# ITU-ML-PS-002

# WALDO: Wireless Artificial Intelligence based Location Detection

**Team name:** The Sixth Sense

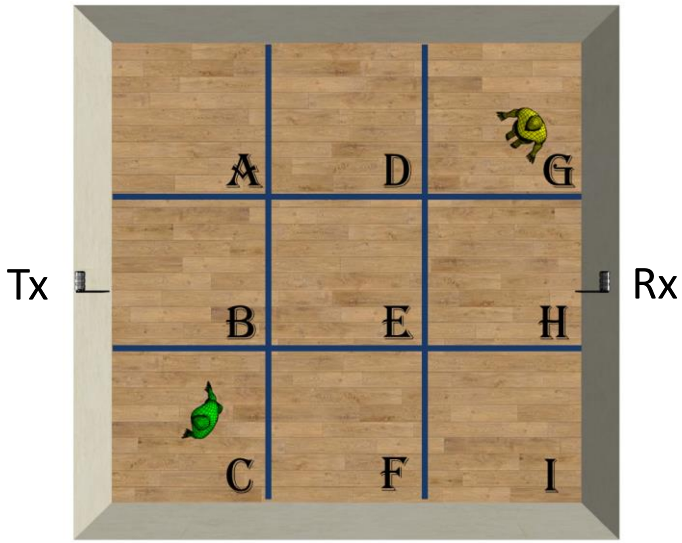
**Team Github Repo**: ITU-AI-ML-in-5G-Challenge/ITU-ML-5G-PS-002-WALDO_TheSixthSense_SRIB_Final

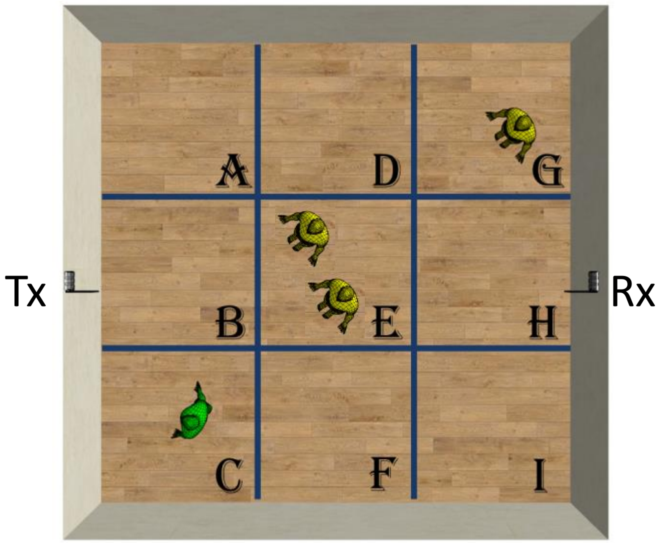

**Team members:**

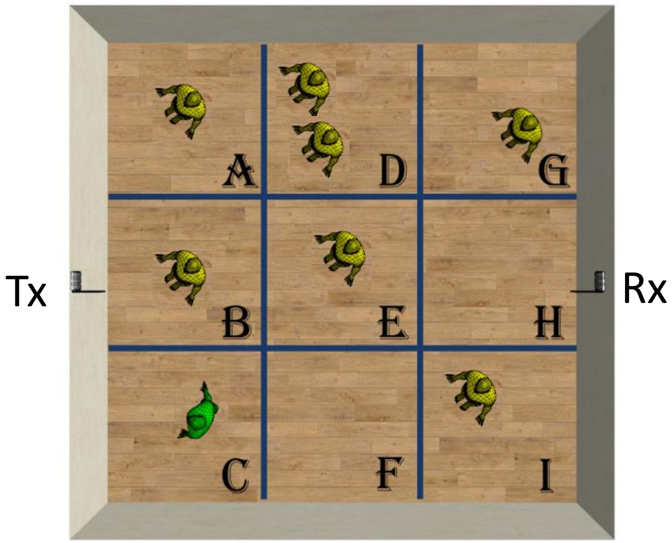
Shubham Khunteta
Avani Agrawal
Ashok Kumar Reddy Chavva
Ashok kumar Sahoo

*Affiliation : Beyond 5G Team, Samsung R&D Institute India, Bengaluru.*

# Problem statement




Given Received Signal : Channel estimation field of conventional IEEE 802.11ay packets

| Sector | No of people |
|---|---|
| C | 1 |
| G | 1 |
| Rest of the sectors | 0 |
| Total | 2 |

| Sector | No of people |
|---|---|
| C | 1 |
| E | 2 |
| G | 1 |
| Rest of the sectors | 0 |
| Total | 4 |

| Sector | No of people |
|---|---|
| A | 1 |
| B | 1 |
| C | 1 |
| D | 2 |
| E | 1 |
| G | 1 |
| I | 1 |
| Rest of the sectors | 0 |
| Total | 8 |

