NEWCOM Autumn School

Estimation Theory for wireless communications

Ecole Nationale Supérieure des Télécommunications, Paris, France
October 24th - October 28th, 2005

- *Brief review on Estimation theory* by Karim ABED-MERAIM
- *Synchronization and Digital receivers* by Marie-Laure BOUCHERET
- *Synchronization in OFDM systems* by Marie-Laure BOUCHERET
- *Bibliography on Training based channel estimation* by Walid HACHEM
- *Recent developments on Multi-Channel blind system identification* by Karim ABED-MERAIM
- *Blind Carrier Frequency Offset estimation and Mean Square Error Lower bounds* by Philippe CIBLAT
- *Channel estimation and Superresolution in UWB system* by Philippe CIBLAT
- *Channel estimation for cyclic prefixed block transmissions* by Mounir GHOGHO
- *Carrier frequency offset for OFDM and related multi-carrier systems* by Mounir GHOGHO
- *Turbo principles and turbo-estimation* by Luc VANDENDORPE
- *Mobile Localization* by Karim ABED-MERAIM

Brief Review on Estimation Theory

K. Abed-Meraim

ENST PARIS, Signal and Image Processing Dept.
abed@tsi.enst.fr

This presentation is essentially based on the course 'BASTA' by E. Moulines

Presentation Outline

- Basic concepts and preliminaries
- Parameter estimation
- Asymptotic theory
- Estimation methods (ML, moment, ...)

Basic Concepts and Preliminaries

Definition and applications

The statistics represent the set of methods that allow the analysis (and information extration) of a given set of observations (data). Application examples include:

- The determination of the production quality by a probing study.
- The measure of the visibility impact of a web site (i.e. number of readed pages, visiting strategies, ...).
- The modelisation of the packets flow at a high-speed network gate.
- The descrimination of important e-mails from spam.
- The prediction of missing data for the restoration of old recordings.
- The estimation and tracking of a mobile position in a cellular system.
- etc, etc, ...

Some history...

One can distinguish 3 phases of development:

- Begining of XIX-th century, apprition of the first data analysis experiments (Prony, Laplace) and the first canonical method in statistics (Gauss, Bayes).
- In the first part of the XX-th century (until the 1960s) the basis of the statistical inference theory have been established by (Pearson, Fisher,, Neyman, Cramer,...). However, due to the lack of powerful calculation machines, the applications and the impact of the statistics were quite limited.
- With the fast development of computers and data bases, the statistic has seen a huge expansion and the number of its applications covers a very large number of domains either in the industry or in research labs.

Statistical model

- In statistics, the observation $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ are seen as a realization of a random vector (process) $\mathbf{X}_n = (X_1, X_2, \cdots, X_n)$ which law P is *partially* known.
- The observation model translates the *a priori knowledge* we have on the data.
- The nature and complexity of the model varies considerably from one application to another...

Parametric model

- *Parametric model*: is a set of probability laws $(P_\theta, \theta \in \Theta)$ indexed by scalar or vectorial parameter $\theta \in \mathbb{R}^d$.
- *Observation*: the observation X is a random variable of distribution P_θ, where the parameter θ is unknown.
- The probability of a given event is a function of θ and hence we'll write: $P_\theta(A), E_\theta(X)$, ...

Objectives

When considering parametric models, the objectives are often:

- *The estimation*: which consists to find an approximate value of parameter θ.
- *The testing*: which is to answer the following type of questions... Can we state, given the observation set, that the proportion of defective objects θ is smaller that $).)1$ with a probability higher than 99%?

Example: Gaussian model

- A random variable X is said standard gaussian if it admits a p.d.f.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}).$$

which is referred to as $X = \mathcal{N}(0,1)$.

- X is a gaussian random variable of mean μ and variance σ^2 if

$$X = \mu + \sigma X_0$$

where X_0 is a standard gaussian.

- *Gaussian model*: the observation $(X_1, \cdots, X_n)$ are n gaussian iid random variables of mean μ and variance σ^2 (i.e. $\theta = (\mu, \sigma)$).

Statistic's concept

- To build statistical estimators or tests, one has to evaluate certain function of the observation: $T_n = T(X_1, \cdots, X_n)$. Such a function is called *statistic*.
- It is crucial that the defined statistic is not a function of the parameter θ or the exact p.d.f. of the observations.
- A statistic is a random variable which distribution can be computed from that of the observations.
- Note that a statistic is a random variable but not any random variable is a statistic.

Examples

- Empirical mean: $T_n = \sum_{i=1}^{n} X_i/n$.
- Median value: $T_n = (X)_n$.
- Min + Max: $T_n = 0.5\,(\max(X_1, \cdots, X_n) + \min(X_1, \cdots, X_n))$.
- Variance: $T_n = \sum_{i=1}^{n} X_i^2/n$.

Parametric Estimation

Parametric versus non-parametric

- *Non-parametric*: The p.d.f. f of X is unknown but belongs to a known function space $\mathcal{F}$, e.g.

$$\mathcal{F} = \{f : \mathbb{R} \to \mathbb{R}^+, \text{twice differentiable and} f" \leq M\}.$$

  leads to difficult estimation problems !!

- *Semi-parametric*: Consider for example a set of observations $\{(X_i, z_i)\}$ following the regression model $X_i = g(\theta, z_i) + \epsilon_i$ where g is a known function and ϵ_i are iid random variables. This model is said semi-parametric if the p.d.f. of ϵ_i is completely unknown.
- *Parametric*: The previous model is parametric if the p.d.f. of ϵ_i is known (up to certain unknown point parameters).

Parametric estimation

- Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ be an observation of a statistical model $(P_\theta, \theta \in \Theta)$.
- An estimator is a function of the observation

$$\hat{\theta}_n(\mathbf{X}) = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$$

  used to infer (approximate) the value of the unknown parameter.

Example: Estimation of the mean value

- Let $(X_1, X_2, \cdots, X_n)$ be a n-sample iid observation given by $X_i = \theta + X_{i0}$, $\theta \in \mathbb{R}$ and X_{i0} are iid zero-mean random variables.
- Mean estimators:

1- Empirical mean: $\hat{\theta}_n = \sum_{i=1}^{n} X_i/n$.

2- Median value: $\hat{\theta}_n = (X)_n$.

3- (Min + Max)/2: $\hat{\theta}_n = \frac{\max(X_1,\cdots,X_n)+\min(X_1,\cdots,X_n)}{2}$.

Estimator

- A statistic is referred to as ‘estimator’ to indicate that it is used to ‘estimate’ a given parameter.
- The estimation theory allows us to characterize ‘good estimators’.
- For that one needs ‘performance measures’ of a given estimator.
- Different performance measures exist that sometimes might lead to different conclusions: i.e. an estimator might be ‘good’ for a first criterion and ‘bad’ for another.

Bias

- An estimator T of parameter θ is said *unbiased* if θ is the mean-value of the distribution of T (θ being the exact value of the parameter): i.e. $E_\theta(T) = \theta$.
- Otherwise, the estimator T is said 'biased' and the difference $b(T, \theta) = E_\theta(T) - \theta$ represents the estimation bias.

Example: variance estimation

- Let $(X_1, \cdots, X_n)$ be an iid observation of pdf $p_\theta(x) = \frac{1}{\sigma} p(x - \mu)$, $\theta = (\mu, \sigma^2)$, and p satisfies $\int x^2 p(x) dx = 1$ and $\int x p(x) dx = 0$.
- $S_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$ is an unbiased estimator of σ^2.
- $V_n = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2$ is a biased estimator of σ^2 which bias is given by $b = -\sigma^2/n$. It is however said *asymptotically* unbiased as the bias goes to zero when n tends to infinity.

Unbiased estimator

- Instead of θ, one might be interested by a function of this parameter... For example in the previous example, the objective can be to estimate $\sigma = \sqrt{\theta_2}$ instead of $\sigma^2 = \theta_2$. When θ is a parameter vector, one might, in particular, be interested in estimating only a sub-vector of θ.

- T is an unbiased estimator of $g(\theta)$ if $E_\theta(T) = g(\theta)$ for all $\theta \in \Theta$.

- Otherwise, $b(T, \theta, g) = E_\theta(T) - g(\theta)$ would represent the bias of this estimator.

Bias and transforms

- Non-lineat transforms of unbiased estimators are not necessarily unbiased: i.e. if T is an unbiased estimator of θ, $g(T)$ is not in general an unbiased estimate of $g(\theta)$.

- For example, if T is an unbiased estimate og θ then T^2 is not an unbiased estimate of θ^2. Indeed, we have

$$E_\theta(T^2) = var_\theta(T) + (E_\theta(T))^2 = var_\theta(T) + \theta^2.$$

Mean squares error

Another pertinent performance measure is the mean squares error (MSE). The MSE measures the dispersion of the estimator arround the 'true' value of the parameter:

$$MSE(T, \theta) = R(T, \theta) = E(T(X) - \theta)^2.$$

The MSE can be decomposed into:

$$MSE(T, \theta) = (b(T, \theta))^2 + var_\theta(T).$$

Example: MSE of the empirical mean

- $(X_1, \cdots, X_n)$ n-sample iid observation of law $\mathcal{N}(\mu, \sigma^2)$.
- Empirical mean: $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$.
- Unbiased estimator and

$$var(\overline{X}) = \frac{\sigma^2}{n}.$$

Estimator's comparison: Risk measure

- We have considered previously the *quadratic risk (loss) function*:

$$l(\theta, \alpha) = (\theta - \alpha)^2.$$

- Other risk (loss) functions are possible and sometimes more suitable:

  1- Absoluve-value error: $l(\theta, \alpha) = \theta - \alpha$,
  2- Truncated quadratic risk function: $l(\theta, \alpha) = \min((\theta - \alpha)^2, d^2)$.
  3- The 0-1 risk function: $l(\theta, \alpha) = 0$ if $\theta - \alpha \leq \epsilon$ and $l(\theta, \alpha) = 1$ otherwise.

- The mean risk value for an estimator is defined as $E_\theta(l(T(X), \theta))$.

Estimator's comparison

One can compare 2 estimators w.r.t. their risk values.

- An estimator T is said 'better' than another estimator T' if

$$R(T, \theta) \leq R(T', \theta), \quad \forall \ \theta \in \Theta$$

  with strict inequality for at least one value of the parameter θ.

- Except for 'very particular cases', it does not exist an estimator *uniformly* better than all other estimators.

Reducing the class of estimators

- *Unbiased estimators*: we seek for the unbiased estimator with the minimum quadratic risk value.
- *Invariance*: One might be interested in estimators satisfying certain invariance property. For example, in a translation model, one is interested in the estimators that satisfy:
  $T(X_1 + c, \cdots, X_n + c) = c + T(X_1, \cdots, X_n)$.
- *Linearity*: One seeks here for the best linear estimator. This is the case, for example, in the linear regression problem (e.g. Theorem of Gauss-Markov).

Cramer Rao Bound: regular model

- For 'regular' statistical models it is possible to determine a lower bound for the quadratic risk (MSE). It is the Cramer-Rao Bound (CRB).
- A statistical model is regular if:

  1- The model is dominated: i.e. $P_\theta(A) = \int_A p_\theta(x)\mu(dx)\ \forall\ A \in \mathcal{B}(X)$.
  2- Θ is an open set of $\mathbb{R}^d$ and $\partial p(x;\theta)/\partial\theta$ exists for all x and all θ.
  3- The pdfs have the same support for all values of θ, i.e. for $A \in \mathcal{B}(X)$, we have either $P_\theta(A) = 0\ \forall\ \theta$ or $P_\theta(A) > 0\ \forall\ \theta$.
  4- $\int_X \frac{\partial}{\partial\theta} p(x;\theta)\mu(dx) = \frac{\partial}{\partial\theta}\int_X p_\theta(x)\mu(dx) = 0$.

Cramer Rao Bound: likelihood & score function

- The function $\theta \rightarrow p(x;\theta)$ is called *likelihood* of the observation.

- For a regular model, the function $\theta \rightarrow S(x;\theta) = \nabla_\theta \log p(x;\theta)$ is called *score* function of the observation.

- When for all θ, $E(S(X;\theta)^2) < \infty$, one define the *Fisher Information Matrix (FIM)* as:

$$I(\theta) = E_\theta[S(X;\theta)S(X;\theta)^T].$$

Fisher information: Properties

- Additivity for iid observations:

$$I_n(\theta) = Cov_\theta(\nabla_\theta \log p(X_1, \cdots, X_n;\theta)) = ni(\theta)$$

  where

$$i(\theta) = Cov_\theta(\nabla_\theta \log p(X_1;\theta))$$

  in other words, each new information contributes in an identical way to the global information.

- When the score function is twice differentiable, we have:

$$I_n(\theta) = -E_\theta(\nabla_\theta^2 \log p(X_1, \cdots, X_n;\theta)).$$

Cramer Rao Bound

- Let $T(X)$ be a statistic such that $E_\theta(T(X)^2) < \infty$, $\forall\, \theta$ and assume that the considered statistical model is regular.

- Let $\psi(\theta) = E_\theta(T(X))$. Then

$$var_\theta(T(X)) \geq \nabla_\theta \psi(\theta)^T I^{-1}(\theta) \nabla_\theta \psi(\theta).$$

- If T is an unbiased estimator of θ, then the CRB becomes:

$$var_\theta(T(X)) \geq I^{-1}(\theta).$$

Example: Empirical mean for gaussian process

- $(X_1, \cdots, X_n)$ n-sample iid observation of law $\mathcal{N}(\mu, \sigma^2)$ (σ^2 known).

- The Fisher information for the mean parameter is given by:

$$I_n(\theta) = n/\sigma^2.$$

- The empirical mean MSE reaches the CRB and hence it is the best estimator (for the quadratic risk) in the class of unbiased estimates.

Example: Linear model

- Observation model: $X = Z\theta + \epsilon$ where $X = [X_1, \cdots, X_n]^T$ is the observation vector, Z is a full rank known matrix and ϵ is the error vector of zero-mean and covariance $E(\epsilon\epsilon^T) = \sigma^2 I$.
- The least squares estimate of θ given by

$$\hat{\theta} = Z^{\#} X$$

  is unbiased and of MSE

$$Var_\theta(\hat{\theta}) = \sigma^2 (Z^T Z)^{-1}.$$

- If ϵ is a gaussian noise, then the FIM is given by $I(\theta) = (Z^T Z)/\sigma^2$ and hence the LS estimate is the best unbiased estimate w.r.t. the quadratic risk.

Efficiency

- An unbiased estimate of θ which reaches the CRB is said *efficient*. It is an unbiased estimate with minimum error variance.
- Efficient estimators exist for the class of exponential distributions where

$$p(x; \theta) \propto \exp(A(\theta)T(x) - B(\theta)).$$

Asymptotic Theory

Asymptotic approach

- Study of the estimators in the limit of ‘large sample sizes’, i.e. $n \to \infty$.
- For usual models, the estimates converge to the exact value of the parameter: *consistency*.
- We then study the dispersion of the estimators around the limit value θ.
- Our tools are: the law of large numbers and the central limit theorem.

Consistency

- Let $(X_1, \cdots, X_n)$ be an observation of a statistical model $(P_\theta, \theta \in \Theta)$.
- $T_n = T_n(X_1, \cdots, X_n)$ is a sequence of consistent estimators of θ if for all θ the sequence of random variables T_n converges in probability to θ:

$$\lim_{n\to\infty} P_\theta(T_n - \theta \geq \delta) = 0 \quad \forall\theta \in \Theta, \delta > 0.$$

.

Large numbers law

- The consistency is often a consequence of the *large numbers law*.
- *Large numbers law*: Let $(X_1, \cdots, X_n)$ be a sequence of iid random variables such that $E(X_1) < \infty$. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow_P E(X).$$

Consistency & continuous transform

- Let T_n be a consistent sequence of estimators of θ, $T_n \to_p \theta$.
- Let ϕ be a continuous function in Θ.
- $\phi(T_n)$ is then a sequence of consistent estimators of $\phi(\theta)$.

Convergence rate

- The consistency is an interesting property but does not give us information on how fast the estimator converges to the limit value.
- In the case of the empirical mean one can easily verify that $\sqrt{n}(\overline{X}_n - \mu)$ is bounded in probability which gives us a rough idea on the convergence speed!!

Asymptotically normal estimator

- An estimator sequence T_n of θ is said asymptotically normal if

$$\sqrt{n}(T_n - \theta) \rightarrow_d \mathcal{N}(0, \sigma^2(\theta)).$$

  where $\sigma^2(\theta)$ is the *asymptotic variance* of the considered estimator.

- This asymptotic result allows us to evaluate (often in a simpler way) the dispersion of the estimators aroud the true value of the parameter.

Convergence in distribution

Let $(X_n, n \geq 0)$ be a sequence of random variables. X_n is said to converge in distribution to X (i.e. $X_n \rightarrow_d X$) if one of the following equivalent properties is verified:

- For any bounded continuous function f: $\lim_{n\to\infty} E(f(X_n)) = E(f(X))$.
- For all u, $\lim_{n\to\infty} E(e^{iuX_n}) = E(e^{iuX})$
- For all subsets $A \in \mathcal{B}(\mathbb{R})$ such that $P(X \in \partial A) = 0$ we have $\lim_{n\to\infty} P(X_n \in A) = P(X \in A)$.

Confidence interval

- Let $(T_n, n \geq 0)$ be a sequence of random variables such that $\sqrt{n}(T_n - \theta) \rightarrow_d T\tilde{\mathcal{N}}(0, \sigma^2)$.
- Let $A = [-a,\ a]$ such that $P(T \in \{a, -a\}) = 0$, then we have

$$\lim_n P_\theta(\sqrt{n}(T_n - \theta) \in [-a,\ a]) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-a}^{a} \exp(-x^2/2\sigma^2)dx = \alpha,\ \forall\,\theta.$$

- Consequently,

$$\lim_n P_\theta(\theta \in [T_n - a/\sqrt{n},\ T_n + a/\sqrt{n}]) = \alpha,\ \ \forall\,\theta$$

which represents a confidence interval of level α for θ.

Central limit theorem

The asymptotic normality of the estimators comes from the *central limit theorem* that can be stated as follows:

Let $(X_1, \cdots, X_n)$ a sequence of iid random variables of mean μ and variance $\sigma^2 = E(X^2) < \infty$. Then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \rightarrow_d \mathcal{N}(0, \sigma^2).$$

The δ-method

- Let T_n a consistent sequence of estimators of θ.
- The continuity theorem states that $g(T_n)$ is a consistent estimate of $g(\theta)$.
- However, this result does not give any information about the convergence rate nor about the asymptotic normality of the estimator $g(T_n)$??

The δ-method

- Suppose that $\sqrt{n}(T_n - \theta) \rightarrow_d T$ and let g be a locally differentiable function at θ. Then:

$$\sqrt{n}(g(T_n) - g(\theta)) \rightarrow_d g'(\theta)T.$$

- If $T = \mathcal{N}(0, \sigma^2), then \sqrt{n}(g(T_n) - g(\theta))$ is asymptotically normal $\mathcal{N}(0, g'(\theta)^2\sigma^2)$.

Relative asymptotic efficiency

- Let T_n and S_n be two asymptotically normal estimators of θ:

$$\sqrt{n}(T_n - \theta) \to_d \mathcal{N}(0, \sigma_T^2(\theta))$$
$$\sqrt{n}(S_n - \theta) \to_d \mathcal{N}(0, \sigma_S^2(\theta))$$

- T_n is said 'asymptotically better' that S_n if

$$\sigma_T^2(\theta) \le \sigma_S^2(\theta) \ \forall \, \theta.$$

Estimation methods

Moments method

- $(X_1, \cdots, X_n)$ n iid random variables $(P_\theta, \theta \in \Theta)$.
- Let $\mu_i(\theta) = E_\theta(g_i(X))$ ($g_i, i = 1, \cdots d$ are given functions).
- Moments method consists in solving in θ the equations

$$\mu_i(\theta) = \hat{\mu}_i, \quad i = 1, \cdots d.$$

where $\hat{\mu}_i$ are empirical (sample averaged) moments.

Moments method

- Several moment choices exist. They should be chosen such that:

  1- One can express explicitely the considered moment function in terms of θ.

  2- Insure a bi-univoque relation between the moments and the desired parameter θ.

- The method is applicable in simple cases only where we have a small number of parameters and there is no ambiguity w.r.t. the choice of the statistics.

Consistency of the moment's estimator

- Using the large numbers law, we have:

$$\frac{1}{n}\sum_{i=1}^{n} g_l(X_i) \to_d E_\theta(g_l(X)).$$

- If the function μ: $\Theta \to \mathbb{R}^d$ is invertible with a continuous inverse function, then the continuity theorem states that

$$\hat{\theta} = \mu^{-1}(\hat{\mu})$$

is a consistent estimate of θ. Similarly, one can establish the asymptotic normality of the moment's estimator using the central limit theorem and the δ-method.

Maximum likelihood method

- Let $X = (X_1, \cdots, X_n)$ a sequence of random variables corresponding to the model $(P_\theta, \theta \in \Theta)$. Let p_θ represents the pdf of X.

- *Likelihood*: $\theta \to p(x;\theta)$ seen as a function of θ.

