Aalto University
School of Electrical
Engineering

Large-System Analysis of Rate Regions in Bidirectional Full-Duplex MIMO Link: Suppression versus Cancellation

Taneli Riihonen, Mikko Vehkaperä, and Risto Wichman
Aalto University School of Electrical Engineering, Finland

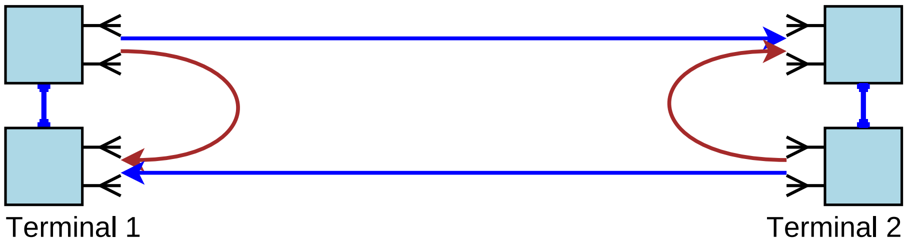
Session WA5 "Communication Over Full-duplex Channels", March 20, 2013
47th Annual Conference on Information Sciences and Systems (CISS)

Introduction

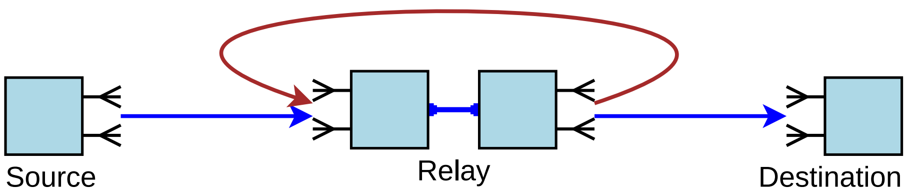
General Topic: Full-Duplex Wireless

- *"Full-duplex"* wireless communication
  - = systems where some node(s) may transmit and receive simultaneously on a *single* frequency band
- Progressive physical/link-layer *frequency-reuse* concept
  - = up to double spectral efficiency at system level, if the significant technical problem of *self-interference* is tackled
- Transmission and reception should use the band for the same amount of time to make the most of full duplex
  - ▷ (a)symmetry of traffic pattern, i.e., *requested* rates in the two simultaneous directions
  - ▷ (a)symmetry of channel quality, i.e., *achieved* rates in the two simultaneous directions

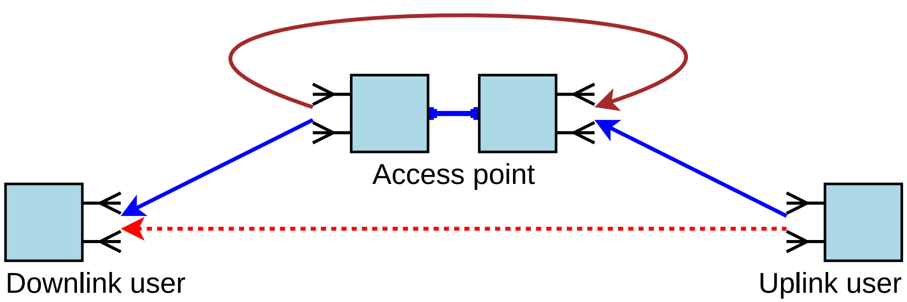
Full-Duplex Communication Scenarios



1) Bidirectional communication link between two terminals
   - Asymmetric traffic (typically)
   - Symmetric channels (roughly)

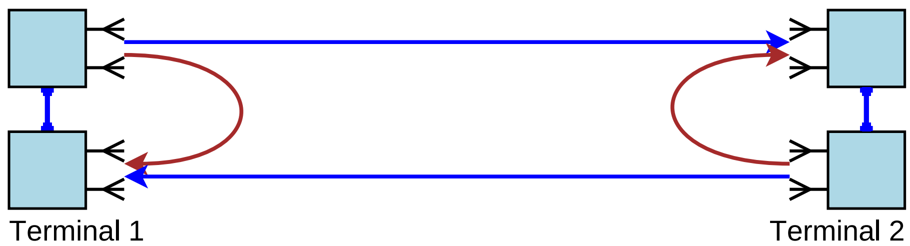


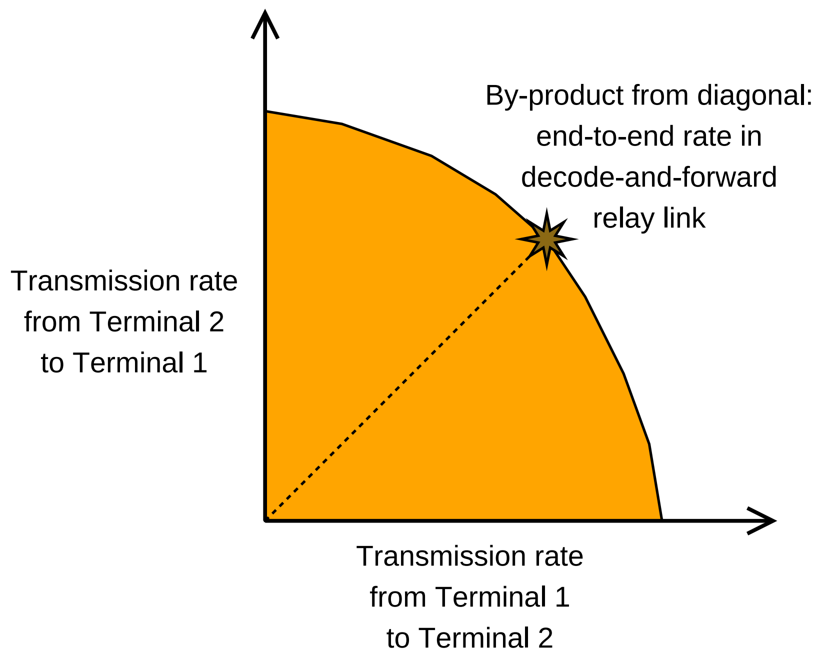
2) Multihop relay link
   - Symmetric traffic
   - Asymmetric channels
   - Direct link may be useful



3) Simultaneous down- and uplink for two half-duplex users
   - Asymmetric traffic
   - Asymmetric channels
   - Inter-user interference!

