Channel Estimation Using Joint Dictionary Learning in FDD Massive MIMO Systems

Yacong Ding and Bhaskar Rao

Department of Electrical and Computer Engineering
University of California, San Diego

December 14, 2015
Outline

1 Introduction
   • Motivation
   • Previous Work

2 Joint Uplink/Downlink Dictionary Learning
   • Joint Sparse Representation
   • Joint Compressed Channel Estimation

3 Simulation
   • Simulation Setting
   • Low Dimension Representation
   • Compressed Channel Estimation

4 Conclusion
Introduction
Motivation

- Acquiring downlink channel information $h^d$ at base station (CSIT):
  - TDD: easy, FDD: difficult.
  - BS has N antennas, user has 1 antennas: $h^d, h^u \in \mathbb{C}^{N \times 1}$

<table>
<thead>
<tr>
<th>TDD: $h^d = h^u$, uplink training (UE: $\phi \in \mathbb{C}^{1 \times T^u}$, BS: $Y^u = h^u \phi + n^u$)</th>
<th>Training $T^u \geq 1$</th>
<th>Feedback No</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDD: $h^d \neq h^u$, downlink training (BS: $\Phi \in \mathbb{C}^{T^d \times N}$, UE: $Y^d = \Phi h^d + n^d$)</td>
<td>$T^d \geq N$</td>
<td>$\propto N$</td>
</tr>
</tbody>
</table>
Motivation

- Acquiring downlink channel information $h^d$ at base station (CSIT):
  - TDD: easy, FDD: difficult.
  - BS has N antennas, user has 1 antennas: $h^d, h^u \in \mathbb{C}^{N \times 1}$

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDD: $h^d = h^u$, uplink training</td>
<td>$T^u \geq 1$</td>
<td>No</td>
</tr>
<tr>
<td>(UE: $\phi \in \mathbb{C}^{1 \times T^u}$, BS: $Y^u = h^u \phi + n^u$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDD: $h^d \neq h^u$, downlink training</td>
<td>$T^d \geq N$</td>
<td>$\propto N$</td>
</tr>
<tr>
<td>(BS: $\Phi \in \mathbb{C}^{T^d \times N}$, UE: $Y^d = \Phi h^d + n^d$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- FDD is widely employed in existing communication systems:
  - Beneficial if directly adopt Massive MIMO to FDD.
Motivation

- Acquiring downlink channel information $h^d$ at base station (CSIT):
  - TDD: easy, FDD: difficult.
  - BS has N antennas, user has 1 antennas: $h^d, h^u \in \mathbb{C}^{N \times 1}$

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDD: $h^d = h^u$, uplink training</td>
<td>$T^u \geq 1$</td>
<td>No</td>
</tr>
<tr>
<td>(UE: $\phi \in \mathbb{C}^{1 \times T^u}$, BS: $Y^u = h^u \phi + n^u$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDD: $h^d \neq h^u$, downlink training</td>
<td>$T^d \geq N$</td>
<td>$\propto N$</td>
</tr>
<tr>
<td>(BS: $\Phi \in \mathbb{C}^{T^d \times N}$, UE: $Y^d = \Phi h^d + n^d$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- FDD is widely employed in existing communication systems:
  - Beneficial if directly adopt Massive MIMO to FDD.

- FDD Downlink training: $Y^d = \Phi h^d + n^d$, $\Phi \in \mathbb{C}^{T^d \times N}$:
  - To be practical: $T^d$ small.
Acquiring downlink channel information $h^d$ at base station (CSIT):

- TDD: easy, FDD: difficult.
- BS has $N$ antennas, user has 1 antenna: $h^d, h^u \in \mathbb{C}^{N \times 1}$

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDD: $h^d = h^u$, uplink training (UE: $\phi \in \mathbb{C}^{1 \times T^u}$, BS: $Y^u = h^u \phi + n^u$)</td>
<td>$T^u \geq 1$</td>
<td>No</td>
</tr>
<tr>
<td>FDD: $h^d \neq h^u$, downlink training (BS: $\Phi \in \mathbb{C}^{T^d \times N}$, UE: $Y^d = \Phi h^d + n^d$)</td>
<td>$T^d \geq N$</td>
<td>$\propto N$</td>
</tr>
</tbody>
</table>

FDD is widely employed in existing communication systems:

- Beneficial if directly adopt Massive MIMO to FDD.