- *Maximum likelihood estimation*: estimation of $\hat{\theta}$ such that

$$p(x;\hat{\theta}) \geq \max_\theta p(x;\theta).$$

- If $p(x;\theta)$ is differentiable, then $\hat{\theta}$ is a solution of

$$\Delta_\theta log p(x;\hat{\theta}) = 0.$$

Log-likelihood function

- *Log-likelihood*: $L(x;\theta) = \log p(x;\theta)$.
- In the case of iid observations:

$$\frac{1}{n}\log p(x;\theta)_p - K(\theta_0,\theta)$$

where $K(\theta_0,\theta)$ is the Kullback-Leibler information defined by

$$K(\theta_0,\theta) = -E_{\theta_0}\left[\log\frac{p(X;\theta)}{p(X;\theta_0)}\right].$$

Kullback information

The Kullback-Leibler information is a 'distance' measure between two pdf satisfying:

- $K(p_{\theta_0}, p_\theta) \geq 0$
- $K(p_{\theta_0}, p_\theta) = 0$ iff

$$P_{\theta_0}(x : p(x;\theta_0) = p(x;\theta)) = 1.$$

Mean and variance of a gaussian

- Log-likelihood:

$$\log p(x;\mu,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

- Likelihood equations:

$$\frac{\partial p}{\partial \mu}(x;\hat{\mu};\hat{\sigma}^2) = 0, \quad \frac{\partial p}{\partial \sigma^2}(x;\hat{\mu};\hat{\sigma}^2) = 0.$$

- Solutions:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu})^2.$$

Non-unicity of ML estimate: Uniform distribution

- $(X_1,\cdots,X_n)$ iid random variables of uniform distribution in $[\theta-0.5 \;\; \theta+0.5]$.

- Likelihood

$$p(x;\theta) = \begin{cases} 1 \quad if \;\; \theta \in [\max(X_i)-0.5, \min(X_i)+0.5] \\ 0 \qquad\qquad\qquad \text{otherwise} \end{cases}$$

- The likelihood is constant in the interval

$$[\max(X_i)-0.5, \min(X_i)+0.5].$$

Other methods

- Minimum contrast method,
- M-estimation,
- Z-estimation
- Robust estimation
- $\vdots$

Synchronization
and
Digital Receivers

Marie-Laure BOUCHERET
IRIT/ENSEEIHT
E-mail : Marie-Laure.Boucheret@enseeiht.fr

Synchronization algorithms
(Single carrier systems, Gaussian channels)

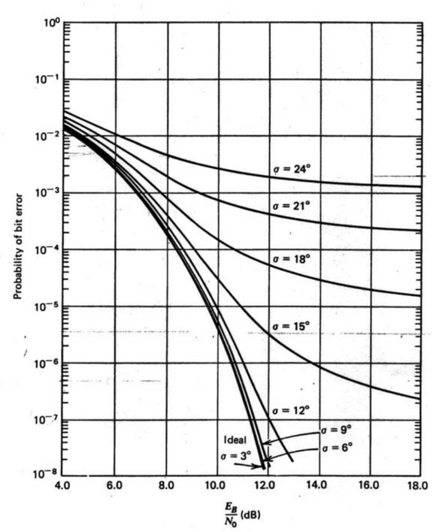
Contents

Impact of synchronization errors (1)

- Carrier phase error:

BPSK, « NRZ »filter

- Maximum phase jitter is determined by the implementation loss in the link budget.

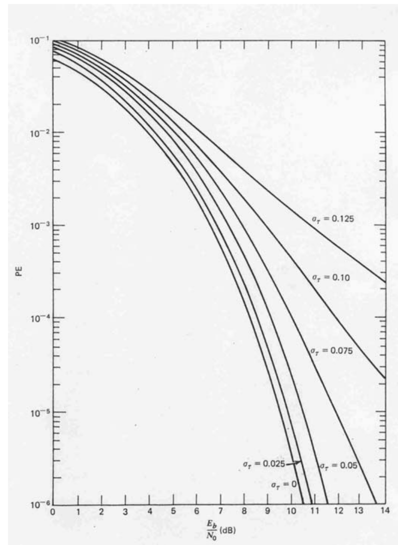


Impact of synchronization errors (2)

- Timing error

BPSK, « NRZ »filter

- Maximum timing jitter is determined by the implementation loss in the link budget.



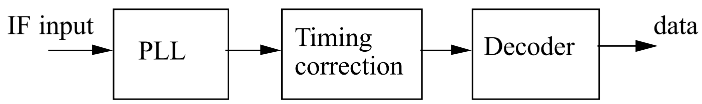
Demodulation

- Functions to be implemented
  - ❑ Baseband conversion
  - ❑ I,Q generation
  - ❑ Carrier recovery
  - ❑ Timing recovery
  - ❑ Matched filtering
  - ❑ Demodulation/decoding

Analog demodulators

• Typical analog demodulator architecture

PLL : baseband conversion + carrier frequency/phase correction



PLL : baseband conversion + carrier frequency/phase correction

Timing correction : FF/FB structure **AFTER PLL**

Digital demodulators

A digital demodulator is **NOT** the sampled version of the equivalent analog demodulator.

⇒ Specific algorithms suited to digital implementation have been developped.

Main differences between digital and analog demodulators:

- Down conversion is INDEPENDENT from phase recovery
- Timing recovery is performed BEFORE phase recovery

Receiver input signal

$$y(t) = \mathrm{Re}\left(x(t)e^{2j\pi f_0 t}\right) + \mathrm{Re}\left(n(t)e^{2j\pi f_0 t}\right)$$
$$x(t) = e^{j\varphi(t)} \sum_k d_k h(t - kT - \tau)$$
$$\varphi(t) = 2\pi\Delta f t + \varphi_0$$

f_0 : carrier frequency, Δf :carrier frequency uncertainty

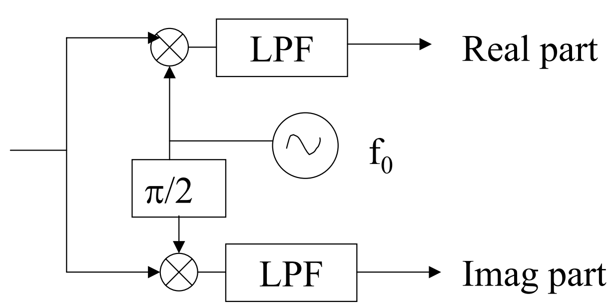
ϕ_0 : phase offset, τ : timing offset

d_k : emitted symbols

h(t): emission filter (wideband channel assumed)

Baseband signal generation (1)

➢ Analog implementation



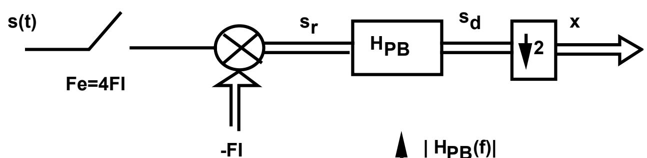
This process can be digitally implemented

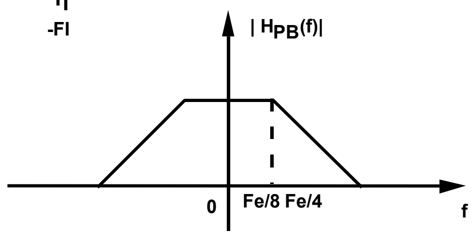
(DAF : digital anti aliasing filter)

Baseband signal generation (2)

➢ Digital implementation (1)

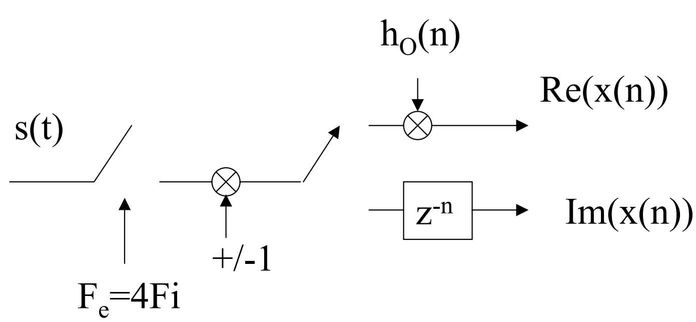
s(t) is the real received passband signal (allocated bandwidth : FI, centred at f_0=FI)





Baseband signal generation (3)

➢ Digital implementation (2)



Likelihood functions (1)

$$r(t) = x(t) + n(t)$$

$$x(t) = e^{j\varphi(t)} \sum_k d_k h(t - kT - \tau)$$

$$\varphi(t) = 2\pi \Delta f t + \varphi_0$$

$\rho(T_0)$: signal observed during a period of duration T_0

$\Phi = \{\varphi_0, \Delta f, \tau, \{d_k\}\}$ Vector of unknown parameters

$\hat{\Phi} = \{\hat{\varphi}_0, \Delta\hat{f}, \hat{\tau}, \{\hat{d}_k\}\}$ Vector of parameters estimates

Synchronization (SC, Gaussian) 13

Likelihood functions (1)

$$\Lambda(\tilde{\Phi}) = \Pr(\rho(T_0) / \tilde{\Phi})$$

In Gaussian channel:

$$\Lambda(\tilde{\Phi}) = \exp\left(-\frac{1}{N_0} \int_{T_0} \left| r(t) - s(t, \tilde{\Phi}) \right|^2 dt \right)$$

$$s(t, \tilde{\Phi}) = A e^{2j\pi \Delta\tilde{f} t + j\tilde{\varphi}_0} \sum_k \tilde{d}_k h(t - kT - \tilde{\tau})$$

$s(t, \tilde{\Phi})$: signal replica

Synchronization (SC, Gaussian) 14

Likelihood functions (2)

<u>Sub-optimal likelihood functions :</u>

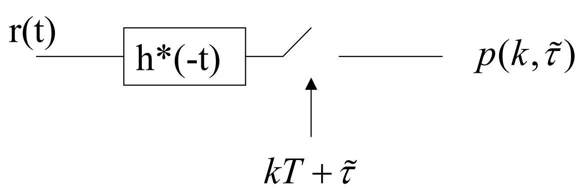
- DD : Decision Directed
- NDA : Non-data aided (depends on modulation)

These sub-optimal likelihood functions are derived for timing, phase and frequency.

Likelihood functions (3)

<u>Timing :</u>

$$L_{NDA}(\tilde{\tau}) = \sum_k |p(k,\tilde{\tau})|^2$$



Timing recovery is performed prior to phase recovery

Likelihood functions (4)

Carrier phase:

- DD likelihood function

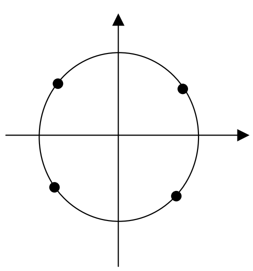
$$L_{DD}(\tilde{\varphi}) = \sum_k \hat{a}_k \operatorname{Re}\left(p(k,\hat{\tau})e^{-j\tilde{\varphi}}\right) + \sum_k \hat{b}_k \operatorname{Im}\left(p(k,\hat{\tau})e^{-j\tilde{\varphi}}\right)$$

- NDA lokelihood function for general rotationnaly symetric signal constellation ($2\pi/N$ symetry)

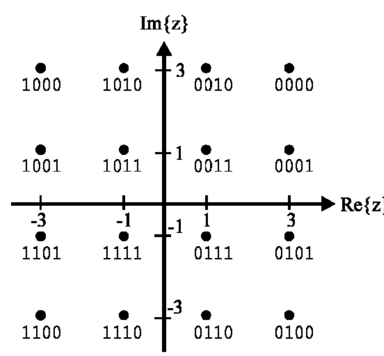
$$L_{NDA}(\tilde{\varphi}) = \operatorname{Re}\left(E\left(d_k^{*N}\right) \sum_k p^N(k,\hat{\tau})e^{-jN\tilde{\varphi}} \right)$$

Likelihood functions (5)

Examples of general rotationnaly symetric signal constellation



QPSK N=4



16QAM N=4

Likelihood functions (6)

- Carrier frequency recovery

QAM

$$L(\{a_k\},\Delta f,\varphi)=\sum_k |d_k|^2+2\sum_k \mathrm{Re}\left\{p(k,\hat{\tau})d_k^* e^{-j\left(2\pi\Delta f kT+\varphi\right)}\right\}$$

PSK

$$L(\{a_k\},\Delta f,\varphi)=\sum_k \mathrm{Re}\left\{p(k,\hat{\tau})d_k^* e^{-j\left(2\pi\Delta f kT+\varphi\right)}\right\}$$

Carrier phase recovery : DDMLFB (1)

Derivation of detector expression from Likelihood function

$$\frac{d}{d\tilde{\varphi}} L_{DD}(\tilde{\varphi}) = 0 \text{ for } \tilde{\varphi}=\hat{\varphi}$$

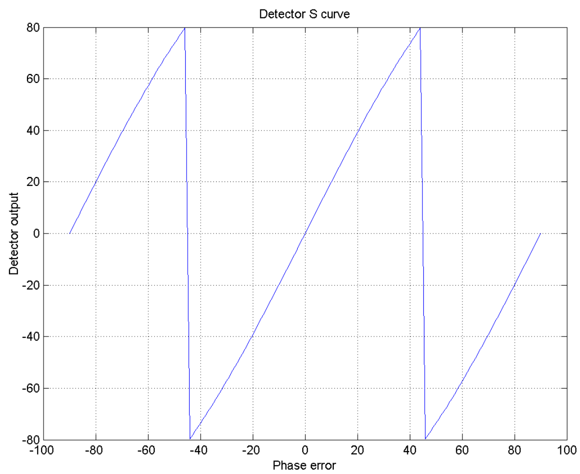
$$\Rightarrow \sum_k \mathrm{Im}\left(d_k^* p(k,\hat{\tau})e^{-j\tilde{\varphi}}\right) = 0 \text{ for } \tilde{\varphi}=\hat{\varphi}$$

$$\boxed{\Rightarrow\ u(k) = \mathrm{Im}\left(d_k^* p(k,\hat{\tau})e^{-j\tilde{\varphi}}\right) \text{ is a phase detector}}$$

Carrier phase recovery : DDMLFB (2)

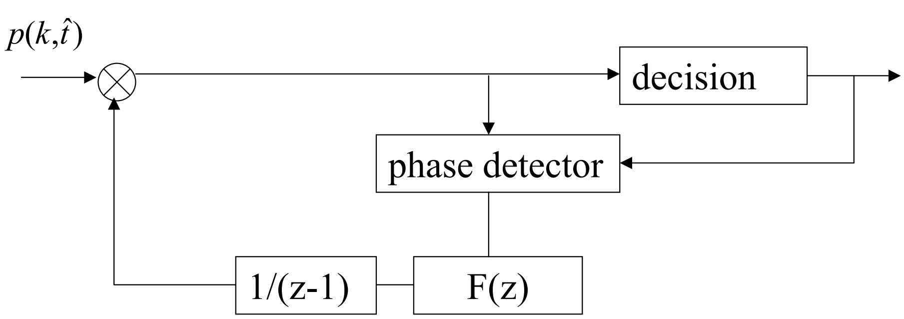
S curve (example for QPSK)

=> Phase ambiguity (solved by using differential encoding/decoding)



Carrier phase recovery : DDMLFB (3)

DPLL



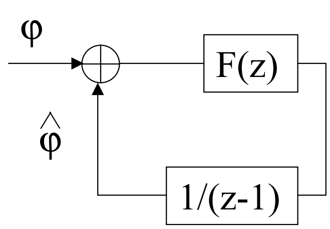
Carrier phase recovery : DDMLFB (4)

Other possible detectors

$$
\begin{aligned}
&p(k,\hat{\tau}) = w(k)\\
&u_1(k) = \mathrm{Im}\left(w^*(k).\mathrm{sgn}\left\{ w(k) - \hat{d}_k \right\} \right)\\
&u_2(k) = \mathrm{Im}\left(\hat{d}_k^* \right)\left[w(k) - \hat{d}_k \right]\\
&u_3(k) = \mathrm{Im}\left(d_k^*.c\mathrm{sgn}\left\{ w(k) - \hat{d}_k \right\} \right)\\
&u_4(k) = \mathrm{Im}\left(\hat{d}_k^* \right)\mathrm{sgn}\left[w(k) - \hat{d}_k \right]
\end{aligned}
$$

Carrier phase recovery : DDMLFB (5)

Phase equivalent scheme



$$
H(z) = \frac{\hat{\varphi}(z)}{\varphi(z)}
$$

$$
2B_L T = \frac{1}{2j\pi} \oint_\gamma H(z) H^*(z^{-1}) \frac{dz}{z}
$$

$$
\sigma^2 \propto \frac{B_L T}{E_s / N_0}
$$

Carrier phase recovery : NDAMLFB (1)

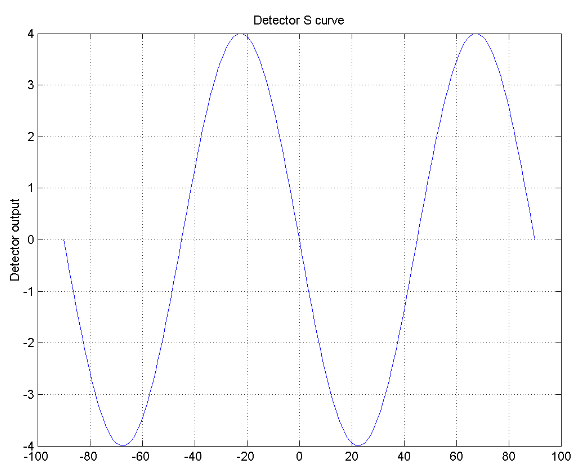
<u>Example for QPSK</u>

$$\frac{d}{d\tilde{\varphi}} L_{NDA}(\tilde{\varphi}) = 0 \text{ for } \tilde{\varphi} = \hat{\varphi}$$

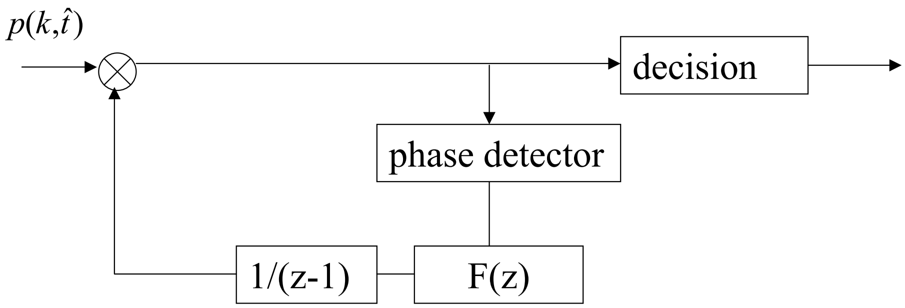
$$\Rightarrow \sum_k \mathrm{Im}\left(\left\{ p(k,\hat{\tau}) e^{-j\tilde{\varphi}} \right\}^4 \right) = 0 \text{ for } \tilde{\varphi} = \hat{\varphi}$$

$$\boxed{\Rightarrow u(k) = \mathrm{Im}\left(\left\{ p(k,\hat{\tau}) e^{-j\tilde{\varphi}} \right\}^4 \right) \text{ is a phase detector}}$$

Carrier phase recovery : NDAMLFB (2)



Carrier phase recovery : NDAMLFB (3)



Carrier phase recovery : NDAMLFF (1)

- Suited for burst transmission
- Two types of structures : block window, sliding window
- Example for QPSK

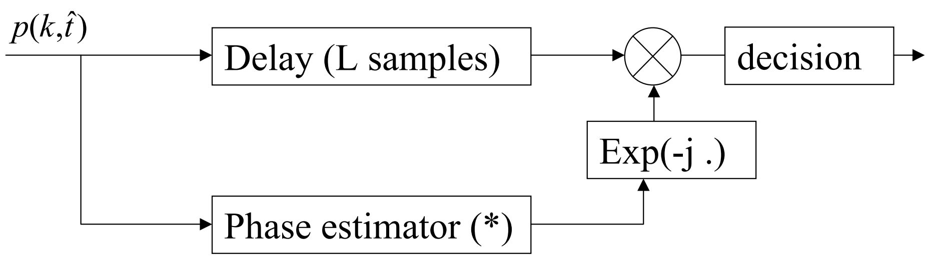
$$\sum_k \mathrm{Im}\left(\left\{p(k,\hat{\tau})e^{-j\tilde{\varphi}}\right\}^4\right)=0 \text{ for } \tilde{\varphi}=\hat{\varphi}$$

$$\Rightarrow \boxed{\hat{\varphi}=\frac{1}{4}Arg\left(\sum_k p^4(k,\hat{\tau})\right)+k\frac{\pi}{2}}$$

$\Rightarrow$ Phase ambiguity (kπ/2)

Carrier phase recovery : NDAMLFF (2)

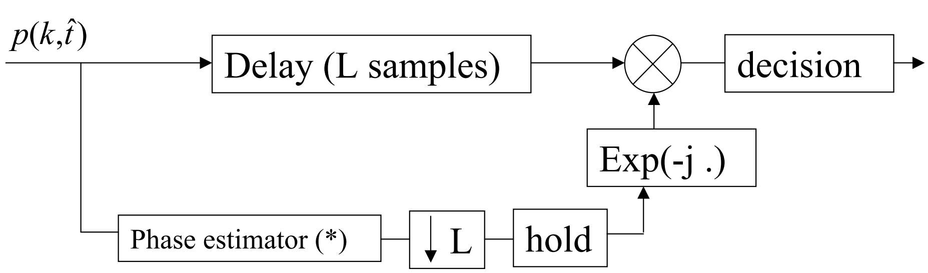
<u>« Sliding window » estimator</u>



(*): averaging over 2L+1 samples

Carrier phase recovery : NDAMLFF (3)