Scope: Rate Regions in Two-Way Communication

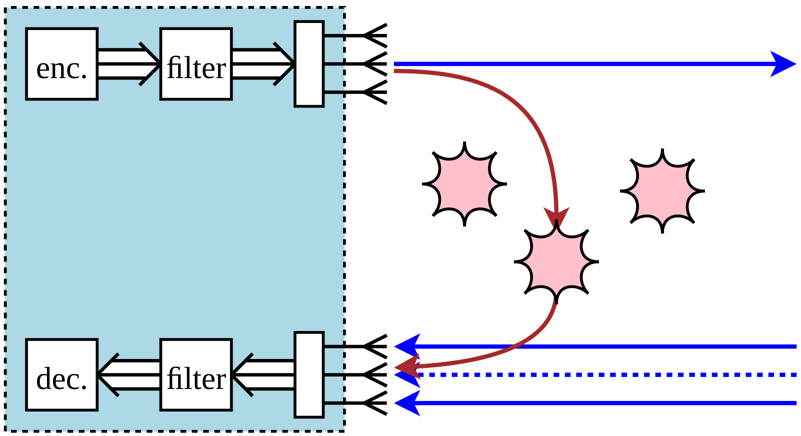




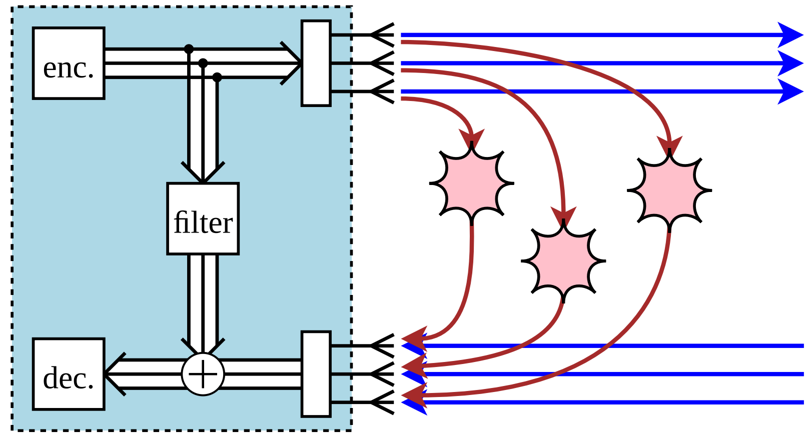
- Bidirectional full-duplex multiantenna (MIMO) link
  - at the large-system limit
  - with asymmetric traffic
  - assuming symmetric setup for numerical results
- Achievable rate regions by controlling
  - spatial multiplexing
  - time sharing
- The analysis is based on the *replica method* borrowed from statistical physics

Focus: Suppression vs. Cancellation

Spatial-domain suppression:



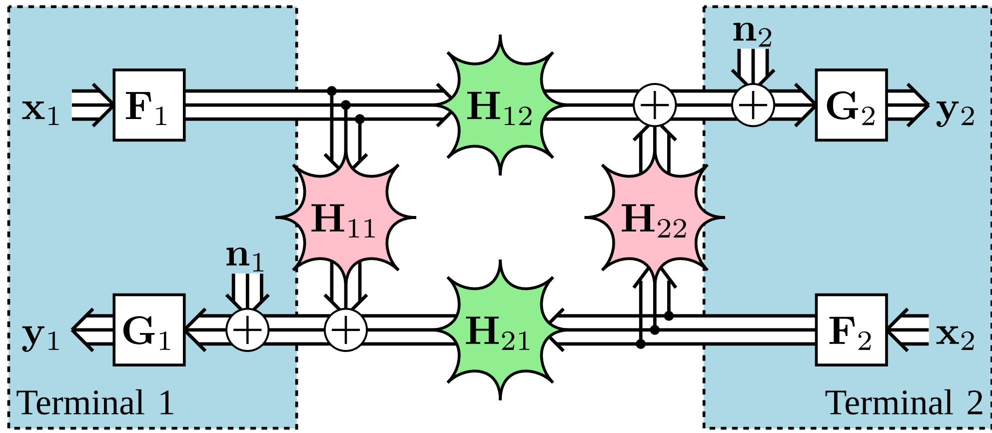
Time-domain cancellation:



- The link needs efficient self-interference mitigation at both ends
  - **Suppression**: forming eigenbeams to transmit and receive in orthogonal directions ("null-space projection")
  - **Cancellation**: subtracting the interfering signal before decoder
- Both schemes can eliminate interference, but suppression is possible only at the cost of consuming spatial degrees of freedom

System Model

Signal Model



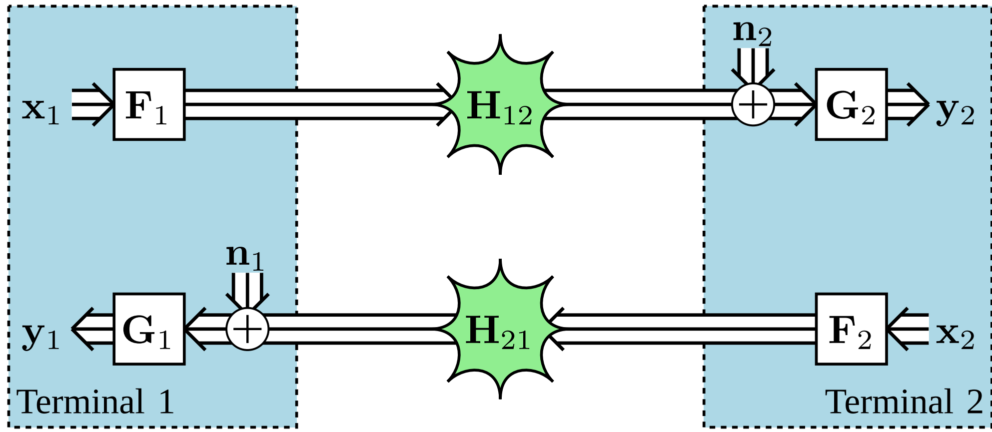
- Terminal $i \in \{1, 2\}$ has M_i transmit and N_i receive antennas
- In communication direction $ij \in \{12, 21\}$:

$$\mathbf{y}_j = \mathbf{G}_j \mathbf{H}_{ij} \mathbf{F}_i \mathbf{x}_i + \mathbf{G}_j \mathbf{H}_{jj} \mathbf{F}_j \mathbf{x}_j + \mathbf{G}_j \mathbf{n}_j$$

