Degrees-of-Freedom Robust Transmission for the K-user Distributed Broadcast Channel

Presented by Paul de Kerret
Joint work with Antonio Bazco, Nicolas Gresset, and David Gesbert

ESIT 2017 in Madrid,
10/05/2017
Our Focus: Decentralized Broadcast Channel with Imperfect CSIT Sharing

sharing/caching of user’s data symbols

Imperfect CSI sharing

\[ x_1 = w_1 (H_1) s \]
\[ x_2 = w_2 (H_2) s \]
\[ x_3 = w_3 (H_3) s \]
Broadcast Channel (JP-CoMP, Network MIMO)

- Some simplifying assumptions:
  1. \( K \) single-antenna TXs and \( K \) single-antennas RXs
  2. Perfect CSI at the RX
  3. Gaussian data symbols
  4. Block fading channel

- A key assumption: User’s data symbols are available at all TXs
Broadcast Channel (JP-CoMP, Network MIMO)

- Some simplifying assumptions:
  1. $K$ single-antenna TXs and $K$ single-antennas RXs
  2. Perfect CSI at the RX
  3. Gaussian data symbols
  4. Block fading channel

- A key assumption: User’s data symbols are available at all TXs

- Received signal at user $i$

  \[
  y_i = h_i^H x + \eta_i
  \]
Some simplifying assumptions:

(i) $K$ single-antenna TXs and $K$ single-antennas RXs
(ii) Perfect CSI at the RX
(iii) Gaussian data symbols
(iv) Block fading channel

A key assumption: User’s data symbols are available at all TXs

Received signal at user $i$

$$y_i = \begin{bmatrix} h_i^H \end{bmatrix} x + \eta_i$$
Broadcast Channel (JP-CoMP, Network MIMO)

- Some simplifying assumptions:
  1. $K$ single-antenna TXs and $K$ single-antennas RXs
  2. Perfect CSI at the RX
  3. Gaussian data symbols
  4. Block fading channel
- A key assumption: User’s data symbols are available at all TXs
- Received signal at user $i$

\[ y_i = h_i^H \begin{bmatrix} x \end{bmatrix} + \eta_i \]

with \( \{x\}_j \) sent from TX $j$
Some simplifying assumptions:

(i) $K$ single-antenna TXs and $K$ single-antennas RXs
(ii) Perfect CSI at the RX
(iii) Gaussian data symbols
(iv) Block fading channel

A key assumption: User’s data symbols are available at all TXs

Received signal at user $i$

$$y_i = h_i^H x + \eta_i$$

Additive white Gaussian Noise $\mathcal{N}(0, 1)$
Some simplifying assumptions:
(i) $K$ single-antenna TXs and $K$ single-antennas RXs
(ii) Perfect CSI at the RX
(iii) Gaussian data symbols
(iv) Block fading channel

A key assumption: User’s data symbols are available at all TXs

Received signal at user $i$

$$y_i = h_i^H x + \eta_i$$

Additive white Gaussian Noise $\mathcal{N}(0, 1)$

For a given transmit power $P$, let $C(P)$ denote the sum capacity

Our figure of merit will be the Degrees-of-Freedom:

$$\text{DoF} \triangleq \lim_{P \to \infty} \frac{C(P)}{\log_2(P)}$$
Is DoF Useful?

- First order approximation in the SNR

\[ R^* = \text{DoF} \log_2(SNR) + o(\log_2(SNR)) \]

+ Closed form results
+ New insights and new paradigms
+ First step towards capacity

- Inaccurate if strong pathloss differences
- Results not always relevant at finite SNR

Very successful to discover new approaches/insights (MIMO [Telatar, 1999, ETC], IA [Cadambe and Jafar, 2008, TIT], delayed CSIT [Maddah-Ali and Tse, 2012, TIT], ...)
Example

- 2-user IC, single-antenna nodes, $\alpha^2 = 10^{-12}$, $H_{i,j} \sim \mathcal{N}_C(0, 1)$