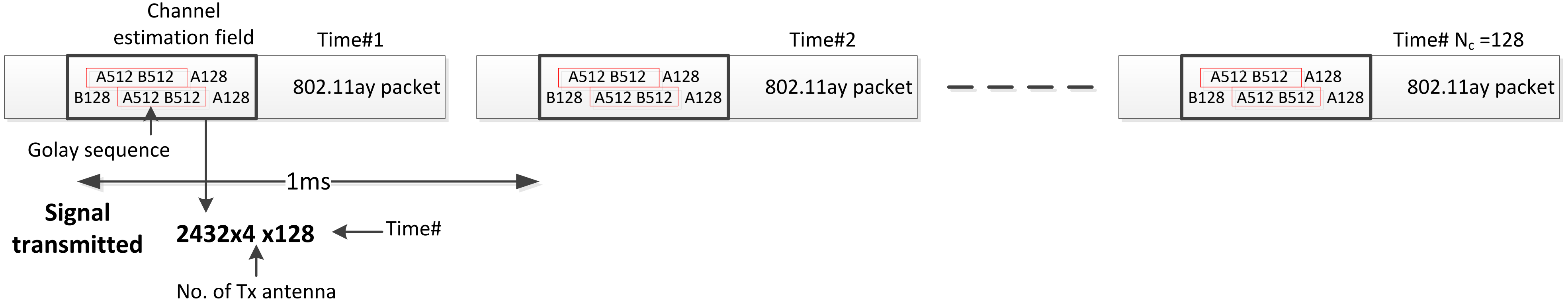
Problem: Find number of persons in each sector and the total number of persons in the room.