  - The link reserves $\hat{M}_i$ transmit and $\hat{N}_j$ receive streams for spatial multiplexing after self-interference mitigation

- Terminal i does not know $\mathbf{H}_{ij}$ but Terminal j knows $\mathbf{H}_{ij}$ and $\mathbf{H}_{jj}$

Spatial-Domain Suppression

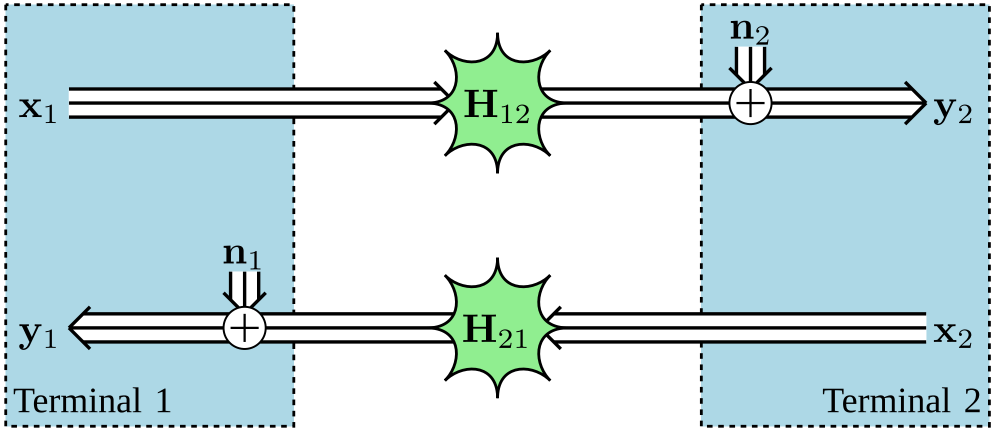


- Suppression exploits the transmit and receive beamforming filters:

$$\mathbf{F}_j \in \mathbb{C}^{M_j \times \hat{M}_j} \quad \text{and} \quad \mathbf{G}_j \in \mathbb{C}^{\hat{N}_j \times N_j}$$

  - ▷ Orthonormal spatial streams: $\mathbf{F}_j^H \mathbf{F}_j = \mathbf{I}$ and $\mathbf{G}_j \mathbf{G}_j^H = \mathbf{I}$
    - – Maximum for full-rank $\mathbf{H}_{jj}$ is $\hat{M}_j + \hat{N}_j = \max\{M_j, N_j\}$
- Self-interference is eliminated in Terminal j if $\boxed{\mathbf{G}_j \mathbf{H}_{jj} \mathbf{F}_j = \mathbf{0}}$
  - ▷ Implemented using the SVD of $\mathbf{H}_{jj}$ (for instance)

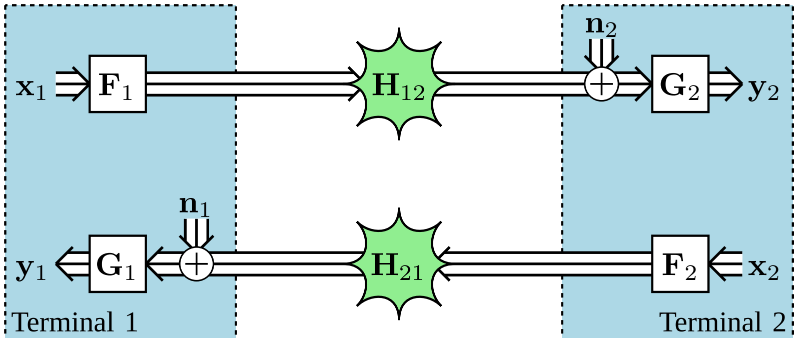
Time-Domain Cancellation



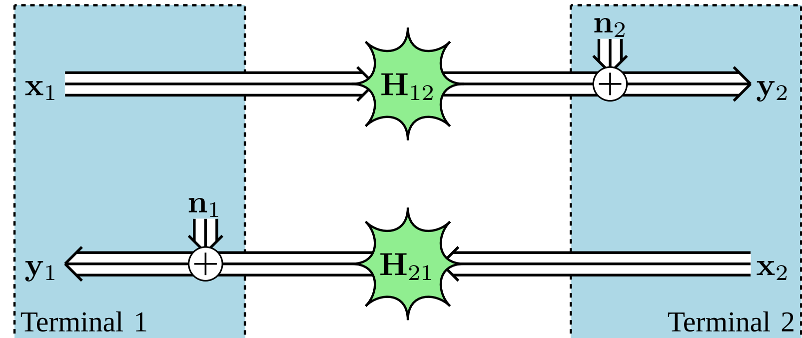
- Cancellation is based on the subtraction of the interfering signal so that decoder input becomes $\boxed{\mathbf{y}_j - \mathbf{G}_j \mathbf{H}_{jj} \mathbf{F}_j \mathbf{x}_j}$
  - ▷ Terminal $j \in \{1, 2\}$ needs to know its own transmitted signal $\mathbf{x}_j$ which is not required with spatial-domain suppression
- All spatial degrees of freedom can be reserved for multiplexing
  - ▷ $\hat{M}_j = M_j$, $\hat{N}_j = N_j$ and $\mathbf{F}_j = \mathbf{I}$, $\mathbf{G}_j = \mathbf{I}$ in the analysis

Spatial-Division Multiplexing

Spatial-domain *suppression*:



Time-domain *cancellation*:



- After mitigation, the signal model is transformed to

$$\mathbf{y}_j = \mathbf{G}_j \mathbf{H}_{ij} \mathbf{F}_i \mathbf{x}_i + \mathbf{G}_j \mathbf{n}_j, \qquad \mathcal{E}\{\mathbf{x}_i \mathbf{x}_i^H\} = (1/\hat{M}_i)\,\mathbf{I}$$

