Convergence Results for the Particle PHD Filter

Daniel Edward Clark, Judith Bell

***Abstract*— Bayesian single-target tracking techniques can be extended to a multiple-target environment by viewing the multiple-target state as a Random Finite Set, but evaluating the multiple-target posterior distribution is currently computationally intractable for real-time applications. A practical alternative to the optimal Bayes multi-target filter is the PHD (Probability Hypothesis Density) filter, which propagates the first-order moment of the multi-target posterior instead of the posterior distribution itself. It has been shown that the PHD is the best-fit approximation of the multi-target posterior in an information-theoretic sense. The method avoids the need for explicit data association, as the target states are viewed as a single global target state, and the identities of the targets are not part of the tracking framework. Sequential Monte Carlo approximations of the PHD using particle filter techniques have been implemented, showing the potential of this technique for real-time tracking applications. This article presents mathematical proofs of convergence for the particle filtering algorithm and gives bounds for the mean square error.**

I. Introduction

Sequential Monte Carlo approximations of the optimal multiple-target filter are computationally expensive. A practical suboptimal alternative to the optimal filter is the Probability Hypothesis Density (PHD) filter, which propagates the first-order statistical moment instead of the full multiple-target posterior. The integral of the PHD in any region of the state space is the expected number of targets in that region [1].

Particle filter methods for the PHD-filter have been devised by Vo [2] and Zajic [3]. Practical applications of the filter include tracking vehicles in different terrains [4], tracking targets in passive radar located on ellipses [5], and tracking a variable number of targets in forward-scan sonar [6]. These have demonstrated its potential in real-time multiple-target tracking applications, avoiding the need for explicit data association, although track continuity is not maintained.

The Sequential Monte Carlo implementation of the PHD is relatively recent, and there have been no results showing asymptotic convergence of the technique. This article shows some of these results, based on results provided by Crisan [7] [8] [9] for particle filters. We present results for the convergence of the mean square error as well as weak convergence of the empirical particle measure to the true PHD measure.

The paper first explains the theoretical foundation for the PHD Filter and how it relates to Point Process theory before giving the model used for tracking multiple targets and the PHD Filter equations. The Sequential Monte Carlo implementation, or Particle PHD Filter algorithm, is given in Section IV. Section V presents our convergence results for the Particle PHD Filter. The proofs of these results are given as appendices.

Daniel Edward Clark and Judith Bell are in the Ocean Systems Lab, Electrical and Computer Engineering, Heriot-Watt University, Edinburgh. dec30@cantab.net, j.bell@hw.ac.uk

Published with kind permission of QinetiQ.

II. Point Processes

A *stochastic process* is a mathematical model for the occurrence of a random phenomenon at each moment after an initial point in time. We will be considering the discrete time case, where a stochastic process is defined as a collection of random variables $X = \{X_t : t \in \mathbb{N}\}$ in time on a *sample space* $(\Omega, \mathcal{F})$, which take values in a *state space* $(\mathcal{S}, \mathcal{G})$. The state space $(\mathcal{S}, \mathcal{G})$ is based on the Euclidean space in which our targets are located, i.e. $\mathcal{S} = \mathbb{R}^d$, where d is the dimension of the space, and $\mathcal{G} = \mathcal{B}(\mathbb{R}^d)$ is the σ-algebra of Borel sets. The sample space $(\Omega, \mathcal{F})$ is the space in which the observations of the targets are taken. The observations may also be from Euclidean space, for example, although the number of dimensions is not necessarily the same, e.g. $\Omega = \mathbb{R}^k$ and $\mathcal{F} = \mathcal{B}(\mathbb{R}^k)$, where k is the number of dimensions. The *point process* is viewed as a stochastic process where the random variables X_t are random sets. The basic idea of point processes is to study collections of point occurrences, the simplest of which is a Poisson process.

A. Random Counting Measures

Every point process on $\mathbb{R}^d$ can be interpreted as a random integer-valued measure, $\mathcal{N}$, on the Borel subsets $\mathcal{B}(\mathbb{R}^d)$. Alternatively, the *counting measures* are defined as those members S for which $\mathcal{N}(S) \in \mathbb{Z}$. (See [10], [11] for further details.) In Vo *et al.* [2], the random finite set Θ was represented by a random counting measure $\mathcal{N}_\Theta$ defined by $\mathcal{N}_\Theta(S) = |\Theta \cap S|$, where the $|\cdot|$ notation represents the number of elements. The PHD is defined as the first-order statistical moment, or expectation, of a random counting measure.

Suppose that the mean number of targets is finite. Then, for any Borel set $A \in \mathcal{B}(\mathbb{R}^d)$, we can define the *expectation measure* $\mathcal{M}(.)$ by

$$\mathcal{M}(A) := \mathcal{M}_1(A) = E[\mathcal{N}(A)], \tag{1}$$

where $\mathcal{M}_1$ is the first-order moment. $\mathcal{M}(.)$ inherits countable additivity from $\mathcal{N}(.)$; hence, the PHD defines a measure on $\mathcal{B}(\mathbb{R}^d)$.

III. Multiple Target Tracking Model

The multiple target filtering problem is to estimate the positions of an unknown varying number of targets, based on observations which may also include false alarms due to erroneous measurements. In standard tracking applications, this is usually solved by assigning a single-target stochastic filter, such as a Kalman filter, to each target and managing the different tracks with a data association technique.

We now present the multiple tracking model used in the paper. Define two random-set stochastic processes $X = \{X_t : t \in \mathbb{N}\}$ and $Y = \{Y_t : t \in \mathbb{N}\backslash\{0\}\}$, where process X is called the *state* process and process Y is called the *observation* process. These processes will be used to formulate the multiple-target Bayesian Filtering equations.

The multiple target state at time t is represented by random-set $X_t = \{x_{t,1},...,x_{t,T_t}\}$, where $x_{t,i}$ represents the state of an individual target and T_t is the number of targets at time t. The multiple target measurement at time t is given by $Z_t = \{z_{t,1},...,z_{t,m_t}\}$, where $z_{t,j}$ represents a single-target measurement or false alarm and m_t is the number of observations at time t.

The filtering problem is then to estimate the unobserved signal process $X_{0:t} = \{X_0,...,X_t\}$ based on observations $Z_{1:t} = \{Z_1,...,Z_t\}$, i.e. to obtain $\hat{X}_t = \{\hat{x}_{t,1},...,\hat{x}_{t,\hat{T}_t}\}$, where $\hat{x}_{t,i}$ are the individual target estimates and $\hat{T}_t$ is the estimate of the number of targets at time t.

The Bayesian recursion for the multiple-target tracking model is determined from the following prior and posterior calculations:

$$p(X_t|\sigma(Z_{1:t-1})) = \int p(X_t|X_{t-1})p(X_{t-1}|\sigma(Z_{1:t-1}))\delta X_{t-1} \quad (2)$$

$$p(X_t|\sigma(Z_{1:t})) = \frac{p(Z_t|X_t)p(X_t|\sigma(Z_{1:t-1}))}{\int p(Z_t|X_t)p(X_t|\sigma(Z_{1:t-1}))\delta X_t} \quad (3)$$

where $\sigma(Z_{1:t})$ is the σ-algebra generated by $Z_{1:t}$, and the integral is the set integral from Finite Set Statistics [1].

