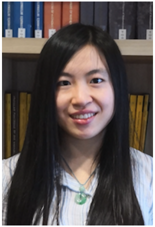
Fast Beam Alignment via Pure Exploration in Multi-Armed Bandits

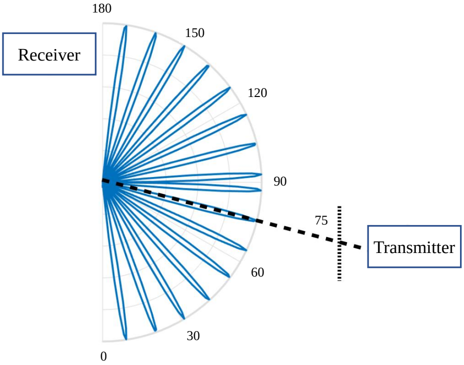
Yi Wei   Zixin Zhong   **Vincent Y. F. Tan**



National University of Singapore

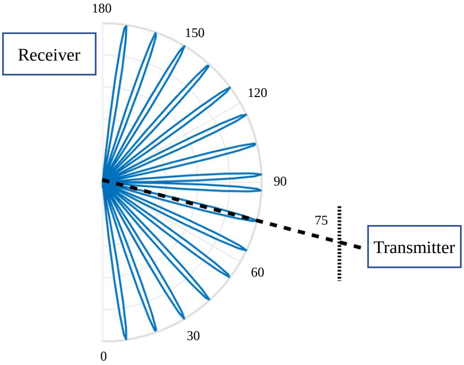
June 29, 2022

The Beam Alignment Problem



- mmWave Communications
- Beams at Tx and Rx are narrow directional

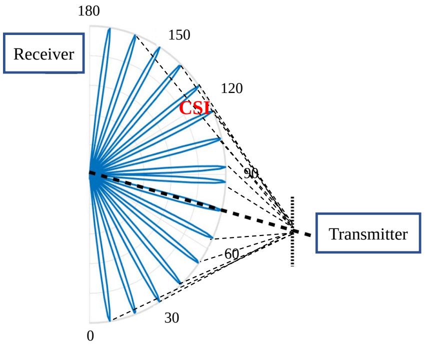
The Beam Alignment Problem



- mmWave Communications
- Beams at Tx and Rx are narrow directional

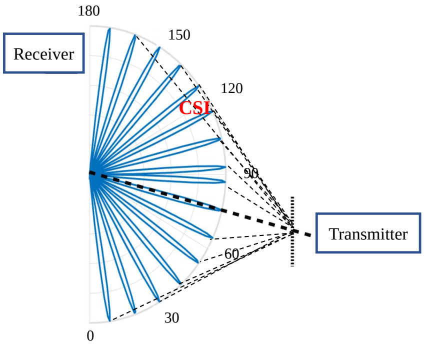
Beam alignment (BA) is to ensure the transmitter and receiver beams are accurately aligned to establish a reliable communication link.

Some Fundamental Challenges in Beam Alignment



- To find optimal Rx/Tx pair, CSI corresponding to each pair is measured
- Frequency of each measurement is high
- High beam alignment latency

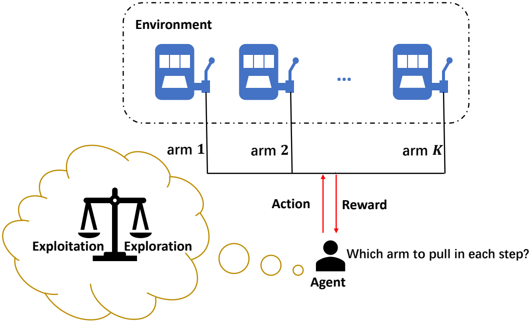
Some Fundamental Challenges in Beam Alignment



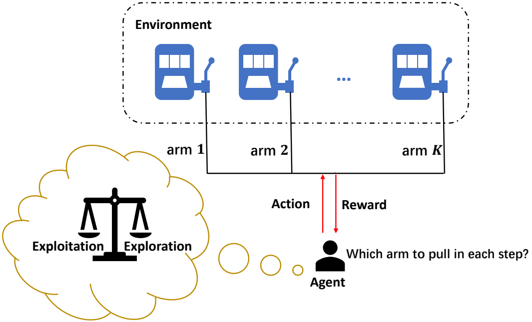
- To find optimal Rx/Tx pair, CSI corresponding to each pair is measured
- Frequency of each measurement is high
- High beam alignment latency

Beam alignment latency will increase with the number of antennas at the receivers and transmitters.

