Algorithms for Wireless Communications II

Andreas Burg

Institute for Integrated Signal Processing Systems, RWTH Aachen

Telecommunications Circuits Laboratory, EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

OFDM Channel Estimation

Notation Conventions

Type	Notation	Example
Scalars	Lower case, italic	a, n, b, x, y
Column vectors	Lower case, bold	$\mathbf{t}, \boldsymbol{\mathscr{t}}$
Matrices	Upper case, bold	$\mathbf{T}, \boldsymbol{\mathcal{T}}$
Time domain signals	Regular font	$\mathbf{H}, H, \mathbf{h}, h$
Frequency domain signals	Script font	$\boldsymbol{\mathscr{h}}, \mathscr{h}, \boldsymbol{\mathcal{H}}, \mathcal{H}, \boldsymbol{\mathcal{T}}, \mathcal{T}, \boldsymbol{\mathscr{t}}, \mathscr{t}$

We distinguish between

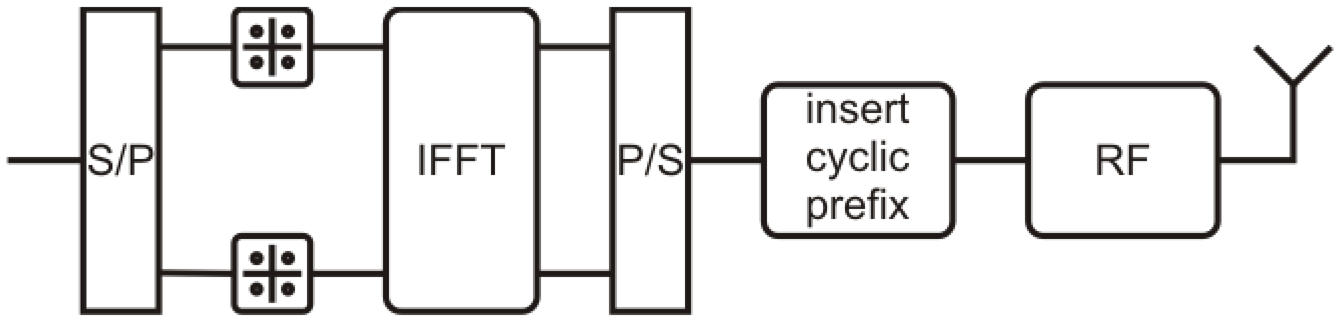
- Scalars and matrices or vectors by using regular and bold upper/lower case letters
- Time and frequency domain representation using regular and script font

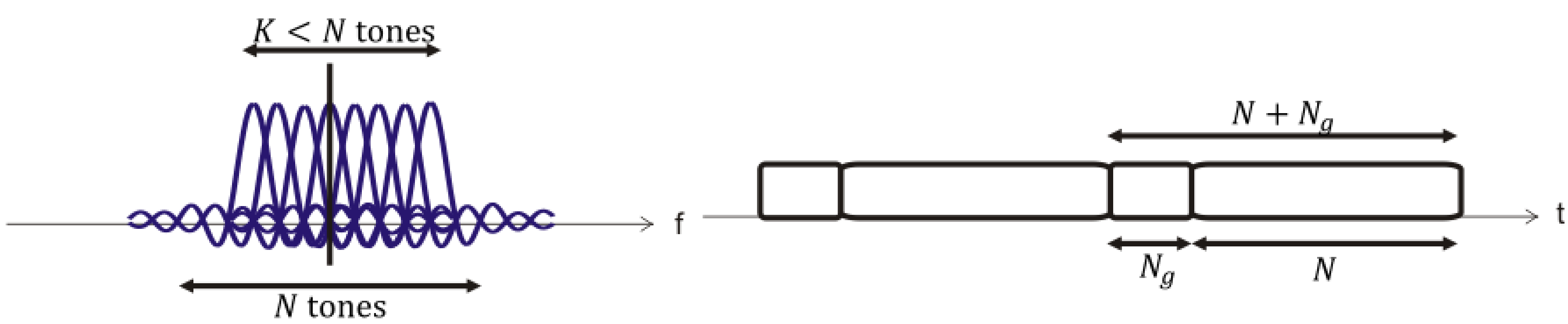
We carefully distinguish between a variable $\mathscr{h}$ and its estimate $\hat{\mathscr{h}}$ (denoted by adding a *hat*)

OFDM Reminder (1/2)

Wideband transmission using many (orthogonal) narrow-band channels

- Transmitted signal is divided into OFDM symbols, each comprising N tones
- Each tone is modulated separately with any modulation scheme
- IDFT or better IFFT generates time-domain signal
- A cyclic prefix is prepended to each OFDM symbol





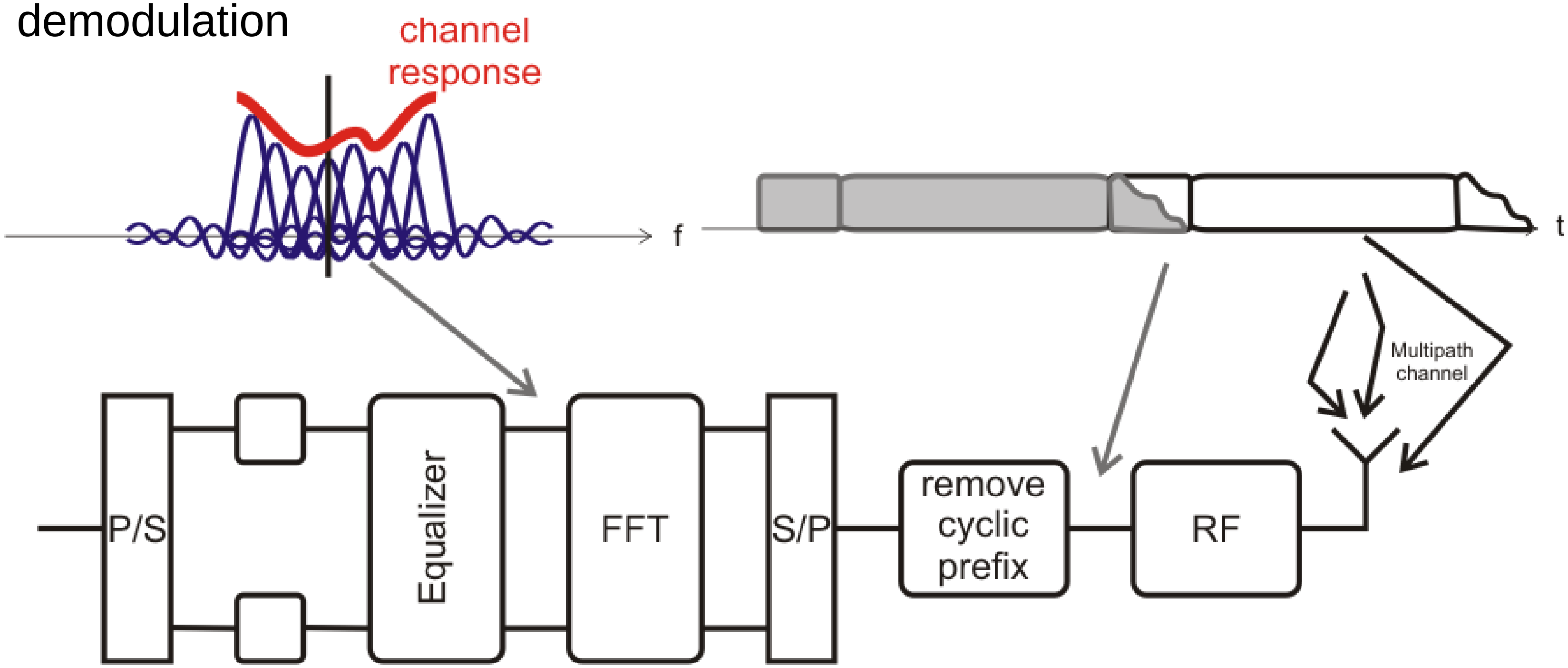
OFDM Reminder (2/2)

Transmission over multipath (frequency selective) channel

- Each tone experiences a different phase and attenuation
- Cyclic prefix captures ISI from previous symbol

Receiver

- Removes the cyclic prefix and transforms back into frequency domain (DFT or FFT)
- Per-tone equalization to undo the impact of the channel on each tone before demodulation



OFDM System Model

Recall the OFDM system model

$$\boldsymbol{y} = \mathbf{F}\left(\mathbf{H}\left(\mathbf{F}^{-1}\boldsymbol{s}\right) + \mathbf{n}\right) = \underbrace{\mathbf{F}\mathbf{H}\mathbf{F}^{-1}}_{diag(\mathbf{Fh})=\boldsymbol{h}} \boldsymbol{s} + \mathbf{F}\mathbf{n}$$

$\mathbf{F}^{-1}\boldsymbol{s}$: Frequency domain modulated data signal after IFFT