A. The PHD Filter

The Probability Hypothesis Density (PHD) is the first moment of the multiple target posterior distribution [1]. The PHD represents the expectation, the integral of which in any region of the state space S is the expected number of objects in S. The PHD is used, instead of the multiple target posterior distribution, as it is much less computationally expensive to do so. The time-complexity required for calculating joint multi-target likelihoods grows exponentially with the number of targets and is thus not very practical for real-time sequential target estimation. Since the PHD filter consists of the first-order moment of the targets, it is like single-target tracking in an augmented state space and hence offers a significant computational improvement over explicitly calculating joint multi-target likelihoods [12].

The PHD is defined as the density, $D_{t|t}(x_t|Z_{1:t})$, whose integral,

$$\int_A D_{t|t}(x_t|Z_{1:t})dx_t = E\left[\mathcal{N}(A)\right], \quad (4)$$

on any region A of the state space is the expected number of targets in A. The estimated object states can be detected as peaks of this distribution.

The derivation of the PHD equations is provided by Mahler [1]. The prediction and update equations are:

$$D_{t|t-1}(x|Z_{1:t-1}) = \gamma_t(x) + \quad (5)$$
$$\int \phi_{t|t-1}(x,x_{t-1})D_{t-1|t-1}(x_{t-1}|Z_{1:t-1})dx_{t-1},$$

$$D_{t|t}(x|Z_{1:t}) = \left[\nu(x) + \sum_{z\in Z_t}\frac{\psi_{t,z}(x)}{\kappa_t(z)+\langle D_{t|t-1},\psi_{t,z}\rangle}\right]D_{t|t-1}(x|Z_{1:t-1}), \quad (6)$$

where $\phi_{t|t-1}(x,x_{t-1}) = P_S(x_{t-1})f_{t|t-1}(x|x_{t-1}) + b_{t|t-1}(x|x_{t-1}), \nu(x) = 1 - P_D(x)$, $\kappa_t(z) = \lambda_t c_t(z)$ and $\psi_{t,z} = P_D(x)g(z|x)$.

In the prediction equation, γ_t is the PHD for spontaneous birth of a new target at time t, $b_{t|t-1}$ is the PHD of a spawned target, P_S is the probability of target survival and $f_{t|t-1}(x_t|x_{t-1})$ is the single-target motion distribution. In the data update equation, g is the single-target likelihood function, P_D is the probability of detection, λ_t is the Poisson parameter specifying the expected number of false alarms and c_t is the probability distribution over the state space of clutter points. The $\langle .,. \rangle$ notation is defined as the inner product $\langle D_{t|t},\varphi\rangle = \int D_{t|t}(x_t|Z_{1:t})\varphi(x_t)dx_t$.

IV. Particle PHD Filter Algorithm

Our implementation of the PHD Particle filter is an adaptation of the method described by Vo *et al.* [2], based on a Sequential Monte Carlo algorithm for multitarget tracking. The algorithm can be informally described by the following stages. In the initialisation stage, particles are uniformly distributed across the field of view. The particles are propagated in the prediction stage using the dynamic model with added process noise and, in addition, particles are added to allow for incoming targets. When the measurements are received, weights are calculated for the particles based on their likelihoods, which are determined by the statistical distance of the particles to the set of observations. The sum of the weights gives the estimated number of targets. Particles are then resampled from the weighted particle set to give an unweighted representation of the PHD.

The Sequential Monte Carlo implementation of the PHD Filter is given here. The algorithm is initialized in Step 0 and then iterates through Steps 1 to 3.

Step 0: *Initialization at t=0*

The filter is initialized with N_0 particles drawn from a prior distribution. The number of particles is adapted at each stage so that it is proportional to the number of targets. Let N be the number of particles per target. The mass associated with each particle is $\hat{T}_0/N_0$, where $\hat{T}_0$ is the expected initial number of targets, which will be updated after an iteration of the algorithm.

• $\forall i = 1,\ldots,N_0$ sample $x_0^{(i)}$ from $D_{0|0}$ and set $t = 1$.

Let $D_{0|0}^{N_0}$ be the measure:

$$D_{0|0}^{N_0}(dx_t) := \frac{1}{N_0}\sum_{i=1}^{N_0}\delta_{x_t^{(i)}}(dx_t), \quad (7)$$

where $\delta_{x_t^{(i)}}$ is the Dirac delta function centred at $x_t^{(i)}$.

Step 1: *Prediction Step, for* $t \geq 0$

In the prediction step, samples are obtained by two importance sampling proposal densities, q_t and p_t:

• $\forall i = 1,..,N_{t-1}$, sample $\tilde{x}_t^{(i)}$ from a proposal density $q_t(.|x_{t-1}^{(i)}, Z_t)$.

•$\forall i=1,..,N_{t-1}$, evaluate the predicted weights $\tilde{\omega}^{(i)}_{t|t-1}$:

$$\tilde{\omega}^{(i)}_{t|t-1}=\frac{\phi_{t|t-1}(\tilde{x}^{(i)}_t,x^{(i)}_{t-1})}{q_t(\tilde{x}^{(i)}_t|x^{(i)}_{t-1},Z_t)}\omega^{(i)}_{t-1}. \tag{8}$$

M new-born particles are also introduced from the spontaneous birth model to detect new targets entering the state space.
•$\forall i=N_{t-1}+1,..,N_{t-1}+M$, sample $\tilde{x}^{(i)}_t$ from another proposal density $p_t(.|Z_t)$.
•$\forall i=N_{t-1}+1,..,N_{t-1}+M$, compute the weights of new born particles $\tilde{\omega}^{(i)}_{t|t-1}$:

$$\tilde{\omega}^{(i)}_{t|t-1}=\frac{1}{M}\frac{\gamma_t(\tilde{x}^{(i)}_t)}{p_t(\tilde{x}^{(i)}_t|Z_t)}. \tag{9}$$

Let $D^{N_{t-1}}_{t|t-1}$ and $D^{N_{t-1},M}_{t|t-1}$ be the measures:

$$D^{N_{t-1}}_{t|t-1}(dx_t):=\sum_{i=1}^{N_{t-1}}\tilde{\omega}^{(i)}_{t|t-1}\delta_{\tilde{x}^{(i)}_t}(dx_t), \tag{10}$$

$$D^{N_{t-1},M}_{t|t-1}(dx_t):=\sum_{i=1}^{N_{t-1}+M}\tilde{\omega}^{(i)}_{t|t-1}\delta_{\tilde{x}^{(i)}_t}(dx_t). \tag{11}$$

Step 2: *Update Step, for* $t\geq 0$

After the new measurements are obtained, the weights are recalculated using the likelihood function $g(\cdot|\cdot)$ to update the distribution based on new information:
• Let $R_t=N_{t-1}+M$. $\forall z\in Z_t$, compute:

$$\langle\tilde{\omega}_{t|t-1},\psi_{t,z}\rangle=\sum_{i=1}^{R_t}\psi_{t,z}(\tilde{x}^{(i)}_t)\tilde{\omega}^{(i)}_{t|t-1}. \tag{12}$$

•$\forall i=1,\ldots,R_t$, update weights:

$$\tilde{\omega}^{(i)}_t=\left[\nu(\tilde{x}^{(i)}_t)+\sum_{z\in Z_t}\frac{\psi_{t,z}(\tilde{x}^{(i)}_t)}{\kappa_t(z)+\langle\tilde{\omega}_{t|t-1},\psi_{t,z}\rangle}\right]\tilde{\omega}^{(i)}_{t|t-1}. \tag{13}$$

Let $\tilde{D}^{R_t}_{t|t}$ be the measure:

$$\tilde{D}^{R_t}_{t|t}(dx_t):=\sum_{i=1}^{R_t}\tilde{\omega}^{(i)}_t\delta_{x^{(i)}_t}(dx_t). \tag{14}$$