FDD Downlink training: $Y^d = \Phi h^d + n^d, \Phi \in \mathbb{C}^{T^d \times N}$:

- To be practical: $T^d$ small.
- $T^d < N$: underdetermined inverse problem, infinite solutions.
Motivation

- Acquiring downlink channel information $h^d$ at base station (CSIT):
  - TDD: easy, FDD: difficult.
  - BS has $N$ antennas, user has 1 antennas: $h^d, h^u \in \mathbb{C}^{N\times1}$

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDD: $h^d = h^u$, uplink training</td>
<td>$T^u \geq 1$</td>
<td>No</td>
</tr>
<tr>
<td>(UE: $\phi \in \mathbb{C}^{1\times T^u}$, BS: $Y^u = h^u\phi + n^u$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDD: $h^d \neq h^u$, downlink training</td>
<td>$T^d \geq N$</td>
<td>$\propto N$</td>
</tr>
<tr>
<td>(BS: $\Phi \in \mathbb{C}^{T^d\times N}$, UE: $Y^d = \Phi h^d + n^d$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- FDD is widely employed in existing communication systems:
  - Beneficial if directly adopt Massive MIMO to FDD.

- FDD Downlink training: $Y^d = \Phi h^d + n^d, \Phi \in \mathbb{C}^{T^d\times N}$:
  - To be practical: $T^d$ small.
  - $T^d < N$: underdetermined inverse problem, infinite solutions.

- Explore channel structure to regularize the problem?
Motivation

- Acquiring downlink channel information $h^d$ at base station (CSIT):
  - TDD: easy, FDD: difficult.
  - BS has N antennas, user has 1 antennas: $h^d, h^u \in \mathbb{C}^{N \times 1}$

<table>
<thead>
<tr>
<th>Training</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T^u \geq 1$</td>
<td>No</td>
</tr>
<tr>
<td>$T^d \geq N$</td>
<td>$\propto N$</td>
</tr>
</tbody>
</table>

- FDD is widely employed in existing communication systems:
  - Beneficial if directly adopt Massive MIMO to FDD.

- FDD Downlink training: $Y^d = \Phi h^d + n^d, \Phi \in \mathbb{C}^{T^d \times N}$:
  - To be practical: $T^d$ small.
  - $T^d < N$: underdetermined inverse problem, infinite solutions.

- Explore channel structure to regularize the problem?
  - **Sparse** channel structure: compressive sensing.
Compressed Channel Estimation

- Low dimensional representation of high dimensional signal:
  - Find a $\Psi$ such that $h^d = \Psi \beta^d$, $\|\beta^d\|_0 < N$. 

Downlink training:

$Y^d = \Phi h^d + n^d = \Phi \Psi \beta^d + n^d$

Apply compressive sensing algorithm to estimate the sparse coefficient $\hat{\beta}^d$.

Compressed Channel Estimation:

$\hat{\beta}^d = \text{arg min}_{\beta^d} \|\beta^d\|_0 \text{ subject to } \|Y^d - \Phi \Psi \beta^d\|_2^2 \leq \sigma^2$

$\hat{h}^d = \Psi \hat{\beta}^d$ (1)

Many practical algorithms.

Measurements: $T^d \propto \|\beta^d\|_0 < N$

Core requirement: find $\Psi$. 

Yacong Ding and Bhaskar Rao (ECE UCSD)  GlobalSIP 2015  December 14, 2015  5 / 21
Compressed Channel Estimation

- Low dimensional representation of high dimensional signal:
  - Find a $\Psi$ such that $h^d = \Psi \beta^d$, $\|\beta^d\|_0 < N$.
  - Downlink training: $Y^d = \Phi h^d + n^d = \Phi \Psi \beta^d + n^d$
Low dimensional representation of high dimensional signal:
- Find a $\Psi$ such that $h^d = \Psi \beta^d$, $\|\beta^d\|_0 < N$.
- Downlink training: $Y^d = \Phi h^d + n^d = \Phi \Psi \beta^d + n^d$

Apply compressive sensing algorithm to estimate the sparse coefficient $\beta^d$.