# Information provided

SAMSUNG

**Signal transmitted :**

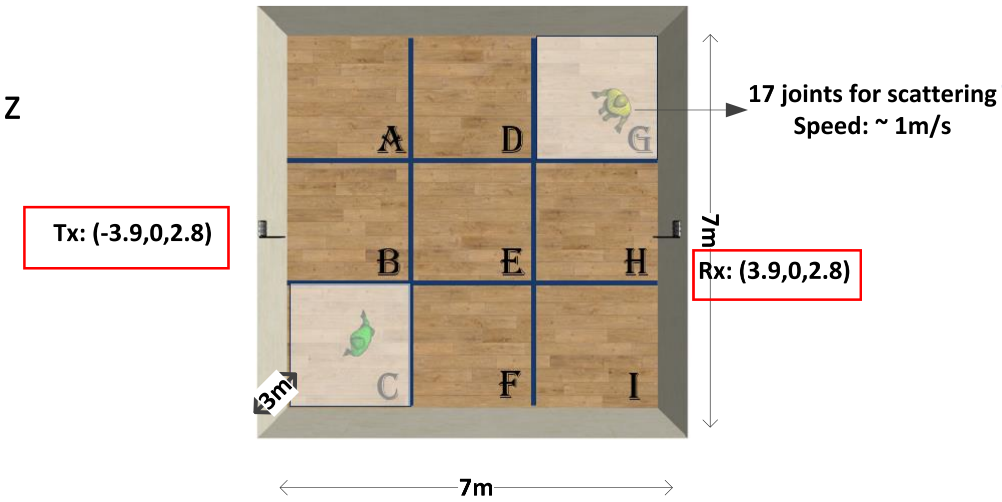
- Size : $N_s$ x $N_a$ x $N_c$

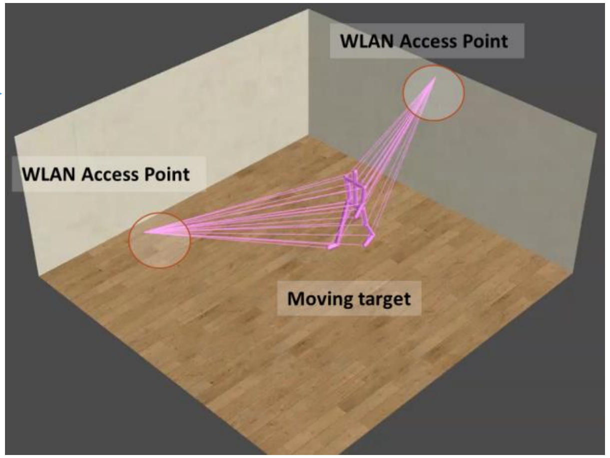



**Channel :**

Frequency: 60GHz

Bandwidth: 1.76GHz

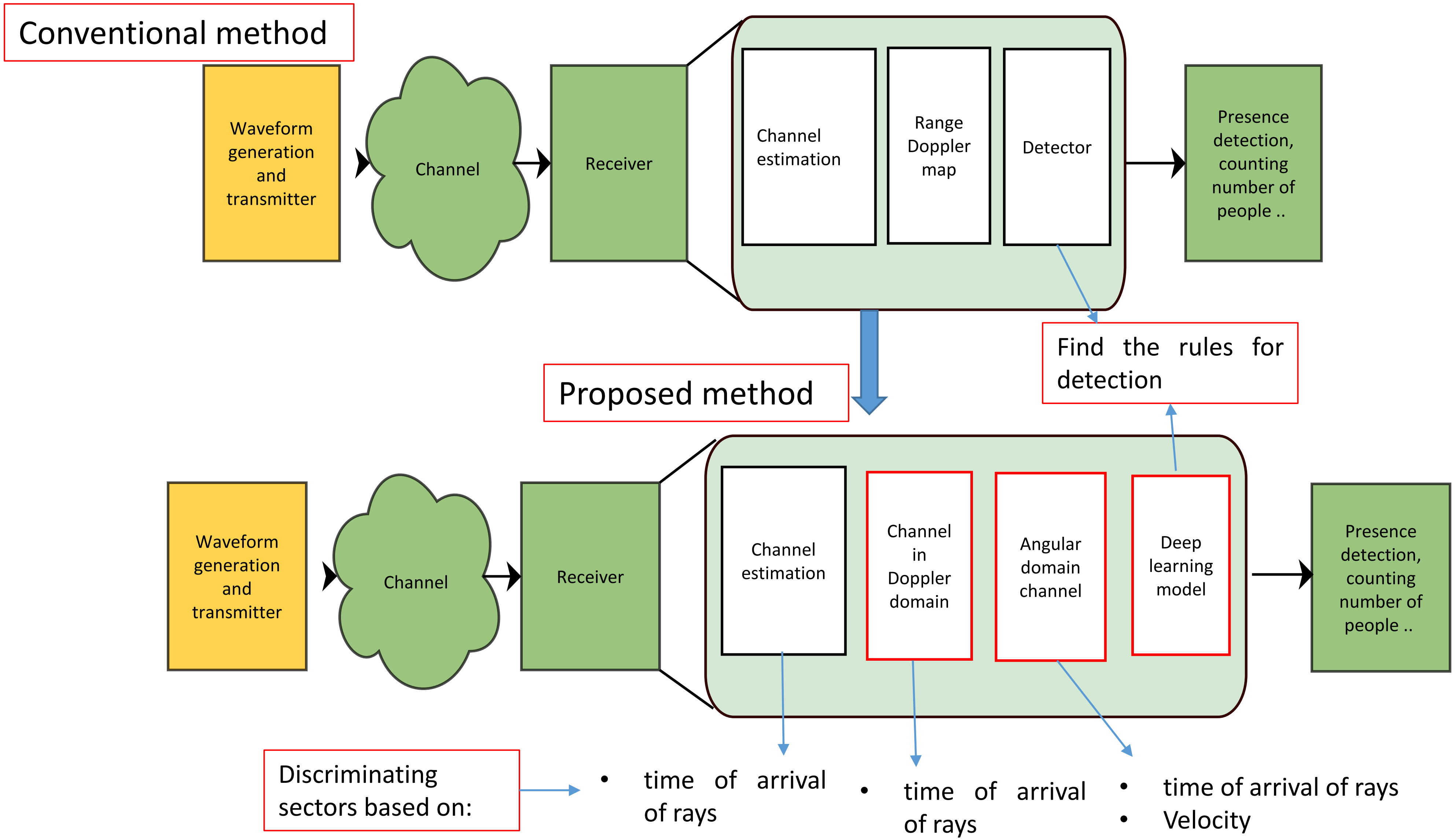



**Signal Received (Provided for the problem) :**

- Size : $(N_s + \mathrm{L} - 1)$ x $N_a$ x $N_c$ where channel has ‘L’ taps (here L=45) and $N_a$ is number of Rx antenna which is 4 here.
- Size : 2476x 4x 128

