MSE-Optimal Training for Linear Time-Varying Channels

Arun Kannu and Phil Schniter

THE OHIO STATE UNIVERSITY

Mar. 21, 2005

Pilot Aided Transmission:

- Assume that the transmitter and receiver both know the channel statistics but not the channel realization.
- Pilot-aided Transmission (PAT) defined as follows.
  1. The transmitter sends an N-block including both data and pilots.
  2. The receiver estimates the channel *once* using only the pilots.
  3. The receiver attempts coherent data detection using the estimated channel matrix.
- Key observations about our definition of PAT:
  1. Iterative channel/data estimation is prohibited.
  2. PAT is actually a form of "non-coherent communication."

MMSE-PAT:

- Many authors have suggested PAT with:
  1. Wiener channel estimation at the receiver.
  2. The pilot sequence chosen to minimize the MSE of channel estimates (subject to a pilot power constraint),
  3. The pilot power chosen via some other criterion.
- This problem has been investigated for various channel classes (e.g., time-selective, frequency-selective, doubly-selective).
- However, most investigations have assumed *non-superimposed* (NSI) pilot/data patterns.

$\rightsquigarrow$ *What about MMSE-PAT with superimposed pilot/data?*

Problem Setup:

Observation:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{H}(\boldsymbol{p}+\boldsymbol{d})+\boldsymbol{v} \qquad \boldsymbol{p},\boldsymbol{d}\in\mathbb{C}^N \\ &= (\boldsymbol{P}+\boldsymbol{D})\boldsymbol{h}+\boldsymbol{v} \end{aligned}$$

Channel estimate:

$$\begin{aligned} \hat{\boldsymbol{h}} &= f(\boldsymbol{y},\boldsymbol{P}) \\ \tilde{\boldsymbol{h}} &:= \boldsymbol{h}-\hat{\boldsymbol{h}} \end{aligned}$$

Detection:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{P}\hat{\boldsymbol{h}}+\boldsymbol{P}\tilde{\boldsymbol{h}}+\boldsymbol{D}\hat{\boldsymbol{h}}+\boldsymbol{D}\tilde{\boldsymbol{h}}+\boldsymbol{v} \\ \underbrace{\boldsymbol{y}-\boldsymbol{P}\hat{\boldsymbol{h}}}_{\boldsymbol{y}_{\text{eff}}} &= \boldsymbol{D}\hat{\boldsymbol{h}}+\underbrace{\boldsymbol{P}\tilde{\boldsymbol{h}}+\boldsymbol{D}\tilde{\boldsymbol{h}}+\boldsymbol{v}}_{\boldsymbol{v}_{\text{eff}}} \\ \boldsymbol{y}_{\text{eff}} &= \hat{\boldsymbol{H}}\boldsymbol{d}+\boldsymbol{v}_{\text{eff}} \end{aligned}$$

The structures of $\boldsymbol{H},\boldsymbol{P},\boldsymbol{D}$ depend on the modulation scheme (e.g., CP-OFDM, SCCP) and the channel properties (e.g., TS, FS, DS).

Generic Conditions for MMSE-PAT:

Say

$$\begin{aligned} \boldsymbol{y} &= (\boldsymbol{P}+\boldsymbol{D})\boldsymbol{h}+\boldsymbol{v} \\ \boldsymbol{h} &= \boldsymbol{U}\boldsymbol{\lambda} \end{aligned}$$

where

$\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_M$, $\mathrm{E}[\boldsymbol{\lambda}] = \boldsymbol{0}$, $\mathrm{E}[\boldsymbol{\lambda}\boldsymbol{\lambda}^H] = \mathrm{diag}(\sigma^2_{\lambda_0}, \dots, \sigma^2_{\lambda_{M-1}}) \geq 0$,
$\mathrm{E}[\boldsymbol{D}] = \boldsymbol{0}$, $\mathrm{E}[\boldsymbol{v}] = \boldsymbol{0}$, $\mathrm{E}[\boldsymbol{v}\boldsymbol{v}^H] = \sigma_v^2\boldsymbol{I}$, uncorrelated $\{\boldsymbol{D}, \boldsymbol{\lambda}, \boldsymbol{v}\}$,
and $\|\boldsymbol{p}\|^2 \leq E_p$.