**Compressed Channel Estimation**: 

$$\hat{\beta}^d = \text{arg min}_{\beta^d} \|\beta^d\|_0 \text{ subject to } \|Y^d - \Phi \Psi \hat{\beta}^d\|_2^2 \leq \sigma^2$$  \hspace{1cm} (1)

$$\hat{h}^d = \Psi \hat{\beta}^d$$
Compressed Channel Estimation

- Low dimensional representation of high dimensional signal:
  - Find a $\Psi$ such that $h^d = \Psi \beta^d$, $\|\beta^d\|_0 < N$.
  - Downlink training: $Y^d = \Phi h^d + n^d = \Phi \Psi \beta^d + n^d$

- Apply **compressive sensing** algorithm to estimate the sparse coefficient $\beta^d$.

  **Compressed Channel Estimation**:
  $\hat{\beta}^d = \arg \min_{\beta^d} \|\beta^d\|_0$ subject to $\|Y^d - \Phi \Psi \beta^d\|_2^2 \leq \sigma^2$ (1)

  $\hat{h}^d = \Psi \hat{\beta}^d$

- Many practical algorithms. **Measurements**: $T^d \propto \|\beta^d\|_0 < N$
Compressed Channel Estimation

- Low dimensional representation of high dimensional signal:
  - Find a $\Psi$ such that $h^d = \Psi \beta^d$, $\|\beta^d\|_0 < N$.
  - Downlink training: $Y^d = \Phi h^d + n^d = \Phi \Psi \beta^d + n^d$

- Apply **compressive sensing** algorithm to estimate the sparse coefficient $\beta^d$.

**Compressed Channel Estimation**

$$\hat{\beta}^d = \arg\min_{\beta^d} \|\beta^d\|_0 \text{ subject to } \|Y^d - \Phi \Psi \beta^d\|_2^2 \leq \sigma^2$$  \hfill (1)

$$\hat{h}^d = \Psi \hat{\beta}^d$$

- Many practical algorithms. **Measurements**: $T^d \propto \|\beta^d\|_0 < N$
- Core requirement: find $\Psi$. 
Sparse Channel Representation
Drawbacks of Existing Work

The existing work that applies compressed channel estimation use **orthogonal** DFT basis as $\Psi$:

- Agree with array manifold using ULA.
- **Infinite** number of antennas, **limited** scattering environment.

For common channels models, such as 3GPP SCM channels:

$$h_d = \Psi_{DFT} \beta_d T_d \propto \| \beta_d \|_0 :$$

lose benefits of compressive sensing

One easy better choice is **overcomplete** DFT matrix: redundancy in basis

Only applicable to ULA.

Can not adapt to specific cell characteristics: urban, rural, hills, etc.

How can we design a dictionary/basis such that:

- Adapt to specific cell properties (antenna, environment).
- Lead to **sparse** representation.
Drawbacks of Existing Work

The existing work that applies compressed channel estimation use **orthogonal** DFT basis as $\Psi$:

- Agree with array manifold using ULA.
- **Infinite** number of antennas, **limited** scattering environment.

For common channels models, such as 3GPP SCM channels:

- High $\|\beta^d\|_0$: $h^d = \Psi_{DFT} \beta^d$
- $T^d \propto \|\beta^d\|_0$: lose benefits of compressive sensing
Drawbacks of Existing Work

The existing work that applies compressed channel estimation use orthogonal DFT basis as $\Psi$:

- Agree with array manifold using ULA.
- **Infinite** number of antennas, **limited** scattering environment.

For common channels models, such as 3GPP SCM channels:

- High $\|\beta^d\|_0$: $h^d = \Psi_{DFT} \beta^d$
- $T^d \propto \|\beta^d\|_0$: lose benefits of compressive sensing

One easy better choice is **overcomplete** DFT matrix: redundancy in basis

- Only applicable to ULA.
- Can not adapt to **specific** cell characteristics: urban, rural, hills, etc.
Drawbacks of Existing Work

The existing work that applies compressed channel estimation use **orthogonal** DFT basis as $\Psi$:

- Agree with array manifold using ULA.
- **Infinite** number of antennas, **limited** scattering environment.