# Solution outline




- time of arrival of rays

- time of arrival of rays
- Velocity

- time of arrival of rays
- Velocity
- In Spatial domain

# Channel Estimation : L-Tap channel

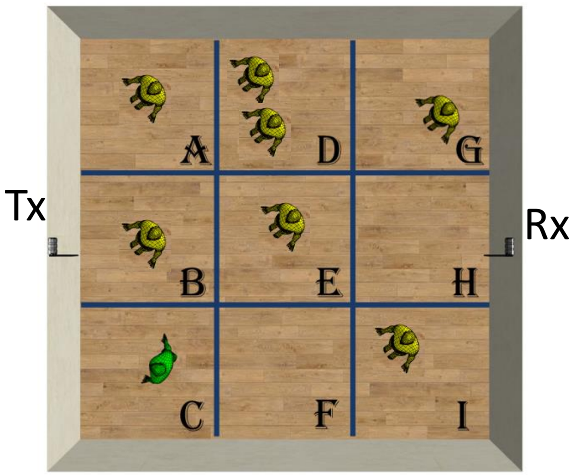


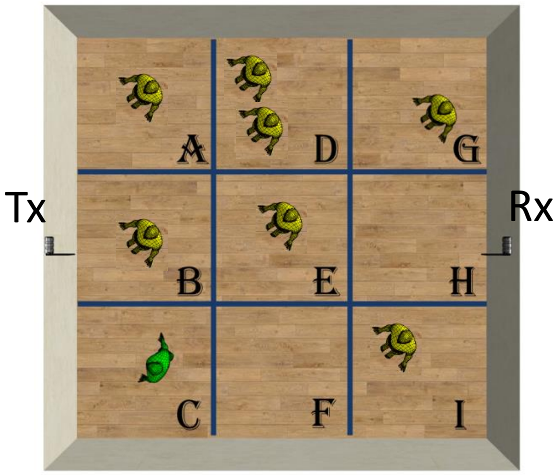
## Signal Model

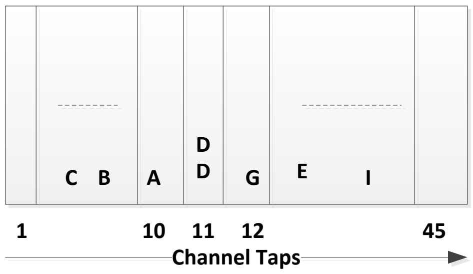
- For each repetition (Total $N_c$ ) of CEF, we considered an 'L' tap channel impulse response (CIR) with
  - tap index $l \in [0, L-1]$,
  - transmit and receive antenna index $\mathrm{n_t}, \mathrm{n_r} \in [0, N_a]$,
  - training symbol index $\mathrm{n} \in [0, N_s]$ and
  - signal index at the receiver $k \in [0, N_s + L - 1]$
- Let $\mathrm{H} = [H_0, .., H_{L-1}]$be the CIR matrix of the MIMO frequency selective channel, where
  - $\mathrm{H_l} = \begin{bmatrix} H_{1,1}(l) \ldots & H_{1,N_a}(l) \\ \vdots & \vdots \\ H_{N_a,1}(l) \ldots & H_{N_a,N_a}(l) \end{bmatrix}$
  - in which $H_{n_r,n_t}(l)$ … is the l-th tap of the CIR between the $n_r$-th receive antenna and the $n_t$-th transmit antenna.
- Y =HX+N where X is
  - $\mathrm{X} = \begin{bmatrix} x(0).. & x(N_s-1)\ 0_{N_aX1}.. & & 0_{N_aX1} \\ 0_{N_aX1} & & & \vdots \\ \vdots & \ddots & & 0_{N_aX1} \\ 0_{N_aX1}.. & 0_{N_aX1} & x(0).. & x(N_s-1) \end{bmatrix}$
  - $Y = [y(0) \ldots y(N_s + L - 1)]$
- Channel H can be estimated as LSE [1]:
  - $H = (YX^T) * (XX^T)^{-1}$

Size of channel estimated : 4x4x45x128

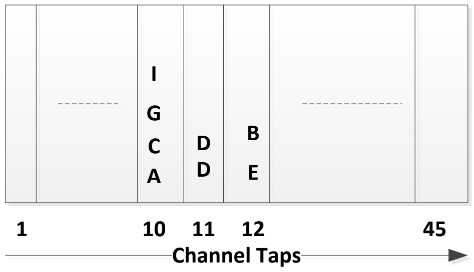
Rx, Tx antenna    Taps    Repetition in time




