Parallel friendly particle filtering approach with disturbance sampling

Deepthi Kattula*, Rajesh K. Pullagura, Praveen B. Choppala*
Dept. of Electronics & Communication Engineering
Andhra University, Visakhapatnam, India
{kdeepthi.rs, prof.prkumar, praveenchoppala}@andhrauniversity.edu.in

***Abstract*—Particle filtering is Bayesian filtering technique used for tracking moving targets from nonlinear dynamics. The resampling step of the filter is computationally complex and mitigates its use in real-time tracking applications. This paper presents a methodology that bypasses the resampling step and fairly parallelises the particle filter. This is done by sampling multiple target heading disturbances for each particles iteratively until unbiasedness is reached. This idea is highly parallel friendly because all particles are processed in batches, but not sequentially as does the resampling process. The merits of the proposed method is shown using a simulated study.**

***Index Terms*—particle filter, importance sampling, resampling, multiple sampling, root mean square error, parallel-friendly.**

I. Introduction

The particle filter (PF) is a nonlinear nonGaussian Bayesian filtering tool used for tracking latent dynamic target states from noisy sensor measurements [1], [2]. The filter uses a set of Monte Carlo samples to represent the posterior probability density function (PDF) of the desired target state. It is known that the accuracy of the representation of the posterior by the PF increases asymptotically with the increase in the number of particles. However, the sequential resampling step of the PF is a major bottleneck for the use of a large number of particles or the use of multicore parallel systems for real-time PF implementation [3]–[5]. The interest of this paper is to develop a PF strategy that represents the posterior PDF accurately using fewer particles and that is extremely parallel-friendly.

The PF resamplers that have found extensive use in the literature are the stochastic resamplers [1], [2], [6]. They employ sequential multinomial selection over a normalised set of weights. Using random variables, i.e., samples drawn from a uniform distribution in the interval $(0, 1]$ for each particle involves a successive operation that requires extensive communication within the particles when a large number of particles are used by the filter [7]. The consequent time complexity prohibits the use of a large number of particles which otherwise would lead to a more accurate approximation of the posterior PDF. As another case, the residual resampler [8] improved upon the stochastic resampling approach by replicating the particles deterministically using principle of proportional allocation, which is also treated as the unbiasedness criterion [9], [10]. Only a small proportion of the particles needed to be stochastically resampled. This criterion was used in soft resampling [11] and the rounding copy PF (RCPF) [12] to develop fast PF resampling strategies that replicate particles based on the unbiasedness rule and follow random sampling for any residuals. Alternative to random resampling, a mixed class of stochastic and deterministic resamplers can be found in [13]. A new class of resamplers that are suitable for hardware architectures have been recently developed in [14]–[16], the key idea of which is to reduce the communication overhead within the particles using architectures that support parallel execution within the minimised set.

Now that we studied resampling methods that aim to mitigate degeneracy and accelerate the PF, we shall focus on PF strategies that aim to achieve accurate tracking using fewer particles. The auxiliary particle filter (APF) [17] is one of the most widely used methods in this category. It generates a set of look-ahead particles, assigns weights to them, and then performs resampling. The resampling indices are subsequently used to guide the propagation of the original particles to the next time step. This allows particles to intelligently explore the state space and hence mitigates the need for many particles. The recently proposed improved APF (IAPF) [18] and other lookahead strategies allow use of fewer particles.

Our contribution: In this paper, we propose to sample multiple children particles for each parent particle corresponding to the previous time step. This can be accomplished by generating multiple target heading disturbances from the Markov state transition density. We propose to sample disturbances until the rule of proportional allocation is reached. The key merit of this proposal is twofold: (a) the resultant particle set represents the posterior accurately, and (b) sampling of excess children particles from parent particles using disturbances is a fully parallel operation.

The rest of the paper is organised as follows. The Bayesian PF estimation methodology is first presented in section II followed by the the proposed iterative disturbance sampling approach in section III, followed by simulation study in section IV and

* Corresponding Author.

concluding remarks in section V.