- <u>« Block » estimator</u>



FF structures vs FB structures

- Advantage
  - No acquisition time
- Drawbacks
  - Smaller $B_L T$ => higher jitter, higher cycle slip probability
  - Sensitivity to frequency deviation

Timing recovery (1)

$$L_{NDA}(\tilde{\tau}) = \sum_k \left|p(k,\tilde{\tau})\right|^2 = \sum_k \mathrm{Re}^2\left(p(k,\tilde{\tau})\right) + \sum_k \mathrm{Im}^2\left(p(k,\tilde{\tau})\right)$$

$$\frac{d}{d\tau} L_{NDA}(\tilde{\tau}) = 2\sum_k \mathrm{Re}(p(k,\tilde{\tau}))\frac{d}{d\tilde{\tau}}\mathrm{Re}(p(k,\tilde{\tau})) + 2\sum_k \mathrm{Im}(p(k,\tilde{\tau}))\frac{d}{d\tilde{\tau}}\mathrm{Im}(p(k,\tilde{\tau}))$$

$\Rightarrow$Derivative vs timing is approximated by a difference

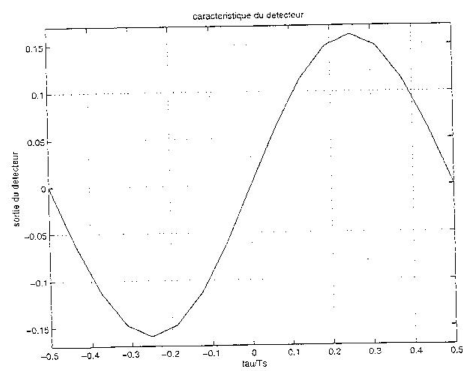
$$\mathrm{Re}(p(k,\tilde{\tau})) \propto \mathrm{Re}(p(k+\lambda,\tilde{\tau})) - \mathrm{Re}(p(k-\lambda,\tilde{\tau}))$$
$$\mathrm{Im}(p(k,\tilde{\tau})) \propto \mathrm{Im}(p(k+\lambda,\tilde{\tau})) - \mathrm{Im}(p(k-\lambda,\tilde{\tau}))$$

Timing recovery (2)

Gardner:

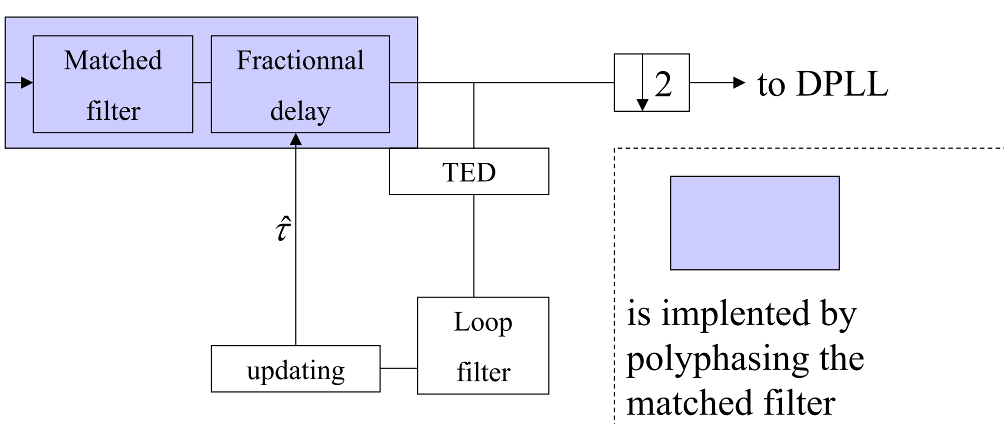
λ=1/2 => detector output is independent from carrier phase error.

$$GA(k)=\mathrm{Re}(p(k+1/2,\tilde{\tau}))\left\{\mathrm{Re}(p(k,\tilde{\tau}))-\mathrm{Re}(p(k+1,\tilde{\tau}))\right\}$$
$$+\mathrm{Im}(p(k+1/2,\tilde{\tau}))\left\{\mathrm{Im}(p(k,\tilde{\tau}))-\mathrm{Im}(p(k+1,\tilde{\tau}))\right\}$$



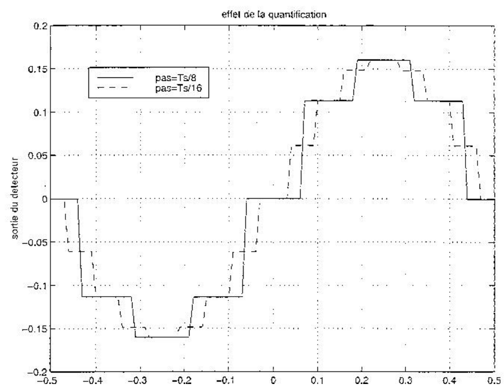
33

Timing recovery (3)



Timing recovery (4)

- S curve (Gardner, quantized)



Timing recovery (5)

- <u>Timing estimator (Oerder and Meyr)</u>

$$\frac{\hat{\tau}}{T} = -\frac{1}{2\pi} Arg\left(\sum_{k=0}^{L-1} \sum_{n=0}^{N-1} \left| p(k,n) \right|^2 e^{2j\frac{\pi n}{N}} \right)$$

$$p(k,n) \triangleq p(kT + nT/N)$$

where N is the number of samples per second

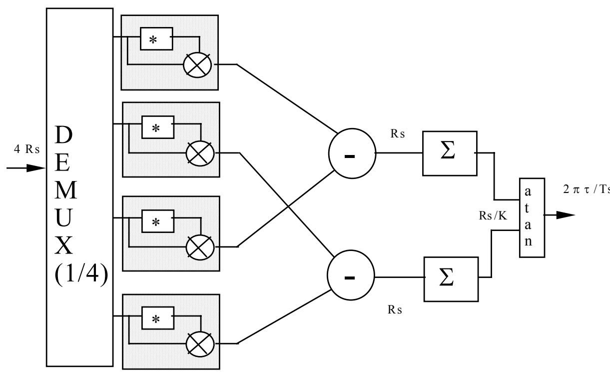
<u>Example :</u> N=4

$$\frac{\hat{\tau}}{T} = -\frac{1}{2\pi} Arg\left(\sum_{k=0}^{L-1} \sum_{n=0}^{3} \left| p(k,n) \right|^2 j^n \right)$$

Timing recovery (6)

Implementation of Oerder and Meyr

$$\frac{\hat{\tau}}{T} = -\frac{1}{2\pi} Arg\left(\sum_{k=0}^{L-1} \left\{ |p(k,0)|^2 - |p(k,2)|^2 \right\} + j \sum_{k=0}^{L-1} \left\{ |p(k,1)|^2 - |p(k,3)|^2 \right\} \right)$$

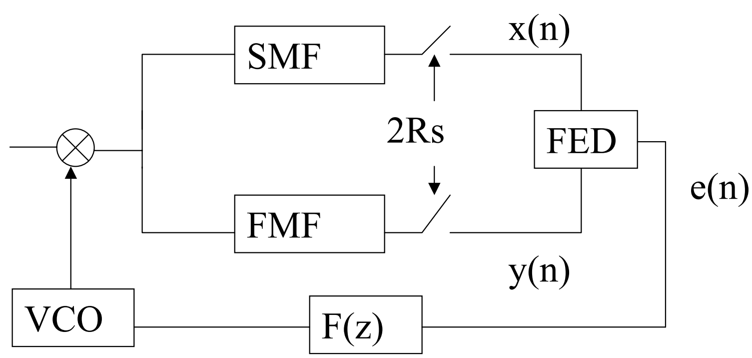


Frequency recovery : general

- Feedback structures
  - « Frequency » detectors
  - « Time » detectors
- Feedforward structures
  - Type 1
  - Type 2

Frequency recovery:FB structures (1)

➢ « Frequency » detector (1)



Frequency recovery:FB structures (2)

➢ « Frequency » detector (2)

SMF : signal matched filter : g(t)
FMF : frequency matched filter : -2jπtg(-t)

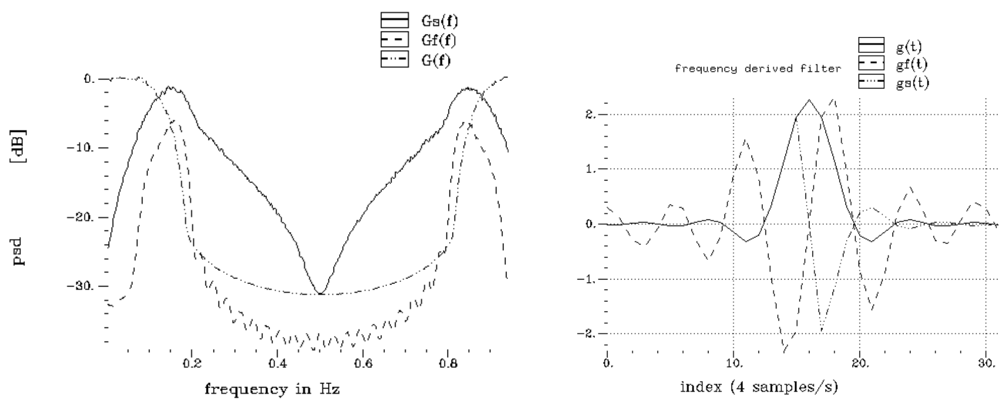
e(n)=Im(x(n)y*(n))

A simpler filter (SFMF) derived from FMF can be used
(g(t)=-j sgn(t) g(-t)

Acquisition range : +/-$(1+\alpha)R_s$

No prior timing correction required

Frequency recovery:FB structures (3)



Synchronization (SC, Gaussian) 41

Frequency recovery:FB structures (3)

- <u>« Time » detectors</u>

Any estimator can be used as a time detector.

Frequency offset range is +/- R_s/M

Timing has to be corrected prior to frequency detection

1 sample/symbol is sufficient.

Synchronization (SC, Gaussian) 42

Frequency recovery:FF structures (1)

➤ Bellini

$$\Delta\hat{f}T = \left(\sum_{-N}^{N} i\alpha_i\right) \Big/ \left(8\pi T \sum_{-N}^{N} i^2\right)$$

=> Cycle slip

α_i: unwrapped phase

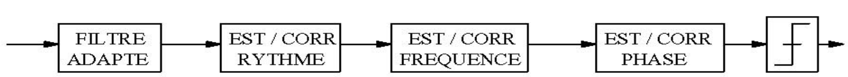
➤ RCFE (reduced complexity frequency estimator)

$$2\pi\Delta\hat{f}T = \frac{1}{MD} Arg\left\{\sum_k d_k \left(r_k r_{k-D}^*\right)^M\right\}$$

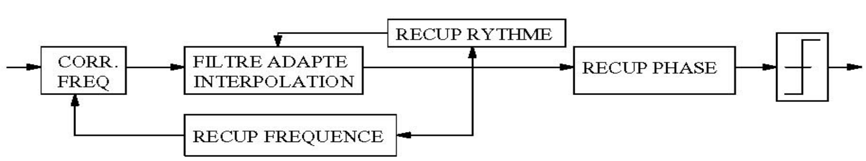
$$r_k = p(k,\hat{\tau})$$

Large D leads to better performances but to lower frequency range.

Digital demodulators (1)



Typical FeedForward Architecture



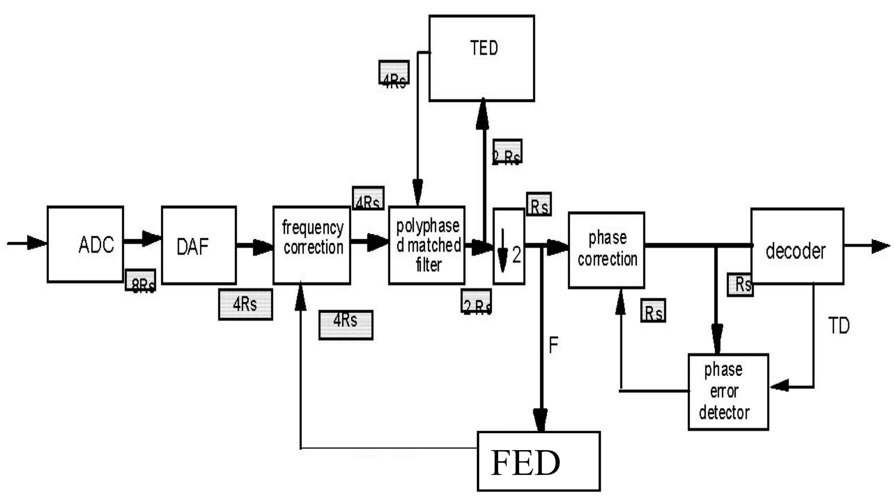
Typical Feedback Architecture

Digital demodulators (2)

Choice of algorithms depends on specifications such as:

- Acquisition time (=> FF/FB structures)
- Maximum frequency deviation (=> frequency circuitry needed)
- Eb/No (=> use of TD if low)
- .....

Digital demodulators (3)



Example: Receiver for TCM (in cooperation with CNES)

Advanced topics

➢Evolutions of input specifications (for satellite communications)

- Low Eb/No (use of efficient coding schemes such as Turbo-Codes and LDPC)
- Bursty transmission
- Large frequency deviation (low-cost terminals, non GEO sat.)

➢Critical function : **phase recovery** (classical algorithms fail)

➢There is a need to develop new synchronisation schemes

References (1)

Reports, books on synchronization

➢ F.M Gardner « Demodulator reference recovery techniques suited for digital implementation » ESTEC contract n° 6847/86/NL/GG,1988

➢ F.M Gardner « Frequency detectors for digital demodulators via ML derivation », ESTEC contract n° 8022/88/NL/DG,part 2, june 1990

➢ T Jesuprret, M Moeneclaey, G Asheid « Digital demodulator synchronization : performance analysis », ESTEC contract n° 8437/89/NL/RE, June 1991

➢ H. Meyr, M Moeneclaey, S.A Fechtel « Digital communications receivers : synchronization, channel estimation and signal processing », J. Wiley and Sons, 1998

PhD Dissaertations

➢ D Mottier « Association des fonctions d'égalisation, de synchronisation et de décodage canal pour les transmissions numériques à grande efficacité spectrale », PhD disertation (in French), INSA de Rennes

➢ Ivar Mortensen "Traitement en bande de base pour charges utiles à régénération bord" ,PhD dissertation (in French), ENST, 1997

References (2)

- Catherine Morlet " Démodulateur embarqué multiporteuses pour applications multimédia par satellites », PhD dissertation (in French), ENST, 2000

Phase estimation

- A.J Viterbi and A.M Viterbi « Non-linear Estimation of PSK-modulated carrier phase with applications to burst digital transmission », IEEE on IT, 1983
- M Moeneclaey, G de Jonghe, « ML oriented NDA carrier Synchronization for General Rotationnally Symmetric Signal Constellations », IEEE on COM, August 1994
- C.N Georghiades « Blind carrier Phase Acquisition for QAM constellations »,IEEE on COM, November 1997

Frequency estimation

- S Bellini and al « Digital frequency estimation in burst mode QPSK transmission », IEEE on COM, vol COM 38, July 1990
- A.N D'Andrea, U mengali « Performance of a frequency detector based on the Maximu Likelihood principle »
- F Classen and al. « Maximum Likelihood Open Loop Carrier Synchronizer for Digital Radio», ICC 93
- J Zhang and al « Data-Aided estimation of carrier frequency for burst detection of QAM », Electronics Lettters, October 200

References (3)

Timing estimation

- F.M Gardner « A BPSK/QPSK timing error detector for sampled reciver », IEEE on COM, may 1986
- Oerder, Meyr « Digital filter and square timing recovery », IEEE on COM, May 1988
- M.K Nezami, R Sudhakar « New schemes for improving Non-Data-Aided symbol Timing Recovery for QAM receivers in flat fading channels », 2000

OFDM Systems

Marie-Laure Boucheret
IRIT/ENSEEIHT
Email : Marie-Laure.Boucheret@enseeiht.fr

Contents

- Recall : OFDM systems
  - multipath mobile channel
  - Principles of OFDM systems
  - OFDM systems and filter banks
  - OFDM systems with guard interval
  - Advantages/drawbacks of OFDM systems
- Synchronization aspects in OFDM systems
  - Specificity of OFDM system w.r.t synchronization
  - Impact of synchronization errors (frequency, sampling time) on OFDM systems
  - Synchronization algorithms

Recall on multipath mobile channels (1)

➢ Coherence bandwidth : $(\Delta f)_c$

–Two carriers separated by $(\Delta f)_c$ are affected by « more or less » the same attenuation.

$$T_m = \frac{1}{(\Delta f)_c}$$

W : occupied bandwidth

$W \ll (\Delta f)_c$ => non frequency selective channels

$W \gg (\Delta f)_c$ => frequency selective channels

Nota : $(\Delta f)_c$ is not related to the relative mobility emitter/receiver (ex: cables)

Recall on multipath mobile channels (2)

➢ Coherence time $(\Delta t)_c$

Two signal samples separated by less than $(\Delta t)_c$ are affected by « more or less « the same attenuation.

$$B_d = \frac{1}{(\Delta t)_c}$$

B_d: doppler bandwidth

Principles of OFDM systems (1)

ENSEIHT

➢ Frequency selective channels

⇒ Use of multiple carriers

The « elementary channel » (one carrier) is now non frequency selective.

➢ Spectral efficiency

⇒ Use of overlapping orthogonal carriers

➢ Diversity

⇒ Use of ECC

COFDM

Synchronization / OFDM systems 5

Principles of OFDM systems (2)

ENSEIHT

➢ Expression of OFDM signal (complex envelop)

Carrier #i :

$$x_i(t) = \sum_k d_{ik} h(t - kT) \exp\left(2 j\pi f_i t\right)$$

h(t): rectangle of width T (NRZ)

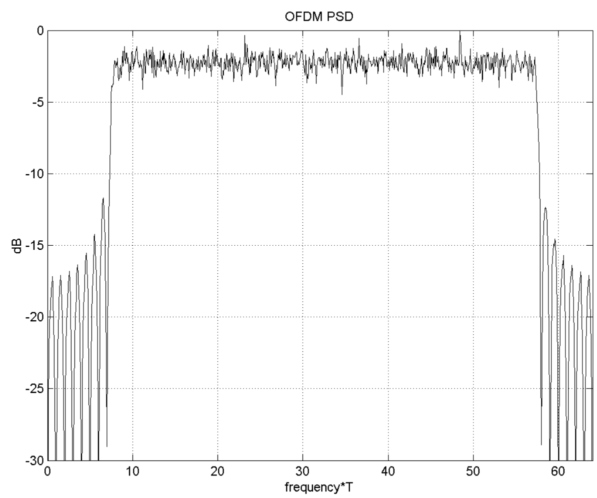
f_i=i/T

Frequency multiplex

$$x(t) = \sum_{i=0}^{N-1} \sum_k d_{ik} h(t - kT) \exp\left(2 j\pi f_i t\right)$$

Synchronization / OFDM systems 6

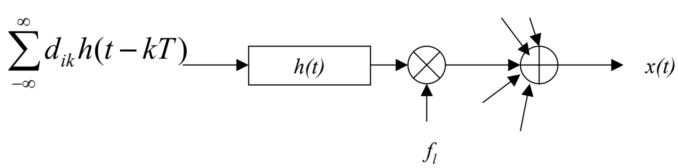
Principles of OFDM systems (3)

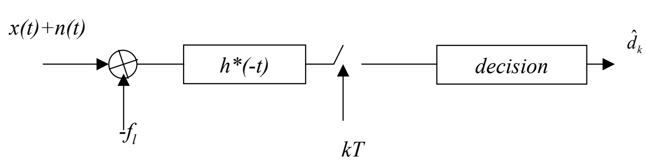


Synchronization / OFDM systems 7

Principles of OFDM systems (4)

➢ Modulator / demodulator for carrier # l (ideal case)

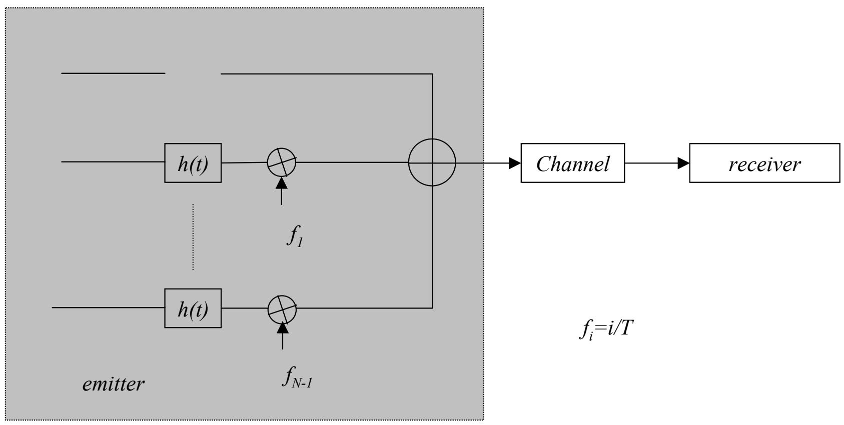




Synchronization / OFDM systems 8

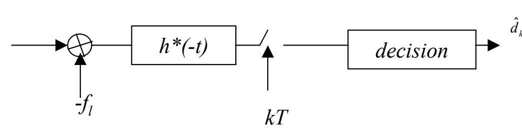
OFDM systems and filter banks (1)