DoF analysis: $\text{DoF} = 1$ [Etkin et al., 2008, TIT]

Not the expected behaviour
With Generalized DoF, model the pathloss difference

$$\mathbb{E}[|H_{i,j}|^2] = P^{-\gamma_{i,j}}$$

Generalized DoF (GDoF) then defined as

$$\text{DoF} (\{\gamma_{i,j}\}_{i,j}) \triangleq \lim_{P \to \infty} \frac{C(P, \{\gamma_{i,j}\}_{i,j})}{\log_2(P)}$$

Example continued: For $P = 20\,\text{dB}$,

$$\gamma_{1,2} = \gamma_{2,1} = 6$$

and

$$\text{DoF} (\{\gamma_{i,j}\}_{i,j}) = 2$$

Expected behaviour!

Remark

GDoF not discussed here but extension for the 2 users case in [Bazco et al., 2017, ISIT]
Centralized VS Distributed CSI

- **Centralized –TX Independent–**: Conventional model

  \[ \hat{H} = H + \sigma N \]

- **Distributed –TX Dependent–**: Our focus here

  \[ H^{(1)} = H + \sigma^{(1)} N^{(1)} \]
  \[ H^{(2)} = H + \sigma^{(2)} N^{(2)} \]
  \[ H^{(3)} = H + \sigma^{(3)} N^{(3)} \]
Outline

1 Review of the Perfect CSIT Configuration

2 Review of the Centralized CSIT Configuration

3 Towards the Distributed CSIT Configuration

4 Warming up: The 2-User Case

5 The K-user Case
DoF with Perfect CSIT

- All TXs have perfect knowledge of $H$: Optimal DoF is $\text{DoF}^{\text{PCSI}} = K$
- DoF-optimal transmission scheme is Zero Forcing:

$$x = \sqrt{P} \frac{H^{-1}}{\|H^{-1}\|_F} \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix}$$

- Received signal is

$$y_i = \sqrt{P} \frac{1}{\|H^{-1}\|_F} s_i + \eta_i$$

**SNR scales in $P$:** asymptotically possible to decode $s_i$ with the rate $\log_2(P)$ bits

**Remark**

Importantly, $x$ can also be chosen as $x = \sum_{i=1}^{K} \sqrt{\frac{P}{K}} \frac{t_i}{\|t_i\|} s_i$ where the beamformer/precoder $t_i \in \mathbb{C}^{K \times 1}$ is

$$t_i = \prod_{h_1,,h_{i-1},h_{i+1},...,h_K}^{\perp} h_i$$
Outline

1. Review of the Perfect CSIT Configuration
2. Review of the Centralized CSIT Configuration
3. Towards the Distributed CSIT Configuration
4. Warming up: The 2-User Case
5. The K-user Case
Imperfect CSIT in the Centralized Case

- Conventional high SNR parameterization

\[ \hat{h}_i = h_i + \sqrt{P^{-\alpha}} \delta_i \]

- \( \alpha \in [0, 1] \) is called the CSIT quality exponent. Intuitively, equal to the ratio between the "available CSIT" over the "needed CSIT"
  - \( \alpha = 0 \approx \) no CSIT
  - \( \alpha = 1 \approx \) perfect CSIT

Some practical motivation:
- Quantization noise with VQ for \( B >> 1 \), with \( B = \# \) quantization bits,
  \[ \sigma^2 \approx 2^{-\frac{B}{M-1}} \]
- If \( B = \alpha(M - 1) \log_2(P) \), \( \alpha \in [0, 1] \)
  \[ \sigma^2 \approx P^{-\alpha} \]
DoF Analysis of the Centralized Configuration

$$\text{DoF}^{\text{CSI}}(\alpha) = 1 + (K - 1)\alpha$$

- Outerbound recently proven in [Davoodi and Jafar, 2016, TIT]
- Achievable scheme in [Jindal, 2006, TIT][Hao et al., 2015, TCOM]
DoF-Optimal Scheme for the Centralized Case (1) [Jindal, 2006, TIT][Hao et al., 2015, TCOM]

