Multitarget Bayes Filtering via First-Order Multitarget Moments

RONALD P. S. MAHLER
Lockheed Martin

The theoretically optimal approach to multisensor-multitarget detection, tracking, and identification is a suitable generalization of the recursive Bayes nonlinear filter. Even in single-target problems, this optimal filter is so computationally challenging that it must usually be approximated. Consequently, multitarget Bayes filtering will never be of practical interest without the development of drastic but principled approximation strategies. In single-target problems, the computationally fastest approximate filtering approach is the constant-gain Kalman filter. This filter propagates a first-order statistical moment—the posterior expectation—in the place of the posterior distribution. The purpose of this paper is to propose an analogous strategy for multitarget systems: propagation of a first-order statistical moment of the multitarget posterior. This moment, the probability hypothesis density (PHD), is the function whose integral in any region of state space is the expected number of targets in that region. We derive recursive Bayes filter equations for the PHD that account for multiple sensors, nonconstant probability of detection, Poisson false alarms, and appearance, spawning, and disappearance of targets. We also show that the PHD is a best-fit approximation of the multitarget posterior in an information-theoretic sense.

Manuscript received September 11, 2000; revised March 27, 2003; released for publication May 7, 2003.

IEEE Log No. T-AES/39/4/822048.

Refereeing of this contribution was handled by X. R. Li.

This work was supported by the U.S. Army Research Office under Contract DAAG55-98-C-0039 and the U.S. Air Force Office of Research under Contract F49620-01-C-0031.

This work does not necessarily reflect the position or policy of the U.S. Government.

Author's address: Lockheed Martin NE&SS Tactical Systems, 3333 Pilot Knob Rd., Eagan, MN 55121, E-mail: (ronald.p.mahler@lmco.com).

NOMENCLATURE

$v(S)$	Lebesgue measure of set S (IIB6)
$\mathbf{1}_S(\mathbf{x})$	Indicator function of set S (IIIA)
$\delta_{\mathbf{w}}(\mathbf{x})$	Dirac delta function concentrated at $\mathbf{w}$
$\delta_X(\mathbf{x})$	Sum of Dirac deltas at elements of X (IVC2)
$\Delta_{\mathbf{w}}(S)$	Dirac measure concentrated at $\mathbf{w}$ (VIA)
h^X	Product of $h(\mathbf{x})$ with $\mathbf{x}$ in X (IIIA)
$\mathbf{x}$	Single-target state-vector (IIB1)
$\mathbf{X}$	Random state-vector
X	Single-target state space (IIB1)
$\mathbf{x}^*, \mathbf{x}^{*[j]}$	State of jth sensor (VD)
$\mathbf{z}, \mathbf{z}_k$	Single observation collected at time-step k (IIB2)
$\mathbf{z}^{[j]}$	Single observation from jth sensor (IIB2)
$\mathbf{Z}$	Random observation-vector
$\mathsf{Z}^{[j]}$	Observation space for jth sensor (IIB2)
X	Finite set of target state-vectors
Z, Z_k	Finite set of observations collected at time-step k
$Z^{[j]}$	Observation-set collected by jth sensor (IIB2)
$Z^k : \mathbf{z}_1,\ldots,\mathbf{z}_k$	Time-sequence of observations
$Z^{(k)} : Z_1,\ldots,Z_k$	Time-sequence of observation-sets
Ξ	Random state-set (finite subset of state space) (IIB8)
Σ	Random observation-set (IIB6)
$\int_S f(Y)\delta Y$	Set integral on region S of $\mathbf{Y}$ (IIB5)
$\int f(Y)\delta Y$	Set integral on $S = \mathbf{Y}$ (IIB5)
$\beta_\Psi(S)$	Belief-mass function of random finite set Ψ (IIB6)
$\frac{\delta\beta_\Psi}{\delta\mathbf{y}}(S)$	First-order set derivative of $\beta_\Psi(S)$ (IIB7)
$\frac{\delta\beta_\Psi}{\delta Y}(S)$	General set derivative of $\beta_\Psi(S)$ (IIB7)
$\frac{\delta^n\beta_\Psi}{\delta\mathbf{y}_1\cdots\delta\mathbf{y}_n}(S)$	General set derivative of $\beta_\Psi(S)$ (IIB7)
$f_\Psi(X)$	Multiobject probability density of Ψ (IIB5)
$D_\Psi(X)$	Multitarget moment density of Ψ (IIIA)
$G_\Psi[h]$	Probability generating functional of Ψ (IIIA)
$\frac{\delta^n G_\Psi}{\delta\mathbf{y}_1\cdots\delta\mathbf{y}_n}[h]$	Iterated functional derivative of $G_\Psi[h]$ (IIIB)
$f_{k+1\vert k}(\mathbf{x}\mid\mathbf{y})$	Single-target Markov transition density
$L_{\mathbf{z}}(\mathbf{x}) = f_k(\mathbf{z}\mid\mathbf{x})$	Single-sensor/target likelihood density
$c_k(\mathbf{z})$	Distribution of Poisson clutter process (VC)

λ_k	Average number of Poisson clutter observations (VC)
$p_D(\mathbf{x}) = p_D(\mathbf{x},\mathbf{x}^*)$	Probability of detection (VC)
$f_{k\|k}(\mathbf{x} \mid Z^k)$	Single-target posterior density, the distribution of random vector $\mathbf{X}_{k\|k}$
$f_{k+1\|k}(\mathbf{x} \mid Z^k)$	Single-target predicted posterior density, the distribution of random vector $\mathbf{X}_{k+1\|k}$ (IIA3)
$L_Z(X) = f_k(Z \mid X)$	Multisensor/target likelihood density (IIB4)
$f_{k+1\|k}(X \mid Y)$	Multitarget Markov transition density (IIB4)
$f_{k\|k}(X \mid Z^{(k)})$	Multitarget posterior density, the distribution of random state-set $\Xi_{k\|k}$ (IIB7)
$f_{k+1\|k}(X \mid Z^{(k)})$	Predicted multitarget posterior, the distribution of random state-set $\Xi_{k+1\|k}$ (IIB7)
$D_{k\|k}(\mathbf{x} \mid Z^{(k)})$	PHD of $\Xi_{k\|k}, f_{k\|k}(X \mid Z^{(k)})$ (IA, IVC)
$D_{k+1\|k}(\mathbf{x} \mid Z^{(k)})$	PHD of $\Xi_{k+1\|k}, f_{k+1\|k}(X \mid Z^{(k)})$ (IA, IVC)
$D_{k\|k}(X \mid Z^{(k)})$	Multitarget moment density of $\Xi_{k\|k}$ (IVA)
$G_{k\|k}[h]$	Probability generating functional of $\Xi_{k\|k}$ (IIIA)
$b_{k+1\|k}(Y \mid \mathbf{x})$	Multitarget posterior of spawned targets (VB)
$b_{k+1\|k}(\mathbf{y} \mid \mathbf{x})$	PHD of $b_{k+1\|k}(Y \mid \mathbf{x})$ (VB)
$b_{k+1\|k}(Y)$	Multitarget posterior of entering targets (VB)
$b_{k+1\|k}(\mathbf{y})$	PHD of $b_{k+1\|k}(Y)$ (VB).

I. INTRODUCTION

The following Bayesian discrete-time recursive nonlinear filtering equations constitute the theoretical foundation for optimal single-sensor, single-target detection, tracking and identification [11; 55; 4, pp. 373–377; 16, p. 174]:

$$f_{k+1|k}(\mathbf{x} \mid Z^k) = \int f_{k+1|k}(\mathbf{x} \mid \mathbf{w}) f_{k|k}(\mathbf{w} \mid Z^k) d\mathbf{w} \quad (1)$$

$$f_{k+1|k+1}(\mathbf{x} \mid Z^{k+1}) = K^{-1} f_k(\mathbf{z}_{k+1} \mid \mathbf{x}) f_{k+1|k}(\mathbf{x} \mid Z^k) \quad (2)$$

where

1) $\mathbf{x}$ is the state-vector of the target at time-step k and $\mathbf{z}_k$ is the sensor observation collected at time-step k;

2) $f_{k|k}(\mathbf{x} \mid Z^k)$ is the Bayes posterior distribution conditioned on the time-sequence $Z^k : \mathbf{z}_1,\ldots,\mathbf{z}_k$ of measurements accumulated at time-step k;

3) $L_{\mathbf{z},k}(\mathbf{x}) = f_k(\mathbf{z} \mid \mathbf{x})$ is the sensor likelihood function;

4) $f_{k+1|k}(\mathbf{x} \mid \mathbf{w})$ is the Markov transition density that models the between-measurements motion of the target;

5) $f_{k+1|k}(\mathbf{x} \mid Z^k)$ is the time-prediction of the posterior $f_{k|k}(\mathbf{x} \mid Z^k)$ to the time-step of the next measurement $\mathbf{z}_{k+1}$; and

6) $K = f_{k+1}(\mathbf{z}_{k+1} \mid Z^k) = \int f_k(\mathbf{z}_{k+1} \mid \mathbf{x}) f_{k+1|k}(\mathbf{x} \mid Z^k) d\mathbf{x}$ is the Bayes normalization factor.

If $= f_k(\mathbf{z} \mid \mathbf{x})$, $f_{k+1|k}(\mathbf{x} \mid \mathbf{w})$, and $f_{0|0}(\mathbf{x})$ are linear and Gaussian then (1), (2) reduce to the Kalman time-update and information-update equations [11].

Similarly, the general, theoretically optimal approach to multisensor-multitarget detection, tracking, and identification is the following generalization of the recursive Bayes filter:

$$f_{k+1|k}(X \mid Z^{(k)}) = \int f_{k+1|k}(X \mid W) f_{k|k}(W \mid Z^{(k)}) \delta W \quad (3)$$

$$f_{k+1|k+1}(X \mid Z^{(k+1)}) = K^{-1} f_{k+1}(Z_{k+1} \mid X) f_{k+1|k}(X \mid Z^{(k)}) \quad (4)$$

where

7) X is the multitarget state-set, i.e., the set of unknown target states (which are also of unknown number), and has the form $X = \emptyset, \{\mathbf{x}_1\}, \{\mathbf{x}_1,\mathbf{x}_2\},\ldots,\{\mathbf{x}_1,\ldots,\mathbf{x}_n\},\ldots$ where $X = \emptyset$ means no target is present, $X = \{\mathbf{x}_1\}$ means that one target with state-vector $\mathbf{x}_1$ is present, $X = \{\mathbf{x}_1,\mathbf{x}_2\}$ means that two targets with state-vectors $\mathbf{x}_1 \neq \mathbf{x}_2$ are present, etc.;

8) Z_k is the observation-set consisting of all measurements collected from all targets by all sensors at time-step k;

9) $f_{k|k}(X \mid Z^{(k)})$ is the multitarget posterior density at time-step k, conditioned on the time-sequence $Z^{(k)} : Z_1,\ldots,Z_k$ of observation-sets accumulated at time-step k;

10) $L_{Z,k}(X) = f_k(Z \mid X)$ is the multisensor, multitarget likelihood function that describes the likelihood of observing the observation-set Z given that the targets have multitarget state-set X;

11) $f_{k+1|k}(X \mid W)$ is the multitarget Markov transition density that describes the likelihood that the targets will have state-set X at time-step $k+1$ given that they had state-set W at time-step k (it models between-measurements multitarget motion, including individual target motions and appearance and disappearance of targets);

12) $f_{k+1|k}(X \mid Z^{(k)})$ is the time-prediction of the multitarget posterior $f_{k|k}(X \mid Z^{(k)})$ to the time-step of the next observation-set Z_{k+1}; and

13) $K = f_{k+1}(Z_{k+1} \mid Z^{(k)}) = \int f_{k+1}(Z_{k+1} \mid X) f_{k+1|k} \cdot (X \mid Z^{(k)}) \delta X$ is the Bayes normalization factor.

The multitarget filter equations (3), (4) cannot be copied from the single-target filter equations (1), (2) in the blind fashion just indicated but, rather, require the tools of finite-set statistics (FISST) (see Section II).

The single-sensor, single-target Bayes filter is already so computationally demanding that it must usually be approximated. Consequently, the multisensor-multitarget Bayes filter will have no practical utility without drastic but intelligent approximation strategies. This paper proposes a multitarget statistical analog of the constant-gain Kalman filter, i.e., propagation of a multitarget first-order moment statistic rather than the entire multitarget posterior itself.

A. First-Order Multitarget Statistical Moments

In single-target problems the two most familiar statistical moments of the posterior $f_{k|k}(\mathbf{x} \mid Z^k)$ are the first-moment vector (posterior expectation) and second-order moment matrix:

$$\mathbf{x}_{k|k} = \int \mathbf{x} \cdot f_{k|k}(\mathbf{x} \mid Z^k) d\mathbf{x}, \qquad C_{k|k} = \int \mathbf{x}\mathbf{x}^{\mathrm{T}} \cdot f_{k|k}(\mathbf{x} \mid Z^k) d\mathbf{x}$$

where "T" denotes matrix transpose. If we assume that the higher-order moment-tensors can be neglected, then $\mathbf{x}_{k|k}$ and $C_{k|k}$ are approximate sufficient statistics, i.e.:

$$f_{k|k}(\mathbf{x} \mid Z^k) \cong f(\mathbf{x} \mid \mathbf{x}_{k|k}, C_{k|k}) = N_{P_{k|k}}(\mathbf{x} - \mathbf{x}_{k|k})$$

where $N_{P_{k|k}}(\mathbf{x} - \hat{\mathbf{x}}_{k|k})$ denotes a multidimensional Gaussian distribution with covariance matrix $P_{k|k} = C_{k|k} - \mathbf{x}_{k|k}(\mathbf{x}_{k|k})^{\mathrm{T}}$. In this case we can propagate $\mathbf{x}_{k|k}$ and $P_{k|k}$ instead of the full distribution $f_{k|k}(\mathbf{x} \mid Z^k)$, using a Kalman filter. If we assume that the second-order moment can be neglected as well, then $f_{k|k}(\mathbf{x} \mid Z^k) \cong f(\mathbf{x} \mid \mathbf{x}_{k|k})$ and we can propagate $\mathbf{x}_{k|k}$ alone using a constant-gain Kalman filter (e.g., an α-β-γ filter):

$$\begin{array}{ccccccccc}
\cdots \rightarrow & f_{k|k}(\mathbf{x} \mid Z^k) & \overset{\text{time prediction}}{\rightarrow} & f_{k+1|k}(\mathbf{x} \mid Z^k) & \overset{\text{Bayes' rule}}{\rightarrow} & f_{k+1|k+1}(\mathbf{x} \mid Z^{k+1}) & \rightarrow \cdots \\
& \downarrow & & \downarrow & & \downarrow & \\
\cdots \rightarrow & \mathbf{x}_{k|k} & \overset{\text{predictor}}{\rightarrow} & \mathbf{x}_{k+1|k} & \overset{\text{corrector}}{\rightarrow} & \mathbf{x}_{k+1|k+1} & \rightarrow \cdots
\end{array}$$

Here the top row portays the time-evolution of the single-target Bayes filter equations (1)–(2); the vertical arrows indicate the collapse of posteriors into their corresponding expectations; and the bottom row portrays the time-evolution of the constant-gain Kalman filter.

Under what sensing conditions can we neglect higher-order moments? A unimodal likelihood function $L_{\mathbf{z},k}(\mathbf{x}) = f_{k+1}(\mathbf{z} \mid \mathbf{x})$ may be non-Gaussian (e.g. heavy tails, large skew or kurtosis, etc.), but if its covariance is small this doesn't greatly matter because small-covariance, unimodal distributions all look much the same. Stated differently: $L_{\mathbf{z},k}(\mathbf{x})$ is so highly concentrated around some $\mathbf{x}$ that all of the posteriors $f_{k|k}(\mathbf{x} \mid Z^k)$ constructed from it are similarly concentrated, and therefore characterized by $\mathbf{x}_{k|k}$.

Our goal is to extend this reasoning to multitarget problems. We assume that some first-order statistical moment $D_{k|k}$ is an approximate sufficient statistic—i.e., $f_{k|k}(X \mid Z^{(k)}) \cong f(X \mid D_{k|k})$—and then "fill in the question marks" in the following diagram:

$$\begin{array}{ccccccccc}
\cdots \rightarrow & f_{k|k}(X \mid Z^{(k)}) & \overset{\text{time prediction}}{\rightarrow} & f_{k+1|k}(X \mid Z^{(k)}) & \overset{\text{Bayes' rule}}{\rightarrow} & f_{k+1|k+1}(X \mid Z^{(k+1)}) & \rightarrow \cdots \\
& \downarrow & & \downarrow & & \downarrow & \\
\cdots \rightarrow & D_{k|k} & \overset{\text{predictor?}}{\rightarrow} & D_{k+1|k} & \overset{\text{corrector?}}{\rightarrow} & D_{k+1|k+1} & \rightarrow \cdots
\end{array}$$

The top row portrays the time-evolution of the multitarget Bayes filter equations (3), (4); the downward-pointing arrows indicate the collapse of multitarget posteriors into their first-order moments; and the bottom row portrays the evolution of the approximate first-moment filter.

If either the moments $D_{k|k}$ and $D_{k+1|k}$ or the predictor or the corrector are poorly chosen, useful information will be unnecessarily discarded. First, information loss in the $f \rightarrow D$ should be minimized. Second, minimal-loss $f \rightarrow D$ can be fully exploited only if the predictor is "lossless," i.e. it produces the first moment $D_{k+1|k}$ of $f_{k+1|k}$ as its answer. Third, the corrector should be similarly lossless. More detailed discussion is deferred until Section V.

The moment $D_{k|k}$ used here was first proposed in 1993 by M. C. Stein and C. L. Winter [59] and Stein and R. R. Tenney [60] for group target applications [67]:

DEFINITION 1 (Probability Hypothesis Density) The PHD is the density $D_{k|k}(\mathbf{x} \mid Z^{(k)})$ whose integral

$\int_S D_{k|k}(\mathbf{x} \mid Z^{(k)})d\mathbf{x}$ on any region S of state space is $N_{k|k}(S) = \int |X \cap S| f_{k|k}(X \mid Z^{(k)})\delta X$, the expected number of targets contained in S.

This property characterizes the PHD uniquely: if $g_{k|k}(\mathbf{x})$ is any other density function such that $\int_S g_{k|k}(\mathbf{x})d\mathbf{x} = N_{k|k}(S)$, it is just the PHD. For, since $\int_S D_{k|k}(\mathbf{x} \mid Z^{(k)})d\mathbf{x} = \int_S g_{k|k}(\mathbf{x})d\mathbf{x}$ for all measurable S then $D_{k|k}(\mathbf{x} \mid Z^{(k)}) = g_{k|k}(\mathbf{x})$ almost everywhere.

The value $D_{k|k}(\mathbf{x} \mid Z^{(k)})d\mathbf{x}$ is the expected number of targets in an infinitesimally small region $d\mathbf{x}$ of $\mathbf{x}$, i.e, $D_{k|k}(\mathbf{x} \mid Z^{(k)})$ is the expected target density at $\mathbf{x}$. Noting that $f_{k|k}(X \mid Z^{(k)})$ is the probability distribution of some random state-set $\Xi_{k|k}$ (Section IIB8), $D_{k|k}(\mathbf{x} \mid Z^{(k)})$ represents the zero-probability event $\Pr(\mathbf{x} \in \Xi_{k|k})$ just as the density $f_{\mathbf{X}}(\mathbf{x})$ of a continuous random vector $\mathbf{X}$ represents the zero-probability event $\Pr(\mathbf{X} = \mathbf{x})$ (Section IVD3). Consequently, $D_{k|k}(\mathbf{x} \mid Z^{(k)})$ will be multimodal with its peaks located near the actual target states.

B. Poisson Approximation

If the moments $D_{k|k}$ and $D_{k+1|k}$ are chosen to be the PHDs of $f_{k+1|k}(X \mid Z^{(k)})$ and $f_{k+1|k+1}(X \mid Z^{(k+1)})$, respectively, then it will be shown that $D_{k+1|k}$ and $D_{k+1|k+1}$ are "best fit" approximations of $f_{k+1|k}(X \mid Z^{(k)})$ and $f_{k+1|k+1}(X \mid Z^{(k+1)})$, respectively, in that Kullback-Leibler information functionals are minimized (Theorem 4). Also, the predictor $D_{k|k} \to D_{k+1|k}$ will be lossless. The corrector $D_{k+1|k} \to D_{k+1|k+1}$ will not be lossless because, to derive a closed-form formula, we must assume the following: The predicted multitarget posterior $f_{k+1|k}(X \mid Z^{(k)})$ is approximately Poisson (Section IVD2):

$$f_{k+1|k}(X \mid Z^{(k)}) \cong e^{-N} D(\mathbf{x}_1)\cdots D(\mathbf{x}_n) \tag{5}$$

for any $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ with $\mathbf{x}_1,\ldots,\mathbf{x}_n$ distinct, where $D(\mathbf{x}) = D_{k+1|k}(\mathbf{x} \mid Z^{(k)})$ is the PHD of $f_{k+1|k}(X \mid Z^{(k)})$ and $N = \int D(\mathbf{x})d\mathbf{x}$.

The conditions under which this approximation is justifiable resemble those underlying the Kalman approximation. Both the multisensor-multitarget likelihood function $L_{Z,k}(X) = f_{k+1}(Z \mid X)$ and the multitarget posteriors $f_{k|k}(X \mid Z^{(k)})$ will be non-Poisson in general. But if both sensor covariances and sensor false alarm densities are small then observations will be tightly clustered around target states, confusion due to false alarms will be small, and so the time-evolving multitarget posteriors will be roughly characterized by their first-order moments.

C. Simple Example

Two targets on the unitless real-number line are to be located using a data-scan $Z = \{z_1,z_2\}$ collected by a position-measuring sensor. Suppose that the multitarget posterior is $f(X) = 0$ unless $X = \{x_1,x_2\}$, $x_1 \neq x_2$, in which case

$$f(\{x_1,x_2\}) = N_\sigma(x_1 - z_1)N_\sigma(x_2 - z_2) + N_\sigma(x_1 - z_2)N_\sigma(x_2 - z_1)$$

where $N_\sigma(x)$ is a Gaussian distribution with mean 0 and variance σ^2. If z_1,z_2 are well separated, $f(X)$ is maximized near $(x_1,x_2) = (z_1,z_2)$ and $(x_1,x_2) = (z_2,z_1)$, i.e., near $X = \{z_1,z_2\}$. The PHD of $f(X)$ is

$$D(x) = \int f(\{x,x_2\})dx_2 = N_\sigma(x - z_1) + N_\sigma(x - z_2). \tag{6}$$

Expected target number is $\int D(x)dx = 2$, and $D(x)$ is bimodal with maxima near z_1 and z_2 if z_1,z_2 are well separated.