II. Bayesian estimation

Let the set of latent target states and the noisy observations until the tth time step be denoted as $\mathbf{x_{1:t}} = \{\mathbf{x}_1, \cdots, \mathbf{x}_t\}$ and $\mathbf{y_{1:t}} = \{\mathbf{y}_1, \cdots, \mathbf{y}_t\}$. The dynamics of the target and the observations are encapsulated in the state space model, defined as

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{a}_t) \tag{1}$$

$$\mathbf{y}_t = h(\mathbf{x}_t, \mathbf{e}_t) \tag{2}$$

where the nonlinear functions $f(.)$ and $h(.)$ respectively are the state transition and sensor observation functions, and $\mathbf{a}_t$ and $\mathbf{e}_t$ respectively are the i.i.d. target heading disturbance and the observation noise. The aim of Bayesian filtering is to construct a PDF of the desired target state using all available observations. This is described according to

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|y_{1:t-1})\, d\mathbf{x}_{t-1} \tag{3}$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is the Markovian state transition density obtained from (1), $p(\mathbf{y}_t|\mathbf{x}_t)$ is the observation density obtained from (2) and $p(\mathbf{x}_{t-1}|y_{1:t-1})$ is the PDF at time $t-1$.

Assuming that the posterior PDF of the target state $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ at time $t-1$ is represented by a set of particles and their corresponding weights $\{\mathbf{x}^i_{t-1}, w^i_{t-1}\}^N_{i=1}$, the PF then advances to time t by generates a new set of particles from the importance distribution according to

$$\bar{\mathbf{x}}^i_t \sim q(\mathbf{x}_t|\mathbf{x}^i_{t-1}, \mathbf{y}_t),\ i = 1, \cdots, N \tag{4}$$

If the importance sampling function $q(.)$ is conveniently chosen to be the Markov transition prior, we obtain,

$$q(\mathbf{x}_t|\mathbf{x}^i_{t-1}, \mathbf{y}_t) = p(\mathbf{x}_t|\mathbf{x}^i_{t-1}), i = 1, \cdots, N \tag{5}$$

The new particles are then weighted as

$$\bar{w}^i_t \propto w^i_{t-1} p(\mathbf{y}_t|\bar{\mathbf{x}}^i_t),\ i = 1, \cdots, N \tag{6}$$

Once normalised, these weighted set of particles are representative of the posterior PDF at time t as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{N} \bar{w}^i_t \delta(\mathbf{x}_t - \bar{\mathbf{x}}^i_t) \tag{7}$$

Over multiple iterations of time, the weight distribution becomes uneven, resulting in degeneracy. Degeneracy is a case where in particle grabs all the weight and the rest are rendered useless in representing the posterior. This problem is addressed by resampling where particles having lower weights are duplicated by those having higher weights. That is, for $i = 1, \cdots, N$, an index $j(i)$ is sampled with probability $\mathrm{P}(j(i) = m) = \bar{w}^m_t, m = 1, \cdots, N$, and the corresponding particle is replaced as $\mathbf{x}^i_t = \bar{\mathbf{x}}^{j(i)}_t$ and the weight is set to $w^i_t = 1/N$. It can be observed that the resampling step is a sequential operation over the particles and hence impedes the fast operation of the PF, especially when using a large number of particles for an accurate representation of the posterior. In the next section, we present a method that renders the PF completely parallel friendly while being able to maintain unbiassedness and high accuracy in representing the posterior PDF.

III. Proposed iterative disturbance sampling

In this section, we present the iterative disturbance sampling method that completely bypasses the need to resample within a sequential framework. Following (4) and (6), we predict a new particle set and their updated normalised weights $\{\bar{\mathbf{x}}^i_t, \bar{w}^i_t\}^N_{i=1}$. As a general principle of unbiassedness, the resampling of this set should preserve the original particle distribution if there is no more information being considered in the process. That is, the expected number of times that each particle is resampled, n^i_t, is proportional to its weight as

$$n^i_t = \lfloor N\bar{w}^i_t \rfloor, i = 1, \cdots, N \tag{8}$$

Using this principle, we propose the following PF approach.