For common channels models, such as 3GPP SCM channels:

- High $\|\beta^d\|_0$: $h^d = \Psi_{DFT} \beta^d$
- $T^d \propto \|\beta^d\|_0$: lose benefits of compressive sensing

One easy better choice is **overcomplete** DFT matrix: redundancy in basis

- Only applicable to **ULA**.
- Can not adapt to **specific** cell characteristics: urban, rural, hills, etc.

How can we design a dictionary/basis such that:
Drawbacks of Existing Work

The existing work that applies compressed channel estimation use **orthogonal** DFT basis as $\Psi$:

- Agree with array manifold using ULA.
- **Infinite** number of antennas, **limited** scattering environment.

For common channels models, such as 3GPP SCM channels:

- High $\|\beta^d\|_0$: $h^d = \Psi_{DFT} \beta^d$
- $T^d \propto \|\beta^d\|_0$: lose benefits of compressive sensing

One easy better choice is **overcomplete** DFT matrix: redundancy in basis

- Only applicable to **ULA**.
- Can not adapt to **specific** cell characteristics: urban, rural, hills, etc.

How can we design a dictionary/basis such that:

- Adapt to **specific** cell properties (antenna, environment).
Drawbacks of Existing Work

The existing work that applies compressed channel estimation use **orthogonal** DFT basis as $\Psi$:

- Agree with array manifold using ULA.
- **Infinite** number of antennas, **limited** scattering environment.

For common channels models, such as 3GPP SCM channels:

- High $\|\beta^d\|_0$: $h^d = \Psi_{DFT}\beta^d$
- $T^d \propto \|\beta^d\|_0$: lose benefits of compressive sensing

One easy better choice is **overcomplete** DFT matrix: redundancy in basis

- Only applicable to **ULA**.
- Can not adapt to **specific** cell characteristics: urban, rural, hills, etc.

How can we design a dictionary/basis such that:

- Adapt to **specific** cell properties (antenna, environment).
- Lead to **sparse** representation.
Rather than using predefined dictionary/basis, learn cell specific $D^d$ from data:
Rather than using **predefined** dictionary/basis, **learn** cell specific $D^d$ from data:

- **Overcomplete**: $D^d \in \mathbb{C}^{N \times M}, N < M$
Rather than using predefined dictionary/basis, learn cell specific $D^d$ from data:

- **Overcomplete:** $D^d \in \mathbb{C}^{N \times M}, N < M$
- **Fit** model to data: $h_i^d \approx D^d \beta_i^d, i = 1, \ldots, L$
Rather than using predefined dictionary/basis, learn cell specific $D^d$ from data:

- Overcomplete: $D^d \in \mathbb{C}^{N \times M}, N < M$
- Fit model to data: $h^d_i \approx D^d \beta^d_i, i = 1, \ldots, L$
- Encourage sparsity: $\| \beta^d_i \|_0 \ll M, \forall i$. 

What data can be utilized?

- Channel measurements: collected within a specific cell.
- Effect of environment on the transmitted electromagnetic waves represented at antennas.

Big data paradigm in wireless communication.
Rather than using predefined dictionary/basis, learn cell specific $D^d$ from data:

- **Overcomplete**: $D^d \in \mathbb{C}^{N \times M}, N < M$
- **Fit** model to data: $h^d_i \approx D^d \beta^d_i, i = 1, \ldots, L$
- **Encourage** sparsity: $\|\beta^d_i\|_0 \ll M, \forall i.$

What data can be utilized?

Channel measurements: collected within a specific cell.
Effect of environment on the transmitted electromagnetic waves represented at antennas.
Big data paradigm in wireless communication.
Rather than using predefined dictionary/basis, learn cell specific $D^d$ from data:

- **Overcomplete:** $D^d \in \mathbb{C}^{N \times M}, N < M$
- **Fit** model to data: $h^d_i \approx D^d \beta^d_i, i = 1, \ldots, L$
- **Encourage** sparsity: $\| \beta^d_i \|_0 \ll M, \forall i$.