- DoF-optimal scheme: Zero-Forcing (ZF) + Rate Splitting (RS)

\[ y_1 = \sqrt{P} h_1^H \begin{bmatrix} 1 \\ 0 \end{bmatrix} s_0 + \sqrt{P^\alpha} h_1^H t_1^{ZF} s_1 + \sqrt{P^\alpha} h_1^H t_2^{ZF} s_2 \]

Common symbol \( \doteq P \)

Private symbol \( \doteq P^\alpha \)

Interference \( \doteq P^0 \)

with \( t_i^{ZF} = \frac{\nabla^\perp h_i}{\| \nabla^\perp h_i \|} \) and with

\[ |h_1^H t_2^{ZF}|^2 = |\hat{h}_1^H t_2^{ZF} + \sqrt{P^{-\alpha}} \delta_1^H t_2^{ZF}|^2 \]

\[ = P^{-\alpha} |\delta_1^H t_2^{ZF}|^2 \]
DoF-Optimal Scheme for the Centralized Case (2) [Jindal, 2006, TIT][Hao et al., 2015, TCOM]

$$y_1 = \sqrt{P h_1^H \begin{bmatrix} 1 \\ 0 \end{bmatrix}} s_0 + \sqrt{P^\alpha h_1^H t_1^Z} s_1 + \sqrt{P^\alpha h_1^H t_2^Z} s_2$$

- Common symbol $\equiv P$
- Private symbol $\equiv P^\alpha$
- Interference $\equiv P^0$

**Successive Decoding**
- Decode first $s_0$ with rate $(1 - \alpha) \log_2(P)$ bits ($\text{SNR} \equiv P^{1-\alpha}$)
- Decode then $s_1$ with rate $\alpha \log_2(P)$ bits ($\text{SNR} \equiv P^\alpha$)

**Sum DoF** is $(1 - \alpha) + K\alpha$
Outline

1. Review of the Perfect CSIT Configuration
2. Review of the Centralized CSIT Configuration
3. Towards the Distributed CSIT Configuration
4. Warming up: The 2-User Case
5. The K-user Case
Towards the Distributed CSIT Configuration

**Distributed CSIT Configuration**

- With imperfect CSIT sharing extends to

  \[ \hat{h}^{(1)} = h + \sqrt{P^{-\alpha^{(1)}}} \delta^{(1)} \]

  \[ \hat{h}^{(2)} = h + \sqrt{P^{-\alpha^{(2)}}} \delta^{(2)} \]

  \[ \hat{h}^{(3)} = h + \sqrt{P^{-\alpha^{(3)}}} \delta^{(3)} \]

- CSIT configuration characterized by

  \[ 1 \geq \alpha^{(1)} \geq \alpha^{(2)} \geq \ldots \geq \alpha^{(K)} \geq 0 \]

**Remark**

Arbitrary CSIT configuration
An Intuitive Outerbound [de Kerret and Gesbert, 2016, ISIT]

Theorem (The Centralized Outerbound)

\[ \text{DoF}^{\text{DCSI}}(\alpha) \leq 1 + (K - 1) \max_{j \in \{1, \ldots, K\}} \alpha(j) \]

- DoF upperbounded by DoF achieved by full CSIT exchange
- Having \( \hat{H}^{\alpha(1)} \), \ldots, \( \hat{H}^{\alpha(K)} \) doesn't help over having just best CSIT \( \hat{H}^{\alpha(1)} \)
Conventional Zero Forcing [de Kerret and Gesbert, 2012, TIT]

- First idea: Use ZF (DoF-optimal for Centralized CSIT)
  \[ \text{DoF}^{\text{ZF}} = 1 + (K - 1) \min_{j \in \{1, \ldots, K\}} \alpha^{(j)} \]

Very inefficient!