  - ▷ Transmitter side: standard open-loop spatial multiplexing of independent Gaussian streams into $\mathbf{x}_i$
  - ▷ Receiver side: joint decoding for $\mathbf{y}_j$ knowing $\mathbf{G}_j \mathbf{H}_{ij} \mathbf{F}_i \in \mathbb{C}^{\hat{N}_j \times \hat{M}_i}$
- Time sharing between different stream configurations in order to make the achievable rate region convex with suppression

Analytical Results

Mutual Information

- We are interested in evaluating the average transmission rate as

$$\boxed{R_{ij} = \mathcal{E}\{\log\det(\mathbf{I} + \frac{1}{\hat{M}_i}\mathbf{G}_j\mathbf{H}_{ij}\mathbf{F}_i(\mathbf{G}_j\mathbf{H}_{ij}\mathbf{F}_i)^H)\}}$$

  over the joint distribution of random matrices $\mathbf{G}_j$, $\mathbf{H}_{ij}$, and $\mathbf{F}_i$
- Instead, we begin from the definition of mutual information:

$$\frac{R_{ij}}{\hat{M}_i} = \frac{\mathcal{E}\{\log p(\mathbf{y}_j \mid \mathbf{x}_i, \mathbf{G}_j, \mathbf{H}_{ij}, \mathbf{F}_i)\}}{\hat{M}_i} - \frac{\mathcal{E}\{\log \mathcal{E}_{\mathbf{x}_i}\{p(\mathbf{y}_j \mid \mathbf{x}_i, \mathbf{G}_j, \mathbf{H}_{ij}, \mathbf{F}_i)\}\}}{\hat{M}_i}$$

  where $p(\cdot \mid \cdot)$ is the Gaussian posterior probability
- The above expression can be transformed to

$$\frac{R_{ij}}{\hat{M}_i} = -\frac{\hat{N}_j}{\hat{M}_i} - \frac{1}{\hat{M}_i}\lim_{u\to 0}\frac{\partial}{\partial u}\log\mathcal{E}\{\mathcal{E}_{\mathbf{x}_i}\{\exp(-\|\mathbf{y}_j - \mathbf{G}_j\mathbf{H}_{ij}\mathbf{F}_i\mathbf{x}_i\|^2)\}^u\}$$

  where the first term is trivial and the second term comes from the identity $\lim_{u\to 0}\frac{\partial}{\partial u}\log\mathcal{E}\{Z^u\} = \mathcal{E}\{\log Z\}$

Replica Method and Integration

- With $\Delta\mathbf{x}_a = \mathbf{x}_0 - \mathbf{x}_a$, the replica trick amounts to evaluating

$$\frac{R_{ij}}{\hat{M}_i} = -\frac{\hat{N}_j}{\hat{M}_i} - \lim_{\hat{M}_i\to\infty}\frac{1}{\hat{M}_i}\lim_{u\to 0}\frac{\partial}{\partial u}\log\mathcal{E}\{\prod_{a=1}^{u}\mathrm{e}^{-\|\hat{M}_i^{-1/2}\mathbf{G}_j\mathbf{H}_{ij}\mathbf{F}_i\Delta\mathbf{x}_a+\mathbf{G}_j\mathbf{n}_j\|^2}\}$$

  where u is an integer inside $\log$ but a real number outside $\log$ (!?)
- After Gaussian integration over $\mathbf{n}_j$ and $\mathbf{v}_a = \hat{M}_i^{-1/2}\mathbf{H}_{ij}\mathbf{F}_i\Delta\mathbf{x}_a$,

$$\frac{R_{ij}}{\hat{M}_i} = -\frac{\hat{N}_j}{\hat{M}_i} - \lim_{\hat{M}_i\to\infty}\frac{1}{\hat{M}_i}\lim_{u\to 0}\frac{\partial}{\partial u}\log\mathcal{E}\{\mathrm{e}^{G(\mathbf{Q},\mathbf{D}_j)}\}$$

  where $\{\mathbf{Q}\}_{a,b} = \frac{1}{\hat{M}_i}\mathbf{x}_b^H\mathbf{x}_a$ and $\mathbf{D}_j = \mathbf{T}_j^T\mathbf{T}_j$ is binary and diagonal
- If the limits can be swapped, the saddle-point method implies

$$\frac{R_{ij}}{\hat{M}_i} = -\frac{\hat{N}_j}{\hat{M}_i} - \lim_{u\to 0}\frac{\partial}{\partial u}\lim_{\hat{M}_i\to\infty}\frac{1}{\hat{M}_i}\log\mathcal{E}_{\mathbf{D}_j}\{\exp(\hat{M}_i\operatorname*{extr}_{\mathbf{Q},\tilde{\mathbf{Q}}}T(\mathbf{Q},\tilde{\mathbf{Q}},\mathbf{D}_j))\}$$

  where $T(\mathbf{Q},\tilde{\mathbf{Q}},\mathbf{D}_j) = \frac{1}{\hat{M}_i}G(\mathbf{Q},\mathbf{D}_j) - \mathrm{tr}(\mathbf{Q}\tilde{\mathbf{Q}}) + \log M(\tilde{\mathbf{Q}})$

Replica Symmetry Assumption and Extremization

- Before extremization, $T(\mathbf{Q}, \tilde{\mathbf{Q}}, \mathbf{D}_j)$ is transformed by replica symmetry ($\mathbf{Q} = \mathbf{I}_{u+1}(p-q) + \mathbf{1}_{(u+1)\times(u+1)}q$ and $\tilde{\mathbf{Q}} = \mathbf{I}_{u+1}(\tilde{p}-\tilde{q}) + \mathbf{1}_{(u+1)\times(u+1)}\tilde{q}$) to $T_u(p,q,\tilde{p},\tilde{q}) = -u\,\frac{\hat{N}_j}{\hat{M}_i}\log(1+\bar{\gamma}_{ij}(p-q)) - (u+1)(p\tilde{p}+uq\tilde{q}) + \log M(\tilde{\mathbf{Q}})$
- Matrix $\mathbf{D}_j$ also disappears and we get a tractable form as

$$\frac{R_{ij}}{\hat{M}_i} = -\lim_{u\to 0}\frac{\partial}{\partial u}\operatorname*{extr}_{p,q,\tilde{p},\tilde{q}} T_u(p,q,\tilde{p},\tilde{q})$$