- What data can be utilized?
  - Channel measurements: collected within a specific cell.
Our Previous Work: Learning Good Representation

- Rather than using **predefined** dictionary/basis, **learn** cell specific $D^d$ from data:
  - **Overcomplete:** $D^d \in \mathbb{C}^{N \times M}, N < M$
  - **Fit** model to data: $h_i^d \approx D^d \beta_i^d$, $i = 1, \ldots, L$
  - **Encourage** sparsity: $\|\beta_i^d\|_0 \ll M, \forall i$.

- What data can be utilized?
  - Channel measurements: collected within a **specific** cell.
  - Effect of **environment** on the transmitted electromagnetic waves represented at **antennas**.
Our Previous Work: Learning Good Representation

- Rather than using **predefined** dictionary/basis, **learn** cell specific $D^d$ from data:
  - **Overcomplete**: $D^d \in \mathbb{C}^{N \times M}, N < M$
  - **Fit** model to data: $h_i^d \approx D^d \beta_i^d, i = 1, \ldots, L$
  - **Encourage** sparsity: $\|\beta_i^d\|_0 \ll M, \forall i$.

- What data can be utilized?
  - Channel measurements: collected within a **specific** cell.
  - Effect of **environment** on the transmitted electromagnetic waves represented at **antennas**.
  - **Big data** paradigm in wireless communication.
Combine data fitting and sparsity encouragement, dictionary learning can be formulated:

\[
\min_{D^d, \beta^d_1, \ldots, \beta^d_L} \lambda \|H^d - D^d B^d\|_F^2 + \sum_{i=1}^{L} \|\beta^d_i\|_0
\]  

where \( H^d = [h^d_1, \ldots, h^d_L] \) \( B^d = [\beta^d_1, \ldots, \beta^d_L] \).
Combine data fitting and sparsity encouragement, dictionary learning can be formulated:

$$
\min_{D^d, \beta^d_1, \ldots, \beta^d_L} \lambda \| H^d - D^d B^d \|_F^2 + \sum_{i=1}^L \| \beta^d_i \|_0
$$

where \( H^d = [h^d_1, \ldots, h^d_L] \) \( B^d = [\beta^d_1, \ldots, \beta^d_L] \).

Benefits of dictionary learning and compressed channel estimation:

- Applicable to any antenna configurations: no assumed structure.
Dictionary Learning

- Combine data fitting and sparsity encouragement, dictionary learning can be formulated:

\[
\min_{D^d, \beta_1^d, \ldots, \beta_L^d} \lambda \|H^d - D^d B^d\|_F^2 + \sum_{i=1}^L \|\beta_i^d\|_0
\]  

(2)

where \(H^d = [h_1^d, \ldots, h_L^d]\) \(B^d = [\beta_1^d, \ldots, \beta_L^d]\).

- Benefits of dictionary learning and compressed channel estimation:
  - Applicable to any antenna configurations: no assumed structure.
  - Robust to any irregularities: mismatched antennas, non-plane wave.
Dictionary Learning

- Combine data fitting and sparsity encouragement, dictionary learning can be formulated:

\[
\min_{D^d, \beta_1^d, \ldots, \beta_L^d} \lambda \|H^d - D^d B^d\|_F^2 + \sum_{i=1}^L \|\beta_i^d\|_0
\]  

(2)

where \(H^d = [h_1^d, \ldots, h_L^d]\) and \(B^d = [\beta_1^d, \ldots, \beta_L^d]\).

- Benefits of dictionary learning and compressed channel estimation:
  - Applicable to any antenna configurations: no assumed structure.
  - Robust to any irregularities: mismatched antennas, non-plane wave.
  - Training and feedback overhead: proportional to channel sparsity \(S\).
Joint Uplink/Downlink Dictionary Learning and Compressed Channel Estimation
In compressive sensing, more measurements are always better:

- More information about the underlying sparse coefficients.
- Better recovery performance.
Utilizing Uplink Channel Information

In compressive sensing, more measurements are always better:
- More information about the underlying sparse coefficients.
- Better recovery performance.

In Massive MIMO, it implies $T^d$ to be larger:
- Larger training and feedback overhead.
- Waste of resources.
In compressive sensing, more measurements are always better:
- More information about the underlying sparse coefficients.
- Better recovery performance.