$\mathbf{H}$: Circulant matrix with time-domain channel coefficients

$\mathbf{F}$: FFT matrix

$\mathbf{n}$: i.i.d. Gaussian noise samples $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \mathbf{I}\sigma_n^2$

Frequency domain representation

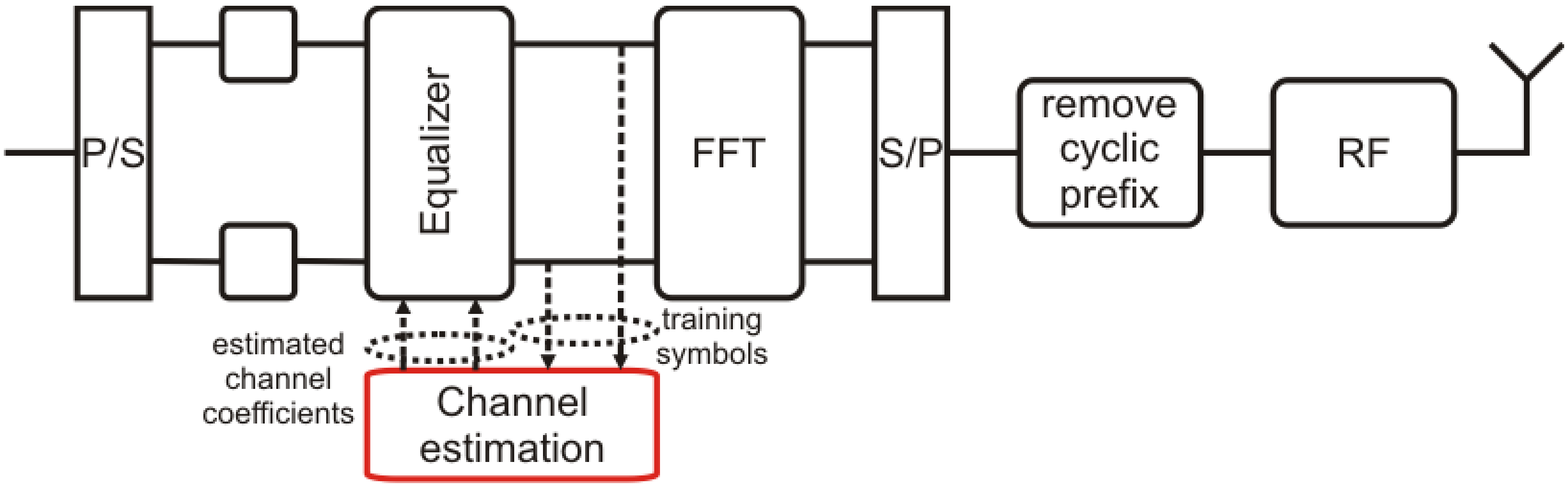
$$\boldsymbol{y} = \boldsymbol{\mathcal{H}}\boldsymbol{s} + \boldsymbol{n} \quad \text{with} \quad \mathcal{H} = \begin{bmatrix} h_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & h_{N-1} \end{bmatrix} = diag(\mathbf{Fh})$$

$$y_f = h_f s_f + n_f$$

- Note: noise properties remain unchanged $\mathbb{E}\{\boldsymbol{n}\boldsymbol{n}^H\} = \sigma_n^2\mathbf{I}$ since $\mathbf{F}$ is an orthonormal transform

Channel Estimation

Task: obtain knowledge (estimate) of the channel coefficients in the frequency domain h_f based on the received signal y_f



Channel Estimation

Estimation strategies

- Training/Pilot based: Transmitter sends known data (training/pilot symbols) so that h_f and n_f are the only unknown parameters
  - Overhead: training/pilot symbols do not carry data!!

- Data aided: channel estimation based on feedback of detected data

- Blind: Estimation only based on statistical properties

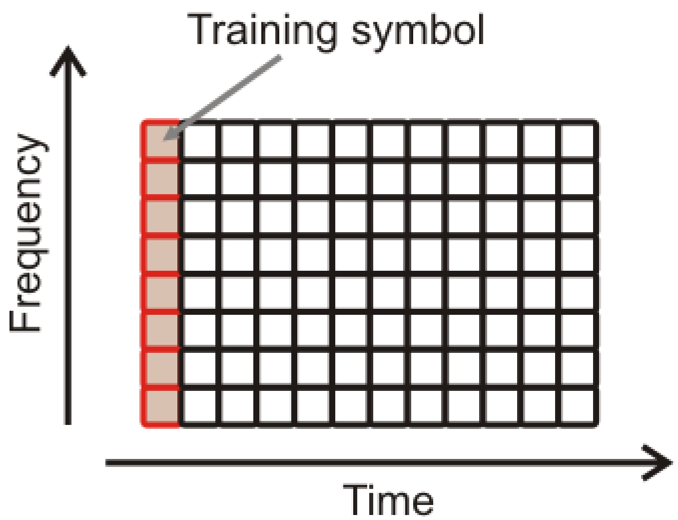
Here, we consider only training/pilot based channel estimation

- Most frequently used approach (very robust)
- Overhead is usually low when channels do not change to quickly

Common Training/Pilot Patterns

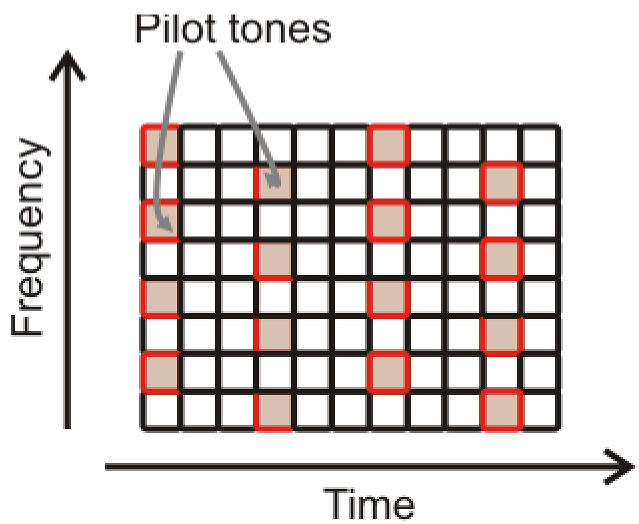
OFDM-training structures differ by how training/pilot symbols are distributed across time and frequency

Block type training (e.g., 802.11a/g/n)



- Dedicated OFDM training symbol
- Used in systems with relatively static channel conditions
- Packet based systems (burst traffic)

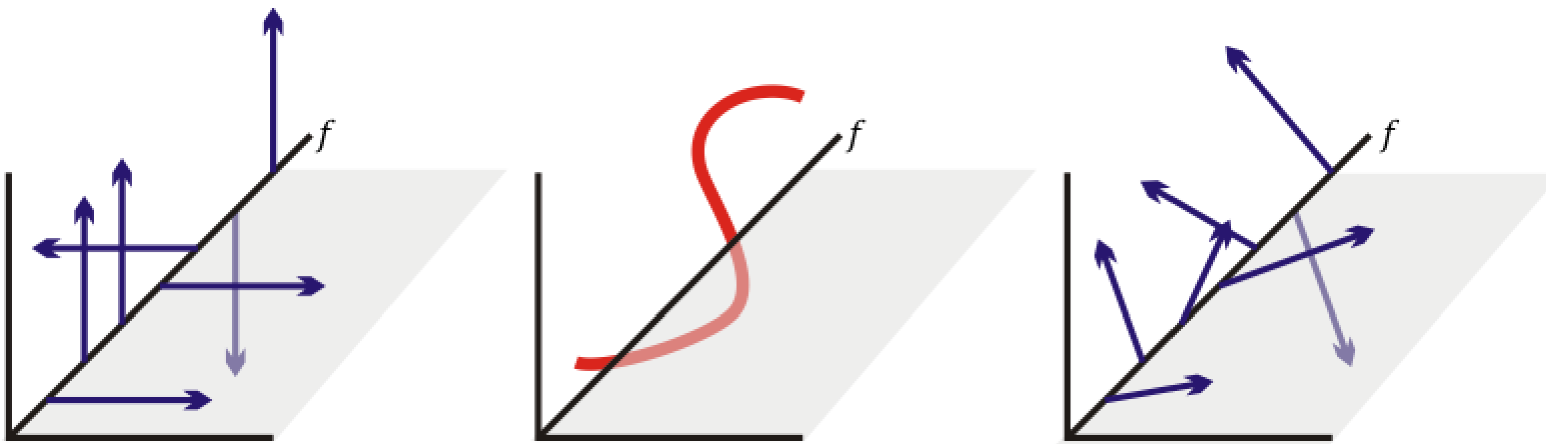
Comb type training (e.g., DVB-T, 3GPP LTE)