The state space model described in (1) and the corresponding state transition density $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \Theta)$ defines the process of generating new particles (read, children) at time t from the previous particles (read, parents) at time $t-1$. The target heading disturbance in (1) which models the randomness associated with the target state transition from $t-1$ to t is encapsulated within the i.i.d. disturbance variable $\mathbf{a}_t$ generated from its distribution $p(\mathbf{a}_t|\Theta)$. In our proposed method, instead of generating one particle, i.e., child, from each previous particle, i.e., parent, as is the case in the traditional PF, we propose to generate multiple children at time t for each parent particle at time $t-1$. This can be be achieved by *(a)* generating multiple disturbances $\mathbf{a}^{i,j}_t, i = 1, \cdots, N, j = 1, \cdots, M$ as $\mathbf{a}^{i,j}_t \sim p(\mathbf{a}_t|\Theta)$, i.e., M disturbances for each ith particle, *(b)* generating one child particle at t according to (4) for each parent particle $\mathbf{x}^i_{t-1}$, and *(c)* then appending the generated disturbances to each child particle as follows

$$\bar{\mathbf{x}}^{i,j}_t = f(\mathbf{x}^i_{t-1}, \mathbf{a}^{i,j}_t),\ i = 1, \cdots, N, j = 1, \cdots, M \tag{9}$$

This will implicitly form a set of MN particles $\bar{\mathbf{x}}^{i,j}_t, i = 1, \cdots, N, j = 1, \cdots, M$ completely in a parallelised fashion. i.e., generate M disturbances for each of the N particles, generate N child particles and combine the two using (9). This set can be rewritten as $\{\bar{\mathbf{x}}^i_t\}^{MN}_{i=1}$. We then compute the weights of the generated particles according to (6) for $i = 1, \cdots, MN$ particles and then normalise them to sum to one. The weighted particle set $\{\bar{\mathbf{x}}^i_t, \bar{w}_t\}^{MN}_{i=1}$ will now be representative of the posterior PDF at time t as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{MN} \bar{w}^i_t \delta(\mathbf{x}_t - \bar{\mathbf{x}}^i_t) \tag{10}$$

We then rearrange the particles and their corresponding weights in accordance to the descending order of the weights and obtain a new set $\{\bar{\mathbf{x}}_t^i, \bar{w}_t\}_{i=1}^{MN}$ such that $\bar{w}_t^1 \geq \bar{w}_t^2 \geq \cdots \geq \bar{w}_t^N$. The number of replications for each particle is computed according to (8). This implies that $1/N$ is the minimum weight for a particle to have at least one replication. Then we compute the aggregate of the replications as $S_t = \sum_{i=1}^N n_t^i$. The value $S_t \geq N$ indicates that there are at least N particles, after replication, that satisfy the unbiasedness criterion of resampling. If $S_t < N$, then the value of M should be incremented iteratively to satisfy $S_t \geq N$, hence the name, "disturbance sampling". Since only N particles are required at the output stage, we discard the $MN - N$ particles and their weights to obtain a new weighted particle set as

$$\{\mathbf{x}_t^i\}_{i=1}^N = \{\bar{\mathbf{x}}_t^i\}_{i=1}^N, \ \{w_t^i\}_{i=1}^N = \{\bar{w}_t^i\}_{i=1}^N \tag{11}$$

Since the weights $\{w_t^i\}_{i=1}^N$ do not sum to one, we redistribute the discarded weight, given by,

$$w_t^{\text{discard}} = 1 - \sum_{i=1}^N w_t^i \tag{12}$$

to all the weights as $w_t^i = w_t^i + w_t^{\text{discard}}/N$. This particle set $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$ will then be an accurate and unbiassed representation of the posterior PDF. The key merits of the proposal are that it bypasses the need for the computationally expensive resampling step and works solely on sampling multiple disturbances in an iterative fashion until the unbiasedness condition is achieved. This approach is highly suitable for parallel architectures. Moreover, the proposed method is expected to yield fairly reliable results with very few particles as each particle is drawn with conformity to the unbiasedness property.

IV. Simulation Study

In this section, we present the simulation results for the proposed method. We compare the proposed method against the systematic resampler [2], the residual resampler [8], the APF [17], the RCPF [12], the soft resampler [11] and the recently random network resampler [16]. All the results of this paper average over 5000 for Monte Carlo iterations.