Can show that $\mathrm{E}\{\|\tilde{\boldsymbol{h}}\|^2\}$ is minimized if and only if

$$\begin{aligned} \forall \boldsymbol{D},\ (\boldsymbol{P}\boldsymbol{U})^H\boldsymbol{D}\boldsymbol{U} &= \boldsymbol{0} \qquad (1) \\ (\boldsymbol{P}\boldsymbol{U})^H\boldsymbol{P}\boldsymbol{U} &= \mathrm{diag}(\alpha_0, \dots, \alpha_{M-1}) \qquad (2) \end{aligned}$$

where the "water-filling" coefs $\{\alpha_m\}$ depend on $\{\sigma^2_{\lambda_m}\}$, σ_v^2, and E_p.

Generic Conditions for MMSE-PAT (cont.):

Interpretation of (1)-(2):

1. $\forall \boldsymbol{D},\ (\boldsymbol{P}\boldsymbol{U})^H \boldsymbol{D}\boldsymbol{U} \ = \ \boldsymbol{0}$:
   Pilot/data subspaces remain orthogonal at channel output.
2. $(\boldsymbol{P}\boldsymbol{U})^H \boldsymbol{P}\boldsymbol{U} \ = \ \mathrm{diag}(\alpha_0, \ldots, \alpha_{M-1})$:
   Pilot excitation proportional to strength of channel mode.

Implication:

Pilot/data superposition is tolerated as long as pilot/data can be separated by a linear receiver,

a consequence of our not allowing iterative channel estimation.

Application: The Doubly-Dispersive Channel:

- Consider a SISO, WSSUS, Rayleigh fading channel.
- Assume N_t ISI coefficients, i.i.d. with uniform Doppler spectrum over $[-f_{\mathsf{d}}, f_{\mathsf{d}})$ Hz, approximated by a basis expansion model:

$$h(n,\ell) \;=\; \frac{1}{\sqrt{N}} \sum_{k=-(N_f-1)/2}^{(N_f-1)/2} \lambda(k,\ell) e^{j\frac{2\pi}{N}kn}, \text{ for } 0 \le n < N.$$

where $N_f := \lfloor 2 f_{\mathsf{d}} T_s N \rfloor + 1$.

- For $\boldsymbol{y} = (\boldsymbol{P}+\boldsymbol{D})\boldsymbol{h} + \boldsymbol{v}$ with length-(N_t-1) CP, this implies

$$\begin{aligned} \boldsymbol{h} &= \boldsymbol{U}\boldsymbol{\lambda} \\ \boldsymbol{U} &= \boldsymbol{I}_{N_t} \otimes \boldsymbol{F}_N^*\big(:, -\tfrac{N_f-1}{2} : \tfrac{N_f-1}{2}\big) \\ \boldsymbol{\lambda} &\sim \mathcal{CN}\big(\boldsymbol{0}, \tfrac{N}{N_f N_t}\boldsymbol{I}_{N_f N_t}\big), \end{aligned}$$

where $\boldsymbol{F}_N$ is the unitary N-DFT matrix. Note $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_{N_f N_t}$.

DS-Channel Conditions for MMSE-PAT:

With this N-block DS model, the necessary and sufficient conditions for MMSE-PAT become: $\forall k \in \mathcal{N}_t, \forall m \in \mathcal{N}_f$,

$$\begin{aligned} E_p \delta(k)\delta(m) &= \textstyle\sum_{n=0}^{N-1} p(n)p^*(n-k)e^{-j\frac{2\pi}{N}mn} && (3) \\ 0 &= \textstyle\sum_{n=0}^{N-1} d(n)p^*(n-k)e^{-j\frac{2\pi}{N}mn} && (4) \\ \mathcal{N}_t &:= \{-N_t+1,\ldots,N_t-1\} \\ \mathcal{N}_f &:= \{-N_f+1,\ldots,N_f-1\}. \end{aligned}$$

To construct such a pilot/data pattern,

1. Find pilot sequence $\boldsymbol{p}$ satisfying (3).
2. Write (4) as $\boldsymbol{W_p d} = \boldsymbol{0}$ and set $\boxed{\boldsymbol{d} = \boldsymbol{B s}}$, where the N_s columns of $\boldsymbol{B}$ form an ON basis for null$(\boldsymbol{W_p})$.