Step 3: *Resampling Step*

The particles are resampled to obtain an unweighted representation of $D_{t|t}$. This is unweighted since the resampled representation of $D_{t|t}$ is given by the particle density.
• Compute the mass of the particles:

$$\hat{T}_t=\sum_{i=1}^{R_t}\tilde{\omega}^{(i)}_t, \tag{15}$$

and set $N_t=N\cdot int(\hat{T}_t)$ (where $int(\hat{T}_t)$ is the integer nearest to $\hat{T}_t$). Target estimates are taken at this stage because the resampling stage introduces further approximations, resulting in less descriptive posterior distributions. In the PHD Filter algorithm, the weights are not normalized as in the standard particle filter algorithm as they do not sum to one but, instead, to the expected number of targets.
• Resample $\left\{\frac{\tilde{\omega}^{(i)}_t}{\hat{T}_t},\tilde{x}^{(i)}_t\right\}_{i=1}^{R_t}$ to get $\left\{\frac{\omega^{(i)}_t}{\hat{T}_t},x^{(i)}_t\right\}_{i=1}^{N_t}$.
The particles each have weight $\hat{T}_t/N_t$ after resampling. Let $D^{N_t}_{t|t}$ be the measure:

$$D^{N_t}_{t|t}(dx_t):=\sum_{i=1}^{N_t}\omega^{(i)}_t\delta_{x^{(i)}_t}(dx_t). \tag{16}$$

V. Convergence for the Particle PHD Filter Algorithm

We now establish some convergence properties for the Particle PHD Filter. First, we consider the rate of convergence of the average mean square error $E\left[(\langle D^{N_t}_{t|t},\varphi\rangle-\langle D_{t|t},\varphi\rangle)^2\right]$ for any function $\varphi\in B(\mathbb{R}^d)$, where $B(\mathbb{R}^d)$ is the set of bounded Borel measurable functions on $\mathbb{R}^d$. Then we show almost-sure convergence of $D^{N_t}_{t|t}$ to $D_{t|t}$. When the measure in the inner product $\langle .,.\rangle$ is continuous, it defines the integral inner product, and when it is discrete, it defines the summation inner product, so that:

$$\langle D_{t|t},\varphi\rangle=\int D_{t|t}(x_t|Z_{1:t})\varphi(x_t)dx_t \tag{17}$$

and

$$\langle D^{N_t}_{t|t},\varphi\rangle=\sum_{i=1}^{N_t}\omega^{(i)}_t\varphi(x^{(i)}_t) \tag{18}$$

The norm $\|\varphi\|$ used here is the supremum norm. (Proofs of the lemmas and theorems used to demonstrate convergence are given in the appendix.)

A. Criteria for Convergence

To show convergence, certain conditions on the functions need to be met:

• The transition kernel $\phi_{t|t-1}$ satisfies the *Feller Property*, i.e. $\forall t>0$, $\int\varphi(y)\phi_{t|t-1}(x,dy)$ is continuous $\forall\varphi\in C_b(\mathbb{R}^d)$, where $C_b(\mathbb{R}^d)$ are the continuous bounded functions on $\mathbb{R}^d$.
•$\psi_{t,z}\in C_b(\mathbb{R}^d)$
• $Q^{(i)}_t$ are rational-valued random variables such that there exists $p>1$, some constant C, and $\alpha<p-1$ so that:

$$E\left[\left|\sum_{i=1}^{N}(Q^{(i)}_t-N\omega^{(i)}_t)q^{(i)}\right|^p\right]\leq CN^{\alpha}\|q\|^p \tag{19}$$

for all vectors $q=(q^{(1)},..,q^{(N)})$ and $\sum_{i=1}^{N}Q^{(i)}_t=N$.
• We assume that the importance sampling ratios are bounded, i.e. there exists constants B_1 and B_2 such that $\|\gamma_t/p_t\|\leq B_1$ and $\|\phi_{t|t-1}/q_t\|\leq B_2$.
• The resampling strategy is multinomial and hence unbiased [7], i.e. the resampled particle set is i.i.d. according to the empirical distribution before resampling.

The data update equation assumes a Poisson model, and hence is only an approximation. The clutter parameter $\kappa_{t,z}$ needs to be determined from the data and cannot be inferred from the recursion. For the purpose of these proofs, it has been assumed that we know the correct density c_t and average number of Poisson clutter points λ_t.

B. Convergence of the Mean Square Errors

If $\mu^N, N = 1, \ldots, \infty$, is a sequence of measures that depend on the number of particles, then we say μ^N converges to μ if $\forall \varphi \in B(\mathbb{R}^d)$,

$$\lim_{N \to \infty} E\left[(\langle \mu^N, \varphi\rangle - \langle \mu, \varphi\rangle)^2\right] = 0. \tag{20}$$

We show that, in the case of the PHD, this depends only on T, the number of targets, and N_t, the number of particles. Let the likelihood function $g \in B(\mathbb{R}^d)$ be a bounded function. At each stage of the algorithm, the approximation admits a mean square error on the order of the number of particles. In particular, we show that given one of the first three properties below, the property below it holds after the next stage of the algorithm.

$$E\left[(\langle D^{N_{t-1}}_{t-1|t-1}, \varphi\rangle - \langle D_{t-1|t-1}, \varphi\rangle)^2\right] \le c_{t-1|t-1}\frac{\|\varphi\|^2}{N_{t-1}}, \tag{21}$$

$$E\left[(\langle D^{N_{t-1},M}_{t|t-1}, \varphi\rangle - \langle D_{t|t-1}, \varphi\rangle)^2\right] \le \|\varphi\|^2\left(\frac{c_{t|t-1}}{N_{t-1}} + \frac{d_t}{M}\right), \tag{22}$$

$$E\left[(\langle \tilde{D}^{R_t}_{t|t}, \varphi\rangle - \langle D_{t|t}, \varphi\rangle)^2\right] \le \tilde{c}_{t|t}\frac{\|\varphi\|^2}{R_t}, \tag{23}$$

$$E\left[(\langle D^{N_t}_{t|t}, \varphi\rangle - \langle D_{t|t}, \varphi\rangle)^2\right] \le c_{t|t}\frac{\|\varphi\|^2}{N_t}. \tag{24}$$

Lemma 0
For any $\varphi \in B(\mathbb{R}^d)$, there exists some real number $c_{0|0}$ such that at **Step 0** (Initialization), condition (24) holds at time $t = 0$.

Lemma 1
Assume that for any $\varphi \in B(\mathbb{R}^d)$, (21) holds. Then, after **Step 1** (Prediction), for any $\varphi \in B(\mathbb{R}^d)$, (22) holds for some constant d_t and some real number $c_{t|t-1}$ that depends on the number of spawned targets.

Lemma 2
Assume that for any $\varphi \in B(\mathbb{R}^d)$, (22) holds. Then, after **Step 2** (Data Update), for any $\varphi \in B(\mathbb{R}^d)$, (23) holds for some real number $\tilde{c}_{t|t}$ that depends on the number of targets.

Lemma 3
Assume that $\forall \varphi \in B(\mathbb{R}^d)$, (23) holds. Then after **Step 3** (Resampling), there exists a real number $c_{t|t}$, that depends on the number of targets, such that $\forall \varphi \in B(\mathbb{R}^d)$, (24) holds.