- Pilot tones interleaved with data
- Pilot location varies to cover all tones
- Used in rapidly time-varying channels
- Streaming (continuous transmission)

Input Output Relationship for Channel Estimation

Consider the transmission of a single training symbol t



Training symbol t_f

Channel frequency response h_f

Received symbol $y_f = h_f t_f + n_f$

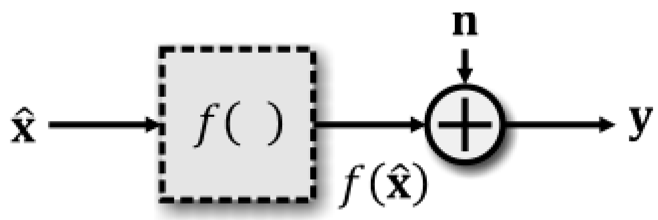
- Two options to formulate the input-output relationship

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} h_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & h_{N-1} \end{bmatrix} \begin{bmatrix} t_0 \\ \vdots \\ t_{N-1} \end{bmatrix} + \begin{bmatrix} n_0 \\ \vdots \\ n_{N-1} \end{bmatrix} = \boxed{\begin{bmatrix} t_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & t_{N-1} \end{bmatrix} \begin{bmatrix} h_0 \\ \vdots \\ h_{N-1} \end{bmatrix} + \begin{bmatrix} n_0 \\ \vdots \\ n_{N-1} \end{bmatrix}}$$

For our purposes (estimate t) the 2nd expression is more appropriate

Estimation Theory Basics

Problem statement: consider the following simplified system model



- Observation $\boldsymbol{y}$ is some function of an unknown parameter $\hat{\mathrm{x}}$ distorted by noise
  - Often, $f(\mathbf{x})$ is a linear function of $\mathbf{x}$
  - Often, the observation is distorted by additive Gaussian noise

- Objective: estimate $\mathbf{x}$ from $\mathbf{y}$ using some cost metric

- We may consider different types of prior knowledge on the variable to be estimated
  - A priori probabilities
  - Knowledge on 1st and 2nd order statistics
- OFDM channel estimation:
  - Observation: received training symbol/tones
  - $f(\quad)$: Impact of the channel on the (known) transmitted training symbol
  - Noise: thermal noise, additive Gaussian
  - A priori knowledge: **what can we say about the channel ??**

Estimation Theory Basics

Maximum likelihood estimation: Maximize likelihood of the observation under hypothesis $\hat{x}$

$$\hat{\mathbf{x}} = \underset{\hat{\mathbf{x}}}{\operatorname{argmax}}\ p(\mathbf{y}|\hat{\mathbf{x}})$$

Least squares estimation: Minimize distance between observation and noise free observation under hypothesis $\hat{x}$

$$\hat{\mathbf{x}} = \underset{\hat{\mathbf{x}}}{\operatorname{argmin}}\ |\mathbf{y} - f(\hat{\mathbf{x}})|^2$$

- Linear least squares approach: $\hat{\mathbf{x}} = \mathbf{W}_{LS}\mathbf{y}$
- No assumptions on 2nd order statistics
- For additive Gaussian noise: equivalent to ML estimation

Minimum mean squared error (MMSE) estimation

$$\hat{\mathbf{x}} = \underset{\hat{\mathbf{x}}}{\operatorname{argmin}}\, \mathbb{E}\{|\mathbf{x} - \hat{\mathbf{x}}|^2\}$$

- Linear least squares approach: $\hat{\mathbf{x}} = \mathbf{W}_{\mathrm{MMSE}}\mathbf{y}$

Per Tone (Linear) Least Squares Method

Linear least squares estimation: obtain the estimate $\hat{\boldsymbol{h}}$ of $\boldsymbol{h}$ according to

$$\hat{\boldsymbol{h}} = \underset{\hat{\boldsymbol{h}}}{\operatorname{argmin}} \left\{ \left| \boldsymbol{y} - \boldsymbol{\mathcal{T}} \hat{\boldsymbol{h}} \right|^2 \right\} \text{ with } \boldsymbol{\mathcal{T}} = \begin{bmatrix} t_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & t_{N-1} \end{bmatrix}$$

- Minimize Euclidean distance between $\boldsymbol{y}$ and $\boldsymbol{\mathcal{T}}\hat{\boldsymbol{h}}$
- Gaussian noise: least squares solution corresponds to the maximum likelihood solution

No assumptions are made on correlation of the individual tones:

- Treat all tones as if they had an independent channel coefficient
- Since $\boldsymbol{\mathcal{T}}$ is diagonal, tones can be estimated independently (locally optimal solutions leads to a global optimal solution)

$$\hat{\boldsymbol{h}} = \boldsymbol{\mathcal{T}}^{-\mathbf{1}} \boldsymbol{y} \text{ or } \hat{h}_f = \frac{1}{t_f} y_f$$

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Estimation Error for Per Tone Least Squares Estimation

- Define the per tone estimation error as

$$\Delta h_f = h_f - \hat{h}_f$$

We are interested in the mean squared error

- Mean: taken over many noise realizations (average error we expect over many independent channel estimations)

$$\sigma_{h_f}^{2} = \mathbb{E}\left\{\left|\Delta h_f\right|^{2}\right\} = \mathbb{E}\left\{\left|h_f - \frac{1}{t_f} y_f\right|^{2}\right\}$$

- Substitute the signal model $y_f = h_f t_f + n_f$

$$\sigma_{h_f}^{2} = \mathbb{E}\left\{\left|h_f - \frac{1}{t_f}\left(h_f t_f + n_f\right)\right|^{2}\right\} = \frac{\mathbb{E}\left\{\left|n_f\right|^{2}\right\}}{\left|t_f\right|^{2}} = \frac{\sigma_{n}^{2}}{\left|t_f\right|^{2}}$$

Per Tone Least Squares: remember from part 1 of the course

Channel Estimation

Quality of the channel estimate

$$\frac{r_k}{t} = h + w_k/t$$

Noise component on the channel est.

$$\frac{\mathcal{E}\{|h|^2\}}{\mathcal{E}\{|w_k/t|^2\}} = \frac{\mathcal{E}\{|h|^2\}}{\sigma_n^2}$$

Improving the channel estimate by averaging

- If the channel is static, we can use multiple training symbols and average

$$\hat{h} = \frac{1}{N}\sum_{k=1}^{N}\frac{r_k}{t} = \frac{1}{N}\sum_{k=1}^{N} h + \frac{1}{N}\sum_{k=1}^{N}\frac{w_k}{t}$$

N-fold averaging reduces noise on the channel estimate by a factor of N

If the channel is time varying, averaging is replaced by a low-pass filter

- Filter bandwidth determined by the Doppler frequency
- Remove the frequency components outside the Doppler bandwidth (noise!!)

RWTHAACHEN UNIVERSITY

EPFL

Design of the Training Symbol

Quality metric: average estimation MSE (across tones)

$$\sigma_{\boldsymbol{h}}^2 = \frac{\sigma_{\boldsymbol{n}}^2}{N} \sum \frac{1}{|t_f|^2}$$

- Average power constraint: $\sum |t_f|^2 = N P_t$ with P_t: transmit power during training

Distribution of power across tones has an impact on the average channel estimation error

Optimum power allocation

- Distribute power equally to all tones of the training symbol

$$|t_f|^2 \equiv P_t$$

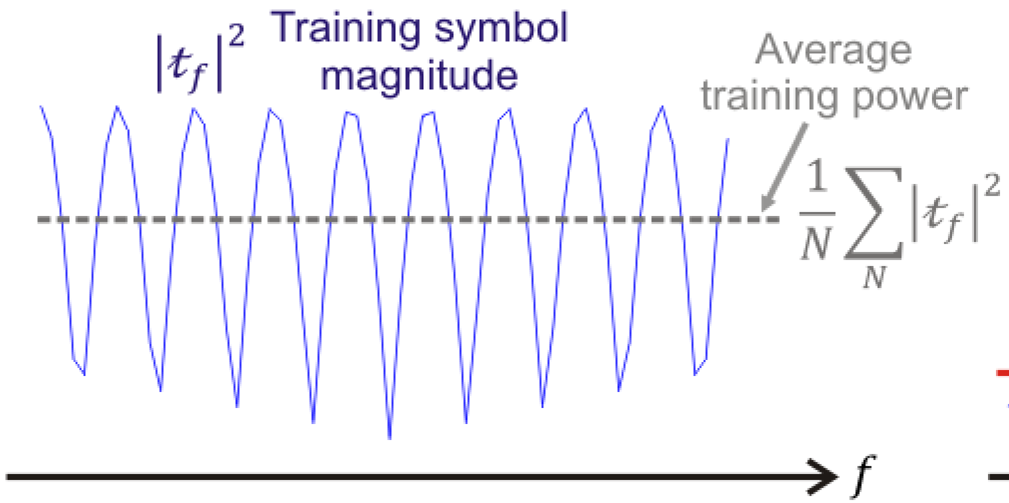
Channel estimation MSE: $\sigma_{\boldsymbol{h}}^2 = \sigma_{h_f}^2 = \sigma_{\boldsymbol{n}}^2 / P_t$