We call this the $(\boldsymbol{p}, \boldsymbol{B})$ MMSE-PAT pattern.

The "Data Dimension" N_s:

- In MMSE-PAT, the data is represented by the N_s symbols in $\boldsymbol{s}$.
- It is relatively easy to bound the data dimension N_s as

$$N - (2N_f - 1)(2N_t - 1) \;\le\; N_s \;\le\; N - N_f N_t.$$

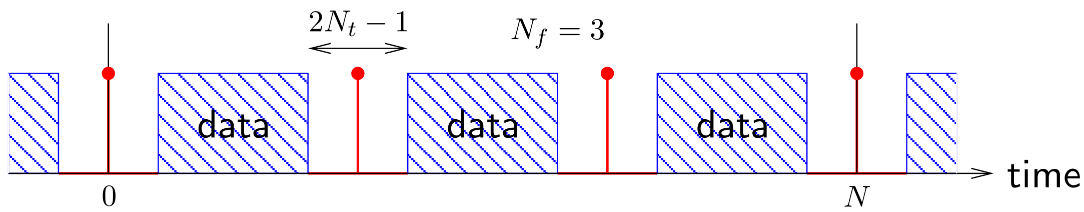
- A more careful analysis, however, reveals the strict upper bound

$$\boxed{N_s \;<\; N - N_f N_t}$$

when $N_t > 1$ and $N_f > 1$ (i.e., the strictly-DS case).

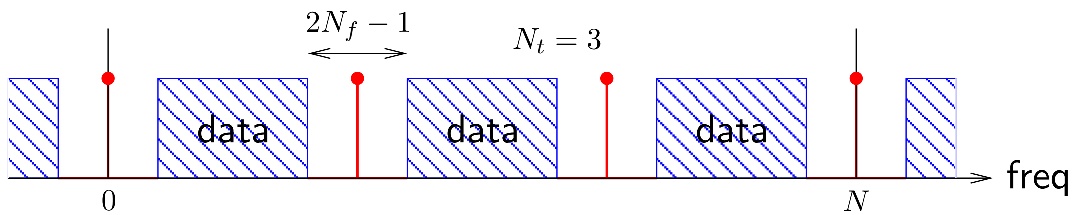
Example MMSE-PAT Pilot Patterns:

1. Time-domain Kronecker Delta (TDKD): $\boxed{N_s = N - N_f(2N_t - 1)}$.



(This non-superimposed PAT was suggested by Ma/Giannakis/Ohno.)

2. Freq-domain Kronecker Delta (FDKD): $\boxed{N_s = N - N_t(2N_f - 1)}$.



3. Orthogonal Chirps: $\boxed{N_s = N - 2N_tN_f + 1}$.

$$
\begin{aligned}
p(n) &= \frac{E_p}{N} e^{j\frac{2\pi}{N}\frac{N_f}{2}n^2} \\
b_k(n) &= \frac{\sqrt{N}}{E_p} p(n) e^{j\frac{2\pi}{N}(k+N_fN_t)n}, \quad 0 \le k < N_s
\end{aligned}
$$

Capacity of $(\boldsymbol{p}, \boldsymbol{B})$ MMSE-PAT:

Say $\|\boldsymbol{p}\|^2 \leq E_p$, $\mathrm{E}[\|\boldsymbol{s}\|^2] \leq E_s$, and define $\sigma_s^2 := \frac{E_s}{N_s}$, $\sigma_p^2 := \frac{E_p}{N_t N_f}$.

Then

$$\begin{aligned} \underline{C}_{\text{mmse-pat}} &\leq C_{\text{mmse-pat}} \leq \overline{C}_{\text{mmse-pat}} \\ \underline{C}_{\text{mmse-pat}} &:= \frac{1}{N} \mathrm{E} \log \det\big(\boldsymbol{I} + \rho_l \boldsymbol{B}^H \boldsymbol{H}^H \boldsymbol{H} \boldsymbol{B}\big) \\ \overline{C}_{\text{mmse-pat}} &:= \frac{1}{N} \mathrm{E} \log \det\big(\boldsymbol{I} + \rho_u \boldsymbol{B}^H \boldsymbol{H}^H \boldsymbol{H} \boldsymbol{B}\big) \end{aligned}$$

where

$$\rho_l := \frac{\sigma_s^2}{\sigma_v^2} \left(\frac{\sigma_p^2}{\sigma_p^2 + \sigma_s^2 + \sigma_v^2} \right) \text{ and } \rho_u := \frac{\sigma_s^2}{\sigma_v^2}.$$

(For the lower bound, we assumed the worst-case $\tilde{\boldsymbol{h}}$ via independent CWGN, and for the upper bound the best-case $\tilde{\boldsymbol{h}}$ via $\tilde{\boldsymbol{h}} = \boldsymbol{0}$.)