A. Empirical measure of unbiassedness

First we evaluate the accuracy of the proposed method in unbiassedly representing the posterior PDF using the Kalman filter as a benchmark. Therefore we use a linear Gaussian autoregressive AR(1) model described as

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\phi\mathbf{x}_{t-1}, \tau^2)\ , p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_t, \sigma^2) \tag{13}$$

for $t = 1, \cdots, T = 100$, where the model parameters $\Theta = (\phi, \tau^2, \sigma^2) = (0.99, 1, 1)$ and the initial target state is $\mathbf{x}_{t=0} = 0$. The Kolomogorov Smirnov (KS) statistic is a well-known measure used to test the misfit between two distributions. This test was used for the PFs first in [7]. The procedure is as follows:

Following the resampling step, the particles undergo demean and de-correlation operations using the optimal Kalman mean and covariance. This ensures that if the particle set accurately represents the optimal Kalman filter distribution, the resulting set will be uncorrelated with a mean of zero and unit variance. Next, the Kolmogorov-Smirnov (KS) statistic is calculated by comparing the cumulative sum of the weights with the error function of a standard Gaussian distribution. A lower KS statistic value indicates that the particle set closely approximates the posterior probability density function (PDF) of the Kalman filter, and this value reaches zeros asymptotically as $N \longrightarrow \infty$.
Figure 1 shows the KS statistic versus the number of particles for the given model, and it can be observed that our proposed method achieves the lowest KS statistic value with fewer particles. The key reason for this is that we iteratively sample multiple particles (from multiple disturbances) in a way that obeys the unbiassed criterion of resampling. The network resampler performs worse as each particle is allowed to interact with only log2 of the number of particles, and this worsens as the number particles is kept small.

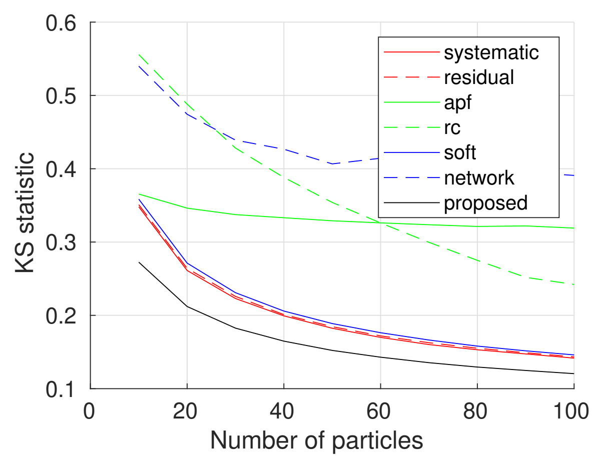

Fig. 1. The KS statistic versus the number of particles.

B. Nonlinear tracking example

The nonlinear state space model is given by

$$\mathbf{x}_t = \frac{\mathbf{x}_{t-1}}{2} + \frac{25\mathbf{x}_{t-1}}{1+\mathbf{x}_{t-1}^2} + 8\cos(1.2t) + a_t \tag{14}$$

$$\mathbf{y}_t = \frac{\mathbf{x}_t^2}{20} + e_t \tag{15}$$

for $t = 1, \cdots, T$, where the heading disturbance is $a_t \sim \mathcal{N}(0, \tau^2)$ and the observation noise is $e_t \sim \mathcal{N}(0, \sigma^2)$ and the total number of time steps is $T = 500$. The model parameters are $\Theta = (\tau^2 = 10, \sigma^2)$. Figure 2 shows the root mean square error (RMSE) versus the number of particles at observation noise variance $\sigma^2 = 1$, and Figure 3 shows the same at $\sigma^2 = 5$. It can be observed that the proposed method exhibits

high tracking accuracy with fewer particles in both noise scenarios. The reason for this is that the sampled particles are made to explore regions of high probability density within the posterior PDF by virtue of sampling iteratively multiple times to conform to the unbiassed criterion in (8). It can also be observed that the network sampler and the RCPF consistently exhibit poor performance as they substantially impede the communication within particles, thus causing information loss.
Now that we have shown that the proposed method faithfully