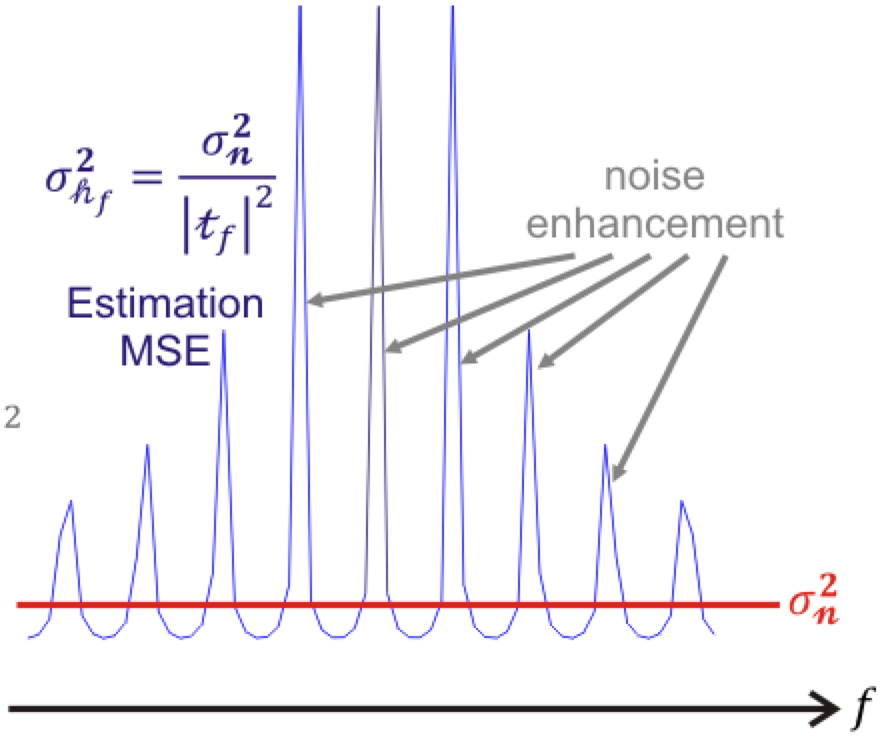
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Implication for the Design of the Training Sequence

Example: training sequence with uneven power allocation across tones

- System constraints limit only the average (across tones) transmit power
- **Example**: training sequence designed in time domain (e.g., for good PAPR)



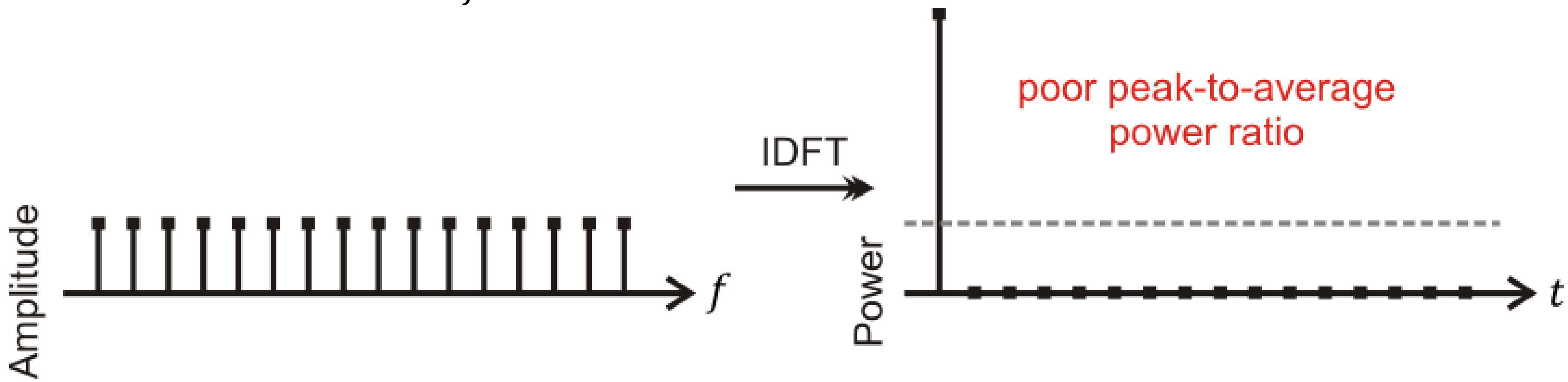


- MSE depends on the power allocated to each pilot tone
- Unequal power allocation -> unequal distribution of the MSE across tones

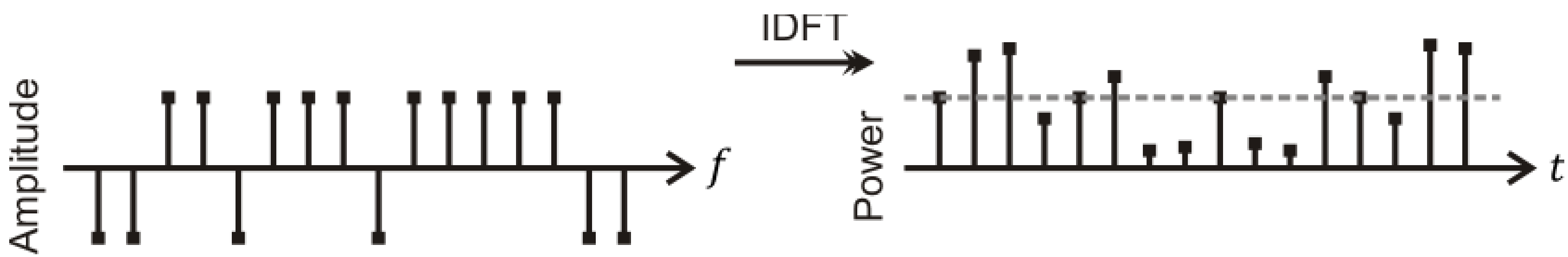
Distribute power equally across tones!

Design of the Training Symbol

Straightforward design: $t_f \equiv 1$



Solution: choose the phase of t_f to achieve an approximately constant envelope (PAPR$\approx$ 1)



- Real-world example: IEEE 802.11a long training symbol (tones -26 to +26)

t_f =++--++-+-++++++--++-+-++++0+--++-+-+-----++--+-+-++++

Impact of Channel Estimation Error on BER

- Channel estimation error $\Delta h_f = h_f - \hat{h}_f$
- Input-output relationship $y_f = h_f s_f + n_f$ (data symbols)

Express the real channel h_f as function of its estimate $\hat{h}_f$ and the error Δh_f

$$h_f = \hat{h}_f + \Delta h_f$$

$$y_f = \left(\hat{h}_f + \Delta h_f\right) s_f + n_f$$

Treat all unknown terms as noise

$$\breve{n}_f = \Delta h_f s_f + n_f$$

$$\sigma_{\breve{n}}^2 = \mathbb{E}\left\{\left|\breve{n}_f\right|^2\right\} = \underbrace{\mathbb{E}\left\{\left|\Delta h_f\right|^2\right\}}_{\sigma_n^2/P_t} \underbrace{\mathbb{E}\left\{\left|s_f\right|^2\right\}}_{P_s} + \underbrace{\mathbb{E}\left\{\left|n_f\right|^2\right\}}_{\sigma_n^2} = \left(\frac{P_s}{P_t} + 1\right) \sigma_n^2$$

Impact of Channel Estimation Error on BER

- Assume training and data have equal power (usually the case) : $P_s = P_t$

$$\sigma_{\breve{n}}^2 = 2\sigma_n^2$$

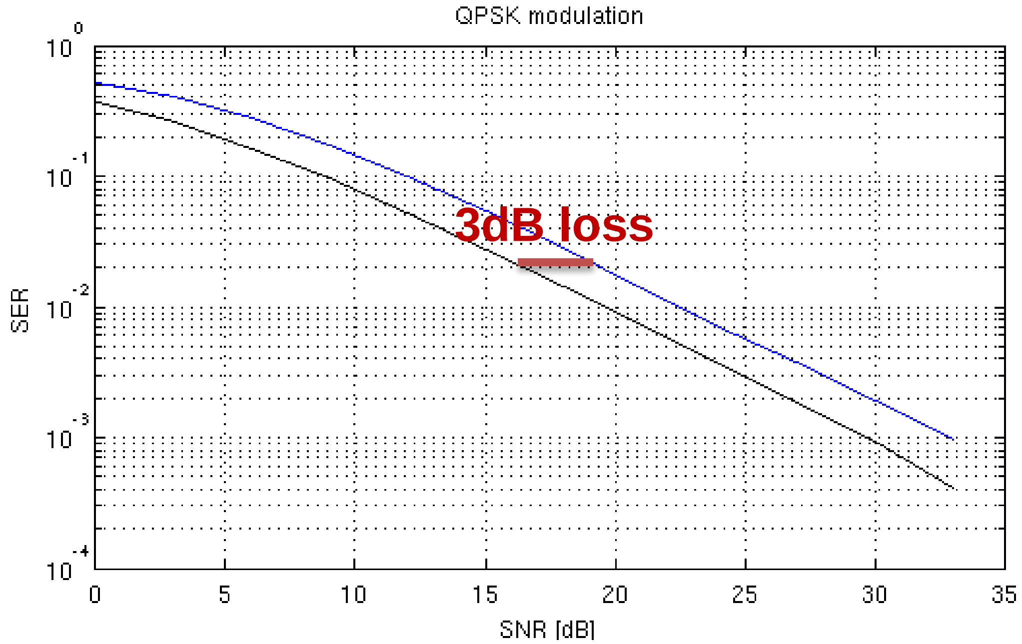
Channel estimation doubles the noise on the signal