  which matches to the case of an i.i.d. Gaussian $\hat{M}_i \times \hat{N}_j$ channel
- Finally, we may exploit existing proofs (e.g., by Verdú) to obtain

$$\boxed{\frac{R_{ij}}{\hat{M}_i} \simeq \log\left(1 + \frac{\hat{N}_j}{\hat{M}_i}\cdot\frac{\bar{\gamma}_{ij}}{1+E}\right) + \frac{\hat{N}_j}{\hat{M}_i}\left(\log(1+E) - \frac{E}{1+E}\right)}$$

  where $E = \bar{\gamma}_{ij}(p-q)$ is a solution to $\frac{\bar{\gamma}_{ij}}{E} = 1 + \frac{\hat{N}_j}{\hat{M}_i}\cdot\frac{\bar{\gamma}_{ij}}{1+E}$
  - The achievable transmission rates of the two directions are indirectly coupled via $\hat{M}_j + \hat{N}_j = \max\{M_j, N_j\}$

Numerical Results

Example Setups

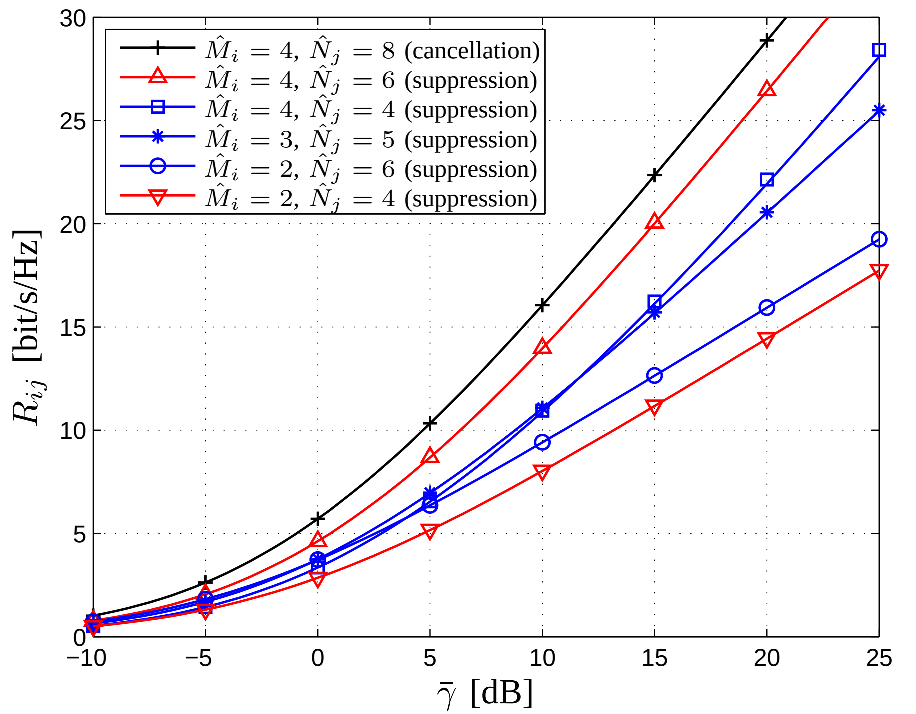
- The numerical results concentrate on symmetric systems where
  - $M = M_1 = M_2$
  - $N = N_1 = N_2$
  - $\bar{\gamma} = \bar{\gamma}_{12} = \bar{\gamma}_{21}$
- However, some asymmetry should be taken into account
  - Requested rates may be different in the two directions, reflecting typical downlink/uplink imbalance (R_{12}/R_{21})
  - There may be transmit/receive antenna imbalance (M/N)
    - At the large-system limit, M and N grow asymptotically
- In summary, there are three key parameters to explore:

$$\boxed{R_{12}/R_{21}} \qquad \boxed{\bar{\gamma}} \qquad \boxed{M/N}$$

Transmission Rate vs. SNR

- When
  - $M = 4$
  - $N = 8$

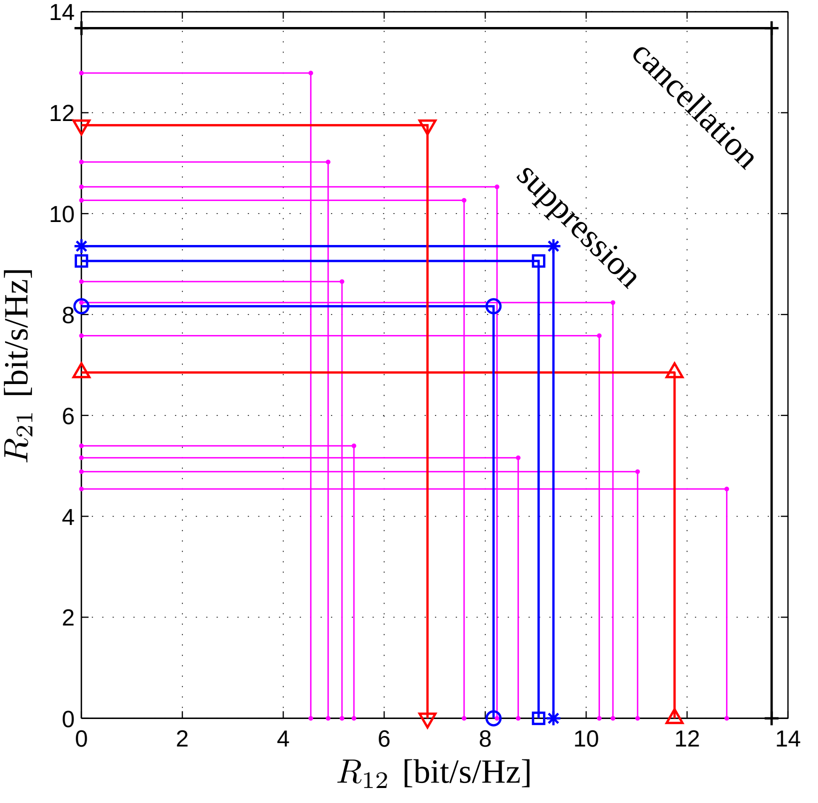
  a) lines:
  asymptotic *analytical* values projected to this finite case
  b) markers:
  accurate *simulated* values