Multi-Armed Bandits

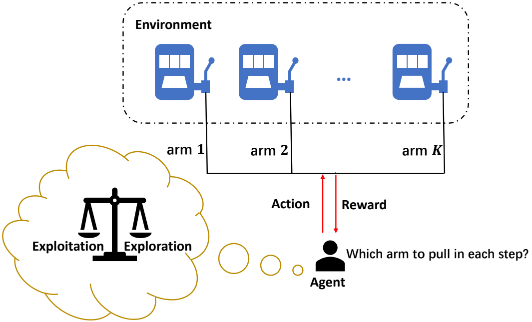


Multi-Armed Bandits



Pure Exploration: Identify the best arm (arm with largest mean) as quickly as possible.

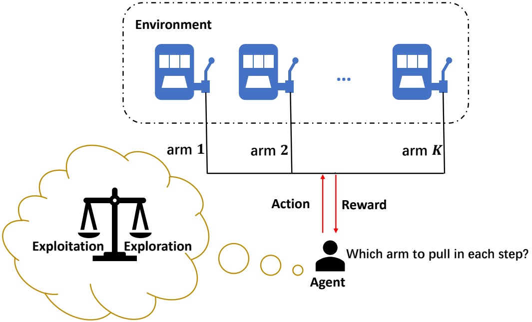
Multi-Armed Bandits



Pure Exploration: Identify the best arm (arm with largest mean) as quickly as possible.

Main Contribution: Formulate and solve the beam alignment problem using ideas in pure exploration in the fixed-confidence setting.

Multi-Armed Bandits



Pure Exploration: Identify the best arm (arm with largest mean) as quickly as possible.

Main Contribution: Formulate and solve the beam alignment problem using ideas in pure exploration in the fixed-confidence setting.

Exploit structure in the beam alignment problem.

System Model

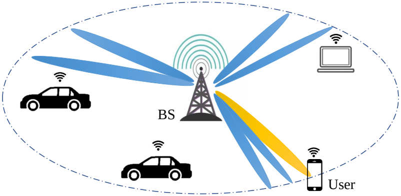


Figure: A mmWave massive MISO system system.

- **Massive mmWave MISO system:** A base station (BS) equipped with N transmit antennas serves a single-antenna user

System Model

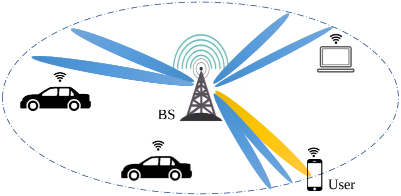


Figure: A mmWave massive MISO system system.

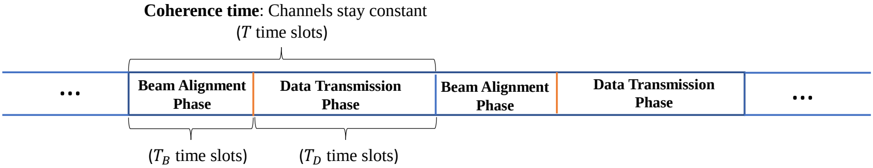
- Massive mmWave MISO system: A base station (BS) equipped with N transmit antennas serves a single-antenna user
- Adopt the widely-used Saleh–Valenzuela channel model

$$\mathbf{h} = \beta^{(1)}\mathbf{a}\left(\theta^{(1)}\right) + \sum_{l=2}^{L} \beta^{(l)}\mathbf{a}\left(\theta^{(l)}\right)$$