Impact on bit error rate

- Approximation: ignore the fact that $\Delta h_f s_f$ is *data dependent* and *not necessarily, but often Gaussian*

Example

- *QPSK modulation*
- *With perfect channel state information (CSI): blue*
- *With estimated channel: black*



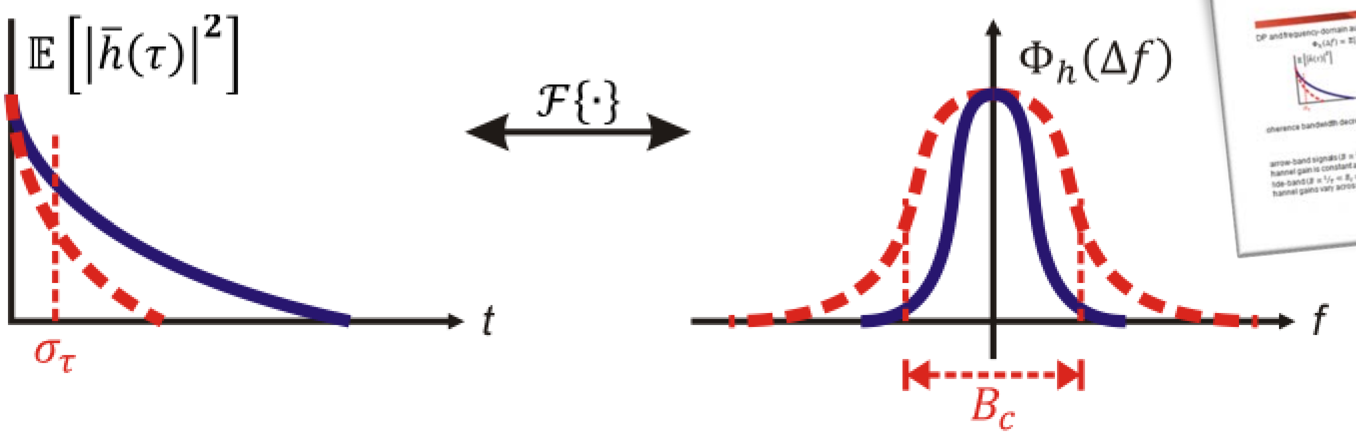
Suboptimality of Per Tone Channel Estimation

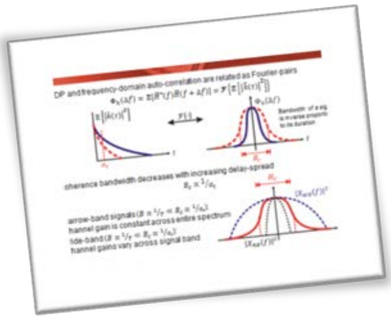
Assumes independence between channel coefficients across all tones

Remember:

- Correlation across tones measured by the **coherence bandwidth** and quantified by

$$\Phi_h(\Delta f) = \mathbb{E}[\bar{H}^*(f)\bar{H}(f+\Delta f)] = \boldsymbol{\mathcal{F}}\left\{\mathbb{E}\left[\left|\bar{h}(\tau)\right|^2\right]\right\}$$





- Coherence bandwidth is inverse proportional to the channel delay spread

- **OFDM** (by design): coherence bandwidth >> carrier spacing

Independence assumption is wrong (->pessimistic)

Least Squares Estimation (Freq. Domain View)

- Vector of channel coefficients spans an N-dimensional space

$$\boldsymbol{h} = [h_0 \quad h_1 \quad \cdots \quad h_{N-1}]^T$$

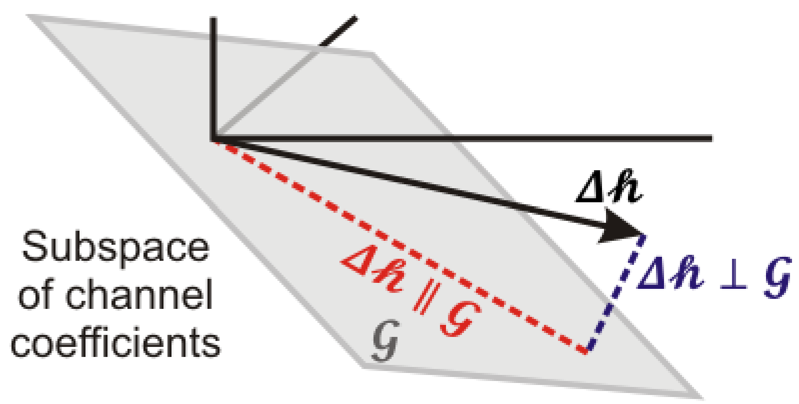
Per-tone channel estimation has two main signal components

$$\hat{\boldsymbol{h}} = \boldsymbol{h} + \boldsymbol{\Delta h}$$

- Uncorrelated (=independent) across tones noise $\boldsymbol{\Delta h}$
- *Correlated* channel across tones channel coefficients $\boldsymbol{h}$

The desired signal $\boldsymbol{h}$ lies in a lower dimensional subspace $\mathcal{G}$ (support)

- Two noise components
  - Orthogonal to the channel subspace $\boldsymbol{\Delta h} \perp \mathcal{G}$
  - Within the channel subspace $\boldsymbol{\Delta h} \parallel \mathcal{G}$
- Per tone estimation: no noise filtering
- But, if we know $\mathbf{G}$, we can remove $\boldsymbol{\Delta h} \perp \mathcal{G}$ completely => better estimate

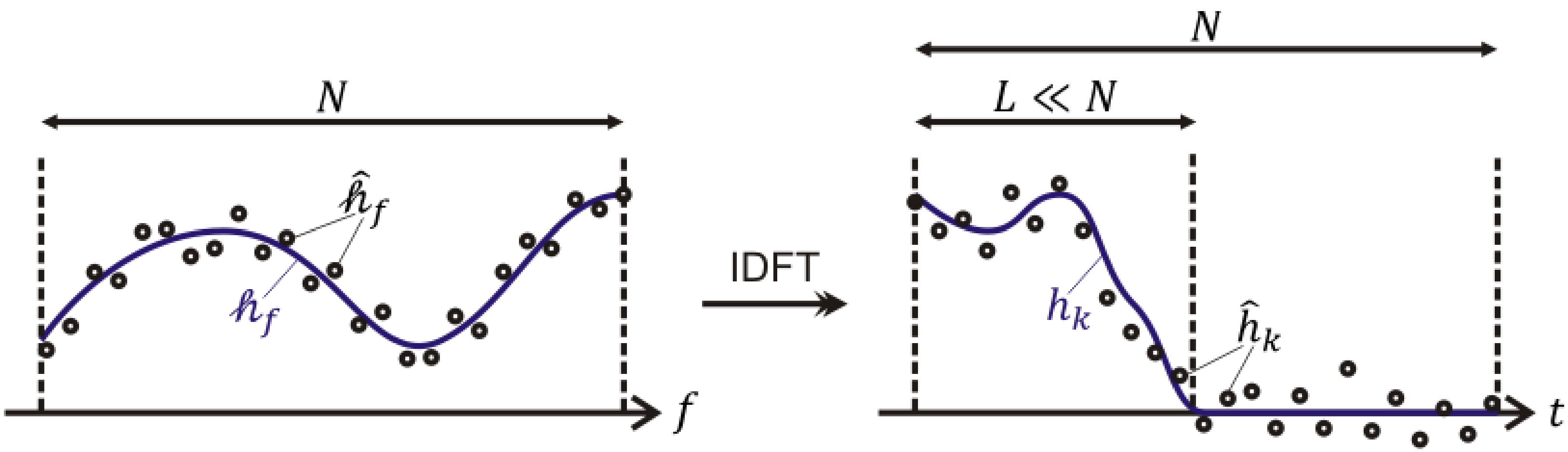


Least Squares Estimation (Time Domain View)