Theorem 1
$\forall t \ge 0$, there is a real number $c_{t|t}$, that depends on the number of new targets but is independent of the number of particles, such that $\forall \varphi \in B(\mathbb{R}^d)$, (24) holds.

C. Convergence of Empirical Measures

To prove that an empirical distribution converges to its true distribution, we need to have a notion of convergence for measures. This type of convergence is called *weak convergence*, which is fundamental to the study of probability and statistics. With this type of convergence, the values of the random variables are not important; it is the probabilities with which they assume those values that matter. Thus, the probability distributions of the random variables will be converging, not the values themselves [13].

Let μ^N and μ be probability measures on $\mathbb{R}^d$. Then, the sequence μ^N converges weakly to μ if $\int f(x)\mu^N(dx)$ converges to $\int f(x)\mu(dx)$ for each real-valued continuous and bounded function f on $\mathbb{R}^d$.

This definition can be extended to more general measures, not just probability distributions. In our case, we will be considering the PHD measure, where the notion still applies. (Further details on weak convergence for measures can be obtained from Billingsley [14].)

The empirical measures considered here are the particles that approximate the true measures, where N_t represents the number of particles. Let $C_b(\mathbb{R}^d)$ be the set of real-valued continuous bounded functions on $\mathbb{R}^d$. If (μ^N) is a sequence of measures, then μ^N converges weakly to μ if:

$$\lim_{N \to \infty} \langle \mu^N, \varphi\rangle = \langle \mu, \varphi\rangle \tag{25}$$

This section shows that after each stage of the PHD Filter algorithm, the measures converge weakly. In particular, we show that given one of the first three properties below, after the next stage of the algorithm, the property below it holds:

$$\lim_{N_{t-1} \to \infty} D^{N_{t-1}}_{t-1|t-1} = D_{t-1|t-1} \;\; a.s. \tag{26}$$

$$\lim_{N_{t-1}, M \to \infty} D^{N_{t-1},M}_{t|t-1} = D_{t|t-1} \;\; a.s. \tag{27}$$

$$\lim_{R_t \to \infty} \tilde{D}^{R_t}_{t|t} = D_{t|t} \;\; a.s. \tag{28}$$

$$\lim_{N_t \to \infty} D^{N_t}_{t|t} = D_{t|t} \;\; a.s. \tag{29}$$

(a.s. stands for almost surely, i.e. true for all values outside the null set.)

We assume that at time $t = 0$, we can sample exactly from the initial distribution $D_{0|0}$. Then, from the Glivenko-Cantelli Theorem [13], which states that empirical distributions converge to their actual distributions almost surely,

$$\lim_{N_0 \to \infty} D^{N_0}_{0|0} = D_{0|0} \;\; a.s. \tag{30}$$

Lemma 4
Suppose (26) holds, then after **Step 1** (Prediction), (27) holds.

Lemma 5
Suppose (27) holds, then after **Step 2** (Data Update), (28) holds.

Lemma 6
Suppose (28) holds, then after **Step 3** (Resampling), (29)

holds.

Theorem 2
For all $t \geq 0, (29)$ holds.

VI. Discussion

We have shown, under the assumption that $\varphi(x)$ is bounded above, that it is possible to find bounds for the mean square error of the PHD Particle filter at each stage of the algorithm. These depend on the number of targets introduced at each iteration, but if the order of the number of targets is much lower than the order of the number of particles, i.e. $T << N$, then the error tends to zero as N tends to infinity.

We have also shown, under the additional assumptions that the transition kernel satisfies the Feller property and the likelihood function is a continuous bounded function, that the empirical distribution, represented by the particles, converges almost surely to the true PHD distribution. These results are not dependent on the state dimension. The data update equation assumes a Poisson model, and hence is only an approximation. The clutter parameter $\kappa_{t,z}$ needs to be determined from the data and cannot be inferred from the recursion. For the purpose of these proofs, it has been assumed that we know the correct density c_t and average number of Poisson clutter points λ_t.

The assumption that $\varphi(x)$ is bounded above may be too restrictive for practioners, and the additional assumptions on the likelihood and transition kernel may be unrealistic for practical applications; although applications of the PHD filter have demonstrated its potential for real-world applications. Despite these reservations, these results give justification to the Sequential Monte Carlo implementation of the PHD filter, and show how the order of the mean squared error is reduced as the number of particles increases.

VII. Acknowledgements

Thanks to Dr. Dan Crisan at Imperial College, London for help understanding his proofs. This work was funded by QinetiQ through an MoD funded programme.

Appendix I
Proofs of Mean Square Error Bounds

In deriving the proofs, we use the Minkowski inequality, which states that for any two random variables X and Y in L^2:

$$E[(X+Y)^2]^{\frac{1}{2}} \leq E[X^2]^{\frac{1}{2}} + E[Y^2]^{\frac{1}{2}}. \tag{31}$$

A. Proof of Lemma 0

We assume that at time $t = 0$, we can sample exactly from the initial distribution $D_{0|0}$. Then,

$$\langle D_{0|0}^{N_0}, \varphi \rangle - \langle D_{0|0}, \varphi \rangle \tag{32}$$

$$= \frac{\hat{T}_0}{N_0} \sum_{i=1}^{N_0} (\varphi(x_t^{(i)}) - \langle D_{0|0}, \varphi \rangle).$$

Let $\xi_i = \varphi(x_t^{(i)}) - \langle D_{0|0}, \varphi \rangle$. Then $E[\xi_i] = 0$, and $\xi_1, \ldots, \xi_N$ is a sequence of independent integrable random variables. From the Marcinkiewicz and Zygmund inequalities (see, for example, p 498 [15]), there exists a constant c such that

$$E\left[(\frac{1}{N} \sum_{i=1}^{N} \xi_i)^2 \right] \leq cE\left[\frac{1}{N^2} \sum_{i=1}^{N} \xi_i^2 \right]. \tag{33}$$

and hence

$$\frac{1}{N^2} E\left[(\sum_{i=1}^{N} \xi_i)^2 \right] \leq \frac{c\|\xi\|^2}{N}. \tag{34}$$

Therefore, at time $t = 0$, there is a real number $c_{0|0}$ such that

$$E\left[(\langle D_{0|0}^{N_0}, \varphi \rangle - \langle D_{0|0}, \varphi \rangle)^2 \right] \leq c_{0|0} \frac{\|\varphi\|^2}{N_0} \tag{35}$$

so condition (24) holds at the beginning of the algorithm.

B. Proof of Lemma 1

Before proving this Lemma, some considerations are given below. The Sequential Monte Carlo implementation involves sampling from two densities: q_t, the density propagated from the previous time step, and p_t, the density for spontaneous birth. Suppose that the spontaneous birth density is sampled by M particles and the propagated density by N_t particles.

To prove convergence, we use the fact that the sum of two sequences converges weakly to the sum of the limits of those sequences, which follows from a basic result of Real Analysis on the convergence of sequences of real numbers. It then suffices to establish weak convergence of the two sequences independently.

We have assumed that we can sample exactly from the spontaneous birth density γ_t, so using the same argument for showing that the initial distribution is bounded (Lemma 0), and using the assumption that the importance ratio $\|\gamma_t/p_t\|$ is bounded, then there is a constant d_t such that

$$E\left[(\langle \gamma_t^M, \varphi \rangle)^2 \right] \leq d_t \frac{\|\varphi\|^2}{M}. \tag{36}$$

Define $D'_{t|t-1}$ to be $D_{t|t-1} - \gamma_t$. We now show that $D'_{t|t-1}(x)$, the density propagated from the previous time step, is bounded.