- The asymptotic results are useful also for not-so-large systems
- Trade-off (indirect coupling) between rates in two directions:
  When choosing $(\hat{M}_i, \hat{N}_j)$ as a stream configuration in one direction, the opposite configuration becomes $(\hat{M}_j, \hat{N}_i) = (8 - \hat{N}_j, 8 - \hat{M}_i)$

Achievable Rate Regions (1)

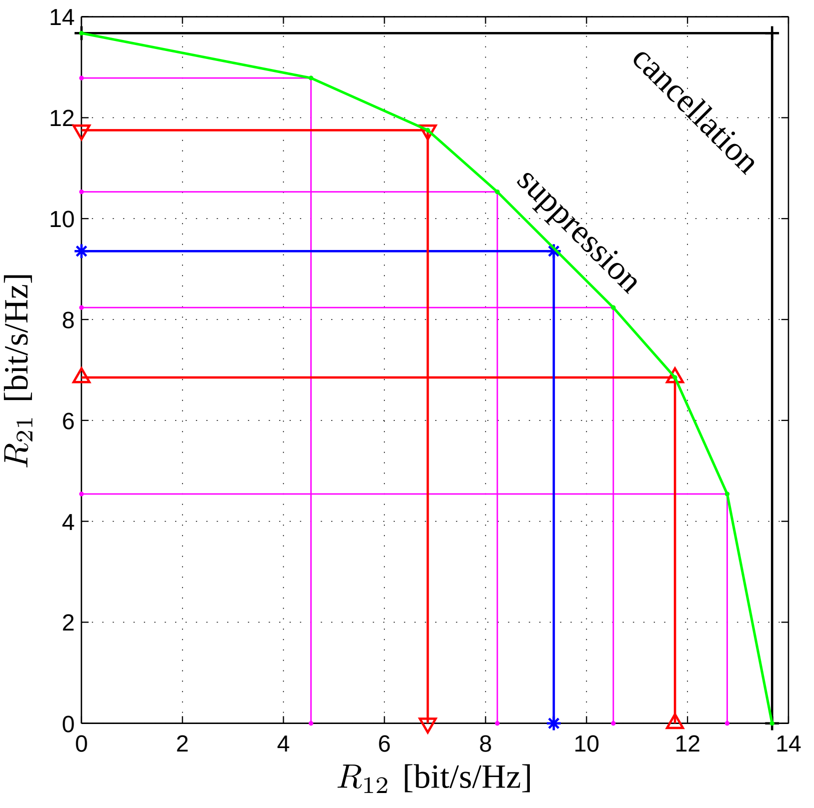
- When
  - ▷ $M = 4$
  - ▷ $N = 8$
  - ▷ $\bar{\gamma} = 8\text{dB}$
- Varying $\hat{M}_1$ and $\hat{M}_2$ which sets $\hat{N}_1 = 8 - \hat{M}_1$ $\hat{N}_2 = 8 - \hat{M}_2$ for suppression



- Each stream configuration $(\hat{M}_1, \hat{M}_2)$ renders a rectangular region
  - ▷ Suppression: 16 different two-way regions and two degenerate cases where data is transmitted in one direction only

Achievable Rate Regions (2)

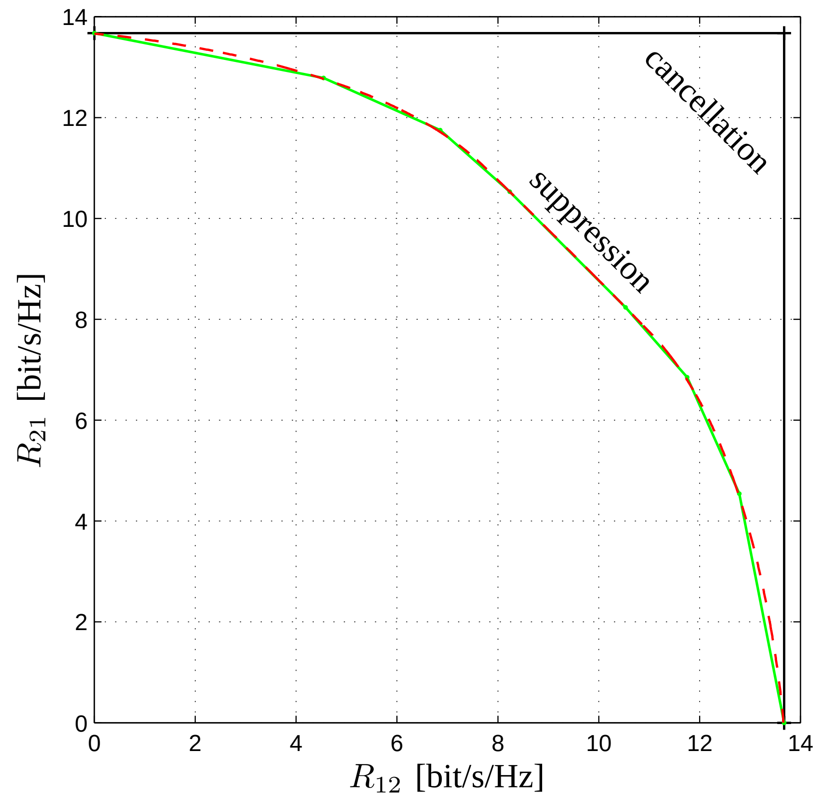
- When
  - $M = 4$
  - $N = 8$
  - $\bar{\gamma} = 8\text{dB}$
- Varying $\hat{M}_1$ and $\hat{M}_2$ which sets $\hat{N}_1 = 8 - \hat{M}_1$ $\hat{N}_2 = 8 - \hat{M}_2$ for suppression



- The complete rate region is achieved by *time sharing* between different fixed stream configurations $(\hat{M}_1, \hat{M}_2)$
  - The convex hull of the union of rectangular rate regions

Achievable Rate Regions (3)