Easy to discriminate sectors



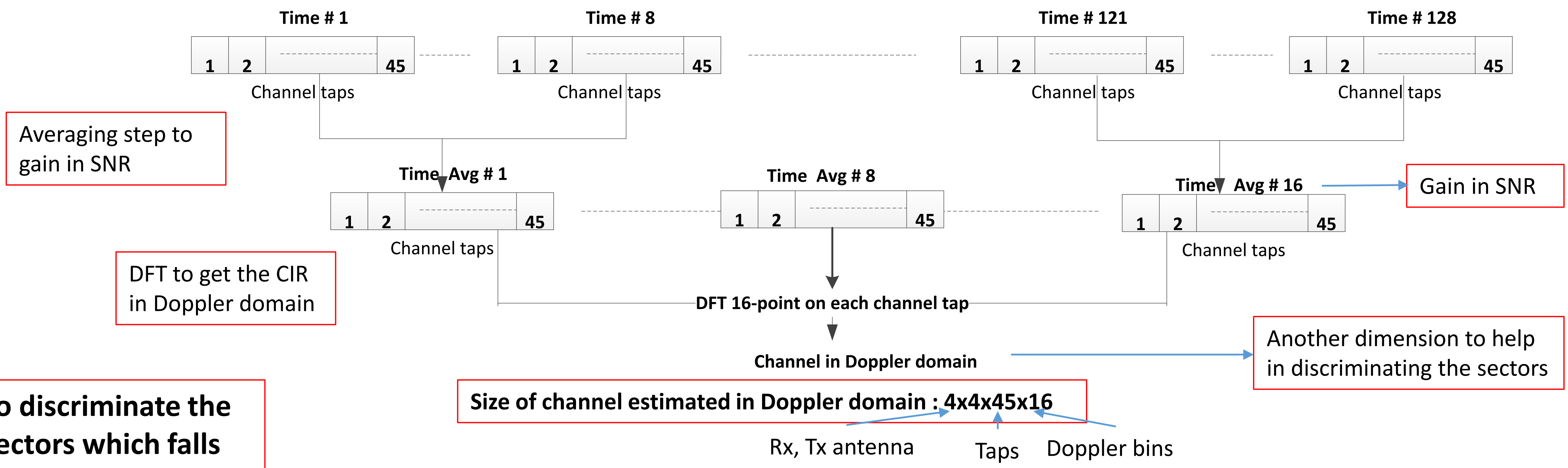

Not sufficient to discriminate the sectors

- With sampling rate =1.76 GHz, range resolution is 17cm

[1] S. Wang and A. Abdi, "Aperiodic Complementary Sets of Sequences-Based MIMO Frequency Selective Channel Estimation".

# Channel Estimation : Doppler domain

SAMSUNG




**To discriminate the sectors which falls under same tap of channel :**

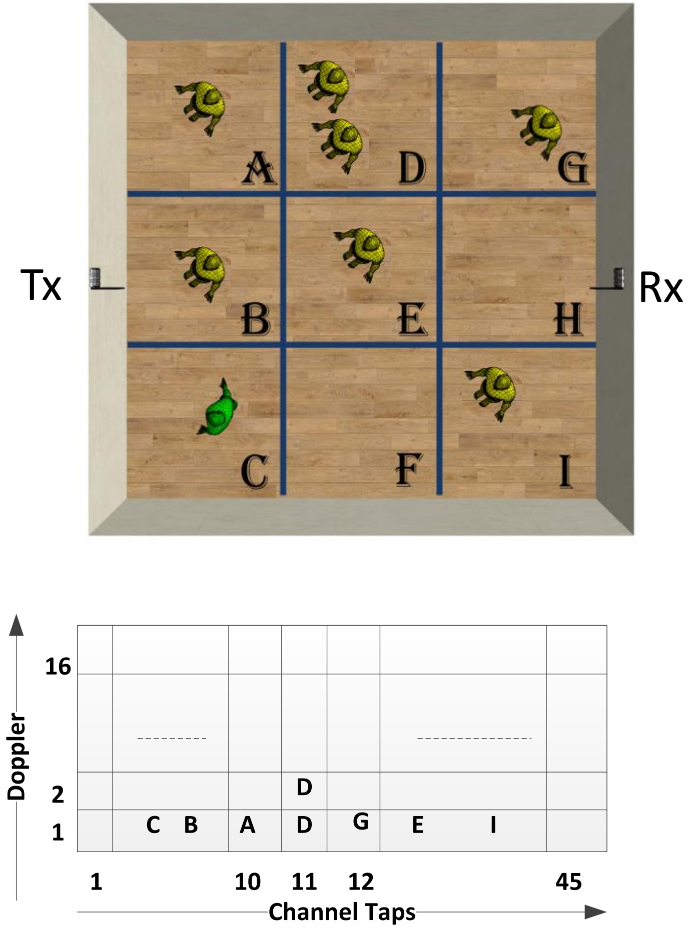
**Doppler domain processing is helpful**