Proof
By the triangle inequality, we have

$$|\langle D_{t|t-1}^{N_{t-1}}, \varphi \rangle - \langle D'_{t|t-1}, \varphi \rangle| \leq |\langle D_{t|t-1}^{N_{t-1}}, \varphi \rangle - \langle D_{t-1|t-1}^{N_{t-1}}, \phi_{t|t-1}\varphi \rangle| \tag{37}$$

$$+ |\langle D_{t-1|t-1}^{N_{t-1}}, \phi_{t|t-1}\varphi \rangle - \langle D_{t-1|t-1}, \phi_{t|t-1}\varphi \rangle|.$$

Let $\mathcal{G}_{t-1}$ be the σ-algebra generated by the particles $\{x_{t-1}^{(i)}\}$. Then

$$E\left[(\langle D_{t,t-1}^{N_{t-1}}, \varphi \rangle | \mathcal{G}_{t-1} \right] = \langle D_{t-1,t-1}^{N_{t-1}}, \phi_{t|t-1}\varphi \rangle, \tag{38}$$

hence

$$E\left[(\langle D_{t,t-1}^{N_{t-1}}, \varphi \rangle - E\left[(\langle D_{t,t-1}^{N_{t-1}}, \varphi \rangle | \mathcal{G}_{t-1} \right])^2 | \mathcal{G}_{t-1} \right] \tag{39}$$

$$= E\left[(\langle D_{t,t-1}^{N_{t-1}}, \varphi \rangle - \langle D_{t-1,t-1}^{N_{t-1}}, \phi_{t|t-1}\varphi \rangle)^2 | \mathcal{G}_{t-1} \right]$$

$$= E\left[(\langle D^{N_{t-1}}_{t,t-1},\varphi\rangle(\langle D^{N_{t-1}}_{t,t-1},\varphi\rangle - \langle D^{N_{t-1}}_{t-1,t-1},\phi_{t|t-1}\varphi\rangle)\right] \quad (40)$$

$$- \langle D^{N_{t-1}}_{t-1,t-1},\phi_{t|t-1}\varphi\rangle E\left[\langle D^{N_{t-1}}_{t,t-1},\varphi\rangle - \langle D^{N_{t-1}}_{t-1,t-1},\phi_{t|t-1}\varphi\rangle|\mathcal{G}_{t-1}\right].$$

The second term in (40) is zero, so the above simplifies to

$$E\left[\langle D^{N_{t-1}}_{t,t-1},\varphi\rangle^2\right] - \langle D^{N_{t-1}}_{t-1,t-1},\phi_{t|t-1}\varphi\rangle^2 \quad (41)$$

Writing out this as a sum, and using the independence of the particles, (41) equals

$$\left(\frac{\hat{T}_{t-1}}{N_{t-1}}\right)^2 \sum_{i=1}^{N_{t-1}} (E\left[\left(\varphi(\tilde{x}_t^{(i)})\frac{\phi_{t|t-1}(\tilde{x}_t^{(i)},x_{t-1}^{(i)})}{q_t(\tilde{x}_t^{(i)}|x_{t-1}^{(i)},Z_t)}\right)^2|\mathcal{G}_{t-1}\right] \quad (42)$$
$$-(\phi_{t|t-1}\varphi)(x_{t-1}^{(i)})^2)$$

$$\leq \frac{\hat{T}^2_{t-1}}{N_{t-1}}\|\varphi\|^2\left(\left\|\frac{\phi_{t|t-1}}{q_t}\right\|^2 + \|\phi_{t|t-1}\|^2\right). \quad (43)$$

Using Minkowski's inequality, we obtain

$$E\left[(\langle D^{N_{t-1}}_{t|t-1},\varphi\rangle - \langle D'_{t|t-1},\varphi\rangle)^2\right]^{\frac{1}{2}} \quad (44)$$

$$\leq E\left[(\langle D^{N_{t-1}}_{t|t-1},\varphi\rangle - \langle D^{N_{t-1}}_{t-1|t-1},\phi_{t|t-1}\varphi\rangle)^2\right]^{\frac{1}{2}} \quad (45)$$

$$+ E\left[(\langle D^{N_{t-1}}_{t-1|t-1},\phi_{t|t-1}\varphi\rangle - \langle D_{t-1|t-1},\phi_{t|t-1}\varphi\rangle)^2\right]^{\frac{1}{2}}$$

$$\leq \frac{1}{\sqrt{N_{t-1}}}\|\varphi\|\left(\hat{T}_{t-1}\left(\left\|\frac{\phi_{t|t-1}}{q_t}\right\|^2 + \|\phi_{t|t-1}\|^2\right)^{\frac{1}{2}} + \sqrt{c_{t-1|t-1}}\right). \quad (46)$$

The transition kernel $\phi_{t|t-1}$ is bounded by the single-target transition, $f_{t|t-1}$, and the PHD of spawned targets, $b_{t|t-1}$:

$$\phi_{t|t-1}(x,x_{t-1}) = P_S(x_{t-1})f_{t|t-1}(x|x_{t-1}) + b_{t|t-1}(x|x_{t-1}). \quad (47)$$

Therefore $\|\phi_{t|t-1}\varphi\|^2 \leq 1 + T_{t|t-1}$, where $T_{t|t-1}$ is the number of spawned targets. By assumption, the ratio $\|\phi_{t|t-1}/q_t\|$ is bounded by some constant B_2, and so the lemma is proved:

$$E\left[(\langle D^{N_{t-1},M}_{t|t-1},\varphi\rangle - \langle D_{t|t-1},\varphi\rangle)^2\right] \leq \|\varphi\|^2\left(\frac{c_{t|t-1}}{N_{t-1}} + \frac{d_t}{M}\right), \quad (48)$$

where $c_{t|t-1} = \left(\hat{T}_{t-1}(B_2^2 + (1+T_{t|t-1})^2)^{\frac{1}{2}} + \sqrt{c_{t-1|t-1}}\right)^2$.

C. Proof of Lemma 2

From the definitions (17) and (18), we have:

$$\langle \tilde{D}^{R_t}_{t|t},\varphi\rangle - \langle D_{t|t},\varphi\rangle \quad (49)$$

$$= \left\langle\left[\nu + \sum_{z\in Z_t}\frac{\psi_{t,z}}{\kappa_{t,z} + \langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle}\right]D^{N_{t-1},M}_{t|t-1},\varphi\right\rangle \quad (50)$$

$$-\left\langle\left[\nu + \sum_{z\in Z_t}\frac{\psi_{t,z}}{\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle}\right]D_{t|t-1},\varphi\right\rangle$$

(by linearity)

$$= \left(\langle D^{N_{t-1},M}_{t|t-1},\varphi\nu\rangle - \langle D_{t|t-1},\varphi\nu\rangle\right) \quad (51)$$

$$+ \sum_{z\in Z_t}\left(\frac{\langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle} - \frac{\langle D_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle}\right)$$

(adding and subtracting a new term)

$$= \left(\langle D^{N_{t-1},M}_{t|t-1},\varphi\nu\rangle - \langle D_{t|t-1},\varphi\nu\rangle\right) \quad (52)$$

$$+ \sum_{z\in Z_t}(\left(\frac{\langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle} - \frac{\langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle}\right)$$

$$+ \left(\frac{\langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle} - \frac{\langle D_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle}\right)).$$