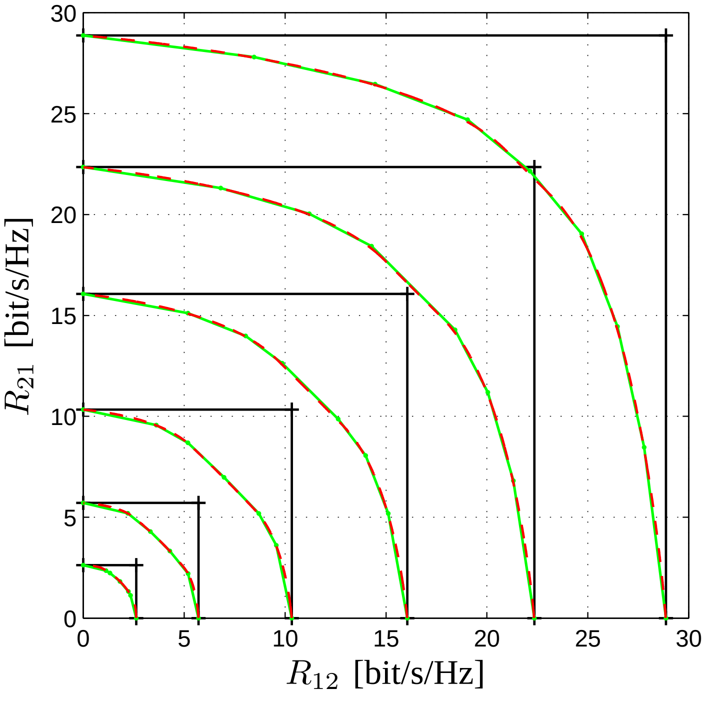
- When
  - $M = 4$
  - $N = 8$
  - $\bar{\gamma} = 8\text{dB}$
- Varying continuously $\hat{M}_1/M$ and $\hat{M}_2/M$
- Using time sharing when $R_{12} \sim R_{21}$



- Rate region projected from the asymptotic analytical results (dashed line) matches well with the finite-case simulations

Achievable Rate Regions vs. SNR

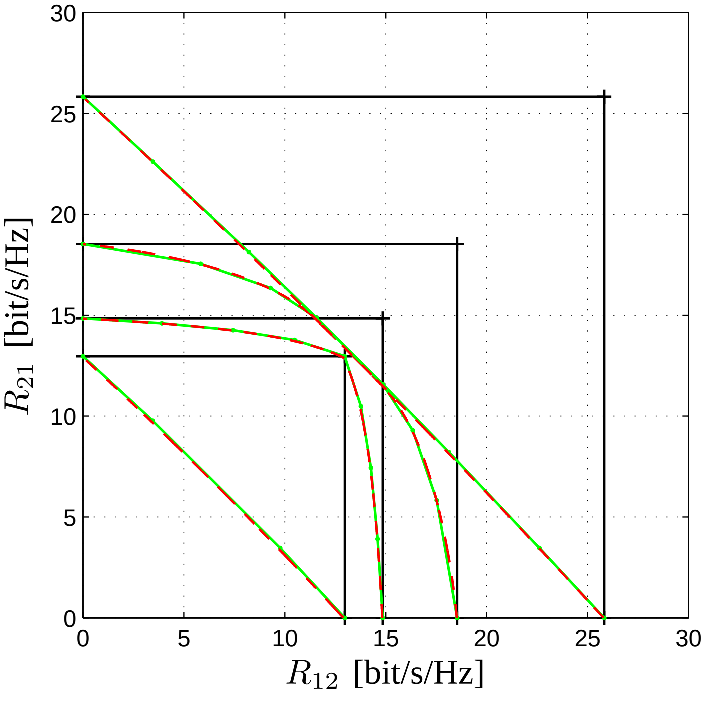
- When
  - $(M, N) = (4, 8)$
  - $$\begin{aligned} \bar{\gamma} &= 20\text{dB} \\ &= 15\text{dB} \\ &= 10\text{dB} \\ &= 5\text{dB} \\ &= 0\text{dB} \\ &= -5\text{dB} \end{aligned}$$



- The absolute rates increase with the SNR value, as expected, but otherwise it affects only slightly the shape of rate regions
- Asymmetric traffic can be supported without time sharing

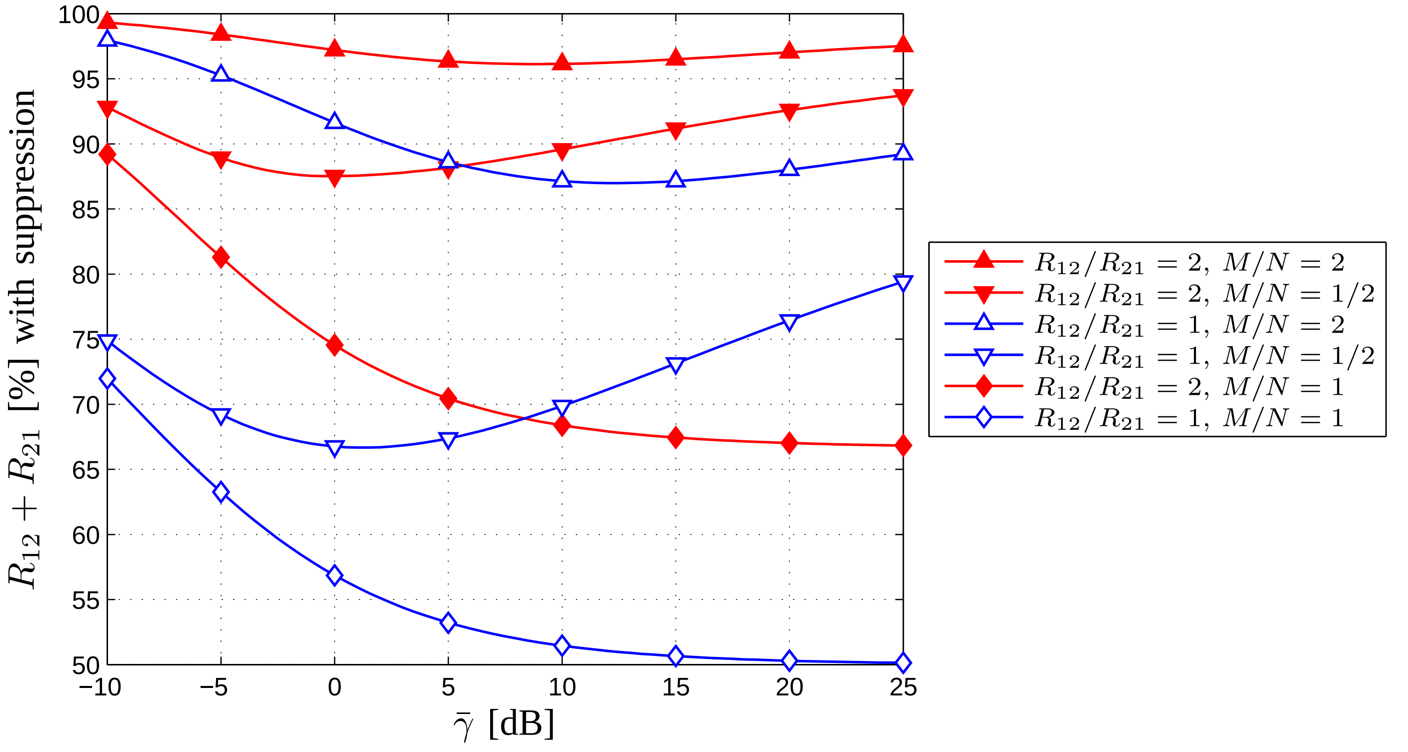
Achievable Rate Regions vs. Antenna Imbalance