**Objects which fall in same range bucket (same tap), can be differentiated with Doppler FFT which is as follows:**

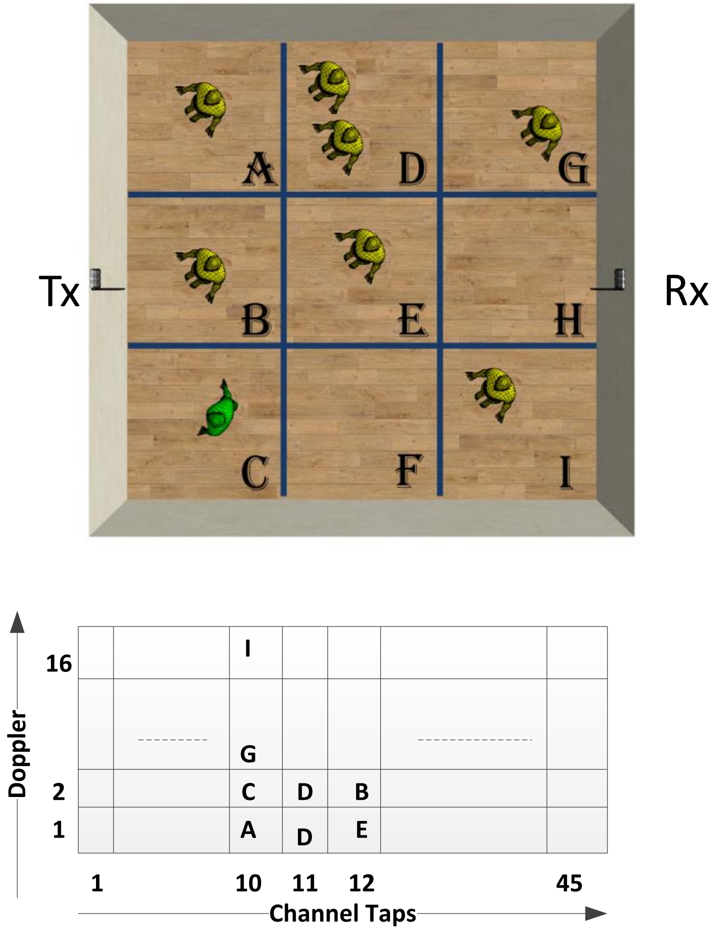
- First, H in time domain is averaged across multiple instances (Here over K=8 instance where $N_c$=128 ):
  - $H_{time} = [H_{T_1}, \dots, H_{T_K}, \dots, H_{T_{N_c}}]$
  - $H_{time_{AV}} = \left[mean(H_{T_1} \dots H_{T_K}), mean(H_{T_{K+1}} \dots H_{T_{2K}}), \dots, mean(H_{T_{K(\frac{N_c}{K}-1)}} \dots H_{T_{N_c}})\right]$
  - Size of $H_{time_{AV}}$ = $(N_a, N_a, L, \frac{N_c}{K})$
- FFT is performed on $H_{time_{AV}}$:
  - $H_{Dopp} = FFT(H_{time_{AV}})$ in time dimension