The first bracket in the summation from (52) is:

$$\left|\frac{\langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle} - \frac{\langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle}\right| \quad (53)$$

$$= \frac{\left|\langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle(\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle) - \langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle(\kappa_{t,z} + \langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle)\right|}{(\kappa_{t,z} + \langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle)(\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle)} \quad (54)$$

$$\leq \frac{\left|\langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle(\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle) - \langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle(\kappa_{t,z} + \langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle)\right|}{\langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle\langle D_{t|t-1},\psi_{t,z}\rangle} \quad (55)$$

$$= \frac{\langle D^{N_{t-1},M}_{t|t-1},\varphi\psi_{t,z}\rangle\left|eD_{t|t-1},\psi_{t,z}\rangle - \langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle\right|}{\langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle\langle D_{t|t-1},\psi_{t,z}\rangle} \quad (56)$$

$$\leq \frac{\|\varphi\|}{\langle D_{t|t-1},\psi_{t,z}\rangle}\left|\langle D_{t|t-1},\psi_{t,z}\rangle - \langle D^{N_{t-1},M}_{t|t-1},\psi_{t,z}\rangle\right|. \quad (57)$$

The second bracket in the summation from (52) is

$$\left|\frac{\langle D_{t|t-1}^{N_{t-1},M}, \varphi\psi_{t,z}\rangle}{\kappa_{t,z}+\langle D_{t|t-1},\psi_{t,z}\rangle} - \frac{\langle D_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z}+\langle D_{t|t-1},\psi_{t,z}\rangle}\right| \quad (58)$$

$$= \frac{\left|\langle D_{t|t-1}^{N_{t-1},M}, \varphi\psi_{t,z}\rangle - \langle D_{t|t-1},\varphi\psi_{t,z}\rangle\right|}{\kappa_{t,z}+\langle D_{t|t-1},\psi_{t,z}\rangle}$$

$$\leq \frac{\left|\langle D_{t|t-1}^{N_{t-1},M}, \varphi\psi_{t,z}\rangle - \langle D_{t|t-1},\varphi\psi_{t,z}\rangle\right|}{\langle D_{t|t-1},\psi_{t,z}\rangle}. \quad (59)$$

Combining these, we get

$$\left(\langle D_{t|t-1}^{N_{t-1},M}, \varphi\nu\rangle - \langle D_{t|t-1},\varphi\nu\rangle\right) + \quad (60)$$

$$\sum_{z\in Z_t}\Bigg(\left(\frac{\langle D_{t|t-1}^{N_{t-1},M}, \varphi\psi_{t,z}\rangle}{\kappa_{t,z}+\langle D_{t|t-1}^{N_{t-1},M},\psi_{t,z}\rangle} - \frac{\langle D_{t|t-1}^{N_{t-1},M}, \varphi\psi_{t,z}\rangle}{\kappa_{t,z}+\langle D_{t|t-1},\psi_{t,z}\rangle}\right)$$

$$+\left(\frac{\langle D_{t|t-1}^{N_{t-1},M}, \varphi\psi_{t,z}\rangle}{\kappa_{t,z}+\langle D_{t|t-1},\psi_{t,z}\rangle} - \frac{\langle D_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z}+\langle D_{t|t-1},\psi_{t,z}\rangle}\right)\Bigg)$$

$$\leq |\langle D_{t|t-1}^{N_{t-1},M}, \varphi\nu\rangle - \langle D_{t|t-1},\varphi\nu\rangle| \quad (61)$$

$$+\sum_{z\in Z_t}\Bigg(\frac{\|\varphi\|}{\langle D_{t|t-1},\psi_{t,z}\rangle}\left|\langle D_{t|t-1},\psi_{t,z}\rangle - \langle D_{t|t-1}^{N_{t-1},M},\psi_{t,z}\rangle\right|$$

$$+\frac{\left|\langle D_{t|t-1}^{N_{t-1},M}, \varphi\psi_{t,z}\rangle - \langle D_{t|t-1},\varphi\psi_{t,z}\rangle\right|}{\langle D_{t|t-1},\psi_{t,z}\rangle}\Bigg).$$

From Minkowski's inequality,

$$E\left[(\langle \tilde{D}_{t|t}^{R_t},\varphi\rangle - \langle D_{t|t},\varphi\rangle)^2\right]^{\frac{1}{2}} \quad (62)$$

$$\leq E\left[(\langle D_{t|t-1}^{N_{t-1},M}, \varphi\nu\rangle - \langle D_{t|t-1},\varphi\nu\rangle)^2\right]^{\frac{1}{2}}$$

$$+\sum_{z\in Z_t}\Bigg(\frac{\|\varphi\|}{\langle D_{t|t-1},\psi_{t,z}\rangle}E\left[(\langle D_{t|t-1},\psi_{t,z}\rangle - \langle D_{t|t-1}^{N_{t-1},M},\psi_{t,z}\rangle)^2\right]^{\frac{1}{2}}$$

$$+\frac{E\left[(\langle D_{t|t-1}^{N_{t-1},M}, \varphi\psi_{t,z}\rangle - \langle D_{t|t-1},\varphi\psi_{t,z}\rangle)^2\right]^{\frac{1}{2}}}{\langle D_{t|t-1},\psi_{t,z}\rangle}\Bigg)$$

$$\leq \frac{\sqrt{c_{t|t-1}}\|\varphi\|\|\nu\|}{\sqrt{R_t}} + \sum_{z\in Z_t}\left(\frac{2\|\varphi\|\|\psi_{t,z}\|\sqrt{c_{t|t-1}}}{\langle D_{t|t-1},\psi_{t,z}\rangle\sqrt{R_t}}\right) \quad (63)$$

$$\leq \frac{\sqrt{c_{t|t-1}}\|\varphi\|}{\sqrt{R_t}}\left[1+\sum_{z\in Z_t}\left(\frac{2\|\psi_{t,z}\|}{\langle D_{t|t-1},\psi_{t,z}\rangle}\right)\right]. \quad (64)$$

$\psi_{t,z}$ is a bounded function, since g is bounded by assumption. Lemma 2 follows from this where $\tilde{c}_{t|t} = c_{t|t-1}\left[1+\sum_{z\in Z_t}\left(\frac{2\|\psi_{t,z}\|}{\langle D_{t|t-1},\psi_{t,z}\rangle}\right)\right]^2$.

D. Proof of Lemma 3

Adding and subtracting the term $\langle \tilde{D}_{t|t}^{R_t},\varphi\rangle$ from the Data Update step, we have

$$\langle D_{t|t}^{N_t},\varphi\rangle - \langle D_{t|t},\varphi\rangle \quad (65)$$

$$= (\langle D_{t|t}^{N_t},\varphi\rangle - \langle \tilde{D}_{t|t}^{R_t},\varphi\rangle) + (\langle \tilde{D}_{t|t}^{R_t},\varphi\rangle - \langle D_{t|t},\varphi\rangle),$$

so by Minkowski's inequality,

$$E\left[(\langle D_{t|t}^{N_t},\varphi\rangle - \langle D_{t|t},\varphi\rangle)^2\right]^{\frac{1}{2}} \quad (66)$$

$$\leq E\left[(\langle D_{t|t}^{N_t},\varphi\rangle - \langle \tilde{D}_{t|t}^{R_t},\varphi\rangle)^2\right]^{\frac{1}{2}} + E\left[(\langle \tilde{D}_{t|t}^{R_t},\varphi\rangle - \langle D_{t|t},\varphi\rangle)^2\right]^{\frac{1}{2}}.$$