Frequency domain: subspace of the channel is not immediately obvious

- **OFDM** (by design): delay spread < guard interval L << number of carriers N

Consider the time domain representation



$$\mathbf{h} = [h_0 \quad \cdots \quad h_{L-1} \quad 0 \quad \cdots \quad 0]^T$$

$$\mathcal{G} = \mathbf{F}\boldsymbol{G} \xleftarrow{\text{DFT}} \boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Least Squares Estimation (Time Domain View)

- Write $\boldsymbol{\mathcal{h}}$ in terms of its (non-zero) time-domain representation $\mathbf{h}_L$

$$\widehat{\boldsymbol{\mathcal{h}}} = \mathbf{F}\boldsymbol{G}\hat{\mathbf{h}}_L = \mathbf{F}_{NxL}\hat{\mathbf{h}}_L$$

$\hat{\mathbf{h}}_L$ = $[\hat{h}_0 \quad \cdots \quad \hat{\mathrm{h}}_{L-1}]^T \quad (\hat{\mathbf{h}} = \boldsymbol{G}\hat{\mathbf{h}}_L)$

$\mathbf{F}_{NxL}$: Truncated Fourier matrix (only first L columns)

$$\boldsymbol{\mathcal{y}} = \boldsymbol{\mathcal{T}}\mathbf{F}_{NxL}\mathbf{h}_L + \mathbf{n}$$

Overdetermined system!!

Least squares solution

$$\hat{\mathbf{h}}_L = \underset{\hat{\mathbf{h}}_L}{\operatorname{argmin}}\left\{\left|\boldsymbol{\mathcal{y}} - \boldsymbol{\mathcal{T}}\mathbf{F}_{NxL}\hat{\mathbf{h}}_L\right|^2\right\}$$

- Pseudo inverse yields impulse response

$$\hat{\mathbf{h}}_L = (\mathbf{F}_{NxL}^H\boldsymbol{\mathcal{T}}^H\boldsymbol{\mathcal{T}}\mathbf{F}_{NxL})^{-1}\mathbf{F}_{NxL}^H\boldsymbol{\mathcal{T}}^H\boldsymbol{\mathcal{y}}$$

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Least Squares Estimation (Time Domain View)

Back to frequency domain ($\widehat{\boldsymbol{h}} = \mathbf{F}_{NxL}\hat{\mathbf{h}}_L$)

$$\widehat{\boldsymbol{h}} = \mathbf{F}_{NxL}(\mathbf{F}_{NxL}^H \boldsymbol{\mathcal{T}}^H \boldsymbol{\mathcal{T}} \mathbf{F}_{NxL})^{-1} \mathbf{F}_{NxL}^H \boldsymbol{\mathcal{T}}^H \boldsymbol{y}$$

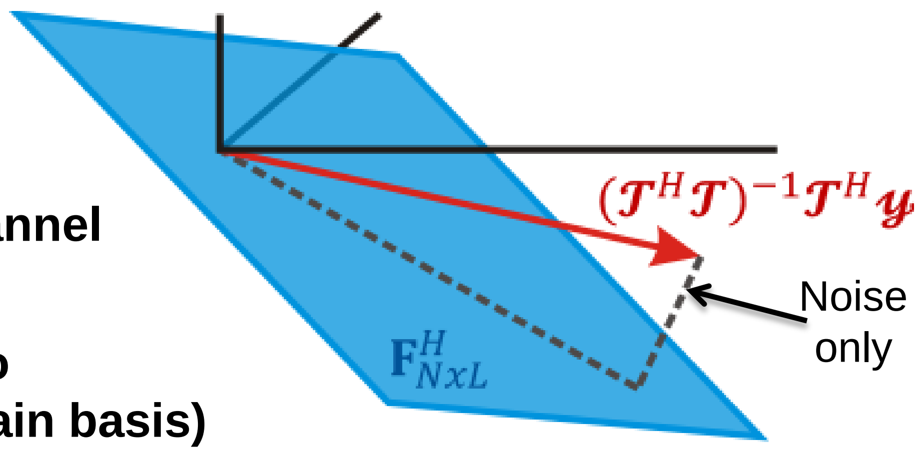
- $\boldsymbol{\mathcal{T}}^H \boldsymbol{\mathcal{T}} = \mathrm{P}_t \boldsymbol{I}$ is a scaled identity matrix
- $\mathbf{F}_{NxL}^H \mathbf{F}_{NxL} = \boldsymbol{I}_L$

support of the channel

$$\widehat{\boldsymbol{h}} = \mathbf{F}_{NxL} \mathbf{F}_{NxL}^H \underbrace{(\boldsymbol{\mathcal{T}}^H \boldsymbol{\mathcal{T}})^{-1} \boldsymbol{\mathcal{T}}^H \boldsymbol{y}}_{\text{per-tone estimate}}$$

Assume $\mathrm{P}_t = 1$

1. **Per tone estimate**
2. **Projection onto the support $\mathbf{F}_{NxL}^H$ of the channel**
3. **Transformation back to the original (freq. domain basis)**



Interpretation in the Frequency Domain

$$\widehat{\boldsymbol{h}} = \mathbf{F}_{NxL}\mathbf{F}_{NxL}^{H}(\boldsymbol{\mathcal{T}}^{H}\boldsymbol{\mathcal{T}})^{-1}\boldsymbol{\mathcal{T}}^{H}\boldsymbol{y}$$

- $(\boldsymbol{\mathcal{T}}^{H}\boldsymbol{\mathcal{T}})^{-1}\boldsymbol{\mathcal{T}}^{H}\boldsymbol{y}$: Per-tone channel estimation
- $\mathbf{F}_{NxL}\mathbf{F}_{NxL}^{H}$: Low-pass filter in frequency domain

$$\mathbf{F}_{NxL}\mathbf{F}_{NxL}^{H} = \mathbf{F}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}\mathbf{F}^{H} = \mathbf{F}\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}}_{diag\{[\underbrace{1 \ \cdots \ 1}_{L} \ \underbrace{0 \ \cdots \ 0}_{N}]\}}\mathbf{F}^{H}$$

- $\mathbf{FDF}^{H}$: Circulant matrix
- First row given by $\mathrm{DFT}\{diag\{\mathbf{D}\}\}$
- Multiplication with a circulant matrix = cyclic convolution with its first row

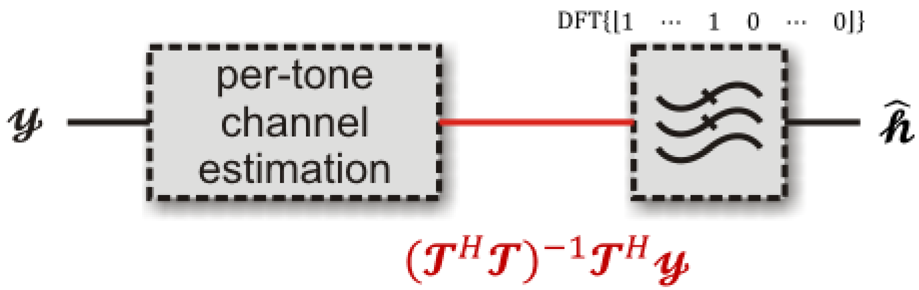
Multiplication of the per-tone estimate with $\mathbf{F}_{NxL}\mathbf{F}_{NxL}^{H}$ corresponds to a cyclic convolution in frequency domain

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Interpretation in the Frequency Domain

Multiplication of the per-tome estimate with $\mathbf{F}_{NxL}\mathbf{F}_{NxL}^{H}$ corresponds to a cyclic convolution in frequency domain

Convolution with $\mathrm{DFT}\{[1 \quad \cdots \quad 1 \quad 0 \quad \cdots \quad 0]\}$ of the per-tone estimates corresponds to low-pass filter in the frequency domain