Power Allocation that Maximizes $\underline{C}_{\text{mmse-pat}}$:

Say $\alpha \in (0,1)$ is used to allocate the total power $E_t = E_s + E_p$:

$$E_s = \alpha E_t \text{ and } E_p = (1-\alpha)E_t.$$

Then $\underline{C}_{\text{mmse-pat}}$ maximized by

$$\begin{aligned}
\alpha &= \begin{cases} \beta - \sqrt{\beta^2 - \beta} & N_s \neq N_t N_f \\ \frac{1}{2} & N_s = N_t N_f \end{cases} \\
\beta &:= \frac{1 + \frac{N_f N_t}{\rho N}}{1 - \frac{N_f N_t}{N_s}} \\
\rho &:= \frac{E_t}{N \sigma_v^2}.
\end{aligned}$$

High-SNR Capacity of MMSE-PAT:

- With the $\underline{C}_{\textsf{mmse-pat}}$-maximizing power allocation,

$$\boxed{C_{\textsf{mmse-pat}}(\rho) \;=\; \frac{N_s}{N}\log(\rho) + O(1), \quad \text{as } \rho \to \infty}$$

- Recall that N_s differed among the different MMSE-PAT examples.
- Note that, when $N_t > N_f$:
  - FDKD-PAT dominates TDKD-PAT and Chirp-PAT.
  - Superimposed PAT has advantages over non-superimposed PAT.

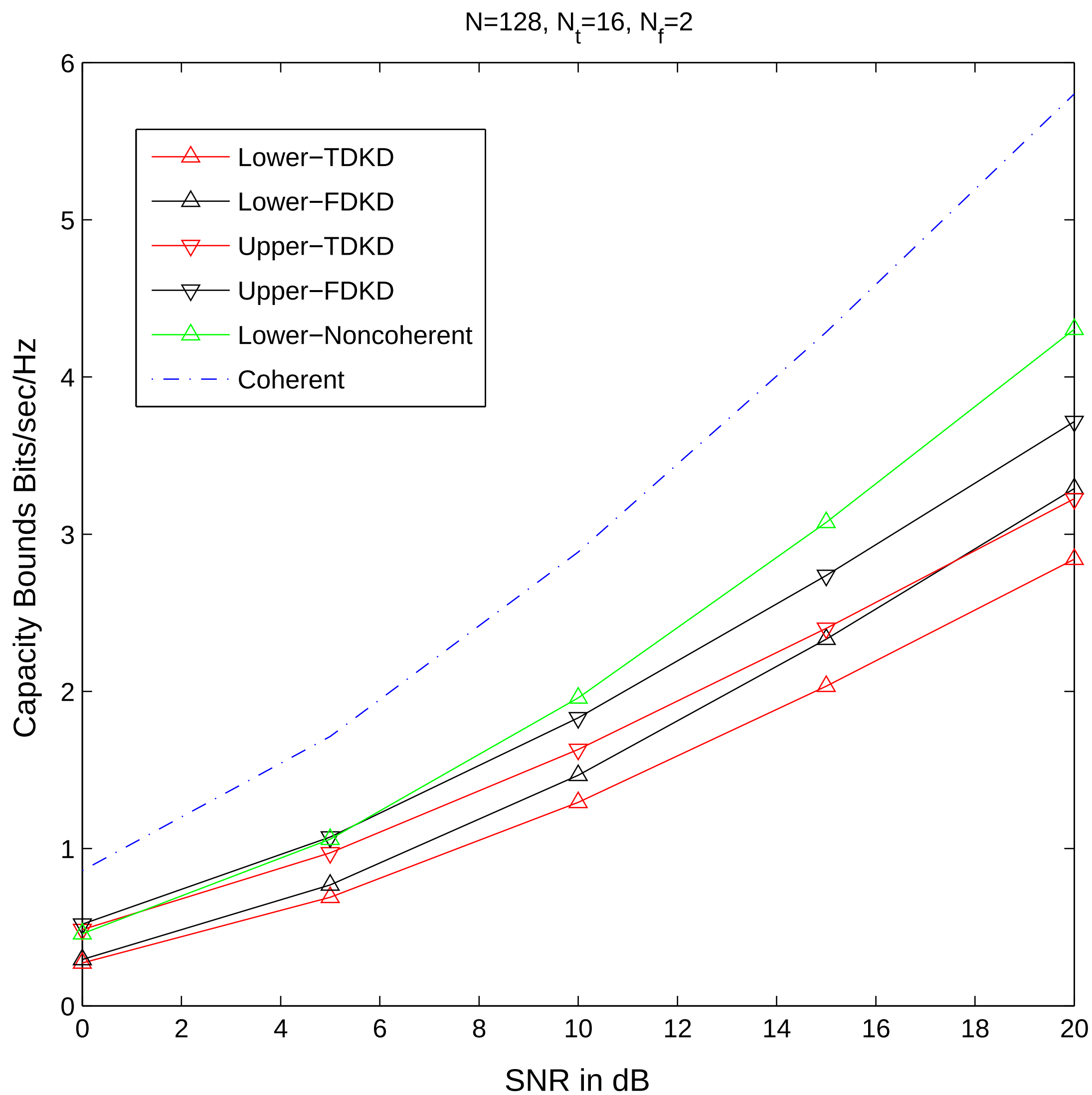
Numerical Example:

Example:

$$f_c = 12\,\text{GHz},$$
$$T_s^{-1} = 2\,\text{MHz},$$
$$v = 3 \times 133\text{kmh},$$
$$T_{\text{delay}} = 8\,\mu\text{s},$$
$$N = 128.$$

Yields:

$$N_t = 16,$$
$$N_f = 2.$$



High-SNR Capacity – Summary of Known Results:

coherent: $\log(\rho) + O(1)$

noncoherent TS: $\dfrac{N - N_f}{N} \log(\rho) + O(1)$

noncoherent FS: $\dfrac{N - N_t}{N} \log(\rho) + O(1)$

noncoherent DS: $\dfrac{N - N_t N_f}{N} \log(\rho) + O(1)$

NSI-PAT TS: $\dfrac{N - N_f}{N} \log(\rho) + O(1)$

NSI-PAT FS: $\dfrac{N - N_t}{N} \log(\rho) + O(1)$

NSI-PAT DS: $\dfrac{N - N_f(2N_t - 1)}{N} \log(\rho) + O(1)$

MMSE-PAT: $\dfrac{N_s}{N} \log(\rho) + O(1)$

On the Non-Optimality of MMSE-PAT:

- Note that for TS and FS channels, $C_{\text{mmse-pat}}(\rho)$ achieves the same slope as $C_{\text{ts}}(\rho)$ and $C_{\text{fs}}(\rho)$ as $\rho \to \infty$.

- But, for DS channels (i.e., $N_f > 1$ and $N_t > 1$) as $\rho \to \infty$,

$$\begin{aligned} C_{\text{ds}}(\rho) &= \frac{N - N_t N_f}{N} \log(\rho) + O(1), \\ C_{\text{mmse-pat}}(\rho) &= \frac{N_s}{N} \log(\rho) + O(1) \quad \text{for} \quad N_s < N - N_t N_f, \end{aligned}$$

  and thus MMSE-PAT is strictly suboptimal.

- This motivates "non-PAT" schemes, e.g., schemes based on iterative channel/data estimation.

Summary:

- Derived nec/suff conditions for MMSE-PAT design in LTV channels.
- Derived nec/suff conditions for MMSE-PAT design in DS channels, yielding novel MMSE-PAT schemes.
- Established bounds on the capacity of MMSE-PAT over DS chans.
- Suggested data/pilot power allocation for MMSE-PAT via $\underline{C}_{\text{mmse-pat}}$ maximization.
- Showed advantages of superimposed over non-superimposed MMSE-PAT when time-spreading dominates frequency-spreading.
- Established high-SNR noncoherent capacity of the DS channel.
- Showed that MMSE-PAT is strictly suboptimal in DS channels.