- OFDM modulator/demodulator can be seen as a synthesis/analysis filter bank (no guard time, no coding)

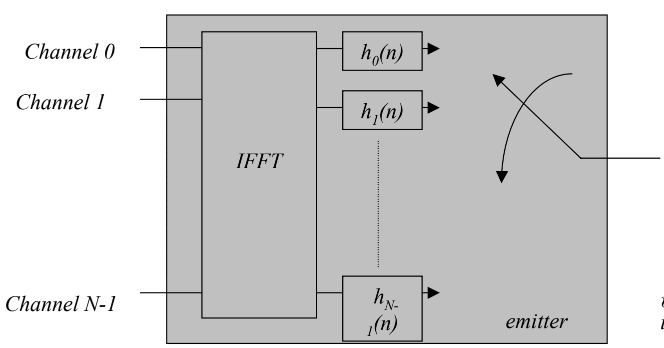


OFDM systems and filter banks (2)

- Receiver for carrier n°1



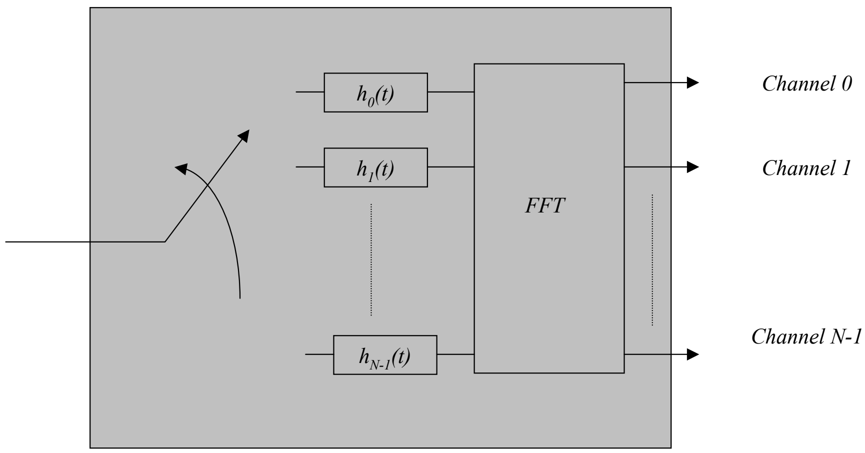
- Efficiently implemented via FFT^{-1} (emitter) and FFT (receiver)



$\{h_i(n)\}$: polyphase implementation of h(n

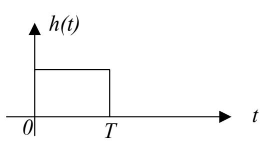
OFDM systems and filter banks (3)

➢OFDM receiver



OFDM systems and filter banks (4)

- Application : classical OFDM



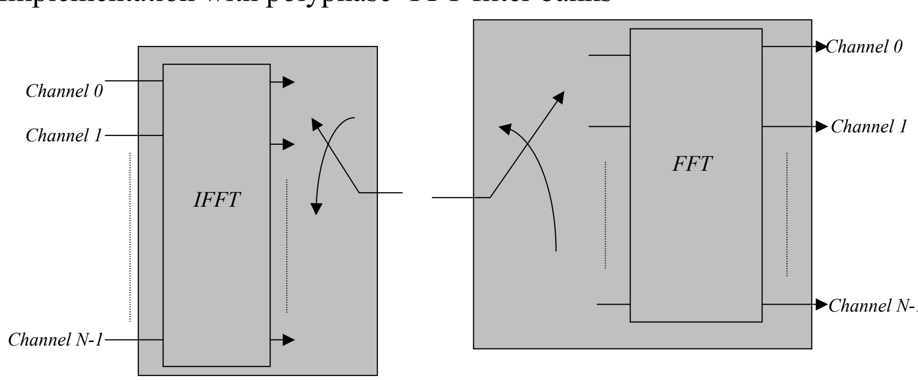
F_e=N/T

h(n)=1 for n=0,…,N-1

h_i(n)=1 for n=0

h_i(n)=0 elsewhere

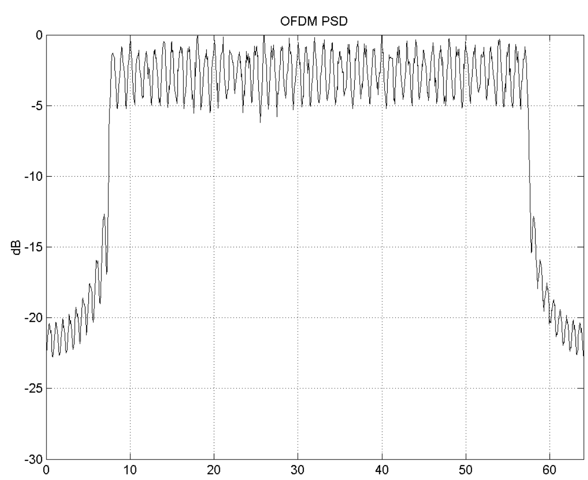
- Implementation with polyphase+FFT filter banks



OFDM system with guard interval (1)

- Guard interval is used to removed residual intersymbol interference (ISI)
- Guard interval is inserted by copying the [kT, kT+ΔT[part of original OFDM symbol => no discontinuity in the signal!
- Resulting OFDM symbol period is T+ΔT (ΔT : guard interval)

OFDM system with guard interval (2)



OFDM system with guard interval (3)

- The FFT output is (symbol # i, carrier #j):

$X_{i,j}=H_j s_{i,j}$ (without noise)

=> flat fading channel at sub-carrier level

- Cyclic prefix is used in order to:
  - Avoid equalization
  - Increase robustness against sampling time error

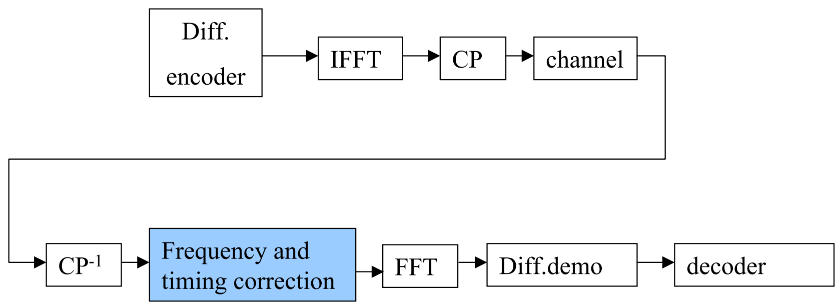
Advantages/drawbacks of OFDM systems

- Advantages:
  - Emitter and receiver are efficiently implemented with FFT/IFFT
  - No equalization is required
  - Spectral efficiency
  - Diversity

- Drawbacks
  - Sensitivity to synchronization errors
  - Sensitivity to non linearities (Amplifiers)
  - Mainly used in broadcasting applications

Receiver Architecture (1)

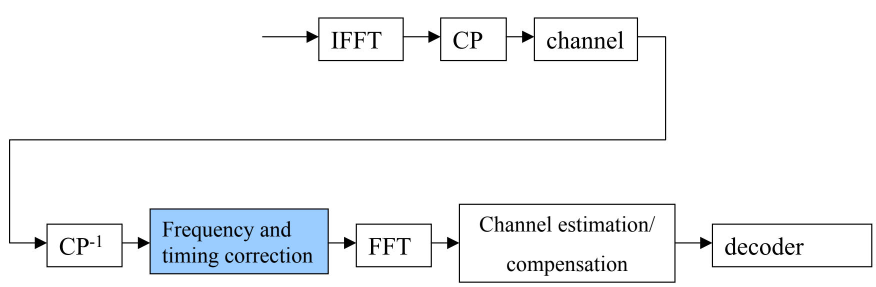
➢ Differential demodulation (ex: DAB)



In non-coherent communication, differential encoding/decoding avoids the use of channel estimation.

Receiver Architecture (2)

➢ Coherent demodulation (ex: DVB-T)



➢ Specificity of OFDM system w.r.t synchronization issue

•OFDM systems are much more sensitive to synchronization errors than single carrier systems.

•Synchronization algorithms suited to single carrier systems are inefficient for OFDM.

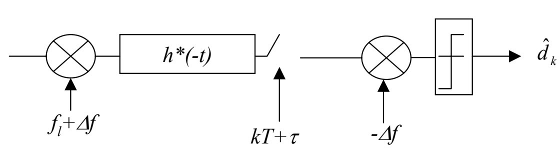
Impact of a synchronization error (1)

➢ System model (Gaussian channel)

–Carrier : n° l

–Frequency offset : Δf

–Timing error : τ



Impact of a synchronization error (2)

➢ Timing error τ

–$\tau<\Delta$-L : phase rotation (compensated by channel estimation/correction=

–$\tau>\Delta$-L : n^{th} symbol, carrier n° i

$$Y_{i,n} = e^{2j\pi(n/N)\tau} \frac{N-\tau}{N} X_{i,n} H_{i,n} + n_{i,n} + n_{\tau}(i,n)$$

➡ SNR loss

➡ ICI/ISI

Impact of a synchronization error (3)

➢ Frequency error : Δf

$Y_{m,l}$=p(Δf)exp[2jπ(m+1/2)ΔfT]d_{ml}+ICI

with

$$ICI = \sum_{n \neq l} \exp\left(2j\pi(k-l)(m+1/2)\right)\sin_c\left(\pi\left(n-l+\Delta fT\right)\right), \;\; p(\Delta f) = \sin_c\left(\pi \Delta fT\right)$$

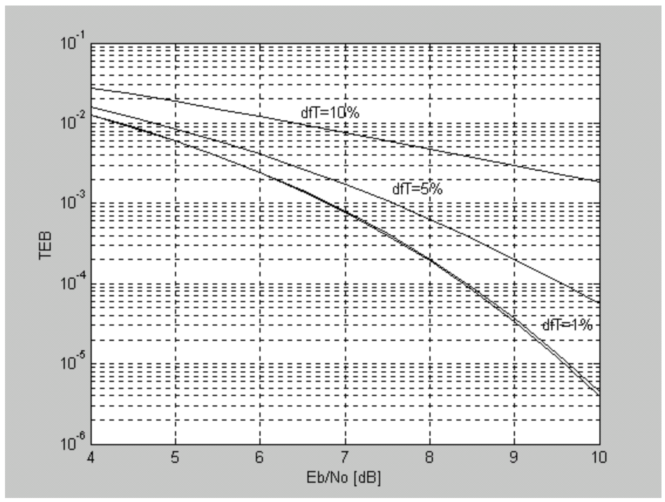
For $|\tau| < G$ (G: guard time)

$$\left|I_{n,i,k}\right| = \left|\frac{\sin\left[\pi\left\{(n-l)+\Delta f \times T\right\}\right]}{\pi\left[(n-l)+\Delta f \times T\right]}\right|$$

$$TEB = \frac{1}{4} erfc\left(\sqrt{\frac{E_b}{N_0}}\left|I_{n,n,k}\right| \times \left(1+2\frac{E_b}{N_0}\sum_{i \neq n}\left|I_{n,i,k}\right|^2\right)^{-\frac{1}{2}}\right)$$

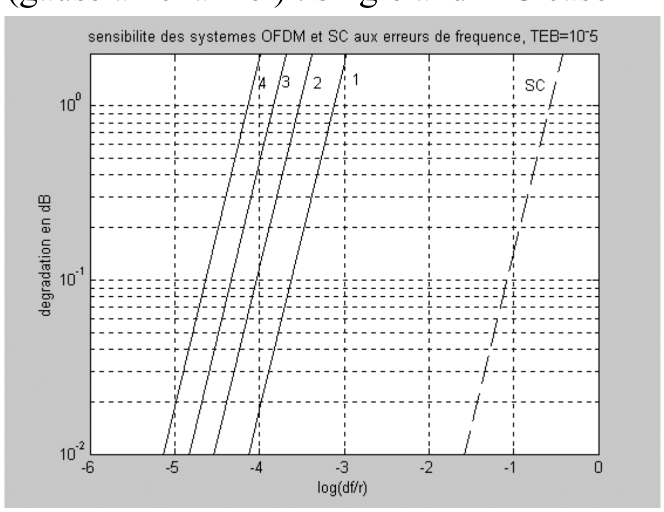
Impact of a synchronization error (4)

BER degradation due to a frequency error
(gaussian channel)



Impact of a synchronization error (5)

BER degradation due to a frequency error
(gaussian channel) : single and MC case



1: single carrier

2: OFDM, N=100

3: OFDM, N=256

3: OFDM, N=512

4: OFDM, N=1024

Impact of a synchronization error (6)

Impact of phase noise

$$D \approx \begin{cases} \dfrac{10}{\ln 10}\dfrac{11}{60}\left(4\pi N\dfrac{\beta}{R}\right)\dfrac{E_s}{N_O} & \text{(OFDM)} \\[2ex] \dfrac{10}{\ln 10}\dfrac{11}{60}\left(4\pi\dfrac{\beta}{R}\right)\dfrac{E_s}{N_O} & \text{(SC)} \end{cases}$$

β : 3 dB BW (SSB) in Wiener model

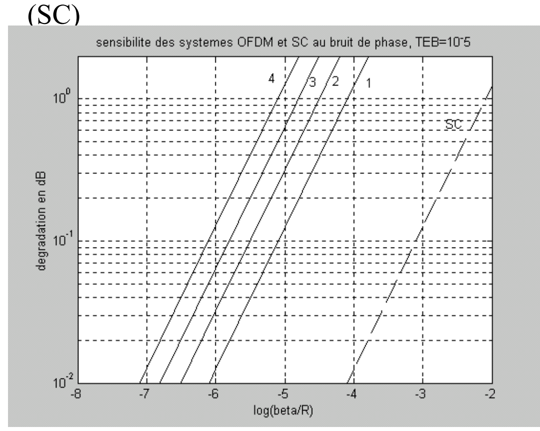
1: single carrier

2: OFDM, N=100

3: OFDM, N=256

3: OFDM, N=512

4: OFDM, N=1024



Timing/frequency estimators (1)

- Estimators using pilot symbols
  - Moose
  - Schmidl et Cox
- Estimators not using pilot symbols
  - Van de Beek
- These estimators are suited to frequency selective channels
  - Guard time is necessary for other reason
  - Each elementary channel (FFT output) is modelled by a different complex multiplicative coefficient.

Moose estimator (1)

- Principle : Emission of 2 identical OFDM symbols
- Timing has to be corrected first
- Hypothesis : the channel impulse response is constant over some OFDM symbols

Moose estimator (2)

First OFDM received symbol : $[r_0\ r_1 \ldots r_{N-1}]$

Second OFDM received symbol : $[r_N\ r_{N+1} \ldots r_{2N-1}]$

CIR constant over 1 OFDM symbol => $r_{n+N} = r_n \exp(2j\pi\Delta f N T_e) = r_n \exp(2j\pi\varepsilon)$

with $\varepsilon = 1/T$ (inter carrier spacing)

FFT output (first symbol) : $y(k) = \sum_{n=0}^{N-1} r_n \exp\left(2j\pi \frac{nk}{N}\right)$

FFT output (second symbol): $y(k+N) = \sum_{n=0}^{N-1} r_{n+N} \exp\left(2j\pi \frac{nk}{N}\right)$

$y(k+N) = y(k)\exp(2j\pi\varepsilon)$ $k \in \{0,1,\ldots,N-1\}$ =>The signal and ICI are affected exactly in the same way by the frequency offset.

Moose estimator (3)

MLE estimator:

$$\hat{\varepsilon} = \frac{1}{2\pi}\mathrm{Arg}\left\{\sum_{\mathrm{k=0}}^{\mathrm{N-1}} \mathrm{y(k+N)y}^*(k)\right\}$$

$$|\varepsilon| < 1 \;\Rightarrow\; |\Delta f| < \frac{1}{T} \;\Rightarrow -\frac{1}{2T} < \Delta f < \frac{1}{2T}$$

Frequency unbiguity has to be removed.

Schmidl et Cox estimator (1)

❑ Estimation of both timing and frequency errors

❑Principle:

➢ 2 dedicated pilot symbols

•First symbol : null odd carriers

•Second symbol : 2 interleaved PN sequences (odd/even carriers)

➢Estimation

•First symbol is used for timing and frequency estimation (2/T ambiguity)

•Second symbol is used to remove ambiguity on frequency estimation

Schmidl et Cox estimator (2)

First symbol : null odd carriers

$$y_n = \sum_{k=0}^{N-1} x_k \exp\left\{2j\pi \frac{nk}{N}\right\}$$
$$= \sum_{k=0}^{N/2-1} x_{2k} \exp\left\{2j\pi \frac{nk}{N/2}\right\}$$

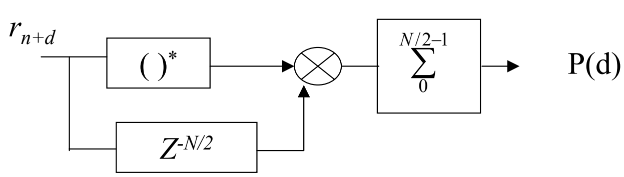
$y_{n+N/2}=y_n$ => OFDM symbols with 2 identical halves

Schmidl et Cox estimator (3)

Received OFDM symbol: $r_n, 0 \le n \le N-1$

Timing metric:

$$M(d) = \frac{|P(d)|^2}{(R(d))^2} \qquad R(d) = \sum_{m=0}^{N/2-1} |r_{d+m+N/2}|^2$$

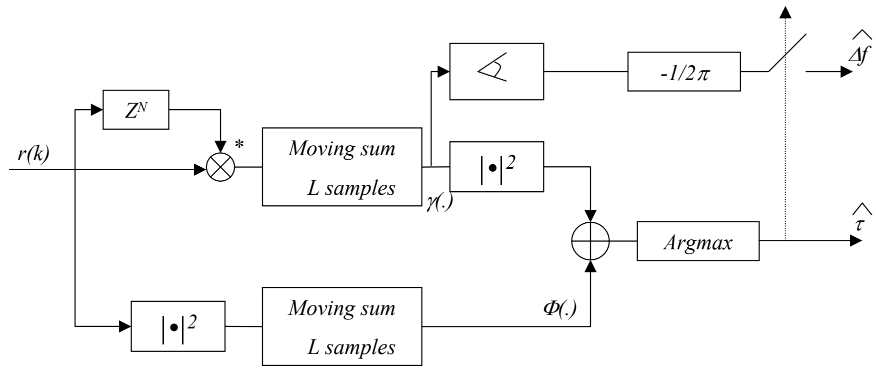


Timing estimate: $\hat{d} = \arg_d \{\max(M(d))\}$

Frequency estimate: $\hat{\varepsilon} = \frac{1}{\pi} angle\{P(\hat{d})\}$

$$|\varepsilon/2| < 1 \;\Rightarrow\; |\Delta f| < \frac{2}{T} \;\Rightarrow -\frac{1}{T} < \Delta f < \frac{1}{T}$$

Van de Beek estimator

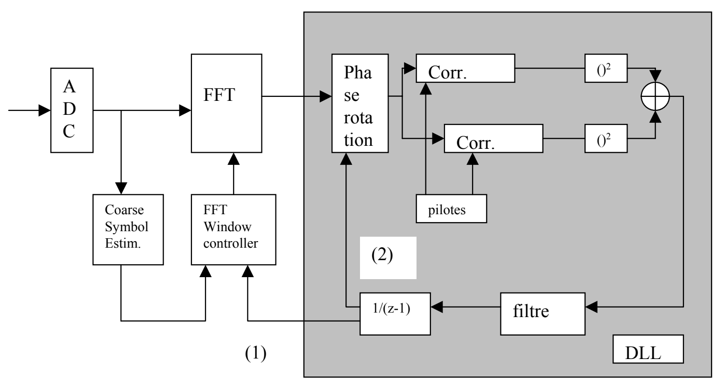


$$\hat{\tau}_{ML} = \arg\max\left\{\left|\gamma(\theta)\right| - \Phi(\theta)\right\}$$

$$\Delta\hat{f}_{ML} = -\frac{1}{2\pi}\measuredangle\gamma\left(\hat{\theta}_{ML}\right)$$

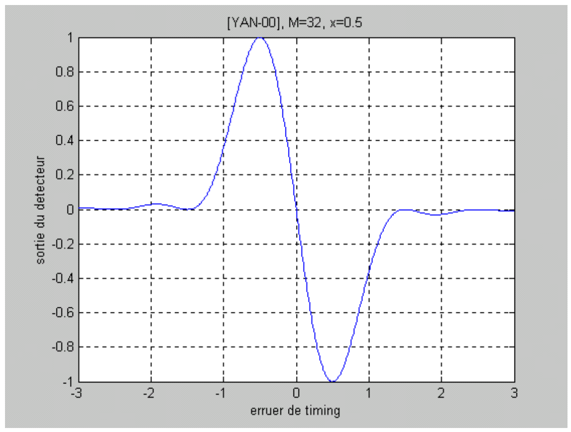
Yang estimator (timing) (1)

Idea : exploit the fact that a timing error introduces a phase error at the FFT output which depends on the carrier number.