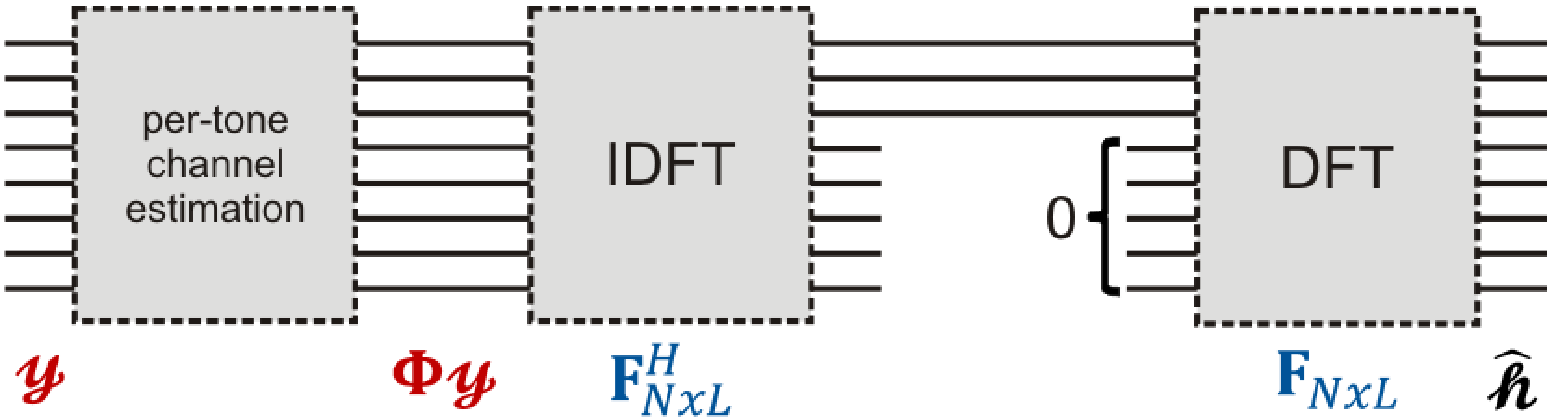
- Computational complexity (multiplications) $\propto N^2$

Low Complexity Implementation

More economic implementation

$$\hat{\boldsymbol{h}} = \mathbf{F}_{NxL}\mathbf{F}_{NxL}^{H}\underbrace{(\boldsymbol{\mathcal{T}}^{H}\boldsymbol{\mathcal{T}})^{-1}\boldsymbol{\mathcal{T}}^{H}}_{\boldsymbol{\Phi}}\boldsymbol{y}$$

- $\mathbf{F}_{NxL}^{H}$: Truncated IFFT (keep only L of N outputs)
- $\mathbf{F}_{NxL}$: Zero-padded FFT (Only first L coefficients are non-zero)



- Computational complexity (multiplications) $\propto N \log N$ (complexity of an FFT/IFFT)

Impact on Channel Estimation Error

Reminder: total channel estimation error with per-tone estimation

$$(\boldsymbol{\mathcal{T}}^H\boldsymbol{\mathcal{T}})^{-1}\boldsymbol{\mathcal{T}}^H\boldsymbol{\mathcal{y}} = \boldsymbol{\mathcal{h}} + \boldsymbol{\Delta\mathcal{h}}$$

$$\mathbb{E}\{|\boldsymbol{\Delta\mathcal{h}}|^2\} = N\sigma_{\boldsymbol{\mathcal{k}}}^{\boldsymbol{2}} = N\sigma_{\mathcal{h}_f}^{\boldsymbol{2}} = N\sigma_{\boldsymbol{n}}^{\boldsymbol{2}}/P_t$$

Time-domain least squares $\widehat{\boldsymbol{\mathcal{h}}} = \mathbf{F}_{NxL}\mathbf{F}_{NxL}^H(\boldsymbol{\mathcal{h}} + \boldsymbol{\Delta\mathcal{h}})$

- Consider the noise $\boldsymbol{\Delta\mathcal{k}}$ after projection onto the support of the channel $\mathbf{F}_{NxL}^H$

$$\mathbb{E}\left\{\left|\mathbf{F}_{NxL}^H\boldsymbol{\Delta\mathcal{h}}\right|^{\boldsymbol{2}}\right\} = \mathbb{E}\{|\boldsymbol{\Delta}\mathbf{h}|^2\} \text{ with } \boldsymbol{\Delta}\mathbf{h} = \mathbf{F}_{NxL}^H\boldsymbol{\Delta\mathcal{h}}$$

- Since $\mathbf{F}_{NxL}^H$ is an orthonormal transform: $\sigma_{h_k}^{\boldsymbol{2}} = \sigma_{\mathcal{h}_f}^{\boldsymbol{2}} = \sigma_{\boldsymbol{n}}^{\boldsymbol{2}}/P_t$ and all Δh_k are independent

$$\mathbb{E}\{|\boldsymbol{\Delta}\mathbf{h}|^2\} = \sum_{k=1}^{L}\mathbb{E}\{|\Delta h_k|^2\} = L\sigma_{\boldsymbol{n}}^{\boldsymbol{2}}/P_t$$

- Transformation back to frequency domain only re-distributes the noise across tones

$$\mathbb{E}\left\{\left|\widehat{\boldsymbol{\mathcal{h}}} - \boldsymbol{\mathcal{h}}\right|^2\right\} = \frac{L\sigma_{\boldsymbol{n}}^{\boldsymbol{2}}}{P_t} = \frac{L}{N}\mathbb{E}\{|\boldsymbol{\Delta\mathcal{h}}|^2\}$$

Improvement compared to per-tone estimation

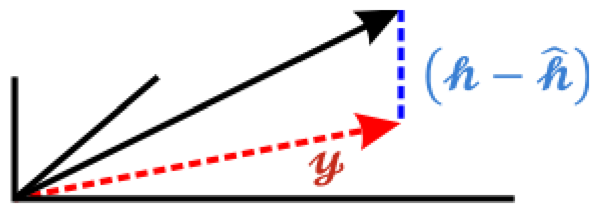
Linear MMSE Channel Estimation (Frequ. Domain View)

$$\hat{\boldsymbol{h}} = \underset{\hat{\boldsymbol{h}}}{\operatorname{argmin}}\, \mathbb{E}\left\{\left|\boldsymbol{h} - \hat{\boldsymbol{h}}\right|^2\right\}$$

- Linear MMSE estimator : $\hat{\boldsymbol{h}} = \boldsymbol{W}\boldsymbol{y}$

Solving the MMSE estimation problem:

- **Orthogonality principle**: ***residual error must be orthogonal to the observation***



$$\mathbb{E}\left\{\left(\boldsymbol{h} - \hat{\boldsymbol{h}}\right)\boldsymbol{y}^{\mathrm{H}}\right\} = \mathbf{0}$$

- Substitute $\boldsymbol{y}$ and $\hat{\boldsymbol{h}} = \boldsymbol{W}\boldsymbol{y}$, then take the expectation

$$\mathbb{E}\left\{\left(\boldsymbol{h} - \hat{\boldsymbol{h}}\right)\boldsymbol{y}^{\mathrm{H}}\right\} = \mathbb{E}\{(\boldsymbol{h} - \boldsymbol{W}\boldsymbol{y})\boldsymbol{y}^{\mathrm{H}}\} = 0 \qquad \overset{\boldsymbol{\Phi}_{\boldsymbol{y}\boldsymbol{h}}}{\mathbb{E}\{\boldsymbol{h}\boldsymbol{y}^{\mathrm{H}}\}} = \boldsymbol{W}\overset{\boldsymbol{\Phi}_{\boldsymbol{y}\boldsymbol{y}}}{\mathbb{E}\{\boldsymbol{y}\boldsymbol{y}^{\mathrm{H}}\}}$$

$$\boldsymbol{W} = \boldsymbol{\Phi}_{\boldsymbol{h}\boldsymbol{h}}\boldsymbol{\mathcal{T}}^{\mathrm{H}}\left(\boldsymbol{\mathcal{T}}\boldsymbol{\Phi}_{\boldsymbol{h}\boldsymbol{h}}\boldsymbol{\mathcal{T}}^{\mathrm{H}} + \boldsymbol{\Phi}_{\boldsymbol{n}\boldsymbol{n}}\right)^{-1}$$

$$\boldsymbol{W} = \boldsymbol{\Phi}_{\boldsymbol{y}\boldsymbol{h}}\boldsymbol{\Phi}_{\boldsymbol{y}\boldsymbol{y}}^{-1}$$

Covariance matrix of the channel coeffs

$$\boldsymbol{\Phi}_{\boldsymbol{h}\boldsymbol{h}} = \mathbb{E}\{\boldsymbol{h}\boldsymbol{h}^{\mathrm{H}}\} = ?$$

Covariance matrix of the thermal noise

$$\boldsymbol{\Phi}_{\boldsymbol{n}\boldsymbol{n}} = \mathbb{E}\{\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}}\} = \sigma_n^2\mathbf{I}$$

Linear MMSE Channel Estimation

How do we know the frequency domain covariance matrix of the channel?