- Why? Goal is to design \( T_1 \) and \( T_2 \) such that
  \[ h_1^H \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \approx 0, \]
  (Zero Forcing constraint at RX 1)

i.e., find a vector orthogonal to \( h_1^H \)
Problem Statement

Towards the Distributed CSIT Configuration

DoF

Centralized outer-bound

Optimal DoF?
Outline

1 Review of the Perfect CSIT Configuration

2 Review of the Centralized CSIT Configuration

3 Towards the Distributed CSIT Configuration

4 Warming up: The 2-User Case

5 The K-user Case
Active-Passive Zero-Forcing (AP-ZF) [de Kerret and Gesbert, 2012, TIT]

- **Main Idea:** Less informed TX generates interference, more informed TX removes it

\[\{(h_{1}^{(1)})^H\}_1 T_1 + \{(h_{1}^{(1)})^H\}_2 T_2 = 0 \rightarrow T_1 = -\frac{\{(h_{1}^{(1)})^H\}_2}{\{(h_{1}^{(1)})^H\}_1} T_2\]

- Achieves the DoF

\[\text{DoF}^{\text{APZF}} = 1 + \alpha^{(1)}\]
Active-Passive Zero-Forcing (AP-ZF)

- Achieves the DoF

\[ \text{DoF}^{\text{APZF}} = 1 + \alpha^{(1)} \]

Remark

In fact achieves also Generalized DoF [Bazco et al., 2017]
Outline

1. Review of the Perfect CSIT Configuration
2. Review of the Centralized CSIT Configuration
3. Towards the Distributed CSIT Configuration
4. Warming up: The 2-User Case
5. The K-user Case
Generalization of AP-ZF?

- **Problem**: AP-ZF doesn’t help much with more users

\[ \text{DoF}^{\text{APZF}} = 1 + (K - 1)\alpha^{(K-1)} \]

- Need for a different approach

**Main Idea**

Exploit interference as side information: Interference useful for both the *interfered user* and the *desired user*

- Analogy to the use of delayed CSIT [Maddah-Ali and Tse, 2012, TIT]
A Multi-layer Transmission Scheme [de Kerret and Gesbert, 2016, ISIT]

- All TXs serve all users with power $P^{\alpha(1)}$ using Active-Passive ZF

$$\mathbf{x} = \sqrt{P^{\alpha(1)}} \sum_{i=1}^{K} T_{i}^{\text{APZF}} \mathbf{s}_{i}$$

- Generate interferences of power $P^{\alpha(1)}$
TX 1 estimates and quantizes the interference terms before their generations.

\[ x = \sqrt{P^{(1)}} \sum_{i=1}^{K} T_{i}^{\text{APZF}} s_{i} \]
The K-user Case

A Multi-layer Transmission Scheme [de Kerret and Gesbert, 2016, ISIT]

TX 1 then transmits them via a common data symbol at the same time as the private data symbols

\[ x = \sqrt{P} \begin{bmatrix} 1 \\ 0_{K-1} \end{bmatrix} s_0 + \sqrt{P \alpha^{(1)}} \sum_{i=1}^{K} T_i^{APZF} s_i \]
Signal Processing at TX 1

Interference estimation at TX 1:

\[
\sqrt{P^{\alpha(1)}} (\hat{h}^{(1)}_1)^H T_2^{\text{APZF}} s_2 = \sqrt{P^{\alpha(1)}} (\hat{h}^{(1)}_1) + \sqrt{P^{-\alpha(1)}} \delta^{(1)}_1)^H T_2^{\text{APZF}} s_2
\]

\[
= \sqrt{P^{\alpha(1)}} h_1^H T_2^{\text{APZF}} s_2 + \sqrt{P^{-\alpha(1)}} (\delta^{(1)}_1)^H T_2^{\text{APZF}} s_2
\]

TX 1 can compute DoF-perfect estimate of the interference terms!