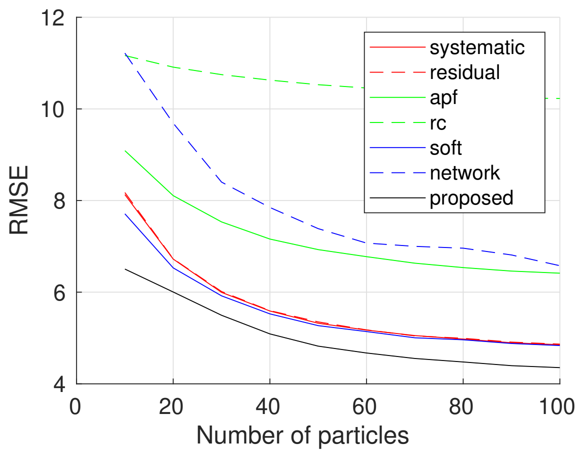

Fig. 2. The RMSE versus the number of particles at $\sigma^2 = 1$.

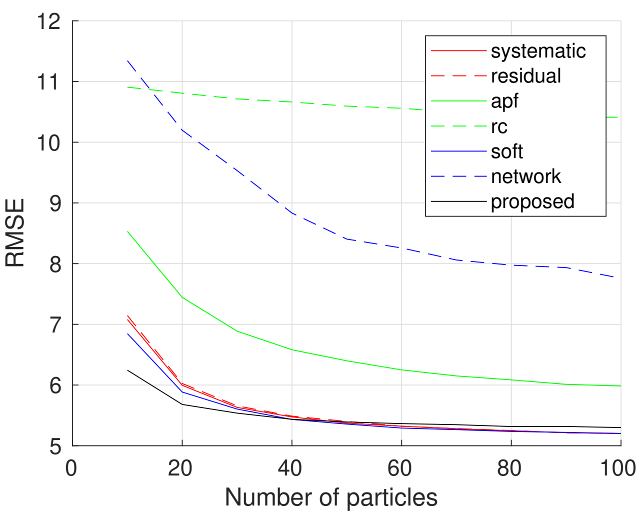

Fig. 3. The RMSE versus the number of particles at $\sigma^2 = 5$.

represents the posterior PDF and also tracks accurately with fewer particles, we now show its computational complexity for varying numbers of particles. This is shown in Figure 4. It can be observed that the proposed method is computationally highly scalable when compared to the systematic resampler. The network and soft resamplers consume a large computational time as they involve a nested sequential operation within particles. However, if the proposed method were to be implemented in parallel hardware architectures, it would consume the least time as it does not involve any sequential nature within the particle regeneration process.

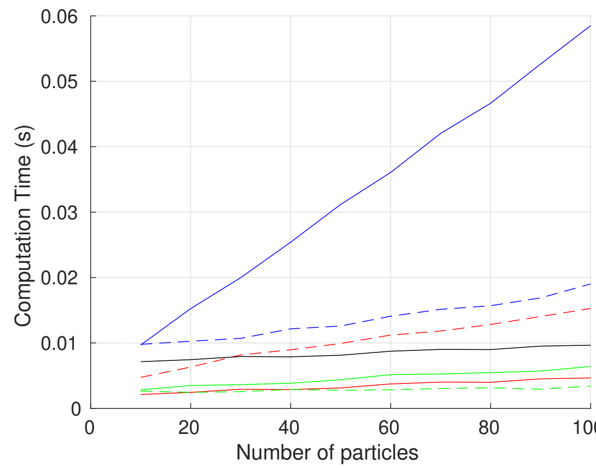

Fig. 4. The computational time (in seconds) versus the number of particles at $\sigma^2 = 5$. The legend of Figure 3 applies to this figure also.

V. Conclusion

This paper proposed a parallel-friendly PF approach for tracking latent dynamic target states by replacing the resampling step with sampling multiple disturbances. The main advantages of the proposed method, shown using simulations, is that it is able to faithfully represent the posterior PDF with fewer particles, and its execution is completely parallel-friendly. The parallel hardware implementation of the proposed method will be explored in the future.

References

[1] N. Gordon, D.J. Salmond, and A.F.M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," In IEE proceedings F (radar and signal processing), vol. 140, no. 2, pp. 107–113. 1993.