Yang estimator (timing) (2)

$$S(\varepsilon,\xi)=\left(\frac{\sin(\pi(\varepsilon+\xi))}{M\sin[\pi(\varepsilon+\xi)/M]}\right)^2-\left(\frac{\sin(\pi(\varepsilon-\xi))}{M\sin[\pi(\varepsilon-\xi)/M]}\right)^2$$



Bibliography (1)

➢ <u>Impact of synchronization errors on performances</u>

[BOU-01] S Bougeard "Modélisation du bruit de phase des oscillateurs hyperfréquence et optimisation des systèmes de communications numériques", Thèse de Doctorat, INSA Rennes, décembre 2001.

[MOE-97] M Moeneclaey "The effect of synchronisation errors on the performance of orthogonal frequency-division multiplex (OFDM) systems", COST 254 conference, Toulouse, July 1997

[POL-94] T pollet, P Spruyt, M Moeneclaey "The BER performance of OFDM systems using non-synchronised sampling"

[POL-95] T Pollet, M Van Bladel, M Moeneclaey « BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise », IEEE on COM , fev/mars/avril 1995

[SPE-97] M Speth, F Classen, H Meyr "Frame synchronization of OFDM systems in frequency selective fading channels", VTC97

[YAN-00] B Yang, KB Letaief, RS Cheng, Z Cao "Timing recovery for OFDM transmission", IEEE JSAC, Novembre 2000

Bibliography (2)

➢ Synchronization algorithms suited to OFDM

[BEE-97] JJ Van De Beek, M sandell, PO Borjesson « ML estimation of time and frequency offset in OFDM systems », IEEE on signal processing, juillet 1997

[DAF-93] F Daffara, A Chouly “ML frequency detectors for orthogonal multicarrier systems”

[MOR-99] M Morelli, U Mengali “An improved frequency offset estimator for OFDM applications”, IEEE com letters, mars 1999

[MOO-94] PH Moose “A technique for orthogonal frequency division multiplexing frequency offset correction”,IEEE on COM, octobre 1994

[SCH-97]T Schmidl, D cox “robust frequency and timing synchronization for OFDM”, IEEE on COM, décembre 1997

[YAN-00] B Yang, KB Letaief, RS Cheng, Z Cao “Timing recovery for OFDM transmission”, IEEE JSAC, Novembre 2000

NEWCOM Autumn School : Estimation Theory for Wireless Communications

Training Based Channel Estimation: a Short Bibliography

Walid Hachem

1 CRB, ML Estimation, Training Sequence Design

Books

- Lehmann, E. L., and Casella, G., *"Theory of Point Estimation"*, Springer Texts in Statistics.
- Van Trees, H. L., *"Detection, Estimation, and Modulation Theory"*, John Wiley.

Articles

- Ayadi, J., and Slock, D.T.M., "Cramer-Rao bounds and methods for knowledge based estimation of multiple FIR channels", *SPAWC'97*, Paris, France.
- Ciblat, Ph., and Vandendorpe, L., "On the Maximum-Likelihood based data-aided frequency offset and channel estimates", EURASIP European Signal Processing Conference (EUSIPCO), vol. 1, pp. 627-630, Toulouse (France), September 2002.
- Gault S., Hachem W., and Ciblat, Ph., "Joint Sampling Clock Offset and Channel Estimation for OFDM Signals : CRB and Algorithms", to appear in *IEEE Transactions on Signal Processing*.
- Sezginer, S., and Bianchi, P., "Joint Frequency Offset and Channel Estimation in the OFDMA Uplink: Cramér-Rao Bounds and Training Sequence Design", *SPAWC'05*, New-York, USA.
- Stoica, P., and Besson, O., "Training Sequence Design for Frequency Offset and Frequency-Selective Channel Estimation", *IEEE Transactions on Communications*, Vol. 51, No. 11, Nov. 2003, pages 1910–1917.
- Vosoughi, A., and Scaglione, A., "Everything You Always Wanted to Know About Training: Guidelines Derived Using the Affine Precoding Framework and the CRB", to appear in *IEEE Transactions on Signal Processing*.

2 Information Theoretic Approach

Books

- Cover, T., and Thomas, J., *"Elements of Information Theory"*, Wiley.
- Gallager, R., *"Information Theory and Reliable Communication"*, Wiley.

Articles

- Biglieri, E., Proakis, J., and Shamai, S., "Fading channels: information-theoretic and communications aspects" *IEEE Transactions on Information Theory*, Volume 44, Issue 6, Oct. 1998 Page(s):2619–2692.
- Hassibi, B., and Hochwald, B.M., "How much training is needed in multiple-antenna wireless links ?", *IEEE Transactions on Information Theory*, Volume 49, Issue 4, April 2003 Page(s):951 - 963.
- Medard, M., "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel", *IEEE Transactions on Information Theory*, Vol 46, No. 3, May 2000, Page(s):933 –946.
- Ohno, S., and Giannakis, G.B., "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels", *IEEE Transactions on Information Theory*, Volume 50, Issue 9, Sept. 2004, Page(s):2138–2145.

Recent Developments on Multi-Channel Blind System Identification (BSI)

K. Abed-Meraim

ENST PARIS, Signal and Image Processing Dept.

abed@tsi.enst.fr

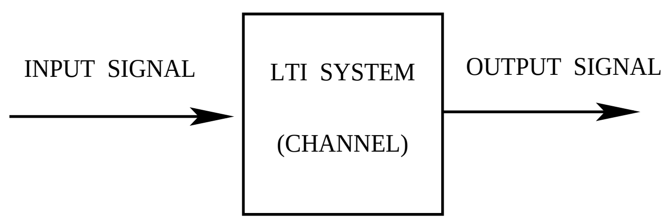
Presentation Outline

- Concepts and preliminaries
- BSI for SISO systems (mono-channel case)
- BSI for SIMO systems
- BSI for MIMO systems
- Concluding remarks

Blind System Identification:

Preliminaries

System identification



OBJECTIVE: Given the output signal and eventually certain side information (training sequence, physical or statistical information, partial channel knowledge, etc.), our objective is to estimate the channel (i.e., system transfer function) and restore the input signal.

Blind processing

We talk about *'BLIND PROCESSING'* in the situation where *'NO TRAINING SEQUENCE'* is available.

BSI $\Longleftrightarrow$ System identification using *only* the output data

Motivations:

- Increased channel throughput in communication systems.
- Robustness against channel modeling errors.
- Blind processing is necessary in certain applications (military applications, seismology, etc.)
- Flexibility and increased system autonomy.

Semi blind processing

Principle: Combining a data-aided (with training sequence) criterion J_{DA} with a blind criterion J_B, i.e:

$$J(h) = \alpha J_{DA}(h) + (1 - \alpha) J_B(h)$$

Criterion choice: The blind criterion should be chosen according to the context. The data-aided criterion is usually chosen as the maximum likelihood (=MMSE) one.

The optimal value of α can be computed based on asymptotic performance analysis (Buchoux et al 1999).

Result: Improve the estimation accuracy and/or shorten the training sequence size and hence increase the 'useful' channel throughput.

Identification versus deconvolution

- **Blind identification**: Estimation of the channel state information using the observation data and certain 'statistical' information on the source signal.

- **Blind deconvolution**: Estimation of the input or the channel inverse (equalizer) using *only* the output (observation) signal. This is also known as the blind equalization problem in communication application.

Channel model

- *Parametric versus non-parametric*: Channel can be simply modeled by certain physical or statistical parameters, e.g. the specular channel model based on the paths delays, attenuations and angle of arrival.

- *Instantaneous versus convolutive*: In communication, convolutive model occurs when the channel delay spread is larger than the symbol duration.

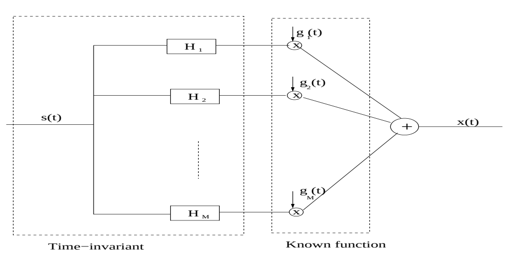
- *Finite (FIR) versus infinite impulse response (IIR) channel*: For long memories channels (this is the case for example in echo cancellation), one model the channel and an IIR one using for example statistical ARMA representation.

Channel model

- *Linear versus non-linear*: Linear-quadratic or post-linear channel models have been considered in the literature. The non-linearity may be due, for example, to amplificator saturation (e.g. satellite communication).
- *Stationary versus non-stationary*: Stationarity is a ‘good’ approximation over a ‘large’ observation period in most real-life applications. Non-stationary model has been considered, for example, in the over-the-horizon channel deconvolution problem.

How to cope with non-stationarity

- By using adaptive and tracking algorithms, e.g. LMS, RLS, PAST, etc.
- By using channel representation with known basis functions.

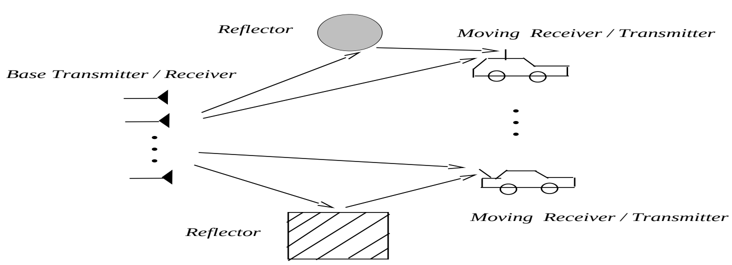


- By using time-frequency signal analysis.

Inherent ambiguities

- **Amplitude**: $y = h \star s = \lambda h \star \frac{1}{\lambda} s$.
- **Phase**: $y = h \star s = e^{j\theta} h \star e^{-j\theta} s$.
- **Delay**: In the stationary source case, $s(t)$ and $s(t+\tau)$ have the same statistical information.
- **Permutation**: This occurs in the multiple input case since the labeling of the source signals is arbitrary.

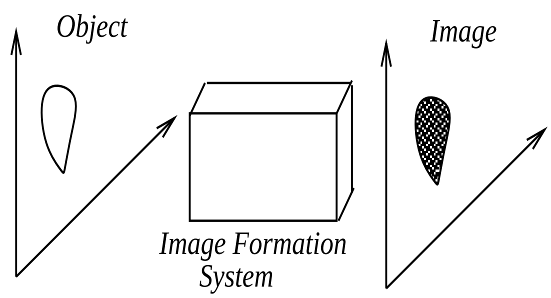
Application example: wireless communication



The objective here is to restore the transmitted signal that has been distorted by the propagation channel.

The blind processing helps in increasing the 'information data' throughput of the channel (for example, in GSM system the training data represents about 25% of transmitted data).

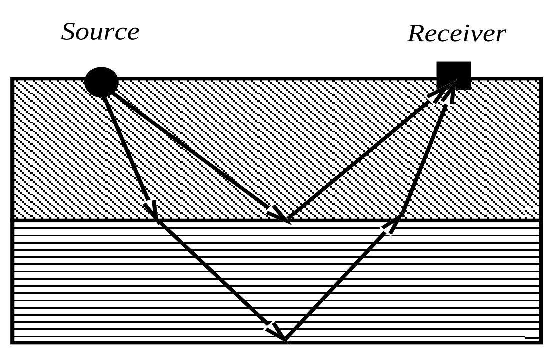
Application example: Image restoration



Objectives: From a blurred image retrieve the original one and/or the point spread function (channel).

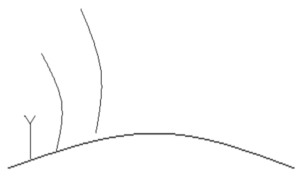
From several 'low quality' images form a 'high or improved quality' image.

Application example: Exploration seismology

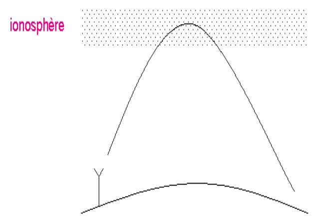


(Blind) channel estimation is used here to get information on the underground structure and the position of the reflectors.

Application example: Over-The-Horizon Radar (OTHR)

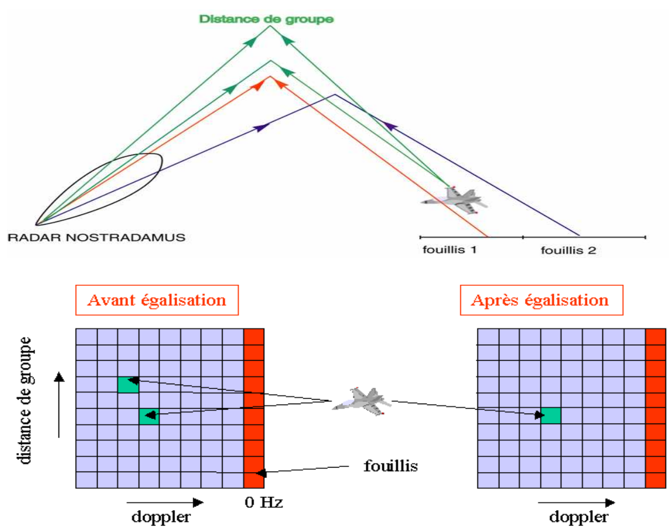


Classical radar: Limited horizon



OTHR: Early detection at all altitudes.

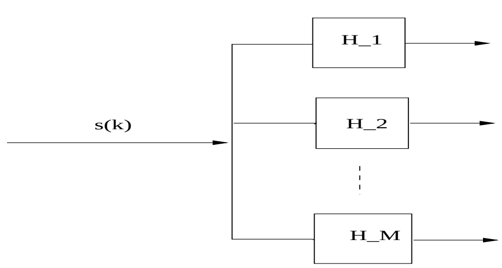
Application example: Over-The-Horizon Radar (OTHR)



Other potential applications

- Blind deconvolution for ultrasonic non-destructive testing (Nandi et al 1997, C.H. Chen et al 2002)
- ECG data processing (Sabry-Rizk et al., 1995): Fetal electro-cardiogram extraction,
- Acoustical and environmental robustness in automatic speech recognition (A. Acero et al 1993)
- Military applications, e.g. interference mitigation (M. Amin et al 1997), signal interception (Ph. Loubaton et al 2000), etc.

Multichannel processing: Diversity

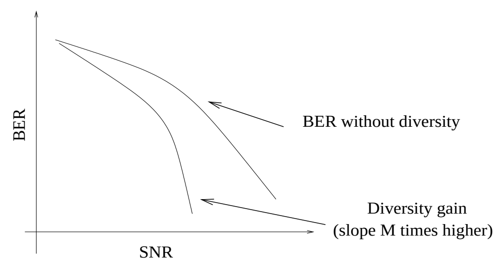


Multichannel processing is intimately linked to the concept of diversity:

Diversity: We would say that we have an order M diversity in the situation where we have several (M) replicas of the same input signal observed through M different and 'independent' channels.

Diversity gain

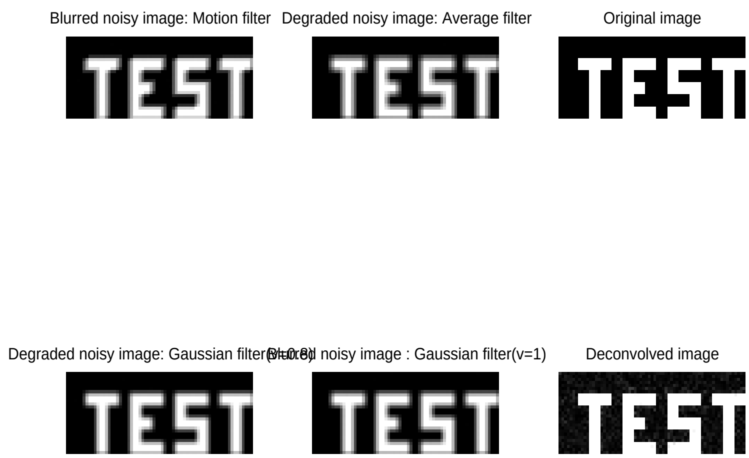
- **Improved restoration quality**: In communication, one can decrease the bit error rate (BER) by a factor of M (M being the diversity order)



- **Increased transmission rate**: The diversity increases the channel capacity.

Example: Monochannel image restoration

Example: Multichannel image restoration



Multichannel system: processing strategy

- *Separate processing*: Perform the blind deconvolution for each channel followed by a maximum ratio combiner of the channel outputs (simplicity, SNR gain but loss of the multichannel diversity).
- *Selective approach*: Deconvolution based only on the ‘best’ channel (simplicity but difficulty to define the best channel in the convolutive case).
- *Joint processing*: Process the channel outputs jointly in order to restore the input signal (leads to the best performance gain).

Channel type (system dimension)

We consider a linear time-invariant finite impulse response channel in the following three cases:

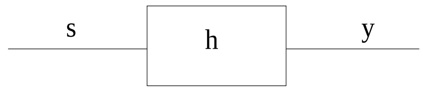
- *Single Input Single Output (SISO)* channel: This model is the most standard and the one considered first in the literature.

- *Single Input Multiple Output (SIMO)* channel: This is the situation for example when a multi-sensor antenna is used at the receiver.

- *Multiple Input Multiple Output (MIMO)* channel: This is an extension of the SIMO case when multiple users (signals) are considered. SIMO and MIMO cases have been studied extensively during the last decade.

Blind system identification for SISO channels

Data model



Noiseless observation:

$$y(k) = h(k) \star s(k)$$

$h(k)$ represents a LTI finite impulse response filter and $s(k)$ is a zero-mean stationary sequence of i.i.d (independent and identically distributed) *non-gaussian* random-variables of variance σ_s^2.

Need for higher order statistics (HOS)

- **Second order statistics (SOS) information**: The power spectral density of the output data is:

  $$S_y(f) = |H(f)|^2 \sigma_s^2$$

  $\Rightarrow$ No channel phase information from the observation SOS. BSI using the data SOS is only possible if the channel is of minimum phase.

- **HOS information**: Data HOS are needed to estimate the missing channel phase information.

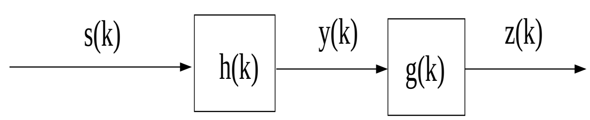
  $\Rightarrow$ the source signal must be *non gaussian*.

HOS-based BSI

- **Explicit HOS methods**: Direct system identification through explicit use of the signal HOS, e.g. 4th order cumulant-based methods (J.A. Cadzow et al 1996), polyspectra based methods (C. Nikias et al 1993, D. Hatzinakos et al 1991), etc.

- **Implicit HOS methods**: Identification of the channel inverse filter (equalizer) through optimization of appropriate non-linear cost functions, Sato algorithm (Y. Sato 1975), CMA algorithm (D. Godard, Treichler et al 1980), Bussgang algorithms (A. Benveniste et al 1980), etc.

Example of an explicit HOS method



Shavi-Weinstein method: Estimate the channel inverse filter $g(k)$ in such a way that we maximize the (absolute value) of its output $z(k)$ fourth order cumulant (under constant power constraint).

Idea: maximize the nongaussianity of $z(k)$ by maximizing its 4th order cumulant.

Example of an implicit HOS method

Constant Modulus Algorithm (CMA): Introduced in communication (initially) for constant modulus constellation signals:

$$g = \arg\min E(|z(k)|^2 - R)^2$$

Idea: Restore the constant modulus property of the source signal at the equalizer output.

General features of HOS-based methods

- In general, HOS based methods require large sample sizes to achieve ‘good’ estimation performances.
- Non-linear optimization techniques are needed to estimate the channel (or the inverse channel) parameters. Often, stochastic gradient techniques are used for the optimization.
- The HOS based criteria suffer from the existence of local-minima.
- Convergence analysis is possible only in the noiseless case.

Blind system identification

for SIMO channels

Motivation for multichannel processing

Blind deconvolution using SOS:

- *Single channel case*: Not possible unless the channel is minimum-phase. The minimum phase condition in the SISO case is a 'strong' condition that is, in general, not met in practice.

- *Multichannel case*: Almost always possible $\Rightarrow$ More robust and more accurate estimation. In fact, the minimum phase condition in the SIMO case is a 'mild' condition that is satisfied when the channels are sufficiently independent from one another.

Motivation for multichannel processing

More channel capacity in communication systems

- *Single channel case*:

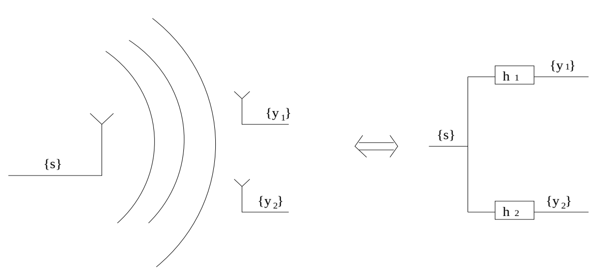
$$C = \log_2(1+\rho)$$

- *Multichannel case*: (M transmit and receive channels)

$$C = \log_2 \det(1 + \frac{\rho}{M}\mathbf{H}\mathbf{H}^H) \overset{M\to\infty}{\longrightarrow} M \log_2(1+\rho)$$

The capacity gain comes from the fact that having several replicas of the transmitted signal observed through independent channels reduces significantly the risk of information loss.