Interference quantization: Use \( \alpha^{(1)} \log_2(P) \) bits to quantize the signal scaling in \( P^{\alpha(1)} \).

Quantization error scaling in \( P^0 \) [Cover and Thomas, 2006]

Transmit \( 3\alpha^{(1)} \log_2(P) \) bits to all users
Signal Processing at RX 1 (w.l.o.g.)

- User 1 has received
  \[ y_1 = \sqrt{P} h_1^H \begin{bmatrix} 1 \\ 0_{K-1} \end{bmatrix} s_0 + \sqrt{P} h_1^H T_1^{APZF} s_1 + \sqrt{P} h_1^H T_2^{APZF} s_2 + \sqrt{P} h_1^H T_3^{APZF} s_3 \]

- User 1 decodes \( s_0 \) and obtains then
  \[
  \sqrt{P} (\hat{h}_1^{(1)})^H T_2^{APZF} s_2, \quad \text{Useful: Remove interference}
  \]
  \[
  \sqrt{P} (\hat{h}_2^{(1)})^H T_3^{APZF} s_3, \quad \text{Useless (for RX 1)}
  \]
  \[
  \sqrt{P} (\hat{h}_3^{(1)})^H T_1^{APZF} s_1, \quad \text{Useful: Desired data}
  \]

- Achieved DoF: If \( 3\alpha^{(1)} \leq 1 - \alpha^{(1)} \), achieves DoF
  \[
  \text{DoF} = 6\alpha^{(1)} + (1 - \alpha^{(1)}) - 3\alpha^{(1)}
  \]
  \[= 2\alpha^{(1)} \text{ per user} + \text{multicast DoF after retransmitting interference} \]
Weak CSIT Regime

Theorem

If \( \max_{j \in \{1, \ldots, K\}} \alpha^{(j)} \leq \frac{1}{1 + K(K-2)} \) (weak CSIT regime),

\[
\text{DoF}^{\text{DCSI}}(\alpha) \geq 1 + (K - 1) \max_{j \in \{1, \ldots, K\}} \alpha^{(j)}
\]

Figure: DoF as a function of \( \alpha^{(1)} \) for \( \alpha^{(2)} = \frac{2}{3} \alpha^{(1)} \) and \( \alpha^{(3)} = 0 \)
Take Home Message

- DoF analysis allows to develop new schemes/insights with simple linear algebra
- Role of each TX adapts to the full multi-TX CSIT configuration
- Multi-layer transmission scheme: Estimate, Quantize & transmit interference at the most informed user
- Many extensions:
  - Developed a new **Hierarchical Zero-Forcing** to extend optimality region
  - Extend further?
  - Improving the centralized-outerbound
  - Going beyond the DoF

![Graph showing DoF analysis results with different CSIT regimes and concentric outer bounds.](image-url)
References I


thanks
Extension of the Weak CSIT Regime for $K = 3$ [de Kerret et al., 2016a, Asilomar]

- Improved scheme building on a new **Hierarchical ZF** precoding paradigm

**Figure:** DoF as a function of $\alpha^{(1)}$ for $\alpha^{(2)} = \frac{2}{3} \alpha^{(1)}$ and $\alpha^{(3)} = 0$
Beyond the Weak CSIT Regime

Definition (Weak CSIT regime)
\[ \alpha^{(1)} \leq \frac{1}{4} + \frac{3}{4} \alpha^{(2)} \]

Definition (Heterogeneous CSIT regime)
\[ \alpha^{(1)} > \min \left( 2\alpha^{(2)}, \frac{1}{4} + \frac{3}{4} \alpha^{(2)} \right) \]

Definition (Intermediate CSIT regime)
\[ \frac{1}{4} + \frac{3}{4} \alpha^{(2)} < \alpha^{(1)} \leq 2\alpha^{(2)} \]
Achievable DoF \cite{deKerretetal2016a,Asilomar}