Let $\mathcal{F}_t$ be the σ-algebra generated by $\{\tilde{x}^{(i)}\}, i=1,\ldots,R_t$ Then the expectation of the inner product $\langle D_{t|t}^{N_t},\varphi\rangle$ conditioned on $\mathcal{F}_t$ is

$$E\left[\langle D_{t|t}^{N_t},\varphi\rangle|\mathcal{F}_t\right] = \langle \tilde{D}_{t|t}^{R_t},\varphi\rangle. \quad (67)$$

Hence there exists a number c such that

$$E\left[(\langle D_{t|t}^{N_t},\varphi\rangle - \langle \tilde{D}_{t|t}^{R_t},\varphi\rangle)^2|\mathcal{F}_t\right] \leq \frac{c}{N_t}\|\varphi\|^2. \quad (68)$$

This follows from the assumption that the resampling strategy is unbiased. Using Minkowski's inequality, as above, we have

$$E\left[(\langle D_{t|t}^{N_t},\varphi\rangle - \langle D_{t|t},\varphi\rangle)^2\right]^{\frac{1}{2}} \leq (\sqrt{c}+\sqrt{\tilde{c}_{t|t}})\frac{\|\varphi\|}{\sqrt{N_t}}. \quad (69)$$

Lemma 3 is then proved with $c_{t|t} = (\sqrt{c}+\sqrt{\tilde{c}_{t|t}})$.

E. Proof of Theorem 1

Combining the above proofs, we have shown that $\forall t \geq 0, \exists c_{t|t}$ independent of N_t, but dependent on the number of targets, such that $\forall \varphi \in B(\mathbb{R}^d)$:

$$E\left[(\langle D_{t|t}^{N_t},\varphi\rangle - \langle D_{t|t},\varphi\rangle)^2\right] \leq c_{t|t}\frac{\|\varphi\|^2}{N_t}. \quad (70)$$

APPENDIX II
PROOFS OF WEAK CONVERGENCE

A. Proof of Lemma 4

Define $D'_{t|t-1}$ to be $D_{t|t-1}-\gamma_t$. It suffices to prove that

$$\lim_{M\to\infty}\gamma_t^M = \gamma_t \ \ a.s. \quad (71)$$

and

$$\lim_{N_{t-1}\to\infty} D_{t|t-1}^{N_{t-1}} = D'_{t|t-1} \ \ a.s. \quad (72)$$

We have assumed that we sample M i.i.d. particles from γ_t for the first of these, so by the Glivenko-Cantelli Theorem, (71) is true. Let $\mathcal{G}_{t-1}$ be the σ-algebra generated by $\{x_{0:t-1}\}_{i=1}^{N_{t-1}}$, then

$$E\left[\langle D_{t|t-1}^{N_{t-1}},\varphi\rangle|\mathcal{G}_{t-1}\right] = \langle D_{t-1|t-1}^{N_{t-1}}, \phi_{t|t-1}\varphi\rangle, \quad (73)$$

Since $E\left[\varphi(x_t^{(i)})|\mathcal{G}_{t-1}\right] = \phi_{t|t-1}\varphi(x_{t-1}^{(i)})$ and $\{\tilde{x}_{0:t}\}_{i=1}^{N_{t-1}}$ are i.i.d. random variables which are conditional on $\mathcal{G}_{t-1}$, we have

$$E\left[(\langle D_{t|t-1}^{N_{t-1}},\varphi\rangle - E\left[\langle D_{t|t-1}^{N_{t-1}},\varphi\rangle|\mathcal{G}_{t-1}\right])^4|\mathcal{G}_{t-1}\right] \quad (74)$$

$$= E\left[\left(\frac{\hat{T}_{t-1}}{N_{t-1}}\sum_{i=1}^{N_{t-1}}(\varphi(\tilde{x}_t^{(i)})\frac{\phi_{t|t-1}(\tilde{x}_t^{(i)},x_{t-1}^{(i)})}{q_t(\tilde{x}_t^{(i)}|x_{t-1}^{(i)},Z_t)} - (\phi_{t|t-1}\varphi)(x_{t-1}^{(i)})\right)^4|\mathcal{G}_{t-1}\right]. \quad (75)$$

For notational simplicity, define the measure Φ as

$$\Phi(\tilde{x}_t^{(i)},x_{t-1}^{(i)}) = \varphi(\tilde{x}_t^{(i)})\frac{\phi_{t|t-1}(\tilde{x}_t^{(i)},x_{t-1}^{(i)})}{q_t(\tilde{x}_t^{(i)}|x_{t-1}^{(i)},Z_t)} - (\phi_{t|t-1}\varphi)(x_{t-1}^{(i)}) \quad (76)$$

Then, expanding the above quartic gives

$$=\left(\frac{\hat{T}_{t-1}}{N_{t-1}}\right)^4 (\sum_{i=1}^{N_{t-1}} E\left[\Phi(\tilde{x}_t^{(i)},x_{t-1}^{(i)})^4|\mathcal{G}_{t-1}\right] \quad (77)$$

$$+\sum_{i\neq j}^{N_{t-1}} E\left[\Phi(\tilde{x}_t^{(i)},x_{t-1}^{(i)})^3\Phi(\tilde{x}_t^{(j)},x_{t-1}^{(j)}) + \Phi(\tilde{x}_t^{(i)},x_{t-1}^{(i)})^2\Phi(\tilde{x}_t^{(j)},x_{t-1}^{(j)})^2|\mathcal{G}_{t-1}\right]$$

$$+\sum_{i,j,k\ distinct}^{N_{t-1}} E\left[\Phi(\tilde{x}_t^{(i)},x_{t-1}^{(i)})^2\Phi(\tilde{x}_t^{(j)},x_{t-1}^{(j)})|\mathcal{G}_{t-1}\right] E\left[\Phi(\tilde{x}_t^{(k)},x_{t-1}^{(k)})|\mathcal{G}_{t-1}\right]$$

$$+\sum_{i,j,k,l\ distinct}^{N_{t-1}} E\left[\Phi(\tilde{x}_t^{(i)},x_{t-1}^{(i)})\Phi(\tilde{x}_t^{(j)},x_{t-1}^{(j)})\Phi(\tilde{x}_t^{(k)},x_{t-1}^{(k)})\Phi(\tilde{x}_t^{(l)},x_{t-1}^{(l)})|\mathcal{G}_{t-1}\right])$$

$$=\left(\frac{\hat{T}_{t-1}}{N_{t-1}}\right)^4 (\sum_{i=1}^{N_{t-1}} E\left[\Phi(\tilde{x}_t^{(i)},x_{t-1}^{(i)})^4|\mathcal{G}_{t-1}\right] \quad (78)$$

$$+\sum_{i\neq j}^{N_{t-1}} E\left[\Phi(\tilde{x}_t^{(i)},x_{t-1}^{(i)})^2\Phi(\tilde{x}_t^{(j)},x_{t-1}^{(j)})^2|\mathcal{G}_{t-1}\right]),$$

where the last equality holds because the particles are mutually independent random variables with mean zero. Taking expectations of (75) and (76), there exists a constant C such that

$$E\left[(\langle D_{t|t-1}^{N_{t-1}},\varphi\rangle - E\left[\langle D_{t|t-1}^{N_{t-1}},\varphi\rangle|\mathcal{G}_{t-1}\right])^4\right] \quad (79)$$

$$\leq \frac{C\hat{T}_{t-1}^4(B_2^4 + (1+T_{t|t-1})^4)\|\varphi\|^4}{N_{t-1}^2},$$

since there are $O(N_{t-1}^2)$ terms bounded by $C\hat{T}_{t-1}^4(B_2^4 + (1+T_{t|t-1})^4)\|\varphi\|^4$, following a similar argument as in Lemma 1. It then follows that

$$E\left[(\langle D_{t|t-1}^{N_{t-1}},\varphi\rangle - \langle D_{t-1|t-1}^{N_{t-1}},\phi_{t|t-1}\varphi\rangle)^4\right] \quad (80)$$

$$\leq \frac{C\hat{T}_{t-1}^4(B_2^4 + (1+T_{t|t-1})^4)\|\varphi\|^4}{N_{t-1}^2},$$

and hence

$$\lim_{N_{t-1}\to\infty}(\langle D_{t|t-1}^{N_{t-1}},\varphi\rangle - \langle D_{t-1|t-1}^{N_{t-1}},\phi_{t|t-1}\varphi\rangle) = 0. \quad (81)$$