Space Diversity (Multiple receivers)

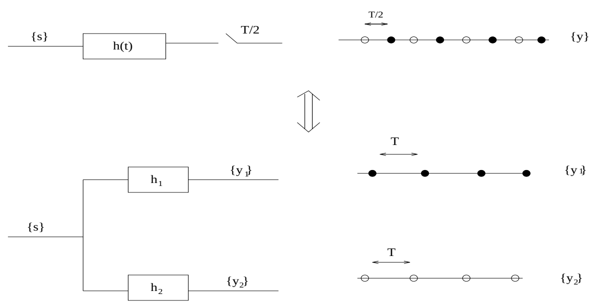


- $s(n)$: the source signal.
- $h_i(z)$: models the propagation between the emitting source and the i-th sensor.

$$h_i(z) = \sum_k h_i(k) z^{-k}$$

- $y_i(n)$: output at the i-th sensor.

Time diversity (oversampling)



$$y(t) = \sum_k h(t - kT)s(k) \quad \textbf{cyclostationary}$$

$\Rightarrow$ Exploit the cyclostationarity (time diversity) by oversampling wrt the symbol duration.

Time diversity (oversampling)

By oversampling we have multiple 'virtual' channels:

$$\begin{cases} h_1(z) &= \sum_k h(kT)z^{-k} \\ h_2(z) &= \sum_k h(kT + T/2)z^{-k} \end{cases}$$

The cyclostationary oversampled signal can be represented as a *stationary multivariate* signal as:

$$\begin{cases} y_1(k) &= x(kT) \\ y_2(k) &= x(kT + T/2) \end{cases} \quad \textbf{stationary}$$

Multichannel model

$$\left\{\begin{array}{lcl} y_1(k) & = & s(k) * h_1(k) \\ y_2(k) & = & s(k) * h_2(k) \\ \vdots & & \\ y_M(k) & = & s(k) * h_M(k) \end{array}\right. \qquad k = 0, \cdots, N-1$$

- $s(n)$: single unknown source signal.
- To each output i corresponds the FIR transfer function $h_i(z)$

$$h_i(z) = \sum_{k=0}^{L} h_i(k) z^{-k}$$

Multichannel model

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_M \end{bmatrix} \begin{bmatrix} s(-L) \\ \vdots \\ s(N-1) \end{bmatrix} = \mathbf{Hs}$$

s is the input vector, $\mathbf{y}_i$ is the observation vector at sensor i and $\mathbf{H}_i$ is the $N \times (N+L)$ Sylvester matrix

$$\mathbf{H}_i = \begin{bmatrix} h_i(L) & \cdots & h_i(0) & \cdots & 0 \\ \vdots & \ddots & & \ddots & \vdots \\ 0 & \cdots & h_i(L) & \cdots & h_i(0) \end{bmatrix}$$

Some properties of SIMO systems

- Weak *minimum phase condition*

$$\mathbf{h}(z) \stackrel{\text{def}}{=} \begin{bmatrix} h_1(z) \\ \vdots \\ h_M(z) \end{bmatrix} \neq 0 \;\; \text{for } |z| > 1$$

Satisfied as soon as $h_i(z),\ 1 \leq i \leq M$ do not share common zeros.

- **Left invertible system**: as soon as the $MN \times (N+L)$ matrix $\mathbf{H}$ is full column rank, i.e. when we have more equations than unknowns.

Some properties of SIMO systems

- **Finite zero-forcing inverse filters**: if $\mathbf{h}(z) \neq 0,\ \forall\, z$, $\exists\ \mathbf{g}(z) = [g_1(z), \cdots, g_M(z)]$ a polynomial vector such that:

$$\underbrace{\mathbf{g}(z)\mathbf{h}(z) = \sum_{k=1}^{M} g_i(z) h_i(z) = 1}_{\textbf{Bezout equality}}$$

- **Exact identification** in the noiseless case from a finite sample size observation vector.

SIMO versus SISO

- FIR equalizer for SIMO versus IIR equalizer for SISO.
- Causal equalizer for SIMO versus non-causal equalizer for SISO.
- Exact estimation using finite sample size for SIMO (not possible for SISO).
- Equalizer delay plays an important role in SIMO case and not in the SISO case.
- SOS-based BSI for SIMO versus HOS-based BSI for SISO.

Strict identifiability

Definition

The system is *strictly identifiable* if a given output $\mathbf{y}$ implies a unique input $\mathbf{s}$ and a unique system matrix $\mathbf{H}$ up to an unknown scalar, i.e.,

$$\mathbf{H}'\mathbf{s}' = \mathbf{H}\mathbf{s} \Longrightarrow \mathbf{s}' = \alpha\mathbf{s} \ \text{ and } \ \mathbf{h}'(z) = \frac{1}{\alpha}\mathbf{h}(z)$$

where α is a given non-zero scalar.

Strict identifiability

Necessary condition: The system is identifiable only if the followings are true:

$$\begin{cases} \mathbf{h}(z) \neq 0, \ \forall z \\ p \geq L+2 \\ N \geq L+2 \end{cases}$$

where p is the number of modes in the input sequence.

Strict identifiability

Sufficient condition : The system is identifiable if the followings are true:

$$\begin{cases} \mathbf{h}(z) \neq 0, \ \forall z \\ p \geq 2L+1 \\ N \geq 3L+1 \end{cases}$$

Strict identifiability

The identifiability conditions shown above essentially ensure the following intuitive requirements:

- *All channels in the system must be different enough from each other*. They can not be identical, for example.
- *The input sequence must be complex enough*. It can not be zero, a constant or a single sinusoid, for example.
- *There must be enough number of output samples available*. A set of available data can not yield sufficient information on a larger set of unknown parameters, for example.

Estimation techniques

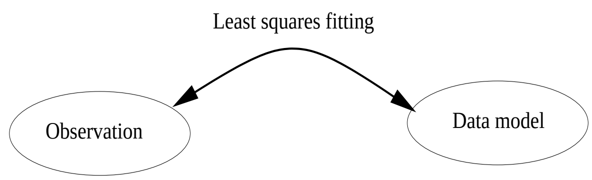
- Direct estimation of **system function**:
  - Maximum likelihood (ML) method.
  - Cross-relations (CR) method.
  - Channel subspace (CS) method.
- Direct estimation of **system input**:
  - Signal subspace (SS) method.
  - Mutually referenced equalizers (MRE) method.
  - Linear prediction (LP) method.

Maximum likelihood method

Principle: Assuming a circular white Gaussian noise vector $\mathbf{w}$

$$p(\mathbf{y}) = \frac{1}{\pi^N \sigma^{2N}} \exp(-\frac{1}{\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2)$$

Thus the ML estimate is given by $\boxed{(\mathbf{H}, \mathbf{s})_{ML} = \arg\min_{\mathbf{H},\mathbf{s}} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2}$



Maximum likelihood method

Separable problem: Minimize over $\mathbf{s}$:

$$\mathbf{s}_{ML} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}$$

Then over $\mathbf{H}$:

$$\mathbf{H}_{ML} = \arg\min_{\mathbf{H}} \|\mathbf{P}_H^{\perp}\mathbf{y}\|^2$$

$\mathbf{P}_H^{\perp}$ = orthogonal projection matrix onto Range$(\mathbf{H})^{\perp}$.

Orthogonal Complement Matrix (OCM)

Idea: One can obtain noise vectors by observing that

$$[0, \cdots, \overset{\text{i-th}}{-h_j(z)}, 0, \cdots, 0, \overset{\text{j-th}}{h_i(z)}, \cdots, 0] \begin{bmatrix} h_1(z) \\ \vdots \\ h_M(z) \end{bmatrix} = 0$$

Result (Y. Hua 1995) : One can form an OCM $\mathbf{G}$ that is a linear function of the channel parameters such that its column vectors form a basis of the noise subspace, i.e.

$$\mathbf{P}_G = \mathbf{P}_H^{\perp}$$

Two step estimation technique

ML criterion:

$$\mathbf{h}_{ML} \quad = \quad \arg\min_{\|\mathbf{h}\|=1} \mathbf{y}^H \mathbf{G}(\mathbf{G}^H\mathbf{G})^{\#}\mathbf{G}^H\mathbf{y}$$

where $\mathbf{h}$ is the vector of all channels' impulse responses. From the *commutativity property* of linear convolution:

$$\mathbf{G}^H\mathbf{y} = \mathbf{Y}\mathbf{h}$$

we obtain

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}} \mathbf{h}^H\mathbf{Y}^H(\mathbf{G}^H\mathbf{G})^{\#}\mathbf{Y}\mathbf{h}$$

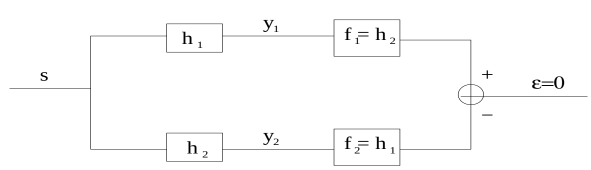
Two step estimation technique

Two Step Maximum Likelihood (TSML):

1. $\mathbf{h}_c = \arg\min\ \mathbf{h}^H \mathbf{Y}^H \mathbf{Y} \mathbf{h}$
2. $\mathbf{h}_e = \arg\min\ \mathbf{h}^H \mathbf{Y}^H (\mathbf{G}_c^H \mathbf{G}_c)^{\#} \mathbf{Y} \mathbf{h}$, where $\mathbf{G}_c$ is $\mathbf{G}$ constructed from $\mathbf{h}_c$.

At each step the solution is given by the least eigenvector associated to the least eigenvalue of the considered quadratic form.

Cross-relations (CR) method



- **Principle**: For every pair of channels, we have

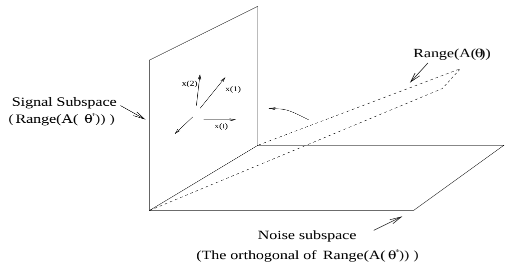
$$y_i(k) * h_j(k) = y_j(k) * h_i(k)$$

- **Algorithm**: By collecting all possible pairs of M channels, one can easily establish a set of linear equations:

$$\mathbf{Y}\mathbf{h} = 0$$

This yields to $\mathbf{h}_{CR} = \arg\min \mathbf{h}^H \mathbf{Y}^H \mathbf{Y} \mathbf{h}$ (first step of TSML).

Subspace method



Principle: Assume the following model: $\mathbf{x}(n) = \mathbf{A}(\theta)\mathbf{s}(n)$ with

$$\mathrm{Range}(\mathbf{A}(\theta)) = \mathrm{Range}(\mathbf{A}(\theta')) \Longleftrightarrow \theta = \theta'$$

Thus, θ can be estimated as:

$$\hat{\theta} = \arg\min_{\theta} \mathrm{d}(\mathrm{Range}\{\mathbf{x}(n)\},\ \mathrm{Range}(\mathbf{A}(\theta)))$$

Channel subspace (CS) method

- **Model**:

$$\begin{aligned}
\mathbf{y}(n) &= \mathbf{H}\mathbf{s}(n)\ n = 0, ..., N - W \\
\mathbf{y}(n) &= [\mathbf{y}_1^T(n), \cdots, \mathbf{y}_M^T(n)]^T \\
\mathbf{y}_i(n) &= [y_i(n), \cdots, y_i(n + W - 1)]^T
\end{aligned}$$

In our case: $\mathbf{A} \longleftrightarrow \mathbf{H}$ and $\theta \longleftrightarrow \mathbf{h}$.

- **Main result**: If $W \geq L + 1$ and the M channels do not share a common zero, then

$$\mathrm{Range}(\mathbf{H}) = \mathrm{Range}(\mathbf{H}') \Longleftrightarrow \mathbf{h}' = \alpha\mathbf{h}$$

where α is a scalar constant.

CS algorithm

- Estimate the signal (resp. noise) subspace as the principal (resp. minor) eigen-subspace of the data covariance matrix $\mathbf{R}_y$:

$$\mathbf{R}_y = \sum_n \mathbf{y}(n)\mathbf{y}^H(n) = [\mathcal{E}_s \ \mathcal{E}_n] \begin{bmatrix} \mathbf{\Lambda}_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{E}_s^H \\ \mathcal{E}_n^H \end{bmatrix}$$

where $\quad \mathrm{Range}(\mathcal{E}_s) = \mathrm{Range}(\mathbf{H}) \perp \mathrm{Range}(\mathcal{E}_n).$

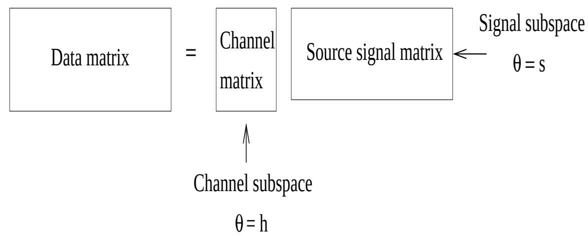
- Compute the least square error solution to

$$\mathbf{h}_{CS} = \arg \min_{\|\mathbf{h}\|=1} \|\mathcal{E}_n^H \mathbf{H}\|^2.$$

Comparison of the ML, CR, and CS methods

- **ML method**
  - Large computational cost
  - Very good estimation accuracy
- **CR method**
  - Low computational cost
  - Moderate estimation accuracy
- **CS method**
  - Moderate computational cost
  - Good estimation accuracy

Signal Subspace (SS) method



- **Model**: $\mathbf{Y} = [\mathbf{y}(0), \cdots, \mathbf{y}(N-W)] = \mathbf{HS}$
  $\mathbf{S}$ being a the source signal matrix of Hankel structure.

- **Principle**: $\mathbf{A} \longleftrightarrow \mathbf{S}$ and $\theta \longleftrightarrow \mathbf{s}$

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} = \mathrm{d}(\mathrm{Row}(\mathbf{Y}),\ \mathrm{Row}(\mathbf{S}))$$

Signal Subspace (SS) method

- **Result** (Xu et al 1995): Assume that $\mathbf{H}$ is full column rank and that the input sequence $\{s(n)\}_{-L \leq n \leq N-1}$ contains more than $W+L+1$ modes, then

$$\mathrm{Row}(\mathbf{S}) = \mathrm{Row}(\mathbf{S}') \Longleftrightarrow \mathbf{s}' = \alpha \mathbf{s}$$

  where α is a scalar constant.

SS algorithm

- Perform the SVD of the data matrix $\mathbf{Y} = [\mathbf{y}(0), \cdots, \mathbf{y}(N-W)]$

$$\mathbf{Y} = \mathbf{U} \begin{bmatrix} \mathbf{\Sigma}_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}$$

  $\mathbf{V}_n$ is the orthogonal matrix to the row space of $\mathbf{S}$

$$\mathbf{V}_n \mathbf{S}^H = 0$$

- Estimate $\mathbf{s}$ by minimizing the quadratic criterion

$$\hat{\mathbf{s}} = \arg \min_{\|\mathbf{s}\|=1} \|\mathbf{V}_n \mathbf{S}^H\|^2$$

Blind equalization

- **Definition**: $\mathbf{g}(z)$ is a blind equalizer iff:

$$\mathbf{g}(n) * \mathbf{y}(n) = \alpha s(n-m) \iff \mathbf{g}(z)\mathbf{h}(z) = \alpha z^{-m}$$

- **Characterization**:
  - *Statistical criterion*: If $s(n)$ is i.i.d.

$$\mathbf{g}(z) \longrightarrow \hat{s}(n) = \mathbf{g}(n) * \mathbf{y}(n) \text{ is i.i.d.}$$

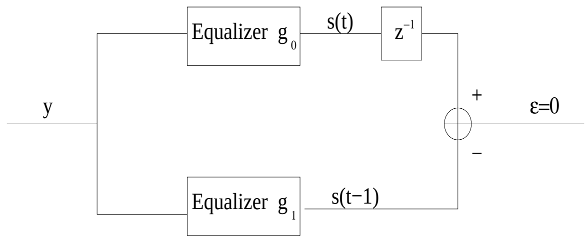
    e.g., Linear prediction , Bussgang , etc.

  - *Geometrical criterion*: If $s(n) \in \mathcal{A}$

$$\mathbf{g}(z) \longrightarrow \hat{s}(n) = \mathbf{g}(n) * \mathbf{y}(n) \in \mathcal{A}$$

    e.g., CMA algorithms.

Mutually referenced equalizers (MRE) method



MRE relations: Let $\mathbf{g}_i(n)$ $i = 0, \cdots, W + L - 1$ be equalizer filters satisfying

$$\mathbf{g}_i(n) \star \mathbf{y}(n) = \alpha s(n - i),\ i = 0, 1, \cdots$$

Then, filters $\mathbf{g}_i$ should satisfy (MRE relations):

$$\mathbf{g}_i \star \mathbf{y}(n) = \mathbf{g}_{i+1} \star \mathbf{y}(n + 1)$$

Mutually referenced equalizers (MRE) method

- **Result** (D. Gesbert et al 1994) : Vice versa, the previous relations characterize uniquely the equalizer filters, i.e. if $\mathbf{g}_0, \cdots, \mathbf{g}_{d-1}$ ($d = W + L$) satisfy the MRE relations, then

$$\mathbf{g}_i(n) \star \mathbf{y}(n) = \alpha s(n - i), \quad \forall\ i$$

- **Algorithm**: $\{\mathbf{g}_i\}$ are estimated by minimizing (under a suitable constraint) the quadratic criterion

$$J = \sum_{n,i} \|\mathbf{g}_i \star \mathbf{y}(n) - \mathbf{g}_{i+1} \star \mathbf{y}(n + 1)\|^2$$

Linear prediction (LP) method

- **Model**:

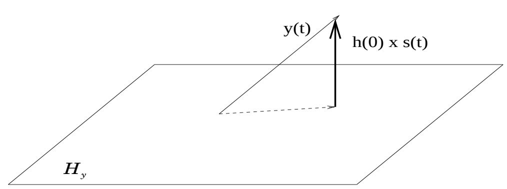
$$\begin{cases} \mathbf{y}(n) = [y_1(n), \cdots, y_M(n)]^T = [\mathbf{h}(z)]s(n), \quad s(n): \;\; i.i.d \\ \mathbf{h}(z) = [h_1(z), \cdots, h_M(z)]^T \neq 0, \forall\, z \end{cases}$$

- **Principle**: **Bezout equality**: $\exists\, \mathbf{g}(z) = [g_1(z), \cdots, g_M(z)]$ such that $\mathbf{g}(z)\mathbf{h}(z) = 1$

$$\Longrightarrow [\mathbf{g}(z)]\mathbf{y}(n) = s(n)$$

- **Result**: $\mathbf{y}(n)$ is an AR process of order L. Its innovation process is $\mathbf{i}(n) = \mathbf{h}(0)s(n)$.

LP algorithm



Prediction (projection) subspace

- Estimate the prediction coefficients of $\mathbf{y}(n)$ by solving the Yule-Walker equations:

$$\mathbf{y}(n) + \sum_{k=1}^{P} \mathbf{A}(k)\mathbf{y}(n-k) = \mathbf{i}(n) = \mathbf{h}(0)s(n)$$

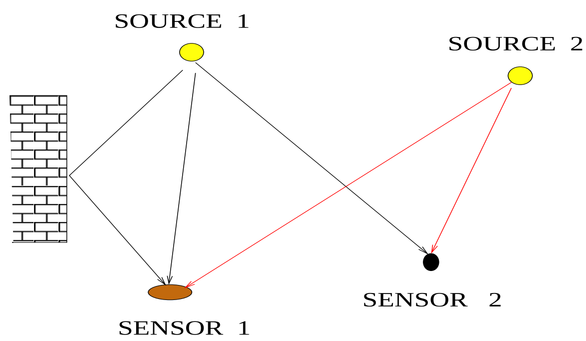
- Estimate vector $\mathbf{h}(0)$ (up to a constant) as the principal eigenvector of the innovation covariance matrix $\mathbf{D} = E(\mathbf{i}(n)\mathbf{i}(n)^H) = \mathbf{h}(0)\mathbf{h}(0)^H$.

Comparison of the SS, MRE, and LP methods

- **SS method**
  - Large computational cost
  - Good estimation accuracy
  - Deterministic input
- **MRE method**
  - Moderate computational cost
  - Good estimation accuracy
  - Deterministic input
- **LP method**
  - Low computational cost
  - Moderate estimation accuracy
  - Stochastic decorrelated input

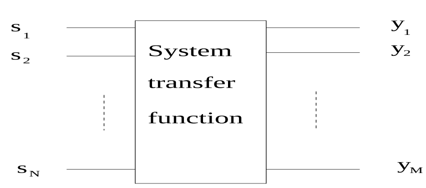
Blind system identification for MIMO channels

The situation of interest



M different (possibly noisy) linear combinations of N independent source signals are observed at the sensors.

Convolutive linear mixture model



$$\mathbf{y}(n) = \mathbf{H}(n) \star \mathbf{s}(n)$$

- $\mathbf{y}(n)$: $M \times 1$ observation vector (array output),
- $\mathbf{s}(n)$: $N \times 1$ *unknown* source vector,
- $\mathbf{H}(z) = \sum_n \mathbf{H}(n) z^{-n}$: $M \times N$ *unknown* transfer function matrix assumed, in general, of finite impulse response, i.e. deg$(\mathbf{H}(z)) = L$.