In Massive MIMO, it implies $T^d$ to be larger:
- Larger training and feedback overhead.
- Waste of resources.

Is it possible to have more information about the underlying sparse coefficient, but without need of larger $T^d$?
In compressive sensing, more measurements are always better:
- More information about the underlying sparse coefficients.
- Better recovery performance.

In Massive MIMO, it implies $T^d$ to be larger:
- Larger training and feedback overhead.
- Waste of resources.

Is it possible to have more information about the underlying sparse coefficient, but without need of larger $T^d$?

From uplink channel $h^u$: 
In compressive sensing, more measurements are always better:
- More information about the underlying sparse coefficients.
- Better recovery performance.

In Massive MIMO, it implies $T^d$ to be larger:
- Larger training and feedback overhead.
- Waste of resources.

Is it possible to have more information about the underlying sparse coefficient, but without need of larger $T^d$?

From uplink channel $h^u$:
- Easy to obtain: $T^u \geq 1$.
- Common sparse channel structure between $h^d$ and $h^u$. 
Joint Uplink/Downlink Channel Representation

- Similar to $h^d = D^d \beta^d : h^u = D^u \beta^u$. 
Joint Uplink/Downlink Channel Representation

- Similar to $h^d = D^d \beta^d : h^u = D^u \beta^u$.
- Duplex distance not large: **similar scattering effect** for uplink and downlink transmission.

Figure 1: Uplink/Downlink Channel Model
Joint Uplink/Downlink Channel Representation

- Similar to $h^d = D^d \beta^d : h^u = D^u \beta^u$.

- Duplex distance not large: similar scattering effect for uplink and downlink transmission.

![Uplink/Downlink Channel Model](image)

**Figure 1: Uplink/Downlink Channel Model**

- In our model, equivalently to assume $\chi(\beta^u) = \chi(\beta^d)$, where $\chi(\beta) = \{i | \beta(i) \neq 0\}$ denotes the set of locations of nonzero entries in $\beta$. 
Joint UL/DL Dictionary Learning

Joint Uplink/Downlink Dictionary Learning

$$\min_{D^u, B^u, D^d, B^d} \left\| H^u - D^u B^u \right\|_F^2 + \left\| H^d - D^d B^d \right\|_F^2$$

subject to

$$\left\| \beta_i^u \right\|_0 = \left\| \beta_i^d \right\|_0 \leq T_0, \ \chi(\beta_i^u) = \chi(\beta_i^d) \ \forall i$$

(3)
Joint UL/DL Dictionary Learning

Joint Uplink/Downlink Dictionary Learning

$$\min_{D^u, B^u, D^d, B^d} \|H^u - D^u B^u\|_F^2 + \|H^d - D^d B^d\|_F^2$$

subject to  \( \|\beta_i^u\|_0 = \|\beta_i^d\|_0 \leq T_0, \quad \chi(\beta_i^u) = \chi(\beta_i^d) \forall i \) \hspace{1cm} (3)

Joint Uplink/Downlink Sparse Representation

$$h^u \approx D^u \beta^u, \quad h^d \approx D^d \beta^d$$

$$\|\beta_i^u\|_0 = \|\beta_i^d\|_0 \leq T_0, \quad \chi(\beta_i^u) = \chi(\beta_i^d) \forall i$$ \hspace{1cm} (4)
Joint Uplink/Downlink Dictionary Learning

$$\min_{D^u, B^u, D^d, B^d} \| H^u - D^u B^u \|_F^2 + \| H^d - D^d B^d \|_F^2$$

subject to
$$\| \beta^u_i \|_0 = \| \beta^d_i \|_0 \leq T_0, \, \chi(\beta^u_i) = \chi(\beta^d_i) \, \forall \, i$$

Joint Uplink/Downlink Sparse Representation

$$h^u \approx D^u \beta^u, \, h^d \approx D^d \beta^d$$

$$\| \beta^u_i \|_0 = \| \beta^d_i \|_0 \leq T_0, \, \chi(\beta^u_i) = \chi(\beta^d_i) \, \forall \, i$$