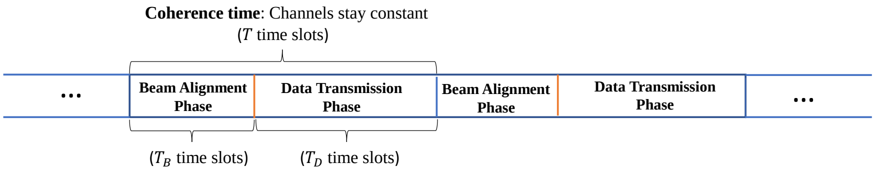
Amplitude

1 line-of-sight (LoS) path $\geq$ L − 1 non-LoS (NLoS) paths

Transmission Scheme



Transmission Scheme



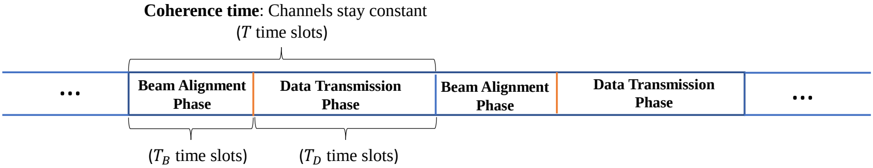
- Beam alignment phase: Fast beam alignment algorithm searches the optimal beam from a given codebook

$$\mathcal{C} \triangleq \{\mathbf{f}_k = \mathbf{a}(-1 + 2k/K) \mid k = 0, 1, \ldots, K-1\}$$

where the array response vector:

$$\mathbf{a}(x) = \frac{1}{\sqrt{N}}\left[1, e^{j\frac{2\pi}{\lambda}dx}, e^{j\frac{2\pi}{\lambda}2dx}, \cdots, e^{j\frac{2\pi}{\lambda}(N-1)dx}\right]^H \in \mathbb{C}^{N\times 1}$$

Transmission Scheme



- Beam alignment phase: Fast beam alignment algorithm searches the optimal beam from a given codebook

$$\mathcal{C} \triangleq \{\mathbf{f}_k = \mathbf{a}(-1 + 2k/K) \mid k = 0, 1, \ldots, K - 1\}$$

where the array response vector:

$$\mathbf{a}(x) = \frac{1}{\sqrt{N}} \left[1, e^{j\frac{2\pi}{\lambda}dx}, e^{j\frac{2\pi}{\lambda}2dx}, \ldots, e^{j\frac{2\pi}{\lambda}(N-1)dx}\right]^{\mathrm{H}} \in \mathbb{C}^{N\times 1}$$

- Data transmission phase: BS transmits effective data using the selected $\mathbf{f}$. Received signal at the user in time slot t is

$$y_t = \sqrt{p}\, \mathbf{h}^{\mathrm{H}}\mathbf{f}\, s_t + n_t,$$

Beam Alignment Phase

- **System Throughput Performance:** Effective achievable rate

$$R_{\text{eff}} \triangleq \left(1 - \frac{T_{\text{B}}}{T_{\text{D}}}\right) \log\left(1 + \frac{p|\mathbf{h}^{\text{H}}\mathbf{f}|^2}{\sigma^2}\right)$$

Time to search for optimal beam T_{B} to be minimized for high R_{eff}

Beam Alignment Phase

- **System Throughput Performance:** Effective achievable rate

$$R_{\text{eff}} \triangleq \left(1 - \frac{T_{\mathrm{B}}}{T_{\mathrm{D}}}\right) \log\left(1 + \frac{p|\mathbf{h}^{\mathrm{H}}\mathbf{f}\ |^2}{\sigma^2}\right)$$

Time to search for optimal beam T_{B} to be minimized for high R_{eff}

- **Measurement:** The received signal power

$$R(\mathbf{f}_k) = |\sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{f}_k + n|^2 = p|\mathbf{h}^{\mathrm{H}}\mathbf{f}_k|^2 + 2\sqrt{p}\Re(\mathbf{h}^{\mathrm{H}}\mathbf{f}_k n^*) + |n|^2$$