Using the result from Real Analysis, we have

$$\lim_{N_{t-1},M\to\infty}\langle D_{t|t-1}^{N_{t-1},M},\varphi\rangle = \lim_{N_{t-1}\to\infty}\langle D_{t|t-1}^{N_{t-1}},\varphi\rangle + \lim_{M\to\infty}\langle\gamma_t^M,\varphi\rangle \quad (82)$$

$$= \langle D'_{t|t-1},\varphi\rangle + \langle\gamma_t,\varphi\rangle = \langle D_{t|t-1},\varphi\rangle. \quad (83)$$

B. Proof of Lemma 5

By definition,

$$\langle\tilde{D}_{t|t}^{R_t},\varphi\rangle = \langle D_{t|t-1}^{N_{t-1}},\varphi\nu\rangle + \sum_{z\in Z_t}\left(\frac{\langle D_{t|t-1}^{N_{t-1}},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D_{t|t-1}^{N_{t-1}},\psi_{t,z}\rangle}\right). \quad (84)$$

By continuity and Lemma 4, we have

$$\lim_{N_{t-1}\to\infty}\langle D_{t|t-1}^{N_{t-1}},\varphi\nu\rangle = \langle D_{t|t-1},\varphi\nu\rangle, \quad (85)$$

$$\lim_{N_{t-1},M\to\infty}\langle D_{t|t-1}^{N_{t-1},M},\varphi\psi_{t,z}\rangle = \langle D_{t|t-1},\varphi\psi_{t,z}\rangle, \quad (86)$$

and

$$\lim_{N_{t-1}\to\infty}\langle D_{t|t-1}^{N_{t-1}},\psi_{t,z}\rangle = \langle D_{t|t-1},\psi_{t,z}\rangle. \quad (87)$$

Hence,

$$\lim_{R_t\to\infty}\langle\tilde{D}_{t|t}^{R_t},\varphi\rangle = \langle D_{t|t-1},\varphi\nu\rangle + \sum_{z\in Z_t}\left(\frac{\langle D_{t|t-1},\varphi\psi_{t,z}\rangle}{\kappa_{t,z} + \langle D_{t|t-1},\psi_{t,z}\rangle}\right) \quad (88)$$

$$= \langle D_{t|t},\varphi\rangle,$$

and therefore

$$\lim_{R_t\to\infty}\tilde{D}_{t|t}^{R_t} = D_{t|t}\ \ a.s. \quad (89)$$

C. Proof of Lemma 6

Let $P_t^{(i)}$ be the number of times that particle $\tilde{x}_t^{(i)}$ is resampled and let $Q_t^{(i)} = P_t^{(i)}\cdot R_t/N_t$. Then, from our assumption, we have

$$E\left[|\langle D_{t|t}^{N_t},\varphi\rangle - \langle\tilde{D}_{t|t}^{R_t},\varphi\rangle|^p\right] = E\left[(\frac{1}{R_t}\sum_{i=1}^{R_t}|(Q_t^{(i)} - R_t\tilde{\omega}_t^{(i)})\varphi(\tilde{x}_t^{(i)})|)^p\right] \quad (90)$$

$$\leq \frac{C\|\varphi\|^p}{R_t^{1+\varepsilon}},$$

where $\varepsilon = p - \alpha - 1 \geq 0$. Hence,

$$\lim_{N_t\to\infty}\langle D_{t|t}^{N_t},\varphi\rangle - \langle\tilde{D}_{t|t}^{R_t},\varphi\rangle = 0\ \ a.s. \quad (91)$$

D. Proof of Theorem 2

The above three proofs have shown that for all $t \geq 0, \lim_{N_t\to\infty} D_{t|t}^{N_t} = D_{t|t}$.

REFERENCES

[1] R. Mahler, “Multitarget Bayes Filtering via First-Order Multitarget Moments,” *IEEE Transactions on Aerospace and Electronic Systems*, 2003.

[2] B.-N. Vo, S. Singh, and A. Doucet, “Sequential Monte Carlo Implementation of the PHD Filter for Multi-target Tracking.” *Proc. FUSION 2003*, pp. 792–799, 2003.

[3] T. Zajic and R. Mahler, “A particle-systems implementation of the PHD multitarget tracking filter,” *SPIE Vol. 5096 Signal Processing, Sensor Fusion and Target Recognition*, pp. 291–299, 2003.

[4] H. Sidenbladh, “Multi-target particle filtering for the Probability Hypothesis Density,” *International Conference on Information Fusion*, pp. 800–806, 2003.

[5] M. Tobias and A. Lanterman, "A Probability Hypothesis Density-based multitarget tracker using multiple bistatic range and velocity measurements," *Proceedings of the Thirty-Sixth Southeastern Symposium on System Theory, March 14-16, 2004*, pp. 205 – 209.
[6] D. Clark and J. Bell, "Bayesian Multiple Target Tracking in Forward Scan Sonar Images Using the PHD Filter," *IEE Radar, Sonar and Navigation, to appear*, 2005.
[7] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering for practitioners," 2002. [Online]. Available: citeseer.ist.psu.edu/crisan02survey.html
[8] ——, "Convergence of sequential Monte Carlo methods," 2000. [Online]. Available: citeseer.ist.psu.edu/crisan00convergence.html
[9] D. Crisan, *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001, ch. 2, pp. 17–41.
[10] D. Daley and D. Vere-Jones, *An introduction to the theory of point processes*. Springer, 1988.
[11] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, 2nd ed. New York: Wiley, 1995.
[12] D. Hall and J. Llinas, Eds., *Handbook of Multisensor Data Fusion*. CRC Press, 2001, ch. 7.
[13] J. Jacod and P. Protter, *Probability essentials*. Springer, 2000.
[14] P. Billingsley, *Convergence of probability measures*. Wiley, New-York, 1968.
[15] A. N. Shiryaev, *Probability*. Springer, 1996.

Daniel Edward Clark was born in Lancashire, England in 1979. He received the B.Sc.(Hons) in Mathematics from the University of Edinburgh in 2001. He subsequently took the Diploma in Computer Science course at the University of Cambridge which he received in 2002.

He is currently working on a Ph.D. in Sonar Signal Processing at Heriot-Watt University in Edinburgh under the supervision of Judith Bell and supported by QinetiQ.

Judith Bell received the M.Eng. degree (with Merit) in Electrical and Electronic Engineering in 1992 and her PhD in 1995, both from Heriot-Watt University. Her thesis examined the simulation of sidescan sonar images. She is currently a Senior Lecturer in the School of Engineering and Physical Sciences at Heriot-Watt University and is extending the modelling and simulation work to include a range of sonar systems and to examine the use of such models for the verification and development of algorithms for processing sonar images.