[2] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," IEEE Trans. Signal Proc., vol 50, no. 2, pp. 174–188, 2002.

[3] Deepthi Kattula, Rajesh K. Pullagura, Praveen B. Choppala, "Multiple sampling with reduced resampling for particle filtering," in Proc. IEEE Signal Processing Algorithms, Architectures, Arrangements, and Applications, Poland, 2024.

[4] Praveen B Choppala, "A Weight-Based Cutoff Resampling Method for Accelerated Particle Filtering," Fluctuation and Noise Letters, Vol. 22, No.5, 2023.

[5] Mohammed Ashik, Ramesh P Manapuram, Praveen B Choppala, "Observation leveraged resampling-free particle filter for tracking of rhythmic biomedical signals", International Journal of Intelligent Systems and Applications in Engineering, Vol. 11, pp. 616–624, 2023.

[6] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," J. of Computational and Graphical Statistics, pp. 1–25,

[7] P. B. Choppala, P. D. Teal, M. R. Frean, "Particle filter parallelisation using random network resampling," in Proc. IEEE Information Fusion, Spain, 2014.

[8] J. Liu, and R. Chen, "Sequential Monte Carlo methods for dynamic systems," J. American Statistical Association, Vol. 93, No. 443, pp. 1032–1044, 1998.

[9] R. Douc, and O Cappe, “Comparison of resampling schemes for particle filtering,” In Proc. IEEE Symp. on Image and Signal Processing and Analysis, pp. 64–69, 2005.

[10] J. D. Hol, T. B. Schon, and F. Gustafsson, “On resampling algorithms for particle filters,” Proc. 2006 IEEE Workshop on Nonlinear Statistical Signal Proc., pp. 79–82. 2006.

[11] P. B. Choppala, P. D. Teal, and M. R. Frean, “Soft systematic resampling for accurate posterior approximation and increased information retention in particle filtering,” In Proc. IEEE Workshop on Statistical Signal Processing, pp. 260-263, 2014.

[12] Tiancheng Li, Tariq P. Sattar, and Dedong Tang, “A fast resampling scheme for particle filters,” in Proc. IEEE Constantinides International Workshop on Signal Processing, London, 2013.

[13] T. Li, M. Bolic, and Petar M. Djuric, “Resampling methods for particle filtering: classification, implementation and strategies” in IEEE Signal processing magazine, Vol. 32, No. 3, pp. 70–86, 2015.

[14] L. M. Murray, Anthony Lee, and Pierre E. Jacob, “Parallel resampling in the particle filter, ” J. of Computational and Graphical Statistics, vol 25, no. 3, pp.789–805, 2016.

[15] A. Varsi, J Taylor, L Kekempanos, E. Knapp, and S. Maskell, “A Fast Parallel Particle Filter for Shared Memory Systems, ” IEEE. Signal Proc. Letters, Vol. 27, pp. 1570–1574, 2020.

[16] P. B. Choppala, P. D. Teal, and M. R. Frean, “Resampling and Network Theory,” IEEE Trans. Signal and Information Processing over Networks, vol 08, pp. 106–119, 2022.

[17] M. Pitt, and N. Shephard, “Filtering via simulation: Auxiliary particle filters,” J. American statistical Association, vol 94, no. 446, pp. 590–599, 1999.

[18] V. Elvira, L. Martino, M. Bugallo, and P. M. Djuric, “Elucidating the auxiliary particle filter via multiple importance sampling, ” IEEE Signal Processing Magazine, vol 36, no. 6, pp. 145–152, 2019.

IEEE INDIA COUNCIL

IEEE Kharagpur Section

21st IEEE INDICON 2024

Certificate of Presentation

This is to certify that

Kattula Deepthi

has presented the paper titled

Parallel friendly particle filtering approach with disturbance sampling

at the 21st IEEE INDICON 2024 held at IIT Kharagpur from 19-21 December 2024.

Dr. Dipankar Debnath, TPC Chair, IEEE INDICON 2024	Dr. Sudipta Mukhopadhyay, TPC Chair, IEEE INDICON 2024	Dr. Rajat Subhra Chakraborty TPC Chair, IEEE INDICON 2024