Heteroscedastic Gaussian Variable $\mathcal{N}(p|\mathbf{h}^{\mathrm{H}}\mathbf{f}_k|^2, 2p|\mathbf{h}^{\mathrm{H}}\mathbf{f}_k|^2\sigma^2)$

Gamma Variable $\Gamma(1, 1/\sigma^2)$

Approximate (Because: noise power $\ll$ transmit power)

$$r_k = p|\mathbf{h}^{\mathrm{H}}\mathbf{f}_k|^2 + 2\sqrt{p}\Re(\mathbf{h}^{\mathrm{H}}\mathbf{f}_k n^*)$$

Relation to MABs and Properties

Beam Alignment	→	Pure Exploration in MAB
Find the optimal beam as soon as possible	→	Find the optimal beam as soon as possible
beams	→	base arms
received signal power	→	rewards

Relation to MABs and Properties

Beam Alignment		Pure Exploration in MAB
Find the optimal beam as soon as possible	→	Find the optimal beam as soon as possible
beams	→	base arms
received signal power	→	rewards

Property: Let $\mu = (\mu_1, \cdots, \mu_K)$, and let $\mu_{(1)} \geq \mu_{(2)} \geq \mu_{(3)} \geq \ldots \geq \mu_{(K)}$ be the sorted sequence of means.

Relation to MABs and Properties

Beam Alignment	→	Pure Exploration in MAB
Find the optimal beam as soon as possible	→	Find the optimal beam as soon as possible
beams	→	base arms
received signal power	→	rewards

Property: Let $\mu = (\mu_1, \cdots, \mu_K)$, and let $\mu_{(1)} \geq \mu_{(2)} \geq \mu_{(3)} \geq \ldots \geq \mu_{(K)}$ be the sorted sequence of means.

1. The means of the rewards associated with close-by arms are close.
2. The variance of the reward of an arm is linearly related to its mean

$$\sigma_k^2 = 2\mu_k\sigma^2.$$

Beam Codebooks Possess the Group Property

$\frac{1}{J}$-lower resolution beam codebook

- Constructed by grouping the nearby beams in the codebook $\mathcal{C}$

$$\mathcal{C}_{(J)} , \left\{ \mathbf{b}_g = \sum_{k=J(g-1)+1}^{Jg} \mathbf{f}_k : g = 0, 1, \ldots, G-1 \right\}$$

Beam Codebooks Possess the Group Property

$\frac{1}{J}$-lower resolution beam codebook

- Constructed by grouping the nearby beams in the codebook $\mathcal{C}$

$$\mathcal{C}_{(J)} \,, \left\{ \mathbf{b}_g = \sum_{k=J(g-1)+1}^{Jg} \mathbf{f}_k \,:\, g = 0, 1, \ldots, G-1 \right\}$$

- Received power for beam $\mathbf{b}_g$

$$R_g = p|\mathbf{h}^{\mathrm{H}}\mathbf{b}_g|^2 + 2\sqrt{p}\,(\mathbf{h}^{\mathrm{H}}\mathbf{b}_g n),$$

follows a heteroscedastic Gaussian distribution.

Beam Codebooks Possess the Group Property

$\frac{1}{J}$-lower resolution beam codebook

- Constructed by grouping the nearby beams in the codebook $\mathcal{C}$

$$\mathcal{C}_{(J)} \triangleq \left\{ \mathbf{b}_g = \sum_{k=J(g-1)+1}^{Jg} \mathbf{f}_k \; : \; g = 0, 1, \ldots, G-1 \right\}$$

- Received power for beam $\mathbf{b}_g$

$$R_g = p|\mathbf{h}^{\mathrm{H}}\mathbf{b}_g|^2 + 2\sqrt{p}\,(\mathbf{h}^{\mathrm{H}}\mathbf{b}_g n),$$

  follows a heteroscedastic Gaussian distribution.