However, it is easily shown that $D(x)$ is unimodal with maximal value at $x = \frac{1}{2}(z_1 + z_2)$ when $|z_1 - z_2| < 2\sigma$. The multitarget posterior $f(\{x_1,x_2\})$, on the other hand, also fails to distinguish two distinct targets when $|z_1 - z_2| < \sqrt{2}\sigma$, in which case its unique maximal value is located at $x_1 = x_2 = \frac{1}{2}(z_1 + z_2)$ (see [10, pp. 248–252]). So, for data separations in the range $\sqrt{2}\sigma < |z_1 - z_2| < 2\sigma$ the multitarget posterior will separate the two targets whereas the PHD cannot. Thus a PHD-based multitarget tracker will experience more difficulty with closely-spaced targets than the multitarget nonlinear filter, unless σ is small compared with target separation.

The multitarget Poisson density that approximates $f(X)$ is

$$g(X) = e^{-2}D(x_1)\cdots D(x_n)$$

for $X = \{x_1,\ldots,x_n\}$ with $x_1,\ldots,x_n$ distinct. In general this does not resemble $f(X)$ except that its expected target number is 2. If σ is small, however, then $g(X) \cong 0$ except near the maxima $X = \{x_1,\ldots,x_n\}$ with all x_i near z_1 or z_2. Even though $g(X)$ is not uniquely maximized at $X = \{z_1,z_2\}$, like $f(X)$ it is concentrated near z_1 and z_2, and so has the same PHD.

D. Summary of Main Results

In this paper we derive recursive Bayes filter equations for the PHD. First, we time-extrapolate the old PHD $D_{k|k}(\mathbf{x}) = D_{k|k}(\mathbf{x} \mid Z^{(k)})$ to the PHD $D_{k+1|k}(\mathbf{x}) = D_{k+1|k}(\mathbf{x} \mid Z^{(k)})$ at the next measurement-collection step $k+1$ using the following predictor equation (Theorem 5, Section VB):

$$D_{k+1|k}(\mathbf{x}) = b_{k+1|k}(\mathbf{x}) + \int (p_S(\mathbf{w})f_{k+1|k}(\mathbf{x} \mid \mathbf{w}) + b_{k+1|k}(\mathbf{x} \mid \mathbf{w}))d\mathbf{w}. \tag{7}$$

Here, $f_{k+1|k}(\mathbf{y} \mid \mathbf{x})$ is the single-target Markov transition density; $p_S(\mathbf{x})$ is the probability that a target with state $\mathbf{x}$ at time-step k will survive into time-step $k+1$; $b_{k+1|k}(\mathbf{y} \mid \mathbf{x})$ describes the spawning of new targets from existing ones; and $b_{k+1|k}(\mathbf{y})$ describes the entry of other new targets.

Suppose now that the targets are interrogated by a single sensor and that a new observation-scan Z_{k+1} has been collected at time-step $k+1$. Then $D_{k+1|k}(\mathbf{x})$ is updated to a new PHD $D_{k+1|k+1}(\mathbf{x}) = D_{k+1|k+1}(\mathbf{x} \mid Z^{(k+1)})$ using the following approximate Bayes corrector equation (Theorem 6, Section VC):

$$D_{k+1|k+1}(\mathbf{x}) \cong F_{k+1}(Z_{k+1} \mid \mathbf{x}) D_{k+1|k}(\mathbf{x}) \tag{8}$$

where

$$F_{k+1}(Z \mid \mathbf{x}) = \sum_{\mathbf{z} \in Z} \frac{p_D(\mathbf{x}) L_{\mathbf{z}}(\mathbf{x})}{\lambda c(\mathbf{z}) + D_{k+1|k}[p_D L_{\mathbf{z}}]} + 1 - p_D(\mathbf{x}) \tag{9}$$

and where $D_{k+1|k}[h] = \int h(\mathbf{x}) D_{k+1|k}(\mathbf{x}) d\mathbf{w}$. Also, $p_D(\mathbf{x})$ is the probability of detection (field of view (FOV)) of the sensor and $L_{\mathbf{z}}(\mathbf{x}) = f_{k+1}(\mathbf{z} \mid \mathbf{x})$ is the sensor likelihood function. It has also been assumed here that the sensor collects an average of $\lambda = \lambda_{k+1}$ Poisson-distributed false alarms per scan, distributed according to the probability density $c(\mathbf{z}) = c_{k+1}(\mathbf{z})$.

Given this, the corrector equation (8) is used to update $D_{k+1|k}(\mathbf{x})$ to $D_{k+1|k+1}(\mathbf{x})$ (Theorem 7, Section VF):

$$D_{k+1|k}(\mathbf{x}) \cong F^{[1]}_{k+1}(Z^{[1]}_{k+1} \mid \mathbf{x}) \cdots F^{[s]}_{k+1}(Z^{[s]}_{k+1} \mid \mathbf{x}) \tag{10}$$

where

$$F^{[j]}_{k+1}(Z^{[j]}_{k+1} \mid \mathbf{x}) = \sum_{\mathbf{z}^{[j]} \in Z^{[j]}_{k+1}} \frac{p^{[j]}_D(\mathbf{x}) L^{[j]}_{\mathbf{z}^{[j]}}(\mathbf{x})}{\tilde{c}^{[j]}(\mathbf{z}^{[j]}) + D_{k+1|k}[p^{[j]}_D L^{[j]}_{\mathbf{z}^{[j]}}]} + 1 - p^{[j]}_D(\mathbf{x}) \tag{11}$$

where $D_{k+1|k}[h] = \int h(\mathbf{x}) D_{k+1|k}(\mathbf{x}) d\mathbf{w}$, and where $Z^{[j]}_{k+1}$ is the subset of Z_{k+1} of observations originating with the jth sensor.

E. Related Approaches and Publications

A short history of the multitarget Bayes filter (eqns. (3), (4)) can be found in Section IIC. The idea of using a single-target density function $g_{k|k}(\mathbf{x})$ (or contour maps of its graph) as a basis for multitarget tracking is a relatively common one. Examples of implemented algorithms are the Naval Research Laboratory's TABS (Tactical Antisubmarine-warfare Battle-management System) tracker, Metron Corp.'s Nodestar tracker [61], [62], and the "probabilistic mapping" multitarget tracking approach of Tao, Abileah, and Lawrence [63]. Stein, Winter, and Tenney's work on the PHD has already been noted.

The work presented here provides a solid theoretical foundation for single-density approaches. Mahler showed that the PHD is the first-order moment of a point process in 1997 [10, pp. 168–170, 179], with proofs first appearing in [38, 42]. He introduced the PHD filter in 2000 [38, 42]. (The same year, Mori suggested the use of Poisson approximations for multitarget tracking [45].) The proof techniques in this work, based on probability generating functionals and functional derivatives (Section III), are more powerful than those used in earlier papers and first appeared in [35]. They lead to more general results than those announced earlier: sensor probabilities of detection need not be constant. Mahler has proposed the PHD filter as a potentially computationally tractable basis for unified group target detection, tracking, and classification [24, 27], and for cluster tracking [25, 43].

Because the PHD filter resembles the usual single-sensor, single-target Bayes filter, it can in principle be implemented using any computational nonlinear filtering technique. El-Fellah et al. [8] describe a PHD filter based on a "spectral compression" technique. Zajic and Mahler [43, 68, 69], Vo, Doucet et al. [65], and Sidenbladh [52] describe particle-systems and sequential Monte Carlo implementations.

F. Organization of the Paper

Section II provides a summary of those aspects of FISST required to understand this paper. Section III introduces additional mathematical concepts needed for the proofs: probability generating functionals (PGFLs) and their functional derivatives. Section IV is devoted to the basic concepts of the PHD approach: multitarget moment densities and verification that the PHD is a first-order statistical moment. The Bayes filtering equations for the PHD are derived in Section V along with an information-theoretic best-fit characterization of the PHD. Proofs of the theorems are in Section VI. Conclusions may be found in Section VII.

II. FINITE SET STATISTICS. A SUMMARY

Progress in single-sensor single-object tracking has been greatly facilitated by the existence of a systematic, rigorous, and yet practical engineering statistics. Until recently multisensor-multitarget applications have lacked a similar statistical basis, despite the decades-long existence of the recognized mathematical foundation for stochastic multiobject problems, point process theory [3, 7, 17, 50, 54, 56]. This section summarizes finite-set statistics (FISST), which is in part an "engineering friendly" formulation of point process theory. That is, it is geometric (models multiobject systems as visualizable images) and preserves the "Statistics 101" formalism that tracking engineers already understand.

FISST provides a way to directly extend single-sensor, single-target Bayes statistics to multisensor-multitarget problems. In Section IIA, we describe this statistics and, in Section IIB, its extension to multisensor-multitarget problems is described. Section IIC provides a short history of the multitarget Bayes filter, and Section IID describes relationships between FISST and point process theory needed later. For futher information regarding FISST, consult [10, 26, 29, 37].

A. Single-Sensor Single-Target Statistics

1) *Single-Sensor Single-Target Modeling*: In single-sensor single-target problems one generally begins with models

$$\mathbf{Z}_k = h_k(\mathbf{x}) + \mathbf{W}_k \tag{12}$$

$$\mathbf{X}_{k+1|k} = g_k(\mathbf{x}) + \mathbf{V}_{k+1|k} \tag{13}$$

of the sensor observations and presumed motion of the target. Here, the random vector $\mathbf{W}_k$ describes sensor noise and the random vector $\mathbf{V}_k$ provides a means of hedging against the fact that actual target dynamics are usually unknown.

2) *Single-Sensor Single-Target Measurement and Markov Densities*: A Bayesian analysis requires that we transform these models into probability density functions. To do this we must first construct the probability mass functions

$$p_k(S \mid \mathbf{x}) = \Pr(\mathbf{Z}_k \in S) \tag{14}$$

$$p_{k+1|k}(S \mid \mathbf{x}) = \Pr(\mathbf{X}_{k+1|k} \in S). \tag{15}$$

Using techniques of the integral and differential calculus, we deduce that the true sensor likelihood function—i.e, the density that faithfully describes the measurement model—is $L_{\mathbf{z},k}(\mathbf{x}) = f_k(\mathbf{z} \mid \mathbf{x}) = f_k(\mathbf{z} - h_k(\mathbf{x}))$ where $f_k(\mathbf{z})$ denotes the density of $\mathbf{W}_k$. One likewise deduces that the true Markov transition density—the one that faithfully describes the motion model—is $f_{k+1|k}(\mathbf{y} \mid \mathbf{x}) = f_{k+1|k}(\mathbf{y} - g_k(\mathbf{x}))$ where $f_{k+1|k}(\mathbf{y})$ denotes the density of $\mathbf{V}_{k+1|k}$. These are probability densities:

$$1 = \int f_k(\mathbf{z} \mid \mathbf{x}) d\mathbf{z}, \qquad 1 = \int f_{k+1|k}(\mathbf{y} \mid \mathbf{x}) d\mathbf{y}. \tag{16}$$

3) *Single-Sensor Single-Target State Estimation*: The posterior $f_{k|k}(\mathbf{x} \mid Z^k)$ contains all known information about the state of the target at time-step k. This information is unavailable for application without a Bayes-optimal state estimator [64, pp. 54–58] e.g., the maximum a posteriori (MAP) and expected a posteriori (EAP) estimators:

$$\hat{\mathbf{x}}^{\text{MAP}}_{k|k} = \arg\sup_{\mathbf{x}} f_{k|k}(\mathbf{x} \mid Z^k), \qquad \hat{\mathbf{x}}^{\text{EAP}}_{k|k} = \int \mathbf{x} \cdot f_{k|k}(\mathbf{x} \mid Z^k) d\mathbf{x}. \tag{17}$$

4) *Single-Sensor Single-Target Bayes Filter*: We now have a solid basis for the recursive Bayes filter equations (1), (2). Without the true sensor likelihood function or a Bayes-optimal state estimator, any claim about "Bayes optimality" or having the "true Bayes posterior" would be hollow or false.

It should also be emphasized that in a Bayesian approach the unknown state $\mathbf{x}$ is a random variable rather than a fixed parameter. So, (1), (2) describe the time-evolution

$$\cdots \to \mathbf{X}_{k|k} \to \mathbf{X}_{k+1|k} \to \mathbf{X}_{k+1|k+1} \to \cdots \tag{18}$$

of the random state-vector of the single-target system, where the posteriors $f_{k|k}(\mathbf{x} \mid Z^k)$ and $f_{k+1|k}(\mathbf{x} \mid Z^k)$ are the probability distributions of $\mathbf{X}_{k|k}$ and $\mathbf{X}_{k+1|k}$, respectively.

B. Multisensor-Multitarget Statistics

A major purpose of FISST is to extend this reasoning to multisensor-multitarget problems. This methodology is employed in the proofs of Theorems 5 and 6.

1) *Multitarget State Spaces*: Let X denote single-target state space. In the terminology of [47] it is a "hybrid space" $\mathsf{X} = \Re^N \times C$ (see [10, pp. 135–137, 220]) whose elements are vectors of the form $\mathbf{x} = (x_1,\ldots,x_N,c)$ where $x_1,\ldots,x_n$ are in the set $\Re$ of real numbers (position, velocity, etc.) and c belongs to some finite set C (target class, threat status, etc.). (Note: in this work at least one state variable must be continuous, $N > 0$.) Integrals of real-valued functions $f(\mathbf{x})$ of state variables involve both summations and continuous integrals. Because of c, $f_k(\mathbf{z} \mid \mathbf{x})$ and $f_{k+1|k}(\mathbf{x} \mid \mathbf{w})$ can encompass different measurement and motion models for different target types. Given this, the state of a multitarget system is a finite set $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ of state-vectors where n is the number of targets and $\mathbf{x}_1,\ldots,\mathbf{x}_n$ are their individual states. The multitarget state space is the class of all finite subsets of X.[1] (A complete discussion of the subtleties of defining multitarget state spaces—sets versus multisets—is on [10, pp. 194–199].)

2) *Multisensor-Multitarget Measurement Spaces*: Suppose that there are s sensors that collect respective observations $\mathbf{z}^{[1]},\ldots,\mathbf{z}^{[s]}$ from respective measurement spaces $\mathsf{Z}^{[1]},\ldots,\mathsf{Z}^{[s]}$. Each is a hybrid space $\mathsf{Z}^{[j]} = \Re^{M(j)} \times D^{[j]} \times \{j\}$ (see [10, p. 220]) consisting of vectors $\mathbf{z} = (z_1,\ldots,z_M,d,j)$ where in general $M = M(j)$

[1] So that random multitarget states—i.e., random finite state-sets—can be defined, this class is endowed with the so-called Mathéron topology (see [10, pp. 131–135]). The same must be done with the multisensor-multitarget measurement spaces. Further discussion is beyond the scope of this paper. Also, note that unlike the multitarget state-space on [61, p. 164], our formulation permits arbitrarily large target number. Without this feature, the Poisson approximation of Section VA would be impossible.

and $\mu = \mu(j)$ will vary with the sensor tag $j = 1,\ldots,s$.[2] Each sensor will collect a finite set $Z^{[j]}$ of individual observations, and so all sensors will collect a finite set $Z = Z^{[1]} \cup \cdots \cup Z^{[s]}$ of observations. Consequently, one must bundle all single observation spaces together into a single "meta-measurement space," the topological sum $\mathsf{Z} = \mathsf{Z}^{[1]} \oplus \cdots \oplus \mathsf{Z}^{[s]}$ where $\oplus$ denotes disjoint union (see [20, Definition 24a, p. 159]).[3] So, any multisensor-multitarget "observation" must be some finite subset Z of Z. The multisensor-multitarget measurement space is, consequently, the class of finite subsets of Z.

Moreover, Z can itself be assumed for conceptual convenience to be a hybrid space ([10, Definition 2, p. 220]) since it is a subset of $\Re^M \times D$ where $M = M(1) + \cdots + M(s)$ and $D = (D^{[1]} \oplus \cdots \oplus D^{[s]}) \times \{1,\ldots,s\}$.

3) *Multisensor-Multitarget Modeling*: By analogy with the single-sensor single-target case, the analysis of a multisensor-multitarget problem should begin with a model

$$\Sigma_k = E_k(X) \cup C_k(X) \tag{19}$$

of the multisensor-multitarget observation and a model

$$\Xi_{k+1|k} = D_k(X) \cup B_k(X) \tag{20}$$

of the presumed motions of the targets. Here, $E_k(X)$ models the self-noise of the sensors and their probabilities of detection, whereas $C_k(X)$ models false alarms and/or clutter. Likewise, $D_k(X)$ models presumed target motion and the persistence/disappearance of existing targets, whereas $B_k(X)$ models the appearance of new targets. (See [29, pp. 17–23] for more details, as well as Theorems 5 and 6.)

4) *Multisensor-Multitarget Measurement and Markov Densities*: In a Bayesian methodology we must convert these models into their respective density functions. That is, we need a general, systematic procedure for constructing the true multisensor-multitarget measurement density $f_k(Z \mid X)$ that faithfully describes the multisensor-multitarget measurement model. It is the likelihood that a multisensor observation-set Z will be collected if targets with state-set X are present. We likewise need a general, systematic procedure for constructing the true multitarget Markov density $f_{k+1|k}(Y \mid X)$, which faithfully describes the multitarget motion model. This is the likelihood that the targets will have state-set Y at time-step $k+1$ if at time-step k they had state-set X.

Provision of such systematic, general procedures is a major purpose of the FISST set integral and set derivative.

5) *Set Integral and Multiobject Probability Density Functions*: If $f(Y)$ is any real-valued function of a finite-set variable $Y \subseteq \mathsf{Y}$, its set integral[4] in a region S is[5]

$$\int_S f(Y)\delta Y = f(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{S^n} f(\{\mathbf{y}_1,\ldots,\mathbf{y}_n\}) d\mathbf{y}_1 \cdots d\mathbf{y}_n. \tag{21}$$

Let $\int \cdot \delta Y = \int_s \cdot \delta Y$ if $S = \mathsf{Y}$. Then $f(Y)$ is a multiobject probability density function [10, pp. 162–168] if

$$1 = \int f(Y)\delta Y. \tag{22}$$

If $S = \mathsf{Y}$ the nth term of (21) is, for $n = 0,1,2,\ldots$, the probability that there are n objects.[6] (For more information on the set integral see [10, pp. 141–144, 159–160] and [29, pp. 28–29].)

Both the likelihood and Markov densities are multiobject probability density functions [10, pp. 162–168]:

$$1 = \int f_k(Z \mid X)\delta Z, \qquad 1 = \int f_{k+1|k}(Y \mid X)\delta Y. \tag{23}$$

6) *Set Derivative*: The construction of true multisensor-multitarget measurement densities and true multitarget Markov densities is made possible by the inverse operation of the set integral called the set derivative. Given the measurement model $\Sigma_k = T_k(X) \cup C_k(X)$ and motion model $\Xi_{k+1|k} = D_k(X) \cup B_k(X)$ we first construct their belief-mass functions [10, pp. 152–157]:

$$\beta_k(S \mid X) = \Pr(\Sigma_k \subseteq S) = \int_S f_k(Z \mid X)\delta Z \tag{24}$$

$$\beta_{k+1|k}(S \mid X) = \Pr(\Xi_{k+1|k} \subseteq S) = \int_S f_{k+1|k}(Y \mid X)\delta Y. \tag{25}$$

---

[2]Contrary to the assertion of [61, p. 204], such vectors need not be "contacts" or "detections" but can be vector models of predetection observations, e.g., vectors whose components are image pixel intensities, radar range-bin intensities, or acoustic frequency-bin intensities.

[3]Again contrary to the assertions of [61, p. 204], it is therefore absolutely necessary that "all measurements take values in the same space with a special topology"; and it is not true that FISST "requires that the measurement spaces be identical for all sensors."

[4]The set integral can be described in terms of the theory of measure-theoretic integrals, but such a discussion is beyond the scope of this paper.

[5]In these integrals $f^*(\mathbf{x}_1,\ldots,\mathbf{x}_n) = f(\mathbf{x}_1,\ldots,\mathbf{x}_n)$ are functions of n variables rather than of a finite set $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$, so adopt the convention $f^*(\{\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_j,\ldots,\mathbf{x}_n\}) = 0$ if $\mathbf{x}_i = \mathbf{x}_j$ for some $i \neq j$. If any state variable is continuous such events have probability zero. (See also Section IIB1.)

[6]Warning: Set integrals of expressions involving multitarget probability densities $f(X)$ may be undefined (see [29, pp. 39, 41] and [10, p. 163]) of [10]). Whereas the units of measurement of a conventional density function $f(\mathbf{x})$ are constant, the units of measurement of a multitarget probability density $f(X)$ vary with the cardinality $|X|$ of X. Consequently, set integrals like $\int f(X)^2 \delta X$ and $\int f(X) \log f(X) \delta X$ can be undefined.