- When
  - $$\begin{aligned}(M,N) &= (8,8)\\ &= (4,8)\\ &= (8,4)\\ &= (4,4)\end{aligned}$$
  - $\bar{\gamma} = 12\text{dB}$



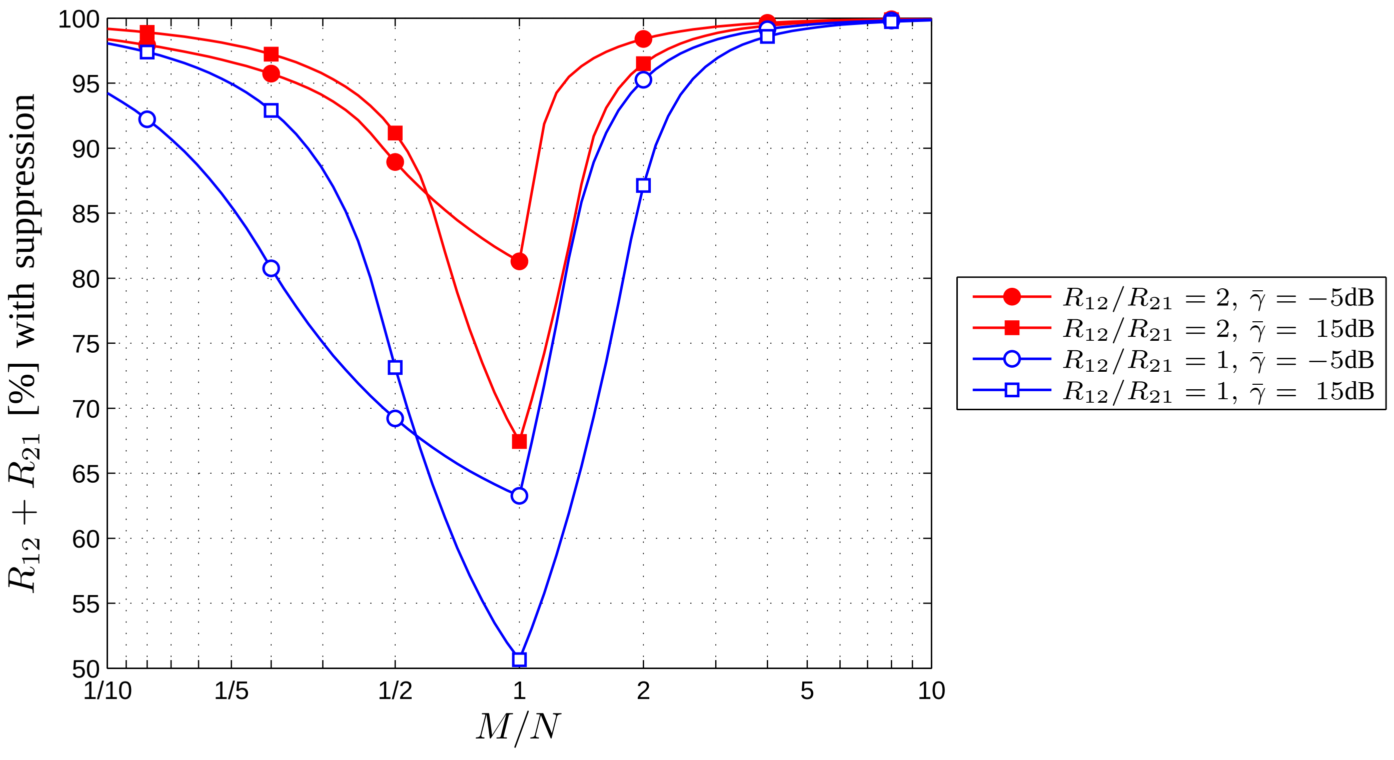
- Transmit/receive antenna imbalance (M/N) affects significantly the shape of the rate regions with spatial-domain suppression
- The rate region of suppression is always inside that of cancellation

Suppression vs. Cancellation (SNR)



- SNR defines whether the performance is limited by transmit-side multiplexing gain or receive-side array gain
- Worst case: equal number of transmit and receive antennas

Suppression vs. Cancellation (Antenna Imbalance)



- Transmit/receive antenna imbalance is a critical factor when characterizing the rate loss of suppression versus cancellation
- Having more transmit antennas than receive antennas is preferred

Conclusion

Conclusion

- Hot emerging research topic: *wireless full-duplex communication*
  - ▷ A progressive frequency-reuse concept: significantly improved spectral efficiency at the cost of self-interference
- Achievable rate regions in bidirectional full-duplex link by controlling stream configurations and time sharing
  - ▷ Comparing spatial suppression to subtractive cancellation
  - ▷ Characterizing the cost of allocating a part of spatial degrees of freedom for self-interference mitigation
- Analysis at the large-system limit based on the replica method
  - ▷ Validation by Monte Carlo simulations with small number of antennas proves that the results have also practical value
- Antenna imbalance is good for suppression especially when there are more transmit than receive antennas

Aalto University
School of Electrical Engineering