- Information of a **set** of beams can be obtained at each time step

Problem Setup for Bandit Beam Alignment

Bandit BA Problem

- K base arms $[K]\ ,\ \{1,\cdots,K\}$: associated with the beam $\mathbf{f}_k$
- $\{[K],J\}$: set of all non-empty consecutive tuples of length $\leq J$

  $\{[6],2\} = \{\{1\},\{1,2\},\{2\},\{2,3\},\{3\},\{3,4\},\{4\},\{4,5\},\{5\},\{5,6\},\{6\}\}$
- (K,J)-super arm: each tuple in $\{[K],J\}$, associated with a "grouped beam" $\mathbf{b}_g \quad \mathcal{C}_{(J)}$.

Bandit BA Problem Setup

In time step t

- Choose an action (or a (K, J)-super arm) $A(t)$ $\{[K], J\}$

Bandit BA Problem Setup

In time step t

- Choose an action (or a (K, J)-super arm) $A(t)$ $\{[K], J\}$
- Observe the reward

$$R(t) = F\left(\sum_{k \;\; A(t)} \mathbf{f}_k, p, \mathbf{h}, n_t\right)$$

where $F(\mathbf{f}, p, \mathbf{h}, n) = p|\mathbf{h}^{\mathrm{H}}\mathbf{f}|^2 + 2\sqrt{p}\;\;(\mathbf{h}^{\mathrm{H}}\mathbf{f}n\;\;)$

Bandit BA Problem Setup

In time step t

- Choose an action (or a (K, J)-super arm) $A(t) \in \binom{[K]}{J}$
- Observe the reward

$$R(t) = F\left(\sum_{k \in A(t)} \mathbf{f}_k, p, \mathbf{h}, n_t\right)$$

where $F(\mathbf{f}, p, \mathbf{h}, n) = p|\mathbf{h}^{\mathrm{H}}\mathbf{f}|^2 + 2\sqrt{p}\,\Re(\mathbf{h}^{\mathrm{H}}\mathbf{f}n^*)$

- Each observed reward $R(t)$ also follows a heteroscedastic Gaussian distribution.

Bandit BA Problem Setup

Algorithm: $\pi := \{(\pi_t)_t, \tau^{\pi}, \psi^{\pi}, J\}$

- Sampling rule π_t: Determines the $([K], J)$-super arm $A(t)$ to pull at time step t based on the observation history and the arm history $\{A(1), R(1), A(2), R(2), \cdots, A(t-1), R(t-1)\}$
- Stopping rule: Leads to a stopping time τ^{π} satisfying $\mathrm{P}(\tau^{\pi} < \infty) = 1$
- Recommendation rule ψ^{π}: Outputs a base arm k^{π} $\quad [K]$.

Bandit BA Problem Setup

Algorithm: $\pi := \{(\pi_t)_t, \tau^\pi, \psi^\pi, J\}$

- Sampling rule π_t: Determines the $([K], J)$-super arm $A(t)$ to pull at time step t based on the observation history and the arm history $\{A(1), R(1), A(2), R(2), \cdots, A(t-1), R(t-1)\}$
- Stopping rule: Leads to a stopping time τ^π satisfying $\mathrm{P}(\tau^\pi < \infty) = 1$
- Recommendation rule ψ^π: Outputs a base arm $k^\pi \quad [K]$.

Aim: Use as few samples as possible to output a guess of the optimal arm with probability at least $1 - \delta$.