Then it can be shown [29, pp. 30–31] and [10, pp. 159–161]:

$$f_k(Z \mid X) = \frac{\delta\beta_k}{\delta Z}(\emptyset \mid X) \tag{26}$$

$$f_{k+1|k}(Y \mid X) = \frac{\delta^n \beta_{k+1|k}}{\delta Y}(\emptyset \mid X) \tag{27}$$

where, for arbitrary functions $F(Y)$ of a finite-set variable Y,

$$\begin{aligned} \frac{\delta F}{\delta \mathbf{y}}(S) &= \lim_{v(E_{\mathbf{y}})\to 0} \frac{F(S \cup E_{\mathbf{y}}) - F(S)}{v(E_{\mathbf{y}})} \\ \frac{\delta\beta}{\delta Y}(S) &= \frac{\delta^n \beta}{\delta \mathbf{y}_n \cdots \delta \mathbf{y}_1}(S) = \frac{\delta}{\delta \mathbf{y}_n}\frac{\delta^{n-1}\beta}{\delta \mathbf{y}_{n-1}\cdots\delta\mathbf{y}_1}(S) \end{aligned} \tag{28}$$

are called set derivatives; where $E_{\mathbf{y}}$ is a small neighborhood of $\mathbf{y}$; and where v(S) is the hypervolume (i.e., Lebesgue measure) of set S.[7] Likewise, multitarget posteriors

$$f_{k|k}(X \mid Z^{(k)}) = \frac{\delta\beta_{k|k}}{\delta X}(\emptyset \mid Z^{(k)}) \tag{29}$$

can be constructed from their belief mass functions

$$\beta_{k|k}(S \mid Z^{(k)}) = \Pr(\Xi_{k|k} \subseteq S) = \int_S f_{k|k}(X \mid Z^{(k)})\delta X. \tag{30}$$

Set derivatives can be computed using "turn the crank" rules similar to those of undergraduate calculus [29, pp. 31–32] and [10, pp. 143, 146, 151]). These include:

14) Sum rule:

$$\frac{\delta}{\delta Y}(a_1\beta_1(S) + a_2\beta_2(S)) = a_1\frac{\delta\beta_1}{\delta Y}(S) + a_2\frac{\delta\beta_2}{\delta Y}(S). \tag{31}$$

15) Product rule:

$$\frac{\delta}{\delta Y}(\beta_1(S)\beta_2(S)) = \sum_{W\subseteq Y}\frac{\delta\beta_1}{\delta W}(S)\frac{\delta\beta_1}{\delta(Y-W)}(S). \tag{32}$$

16) Chain rule:

$$\frac{\delta}{\delta \mathbf{y}}(f(\beta_1(S),\ldots,\beta_n(S)) = \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(\beta_1(S),\ldots,\beta_n(S))\frac{\delta\beta_i}{\delta\mathbf{y}}(S). \tag{33}$$

17) Constant rule:

$$\text{If} \quad Y \neq \emptyset \quad \text{then} \quad \frac{\delta}{\delta Y}K = 0. \tag{34}$$

18) Power rule: If $p(S)$ is a probability mass function with density function $f_p(\mathbf{y})$ then

$$\frac{\delta}{\delta Y}p(S)^n = \begin{cases} \dfrac{n!}{(n-k)!}p(S)^{n-k}f_p(\mathbf{y}_1)\cdots f_p(\mathbf{y}_k) & \text{if} \quad k \leq n \\ 0 & \text{if} \quad k > n \end{cases}. \tag{35}$$

[7]Warning: the first equation (28) has been highly simplified for the sake of clarity. For further details, see [10, pp. 144–151, 157–162]. The notation $\delta/\delta\mathbf{x}$ is a simplified version of a common notation used in physics (see [51, pp. 173–174] and Section IIIB).

7) *Multisensor-Multitarget State Estimation*: Without a Bayes-optimal estimator of the multitarget state X, the information in $f_{k|k}(X \mid Z^{(k)})$ is not available for practical use. Surprisingly, the multitarget analogs of the classical MAP and EAP estimators are not even defined in general; meaning that alternatives must be defined and proved optimal.[8] Once this has been done and we have constructed some estimate

$$\hat{X}_{k|k} = \{\hat{\mathbf{x}}_1,\ldots,\hat{\mathbf{x}}_{\hat{n}}\} \tag{36}$$

of the multitarget state, it follows that the conflicting objectives of detection and estimation have been unified into a single procedure: both the number $\hat{n}$ and states $\hat{\mathbf{x}}_1,\ldots,\hat{\mathbf{x}}_{\hat{n}}$ of targets are determined simultaneously.[9] (For more details, see [30; 34; 29, pp. 40–44; 10, pp. 188–194, 199–205].)

8) *Multisensor-Multitarget Bayes Filter*: We now have a solid basis for the multitarget recursive Bayes filter, (3), (4) of Section I ([29; 10, pp. 237–239].) They cannot even be defined without the set integral $\int \cdot \delta Y$. Without the true multisensor-multitarget likelihood function (i.e., without the set derivative) or without a Bayes-optimal multitarget state estimator, any boast about "Bayes optimality" or having the "true Bayes posterior" would be hollow or false.

It is even more necessary to emphasize that because our approach is Bayesian, the unknown multitarget state X is a random variable rather than a fixed parameter. So, (3), (4) describe the time-evolution

$$\cdots \to \Xi_{k|k} \to \Xi_{k+1|k} \to \Xi_{k+1|k+1} \to \cdots \tag{37}$$

of the random state-set of the multitarget system, where $f_{k|k}(X \mid Z^{(k)})$ and $f_{k+1|k}(X \mid Z^{(k)})$ are the probability distributions of $\Xi_{k|k}$ and $\Xi_{k+1|k}$, respectively.

C. Short History of Multitarget Recursive Bayes Filtering

The concept of multitarget Bayes filtering (3) and (4) is a relatively new one. If target number is assumed known a priori, the earliest exposition appears to be due to Washburn [66]. When the

[8]Consequently and contrary to the assertion of [19, p. 123], because of such difficulties it is not true that "if the target space is discretized into a collection of cells [then] in the continuous case, the cell probabilities can be replaced by densities in the usual way." General multitarget Bayes statistics is not a blind generalization of its discrete-space special case.

[9]Because FISST multitarget state estimation has always addressed unknown target count, it is not true—contrary to the assertion of [61, p. 204]—that FISST "does not provide an explicit method for handling unknown numbers of targets." Moreover, these authors do not themselves provide such a method since their multitarget state estimation procedure specified on [61, pp. 162–163] is erroneous. Contrary to assertion, the multitarget MAP estimator is not defined in general; and multitarget means (expectations) apparently cannot be defined at all.

number n of targets is not known and must be determined along with the individual target states, the earliest work appears to be due to Miller, O'Sullivan, Srivastava, Lanterman, et al. [23, 57, 58]. Their "jump diffusion" approach utilizes solution of stochastic diffusion equations on non-Euclidean manifolds. It is also apparently the only approach to systematically deal with continuous evolution of the multitarget state. (More recently, Lanterman has adopted a random set perspective for the jump diffusion approach [22].) Mahler was apparently the first to systematically deal with the general discrete state-evolution case (Bethel and Paras [5] assume discrete observation and state variables). Portenko et al. used branching-process concepts to model target appearance and disappearance [49]. Challa et al. [6] have shown that the IPDA tracking approach arises directly from the FISST methodology of Sections IIC through IIE, and have established connections with its multitarget extension, JIPDA [44]. In recent years several researchers have implemented equations (3), (4) using particle-systems, Markov chain, and other approximations. Representative instances include Ballantyne, Kouritzin et al. [2], Hue, Le Cadre et al., [12, 13], and Agate et al. [1]. Sidenbladh and Wirkander have applied FISST techniques to ground target tracking [53].

The approach of Stone et al. [61], which essentially consists of citing the multitarget Bayes filter equations (3), (4), is best described as heuristic (see [29, pp. 42, 91–93, footnotes 2, 3, 9], [37, ch. 14, p. 14–24] and [33, pp. 222–223]).

As is acknowledged on [48, pp. 27–28], Kastella's "joint multitarget probabilities (JMP)" approach [18], and the system-level conceptual apparatus surrounding it, are elements of FISST core system-level concepts (e.g. set integrals, multitarget information metrics, multitarget posteriors, joint multitarget state estimators, etc.; see [30 and 28, pp. 256–258]). A JMP itself is just a FISST multitarget posterior density written in the alternative FISST notation of [10, p. 231: $f_{\mathrm{FISST}}(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\} \mid Z) = n! f_{\mathrm{JMP}}(\mathbf{x}_1,\ldots,\mathbf{x}_n \mid Z)$]. Kastella's "multitarget microdensity" approach [19] is likewise a restatement of FISST concepts using random density notation (Section IID1) rather than random set notation for a simple point process (see footnote 12). His assertion in the same paper, that multitarget Bayes filtering requires no "random set concepts," is thereby not only false (see Section IIB; footnote 8; [35, sect. 3.4] and [27, pp. 200–201]) but self-refuting.

D. FISST and Point Process Theory

This section summarizes point process theory [3, 7, 17, 50, 54, 56], and its relationship to FISST. All unreferenced page numbers refer to the textbook by Daly & Vere-Jones [7]. A fuller discussion is on [31, pp. 139–146].

1) *Point Processes*: Let $\delta_{\mathbf{w}}(\mathbf{x})$ be the Dirac delta density concentrated at $\mathbf{w}$ and $\mathbf{1}_S(\mathbf{w})$ the characteristic function of the set $S : \mathbf{1}_S(\mathbf{w}) = 1$ if $\mathbf{w} \in S$ and $\mathbf{1}_S(\mathbf{w}) = 0$ otherwise. A point process on a space Y is a random finite multiset of elements in Y.[10] If $\mathbf{y}_1,\ldots,\mathbf{y}_n$ are distinct elements of Y and the positive integers $\nu_1,\ldots,\nu_n$ are the respective number of copies of $\mathbf{y}_1,\ldots,\mathbf{y}_n$ then equivalent mathematical formulations of the concept of a finite multiset are as follows:

19) finite unordered list $L = \ll \mathbf{y}_1,\ldots,\mathbf{y}_1,\ldots,\mathbf{y}_n,\ldots,\mathbf{y}_n \gg$;

20) density function $\delta(\mathbf{y}) = \nu_1\delta_{\mathbf{y}}(\mathbf{y}_1) + \cdots + \nu_n\delta_{\mathbf{y}}(\mathbf{y}_n)$;

21) measure $N(S) = \nu_1\mathbf{1}_S(\mathbf{y}_1) + \cdots + \nu_n\mathbf{1}_S(\mathbf{y}_n) = |L \cap S| = \int_S \delta(\mathbf{y})d\mathbf{y}$;

22) finite set of pairs $\{(\nu_1,\mathbf{y}_1),\ldots,(\nu_n,\mathbf{y}_n)\}$;

23) density function $\delta(\nu,\mathbf{y}) = \delta_{(\nu,\mathbf{y})}(\nu_1,\mathbf{y}_1) + \cdots + \delta_{(\nu,\mathbf{y})}(\nu_n,\mathbf{y}_n)$;

and so on.[11] In particular, if it is always the case that $\nu_1 = 1,\ldots,\nu_n = 1$ then the above are all equivalent mathematical representations of a finite subset of $\mathbf{Y}$. So, Ξ and $N_\Xi(S) = |\Xi \cap S|$ and $\delta_\Xi(\mathbf{x}) = \Sigma_{\mathbf{w}\in\Xi}\delta_{\mathbf{w}}(\mathbf{x})$ are equivalent mathematical representations of the random finite state-set Ξ, a.k.a. a simple point process.[12] The representation Ξ is more suited for engineering since it is less abstract, represents random multiobject systems as visualizable images, and permits the methodology of Sections IIB3 through IIB7. (However, FISST does encompass multisets, see [10, pp. 194–199].)

2) *Probability Laws of Point Processes*: The statistical behavior of a point process Λ is characterized by its family $j_{\Lambda,1}(\mathbf{x}_1), j_{\Lambda,2}(\mathbf{x}_1,\mathbf{x}_2),\ldots, j_{\Lambda,n}(\mathbf{x}_1,\ldots,\mathbf{x}_n),\ldots$ of Janossy densities (pp. 122–123). The Janossy densities of a simple point process are completely symmetric in all arguments, vanish whenever $\mathbf{x}_i = \mathbf{x}_j$ for some $1 \le i \ne j \le n$ (Prop. 5.4.IV,

[10]More precisely, point processes are random locally finite multisets, but for simplicity's sake I avoid this complication. Point processes can also be defined as random variables on unions of vector spaces [7, p. 121], but this formulation is so restrictive that it is rarely used.

[11]See [17, pp. 5, 16, 412] for the equivalence of point processes and simple point processes of pairs. See [56, pp. 100–102] for the equivalence of random set, density function, and counting measure formalisms. See [3, 50] for the equivalence of random set and random measure formalisms.

[12]Consequently, the approach of [19] is—contrary to the assertion on p. TuB-1 that Bayes multitarget filtering can be accomplished while "requiring no…random set concepts to be introduced"—just FISST rewritten in obfuscated notation. The random "multi-target microdensity" $\rho(\mathbf{x}) = \Sigma_{\mathbf{w}\in X}\delta_{\mathbf{w}}(\mathbf{x}) = \delta_X(\mathbf{x})$ is a random set in random density notation. The "probability density functional" on the microdensity $p\{\rho \mid Y\} = p\{\Sigma_{\mathbf{w}\in X}\delta_{\mathbf{w}}(\mathbf{x}) \mid Y\}$ is a FISST multitarget posterior $f(X \mid Y) = f(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\} \mid Y)$. "Functional integrals" are FISST set integrals: $\int p\{\rho\}D\rho = \int p\{\delta_X\}\delta X = \int p(X)\delta X$. The "expected value of the target density" is the PHD: $\rho(\mathbf{x}) = \int \rho(\mathbf{x})p\{\rho\}D\rho = \int \delta_X(\mathbf{x})p(X)\delta X = D(\mathbf{x})$.

p. 134), and are jointly normalized in the sense that

$$1 = \sum_{n=0}^{\infty} \frac{1}{n!} \int j_{\Lambda,n}(\mathbf{x}_1,\ldots,\mathbf{x}_n) d\mathbf{x}_1 \cdots d\mathbf{x}_n$$

(eqn. 5.4.11, p. 133). So, the FISST multitarget posterior density $f_{k|k}(X \mid Z^{(k)})$ of a random state-set $\Xi_{k|k}$ packages the entire family of Janossy densities of $\Xi_{k|k}$ into a single function defined on a finite-set variable:

$$f_{k|k}(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\} \mid Z^{(k)}) = j_{\Xi,n}(\mathbf{x}_1,\ldots,\mathbf{x}_n). \tag{38}$$

Assume for ease of discussion that $\Lambda = \Xi$ is simple. Then

$$M_{\Xi,1}(S) = E[|\Xi \cap S|] = \int_S E[\delta_\Xi(\mathbf{x})] d\mathbf{x} \tag{39}$$

is the expectation measure or first factorial-moment measure. The expectation density or first factorial-moment density.

$$m_{\Xi,1}(\mathbf{x}) = E[\delta_\Xi(\mathbf{x})] \tag{40}$$

is its density function. Higher order factorial-moment densities

$$m_{\Xi,k}(\mathbf{x}_1,\ldots,\mathbf{x}_k) = \sum_{n=0}^{\infty} \frac{1}{n!} \int j_{\Lambda,k+n}(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{w}_1,\ldots,\mathbf{w}_n) d\mathbf{w}_1 \cdots d\mathbf{w}_n \tag{41}$$

can also be defined (eqn. 5.4.11, p. 133) and, like the $j_{\Lambda,1}, j_{\Xi,2},\ldots,j_{\Xi,n},\ldots$ vanish when $\mathbf{x}_i = \mathbf{x}_j$ for some $1 \le i \ne j \le k$.

III. GENERALIZED FISST MULTITARGET CALCULUS

This section summarizes the generalized FISST multitarget calculus introduced in [31]. The PGFL (Section IIIA) generalizes the belief-mass function of (24), (25). The gradient functional derivative (Section IIIB) generalizes the set derivative of (28). Some of its properties are derived.

A. Probability Generating Functionals

Given a random finite set Ψ of objects in some space Y of such objects and given a measurable subset S of Y let $\mathbf{1}_S(\mathbf{y})$ be the indicator function of S defined by $\mathbf{1}_S(\mathbf{y}) = 1$ if $\mathbf{y} \in S$ and $\mathbf{1}_S(\mathbf{y}) = 0$ otherwise. For any finite subset Y of Y and any real-valued function $h(\mathbf{x})$ define $h^Y = 1$ if $Y = \emptyset$ and

$$h^Y = h(\mathbf{y}_1)\cdots h(\mathbf{y}_n) \tag{42}$$

otherwise, if $Y = \{\mathbf{y}_1,\ldots,\mathbf{y}_n\}$) where $\mathbf{y}_1,\ldots,\mathbf{y}_n$ are distinct. Then the belief-mass function $\beta_\Psi(S)$ of Ψ, equations (24) and (25), can be rewritten as the expected value of $\mathbf{1}_S^\Psi$:

$$\beta_\Psi(S) = \int \mathbf{1}_S^Y f_\Psi(Y) \delta Y. \tag{43}$$

Generalize $\beta_\Psi(S)$ by replacing $\mathbf{1}_S(\mathbf{y})$ with any $h(\mathbf{y})$ such that $h(\mathbf{y}) = h_0(\mathbf{y}) + w_1 \delta_{\mathbf{w}_1}(\mathbf{y}) + \cdots + w_m \delta_{\mathbf{w}_m}(\mathbf{y})$ where $h_0(\mathbf{y})$ has no units of measurement; where $0 \le h_0(\mathbf{y}) \le 1$; where $\delta_{\mathbf{w}}(\mathbf{y})$ is the Dirac delta; where $\mathbf{w}_1,\ldots,\mathbf{w}_m$ are fixed distinct elements of X; and where $w_1,\ldots,w_m$ have the same units of measurement as $\mathbf{y}$. (Note that the definition of h in the earlier paper [36] is a typo; also the restriction on h_0 given here is slightly different.) Then:

DEFINITION 2 (Probability Generating Functionals)

$$G_\Psi[h] = \int h^Y f_\Psi(Y) \delta Y \tag{44}$$

is called the PGFL of Ψ (see [7, pp. 141, 220]). So, $\beta_\Psi(S) = G_\Psi[\mathbf{1}_S]$.

The PGFL is well defined and finite valued because $f_\Psi(\{\mathbf{y}_1,\ldots,\mathbf{y}_i,\ldots,\mathbf{y}_j,\ldots,\mathbf{y}_n\}) = 0$ whenever $\mathbf{y}_i = \mathbf{y}_j$ for $i \ne j$; so undefined products of the form $\delta_{\mathbf{u}}(\mathbf{y})^2$ do not occur.

The intuitive meaning of the PGFL is as follows. Let $\mathsf{Y} = \mathsf{X}$ be single-target state space, $\Psi = \Xi$ a random finite subset of X, and $0 \le h(\mathbf{x}) \le 1$, so that $h(\mathbf{x})$ can be interpreted as the probability of detection or FOV of some sensor. Then it can be shown that $G_\Xi[h]$ is the probability that Ξ is contained in the FOV. Since $h(\mathbf{x})$ is also a fuzzy membership function on $\mathbf{X}$, $G_\Xi[h]$ is a generalization of the belief-mass function $\beta_\Xi(S) = G_\Xi[\mathbf{1}_S]$ from crisp sets S to fuzzy subsets h.

The PGFL shares the following useful property with the belief-mass function. Let $\Xi = \Xi_1 \cup \cdots \cup \Xi_N$ where $\Xi_1,\ldots,\Xi_N$ are statistically independent. Let $G_1[h],\ldots,G_N[h]$ be the respective PGFLs of the $\Xi_1,\ldots,\Xi_N$. Then for all h,

$$G_\Xi[h] = G_1[h]\cdots G_N[h]. \tag{45}$$

B. Functional Derivatives of PGFLs

The gradient derivative (a.k.a. directional or Frechét derivative) of a real-valued function $G(\mathbf{x})$ in the direction of a vector $\mathbf{w}$ is [15, p. 1075]

$$\frac{\partial G}{\partial \mathbf{w}}(\mathbf{x}) = \lim_{\varepsilon \to 0} \frac{G(\mathbf{x} + \varepsilon \cdot \mathbf{w}) - G(\mathbf{x})}{\varepsilon} \tag{46}$$

where for each $\mathbf{x}$ the function $\mathbf{w} \to (\partial G/\partial \mathbf{w})(\mathbf{x})$ is linear and continuous; and so

$$\frac{\partial G}{\partial \mathbf{w}}(\mathbf{x}) = w_1 \frac{\partial G}{\partial w_1}(\mathbf{x}) + \cdots + w_N \frac{\partial G}{\partial w_N}(\mathbf{x})$$

for all $\mathbf{w} = (w_1,\ldots,w_N)$, where the derivatives on the right are ordinary partial derivatives. Likewise, the gradient derivative of a PGFL $G[h]$ in the direction of the function g is

$$\frac{\partial G}{\partial g}[h] = \lim_{\varepsilon \to 0} \frac{G[h + \varepsilon \cdot g] - G[h]}{\varepsilon} \tag{47}$$

where for each h the functional $g \to (\partial G/\partial g)[h]$ is linear and continuous. Gradient derivatives obey the usual "turn the crank" rules of undergraduate calculus, e.g. sum rule, product rule, etc.

PROPOSITION 1 (Set Derivatives are Functional Derivatives) *The set derivative of $\beta_\Xi(S)$ is a gradient derivative of $G_\Xi[h]$*

$$\frac{\delta\beta_\Xi}{\delta\mathbf{x}}(S) = \frac{\partial G_\Xi}{\partial\delta_\mathbf{x}}[\mathbf{1}_S] \tag{48}$$

with $g = \delta_\mathbf{x}$ and $h = \mathbf{1}_S$. Likewise for the iterated derivatives:

$$\frac{\delta\beta_\Xi}{\delta X}(S) = \frac{\delta^n\beta_\Xi}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}(S) = \frac{\partial^n G_\Xi}{\partial\delta_{\mathbf{x}_1}\cdots\partial\delta_{\mathbf{x}_n}}[\mathbf{1}_S] \tag{49}$$

for $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ with $\mathbf{x}_1,\ldots,\mathbf{x}_n$ distinct (see Section VIA).

In physics, gradient derivatives with $g = \delta_\mathbf{x}$ are called "functional derivatives" [51, pp. 173–174] and [31, pp. 140–141]. Using the simplified version of this physics notation used in FISST, we define the following.

DEFINITION 3 (Functional Derivatives) The functional derivatives of a PGFL $G[h]$ are

$$\frac{\delta^0 G}{\delta\mathbf{x}^0}[h] = G[h], \qquad \frac{\delta G}{\delta\mathbf{x}}[h] = \frac{\partial G}{\partial\delta_\mathbf{x}}[h] \tag{50}$$

$$\frac{\delta^n G}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}[h] = \frac{\partial^n G}{\partial\delta_{\mathbf{x}_1}\cdots\partial\delta_{\mathbf{x}_n}}[h] \tag{51}$$

So, the multitarget probability density of random state-set Ξ is

$$f_\Xi(X) = \frac{\delta^n\beta_\Xi}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}(\emptyset) = \frac{\delta^n G_\Xi}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}[0]. \tag{52}$$

C. Some Properties of Functional Derivatives

We need the following facts. First, let $G[h] = h(\mathbf{x}_0)$ for all h and fixed $\mathbf{x}_0 \in \mathbf{X}$. Then from (47) and (50)

$$\frac{\delta G}{\delta\mathbf{x}}[h] = \delta_\mathbf{x}(\mathbf{x}_0). \tag{53}$$

for all $\mathbf{x}$. Second, if $G[h] = \int h(\mathbf{x})f(\mathbf{x})d\mathbf{x}$ then likewise

$$\frac{\delta G}{\delta\mathbf{x}}[h] = f(\mathbf{x}). \tag{54}$$

Third, if $f(\mathbf{x})$ is absolutely bounded and $G[h] = G_\Xi[f\cdot h]$ for all h then (see proof in Section VIB)

$$\frac{\delta G}{\delta\mathbf{x}}[h] = f(\mathbf{x})\cdot\frac{\delta G_\Xi}{\delta\mathbf{x}}[f\cdot h]. \tag{55}$$

IV. MULTITARGET MOMENT DENSITIES AND THE PHD

This section 1) defines the concept of a multitarget moment density function $D_{k|k}(X \mid Z^{(k)})$ (Section IVA), 2) shows how to construct it using the set or functional derivative (Section IVB), 3) shows that the first-order multitarget moment $D_{k|k}(\{\mathbf{x}\} \mid Z^{(k)})$ is the PHD (Section IVC), and 4) provides examples of PHDs (Section IVD). With the exception of the proofs and examples, this material originally appeared in [10, pp. 168–170].