Joint Uplink/Downlink Compressed Channel Estimation:

$$\arg \min_{\beta^u, \beta^d} \| Y^u - \phi^u D^u \beta^u \|_2^2 + \| Y^d - \Phi^d D^d \beta^d \|_2^2$$

subject to
$$\chi(\beta^u) = \chi(\beta^d), \, \| \beta^u \|_0 = \| \beta^d \|_0 \leq T_0$$
Benefits of Joint Sparse Framework

- Joint dictionary learning:
  - Regularize the learning process.
  - Better performance when underlying generative model satisfies joint sparsity.
Benefits of Joint Sparse Framework

- Joint dictionary learning:
  - Regularize the learning process.
  - Better performance when underlying generative model satisfies joint sparsity.

- Joint channel estimation:
  - Better recovery: additional measurements from uplink training.
Benefits of Joint Sparse Framework

- **Joint dictionary learning:**
  - Regularize the learning process.
  - Better performance when underlying generative model satisfies joint sparsity.

- **Joint channel estimation:**
  - Better recovery: additional measurements from uplink training.

- In other words, we can further decrease downlink training duration $T^d$ with the same performance.
Numerical Results
Simulation Settings

- Apply 3GPP SCM: far scatterer clusters and local scatterer clusters.
Simulation Settings

- Apply 3GPP SCM: far scatterer clusters and local scatterer clusters.
- Each channel snapshot:
  - 4 local SC: locations change with user.
  - 2 far SC: fixed locations.
- Training samples: 50000 channel snapshots uniformly sampled in the cell.
- 100 antennas at base station and 1 antenna at user. Apply uniform linear array.
- Pair of uplink/downlink channel: same angles, different amplitudes and phases.
Simulation Settings

- Apply 3GPP SCM: far scatterer clusters and local scatterer clusters.
- Each channel snapshot:
  - 4 local SC: locations change with user.
  - 2 far SC: fixed locations.
- Training samples: 50000 channel snapshots uniformly sampled in the cell.
Simulation Settings

- Apply 3GPP SCM: far scatterer clusters and local scatterer clusters.
- Each channel snapshot:
  - 4 local SC: locations change with user.
  - 2 far SC: fixed locations.
- Training samples: 50000 channel snapshots uniformly sampled in the cell.
- 100 antennas at base station and 1 antenna at user. Apply uniform linear array.
Simulation Settings

- Apply 3GPP SCM: far scatterer clusters and local scatterer clusters.
- Each channel snapshot:
  - 4 local SC: locations change with user.
  - 2 far SC: fixed locations.
- Training samples: 50000 channel snapshots uniformly sampled in the cell.
- 100 antennas at base station and 1 antenna at user. Apply uniform linear array.
- Pair of uplink/downlink channel: same angles, different amplitudes and phases.
Low Dimension Representation
Constrain $T_0$ atoms to be used. Compare $\text{MSE}(E\|h^d - \hat{h}^d\|^2_2)$ between $h^d$ and $\hat{h}^d = D^d\beta^d$. $\|h^d\|_2 = 1$.

Figure 2: MSE comparison.
Compressed Channel Estimation
Dictionary Learning in Joint UL/DL Channel Estimation

Compare MSE between $h^d$ and $\hat{h}^d = D^d \hat{\beta}^d$. $\hat{\beta}^d = \text{OMP}(Y^d, \Phi, D^d)$, or $\hat{\beta}^d = \text{jointOMP}(Y^d, \Phi, D^d; Y^u, \phi, D^u)$. $D^d, D^u$: learned dictionary.

Figure 3: MSE comparison.
In this work we propose a novel downlink channel estimation algorithm in FDD Massive MIMO systems.
In this work we propose a novel downlink channel estimation algorithm in FDD Massive MIMO systems.

Joint uplink/downlink dictionary learning can explore similar scattering effect between the uplink and downlink channel, leading to a joint sparse representation.
In this work we propose a novel downlink channel estimation algorithm in FDD Massive MIMO systems.

Joint uplink/downlink dictionary learning can explore similar scattering effect between the uplink and downlink channel, leading to a joint sparse representation.

Joint compressed channel estimation can further improve the recovery performance by utilizing uplink training information.