# Channel Estimation : Doppler domain

SAMSUNG

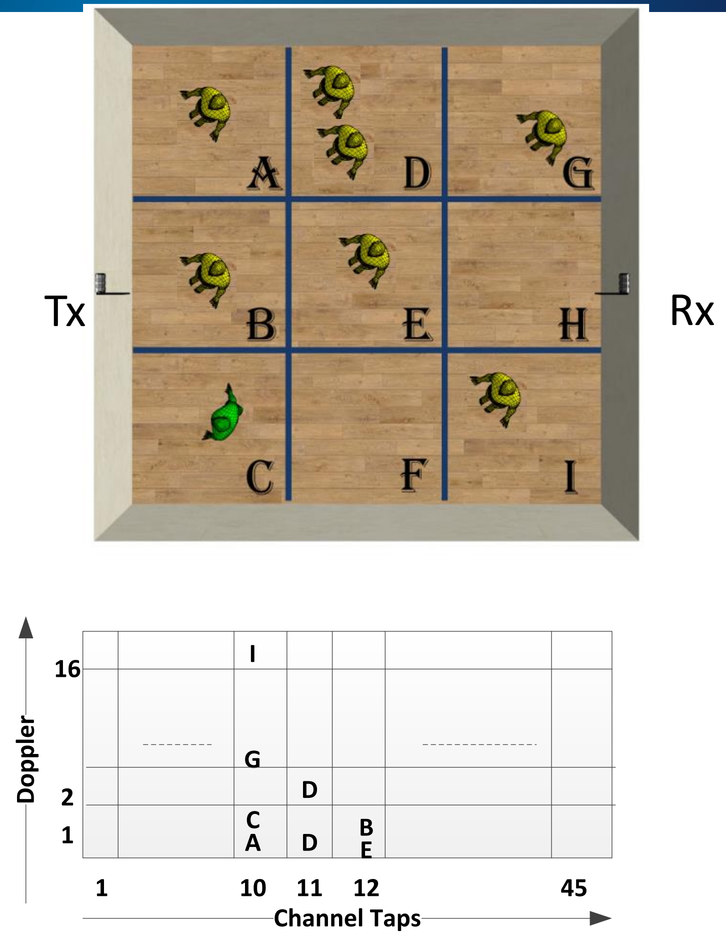



Easy to discriminate sectors




Easy to discriminate sectors

**Now, two object with same tap and different velocity can easily detected.**




Not sufficient to discriminate sectors

**What if, two object falls in the same tap and same Doppler bin … ?**

- $V_{res} = \frac{\lambda}{2T_f}$ where : $\lambda$=5mm, $T_f = \frac{N_c}{K} * 1ms$ = 16ms. $V_{res}$ =.15625 m/s

# Channel Estimation : Angular domain

SAMSUNG

- Size of channel estimated in Doppler domain : 4x4x45x16

?   Explored

- The resultant CIR ($H_D$) in Doppler domain is then converted to Angular domain using unitary matrices.
  - $H_{Ang} = U' * H_D * U$
  - where $U_{kl} = (1/sqrt(N_a))\exp(-j2\,\pi\,\frac{kl}{N_a})$
  - Here *k,l=1…4*

- The above transformation resolve the multiple rays in angular domain [2].

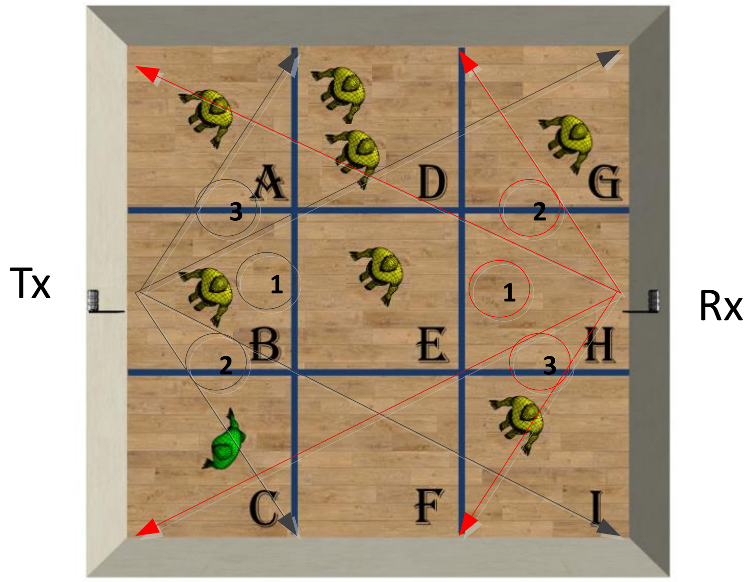
- Size of channel estimated in Angular domain : 4x4x45x16