A. Multitarget Moment Densities

DEFINITION 4 (Multitarget Moment Densities) Given a random state-set Ξ with multitarget probability density $f_\Xi(X)$, its multitarget moment density $D_\Xi(X)$ is

$$D_\Xi(X) = \int f_\Xi(X\cup W)\delta W \tag{56}$$

where the rightmost integral is a set integral (Section IIB5). In particular, if $\Xi = \Xi_{k|k}$ and $f_\Xi(X) = f_{k|k}(X \mid Z^{(k)})$ then write

$$D_{k|k}(X \mid Z^{(k)}) = \int f_{k|k}(X\cup W \mid Z^{(k)})\delta W. \tag{57}$$

Here $D_\Xi(\emptyset) = D_{k|k}(\emptyset \mid Z^{(k)}) = 1$ and the set integral is well defined. That is, $\int D_\Xi(X\cup W)\delta W$ always has the same units of measurement as X and so there is no incommensurability of units (footnote 6 of Section IIB5 or [29, p. 39]). An intuitive interpretation: for any $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$, $D_\Xi(X)$ is the probability (density) that n of the targets in Ξ have states $\mathbf{x}_1,\ldots,\mathbf{x}_n$. (See [7, eqn. 5.4.13, p. 133].)

DEFINITION 5 (Multitarget Moments of Order n) The functions of n vector variables $\mathbf{x}_1,\ldots,\mathbf{x}_n$

$$D_\Xi(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}), \qquad D_{k|k}(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\} \mid Z^{(k)}) \tag{58}$$

are nth-order multitarget moment densities. If $\delta_X(\mathbf{x}) = \Sigma_{\mathbf{w}\in X}\delta_\mathbf{w}(\mathbf{x})$ the first multitarget moment is

$$D_\Xi(\{\mathbf{x}\}) = \int f_\Xi(\{\mathbf{x}\}\cup W)\delta W = \int \delta_X(\mathbf{x})f_\Xi(X)\delta X. \tag{59}$$

The second of these last two equations is a special case of Proposition 2a, setting $h(\mathbf{x}) = \delta_X(\mathbf{x})$. Note that $D_\Xi(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}) = 0$ when $\mathbf{x}_i = \mathbf{x}_j$ for some $i \neq j$ since the same convention holds for $f_\Xi(X)$.

B. Computing Multitarget Moments Using Set and Functional Derivatives

This section proves facts necessary for Theorems 5 and 6.

THEOREM 1 (The Multitarget Moment Density is a Set Derivative) *Let $f_\Xi(X)$ be a multitarget probability density and let $G_\Xi[h] = \int h^X f_\Xi(X)\delta X$ and $\beta_\Xi(S) = G_\Xi[\mathbf{1}_S]$ be its PGFL and belief-mass function, respectively. Then*

$$D_\Xi(X) = \frac{\delta\beta_\Xi}{\delta X}(\mathsf{X}) = \frac{\delta^n\beta_\Xi}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}(\mathsf{X}) = \frac{\delta^n G_\Xi}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}[1] \tag{60}$$

where X *denotes the entire (single-target) state space. In particular, the first-moment density is*

$$D_\Xi(\{\mathbf{x}\}) = \frac{\delta\beta_\Xi}{\delta\mathbf{x}}(\mathsf{X}) = \frac{\delta G_\Xi}{\delta\mathbf{x}}[1]. \quad (61)$$

The proof is in Section VIC.

PROPOSITION 2 *Let $f(Y)$ be a multitarget probability density and $G[h] = \int h^Y f(Y)\delta Y$ its PGFL. Let us be given a real-valued function $h(\mathbf{y})$ and $Y = \{\mathbf{y}_1,\ldots,\mathbf{y}_m\}$ with $\mathbf{y}_1,\ldots,\mathbf{y}_m$ distinct.* a) *Define $h(Y)$ as $h(Y) = 0$ if $|Y| = 0$ and*

$$h(Y) = \sum_{i=1}^{m} h(\mathbf{y}_i)$$

otherwise. Then

$$\int h(Y)f(Y)\delta Y = \int h(\mathbf{y})\frac{\delta G}{\delta\mathbf{y}}[1]d\mathbf{y}. \quad (62)$$

b) *Define $h(Y)$ as $h(Y) = 0$ if $|Y| \leq 1$ and*

$$h(Y) = \sum_{1\leq i\neq j\leq m} h(\mathbf{y}_i)\cdot h(\mathbf{y}_j)$$

otherwise. Then

$$\int h(Y)f(Y)\delta Y = \int h(\mathbf{y}_1)h(\mathbf{y}_2)\frac{\delta^2 G}{\delta\mathbf{y}_1\delta\mathbf{y}_2}[1]d\mathbf{y}_1 d\mathbf{y}_2. \quad (63)$$

The proofs are in Sections VID, VIE, respectively. The next proposition shows how to use functional derivatives to compute the PHD of a transformed PGFL (necessary for Theorem 5). A functional transformation is a mapping Φ that transforms any function h into another function $\Phi[h]$.

PROPOSITION 3 *Let $h \to \Phi[h]$ be a functional transformation such that $\Phi[1](\mathbf{x}) = 1$ identically for all $\mathbf{x}$. Given a functional $G_\Xi[h]$, define the new functional $G[h]$ by $G[h] = G_\Xi[\Phi[h]]$ for all h. Then the PHD of $G[h]$ is*

$$D(\mathbf{x}) = \int D_{\mathbf{w}}(\mathbf{x})D_\Xi(\mathbf{w})d\mathbf{w} \quad (64)$$

where for each fixed $\mathbf{w}$, $\Phi_{\mathbf{w}}$ is the functional defined by $\Phi_{\mathbf{w}}[h] = \Phi[h](\mathbf{w})$; and where the PHD of $\Phi_{\mathbf{w}}[h]$ is

$$D_{\mathbf{w}}(\mathbf{x}) = \frac{\delta\Psi_{\mathbf{w}}}{\delta\mathbf{x}}[1].$$

The proof can be found in Section VIF.

C. The PHD is a 1st-Order Multitarget Statistical Moment

This section shows that 1) the multitarget moment density repackages the point process factorial moment densities of Section IID2, 2) the first-order multitarget moment is the PHD, and so 3) the PHD is a first statistical moment in a mathematically recognized sense. It also provides an inversion formula for transforming multitarget moment densities $D_{k|k}(X \mid Z^{(k)})$ into their multitarget posteriors $f_{k|k}(X \mid Z^{(k)})$.

1) *The PHD is the First-Order Multitarget Moment Density.* We begin by proving the equivalence of the PHD and the first-order multitarget moment density (eqn. (40)).

THEOREM 2 (The 1st-Order Multitarget Moment is the PHD) *Let $\Xi_{k|k}$ denote the state-set whose multitarget distribution is $f_{k|k}(X \mid Z^{(k)})$, and let S be some region of state space. Then*

$$\int_S D_{k|k}(\{\mathbf{x}\} \mid Z^{(k)})d\mathbf{x} = E[|\Xi_{k|k}\cap S|]. \quad (65)$$

So, the first-order moment equals the PHD almost everywhere:

$$D_{k|k}(\{\mathbf{x}\} \mid Z^{(k)}) = D_{k|k}(\mathbf{x} \mid Z^{(k)}). \quad (66)$$

The proof of the first assertion is in Section VIG. The second assertion is easy. For by (65) and Definition 1, $\int_S D_{k|k}(\{\mathbf{x}\} \mid Z^{(k)})d\mathbf{x} = \int_S D_{k|k}(\mathbf{x} \mid Z^{(k)})d\mathbf{x}$ for all measurable S and so $D_{k|k}(\{\mathbf{x}\} \mid Z^{(k)}) = D_{k|tk}(\mathbf{x} \mid Z^{(k)})$ almost everywhere.

2) *The PHD is the First-Order Multitarget Statistical Moment.* Notice that $D_\Xi(X)$ repackages the family of point process factorial moment densities ((41) of Section IID2) into a single function of a finite-set variable:

$$\begin{aligned}
D_\Xi(\{\mathbf{x}_1,\ldots,\mathbf{x}_k\}) &= \int f_\Xi(X\cup W)\delta W \\
&= \sum_{n=0}^{\infty}\frac{1}{n!}\int f_\Xi(\{\mathbf{x}_1,\ldots,\mathbf{x}_k\}\cup\{\mathbf{w}_1,\ldots,\mathbf{w}_n\})d\mathbf{w}_1\cdots d\mathbf{w}_n \\
&= \sum_{n=0}^{\infty}\frac{1}{n!}\int j_{\Xi,k+n}(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{w}_1,\ldots,\mathbf{w}_n)d\mathbf{w}_1\cdots d\mathbf{w}_n \\
&= m_{\Xi,k}(\mathbf{x}_1,\ldots,\mathbf{x}_k). \quad (67)
\end{aligned}$$

In other words, from the point of view of point process theory the $D_{k|k}(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\} \mid Z^{(k)})$ are standard statistical moments of the random state-set $\Xi_{k|k}$. In particular the PHD is a first-order moment statistic in a well-understood mathematical sense.

Engineers tend to react with puzzlement to the idea that the first-order moment of a multitarget posterior is a density function, expecting instead to see some set $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ of state vectors.[13] The mystery is due to the fact that the set integral $\int X\cdot f_{k+1|k}(X \mid Z^{(k)})\delta X$ required to define a conventional expectation does not exist because the class of finite sets is not even a vector space [29, p. 41]. As a result, one has no choice but to construct a multitarget moment indirectly.

[13]The problem of defining a set-valued multitarget first moment is unsolved (see [13, pp. 153–154] for a fuller discussion).

That is, first specify some function θ that transforms state-sets $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ into elements $\theta(X)$ of a suitably well-behaved vector space, and which also preserves additive set-theoretic structure: $\theta(X \cup Y) = \theta(X) + \theta(Y)$ whenever $X \cap Y \neq \emptyset$ (see [10, p. 179]). Then $\int \theta(X) f_{k+1|k}(X \mid Z^{(k)})\delta X$ is an indirect first-order expectation. The PHD arises from the specific choice $\theta(X) = \delta_X$ where $\delta_X(\mathbf{x}) = \delta_{\mathbf{x}_1}(\mathbf{x}) + \cdots + \delta_{\mathbf{x}_n}(\mathbf{x})$.

3) *Inversion Formula for Multitarget Moments*. The multitarget moment density is defined in terms of the multitarget posterior. Conversely, the multitarget posterior can be recovered from the multitarget moment density.

THEOREM 3 (Inversion Formula)

$$D_\Xi(X) = \int (-1)^{|W|} D_\Xi(X \cup W)\delta W. \tag{68}$$

The proof is in Section VIH.

D. Examples of PHDs and PGFLs

This section computes the PHD and PGFL of a multitarget Poisson density (Section IVD1) and of a random state-set (Section IVD2); and shows that the PHD is a fuzzy subset of discrete state spaces X (Section IVD3).

1) *Example 1 (Computing PHD and PGFL of a Random State-Set)*. This example is required for the proofs of Theorems 5 and 6. Suppose that we know that a) there are n random track vectors $\mathbf{X}_1,\ldots,\mathbf{X}_n$, b) the probability distribution of the ith track $\mathbf{X}_i$ is $f_i(\mathbf{x})$, and c) there is a probability $1-\pi_i$ that the ith track may not exist. Then the random state-set is $\Xi = \Xi_1 \cup \cdots \cup \Xi_n$ where $\Xi_i = \{\mathbf{X}_i\} \cap \emptyset_i$ and where $\emptyset_i$ is a random subset of state space X such that $\emptyset_i = \emptyset$ with probability $1-\pi_i$ and $\emptyset_i = \mathsf{X}$ with probability π_i. Consequently $\Xi_i = \emptyset$ (no target) with probability $1-\pi_i$ and, otherwise, $\Xi_i = \{\mathbf{X}_i\}$ (one target with state $\mathbf{X}_i$) with probability π_i. Assume that the $\mathbf{X}_1,\ldots,\mathbf{X}_n,\emptyset_1,\ldots,\emptyset_n$ are statistically independent. Then:

PROPOSITION 4 (PHD and PGFL of a Random State-Set) *The belief-mass function, PHD, and PGFL of X are*

$$\beta_\Xi(S) = (1-\pi_1+\pi_1 p_1(S))\cdots(1-\pi_n+\pi_n p_n(S)) \tag{69}$$

$$D_\Xi(\mathbf{x}) = \pi_1 f_1(\mathbf{x}) + \cdots + \pi_n f_n(\mathbf{x}) \tag{70}$$

$$G_\Xi[h] = (1-\pi_1+\pi_1 p_1[h])\cdots(1-\pi_n+\pi_n p_n[h]) \tag{71}$$

where $p_i[h] = \int h(\mathbf{x}) f_i(\mathbf{x}) d\mathbf{x}$ *and* $p_i(S) = p_i[\mathbf{1}_S]$.

To see this, first note that the belief-mass function of Ξ_i is

$$\begin{aligned}\beta_i(S) &= \Pr(\Xi_i \subseteq S) = \Pr(\Xi_i = \emptyset) + \Pr(\Xi_i \neq \emptyset, \mathbf{X}_i \in S)\\ &= \Pr(\Xi_i = \emptyset) + \Pr(\Xi_i \neq \emptyset)\Pr(\mathbf{X}_i \in S)\\ &= 1-\pi_i+\pi_i p_i(S).\end{aligned}$$

So, the belief-mass function of Ξ is

$$\begin{aligned}\beta_\Xi(S) &= \Pr(\Xi_1 \cup\cdots\cup \Xi_n \subseteq S) = \Pr(\Xi_1 \subseteq S)\cdots\Pr(\Xi_n \subseteq S)\\ &= (1-\pi_1+\pi_1 p_1(S))\cdots(1-\pi_n+\pi_n p_n(S)).\end{aligned}$$

From the "turn the crank" formulas for the FISST calculus (Section IIB6) the first set derivative of $\beta_\Xi(S)$ is

$$\frac{\delta\beta_\Xi}{\delta\mathbf{x}}(S) = \sum_{i=1}^{n} \beta_1(S)\cdots\pi_i f_i(\mathbf{x})\cdots\beta_i(S).$$

Substitute $S = \mathsf{X}$ and use Theorem 1 and $\beta_i(\mathsf{X}) = 1$:

$$D_\Xi(\mathbf{x}) = \pi_1 f_1(\mathbf{x}) + \cdots + \pi_n f_n(\mathbf{x}).$$

The expected number of targets is therefore

$$N_\Xi = \int D_\Xi(\mathbf{x}) d\mathbf{x} = \pi_1 + \cdots + \pi_n.$$

It is n if and only if $\pi_i = 1$ for all i, i.e., if and only if all targets are known to exist with certainty.

Finally, from (29) and the "turn the crank" rules for the set derivative it is easy to show that the multitarget distribution of Ξ_i is $f_i(X) = 0$ if $|X| \geq 2$ and, otherwise, $f_i(\emptyset) = 1-\pi_i$ and $f_i(\{\mathbf{x}\}) = \pi_i f_i(\mathbf{x})$. So, the PGFL of Ξ_i is

$$\begin{aligned}G_i[h] &= \int h^X f_i(X)\delta X = f_i(\emptyset) + \int h(\mathbf{x}) f_i(\{\mathbf{x}\}) d\mathbf{x}\\ &= 1-\pi_i+\pi_i\int h(\mathbf{x}) f_i(\mathbf{x}) d\mathbf{x} = 1-\pi_i+\pi_i p_i[h].\end{aligned}$$

So, the PGFL of Ξ is

$$\begin{aligned}G_\Xi[h] &= G_1[h]\cdots G_n[h]\\ &= (1-\pi_1+\pi_1 p_1[h])\cdots(1-\pi_n+\pi_n p_n[h]).\end{aligned}$$

2) *Example 2 (PHD and PGFL of Multitarget Poisson Distributions)*. A multitarget density $f_I(X)$ is Poisson if

$$f_I(X) = e^{-\lambda} I(\mathbf{x}_1)\cdots I(\mathbf{x}_n) \tag{72}$$

for any $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ with $\mathbf{x}_1,\ldots,\mathbf{x}_n$ distinct, where $I(\mathbf{x})$ is a density function (the "intensity") and $\lambda = \int I(\mathbf{x}) d\mathbf{x}$ (the "parameter"). (See [7, eqn. 7.4.10, p. 225].)

The following establishes some basic properties of multitarget Poisson distributions required for Theorem 6.

PROPOSITION 5 *Let $f_I(X)$ be a multitarget Poisson density with intensity $I(\mathbf{x})$ and parameter λ. Then* a) $\int f_I(X)\delta X = 1$, b) *the multitarget moment density* *of* $f_I(X)$ *is* $D_I(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}) = I(\mathbf{x}_1)\cdots I(\mathbf{x}_n)$, *and* c) *the PGFL of* $f_I(X)$ *is* $G_I[h] = e^{I[h]-\lambda}$ *where* $I[h] = \int h(\mathbf{x})I(\mathbf{x})d\mathbf{x}$.

The proof of these assertions is in Section VII. In particular, the first multitarget moment of a multitarget Poisson distribution—its PHD—is just its intensity function: $D_I(\{\mathbf{x}\}) = I(\mathbf{x})$. In other words, a multitarget Poisson distribution and its PHD contain exactly the same information.

For Proposition 5a, $f_I(X)$ is a multitarget probability density because

$$\begin{aligned}\int f_I(X)\delta X &= \sum_{n=0}^{\infty}\frac{1}{n!}\int f_I(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\})d\mathbf{x}_1\cdots d\mathbf{x}_n\\ &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{1}{n!}\int I(\mathbf{x}_1)\cdots I(\mathbf{x}_n)d\mathbf{x}_1\cdots d\mathbf{x}_n\\ &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} = e^{-\lambda}e^{\lambda} = 1.\end{aligned}$$

For Proposition 5b, the multitarget moment density of $f_I(X)$ is

$$\begin{array}{ccccccc} \cdots\rightarrow & f_{k|k}(X\mid Z^{(k)}) & \overset{\text{time prediction}}{\rightarrow} & f_{k+1|k}(X\mid Z^{(k)}) & \overset{\text{Bayes' rule}}{\rightarrow} & f_{k+1|k+1}(X\mid Z^{(k+1)}) & \rightarrow\cdots\\ & \downarrow & & \downarrow & & \downarrow & \\ \cdots\rightarrow & D_{k|k} & \overset{\text{predictor?}}{\rightarrow} & D_{k+1|k} & \overset{\text{corrector?}}{\rightarrow} & D_{k+1|k+1} & \rightarrow\cdots \end{array}$$

$$\begin{aligned}D_I(X) &= \int f_I(X\cup W)\delta W\\ &= \sum_{i=0}^{\infty}\frac{1}{i!}\int f_I(\{\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{w}_1,\ldots,\mathbf{w}_i\})d\mathbf{x}_1\cdots d\mathbf{x}_i\\ &= e^{-\lambda}I(\mathbf{x}_1)\cdots I(\mathbf{x}_n)\cdot\sum_{i=0}^{\infty}\frac{1}{i!}\int I(\mathbf{w}_1)\cdots I(\mathbf{w}_i)d\mathbf{w}_1\cdots d\mathbf{w}_i\\ &= I(\mathbf{x}_1)\cdots I(\mathbf{x}_n).\end{aligned}$$

For Proposition 5c, the PGFL of $f_I(X)$ is

$$\begin{aligned}G_I[h] &= \sum_{n=0}^{\infty}\frac{1}{i!}\int\left(\prod_{i=1}^{n}h(\mathbf{x}_i)\right)f_I(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\})d\mathbf{x}_1\cdots d\mathbf{x}_n\\ &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{1}{n!}\int h(\mathbf{x}_1)\cdots h(\mathbf{x}_n)\\ &\quad\cdot I(\mathbf{x}_1)\cdots I(\mathbf{x}_n)d\mathbf{x}_1\cdots d\mathbf{x}_n\\ &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\int h(\mathbf{x})I(\mathbf{x})d\mathbf{x}\right)^n\\ &= \exp(-\lambda+\int h(\mathbf{x})I(\mathbf{x})d\mathbf{x}).\end{aligned}$$

3) *Example 3 (PHD When State Space is Discrete).* Suppose that X is a finite set of target-state cells x. Let Ξ be a random state-set and note that $f_\Xi(X) = \Pr(\Xi = X)$. Then:

$$\begin{aligned}D_\Xi(x) &= \sum_{X\supseteq\{x\}}\Pr(\Xi = X)\\ &= \sum_{X}\Pr(x\in X,\Xi = X) = \Pr(x\in\Xi).\end{aligned}\tag{73}$$

This result demonstrates that when X is discrete the PHD is: 1) I. R. Goodman's one-point covering function $\mu_\Xi(x) = \Pr(x\in\Xi)$ of the random set Ξ [9], and 2) a fuzzy subset of X since $\mu_\Xi(x)$ is a fuzzy membership function on X. So, for continuous X the PHD represents the zero-probability event $\Pr(\mathbf{x}\in\Xi)$ just as the density $f_{\mathbf{X}}(\mathbf{x})$ of a continuous random vector $\mathbf{X}$ represents $\Pr(\mathbf{X} = \mathbf{x})$.

V. PHD FILTER EQUATIONS

The purpose of this section is to derive equations for the "predictor" and "corrector" steps of a recursive filter for the PHD, as indicated in the following diagram

where the top row portrays the the multitarget Bayes filter, the downward-pointing arrows indicate the collapse of multitarget posteriors into their PHDs, and the bottom row portrays the to-be-determined approximate filter for PHDs. As noted in Section IA, the approximate filter should have the following properties.[14] For any k,

24) the predictor $D_{k|k}(\mathbf{x})\rightarrow D_{k+1|k}(\mathbf{x})$ is lossless: if $D_{k|k}(\mathbf{x}) = D_{k|k}(\mathbf{x}\mid Z^{(k)})$ then $D_{k+1|k}(\mathbf{x}) = D_{k+1|k}(\mathbf{x}\mid Z^{(k)})$;

25) the corrector $D_{k+1|k}(\mathbf{x})\rightarrow D_{k+1|k+1}(\mathbf{x})$ is lossless: if $D_{k+1|k}(\mathbf{x}) = D_{k+1|k}(\mathbf{x}\mid Z^{(k)})$ then $D_{k+1|k+1}(\mathbf{x}) = D_{k+1|k+1}(\mathbf{x}\mid Z^{(k+1)})$; and

26) $D_{k|k}(\mathbf{x}\mid Z^{(k)})$ and $D_{k+1|k}(\mathbf{x}\mid Z^{(k)})$ are "best-fit" approximations of $f_{k|k}(X\mid Z^{(k)})$ and $f_{k+1|k}(X\mid Z^{(k)})$, respectively.

