory Board, Washington DC, 1995.

189. USAF Scientific Advisory Board. *New World Vista's Attack Volume: Air and Space Power for the 21st Century: Munition Volume.* USAF Scientific Advisory Board, Washington DC, 1995.

190. Valdes, Jacoboand Tarjan, Robert E. and Lawler, Eugene L. The Recognition of Series Parallel Digraphs. *SIAM Journal on Computing*, 11(2):298–313, 1982.

191. Van Slyke, Richard M. Monte Carlo Methods and the PERT Problem. *Operations Research*, 11(5):839–860, 1963.

192. Walker, Air Vice Marshal John R . *Air Superiority Operations.* Brassey's, UK, 1989.

193. Warden, Col(ret) John A. *The Air Campaign: Planning for Combat.* Brassey's, Washington DC, 1989.

194. Watson, G. A. *Approximation Theory and Numerical Methods.* John Wiley & Sons, Chichester, 1980.

195. Weast, Robert C. and Astle, Melvin J. editors. *Handbook of Chemistry and Physics, 63rd edition.* CRC Press, Inc., Boca Raton FL, 1982.

196. Whitehouse, Gary E. *Systems Analysis and Design Using Network Techniques.* Prentice Hall, Englewood Cliffs NJ, 1973.

197. Yang, J. *A Fast Approximation of Homogeneous Stochastic Combat.* PhD thesis, University of Southern California, 1992.

198. Yang, J. and Gafarian, A.V. A Fast Approximation of Homogeneous Stochastic Combat. *Naval Logistics Research Quarterly*, 42:503–533, 1995.

199. Yiannoutsos, C. T. and Gelfand, A. E. Simulation Approaches for Calculations in Directed Graphical Models. Technical report, Department of Statistics, University of Connecticut, 1992.

Vita

Major Denhard is currently a joint air operations analyst at USCENTCOM's Combat Analysis Division at MacDill AFB. Upon completing a Bachelor of Science degree in mechanical engineering from Carnegie Mellon, Major Denhard was commissioned in the United States Air Force through the Reserve Officer Training Corp program in May 1988. Assigned to the Training System Program Office at the Aeronautical Systems Center in January 1989, Major Denhard managed several aircraft simulator programs including the F-15 and F-16 unit training device and F-15, B-1, and B-2 improved visual system upgrade. In March 1995, Major Denhard earned a Master of Science in Operations Research at the Air Force Institute of Technology (AFIT).

Upon completion of his Masters, Major Denhard was selected to follow-on in the Department of Operational Sciences' PhD program. Major Denhard arrived at the Air Force Studies and Analysis Agency (AFSAA) on the Air Staff in 1998 and was assigned to the Air and Space Superiority branches. While at AFSAA, Major Denhard conducted numerous studies including F-22 and JSF air to air and air to ground mission effectiveness as well as assessments of strategic missile defense and space surveillance capabilities.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From – To)
08-24-2001	**Dissertation**	May 1998 – Sep 2001

4. TITLE AND SUBTITLE

THEATER-LEVEL STOCHASTIC AIR-TO-AIR ENGAGEMENT MODELING VIA EVENT OCCURRENCE NETWORKS USING PIECEWISE POLYNOMIAL APPROXIMATION

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

D. R. Denhard, Major, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 P Street, Building 640
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/DS/ENS/01-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Studies and Analyses Agency
1570 Air Force Pentagon   Maj Craig Knierem
Washington DC 20330-1570   craig.knierem.@pentagon.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This dissertation investigates a stochastic network formulation termed an event occurrence network (EON). EONs are graphical representations of the superposition of several terminating counting processes. An EON arc represents the occurrence of an event from a group of (sequential) events before the occurrence of events from other event groupings. Events between groups occur independently, but events within a group occur sequentially. A set of arcs leaving a node is a set of competing events, which are probabilistically resolved by order relations.

An important EON metric is the probability of being at a particular node or set of nodes at time t. Such a probability is formulated as an integral expression (generally a multiple integral expression) involving event probability density functions. This integral expression involves several stochastic operators: subtraction; multiplication; convolution, and integration. For the EON probability metric, simulation is generally computationally costly to obtain accurate estimates for large EONs, transient nodes, or "rare" states. Instead, using research with probabilistic activity networks, a numerical approximation technique using piecewise polynomial functions is developed. The dissertation's application area is air-to-air combat modeling.

15. SUBJECT TERMS

Air Combat, Event Occurrence Networks, Probabilistic Networks, Activity Networks, Piecewise Polynomial Approximation, Simulation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Raymond R. Hill, Lt Col, USAF (AFIT/ENS)
U	U	U	UU	379	19b. TELEPHONE NUMBER (Include area code) (937) 255-6565, ext 4323; e-mail: raymond.hill@afit.edu