- Start from the time-domain representation of the channel and substitute

$$\boldsymbol{\hbar} = \mathbf{F}\mathbf{h}$$
$$\boldsymbol{\Phi}_{\boldsymbol{\hbar}\boldsymbol{\hbar}} = \mathbf{F}\mathbb{E}\{\mathbf{h}\mathbf{h}^{\mathrm{H}}\}\mathbf{F}^{\mathrm{H}} = \mathbf{F}\boldsymbol{\Phi}_{\mathbf{hh}}\mathbf{F}^{\mathrm{H}}$$

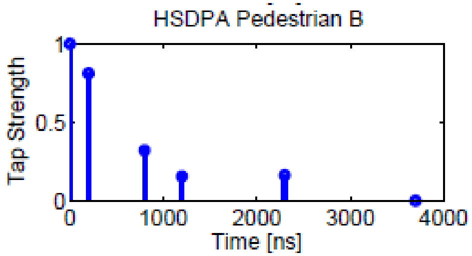
We can write the frequency domain channel covariance matrix in terms of the covariance matrix of the channels time-domain coefficients

Remember from discussion of channel models:

- In many models, individual taps are *uncorrelated* => $\boldsymbol{\Phi}_{\mathbf{hh}}$: diagonal
- Power-delay-profile specifies variance of the individual taps

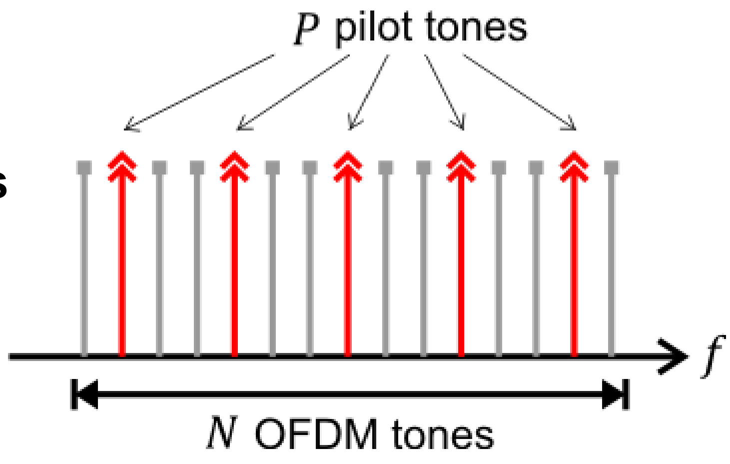
$$\boldsymbol{\Phi}_{\mathbf{hh}} = \mathrm{diag}\{PDP\}$$

- Length of the PDP (including trailing zeros is given by the OFDM symbol duration!!
- $\boldsymbol{\Phi}_{\boldsymbol{\hbar}\boldsymbol{\hbar}}$ is a circulant matrix by construction



Comb-Type Training

Only P out of N tones are used for training



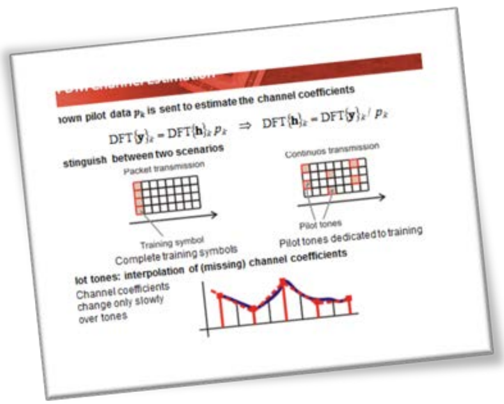
Per tone channel estimation only for P pilot tones

- Set of pilot tone indices $\mathcal{P} = \{p_0, p_1, \dots, p_{P-1}\}$

$$y_{p_k} = h_{p_k} t_k + n_{p_k}$$

- $\boldsymbol{h}_{\mathcal{P}} = [h_{p_0} \quad h_{p_1} \quad \cdots \quad h_{p_{P-1}}]^T$
- $\boldsymbol{y}_{\mathcal{P}} = [y_{p_0} \quad y_{p_1} \quad \cdots \quad y_{p_{P-1}}]^T$

$$\boldsymbol{T}_{\mathcal{P}} = \begin{bmatrix} t_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & t_{P-1} \end{bmatrix}$$



$$\boldsymbol{y}_{\mathcal{P}} = \boldsymbol{T}_{\mathcal{P}} \boldsymbol{h}_{\mathcal{P}} + \boldsymbol{n}_{\mathcal{P}}$$

How can we estimate the channel for all tones?

Time Domain Channel Estimation for Comb-Training

- Write channel frequency response in terms of the channels impulse response
- Consider only the tones occupied by pilot symbols
- Example $N = 8, L = 4$

$$
\begin{bmatrix} \mathscr{h}_0 \\ \mathscr{h}_{p_0} \\ \mathscr{h}_2 \\ \mathscr{h}_{p_1} \\ \mathscr{h}_4 \\ \mathscr{h}_{p_2} \\ \mathscr{h}_6 \\ \mathscr{h}_{p_3} \end{bmatrix}
=
\begin{bmatrix}
f_{0,0} & f_{0,1} & f_{0,2} & f_{0,3} & f_{0,4} & f_{0,5} & f_{0,6} & f_{0,7} \\
f_{1,0} & & & f_{1,3} & f_{1,4} & & & \\
f_{2,0} & & & f_{2,3} & f_{2,4} & & & \\
f_{3,0} & & & f_{3,3} & f_{3,4} & & & \\
f_{4,0} & & & f_{4,3} & f_{4,4} & & & \\
f_{5,0} & & & f_{5,3} & f_{5,4} & & & \\
f_{6,0} & & & f_{6,3} & f_{6,4} & & & \\
f_{7,0} & & & f_{7,3} & f_{7,4} & & &
\end{bmatrix}
\begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ h_3 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
$$

- $\mathbf{F}_{\mathcal{P}xL}$: Truncated Fourier matrix (only first L columns) keeping only the rows with indices in $\mathcal{P}$ (only pilot tones)

Time Domain Channel Estimation for Comb-Training

- System model for pilot subcarriers based on channel impulse response of length L

$$\boldsymbol{y}_{\mathcal{P}} = \boldsymbol{\mathcal{T}}_{\mathcal{P}}\mathbf{F}_{\mathcal{P}xL}\mathbf{h}_L + \boldsymbol{n}_{\mathcal{P}}$$

Solve for the time domain impulse response

$$\widehat{\mathbf{h}_L} = \left(\mathbf{F}_{\mathcal{P}xL}^H\boldsymbol{\mathcal{T}}_{\mathcal{P}}^H\boldsymbol{\mathcal{T}}_{\mathcal{P}}\mathbf{F}_{\mathcal{P}xL}\right)^{-1}\mathbf{F}_{\mathcal{P}xL}^H\boldsymbol{\mathcal{T}}_{\mathcal{P}}^H\boldsymbol{y}$$

- Necessary condition: $\left(\mathbf{F}_{\mathcal{P}xL}^H\boldsymbol{\mathcal{T}}^H\boldsymbol{\mathcal{T}}\mathbf{F}_{\mathcal{P}xL}\right)$ must be invertible
- **Note:** condition of $\left(\mathbf{F}_{\mathcal{P}xL}^H\boldsymbol{\mathcal{T}}^H\boldsymbol{\mathcal{T}}\mathbf{F}_{\mathcal{P}xL}\right)$ depends on $\mathcal{P}$ => improper selection of tones leads to ill conditioned estimator (noise enhancement)

- With $\widehat{\mathbf{h}_L}$ we can easily find $\widehat{\boldsymbol{h}}$ according to

$$\widehat{\boldsymbol{h}} = \mathbf{F}_{NxL}\widehat{\boldsymbol{h}}$$

Time Domain Channel Estimation for Comb-Training

Consider a frequently used special case: $P > L$ training tones are distributed equally spaced in frequency domain

- In this case: $\mathbf{F}_{\mathcal{P}xL}^{H}\mathbf{F}_{\mathcal{P}xL} = \boldsymbol{I}_P$

$$\hat{\boldsymbol{h}} = \mathbf{F}_{NxL}\left(\mathbf{F}_{\mathcal{P}xL}^{H}\boldsymbol{\mathcal{T}}_{\mathcal{P}}^{H}\boldsymbol{\mathcal{T}}_{\mathcal{P}}\mathbf{F}_{\mathcal{P}xL}\right)^{-1}\mathbf{F}_{\mathcal{P}xL}^{H}\boldsymbol{\mathcal{T}}_{\mathcal{P}}^{H}\boldsymbol{y} = \mathbf{F}_{NxL}\mathbf{F}_{\mathcal{P}xL}^{H}\underbrace{\left(\boldsymbol{\mathcal{T}}_{\mathcal{P}}^{H}\boldsymbol{\mathcal{T}}_{\mathcal{P}}\right)^{-1}\boldsymbol{\mathcal{T}}_{\mathcal{P}}^{H}\boldsymbol{y}}_{\text{per-tone estimate}}$$

- $\mathbf{F}_{\mathcal{P}xL}^{H}$: P-point DFT results in a P-periodic impulse response, truncated to the first $L < P$ coefficients

- $\mathbf{F}_{NxL}\mathbf{F}_{\mathcal{P}xL}^{H}$ is a low-pass (interpolation) filter in the frequency domain (cf. Slide 26)

For equally spaced pilot tones, channel estimation corresponds to interpolation of channel coefficients