Some Notations for the General Lower Bound

- Heteroscedastic Gaussian bandit instance:

$$\nu = \left(\mathcal{N}(\mu_1^\nu, 2\mu_1^\nu\sigma^2), \ldots, \mathcal{N}(\mu_K^\nu, 2\mu_K^\nu\sigma^2) \right)$$

- Optimal arm $A^*(\nu)$:

$$\mu_{A^*(\nu)}^\nu = \operatorname*{argmax}_{k \in [K]} \mu_k^\nu$$

- Set of probability distributions

$$W_K := \left\{ \mathbf{w} \in \mathbb{R}_+^K : \sum_{k=1}^K w_k = 1 \right\}$$

- Alternative Set

$$\mathrm{Alt}(\nu) := \{\mathrm{u} \in V : A^*(\mathrm{u}) \neq A^*(\nu)\}$$

General Lower Bound

Theorem

For any $(\delta, J)-$PAC algorithm where $\delta \in (0,1)$, it holds that

$$\mathbb{E}_{\pi}[\tau] \geq c(\nu) \ln\left(\frac{1}{4\delta}\right),$$

where

$$c(\nu)^{-1} = \sup_{\mathbf{w} \in W_K} \inf_{u \in \mathrm{Alt}(\nu)} \sum_{k=1}^{K} w_k D_{\mathrm{HG}}(\mu_k^{\nu}, \mu_k^{u})$$

and the heteroscedastic relative entropy is

$$D_{\mathrm{HG}}(\mu_i, \mu_j) = \frac{1}{2} \ln\left(\frac{\mu_j}{\mu_i}\right) + \frac{\mu_i}{2\mu_j} + \frac{(\mu_j - \mu_i)^2}{4\mu_j \sigma^2} - \frac{1}{2}.$$

Two-Phase Heteroscedastic Track & Stop (2PHT&S)

Main Idea

to exploit the **prior knowledge** which have not been considered by existing bandit-based beam alignment algorithms:

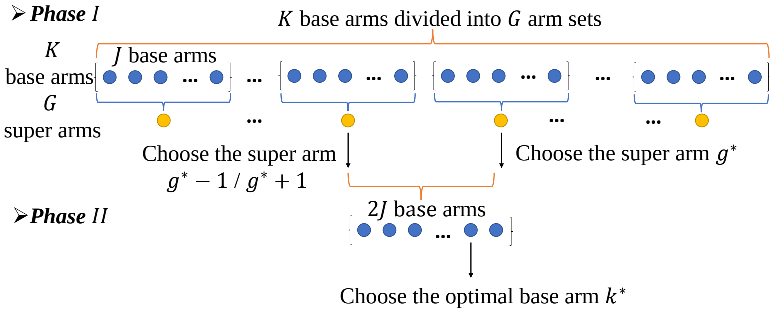
- Close-by correlation
- Heteroscedasticity
- Group property

Two-Phase Heteroscedastic Track & Stop (2PHT&S)

Main Idea

to exploit the **prior knowledge** which have not been considered by existing bandit-based beam alignment algorithms:

- Close-by correlation
- Heteroscedasticity
- Group property



2PHT&S Algorithm

Phase I: Search for the optimal super arm w.p. $1 \quad {}_1$

2PHT&S Algorithm

Phase I: Search for the optimal super arm w.p. 1   1

Group K base arms into G arm sets to reduce the search space

2PHT&S Algorithm

Phase I: Search for the optimal super arm w.p. $1 \quad {}_1$

- Group K base arms into G arm sets to reduce the search space
- At each time,
  - Choose one super arm (beam group) by the sampling rule in a new Heteroscedastic Track and Stop (HT&S) algorithm

2PHT&S Algorithm

Phase I: Search for the optimal super arm w.p. $1 \quad {}_1$

- Group K base arms into G arm sets to reduce the search space
- At each time,
  - Choose one super arm (beam group) by the sampling rule in a new Heteroscedastic Track and Stop (HT&S) algorithm
  - Use the grouped beam to transmit the pilot symbols and observe the grouped reward $R_g(t)$.