We will be able to derive a information-lossless predictor (Theorem 5) and we will be able to show that PHDs are best-fit approximations of their corresponding multitarget posteriors in an information-theoretic sense (Theorem 4).

[14] An approximate filter that satisfies these three conditions is "Bayes-closed" in the terminology of Kulhavý [21] and Iltis [14].

Unfortunately, the corrector equations we derive (Theorems 6 and 7) will only be approximate and therefore not lossless.

The section is organized as follows. Section VA proves a "best fit" approximation characterization of the PHD. Section VB is devoted to the PHD predictor equation and Section VC to the PHD corrector equation in the single-sensor case. The corrector equation for the multisensor case requires further approximation. As explained in Section VD, this is because the Poisson assumption used in the single-sensor case results in a multisensor corrector equation too complex to be tractable. To get a tractable equation we must use an additional approximation, explained in Section VE, which models the entire sensor suite as a single "pseudosensor." This leads, in Section VF, to an approximate PHD corrector equation for the multisensor case.

A. Best Poisson Approximation of Multitarget Posterior

Suppose that $D(\mathbf{x}) = D_{k|k}(\mathbf{x} \mid Z^{(k)})$ is the PHD of the multitarget posterior distribution $f_{k|k}(X \mid Z^{(k)})$, and let $f_D(X)$ be the multitarget Poisson density whose intensity is also $D(\mathbf{x})$ (72). The following shows that the multitarget Poisson distribution that has the closest fit to $f_{k|k}(X \mid Z^{(k)})$, in an information-theoretic sense,[15] is just $f_D(X)$. It is based on the concept of multitarget Kullback-Leibler discrimination [40; 10, pp. 205–209, 297–303], i.e., the PHD of a multitarget posterior is the least lossy collapse of the multitarget posterior to a density function on single-target state space.

THEOREM 4 (Best Poisson Approximation) *Let $f_{k|k}(X) = f_{k|k}(X \mid Z^{(k)})$ be a multitarget posterior distribution and let $f_I(X)$ be a multitarget Poisson density with intensity-function $I(\mathbf{x})$. Then the multitarget Kullback-Leibler discrimination*

$$K(I) = \int f_{k|k}(X) \log\left(\frac{f_{k|k}(X)}{f_I(X)}\right) \delta X \tag{74}$$

is minimized if and only if $I(\mathbf{x}) = D_{k|k}(\mathbf{x} \mid Z^{(k)})$.

The proof is in Section VII and [35, p. 161].

[15] In [46] it was proposed that the Poisson best-fit approximation be that $I = I*$ that minimizes the multitarget square-error

$$\varepsilon(I)^2 = \int (f_{k|k}(X) - f_i(X))^2 \delta X.$$

However, the indicated set integral is ill defined because of the incommensurability of units of measurement problem (see footnote 6 of Section IIB5). Contrary to the last sentence of [46, Sect. 3.3], normalization with respect to a fixed reference unit does not resolve this difficulty since the solution of the optimization problem $I* = \arg\inf_I \varepsilon(I)^2$ will then depend on the choice of reference unit.

B. Recursive Time-Update of PHD

The purpose of this section is to derive an equation for the prediction $D_{k|k}(\mathbf{x} \mid Z^{(k)}) \to D_{k+1|k}(\mathbf{x} \mid Z^{(k)})$ of the PHD to the time of the next observation-collection. Because of Theorem 2 and Definition 4 we know that

$$D_{k|k}(\mathbf{x} \mid Z^{(k)}) = \int f_{k|k}(\{\mathbf{x}\} \cup W \mid Z^{(k)}) \delta W$$

$$D_{k+1|k}(\mathbf{x} \mid Z^{(k)}) = \int f_{k+1|k}(\{\mathbf{x}\} \cup W \mid Z^{(k)}) \delta W.$$

We need an equation for the predictor $D_{k|k}(\mathbf{x}) \to D_{k+1|k}(\mathbf{x})$ such that $D_{k+1|k}(\mathbf{x}) = D_{k+1|k}(\mathbf{x} \mid Z^{(k)})$ if $D_{k|k}(\mathbf{x}) = D_{k|k}(\mathbf{x} \mid Z^{(k)})$. This is achieved as follows. First, show that the PGFL $G_{k+1|k}[h]$ of $f_{k+1|k}(X \mid Z^{(k)})$ is a transformation $G_{k+1|k}[h] = e_h \cdot G_{k|k}[\Phi[h]]$ of the PGFL of $f_{k|k}(X \mid Z^{(k)})$. Second, get a formula for $D_{k+1|k}(\mathbf{x} \mid Z^{(k)})$ in terms of $D_{k|k}(\mathbf{x} \mid Z^{(k)})$ by taking a functional derivative $\delta/\delta\mathbf{x}$ of this equation.

We begin by specifying notation and assumptions regarding between-measurements multitarget motion:

27) Motion of individual targets: $f_{k+1|k}(\mathbf{y} \mid \mathbf{x})$ is the single-target Markov transition density;

28) Disappearance of existing targets: $p_{S,k+1|k}(\mathbf{x})$ is the probability that a target with state $\mathbf{x}$ at time-step k will survive in time-step $k+1$, hereafter abbreviated as $p_S(\mathbf{x})$;

29) Spawning of new targets by existing targets: $b_{k+1|k}(Y \mid \mathbf{x})$ is the likelihood that a group of new targets with state-set Y will be spawned at time-step $k+1$ by a single target that had state $\mathbf{x}$ at time-step k, and its PHD is denoted by $b_{k+1|k}(\mathbf{y} \mid \mathbf{x}) = \int b_{k+1|k}(\{\mathbf{y}\} \cup W \mid \mathbf{x}) \delta W$;

30) Appearance of completely new targets: $b_{k+1|k}(Y)$ is the likelihood that new targets with state-set Y will enter the scene at time-step $k+1$, and $b_{k+1|k}(\mathbf{y}) = \int b_{k+1|k}(\{\mathbf{y}\} \cup W) \delta W$ is its PHD.

Now, apply the methodology of Sections, IIB3, IIB4, and IIB5. To determine the multitarget Markov transition density $f_{k+1|k}(Y \mid X)$ we must first specify a multitarget motion model, i.e., a formula for the predicted random state-set $\Xi_{k+1|k}$ in terms of the previous multitarget state-set $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ at time-step k. Assume that

$$\Xi_{k+1|k} = \Xi(X) \cup \Psi(X) \cup \Psi_0$$

where

$$\Xi(X) = \Xi(\mathbf{x}_1) \cup \cdots \cup \Xi(\mathbf{x}_n)$$

is the set of surviving targets, where

$$\Psi(X) = \Psi(\mathbf{x}_1) \cup \cdots \cup \Psi(\mathbf{x}_n)$$

is the set of spawned targets, and where Ψ_0 is the set of entering targets. Here, $\Xi(\mathbf{x}) = \emptyset$ (target disappearance) with probability $1 - p_S(\mathbf{x})$ and $\Xi(\mathbf{x}) = \{\mathbf{Y}(\mathbf{x})\}$ with probability $p_S(\mathbf{x})$, where $\mathbf{Y}(\mathbf{x})$ is a random vector whose distribution is $f_{k+1|k}(\mathbf{y} \mid \mathbf{x})$.

The multitarget distributions of $\Psi(\mathbf{x})$ and Ψ_0 are $b_{k+1|k}(Y \mid \mathbf{x})$ and $b_{k+1|k}(Y)$, respectively. Further, assume that

31) $\Xi(\mathbf{x}_1),\ldots,\Xi(\mathbf{x}_n)$, $\Psi(\mathbf{x}_1),\ldots,\Psi(\mathbf{x}_n)$, Ψ_0 are statistically independent, i.e., between-measurements target motions are independent when conditioned on the previous state X.

Then:

THEOREM 5 (General Law of Motion for PHDs) *Abbreviate $D_{k+1|k}(\mathbf{x}) = D_{k+1|k}(\mathbf{x} \mid Z^{(k)})$ and $D_{k|k}(\mathbf{x}) = D_{k|k}(\mathbf{x} \mid Z^{(k)})$. Then the predictor equation for the PHD is*

$$D_{k+1|k}(\mathbf{x}) = b_{k+1|k}(\mathbf{x}) + \int (p_S(\mathbf{w}) f_{k+1|k}(\mathbf{x} \mid \mathbf{w}) + b_{k+1|k}(\mathbf{x} \mid \mathbf{w})) D_{k|k}(\mathbf{w}) d\mathbf{w}. \tag{75}$$

This result is proved in Section VIJ.[16] As already noted, it is based on taking a functional derivative of the equation

$$G_{k+1|k}[h] = e_h \cdot G_{k|k}[\Phi[h]] \tag{76}$$

where we show that

$$\Phi[h] = (1 - p_S + p_S p_h) b_h \tag{77}$$

$$p_h(\mathbf{x}) = \int h(\mathbf{x}) f_{k+1|k}(\mathbf{w} \mid \mathbf{x}) d\mathbf{w} \tag{78}$$

where

$$b_h(\mathbf{x}) = \int h^X b_{k+1|k}(X \mid \mathbf{x}) \delta X \tag{79}$$

$$e_h = \int h^X b_{k+1|k}(X) \delta X \tag{80}$$

are the PGFLs of $b_{k+1|k}(X \mid \mathbf{x})$ and $b_{k+1|k}(X)$, respectively.

Now, let $N_{k|k} = \int D_{k|k}(\mathbf{x} \mid Z^{(k)}) d\mathbf{x}$ be the expected number of targets at time-step k. From Theorem 5, the expected number $N_{k+1|k} = \int D_{k+1|k}(\mathbf{x} \mid Z^{(k)}) d\mathbf{x}$ of targets at the next time-step is

$$N_{k+1|k} = B_{k+1|k} + \int (p_S(\mathbf{w}) + B_{k+1|k}(\mathbf{w})) D_{k|k}(\mathbf{w}) d\mathbf{w} \tag{81}$$

where $B_{k+1|k}(\mathbf{w}) = \int b_{k+1|k}(\mathbf{x} \mid \mathbf{w}) d\mathbf{x}$ is the expected number of targets spawned at time-step $k+1$ by a target that had state $\mathbf{w}$ at time-step k; and $B_{k+1|k} = \int b_{k+1|k}(\mathbf{x}) d\mathbf{x}$ is the expected number of new targets entering at time-step $k+1$.

The following is a special case of (75) that employs a Poisson model to account for new targets.

COROLLARY 1 (Law of Motion for PHDs Assuming Poisson Target Births) *Suppose that the multitarget density functions $b_{k+1|k}(X \mid \mathbf{x})$ and $b_{k+1|k}(X)$ are Poisson:*

$$b_{k+1|k}(X \mid \mathbf{w}) = \exp(-\mu_{k+1|k}(\mathbf{w})) \Pi_{\mathbf{x} \in X} \mu_{k+1|k}(\mathbf{w}) s_{k+1|k}(\mathbf{x} \mid \mathbf{w})$$

$$b_{k+1|k}(X) = \exp(-\mu_{k+1|k}) \Pi_{\mathbf{x} \in X} \mu_{k+1|k} s_{k+1|k}(\mathbf{x})$$

where $\mu_{k+1|k}(\mathbf{w}) \geq 0$, $\mu_{k+1|k} \geq 0$, and $s_{k+1|k}(\mathbf{x} \mid \mathbf{w})$ and $s_{k+1|k}(\mathbf{x})$ are probability densities. Then:

$$D_{k+1|k}(\mathbf{x}) = \mu_{k+1|k} s_{k+1|k}(\mathbf{x}) + \int (p_S(\mathbf{w}) f_{k+1|k}(\mathbf{x} \mid \mathbf{w}) + \mu_{k+1|k}(\mathbf{w}) s_{k+1|k}(\mathbf{x} \mid \mathbf{w})) D_{k|k}(\mathbf{w}) d\mathbf{w} \tag{82}$$

and

$$N_{k+1|k} = \lambda_{k+1|k} + \int (p_S(\mathbf{w}) + \mu_{k+1|k}(\mathbf{w})) D_{k|k}(\mathbf{w}) d\mathbf{w}. \tag{83}$$

This formula follows immediately since the PHD of a Poisson distribution $b_{k+1|k}(X \mid \mathbf{w})$ is its intensity $b_{k+1|k}(\mathbf{x} \mid \mathbf{w}) = \mu_{k+1|k}(\mathbf{w}) s_{k+1|k}(\mathbf{x} \mid \mathbf{w})$); and likewise the PHD of $b_{k+1|k}(X)$ is its intensity $b_{k+1|k}(\mathbf{x}) = \mu_{k+1|k} s_{k+1|k}(\mathbf{x})$ (Proposition 5). Finally:

COROLLARY 2 (Simplest Law of Motion for PHDs) *Suppose that there are no target appearances ($b_{k+1|k}(Y \mid \mathbf{x}) = b_{k+1|k}(Y) = 0$) or disappearances ($p_S(\mathbf{w}) = 1$). Then*

$$D_{k+1|k}(\mathbf{x}) = \int f_{k+1|k}(\mathbf{x} \mid \mathbf{w}) D_{k|k}(\mathbf{w}) d\mathbf{w} \tag{84}$$

$$N_{k+1|k} = N_{k|k}. \tag{85}$$

That is, if target number does not change between measurement collections then the time-evolution of the PHD is governed by the same law of motion as that which governs the between-measurements time-evolution of the posterior density of any single target in the multitarget system.

C. Single-Sensor Bayes-Update of PHD

The purpose of this section is to derive an equation for the PHD corrector $D_{k+1|k}(\mathbf{x} \mid Z^{(k)}) \rightarrow D_{k+1|k+1}(\mathbf{x} \mid Z^{(k+1)})$. Because of (56) of Definition 4

$$D_{k+1|k}(\mathbf{x} \mid Z^{(k)}) = \int f_{k+1|k}(\{\mathbf{x}\} \cup W \mid Z^{(k)}) \delta W$$

$$D_{k+1|k+1}(\mathbf{x} \mid Z^{(k+1)}) = \int f_{k+1|k+1}(\{\mathbf{x}\} \cup W \mid Z^{(k+1)}) \delta W.$$

We would like to find an equation for the corrector $D_{k+1|k}(\mathbf{x}) \rightarrow D_{k+1|k+1}(\mathbf{x})$ such that $D_{k+1|k+1}(\mathbf{x}) = D_{k+1|k+1}(\mathbf{x} \mid Z^{(k+1)})$ if $D_{k+1|k}(\mathbf{x}) = D_{k+1|k}(\mathbf{x} \mid Z^{(k)})$. This is not possible because the formula for $f_{k+1|k+1}(X \mid Z^{(k+1)})$ is too complicated to permit closed-form formulas if $f_{k+1|k}(X \mid Z^{(k)})$ is arbitrary. To get around this assume that $f_{k+1|k}(X \mid Z^{(k)})$ is approximately Poisson.

We begin by specifying notation and assumptions:

[16] Theorem 5 is easily extended to the higher order multitarget moments. Specifically, predict $D_{k|k}(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\} \mid Z^{(k)})$ to $D_{k+1|k}(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\} \mid Z^{(k)})$ by applying Theorem 5 to each of its arguments $\mathbf{x}_1,\ldots,\mathbf{x}_n$ in turn.

32) Single-target measurement-generation: $L_{k+1,\mathbf{z}}(\mathbf{x}) = f_{k+1}(\mathbf{z} \mid \mathbf{x})$ is the sensor likelihood function, hereafter abbreviated as $L_{\mathbf{z}}(\mathbf{x})$;

33) Probability of detection (sensor FOV): $p_{D,k+1}(\mathbf{x},\mathbf{x}^*)$ is the probability that an observation will be collected at time-step $k+1$ from a target with state $\mathbf{x}$ if the sensor has state $\mathbf{x}^*$ at that time-step, hereafter abbreviated as $p_D(\mathbf{x})$;

34) Poisson false alarms: at time-step $k+1$ the sensor collects an average number $\lambda_{k+1}(\mathbf{x}^*)$ of Poisson-distributed false alarms whose spatial distribution is the probability density $c_{k+1}(\mathbf{z} \mid \mathbf{x}^*)$, and these are hereafter abbreviated as λ and $c(\mathbf{z})$.

Now apply the methodology described in Sections IIB3, IIB4, IIB5. Let $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_k\}$ be the state-set of the multitarget system at time-step $k+1$. Then the random observation-set Σ collected from these targets has the form

$$\Sigma_{k+1} = \Sigma(\mathbf{x}_1) \cup \cdots \cup \Sigma(\mathbf{x}_n) \cup \Theta$$

where $\Sigma(\mathbf{x}_i)$ is the observation-set produced by the target with state $\mathbf{x}_i$; and where Θ is the Poisson false alarm process. Specifically, $\Sigma(\mathbf{x}) = \emptyset$ (no observation) with probability $1 - p_D(\mathbf{x})$ and $\Sigma(\mathbf{x}) = \{\mathbf{Z}(\mathbf{x})\}$ with probability $p_D(\mathbf{x})$, where $\mathbf{Z}(\mathbf{x})$ is a random vector whose distribution is $f_{k+1}(\mathbf{z} \mid \mathbf{x})$. Because Θ is Poisson, by Proposition 5c its PGFL is $G_\Theta[g] = e^{\lambda c[g]-\lambda}$ where $c[g] = \int g(\mathbf{z})c(\mathbf{z})d\mathbf{z}$. It is also assumed that:

35) $\Sigma(\mathbf{x}_1),\ldots,\Sigma(\mathbf{x}_n),\Theta$ are independent, i.e., observations are independent conditionally on the multitarget state X. Then:

THEOREM 6 (Single-Sensor Bayes Update Formula for PHD) *Let $Z_{k+1} = \{\mathbf{z}_1,\ldots,\mathbf{z}_{k+1}\}$ and abbreviate*

$$D_{k+1|k}(\mathbf{x}) = D_{k+1|k}(\mathbf{x} \mid Z^{(k)})$$

$$D_{k+1|k+1}(\mathbf{x}) = D_{k+1|k+1}(\mathbf{x} \mid Z^{(k+1)})$$

$$D_{k+1|k}[h] = \int h(\mathbf{x})D_{k+1|k}(\mathbf{x})d\mathbf{x}.$$

Assume that $f_{k+1|k}(X \mid Z^{(k)})$ is approximately Poisson, i.e.

$$f_{k+1|k}(X \mid Z^{(k)}) \cong e^{-\mu}\mu^n s(\mathbf{x}_1)\cdots s(\mathbf{x}_n) \tag{86}$$

for some $\mu \geq 0$ and probability density $s(\mathbf{x})$, in which case $D_{k+1|k}(\mathbf{x} \mid Z^{(k)}) = \mu \cdot s(\mathbf{x})$ where $s[h] = \int h(\mathbf{x})s(\mathbf{x})d\mathbf{x}$. Then the PHD approximate Bayes corrector equation is

$$D_{k+1|k+1}(\mathbf{x}) \cong F_{k+1}(Z_{k+1} \mid \mathbf{x})D_{k+1|k}(\mathbf{x}) \tag{87}$$

where

$$F_{k+1}(Z \mid \mathbf{x}) = \sum_{\mathbf{z}\in Z_{k+1}} \frac{p_D(\mathbf{x})L_{\mathbf{z}}(\mathbf{x})}{\lambda c(\mathbf{z}) + D_{k+1|k}[p_D L_{\mathbf{z}}]} + 1 - p_D(\mathbf{x}). \tag{88}$$

This result is proved in Section VIK. It is based on the following construction. By Proposition 5c the assumption that $f_{k+1|k}(X \mid Z^{(k)})$ is approximately Poisson can be restated as

$$G_{k+1|k}[h] = e^{\mu s[h]-\mu}. \tag{89}$$

Define the two-variable PGFL $F[g,h]$ as

$$F[g,h] = \int\int h^X g^Z f_{k+1}(Z \mid X)f_{k+1|k}(X \mid Z^{(k)})\delta X \delta Z. \tag{90}$$

Then it will be shown that

$$F[g,h] = e^{\lambda c[g]-\lambda}G_{k+1|k}[\Phi[g,h]] \tag{91}$$

where

$$\Phi[g,h] = (1 - p_D + p_D p_g)h \tag{92}$$

$$p_g(\mathbf{x}) = \int g(\mathbf{z})f_{k+1}(\mathbf{z} \mid \mathbf{x})d\mathbf{z}. \tag{93}$$

The data-updated PHD is derived by noting that the denominator of Bayes' rule

$$f_{k+1|k+1}(X \mid Z^{(k+1)}) = \frac{f_{k+1}(Z_{k+1} \mid X)f_{k+1|k}(X \mid Z^{(k)})}{f_{k+1}(Z_{k+1} \mid Z^{(k)})} \tag{94}$$

can be expressed as iterated functional derivatives of $F[g,1]$, whereas its numerator can be expressed as iterated functional derivatives of

$$\frac{\delta F}{\delta \mathbf{x}}[g,1].$$

Consider now two special cases. The first, Corollary 3, is the result reported in earlier papers [31, 38, 42, 43].[17]

COROLLARY 3 (PHD Data-Update for Constant Probability of Detection) *Suppose that the sensor has an infinite FOV, i.e. $p_D(\mathbf{x}) = p_D$ is constant. Then:*

$$D_{k+1|k+1}(\mathbf{x}) \cong F_{k+1}(Z_{k+1} \mid \mathbf{x})D_{k+1|k}(\mathbf{x}) \tag{95}$$

where

$$F_{k+1}(Z \mid \mathbf{x}) = \sum_{\mathbf{z}\in Z_{k+1}} \frac{p_D L_{\mathbf{z}}(\mathbf{x})}{\lambda c(\mathbf{z}) + p_D D_{k+1|k}[L_{\mathbf{z}}]} + 1 - p_D. \tag{96}$$

Also, the expected number of targets is, approximately,

$$N_{k+1|k+1} \cong \sum_{\mathbf{z}\in Z} \frac{p_D D_{k+1|k}[L_{\mathbf{z}}]}{\lambda c(\mathbf{z}) + p_D D_{k+1|k}[L_{\mathbf{z}}]} + (1 - p_D)N_{k+1|k}. \tag{97}$$

COROLLARY 4 (PHD Data-Update for No Missed Detections and No False Alarms) *Suppose that $p_D(\mathbf{x}) = 1$ and $\lambda_{k+1} = 0$. Then (87) becomes*

$$D_{k+1k+1}(\mathbf{x}) \cong \sum_{\mathbf{z}\in Z_{k+1}} D_{k+1|k+1}(\mathbf{x} \mid \mathbf{z}) \tag{98}$$

[17] *ERRATUM:* In the papers [8, 24, 31, 38, 42], the second term of (96) was erroneously reported to have the factor $(1-p_D)/(1-(1-p_D)N_{k+1|k})$. The correct factor is $1-p_D$. Also the requirement that $p_D > 1 - N_{k+1|k}^{-1}$ is no longer necessary.

where

$$D_{k+1|k+1}(\mathbf{x} \mid \mathbf{z}) = \frac{p_D(\mathbf{x})L_{\mathbf{z}}(\mathbf{x})}{D_{k+1|k}[p_D L_{\mathbf{z}}]} D_{k+1|k}(\mathbf{x}) \tag{99}$$

is a Bayes' rule-like update of $D_{k+1|k}(\mathbf{x})$ *using* $\mathbf{z}$.