Theorem

*In the 3-user MIMO BC with D-CSIT, it holds that*

\[
\text{DoF}^{\text{DCSI}}(\alpha) \geq \begin{cases} 
1 + 2\alpha^{(1)} & \text{(Weak CSIT)} \\
\frac{3}{2} (1 + \alpha^{(2)}) & \text{(Intermediate CSIT)} \\
1 + \alpha^{(1)} + \frac{3\alpha^{(1)}(1-\alpha^{(1)})+\alpha^{(2)}(5\alpha^{(1)}-3\alpha^{(2)}-1)}{9\alpha^{(1)}-8\alpha^{(2)}} & \text{(Heterogeneous CSIT)}
\end{cases}
\]

Can be achieved building on new precoding scheme: Hierarchical zero-forcing
Hierarchical ZF with $K = 3$: Main Property

**Lemma**

Let $t_3^{HZF}$ be the HZF beamformer towards user 3 with average power $P$. Then:

$$|h_1^H t_3^{HZF}|^2 \leq P^{1-\alpha(1)}$$

$$|h_2^H t_3^{HZF}|^2 \leq P^{1-\alpha(2)}$$

compared with conventional ZF
Roadmap of Hierarchical ZF

1. Make CSIT hierarchical
2. Split precoding in layers
3. Design layer $k$ to reduce interference at user $k$ without reducing interference reduction already realized
(1) Make CSIT Hierarchical

Example

- Example for two transmitters TX1, TX2 with $\alpha^{(1)} \geq \alpha^{(2)}$
- Let $Q_{\alpha^{(2)}}$ be our Hierarchical quantizer using $\alpha^{(2)} \log_2(P)$ bits
- Let us define

$$\hat{H}^{(1)}_{\alpha^{(2)}} \triangleq Q_{\alpha^{(2)}}(\hat{H}^{(1)})$$

$$\hat{H}^{(2)}_{\alpha^{(2)}} \triangleq Q_{\alpha^{(2)}}(\hat{H}^{(2)})$$

- Then, there exists a quantizer $Q_{\alpha^{(2)}}$ such that [de Kerret et al., 2016b, ITW]

$$\lim_{P \to \infty} \Pr \left\{ \hat{H}^{(1)}_{\alpha^{(2)}} = \hat{H}^{(2)}_{\alpha^{(2)}} \right\} = 1$$

$$\mathbb{E} \left[ \|\hat{H}^{(j)}_{\alpha^{(2)}} - H\|_F^2 \right] \leq P^{-\alpha^{(2)}}, \quad j = 1, 2$$

⇒ TX 1 can obtain $\hat{H}^{(2)}_{\alpha^{(2)}}$: CSIT configuration has been made hierarchical

⇒ More generally, TX $i$ knows what TX $i + 1$ knows (post quantizing)
(1) Make CSIT Hierarchical
(2) Split Precoding in Layers

- $t_3^{HZF}$ aimed at user 3 decomposed as

$$t_3^{HZF} = \begin{bmatrix} t_3^{HZF}(1) \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \{t_3^{HZF}(2)\}_1 \\ \{t_3^{HZF}(2)\}_2 \\ 0 \end{bmatrix} + \begin{bmatrix} \{t_3^{HZF}(3)\}_1 \\ \{t_3^{HZF}(3)\}_2 \\ \{t_3^{HZF}(3)\}_3 \end{bmatrix}$$

- e.g., TX 2 needs to be able to compute the 2th row:

$$t_3^{HZF} = \begin{bmatrix} t_3^{HZF}(1) \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \{t_3^{HZF}(2)\}_1 \\ \{t_3^{HZF}(2)\}_2 \\ 0 \end{bmatrix} + \begin{bmatrix} \{t_3^{HZF}(3)\}_1 \\ \{t_3^{HZF}(3)\}_2 \\ \{t_3^{HZF}(3)\}_3 \end{bmatrix}$$
(3) Hierarchical ZF for $K = 3$