Basic assumptions

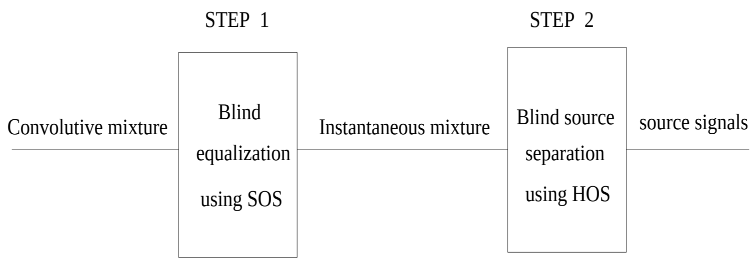
- **System dimension**: We assume here strictly *more sensors than sources*, i.e. $M > N$.
- **Source signals**: They are assumed to be *mutually independent* stationary random processes.
- **System matrix**: The transfer function $\mathbf{H}(z)$ is assumed to be *irreducible* (rank$(\mathbf{H}(z)) = N$ for all z) and *column reduced* (rank$(\mathbf{H}(L)) = N$).

Objectives

- **SIMO case**: In the SIMO case we have to get rid of the inter-symbol interference (ISI) only *(blind equalization problem)*.
- **MIMO case**: In the MIMO case we have to get rid of the ISI *(blind equalization problem)* and to get rid also of the inter-user interferences *(blind source separation (BSS))* .

MIMO blind deconvolution $\Longleftrightarrow$ Blind equalization + BSS.

Deconvolution approach



- **Step 1**: Blind equalization using second order statistics. This step transforms the convolutive mixture into an instantaneous mixture.
- **Step 2**: Application of a BSS algorithm (using, in general, the data HOS) to the instantaneous mixture obtained at the previous step.

Other possible deconvolution approaches

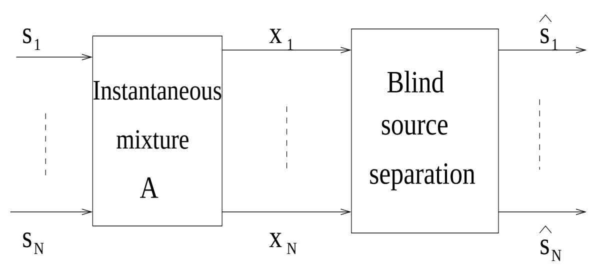
- Blind identification and deconvolution using HOS (Nikias et al 1993, Liu et al 2002, etc.).
- Blind separation followed by M parallel SIMO blind equalization (Bousbia-Salah et al 2000).
- Joint blind equalization and source separation by decorrelation (Y. Hua et al 2000).
- Iterative blind deconvolution with interference cancellation (Delfosse et al 1996).

First step: Blind equalization

The same algorithms (except for certain details) for SIMO blind identification can be applied to MIMO identification.

However, in the SIMO case we estimate the channel transfer function (resp. the source signal) up to a 1×1 constant factor α, i.e. $\hat{\mathbf{h}}(z) = \mathbf{h}(z)\alpha$, while in the MIMO case we estimate the channel transfer function (resp. the source vector) up to a $N \times N$ constant matrix $\mathbf{A}$, i.e. $\hat{\mathbf{H}}(z) = \mathbf{H}(z)\mathbf{A}$.

Second step: Blind source separation



Instantaneous linear mixture model:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

BSS versus ICA (Independent Component Analysis)

1. *BSS = signal synthesis*: Identify the mixture matrix and/or recover the input signals from the observed signal by exploiting the *statistical independence* or other features of the sources.

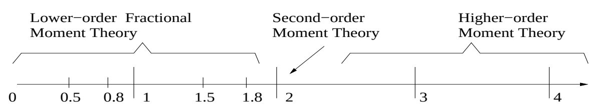
2. *ICA = signal analysis*: Analyse a multi-variate signal by decomposing it into a set of independent components (independent component analysis ICA).

ICA versus PCA

- **Principal component analysis**: seeks directions in feature space that best represent the data in *least squares* sense.

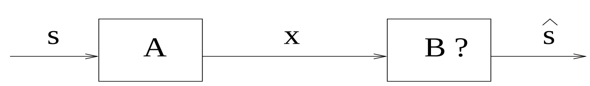
- **Independent component analysis**: seeks directions in feature space that are most *independent* from one another.

BSS approaches



- *HOS-based methods*: Exploit the observations higher order statistics either explicitely by processing their higher order cumulants or implicitely through the optimization of non-linear functions given by information-theoretic criteria.
- *SOS-based methods*: When the sources are 'temporally colored', one can achieve BSS using signal decorrelation.
- *FLOM-based methods*: Dedicated to the separation of impulsive signals, e.g. alpha-stable signals (these signals have infinite 2nd and higher order moments).

Information theoretic principles



$\mathbf{B}$ is computed such that its outputs are most independent from one another:

- By minimizing the mutual information between the components of $\hat{\mathbf{s}}(t)$.
- By minimizing the Kullbak-Leibler distance in between the pdf of $\hat{\mathbf{s}}(t)$ and the product of its components pdfs, i.e.

$$KL(p(\hat{\mathbf{s}}(t)), \prod_k p_k(\hat{s}_k(t)).$$

- My maximizing the nongaussianity of $\hat{\mathbf{s}}(t)$ (measures of nongaussianity include the Kurtosis -fourth order cumulant- and the Negentropy -differential entropy-).

BSS by decorrelation

Basic assumptions:

- The mixing matrix $\mathbf{A}$ is full column rank.

- The sources are temporally coherent but mutually uncorrelated, i.e.,

$$\begin{aligned}
\mathbf{R}_s(\tau) &\stackrel{\text{def}}{=} E(\mathbf{s}(t+\tau)\mathbf{s}(t)^H) = \begin{bmatrix} \rho_1(\tau) & & 0 \\ & \ddots & \\ 0 & & \rho_n(\tau) \end{bmatrix} \\
\mathbf{R}_x(\tau) &= \mathbf{A}\mathbf{R}_s(\tau)\mathbf{A}^H
\end{aligned}$$

Separation by decorrelation

- *Principle*: $\mathbf{B} = \mathbf{A}^{-1}$ is the linear transform that decorrelate the signal components at all time lags, i.e.

$$\mathbf{B}\mathbf{R}_x(\tau)\mathbf{B}^H = \mathbf{R}_s(\tau)$$

is diagonal for all τ.

- *A two step procedure*:
  - Data *whitening*: The whitening matrix transforms $\mathbf{A}$ into a unitary matrix.
  - *Diagonalization*: Estimate the unitary matrix by diagonalizing the non-zero lag correlation matrices.

Whitening

Whitening Matrix: Let $\mathbf{W}$ denotes a $n \times m$ matrix, such that

$$(\mathbf{WA})(\mathbf{WA})^H = \mathbf{UU}^H = \mathbf{I}$$

$\mathbf{W}$ can be computed as an inverse square root of covariance matrix of the observation vector (assuming unit-power sources).

Whitened correlations: Defined as

$$\underline{\mathbf{R}}_x(\tau) \quad = \quad \mathbf{W}\mathbf{R}_x(\tau)\mathbf{W}^H = \mathbf{U}\mathbf{R}_s(\tau)\mathbf{U}^H$$

Diagonalization

- Diagonalization of **one** single normal matrix $\mathbf{M}$
$\Longleftrightarrow$ Minimizing under unitary transform the sum of squared moduli of the off-diagonal elements. This is equivalent to the maximization under unitary transform $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_n]$ the sum of the squared moduli of the diagonal elements:

$$C(\mathbf{M}, \mathbf{V}) = \sum_i |\mathbf{v}_i^* \mathbf{M} \mathbf{v}_i|^2$$

- For **a set** of d matrices:

$$C(\mathbf{V}) = \sum_{k=1}^{d} C(\mathbf{M}_k, \mathbf{V}) = \sum_{k,i} |\mathbf{v}_i^* \mathbf{M}_k \mathbf{v}_i|^2$$

$\Longrightarrow$ **Joint diagonalization criterion**.

Identifiability

Objectives: Given a set of K correlation matrices $\mathbf{R}_x(\tau_1), \cdots, \mathbf{R}_x(\tau_K)$ answer the following:

- Is it possible to separate the sources given this statistics?
- If no, what it the best we can do (partial identifiability)?
- Is it possible to test the identifiability condition?

Theorem 1: Identifiability

- Define for each source i

  $$\boldsymbol{\rho}_i \quad = \quad [\rho_i(\tau_1), \rho_i(\tau_2), \cdots, \rho_i(\tau_K)] \quad \text{and} \quad \tilde{\boldsymbol{\rho}}_i = [\Re(\boldsymbol{\rho}_i), \Im(\boldsymbol{\rho}_i)]$$

  Then, BSS can be achieved using the output correlation matrices at time lags $\tau_1, \tau_2, \cdots, \tau_K$ iff $\forall\ i \neq j$

  $$\tilde{\boldsymbol{\rho}}_i \ \textit{and}\ \tilde{\boldsymbol{\rho}}_j \ \textit{are}\ \underline{\textit{(pairwise)}}\ \textit{linearly independent}$$

- If this condition is satisfied then $\mathbf{B}$ is a separating matrix iff $\forall\ i \neq j$

  $$r_{ij}(k) = 0 \quad \textit{and} \quad \sum_{k=\tau_1}^{\tau_K} |r_{ii}(k)| > 0 \tag{1}$$

  where $r_{ij}(k) \stackrel{\text{def}}{=} E(z_i(t+k)z_j^*(t))$, $\mathbf{z} = \mathbf{Bx}$ and $k = \tau_1, \tau_2, \cdots, \tau_K$.

Discussion

- Theorem 1 gives a **necessary and sufficient** condition to achieve BSS.
- It is possible to separate the sources from only **one** correlation matrix.
- $K \longrightarrow \infty \implies$ 2 sources are separable iff they have **different spectral shape**.
- It is well known that HOS methods can achieve BSS when no more than one Gaussian source is present. In contrast, SOS methods can achieve BSS when no more than one temporally white source is present.

Theorem 2: Partial Identifiability

Assume there are d distinct groups of sources each of them containing d_i source signals with same (up to a scalar) correlation vector $\tilde{\boldsymbol{\rho}}_i$, $i = 1, \cdots, d$, i.e., $\mathbf{s} = [\mathbf{s}_1^T, \cdots, \mathbf{s}_d^T]^T$.

Let $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$ be an $m \times 1$ random vector satisfying equation (1).

Then, there exists a permutation matrix $\mathbf{P}$ and non-singular matrices $\mathbf{U}_i$ such that

$$\begin{aligned} \mathbf{P}\mathbf{z}(t) &= [\mathbf{z}_1^T(t), \cdots, \mathbf{z}_d^T(t)]^T \\ \mathbf{z}_i(t) &= \mathbf{U}_i \mathbf{s}_i(t) \end{aligned}$$

Moreover, sources belonging to the same group, i.e., having same (up to a scalar) correlation vector $\tilde{\boldsymbol{\rho}}_i$ can not be separated using only the correlation matrices $\mathbf{R}_x(k),\ k = \tau_1, \cdots, \tau_K$.

Theorem 3: Testing of Identifiability Condition

Let $\tau_1 < \tau_2 < \cdots < \tau_K$ be K distinct time lags and $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$. Assume that $\mathbf{B}$ is such a matrix that $\mathbf{z}(t)$ satisfies equation (1). Then there exists a generalized permutation matrix $\mathbf{P}$ such that for $k = \tau_1, \cdots, \tau_K$:

$$\mathbf{R}_z(k) = E(\mathbf{z}(t+k)\mathbf{z}^H(t)) = \mathbf{P}\mathbf{R}_s(k)\mathbf{P}^T$$

In other words, $z_1, \cdots, z_m$ have the same (up to a permutation) correlation factors as $s_1, \cdots, s_m$ at time lags $\tau_1, \cdots, \tau_K$.

Discussion

- Two situations may happen:
  1. For all pairs (i, j), $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ are pairwise linearly independent. Then we are sure that the sources have been separated and that $\mathbf{z}(t) = \mathbf{s}(t)$ up to a scalar and a permutation.
  2. A few pairs (i, j) out of all pairs satisfy $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ linearly dependent. Therefore the sources have been separated in blocks.
- The angle between $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ can be used as a measure of the quality of separation between source i and source j.

Simulation Examples

- **Simulation context**:
  - ULA with $M = 5$ sensors, $N = 2$ unit-norm independent sources and $T = 1000$ samples.
- **Criteria**:
  - Rejection level criterion:

$$\mathcal{I}perf_i \stackrel{\text{def}}{=} \sum_{j \neq i} E \left\{ \frac{\rho_j(0)|(\hat{\mathbf{B}}\mathbf{A})_{ij}|^2}{\rho_i(0)|(\hat{\mathbf{B}}\mathbf{A})_{ii}|^2} \right\}$$

  - Identifiability criterion:

$$\vartheta_{\boldsymbol{\rho}} \stackrel{\text{def}}{=} \left| \frac{|\tilde{\boldsymbol{\rho}}_1 \tilde{\boldsymbol{\rho}}_2^T|}{\|\tilde{\boldsymbol{\rho}}_1\| \|\tilde{\boldsymbol{\rho}}_2\|} - 1 \right|$$

Simulation examples

Table 1. Separation performance versus $\vartheta_{\boldsymbol{\rho}}$.

Sources	$\vartheta_{\boldsymbol{\rho}}$	$\mathcal{I}perf$ (dB)
2 AR1 signals	0.213	-26.23
2 CWGP signals	0.007	-5.14

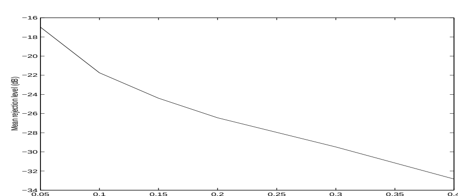


Figure 1. Separation Performance versus $\delta\theta$ (SNR=25dB): 2 AR1 sources with $a_1 = 0.95 \exp(j0.5)$ and $a_2 = 0.5 \exp(j(0.5 + \delta\theta))$.

Simulation examples

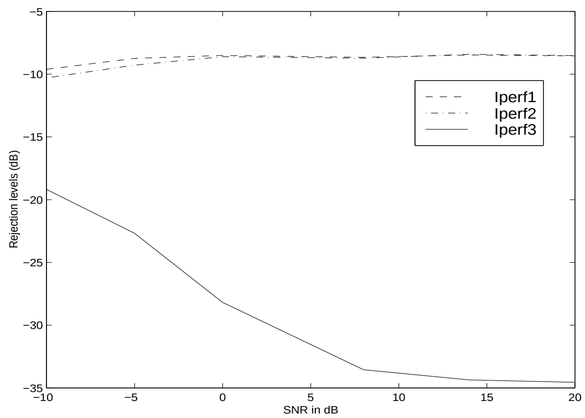


Figure 2. Separation Performance versus the SNRs: $m = 3$ sources 2 of them are CWGP signals and the third is AR1..

Simulation examples

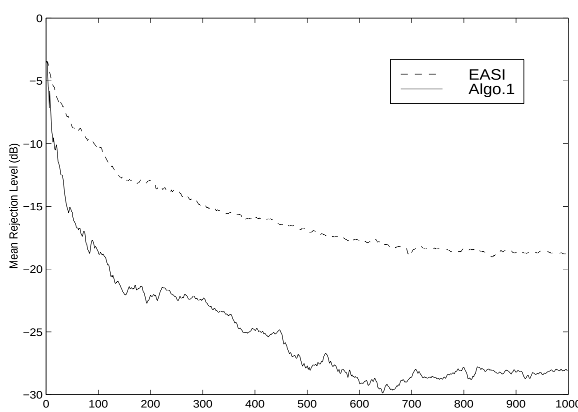


Figure 5. Comparison with EASI (Laheld & Cardoso 1996): 2 AR1 sources with QAM4 innovation processes & $SNR = 30dB$. .

Conclusions ...

Concluding remarks

- A common feature of all presented methods is the use of time and/or space **diversity**.
- **Extension** to IIR case or multiple input case is possible.
- **Partial knowledge** of the channels can be incorporated in the blind criteria, e.g., DOA of multi-paths, pulse-shape filters, spreading sequence in CDMA systems, etc.
- **Robustness** to channel order estimation errors: The last 3 methods (SS, MRE, LP) are more robust than the first 3 methods (ML, CR, CS).

Some hot topics & perspectives

- **Semi-blind** methods: i.e., combining blind and non-blind criteria (i.e., training sequence) to improve the estimation accuracy.
- **Induced cyclostationarity or pre-filtering**: i.e., modify the signal modulation at the transmission side in such a way to suit and simplify the blind system identification (BSI).
- **Space time coding**: BSI is a tool to exploit the diversity at the reception. The space-time coding is to create the diversity at the transmission.

Some hot topics & perspectives

- **Application-oriented BSI methods**: Derive or adapt blind system identification (BSI) methods for specific applications (this allows to exploit a maximum of side-information).
- **Robustness**: Improve the robustness of BSI methods against noise and modellization errors.
- **Under-determined case**: BSI for systems with more sources than sensors.

Blind Carrier Frequency Offset estimation and Mean Square Error Lower bounds

Philippe Ciblat

École Nationale Supérieure des Télécommunications, Paris, France

NEWCOM Autumn School :
Estimation Theory for wireless communications

Outline

1. Blind Carrier Frequency Offset synchronization
   - Harmonic retrieval in multiplicative noise
   - Design of powerful estimates
   - Asymptotic analysis
   - Probability of outliers
2. Mean Square Error Lower bound
   - Standard Cramer-Rao bound
   - Cramer-Rao bound with nuisance parameter
   - Bayesian Cramer-Rao bound
   - Other bounds
     - Deterministic approach : Battacharya, Barankin
     - Random approach : Ziv-Zakai

Harmonic retrieval (I)

We assume

$$y(n) = a(n)e^{2i\pi f_0 n} + b(n), \quad n = 0, \ldots, N-1$$

with

- $y(n)$: the received signal
- $a(n)$: a zero-mean *random process* or a *time-varying amplitude*.
- $b(n)$: circular white Gaussian stationary additive noise.

Goal : Estimating the frequency f_0 in multiplicative and additive noise

Outline :

1. Short review on some estimates
2. Derivations of asymptotic performance and non-asymptotic performance
3. MSE lower bounds associated with this problem

Harmonic retrieval (II)

Previous model holds for

Digital Communications : Non-data-aided/Blind synchronization.

$$a(n) = \sum_{l=0}^{L} h_l s_{n-l}$$

$\rightsquigarrow$ *circular/noncircular* complex-valued MA process
$\rightsquigarrow$ non-Gaussian process

Radar : Jakes model
$\rightsquigarrow$ $a(n)$ *circular* complex-valued Gaussian process.

Direction of Arrival (DOA) : Frequency domain.
$\rightsquigarrow$ $a(n)$ *circular* complex-valued Gaussian process.

Literature on estimator design

Digital Communications community (COM)

- A. Viterbi, U. Mengali, M. Moeneclaey
  - $\rightsquigarrow$ Ad hoc algorithms based on modulation properties (Gaussian channel)

Signal Processing community (SP)

- P. Whittle, D. Brillinger, E. Hannan, A. Walker (1950-1970)
  - $\rightsquigarrow$ Constant amplitude and periodogram analysis.
- O. Besson, P. Ciblat, M. Ghogho, G.B. Giannakis, H. Messer, E. Serpedin, P. Stoica (1990-present)
  - $\rightsquigarrow$ Time-varying amplitude
  - $\rightsquigarrow$ Notion of non-circularity
  - $\rightsquigarrow$ Notion of cyclostationary
  - $\rightsquigarrow$ Asymptotic performance analysis

Definition of circularity

Circularity (strict sense)

Let Z be a zero-mean complex random variable. Z is said circular in strict sense iff

$$Z \quad \text{and} \quad Ze^{i\theta}$$

have the same distribution for any θ.

Property

$$\mathbb{E}[\underbrace{Z\cdots Z}_{p \text{ times}}\underbrace{\overline{Z}\cdots\overline{Z}}_{q \text{ times}}] = 0$$

as soon as $p \neq q$.

Remark Z is M-order noncircular/ $M-1$-order circular random variable if only the moments of order $(M-1)$ or less satisfy the previous property

Second order circular case (I)

Assumptions

- $a(n)$ is second order circular ($=$ circular in wide sense)

$$\mathbb{E}[a(n)^2] = 0$$

- $a(n)$ is Gaussian
- $a(n)$ is colored
- $a(n)$ obeys the Jakes model

$$r_a(\tau) = J_0(2\pi f_d \tau)$$

and so $r_a(\tau)$ is real-valued.

$\rightsquigarrow$ Applications : Radar
$\rightsquigarrow$ SP community

Second order circular case (II)

We get

$$r_y(\tau) = \mathbb{E}\left[y(n+\tau)\overline{y(n)}\right] = r_a(\tau)e^{2i\pi f_0 \tau}, \quad \forall \tau \neq 0$$

As $r_a(\tau)$ is real-valued (as in Jakes model), we obtain

$$\hat{f}_N = \frac{1}{2\pi\tau}\angle \hat{r}_N(\tau)$$

where $\hat{r}_N(\tau)$ is the empirical estimate of $r_y(\tau)$ when N samples are available.