Note ("Weak Evidence Accrual"): If we write $Z_{k+1} = \{\mathbf{z}_1,\ldots,\mathbf{z}_m\}$ then (95) can be rewritten as

$$D_{k+1|1k+1}(\mathbf{x}) \cong \alpha_0 D_{k+1|k}(\mathbf{x}) + \sum_{i=1}^{m} \alpha_i D_{k+1|k+1}(\mathbf{x} \mid \mathbf{z}_i) \tag{100}$$

for nonnegative weights $\alpha_0,\alpha_1,\ldots,\alpha_m$. This additive property is essentially what Stein, Winter, and Tenney called Weak Evidence Accrual (WEA) [59, 60]. WEA thus is a linearizing side-effect of collapsing a multitarget posterior into its PHD, rather than an explicit evidence accrual method.

D. Difficulties With the Multisensor Case

This section shows why the single-sensor PHD corrector cannot be immediately generalized to the multisensor case.

Suppose that at time-step $k+1$ there are s sensors with respective state variables $\mathbf{x}^{*[1]},\ldots,\mathbf{x}^{*[s]}$ that collect respective conditionally independent observations $\mathbf{z}^{[1]},\ldots,\mathbf{z}^{[s]}$ from respective observation spaces $\mathbf{Z}^{[1]},\ldots,\mathbf{Z}^{[s]}$. Assume the following abbreviations. First, likelihood functions:

$$L^{[1]}_{\mathbf{z}^{[1]}}(\mathbf{x}) = f^{[1]}_k(\mathbf{z}^{[1]} \mid \mathbf{x},\mathbf{x}^{*[1]}),\ldots,L^{[s]}_{\mathbf{z}^{[s]}}(\mathbf{x}) = f^{[s]}_k(\mathbf{z}^{[s]} \mid \mathbf{x},\mathbf{x}^{*[s]}). \tag{101}$$

Second, probabilities of detection (FOVs):

$$p^{[1]}_D(\mathbf{x}) = p^{\{1]}_D(\mathbf{x},\mathbf{x}^{*[1]}),\ldots,p^{[s]}_D(\mathbf{x}) = p^{[s]}_D(\mathbf{x},\mathbf{x}^{*[s]}). \tag{102}$$

Third, average numbers of false alarms:

$$\lambda^{[1]} = \lambda^{[1]}_{k+1},\ldots,\lambda^{[s]} = \lambda^{[s]}_{k+1}. \tag{103}$$

Fourth, false alarm spatial distributions:

$$c^{[1]}(\mathbf{z}^{[1]}) = c^{[1]}_{k+1}(\mathbf{z}^{[1]}),\ldots,c^{[s]}(\mathbf{z}^{[s]}) = c^{[s]}_{k+1}(\mathbf{z}^{[s]}). \tag{104}$$

Assume now that the predicted multitarget posterior $f_{k+1|k}(X \mid Z^{(k)})$ is approximately Poisson. Then mathematical complexities prevent equ. (87) from being extended to the multisensor case. For example, in the two-sensor case the analog of the PGFL in (90) is

$$F[g_1,g_2,h] = \int h^X g_1^{Z^{[1]}} g_2^{Z^{[2]}} f^{[1]}_{k+1}(Z^{[1]} \mid X) f^{[2]}_{k+1}(Z^{[2]} \mid X) \cdot f_{k+1|k}(X \mid Z^{(k)}) \delta X \delta Z^{[1]} \delta Z^{[2]}$$

which reduces to

$$F[g_1,g_2,h] = e^{\lambda^{[1]}c^{[1]}[g_1]-\lambda^{[1]}} \cdot e^{\lambda^{[2]}c^{[2]}[g_2]-\lambda^{[2]}} \cdot G_{k+1|k}[\Psi[g_1,g_2.h]]$$

where

$$\Psi[g_1,g_2,h] = (1 - p^{[1]}_D + p^{[1]}_D p^{[1]}_{g_1})(1 - p^{[2]}_D + p^{[2]}_D p^{[2]}_{g_2})h$$

$$p^{[1]}_{g_1}(\mathbf{x}) = \int g_1(\mathbf{z}^{[1]}) f^{[1]}_{k+1}(\mathbf{z}^{[1]} \mid \mathbf{x}) d\mathbf{z}^{[1]}$$

$$p^{[2]}_{g_2}(\mathbf{x}) = \int g_2(\mathbf{z}^{[2]}) f^{[2]}_{k+1}(\mathbf{z}^{[2]} \mid \mathbf{x}) d\mathbf{z}^{[2]}$$

$$c^{[1]}[g_1] = \int g_1(\mathbf{z}^{[1]}) c^{[1]}(\mathbf{z}^{[1]}) d\mathbf{z}^{[1]}$$

$$c^{[2]}[g_2] = \int g_2(\mathbf{z}^{[2]}) c^{[1]}(\mathbf{z}^{[2]}) d\mathbf{z}^{[2]}.$$

Even with the Poisson assumption $G_{k+1|k}[h] = e^{\mu s[h]-\mu}$ the resulting formula for $D_{k+1|k+1}(\mathbf{x} \mid Z^{(k+1)})$ is impractical because of the quadratic dependence on g_1 and g_2 in $F[g_1,g_2,h]$.

One way out of this difficulty, proposed in [31, 38, 42], is as follows. Abbreviate $Z^{[j]} = Z^{[j]}_{k+1}$ for all $j = 1,\ldots,s$. Assuming that $f_{k+1|k}(X \mid Z^{(k)})$ is approximately Poisson, update it with $Z^{[1]}$ to get $f_{k+1|k+1}(X \mid Z^{(k)},Z^{[1]})$. Assume that this posterior is also approximately Poisson, and update it with $Z^{[2]}$ to get $f_{k+1|k+1}(X \mid Z^{(k)},Z^{[1]},Z^{[2]})$. And so on, until we get $f_{k+1|k+1}(X \mid Z^{(k)})$. This $(s+1)$-fold application of the Poisson approximation is rather drastic. A weaker approximation would be desirable.

ERRATUM: In an ill-advised attempt at a less drastic approximation for the multisensor case, in recent conference papers [35, 41] we proposed a "pseudo-sensor approximation." This approximation is incorrect because, assuming it, single targets generate multiple rather than single observations, in violation of the measurement model assumed earlier.

E. Multisensor Bayes Update of PHD

Given this, Theorem 6 generalizes as follows.

THEOREM 7 (Multisensor Bayes-Update Formula for PHD) *Let the next multisensor observation-set* Z_{k+1} *be*

$$Z_{k+1} = Z^{[1]}_{k+1} \cup \cdots \cup Z^{[s]}_{k+1}. \tag{105}$$

Then the data-updated PHD is, approximately,

$$D_{k+1|k}(\mathbf{x}) \cong F^{[1]}_{k+1}(Z^{[1]}_{k+1} \mid \mathbf{x}) \cdots F^{[s]}_{k+1}(Z^{[s]}_{k+1} \mid \mathbf{x}) \tag{106}$$

where

$$F^{[j]}_{k+1}(Z^{[j]}_{k+1} \mid \mathbf{x}) = \sum_{\mathbf{z}^{[j]} \in Z^{[j]}_{k+1}} \frac{p^{[j]}_D(\mathbf{x}) L^{[j]}_{\mathbf{z}^{[j]}}(\mathbf{x})}{\tilde{c}^{[j]}(\mathbf{z}^{[j]}) + D_{k+1|k}[p^{[j]}_D L^{[j]}_{\mathbf{z}^{[j]}}]} + 1 - p^{[j]}_D(\mathbf{x}) \tag{107}$$

where the various quantities are as defined in Section VE.

This result results from substitution into equ. (87).

VI. MATHEMATICAL PROOFS

A. Proof of Proposition 1 (The Set Derivative is a Functional Derivative)

The gradient derivative of $G_\Xi[h]$ is by equ.'s (44) and (47)

$$\frac{\partial G_\Xi}{\partial g}[h] = \int \left(\frac{\partial}{\partial g} h^X\right) f_\Xi(X)\delta X$$

where the product rule and (47) leads to

$$\begin{aligned}\frac{\partial}{\partial g} h^X &= \frac{\partial}{\partial g} h(\mathbf{x}_1)\cdots h(\mathbf{x}_n)\\ &= \sum_{i=1}^n h(\mathbf{x}_1)\cdots\left(\frac{\partial}{\partial g}h(\mathbf{x}_i)\right)\cdots h(\mathbf{x}_n)\\ &= \sum_{i=1}^n h(\mathbf{x}_1)\cdots g(\mathbf{x}_i)\cdots h(\mathbf{x}_n) \qquad (108)\end{aligned}$$

for $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$. Notice that $\mathbf{1}_S(\mathbf{x}) = \Delta_\mathbf{x}(S)$ where $\Delta_\mathbf{x}(S)$ is the Dirac probability measure defined by $\Delta_\mathbf{x}(S) = 1$ if $\mathbf{x} \in S$ and $\Delta_\mathbf{x}(S) = 0$ if otherwise. Also notice that since the set derivative of any probability measure $p(S)$ is its corresponding density function $f(\mathbf{x})$, the set derivative of $\Delta_\mathbf{x}(S)$ is the Dirac delta function $\delta_\mathbf{x}(\mathbf{y})$. Now, the set derivative of $\beta_\Xi(S)$ is

$$\frac{\delta\beta_\Xi}{\delta\mathbf{x}}(S) = \int \left(\frac{\delta}{\delta\mathbf{x}}\mathbf{1}_S^X\right) f_\Xi(X)\delta X \qquad (109)$$

where from the product rule for set derivatives (Section IIB7)

$$\begin{aligned}\frac{\delta}{\delta\mathbf{x}}\mathbf{1}_S^X &= \frac{\delta}{\delta\mathbf{x}}\Delta_{\mathbf{x}_1}(S)\cdots\Delta_{\mathbf{x}_n}(S)\\ &= \sum_{i=1}^n \Delta_{\mathbf{x}_1}(S)\cdots\left(\frac{\delta}{\delta\mathbf{x}}\Delta_{\mathbf{x}_i}(S)\right)\cdots\Delta_{\mathbf{x}_n}(S)\\ &= \sum_{i=1}^n \mathbf{1}_S(\mathbf{x}_1)\cdots\delta_{\mathbf{x}_i}(\mathbf{x})\cdots\mathbf{1}_S(\mathbf{x}_n).\end{aligned}$$

So, the claimed relationship immediately follows by substituting $g = \delta_\mathbf{x}$ and $h = \mathbf{1}_S$ into (115).

B. Proof of Equation (55)

From (46) and (50), if $f(\mathbf{x}) = 0$ then the result is trivial; if $f(\mathbf{x}) > 0$:

$$\begin{aligned}\frac{\delta G}{\delta\mathbf{x}}[h] &= \lim_{\varepsilon\to 0}\frac{G_\Xi[f\cdot h + \varepsilon\cdot f\cdot\delta_\mathbf{x}] - G_\Xi[f\cdot h]}{\varepsilon}\\ &= f(\mathbf{x})\cdot\lim_{\varepsilon\to 0}\frac{G_\Xi[f\cdot h + \varepsilon\cdot f(\mathbf{x})\cdot\delta_\mathbf{x}] - G_\Xi[f\cdot h]}{\varepsilon\cdot f(\mathbf{x})}\\ &= f(\mathbf{x})\cdot\frac{\delta G_\Xi}{\delta\mathbf{x}}[f\cdot h].\end{aligned}$$

C. Proof of Theorem 1 (Computing Multitarget Moments Using Set and Functional Derivatives)

This was originally proved on [38, pp. 113–114]. Here we give a simpler proof based on functional derivatives. From Proposition 1 we know that

$$\frac{\delta\beta_\Xi}{\delta X}(S) = \frac{\delta^n G_\Xi}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}[\mathbf{1}_S]$$

for all S and all $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$. So, it is enough to show that

$$\frac{\delta^n G_\Xi}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}[1] = D_\Xi(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}).$$

First note that, differentiating under the integral sign,

$$\frac{\delta^n G_\Xi}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}[h] = \int\left(\frac{\delta^n}{\delta\mathbf{x}_1\cdots\delta\mathbf{x}_n}h^X\right) f_\Xi(X)\delta X.$$

Now, for each fixed $X = \{\mathbf{y}_1,\ldots,\mathbf{y}_n\}$

$$\frac{\delta}{\delta\mathbf{x}_1}h^X = \sum_{i=1}^n h(\mathbf{y}_1)\cdots\delta_{\mathbf{x}_1}(\mathbf{y}_i)\cdots h(\mathbf{y}_n)$$

and so, setting $h = 1$, from (61) and (66) we get

$$\begin{aligned}\frac{\delta G_\Xi}{\delta\mathbf{x}_1}[1] &= \int\left(\sum_{\mathbf{y}\in X}\delta_{\mathbf{y}_i}(\mathbf{x}_1)\right) f_\Xi(X)\delta X\\ &= \int f_\Xi(\{\mathbf{x}_1\}\cup W)\delta W = D_\Xi(\{\mathbf{x}_1\})\end{aligned}$$

as claimed. Next,

$$\frac{\delta^2}{\delta\mathbf{x}_2\delta\mathbf{x}_1}h^X = \sum_{1\le i_1\ne i_2\le n} h(\mathbf{y}_1)\cdots\delta_{\mathbf{x}_1}(\mathbf{y}_{i_1})\cdots\delta_{\mathbf{x}_2}(\mathbf{y}_{i_2})\cdots h(\mathbf{y}_n)$$

and so setting setting $h = 1$ we get

$$\begin{aligned}\frac{\delta^2 G_\Xi}{\delta\mathbf{x}_2\delta\mathbf{x}_1}[1] &= \int\left(\sum_{1\le i_1\ne i_2\le n}\delta_{\mathbf{x}_1}(\mathbf{y}_{i_1})\delta_{\mathbf{x}_2}(\mathbf{y}_{i_2})\right) f(X)\delta X\\ &= \sum_{n=2}^\infty \frac{2C_{n,2}}{n!}\int f_\Xi(\{\mathbf{y}_1,\mathbf{y}_2,\mathbf{w}_1,\ldots,\mathbf{w}_{n-2}\})d\mathbf{w}_1\cdots d\mathbf{w}_{n-2}\\ &= \sum_{j=0}^\infty \frac{1}{j!}\int f_\Xi(\{\mathbf{y}_1,\mathbf{y}_2,\mathbf{w}_1,\ldots,\mathbf{w}_j\})d\mathbf{w}_1\cdots d\mathbf{w}_j\\ &= D_\Xi(\{\mathbf{y}_1,\mathbf{y}_2\})\end{aligned}$$

as claimed. Continue inductively in this fashion.

D. Proof of Proposition 2a

By definition of $h(Y)$,

$$\int \left(\sum_{\mathbf{y}\in Y} h(\mathbf{y}) \right) f(Y)\delta Y$$
$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int (h(\mathbf{y}_1) + \cdots + h(\mathbf{y}_n)) f(\{\mathbf{y}_1,\ldots,\mathbf{y}_n) d\mathbf{y}_1 \cdots d\mathbf{y}_n$$
$$= \sum_{n=1}^{\infty} \frac{n}{n!} \int h(\mathbf{y}) f(\{\mathbf{y},\mathbf{w}_1 \ldots,\mathbf{w}_{n-1}\}) d\mathbf{w}_1 \cdots d\mathbf{w}_{n-1} d\mathbf{y}$$
$$= \int h(\mathbf{y}) \left(\sum_{j=1}^{\infty} \frac{1}{j!} \int f(\{\mathbf{y},\mathbf{w}_1 \ldots,\mathbf{w}_j\}) d\mathbf{w}_1 \cdots d\mathbf{w}_j \right) d\mathbf{y}$$
$$= \int h(\mathbf{y}) D(\mathbf{y}) d\mathbf{y}.$$

E. Proof of Proposition 2b

By definition of $h(Y)$,

$$\int h(Y) f(Y)\delta Y$$
$$= \sum_{n=2}^{\infty} \frac{1}{n!} \int \left(\sum_{1\le i\neq j\le n} h(\mathbf{y}_i)\cdot h(\mathbf{y}_j) \right) f(\{\mathbf{y}_1,\ldots,\mathbf{y}_n\}) d\mathbf{y}_1 \cdots d\mathbf{y}_n$$
$$= \sum_{n=2}^{\infty} \frac{2C_{n,2}}{n!} \int h(\mathbf{y}_1) h(\mathbf{y}_2) f(\{\mathbf{y}_1,\mathbf{y}_2,\mathbf{w}_1,\ldots,\mathbf{w}_{n-2}\})$$
$$\cdot d\mathbf{w}_1 \cdots d\mathbf{w}_{n-2} d\mathbf{y}_1 d\mathbf{y}_2$$
$$= \int h(\mathbf{y}_1) h(\mathbf{y}_2)$$
$$\cdot \left(\sum_{j=0}^{\infty} \frac{1}{j!} \int f(\{\mathbf{y}_1,\mathbf{y}_2,\mathbf{w}_1,\ldots,\mathbf{w}_j\} d\mathbf{w}_1 \cdots d\mathbf{w}_j \right) d\mathbf{y}_1 d\mathbf{y}_2$$
$$= \int h(\mathbf{y}_1) h(\mathbf{y}_2) \left(\int f(\{\mathbf{y}_1,\mathbf{y}_2\} \cup W)\delta W \right) d\mathbf{y}_1 d\mathbf{y}_2$$
$$= \int h(\mathbf{y}_1) h(\mathbf{y}_2) D(\{\mathbf{y}_1,\mathbf{y}_2\}) d\mathbf{y}_1 d\mathbf{y}_2$$

where $D(\{\mathbf{y}_2,\mathbf{y}_2\})$ is the second-order multitarget moment of $f(Y)$. But by Theorem 1

$$D(\{\mathbf{y}_1,\mathbf{y}_2\}) = \frac{\delta^2 G}{\delta\mathbf{y}_1 \delta\mathbf{y}_2}[1].$$

F. Proof of Proposition 3 (PHD of a Functional Transformation of a PGFL)

The functional derivative of the PGFL $G[h] = G_\Xi[\Phi[h]]$ is

$$\frac{\delta G}{\delta\mathbf{x}}[h] = \int \left(\frac{\delta}{\delta\mathbf{x}} \Psi[h]^X \right) f_\Xi(X)\delta X$$

where for $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$

$$\frac{\delta}{\delta\mathbf{x}} \Psi[h]^X = \frac{\delta}{\delta\mathbf{x}} \Psi[h](\mathbf{x}_1) \cdots \Psi[h](\mathbf{x}_n)$$
$$= \sum_{i=1}^{n} \Psi[h](\mathbf{x}_1) \cdots \frac{\delta\Psi_{\mathbf{x}_i}}{\delta\mathbf{x}}[h] \cdots \Psi[h](\mathbf{x}_n).$$

Substituting $h = 1$ into this equation and using $\Phi[1](\mathbf{x}) = 1$,

$$\left[\frac{\delta}{\delta\mathbf{x}} \Psi[h]^X \right]_{h=1} = \sum_{i=1}^{n} \frac{\delta\Psi_{\mathbf{x}_i}}{\delta\mathbf{x}}[1] = \sum_{i=1}^{n} D_{\mathbf{x}_i}(\mathbf{x}).$$

The claimed result follows from (61) and (66):

$$D_\Xi(\mathbf{x}) = \frac{\delta G_\Xi}{\delta\mathbf{x}}[1] = \int \left(\sum_{\mathbf{w}\in X} D_{\mathbf{w}}(\mathbf{x}) \right) f_\Xi(X)\delta X$$
$$= \int D_{\mathbf{w}}(\mathbf{x}) \cdot D_\Xi(\mathbf{w}) d\mathbf{w}.$$

G. Proof of Theorem 2 (PHD is the First-Order Multitarget Moment)

If $f(X)$ and $D(\{\mathbf{x}\})$ are the multitarget posterior and first-order multitarget moment of Ξ, we are to show that

$$E[|\Xi \cap S|] = \int_S D(\{\mathbf{x}\}) d\mathbf{x}$$

for any region S of state space. From (59) we have $D(\mathbf{x}) = \int \delta_X(\mathbf{x}) f(X)\delta X$ where $\delta_X(\{\mathbf{x}\}) = \Sigma_{\mathbf{w}\in X} \delta_{\mathbf{w}}(\mathbf{x})$. So, if $\mathbf{1}_S(\mathbf{x})$ is the set indicator function we get, as claimed:

$$\int_S D(\{\mathbf{x}\}) d\mathbf{x} = \int \mathbf{1}_S(\mathbf{x}) D(\{\mathbf{x}\}) d\mathbf{x}$$
$$= \int \mathbf{1}_S(\mathbf{x}) \int \delta_X(\mathbf{x}) f(X)\delta X d\mathbf{x}$$
$$= \int \int \mathbf{1}_S(\mathbf{x}) \delta_X(\mathbf{x}) d\mathbf{x} f(X)\delta X$$
$$= \int |X \cap S| f(X)\delta X = E[|\Xi \cap S|].$$

H. Proof of Theorem 3 (Moment Inversion Formula)

Equations (67) and [7, eqn. 5.4.12, p. 133] give

$$f_\Xi(\{\mathbf{x}_1,\ldots,\mathbf{x}_k\}) = j_{\Xi,k}(\mathbf{x}_1,\ldots,\mathbf{x}_k)$$
$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int m_{\Xi,k+n}(\mathbf{x}_1,\ldots,\mathbf{x}_k,\mathbf{w}_1,\ldots,\mathbf{w}_n) d\mathbf{w}_1 \cdots d\mathbf{w}_n$$
$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int D_\Xi(\{\mathbf{x}_1,\ldots,\mathbf{x}_k\} \cup \{\mathbf{w}_1,\ldots,\mathbf{w}_n\}) d\mathbf{w}_1 \cdots d\mathbf{w}_n$$
$$= \int (-1)^{|W|} D_\Xi(X \cup W)\delta W.$$

I. Proof of Theorem 4 (Best Poisson Approximation)

We are to prove that $f_I(X) = e^{-\lambda} I(\mathbf{x}_1)\cdots I(\mathbf{x}_n)$ minimizes

$$K(I) = \int f_{k|k}(X) \log\left(\frac{f_{k|k}(X)}{f_I(X)}\right) \delta X$$

if and only if $I(\mathbf{x}) = D_{k|k}(\mathbf{x})$, where $D_{k|k}(\mathbf{x}) = D_{k|k}(\mathbf{x} \mid Z^{(k)})$ is the PHD of $f_{k|k}(X) = f_{k|k}(X \mid Z^{(k)})$. First, notice that

$$K(I) = \int f_{k|k}(X) \log f_{k|k}(X) \delta X - \int f_{k|k}(X) \log\left(e^{-\lambda} \prod_{\mathbf{x}\in X} I(\mathbf{x})\right) \delta X$$

and so

$$K(I) = K_1 + \lambda - \int (f_{k|k}(X) \sum_{\mathbf{x}\in X} \log I(\mathbf{x})) \delta X$$

where K_1 is constant (has no functional dependence on I). Let $I(\mathbf{x}) = \lambda \cdot s(\mathbf{x})$ and $D_{k|k}(\mathbf{x}) = \lambda_{k|k} \cdot s_{k|k}(\mathbf{x})$ where $s(\mathbf{x})$ and $s_{k|k}(\mathbf{x})$ are probability distributions. Then

$$\int \left(f_{k|k}(X) \sum_{\mathbf{x}\in X} \log I(\mathbf{x})\right) \delta X = \lambda_{k|k} \int s_{k|k}(\mathbf{x}) \log s(\mathbf{x}) d\mathbf{x} + \lambda_{k|k} \log \lambda.$$

For, abbreviate $h(\mathbf{x}) = \log I(\mathbf{x})$. Then Proposition 2a yields

$$\int (f_{k|k}(X) \sum_{\mathbf{x}\in X} h(\mathbf{x})) \delta X = \int h(\mathbf{x}) D_{k|k}(\mathbf{x}) d\mathbf{x} = \int D_{k|k}(\mathbf{x}) \log I(\mathbf{x}) d\mathbf{x}.$$

Substituting $I(\mathbf{x}) = \lambda \cdot s(\mathbf{x})$ and $D_{k|k}(\mathbf{x}) = \lambda_{k|k} \cdot s_{k|k}(\mathbf{x})$, we get the claimed result. So

$$K(I) = K_2 + \lambda_{k|k} \int s_{k|k}(\mathbf{x}) \log s(\mathbf{x}) d\mathbf{x} + \lambda - \lambda_{k|k} \log \lambda = K_2 + \lambda_{k|k} K(s_{k|k}; s) + \lambda - \lambda_{k|k} \log \lambda$$

where K_2 is another constant and $K(s_{k|k}; s) \geq 0$ is the Kullback-Leibler discrimination between $s_{k|k}(\mathbf{x})$ and $s(\mathbf{x})$.