- First “layer” (at TX 1, TX 2 and TX 3)

$$
t_{3}^{HZF}(3) = \lambda^{HZF} \hat{H}^{(3)H} \left( \hat{H}^{(3)}(\hat{H}^{(3)})^{H} + \frac{1}{P} I_{3} \right)^{-1} e_{3}
$$
(3) Hierarchical ZF for $K = 3$

- **First “layer” (at TX 1, TX 2 and TX 3)**

$$\mathbf{t}_3^{\text{HZF}}(3) = \lambda^{\text{HZF}} \hat{\mathbf{H}}(3) \mathbf{H} \left( \hat{\mathbf{H}}(3)(\hat{\mathbf{H}}(3))^H + \frac{1}{P} \mathbf{l}_3 \right)^{-1} \mathbf{e}_3$$

- **Second “layer” (at TX 1 and TX 2)**

$$\mathbf{t}_3^{\text{HZF}}(2) = -\hat{\mathbf{H}}(2) \mathbf{H}_{[1:2,1:2]} \left( \hat{\mathbf{H}}(2) \mathbf{H}_{[1:2,1:2]} + \frac{1}{P} \mathbf{l}_2 \right)^{-1} \hat{\mathbf{H}}(2) \mathbf{H}_{[1:2,1:3]} \mathbf{t}_3^{\text{HZF}}(3)$$
(3) Hierarchical ZF for $K = 3$

- **First “layer” (at TX 1, TX 2 and TX 3)**

  $$t_{3}^{HZF} (3) = \lambda^{HZF} \hat{H}^{(3)H} (\hat{H}^{(3)} (\hat{H}^{(3)})^{H} + \frac{1}{P} I_{3})^{-1} e_{3}$$

- **Second “layer” (at TX 1 and TX 2)**

  $$t_{3}^{HZF} (2) = -\hat{H}^{(2)H} [1:2,1:2] (\hat{H}^{(2)} [1:2,1:2] \hat{H}^{(2)H} [1:2,1:2] + \frac{1}{P} I_{2})^{-1} \hat{H}^{(2)} [1:2,1:3] t_{3}^{HZF} (3)$$

- **Third “layer” (at TX 1)**

  $$t_{3}^{HZF} (1) = -\hat{H}^{(1)H} [1,1] (|\hat{H}^{(1)} [1,1]|^{2} + \frac{1}{P})^{-1} \hat{H}^{(1)H} \left( [t_{3}^{HZF} (2)] + t_{3}^{HZF} (3) \right)$$
(3) Hierarchical ZF for $K = 3$

- First “layer” (at TX 1, TX 2 and TX 3)
  \[ t_3^{HZF}(3) = \lambda^{HZF} \hat{H}^{(3)H} \left( \hat{H}^{(3)} \left( \hat{H}^{(3)} \right)^H + \frac{1}{P} I_3 \right)^{-1} e_3 \]

- Second “layer” (at TX 1 and TX 2)
  \[ t_3^{HZF}(2) = -\hat{H}^{(2)H}_{[1:2,1:2]} \left( \hat{H}^{(2)}_{[1:2,1:2]} \hat{H}^{(2)H}_{[1:2,1:2]} + \frac{1}{P} I_2 \right)^{-1} \hat{H}^{(2)}_{[1:2,1:3]} t_3^{HZF}(3) \]

- Third “layer” (at TX 1)
  \[ t_3^{HZF}(1) = -\hat{H}^{(1)H}_{1,1} \left( |\hat{H}^{(1)}_{1,1}|^2 + \frac{1}{P} \right)^{-1} \hat{H}^{(1)H}_{1,1} \begin{bmatrix} t_3^{HZF}(2) \\ 0 \end{bmatrix} + t_3^{HZF}(3) \]

Main Intuition of the Proof

- Does not increase interference at user 2 when reducing interference at user 1 because
  \[ |t_3^{HZF}(1)|^2 \leq P^{1-\alpha(2)} \]
Transmission Scheme

Less interference bits to convey in second layer
More information bits can be squeezed in first layer