2PHT&S Algorithm

Phase I: Search for the optimal super arm w.p. $1 \quad _1$

- Group K base arms into G arm sets to reduce the search space
- At each time,
  - Choose one super arm (beam group) by the sampling rule in a new Heteroscedastic Track and Stop (HT&S) algorithm
  - Use the grouped beam to transmit the pilot symbols and observe the grouped reward $R_g(t)$.
- Select the optimal super arm

$$g \quad = \operatorname*{argmax}_{g 2 [G]} E[R_g(t)]:$$

2PHT&S Algorithm

Phase II: Search for the optimal base arm w.p. $1 \quad {}_{2}$

2PHT&S Algorithm

Phase II: Search for the optimal base arm w.p. 1 $_2$

Construct a base arm set (including the optimal super arm and its neighboring super arm)

2PHT&S Algorithm

Phase II: Search for the optimal base arm w.p. 1 $_2$

- Construct a base arm set (including the optimal super arm and its neighboring super arm)
- Search for the optimal base arm in the base arm set using the HT&S algorithm

HT&S: An Improved Track & Stop Algorithm

Sampling Rule: Estimate the number of times each arm should be sampled

$$Q(t) = \begin{cases} \arg\min_i T_i(t-1); & \text{if } \min_i T_i(t-1) \le \sqrt{t}; \\ \arg\max_i t\hat{w}_i(t-1) - T_i(t-1); & \text{otherwise.} \end{cases}$$

HT&S: An Improved Track & Stop Algorithm

Sampling Rule: Estimate the number of times each arm should be sampled

$$Q(t) = \begin{cases} \operatorname{argmin}_i T_i(t \quad 1); & \text{if } \min_i T_i(t \quad 1) \quad \sqrt{t}; \\ \operatorname{argmax}_i t\hat{w}_i(t \quad 1) \quad T_i(t \quad 1); & \text{otherwise:} \end{cases}$$

Stopping Rule: Stop when the number of pulls of all arms meet a certain requirement

Threshold to be $(t; \; ; \quad) = \ln(\quad t = \quad)$, and the stopping rule is

$$= \min \quad t \ 2 \ N : Z(t) \qquad (t; \; ; \quad) \ :$$

HT&S: An Improved Track & Stop Algorithm

- **Sampling Rule:** Estimate the number of times each arm should be sampled

$$Q(t) = \begin{cases} \operatorname{argmin}_i T_i(t-1), & \text{if } \min_i T_i(t-1) \le \sqrt{t}, \\ \operatorname{argmax}_i t\hat{w}_i(t-1) - T_i(t-1), & \text{otherwise.} \end{cases}$$

- **Stopping Rule:** Stop when the number of pulls of all arms meet a certain requirement

Threshold to be $\beta(t, \delta, \alpha) = \ln(\alpha t/\delta)$, and the stopping rule is

$$\tau_\delta = \min\left\{ t \in \mathbb{N} : Z(t) \ge \beta(t, \delta, \alpha) \right\}.$$

- **Heteroscedasticity:** Considered explicitly in the calculations of estimated reward $\hat{w}_i(t-1)$ and statistic $Z(t)$

Time Complexity Analysis of 2PHT&S

Theorem

Let the means of the super and base arms be respectively

$$\mathsf{s} := \left(N(\mu_1^{\mathsf{s}}, 2\mu_1^{\mathsf{s}}\sigma^2), \ldots, N(\mu_G^{\mathsf{s}}, 2\mu_G^{\mathsf{s}}\sigma^2) \right) \quad \text{and}$$

$$\mathsf{b} := \left(N(\mu_{S_f(1)}^{\mathsf{b}}, 2\mu_{S_f(1)}^{\mathsf{b}}\sigma^2), \ldots, N(\mu_{S_f(2J)}^{\mathsf{b}}, 2\mu_{S_f(2J)}^{\mathsf{b}}\sigma^2) \right)$$

where

$$\mu_g^{\mathsf{s}} = p \left\| \mathbf{h}^H \left(\sum_{k \in S_g} \mathbf{f}_k \right) \right\|^2 .$$