Elementary calculus shows that the function $f(\lambda) = \lambda - \lambda_{k|k} \log \lambda$ has a unique minimum value $A = f(\lambda_{k|k}) = \lambda_{k|k} - \lambda_{k|k} \log \lambda_{k|k}$ at $\lambda = \lambda_{k|k}$. Consider two cases. Case I: $A \geq 0$. Then $\lambda - \lambda_{k|k} \log \lambda \geq 0$ for all λ and $K(D)$ is minimized if and only if $K(s_{k|k}; s)$ and $\lambda - \lambda_{k|k} \log \lambda$ are minimized separately, i.e., if and only if $s = s_{k|k}$, $\lambda = \lambda_{k|k}$. Case II: $A < 0$. Then $-A + \lambda - \lambda_{k|k} \log \lambda \geq 0$ for all λ and

$$K(I) = K_3 + \lambda_{k|k} K(s_{k|k}; s) - A + \lambda - \lambda_{k|k} \log \lambda$$

for some constant K_3. This is minimized if and only if $K(s_{k|k}; s)$ and $-A + \lambda - \lambda_{k|k} \log \lambda$ are separately minimized, which occurs if and only if $s = s_{k|k}$ and $\lambda = \lambda_{k|k}$.

J. Proof of Theorem 5 (PHD Time-Update Equation)

Let $f_{k|k}(Y \mid Z^{(k)})$ denote the multitarget posterior at time-step k. Then its prediction to the next time-step is (eqn. (3))

$$f_{k+1|k}(Y \mid Z^{(k)}) = \int f_{k+1|k}(Y \mid X) f_{k|k}(X \mid Z^{(k)}) \delta X$$

where $f_{k+1|k}(Y \mid X)$ is the multitarget Markov transition density, (the distribution of $\Xi_{k+1|k}$). The PGFL of $f_{k+1|k}(Y \mid Z^{(k)})$ is

$$G_{k+1|k}[h] = \int h^Y f_{k+1|k}(Y \mid Z^{(k)}) \delta Y = \int \left\{ \int h^Y f_{k+1|k}(Y \mid X) \delta Y \right\} f_{k|k}(X \mid Z^{(k)}) \delta X = \int G_{k+1|k}[h \mid X] f_{k|k}(X \mid Z^{(k)}) \delta X.$$

($G_{k+1|k}[h \mid X]$ is the PGFL of $f_{k+1|k}(Y \mid X)$.) The PHD of $G_{k+1|k}[h]$ is

$$D_{k+1|k}(\mathbf{x}) = \frac{\delta G_{k+1|k}}{\delta \mathbf{x}}[1].$$

Before we can compute this we must first determine the formula for $G_{k+1|k}[h \mid X]$ and then, from this, the formula for $G_{k+1|k}[h]$. By assumption, $\Xi_{k+1|k} = \Xi(X) \cup \Psi(X) \cup \Psi_0$ where $\Xi(X) = \Xi(\mathbf{x}_1) \cup \cdots \cup \Xi(\mathbf{x}_n)$ is the set of surviving targets, $\Psi(X) = \Psi(\mathbf{x}_1) \cup \cdots \cup \Psi(\mathbf{x}_n)$ is the set of spawned targets, and Ψ_0 is the set of entering targets. Also, $\Xi(\mathbf{x}) = \emptyset$ with probability $1 - p_S(\mathbf{x})$ and $\Xi(\mathbf{x}) = \{\mathbf{Y}(\mathbf{x})\}$ with probability $p_S(\mathbf{x})$, where $\mathbf{Y}(\mathbf{x})$ is a random vector whose distribution is $f_{k+1|k}(\mathbf{y} \mid \mathbf{x})$. The multitarget distributions of $\Psi(\mathbf{x})$ and Ψ_0 are $b_{k+1|k}(Y \mid \mathbf{x})$ and $b_{k+1|k}(Y)$, respectively. The $\Xi(\mathbf{x}_1), \ldots, \Xi(\mathbf{x}_n)$, $\Psi(\mathbf{x}_1), \ldots, \Psi(\mathbf{x}_\tau)$, Ψ_0 are independent.

The PGFL $G_{k+1|k}[h \mid X]$ of $\Xi_{k+1|k}$ (i.e., of $f_{k+1|k}(Y \mid X)$) is the product of the PGFLs of $\Xi(\mathbf{x}_1), \ldots, \Xi(\mathbf{x}_n)$, $\Psi(\mathbf{x}_1), \ldots, \Psi(\mathbf{x}_\tau)$, and Ψ_0 (eqn. (45)). The PGFL of $\Xi(\mathbf{x})$ is (Proposition 4)

$$G_{\Xi(\mathbf{x})}[h] = 1 - p_S(\mathbf{x}) + p_S(\mathbf{x}) p_h(\mathbf{x})$$

where $p_h(\mathbf{x}) = \int h(\mathbf{x}) f_{k+1|k}(\mathbf{y} \mid \mathbf{x}) d\mathbf{x}$. So, the PGFL of $\Xi_{k+1|k}$ is

$$G_{k+1|k}[h \mid X] = (1 - p_S + p_S p_h)^X \cdot b_h^X \cdot e_h$$

where $b_h(\mathbf{x}) = \int h^X b_{k+1|k}(X \mid \mathbf{x}) \delta X$ is the PGFL of $\Psi(\mathbf{x})$ and $e_h = \int h^X b_{k+1|k}(X) \delta X$ is the PGFL of Ψ_0.

We now compute the the PGFL of $f_{k+1|k}(X \mid Z^{(k)})$. It is

$$G_{k+1|k}[h] = \int G_{k+1|k}[h \mid X] f_{k|k}(X \mid Z^{(k)}) \delta X$$
$$= e_h \cdot \int (1 - p_S + p_S p_h)^X \cdot b_h^X \cdot f_{k|k}(X \mid Z^{(k)}) \delta X$$
$$= e_h \cdot G_{k|k}[(1 - p_S + p_S p_h) b_h] = e_h \cdot G_{k|k}[\Phi[h]]$$

where $\Phi[h] = (1 - p_S + p_S p_h) b_h$. However,

$$\frac{\delta G_{k+1|k}}{\delta \mathbf{x}}[h] = \left(\frac{\delta}{\delta \mathbf{x}} e_h\right) \cdot G_{k|k}[\Psi[h]] + e_h \cdot \left(\frac{\delta}{\delta \mathbf{x}} G_{k|k}[\Psi[h]]\right)$$

and so by (61) and (66) the PHD of $G_{k+1|k}[h]$ is

$$D_{k+1|k}(\mathbf{x}) = \frac{\delta G_{k+1|k}}{\delta \mathbf{x}}[1] = \left[\frac{\delta}{\delta \mathbf{x}} e_h\right]_{h=1} \cdot G_{k|k}[\Psi[1]]$$
$$+ e_1 \cdot \left[\frac{\delta}{\delta \mathbf{x}} G_{k|k}[\Psi[h]]\right]_{h=1}$$
$$= \left[\frac{\delta}{\delta \mathbf{x}} e_h\right]_{h=1} \cdot 1 + 1 \cdot \left[\frac{\delta}{\delta \mathbf{x}} G_{k|k}[\Psi[h]]\right]_{h=1}$$
$$= b_{k+1|k}(\mathbf{x}) + \left[\frac{\delta}{\delta \mathbf{x}} G_{k|k}[\Psi[h]]\right]_{h=1}$$

where $b_{k+1|k}(\mathbf{x})$ is the PHD of $b_{k+1|k}(X)$. On the other hand, from Proposition 3 we know that the PHD of $G_{k|k}[\Phi[h]]$ is

$$D_{k+1|k}(\mathbf{x}) = \int \frac{\delta \Psi_{\mathbf{w}}}{\delta \mathbf{x}}[1] \cdot D_{k|k}(\mathbf{w}) d\mathbf{w}$$

where $\Phi_{\mathbf{w}}[h] = \Phi[h](\mathbf{w})$ and so

$$\frac{\delta \Psi_{\mathbf{w}}}{\delta \mathbf{x}}[h] = \frac{\delta}{\delta \mathbf{x}}(1 - p_S(\mathbf{w}) + p_S(\mathbf{w}) p_h(\mathbf{w})) b_h(\mathbf{w})$$
$$= p_S(\mathbf{w}) \left(\frac{\delta}{\delta \mathbf{x}} p_h(\mathbf{w})\right) b_h(\mathbf{w})$$
$$+ (1 - p_S(\mathbf{w}) + p_S(\mathbf{w}) p_h(\mathbf{w})) \frac{\delta}{\delta \mathbf{x}} b_h(\mathbf{w})$$

and so from (54)

$$\frac{\delta \Psi_{\mathbf{w}}}{\delta \mathbf{x}}[h] = p_S(\mathbf{w}) \cdot f_{k+1|k}(\mathbf{x} \mid \mathbf{w}) \cdot b_h(\mathbf{w})$$
$$+ (1 - p_S(\mathbf{w}) + p_S(\mathbf{w}) p_h(\mathbf{w})) b_{k+1|k}(\mathbf{x} \mid \mathbf{w}).$$

Setting $h = 1$,

$$\frac{\delta \Psi_{\mathbf{w}}}{\delta \mathbf{x}}[1] = p_S(\mathbf{w}) \cdot f_{k+1|k}(\mathbf{x} \mid \mathbf{w}) \cdot 1$$
$$+ (1 - p_S(\mathbf{w}) + p_S(\mathbf{w}) \cdot 1) b_{k+1|k}(\mathbf{x} \mid \mathbf{w})$$
$$= p_S(\mathbf{w}) \cdot f_{k+1|k}(\mathbf{x} \mid \mathbf{w}) + b_{k+1|k}(\mathbf{x} \mid \mathbf{w})$$

we finally get the claimed result,

$$D_{k+1|k}(\mathbf{x}) = b_{k+1|k}(\mathbf{x}) + \int (p_S(\mathbf{w}) f_{k+1|k}(\mathbf{x} \mid \mathbf{w})$$
$$+ b_{k+1|k}(\mathbf{x} \mid \mathbf{w})) d\mathbf{w}.$$

K. Proof of Theorem 6 (Single-Sensor Bayes Update)

Let $f_{k+1|k}(X \mid Z^{(k)})$ be the time-predicted multitarget posterior and $Z_{k+1} = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$ a new scan of observations. The data-updated multitarget posterior is given by Bayes' rule:

$$f_{k+1|k+1}(X \mid Z^{(k+1)}) = K^{-1} f_{k+1}(Z_{k+1} \mid X) f_{k+1|k}(X \mid Z^{(k)})$$
$$K = f_{k+1}(Z_{k+1} \mid Z^{(k)})$$
$$= \int f_{k+1}(Z_{k+1} \mid X) f_{k+1|k}(X \mid Z^{(k)}) \delta X.$$

Define the two-variable PGFL

$$F[g,h] = \int \int h^X g^Z f_{k+1}(Z \mid X) f_{k+1|k}(X \mid Z^{(k)}) \delta X \delta Z$$
$$= \int h^X G_{k+1}[g \mid X] f_{k+1|k}(X \mid Z^{(k)}) \delta X$$

where $G_{k+1}[g \mid X]$ is the PGFL

$$G_{k+1}[g \mid X] = \int g^Z f_{k+1}(Z \mid X) \delta Z$$

of $f_{k+1|k}(X \mid Z^{(k)})$. From (26) and (52) the denominator of Bayes' rule can be rewritten as

$$f_{k+1}(Z_{k+1} \mid Z^{(k)}) = \frac{\delta^m F}{\delta \mathbf{z}_m \cdots \delta \mathbf{z}_1}[0,1] \qquad (110)$$

for any $Z_{k+1} = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$. Equations (26), (52), (61), and (66) tell us that the PHD of $f_{k+1|k+1}(X \mid Z^{(k+1)})$ can be rewritten as

$$D_{k+1|k+1}(\mathbf{x} \mid Z^{(k+1)}) = \frac{1}{f_{k+1}(Z_{k+1} \mid Z^{(k)})} \cdot \frac{\delta^{m+1} F}{\delta \mathbf{z}_m \cdots \delta \mathbf{z}_1 \delta \mathbf{x}}[0,1] \qquad (111)$$

The random observation-set Σ is $\Sigma_{k+1} = \Sigma(\mathbf{x}_1) \cup \cdots \cup \Sigma(\mathbf{x}_n) \cup \Theta$ where Θ is the Poisson false alarm process and where $\Sigma(\mathbf{x}) = \emptyset$ (no observation) with probability $1 - p_D(\mathbf{x})$ and $\Sigma(\mathbf{x}) = \{\mathbf{Z}(\mathbf{x})\}$ with probability $p_D(\mathbf{x})$, where $\mathbf{Z}(\mathbf{x})$ is a random vector whose distribution is $f_{k+1}(\mathbf{z} \mid \mathbf{x})$. By Proposition 5c the PGFL of Θ is $G_\Theta[g] = e^{\lambda c[g] - \lambda}$ where $c[g] = \int g(\mathbf{z}) c(\mathbf{z}) d\mathbf{z}$. The $\Sigma(\mathbf{x}_1), \ldots, \Sigma(\mathbf{x}_n)$, Θ are statistically independent.

Given this, the PGFL $G[g \mid \mathbf{x}]$ of $\Sigma(\mathbf{x})$ is, from Proposition 4,

$$G[g \mid \mathbf{x}] = 1 - p_D(\mathbf{x}) + p_D(\mathbf{x}) p_g(\mathbf{x})$$

where $p_g(\mathbf{x}) = \int g(\mathbf{z}) f_{k+1}(\mathbf{z} \mid \mathbf{x}) d\mathbf{z}$. From conditional independence and $G_\Theta[g] = e^{\lambda c[g] - \lambda}$, $c[g] = \int g(\mathbf{z}) c(\mathbf{z}) d\mathbf{z}$, the PGFL of Σ_{k+1} is

$$G_{k+1}[g \mid X] = G[g \mid \mathbf{x}_1] \cdots G[g \mid \mathbf{x}_n] \cdot G_\Theta[g]$$

by (45). Consequently, if we abbreviate

$$h_0 = q_D + p_D p_g, \qquad q_D = 1 - p_D$$

then

$$F[g,h] = \int h^X h_0^X e^{\lambda c[g]-\lambda} f_{k+1|k}(X \mid Z^{(k)})\delta X$$
$$= e^{\lambda c[g]-\lambda} \int (h_0 h)^X f_{k+1|k}(X \mid Z^{(k)})\delta X$$
$$= e^{\lambda c[g]-\lambda} G_{k+1|k}[h_0 h]$$
$$= e^{\lambda c[g]-\lambda} G_{k+1|k}[h(q_D + p_D p_g)].$$

Assume now that $f_{k+1|k}(X \mid Z^{(k)})$ is Poisson:

$$G_{k+1|k}[h] = e^{\mu s[h]-\mu}, \qquad s[h]_= \int h(\mathbf{x}) s(\mathbf{x}) d\mathbf{x}$$

so the PHD of $f_{k+1|k}(X \mid Z^{(k)})$ is $D_{k+1|k}(\mathbf{x}) = \mu s(\mathbf{x})$. Then

$$G_{k+1|k}[h(q_D + p_D p_g)] = \exp(\mu s[hq_D] + \mu s[hp_D p_g] - \mu)$$
$$F[g,h] = \exp(\lambda c[g] - \lambda + \mu s[hq_D] + \mu s[hp_D p_g] - \mu).$$

Set $h = 1$ and use $q_D = 1 - p_D$ to get

$$F[g,1] = \exp(\lambda c[g] - \lambda - \mu c[p_D] + \mu c[p_D p_g])$$

and compute $f_{k+1}(Z_{k+1} \mid Z^{(k)})$ in (111). For $Z = \emptyset$,

$$f_{k+1}(Z_{k+1} \mid Z^{(k)}) = F[0,1] = e^{-\lambda-\mu s[pD]}.$$

For $Z_{k+1} = \{\mathbf{z}_1\}$, first compute the functional derivative:

$$\frac{\delta F}{\delta \mathbf{z}_1}[g,1] = F[g,1] \cdot \frac{\partial}{\partial \delta_{\mathbf{z}_1}}(\lambda c[g] - \lambda - \mu s[p_D] + \mu s[p_D p_g])$$
$$= F[g,1] \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}]). \tag{112}$$

Then setting $g = 0$ we get

$$f_{k+1|k+1}(Z \mid Z^{(k)}) = e^{-\lambda-\mu s[p_D]} \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}]). \tag{113}$$

For $Z_{k+1} = \{\mathbf{z}_1, \mathbf{z}_2\}$, compute a functional derivative of the right-hand side of (112):

$$\frac{\delta^2 F}{\delta \mathbf{z}_2 \delta \mathbf{z}_1}[g,1] = \frac{\delta}{\delta \mathbf{z}_2} F[g,1] \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}])$$
$$= F[g,1] \cdot (\lambda c(\mathbf{z}_2) + \mu s[p_D L_{\mathbf{z}_2}])$$
$$\cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}]) \tag{114}$$

which, after setting $g = 0$, leads to:

$$f_{k+1|k+1}(Z \mid Z^{(k)}) = e^{-\lambda-\mu s[p_D]} \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}])$$
$$\cdot (\lambda c(\mathbf{z}_2) + \mu s[p_D L_{\mathbf{z}_2}]). \tag{115}$$

For $Z_{k+1} = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$ the denominator of (111) is

$$f_{k+1|k+1}(Z \mid Z^{(k)}) = e^{-\lambda-\mu s[p_D]} \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}])$$
$$\cdots (\lambda c(\mathbf{z}_m) + \mu s[p_D L_{\mathbf{z}_m}]). \tag{116}$$

Now turn to the numerator of the PHD in (111). Begin by taking a functional derivative with respect to h:

$$\frac{\delta F}{\delta \mathbf{x}}[g,h] = F[g,h] \cdot \frac{\delta}{\delta \mathbf{x}}(\lambda c[g] - \lambda + \mu s[hq_D] + \mu s[hp_D p_g] - \mu)$$
$$= F[g,h] \cdot (\mu q_D(\mathbf{x}) s(\mathbf{x}) + \mu p_D(\mathbf{x}) p_g(\mathbf{x}) s(\mathbf{x})). \tag{117}$$

Then set $h = 1$:

$$\frac{\delta F}{\delta \mathbf{x}}[g,1] = F[g,1] \cdot (\mu q_D(\mathbf{x}) s(\mathbf{x}) + \mu p_D(\mathbf{x}) p_g(\mathbf{x}) s(\mathbf{x})). \tag{118}$$

Repeat the steps of the derivation of the denominator. For $Z_{k+1} = \emptyset$ we set $g = 0$ in (118):

$$\frac{\delta F}{\delta \mathbf{x}}[0,1] = F[0,1] \cdot \mu q_D(\mathbf{x}) s(\mathbf{x}). \tag{119}$$

For $Z_{k+1} = \{\mathbf{z}_1\}$, compute the functional derivative:

$$\frac{\delta^2 F}{\delta \mathbf{z}_1 \delta \mathbf{x}}[g,1] = F[g,1] \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}])$$
$$\cdot (\mu q_D(\mathbf{x}) s(\mathbf{x}) + \mu p_D(\mathbf{x}) p_g(\mathbf{x}) s(\mathbf{x}))$$
$$+ F[g,1] \cdot \mu p_D(\mathbf{x}) L_{\mathbf{z}_1}(\mathbf{x}) s(\mathbf{x})$$

and then set $g = 0$:

$$\frac{\delta^2 F}{\delta \mathbf{z}_1 \delta \mathbf{x}}[0,1] = e^{-\lambda-\mu s[p_D]} \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}]) \cdot \mu q_D(\mathbf{x}) s(\mathbf{x})$$
$$+ e^{-\lambda-\mu s[p_D]} \cdot \mu p_D(\mathbf{x}) L_{\mathbf{z}_1}(\mathbf{x}) s(\mathbf{x}). \tag{120}$$