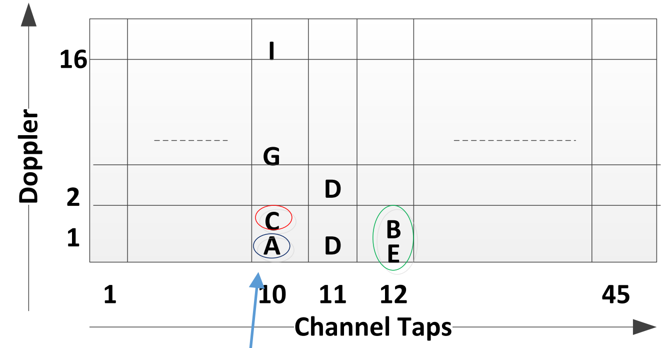
4x4x45x16 → 4x4x45x16

Antenna pairs : Tx and Rx

Resolvable paths in angular domain from Tx and Rx

Another dimension to help in discriminating the sectors

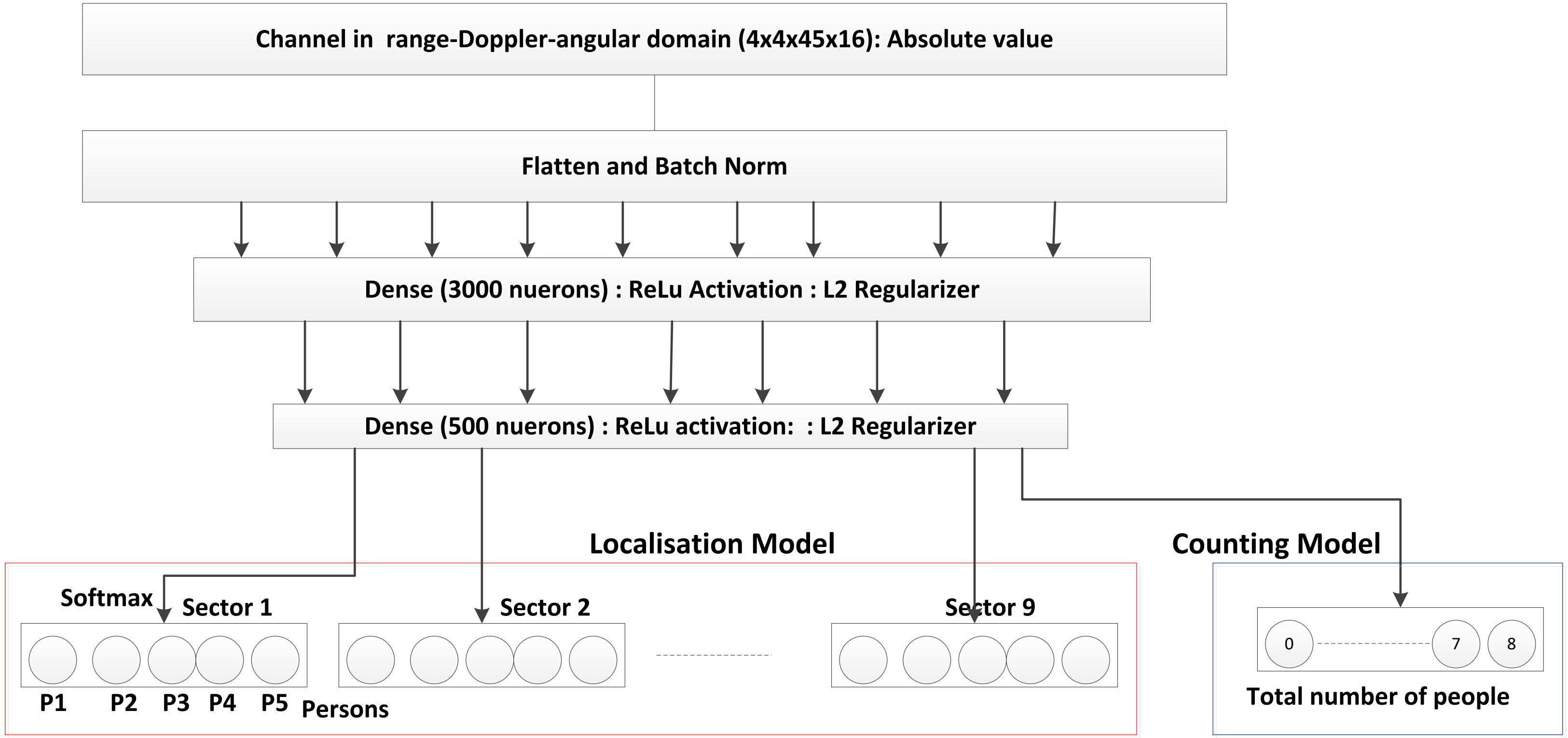



**Separable in angular domain : Same Tap and Doppler bin but different resolvable paths**

[2]. Tse and Vishwanath , “Fundamentals of Wireless Communication”, *Chapter: 7.3.4 Angular Domain Representation of MIMO Channels.*

# ML based detection model




Deep learning based detector

# Performance evaluation

SAMSUNG

**Dataset and Training:**

- Dataset provided for SNR 18, 0 and -18.
- Total of 15578 samples for each snr point and each sample is unique in terms of arrangement of people in the sectors.

**Result:**

**Counting model accuracy**: Number of times the total number of persons in the room are predicted correctly.

**Localisation model accuracy**: Number of times the number of persons in each sectors are predicted correctly.

| Counting Model accuracy | |
|---|---|
| SNR | % Number of samples in which Correctly counting total number of persons in room |
| 18 | 99% |
| 0 | 99% |
| -18 | 39% |

| Parameter List for ML model | |
|---|---|
| Input size | 4x4x45x16 (flatten) |
| Size of Hidden layer 1 | 3000 neurons with ReLu activation |
| Size of Hidden layer 2 | 500 nuerons with ReLu activation |
| Size of output layer | 45 (localization model) and 9 for counting model |
| Epochs | 200 |
| Learning | ADAM optimizer and learning rate 0.0005 with exponential decay |
| Regularizer | L2 |

| Localisation model accuracy | |
|---|---|
| SNR | % Number of samples in which Correctly identify number of persons in each sector |
| 18 | 99% |
| 0 | 85% |
| -18 | 2% |

Thank you.

Questions ?