Time Complexity Analysis of 2PHT&S

Theorem

Let the means of the super and base arms be respectively

$$\mathsf{s} := \left\{ N(\mu_1^{\mathsf{s}}, 2\mu_1^{\mathsf{s}}\sigma^2), \ldots, N(\mu_G^{\mathsf{s}}, 2\mu_G^{\mathsf{s}}\sigma^2) \right\} \qquad \text{and}$$
$$\mathsf{b} := \left\{ N(\mu_{S_f(1)}^{\mathsf{b}}, 2\mu_{S_f(1)}^{\mathsf{b}}\sigma^2), \ldots, N(\mu_{S_f(2J)}^{\mathsf{b}}, 2\mu_{S_f(2J)}^{\mathsf{b}}\sigma^2) \right\}$$

where

$$\mu_g^{\mathsf{s}} = p \left\| \mathbf{h}^H \left(\sum_{k \in S_g} \mathbf{f}_k \right) \right\|^2 .$$

Under 2PHT&S, we have

$$\limsup_{\delta \to 0} \frac{\mathsf{E}[\tau^{2PHT\&S}]}{\ln(1/\delta)} \le C_{\mathsf{s}} + C_{\mathsf{b}},$$

where C_{s} and C_{b} represent time complexities of Phases I and II resp.

Simulation Results

Experiment Setup

- Massive mmWave MISO system
- Base station equipped with 64 transmit antennas serving a single-antenna user
- Size of codebook is set as $K = 128$.

Simulation Results

Experiment Setup

- Massive mmWave MISO system
- Base station equipped with 64 transmit antennas serving a single-antenna user
- Size of codebook is set as $K = 128$.

Baseline Algorithms

- Original Track-and-Stop (T&S) algorithm (Garivier et al. 2016)
- Two-Phase Track-and-Stop (2PT&S) algorithm
- Heteroscedastic Track-and-Stop (HT&S) algorithm

A. Garivier and E. Kaufmann, "Optimal best arm identi cation with xed con dence," in PMLR, 2016, pp. 998-1027.

Numerical Simulations on Synthetic Data

Figure: Mean of the reward of each base arm and super arm (p = 10 dBm).

Table: Average time complexities over 100 experiments when = 0:1

Power	4	6	8	10	12
T&S	1154.3 338:7	654.6 212:1	382.5 129:6	209.4 68:6	133.7 8:9
HT&S	473.2 275:5	271.4 143:4	175.6 69:2	133.2 24:1	123.9 6:5
2PT&S	206.2 60:4	120.2 35:0	68.4 19:4	49.1 4:6	45.2 1:1
2HPT&S	84.3 41:5	58.0 19:6	48.4 6:3	45.5 1:6	45 0

Numerical Simulations on Synthetic Data

Figure: Mean of the reward of each base arm and super arm (p = 10 dBm).

Table: Average time complexities over 100 experiments when = 0:1

Power	4	6	8	10	12
T&S	1154.3 338:7	654.6 212:1	382.5 129:6	209.4 68:6	133.7 8:9
HT&S	473.2 275:5	271.4 143:4	175.6 69:2	133.2 24:1	123.9 6:5
2PT&S	206.2 60:4	120.2 35:0	68.4 19:4	49.1 4:6	45.2 1:1
2HPT&S	84.3 41:5	58.0 19:6	48.4 6:3	45.5 1:6	45 0

Experiments on real data for a practical scenario in a city available in the longer version of our paper.

Conclusions

- Formulate the beam alignment problem as a multi-armed bandit problem under the pure exploration setting

Conclusions

- Formulate the beam alignment problem as a multi-armed bandit problem under the pure exploration setting
- Derived a general lower bound on the sample complexity

Conclusions

- Formulate the beam alignment problem as a multi-armed bandit problem under the pure exploration setting
- Derived a general lower bound on the sample complexity
- Exploited the structure and properties of the beam alignment problem to derive an algorithm 2PHT&S with a reduced time complexity

Conclusions

- Formulate the beam alignment problem as a multi-armed bandit problem under the pure exploration setting
- Derived a general lower bound on the sample complexity
- Exploited the structure and properties of the beam alignment problem to derive an algorithm 2PHT&S with a reduced time complexity
- Simulations demonstrate superior performances over benchmarks