For $Z_{k+1} = \{\mathbf{z}_1, \mathbf{z}_2\}$, compute another functional derivative with respect to g of the left-hand side of (120):

$$\frac{\delta^3 F}{\delta \mathbf{z}_2 \delta \mathbf{z}_1 \delta \mathbf{x}}[g,1] = \left(\frac{\delta}{\delta \mathbf{z}_2} F[g,1]\right) \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}])$$
$$\cdot (\mu q_D(\mathbf{x}) s(\mathbf{x}) + \mu p_D(\mathbf{x}) p_g(\mathbf{x}) s(\mathbf{x}))$$
$$+ F[g,1] \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}])$$
$$\cdot \frac{\delta}{\delta \mathbf{z}_2}(\mu q_D(\mathbf{x}) s(\mathbf{x}) + \mu p_D(\mathbf{x}) p_g(\mathbf{x}) s(\mathbf{x}))$$
$$+ \left(\frac{\delta}{\delta \mathbf{z}_2} F[g,1]\right) \cdot \mu p_D(\mathbf{x}) L_{\mathbf{z}_1}(\mathbf{x}) s(\mathbf{x})$$
$$= F[g,1] \cdot (\lambda c(\mathbf{z}_2) + \mu s[p_D L_{\mathbf{z}_2}])$$
$$\cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}])$$
$$\cdot (\mu q_D(\mathbf{x}) s(\mathbf{x}) + \mu p_D(\mathbf{x}) p_g(\mathbf{x}) s(\mathbf{x}))$$
$$+ F[g,1] \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}])$$
$$\cdot \mu p_D(\mathbf{x}) L_{\mathbf{z}_2}(\mathbf{x}) s(\mathbf{x})$$
$$+ F[g,1] \cdot (\lambda c(\mathbf{z}_2) + \mu s[p_D L_{\mathbf{z}_2}]).$$
$$\cdot \mu p_D(\mathbf{x}) L_{\mathbf{z}_1}(\mathbf{x}) s(\mathbf{x}).$$

Setting $g = 0$ we finally get

$$\frac{\delta^3 F}{\delta \mathbf{z}_2 \delta \mathbf{z}_1 \delta \mathbf{x}}[0,1] = e^{-\lambda-\mu s[p_D]} \cdot (\lambda c(\mathbf{z}_2) + \mu s[p_D L_{\mathbf{z}_2}]) \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}]) \cdot \mu q_D(\mathbf{x}) s(\mathbf{x}) + e^{-\lambda-\mu s[p_D]} \cdot (\lambda c(\mathbf{z}_1) + \mu s[p_D L_{\mathbf{z}_1}]) \cdot \mu p_D(\mathbf{x}) L_{\mathbf{z}_2}(\mathbf{x}) s(\mathbf{x}) + e^{-\lambda-\mu s[p_D]} \cdot (\lambda c(\mathbf{z}_2) + \mu s[p_D L_{\mathbf{z}_2}]) \cdot \mu p_D(\mathbf{x}) L_{\mathbf{z}_1}(\mathbf{x}) s(\mathbf{x}).$$

In general, a proof by induction shows that

$$\frac{\delta^{m+1} F}{\delta \mathbf{z}_m \cdots \delta \mathbf{z}_1 \delta \mathbf{x}}[0,1] = e^{-\lambda-\mu s[p_D]} \cdot \prod_{\mathbf{z}\in Z_{k+1}} (\lambda c(\mathbf{z}) + \mu s[p_D L_{\mathbf{z}}]) \cdot \mu q_D(\mathbf{x}) s(\mathbf{x}) + e^{-\lambda-\mu s[p_D]} \left(\prod_{\mathbf{z}\in Z_{k+1}} (\lambda c(\mathbf{z}) + \mu s[p_D L_{\mathbf{z}}]) \right) \cdot \sum_{\mathbf{z}\in Z_{k+1}} \frac{\mu p_D(\mathbf{x}) L_{\mathbf{z}}(\mathbf{x}) s(\mathbf{x})}{\lambda c(\mathbf{z}) + \mu s[p_D L_{\mathbf{z}}]}. \tag{121}$$

Substituting (116), (121) into (111),

$$D_{k+1|k+1}(\mathbf{x}) = \mu q_D(\mathbf{x}) s(\mathbf{x}) + \sum_{\mathbf{z}\in Z_{k+1}} \frac{\mu p_D(\mathbf{x}) L_{\mathbf{z}}(\mathbf{x}) s(\mathbf{x})}{\lambda c(\mathbf{z}) + \mu s[p_D L_{\mathbf{z}}]} = \left(1 - p_D(\mathbf{x}) + \sum_{\mathbf{z}\in Z_{k+1}} \frac{p_D(\mathbf{x}) L_{\mathbf{z}}(\mathbf{x})}{\lambda c(\mathbf{z}) + \mu s[p_D L_{\mathbf{z}}]}\right) \mu s(\mathbf{x}) \tag{122}$$

and so, as claimed,

$$D_{k+1|k+1}(\mathbf{x}) = \left(1 - p_D(\mathbf{x}) + \sum_{\mathbf{z}\in Z_{k+1}} \frac{p_D(\mathbf{x}) L_{\mathbf{z}}(\mathbf{x})}{\lambda c(\mathbf{z}) + D_{k+1|k}[p_D L_{\mathbf{z}}]}\right) D_{k+1|k}(\mathbf{x}). \tag{123}$$

VII. CONCLUSION

In this paper I used finite-set statistics (FISST) to provide a systematic and rigorous foundation for "single-density" multitarget detection, tracking, and identification approaches that collapse the multitarget posterior into a single density function on single-target state space X. The core concept is the Stein-Winter probability hypothesis density (PHD). After framing the optimal multitarget tracking problem as a multitarget recursive Bayes filtering problem, I showed that the PHD is a first-order statistical moment of the time-evolving random state-set $\Xi_{k|k}$. I derived a recursive Bayes filter for the PHD which generalizes that of earlier papers in that it allows nonconstant probability of detection. This filter can be interpreted as a multitarget statistical analog of a constant-gain Kalman filter (e.g., the α-β-γ filter).

Current work is directed at computational implementation of the PHD filter using particle-systems and related techniques. Such techniques are potentially practical since they have guaranteed-convergence properties and enjoy $O(N)$ computational complexity in the number N of particles. It is anticipated that implemented PHD filters will be most useful in high-density applications where multihypothesis trackers begin to fail, e.g., cluster tracking and group-target tracking [25, 43].

REFERENCES

[1] Agate, C. S., and Sullivan, K. J. (2003)
Particle filtering algorithm for tracking multiple road-constrained targets.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition XII*, SPIE Proceedings, **5096** (2003), 256–266.

[2] Ballantyne, D. J., Chan, H. Y., and Kouritzin, M. A. (2001)
A branching particle-based nonlinear filter for multi-target tracking.
In *Proceedings of 2001 International Conference on Information Fusion*, Montreal, Aug. 7–10, 2001.

[3] Baudin, M. (1984)
Multidimensional point processes and random closed sets.
Journal of Applied Prob., **21** (1984), 173–178.

[4] Bar-Shalom, Y., and Li, X-R. (1993)
Estimation and Tracking: Principles, Techniques, and Software.
Boston: Artech House, 1993.

[5] Bethel, R. E., and Paras, G. J. (1994)
A PDF multitarget-tracker.
IEEE Transactions on Aerospace and Electronic Systems, **30**, 2 (1994), 386–403.

[6] Challa, S., Vo, B-N., and Wang, X. (2002)
Bayesian approaches to track existence—IPDA and random sets.
In *Proceedings of 5th International Conference on Information Fusion*, Vol. II, 2002, 1228–1235.

[7] Daley, D. J., and Vere-Jones, D. (1988)
An Introduction to the Theory of Point Processes.
New York: Springer-Verlag, 1988.

[8] El-Fallah, A., Zajic, T., Mahler, R., Ravichandran, B., and Mehra, R. (2001)
Multitarget nonlinear filtering based on spectral compression and probability hypothesis density.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition X*, SPIE Proceedings, **4380** (2001), 207–216.

[9] Goodman, I. R., and Nguyen, H. T. (1985)
Uncertainty Models for Knowledge Based Systems.
Amsterdam, North-Holland, 1985.

[10] Goodman, I. R., Mahler, R. P. S., and Nguyen, H. T. (1997)
Mathematics of Data Fusion, Boston: Kluwer Academic Publishers, 1997.

[11] Ho, Y. C., and Lee, R. C. K. (1964)
A Bayesian approach to problems in stochastic estimation and control.
IEEE Transactions on Automatic Control, **AC-9** (1964), 333–339.

[12] Hue, C., Le Cadre, J-P., and Pérez, P. (2001)
A particle filter to track multiple objects.
In *Proceedings of 2001 IEEE Workshop on Multi-Object Tracking*, Vancouver, July 8, 2001, 61–68.

[13] Hue, C., Le Cadre, J-P., and Peréz, P. (2002)
Sequential Monte Carlo methods for multiple target tracking and data fusion.
In *IEEE Transactions on Signal Processing*, **50**, 2 (2002), 309–325.

[14] Iltis, R. A. (1999)
State estimation using an approximate reduced statistics algorithm.
IEEE Transactions on Aerospace and Electronic Systems, **35**, 4 (1999), 1161–1172.

[15] Itô, K. (1993)
Encyclopedic Dictionary of Mathematics (2nd ed.), Vol. I.
Cambridge, MA, MIT Press, 1993.

[16] Jazwinski, A. H. (1970)
Stochastic Processes and Filtering Theory.
New York: Academic Press, 1970.

[17] Karr, A. F. (1991)
Point Processes and Their Statistical Inference, (2nd ed.).
Marcel Dekker, 1991.

[18] Kastella, K. (1996)
Discrimination gain for sensor management in multitarget detection and tracking.
In *Proceedings of the 1996 IMACS Multiconference on Comp. and Eng. Appl. (CESA'96): Symposium on Contr., Opt., and Supervision*, Lille France, July 9–12, 1996, 167–172.

[19] Kastella, K. (2000)
A microdensity approach to multitarget tracking.
In *Proceedings of 4th International Conference on Information Fusion*, Paris, 2000, TuB1-3–TuB1-10.

[20] Kowalsky, H. J. (1965)
Topological Spaces.
Academic Press, 1965.

[21] Kulhavý, R. (1990)
Recursive nonlinear estimation: A geometric approach.
Automatica, **26**, 3 (1990), 545–555.

[22] Lanterman, A. D. (2003)
Sampling from multitarget Bayesian posteriors for random sets via jump-diffusion processes.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition XII*, SPIE Proceedings, **5096** (2003), 300–311.

[23] Lanterman, A. D., Miller, M. I., Snyder, D. L., and Miceli, W. J. (1994)
Jump-diffusion processes for the automated understanding of FLIR scenes.
In F. A. Sadjadi (Ed.), *Automatic Target Recognition IV*, SPIE Proceedings, Vol. 2234, 1994, 416–427.

[24] Mahler, R. (2001)
Detecting, tracking, and classifying group targets: A unified approach.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition X*, SPIE Proceedings, Vol. 4380, 2001, 217–228.

[25] Mahler, R. (2003)
Bayesian cluster tracking using a generalized Cheeseman approach.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition XII*, SPIE Proceedings, **5096** (2003), 334–345.

[26] Mahler, R. (2001)
Engineering statistics for multi-object tracking.
In *Proceedings of 2001 IEEE Workshop on Multi-Object Tracking*, Vancouver, July 8, 2001, 53–60.

[27] Mahler, R. (2002)
An extended first-order Bayes filter for force aggregation.
In O. Drummond (Ed.), *Signal and Data Processing of Small Targets 2002*, SPIE Proceedings, **4728** (2002), 196–207.

[28] Mahler, R. (1998)
Global posterior densities for sensor management.
In M. K. Masten and L. A. Stockum (Eds.), *Acquisition, Tracking, and Pointing XII*, SPIE Proceedings, **3365** (1998), 252–263.

[29] Mahler, R. (2000)
An Introduction to Multisource-Multitarget Statistics and Its Applications.
Technical Monograph, Lockheed Martin, Mar. 15, 2000.

[30] Mahler, R. (1999)
Multitarget Markov motion models.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition VIII*, SPIE Proceedings, **3720** (1999), 47–57.

[31] Mahler, R. (2001)
Multitarget moments and their application to multitarget tracking.
Proceedings of the Workshop on Estimation, Tracking, and Fusion: A Tribute to Yaakov Bar-Shalom, Naval Postgraduate School, Monterey, CA, May 17, 2001, 134–166.

[32] Mahler, R. (2003)
Multisensor-multitarget sensor management: A unified Bayesian approach.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition XII*, SPIE Proceedings, **5096** (2003), 222–233.

[33] Mahler, R. (1999)
Multisource-multitarget detection and acquisition: A unified approach.
In O. E. Drummond (Ed.), *Signal and Data Processing of Small Targets 1999*, SPIE Proceedings, **3809** (1999), 218–229.

[34] Mahler, R. (1998)
Multisource-multitarget filtering: A unified approach.
In O. E. Drummond (Ed.), *Signal and Data Processing of Small Targets 1998*, SPIE Proceedings, **3373** (1998), 296–307.

[35] Mahler, R. (2003)
Objective functions for Bayesian control-theoretic sensor management, I: Multitarget first-moment approximation.
In *Proceedings of the 2002 IEEE Conference on Aerospace Systems*, Big Sky MT, Mar. 8–15, 2003.

[36] Mahler, R.
Objective functions for Bayesian control-theoretic sensor management, II: MHC-like approximation.
In S. Butenko, R. Murphey, and P. Paralos (Eds.), *New Developments in Cooperative Control and Optimization*, Boston: Kluwer Academic Publishers, to be published.

[37] Mahler, R. (2002)
Random set theory for target tracking and identification.
In D. L. Hall and J. Llinas (Eds.), *Handbook of Multisensor Data Fusion*, Boca Raton, FL: CRC Press, 2002, ch. 14.

[38] Mahler, R. (2000)
A theoretical foundation for the Stein-Winter probability hypothesis density (PHD) multitarget tracking approach.
In *Proceedings of the 2000 MSS National Symposium on Sensor and Data Fusion*, Vol. I (unclassified), San Antonio TX, June 20–22, 2000, 99–117.

[39] Mahler, R. (2003)
Tractable multistep sensor management via MHT.
In *Proceedings of the Workshop on Multi-Hypothesis Tracking: A Tribute to Samuel Blackman*, San Diego, CA, May 30, 2003, to be published.

[40] Mahler, R. (1994)
A unified foundation for data fusion.
In F. A. Sadjadi (Ed.), *Selected Papers on Sensor and Data Fusion*, SPIE, **MS-124** (1996), 325–345, 1996; reprinted from *Seventh Joint Service Data Fusion Symposium*, Laurel, MD, 1994, 154–174.

[41] Mahler, R. (2003)
Unified sensor anagement for group target processing.
In *Proceedings of the 2003 MSS National Symposium on Sensor and Data Fusion*, San Diego, CA, June 2–5, 2003, to be published.

[42] Mahler, R., and Zajic, T. (2001)
Multitarget filtering using a multitarget first-order moment statistic.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition X*, SPIE Proceedings, **4380** (2001), 184–195.

[43] Mahler, R., and Zajic, T. (2002)
Bulk multitarget tracking using a first-order multitarget tracking filter.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition XI*, SPIE, **4729** (2002), 175–186.

[44] Moreland, M., and Challa, S. (2003)
A multi-target tracking algorithm based on random sets.
In *Proceedings of 6th International Conference on Information Fusion*, July 8–11, 2003, Cairns, Australia, to be published.

[45] Mori, S. (2000)
Random sets in data fusion: Formalism to new algorithms.
In *Proceedings of Third International Conference on Information Fusion*, Paris, July 10–13, 2000.

[46] Mori, S., and Chong, C-Y. (2001)
Quantitative evaluation of Poisson point process approximation.
In *Proceedings of 4th International Conference on Information Fusion*, Montreal, Aug. 7–10, 2001.

[47] Mori, S., Chong, C-Y., Tse, E., and Wishner, R. P. (1986)
Tracking and classifying multiple targets without a priori identification.
IEEE Transactions on Automatic Control, **AC-31**, 5 (1986), 401–409.

[48] Musick, S., Kastella, K., and Mahler, R. (1998)
A practical implementation of joint multitarget probabilities.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition VII*, SPIE Proceedings, **3374** (1998), 26–37.

[49] Portenko, N., Salehi, H., and Skorokhod, A. (1997)
On optimal filtering of multitarget tracking systems based on point processes observations.
Random Operators and Stochastic Equations, **1** (1997), 1–34.

[50] Ripley, B. D. (1976)
Locally finite random sets: foundations for point process theory.
Annals of Probability, **4**, 6 (1976), 983–994.

[51] Ryder, L. H. (1996)
Quantum Field Theory (2nd ed.).
New York: Cambridge University Press, 1996.

[52] Sidenbladh, H. (2003)
Multi-target particle filtering for the probability hypothesis density.
In *Proceedings of 6th International Conference on Information Fusion*, Cairns, Australia, July 8–11, 2003, to be published.

[53] Sidenbladh, H., and Wirkander, S-L. (2003)
Tracking random sets of vehicles in terrain.
In *Proceedings of 2003 IEEE Workshop on Multi-Object Tracking*, Madison, WI, June 21, 2003, to be published.

[54] Snyder, D. L., and Miller, M. I. (1991)
Random Point Processes in Time and Space (2nd ed.).
New York: Springer, 1991.

[55] Sorenson, H. W. (1988)
Recursive estimation for nonlinear dynamic systems.
In J. C. Spall (Ed.), *Bayesian Analysis of Statistical Time Series and Dynamic Models*, Marcel Dekker, 1988.

[56] Stoyan, D., Kendall, W. S., and Meche, J. (1995)
Stochastic Geometry and Its Applications (2nd ed.).
New York: Wiley, 1995

[57] Srivastava, A., Miller, M. I., and Grenander, U. (1991)
Jump-diffusion processes for object tracking and direction finding.
In *Proceedings of the 29th Allerton Conference on Communication, Control, and Computing*, University of Illinois, Urbana, 1991, 563–570.

[58] Srivastava, A., Cutaia, N., Miller, M. I., O'Sullivan, J. A., and Snyder, D. L. (1992)
Multitarget narrowband direction finding and tracking using motion dynamics.
In *Proceedings of the 30th Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sept. 30–Oct.2, 1992, 279–288.

[59] Stein, M. C., and Winter, C. L. (1993)
An additive theory of probabilistic evidence accrual.
Report LA-UR-93-3336, Los Alamos National Laboratories, 1993.

[60] Stein, M., and Tenney, R. R.
What's the difference between PHS and MHT?
Oasis Corp. and Alphatech Corp., working paper, undated.

[61] Stone, L. D., Barlow, C. A., and Corwin, T. L. (1999)
Bayesian Multiple Target Tracking.
Boston: Artech House, 1999.

[62] Stone, L. D., Corwin, T. L., and Hofmann, J. B. (1995)
Technical documentation of Nodestar.
Technical Report NRL/FR/5580-95-9788, Naval Research Laboratory, 1995.

[63] Tao, K. M., Abileah, R., and Lawrence, J. D. (2000)
Multiple-target tracking in dense noisy environments: A probabilistic mapping perspective.
In O. E. Drummond (Ed.), *Signal and Data Processing of Small Targets 2000*, SPIE Proceedings, **4048** (2000), 474–485.

[64] van Trees, H. L. (1968)
Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory.
New York: Wiley, 1968.

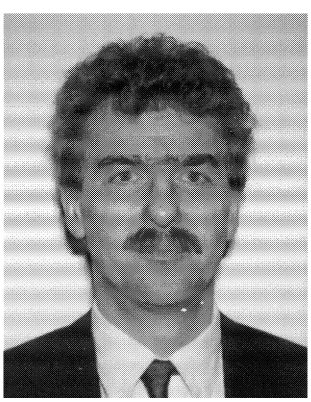
[65] Vo, B., Singh, S., and Doucet, A. (2003)
Sequential implementation of the PHD filter for multi-target tracking.
In *Proceedings of the Sixth International Conference on Information Fusion*, Cairns, Australia, July 8–11, 2003, to be published.

[66] Washburn, R. B. (1995)
A random point process approach to multi-object tracking.
In *Proceedings of the American Control Conference*, Vol. 3, June 10–12, 1987, Minneapolis, MN, 1846–1852.

[67] Winter, C. L., and Stein, M. C. (1995)
IES/BTI system overview.
Proceedings of 8th National Symposium on Sensor Fusion, Vol. I (unclassified), Mar. 15–17, 1995, Dallas TX, 27–46.

[68] Zajic, T., Ravichandran, R., Mahler, R., Mehra, R., and Noviskey, M. (2003)
Joint tracking and identification with robustness against unmodeled targets.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition XII*, SPIE Proceedings, **5096** (2003), 279–290.

[69] Zajic, T., and Mahler, R. (2003)
A particle-systems implementation of the PHD multitarget tracking filter.
In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition XII*, SPIE Proceedings, **5096** (2003), 291–299.



Ronald P. S. Mahler was born in Great Falls, MT, in 1948. He earned a B.A. in mathematics from the University of Chicago, Chicago, IL, in 1970, a Ph.D. in mathematics from Brandeis University, Waltham, MA, in 1974, and a B.E.E. in electrical engineering from the University of Minnesota, Minneapolis, in 1980.

He was an assistant professor of Mathematics at the University of Minnesota, from 1974 to 1979. Since 1980 he has been employed as a research engineer at Lockheed Martin NE&SS Tactical Systems, Eagan, MN. His research interests include data fusion, expert systems theory, multitarget tracking, combat identification, sensor management, random set theory, point process theory, and conditional event algebra.

Dr. Mahler is the author, coauthor, or coeditor of over forty publications, including eleven articles in refereed journals, a book, a hardcover conference proceedings, and a monograph. He has been invited to present at many universities, U.S. government laboratories, and conferences including Harvard, Johns Hopkins, the University of Wisconsin, the University of Massachusetts, the Air Force Institute of Technology, SPAWAR Systems Center, the IEEE Conference on Decision and Control, the International Conference on Information, Decision, and Control, and the International Conference on Information Fusion.