Remark

Estimating frequency boils down to estimating constant phase.

Non-circular case (I)

Assumptions

- $a(n)$ is M-order noncircular

$$\mathbb{E}[a(n)^M] \neq 0$$

- $a(n)$ is Gaussian or not
- $a(n)$ is colored or not
- $a(n)$ is a MA process

$$a(n) = \sum_{l=0}^{L} h_l s_{n-l}$$

where $\{h_l\}$ is the impulse channel response and s_n is the unknown M-order noncircular data

Non-circular case (II)

Remark

Any usual constellation is rotationally symmetric over $2\pi/M$.

Constellation	M
P-PAM	2
P-PSK	P
P-QAM	4

One can prove that any usual constellation is M-order noncircular

$\rightsquigarrow$ Applications : Digital communications
$\rightsquigarrow$ COM and SP community

Second order non-circular case (I)

Deterministic ML based method : Besson 1998

$$\{\hat{\mathbf{a}}_N, \hat{f}_N\} = \arg\min_{\mathbf{a},f} \mathbf{K}_N(\mathbf{a}, f) = \frac{1}{N}\sum_{n=0}^{N-1} |y(n) - a(n)e^{2i\pi fn}|^2$$

Non-linear least square (NLLS) asymptotically equivalent to maximization of periodogram of $y^2(n)$

$$\hat{f}_N = \arg\min_f \mathbf{J}_N(f) = \left|\frac{1}{N}\sum_{n=0}^{N-1} y^2(n)e^{-2i\pi(2f)n}\right|^2$$

$\rightsquigarrow$ Traditional Square-Power estimate in COM community for BPSK

Second order non-circular case (II)

Remark

As $u_a(0) = \mathbb{E}[a^2(n)] \neq 0$, then

$$z(n) = y^2(n) = r_a(0)e^{2i\pi(2f_0)n} + e(n)$$

where $e(n)$ is a *non-Gaussian* and *non-stationary* additive noise.

Conclusion

- $\rightsquigarrow$ Frequency estimation in multiplicative and additive noise

  $\Updownarrow$

  Frequency estimation in additive noise *but non-standard noise*
- $\rightsquigarrow$ Periodogram based on $y^2(n)$ instead of $y(n)$.
- $\rightsquigarrow$ If $a(n)$ colored, periodogram not exhaustive.

Cyclostationary based method

- Let $u_y(n,\tau) = \mathbb{E}[y(n+\tau)y(n)]$ be the *pseudo-correlation*

Definition

$y(n)$ is cyclostationary w.r.t. its pseudo-correlation iff $n \mapsto u_y(n,\tau)$ is periodic of period $1/\alpha_0$. Then

$$u_y(n,\tau) = \sum_k u_y^{(k\alpha_0)}(\tau) e^{2i\pi k\alpha_0 n}$$

with

- $k\alpha_0$: k^{th} cyclic frequency
- $u_y^{(k\alpha_0)}(\tau)$: cyclic pseudo correlation
- $c_a^{(k\alpha_0)}(e^{2i\pi f}) = \text{F.T.}(\tau \mapsto u_y^{(k\alpha_0)}(\tau))$: cyclic pseudo spectrum

- $n \mapsto u_y(n,\tau)$ is periodic of period $1/\alpha_0$ with $\alpha_0 = 2f_0$
- Ciblat & Loubaton 2000

Contrast function

Remark

Estimating frequency in multiplicative and additive noise boils down to estimating a cyclic frequency

$$f_0 = \arg\max_f \mathbf{J}_{\mathbf{W}}(f) = \mathbf{u}_y^{(2f)^{\mathrm{H}}} \mathbf{W} \mathbf{u}_y^{(2f)} = \left\| \mathbf{u}_y^{(2f)} \right\|_{\mathbf{W}}^2$$

with $\mathbf{u}_y^{(\alpha)} = [u_y^{(\alpha)}(-T), \cdots, u_y^{(\alpha)}(T)]^{\mathrm{T}}$.

In practice, $\mathbf{u}_y^{(2f)}$ is not available and needs to be estimated

$$\begin{aligned} u_y^{(\alpha)}(\tau) &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} u_y(n,\tau) e^{-2i\pi\alpha n} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}[y(n+\tau)y(n)] e^{-2i\pi\alpha n} \end{aligned}$$

Contrast process

$$\hat{f}_N = \arg\max_f \mathbf{J}_{N,\mathbf{w}}(f) = \left\| \hat{\mathbf{u}}_N^{(\alpha)} \right\|_{\mathbf{w}}^2$$

with $\hat{\mathbf{u}}_y^{(\alpha)} = [\hat{u}_y^{(\alpha)}(-T), \cdots, \hat{u}_y^{(\alpha)}(T)]^{\mathrm{T}}$ and

$$\hat{u}_N^{(\alpha)}(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} y(n+\tau)y(n)e^{-2i\pi\alpha n}.$$

Then

$$\hat{f}_N = \arg\max_f \mathbf{J}_{N,\mathbf{w}}(f) = \left\| \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{z}(n)e^{-2i\pi(2f)n} \right\|_{\mathbf{w}}^2$$

with $\mathbf{z}(n) = [z_{-T}(n), \ldots, z_T(n)]^{\mathrm{T}}$ and
$z_\tau(n) = y(n+\tau)y(n) = u_y^{(\alpha_0)}(\tau)e^{2i\pi\alpha_0 n} + e_\tau(n)$.

Remarks

- Multi-variate periodogram
- Weighted periodogram
- Extended Square-Power algorithm
- Asymptotic performance
  - Giannakis & Zhou
    1995 : cyclostationarity approach and CRB bounds
  - Besson & Stoica
    1999 : deterministic NLS with white real-valued multiplicative noise
  - Ghogho & Swami
    1999 : deterministic NLS with white real-valued multiplicative noise
  - Ciblat & Loubaton
    2000 : weighted multi-variate periodogram and analysis with colored complex-valued multiplicative noise

High-order noncircular case

***P*-PSK** : Viterbi 1983.

$$\mathbb{E}[a(n)^P] \neq 0 \Leftrightarrow \hat{f}_N = \arg\max_f \left\| \frac{1}{N} \sum_{n=0}^{N-1} y^P(n) e^{-2i\pi(Pf)n} \right\|^2$$

Tutorial done by Morelli-Mengali in 1998.

***P*-QAM** : Moeneclaey 2001 & Serpedin 2004

$$\mathbb{E}[a(n)^4] \neq 0 \Leftrightarrow \hat{f}_N = \arg\max_f \left\| \frac{1}{N} \sum_{n=0}^{N-1} y^4(n) e^{-2i\pi(4f)n} \right\|^2$$

$\Rightarrow$ *The so-called M-power estimate*

Asymptotic analysis

- Consistency

$$\hat{f}_N - f_0 \overset{p.s.}{\rightarrow} 0$$

- Asymptotic normality : it exists p such that

$$N^p(\hat{f}_N - f_0) \overset{\mathcal{D}}{\rightarrow} \mathcal{N}(0, \gamma)$$

  with
  - p the so-called convergence speed
  - γ the so-called asymptotic covariance
- Asymptotic covariance

$$\mathrm{MSE} = \mathbb{E}[(\hat{f}_N - f_0)^2] \sim \frac{\gamma}{N^{2p}}$$

Convergence analysis

- Consistency
- Asymptotic normality (with $p = 3/2$)

are proven in Ciblat & Loubaton for

$$\hat{\alpha}_N = \arg\max_{\alpha} \mathbf{J}_{N,\mathbf{W}}(\alpha) = \left\| \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{z}(n) e^{-2i\pi\alpha n} \right\|_{\mathbf{W}}^2$$

where

$$\mathbf{z}(n) = \mathbf{u} e^{2i\pi\alpha_0 n} + \mathbf{e}(n)$$

whatever the noise process $\mathbf{e}(n)$ satisfying standard mixing conditions

Remarks

- Analysis valid for second order and high order noncircular case
- Derivations of the asymptotic covariance need still to be done

Asymptotic covariance (I)

Second-order noncircular case :
whatever the second-order noncircular process $a(n)$, Ciblat & Loubaton (*IEEE SP 2002*) have proven that

1. $\mathbf{W}_{opt} = \mathbf{Id}_{2T+1}$
2. $T_{opt} = L$ with L the memory size of $a(n)$
3. $$\mathrm{MSE} \sim \frac{3}{4\pi^2 N^3} \cdot \frac{\int_0^1 |c_a(e^{2i\pi f})|^2 \mathcal{X}(e^{2i\pi f}) df}{\left(\int_0^1 |c_a(e^{2i\pi f})|^2 df\right)^2}.$$

   with
   $\mathcal{X}(e^{2i\pi f}) = (s_a(e^{2i\pi f}) + \sigma^2)(\overline{s_a(e^{-2i\pi f})} + \sigma^2) - c_a(e^{2i\pi f})\overline{c_a(e^{-2i\pi f})}$

- if $a(n)$ is a white real-valued process, then asymptotic covariance also available in Ghogho and in Besson

Asymptotic covariance (II)

High-order noncircular case :
Serpedin (*IEEE TCOM 2003* and *IEEE TWIRELESS 2003*) has proven that

$$\mathrm{MSE}_{P\text{-PSK}} \sim \frac{24}{\pi^2 P N^3} \frac{B-D}{C^2}$$

with

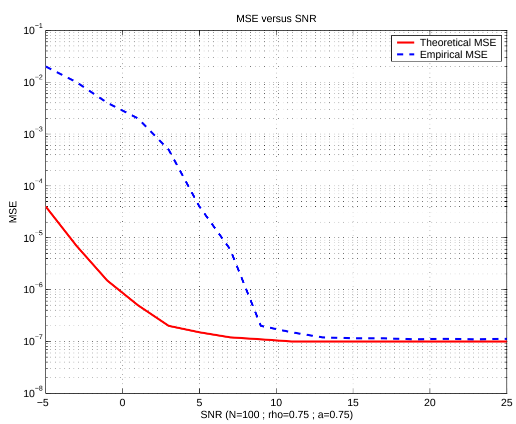
$$\begin{aligned} B &= \sum_{q=0}^{P} (C_P^q)^2 q! \sigma_b^{2q} \\ C &= e^{-1/\sigma_b^2} {}_2F_1(2P+1, 2P+1, 1/\sigma_b^2) \\ D &= \frac{P!}{(2P)!} e^{-1/\sigma_b^2} {}_2F_1(P+1, P+1, 1/\sigma_b^2) \end{aligned}$$

- Similar equations for P-QAM constellation

Numerical illustrations

Set-up :

- $a(n) = s(n) + 0.75 s(n-1)$ with $s(n)$ white Gaussian process
- Performance of “weighted periodogram-based estimate” vs. SNR



Questions :
⇝ How far away from Cramer-Rao Bound we are ?
⇝ Irrelevancy of MSE at low SNR (*outliers effect*).

Outliers effect

We focus on the following M-power estimate

$$\hat{f}_N = \frac{1}{M} \arg \max_{\alpha \in]-1/2, 1/2]} \left| \frac{1}{N} \sum_{n=0}^{N-1} y(n)^M e^{-2i\pi\alpha n} \right|^2$$

with

$$y(n)^M = ue^{2i\pi M f_0 n} + e(n)$$

This periodogram is maximizing by proceeding into two steps

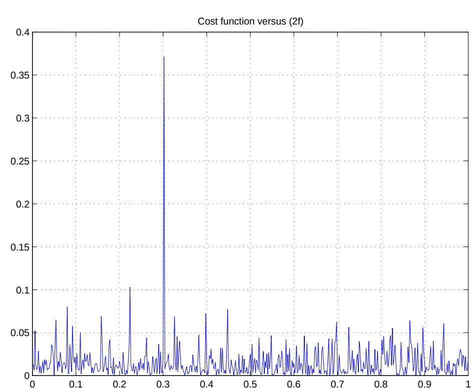
- a "coarse" step detecting the peak
- a "fine" step refining the estimation around the peak

Remark

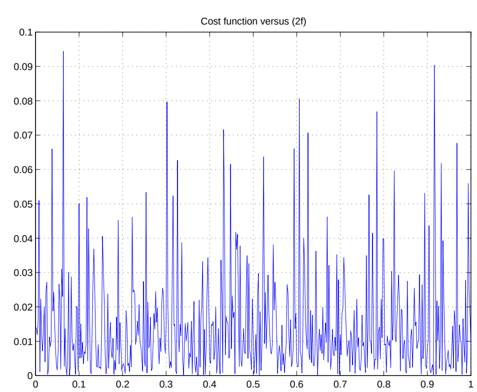
At *low SNR* and/or when *few samples are available*, the coarse step may fail. This leads to the so-called *outliers effect*.

Example

- $a(n)$ is a complex-valued white zero-mean Gaussian process with unit-variance and pseudo-variance $u = \mathbb{E}[a(n)^2]$
  $\rightsquigarrow$ $|u|$ refers to non-circularity rate.
- $SNR = 0\mathrm{dB}$ and $N = 500$



Cost function with $|u| = 2/3$



Cost function with $|u| = 1/3$

Mean Square Error

True MSE

$$\mathrm{MSE} = \frac{p}{12} + (1-p)\mathrm{MSE}_{\mathrm{o.f.}}$$

where

- p is the probability of coarse step failure
- $\mathrm{MSE}_{\mathrm{o.f.}}$ is the standard "outliers effect"-free MSE

Available Results :

1. $\mathrm{MSE}_{\mathrm{o.f.}}$ seen in previous slides
2. p recently derived (Ciblat & Ghogho submitted to TCOM)

Failure probability p (I)

Let Y_k (resp. E_k) be the N-FFT of $y(n)^M$ (resp. $e(n)$)

$$|Y_k| = \begin{cases} |ue^{2i\pi M\phi_0} + E_0| & \text{si} \quad k = 0 \\ |E_k| & \text{si} \quad k \neq 0 \end{cases}, (f_0 = 0)$$

The failure probability may write as follows

$$p = 1 - \mathrm{Pb}(\forall k \neq 0, |Y_k| < |Y_0|) = 1 - \int p_1(x)p_2(x)dx$$

where

$$\begin{aligned} p_1(x) &= \mathrm{Pb}(\forall k \neq 0, |Y_k| < x) \\ &= \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} p_{|Y_1|,\cdots,|Y_{N-1}|}(y_1, \cdots, y_N - 1)dy_1 \cdots dy_{N-1} \\ p_2(x) &= p_{|Y_0|}(x) \end{aligned}$$

$\Rightarrow$ *The distribution of FFT points are needed*

Failure probability p (II)

Constant-amplitude multiplicative noise :

- $a(n) = a, \quad \forall n$
- $M = 1$
- Rife & Boorstyn (*IEEE IT 1974*)

$\rightsquigarrow$ $e(n)$ is white circular Gaussian process

Time-varying multiplicative noise :

- $a(n)$ is white and belongs to an usual constellation

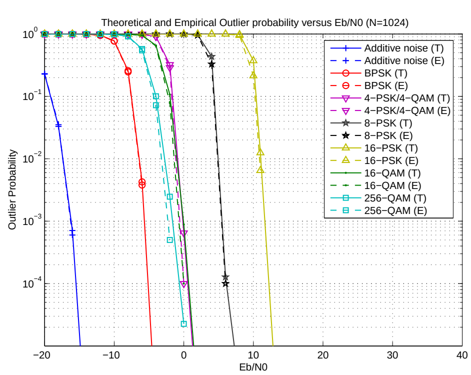
$\rightsquigarrow$ $e(n)$ is white *noncircular* and *non-Gaussian* process

Result

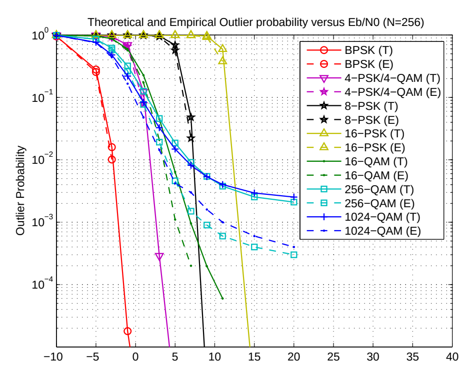
Under Gaussian assumption, a closed-form expression for p can be addressed which strongly depends on

$$\sigma_e^2 = \mathbb{E}[|a(n)|^{2M}] - |\mathbb{E}[a(n)^M]|^2 + \sum_{m=0}^{M-1} (C_M^m)^2 \mathbb{E}[|a(n)|^{2m}]\mathbb{E}[|b(n)|^{2(M-m)}]$$

Simulations : p versus SNR



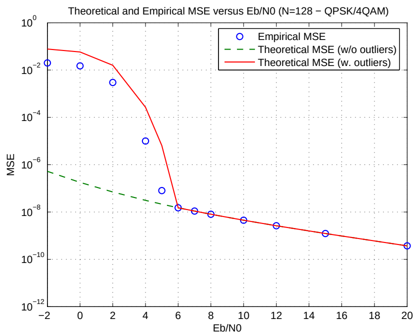
$N = 1024$



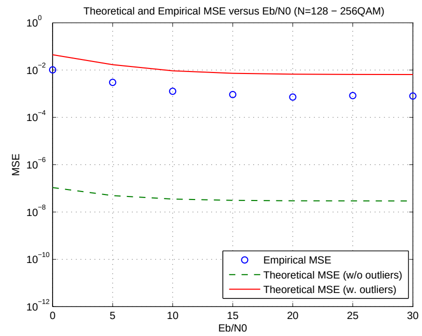
$N = 256$

- p strongly depends on P for P-PSK
- p slightly depends on P for P-QAM
- *Self-noise* for QAM due to $\sigma_e^2 = \mathbb{E}[|a(n)|^8] - |\mathbb{E}[a(n)^4]|^2 \neq 0$ in noiseless case

Simulations : MSE versus SNR



4-QAM and $N = 128$

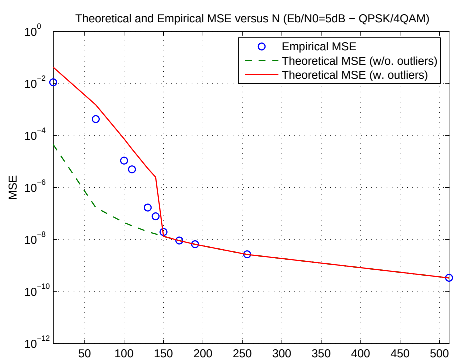


256-QAM and $N = 128$

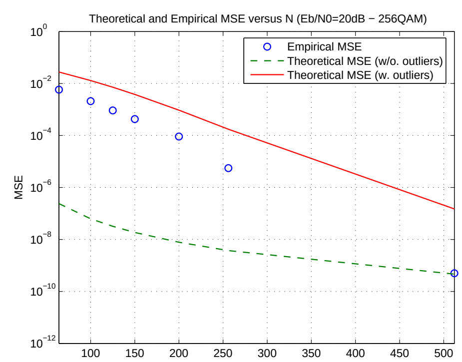
Threshold analysis

- For 4-QAM, $\mathrm{SNR}_{\mathrm{th.}} = 6$dB if $N = 128$
- For P-QAM (with $P > 4$), floor effect for $p \Rightarrow$ no threshold

Simulations : MSE versus N



4-QAM and $E_b/N_0 = 5$dB



256-QAM and $E_b/N_0 = 20$dB

- When N increases, p decreases (without floor effect)
- Any MSE is reachable BUT sometimes with very large N

Remarks

Estimation accuracy

- Data-aided context can improve the performance but outliers effect still exists (Mengali IEEE TCOM 2000)
- Cramer-Rao bound (CRB) with coded scheme is less than CRB without coded scheme (Moeneclaey IEEE COML 2003)
- Turbo-estimation is an appropriate solution (Vandendorpe & al. EURASIP JWCN 2005)

Questions

1. MSE value : is it far away from the lower bound (Cramer-Rao Bound) ?
2. Outliers effect : is it intrinsic to M power estimate or to any estimate ?

Mean Square Error Lower Bounds

Signal Model

$$y(n) = a(n)e^{2i\pi f_0 n} + b(n), \quad n = 0, \dots, N-1 \Leftrightarrow \mathbf{y} = \mathbf{D}(f_0)\mathbf{a} + \mathbf{b}$$

where

- $\mathbf{y} = [y(0), \cdots, y(N-1)]^{\mathrm{T}}$
- $\mathbf{D}(f_0) = \mathrm{diag}([1, \cdots, e^{2i\pi f_0(N-1)}])$
- Noise variance assumed to be known (for sake of simplicity)

f_0 : (deterministic) parameter of interest
$\{a(0), \cdots, a(N-1)\}$: parameters of nuisance

Each assumption on the parameters of nuisance (deterministic/random, etc.) leads to ONE Cramer-Rao-type bound

Unconditional CRB

We consider the likelihood for parameters $\{f_0, \mathbf{a}\}$:

$$\Lambda(f, \mathbf{a}) \quad \left(\propto e^{\frac{-\|\mathbf{y}-\mathbf{D}(f)\mathbf{a}\|^2}{2N_0}}\right)$$

$a(n)$ are viewed as real nuisance $\rightsquigarrow$ *stochastic*

Unconditional CRB or True CRB or Stochastic CRB

Unconditional Likelihood is equal to True-Likelihood

$$\Lambda_u(f) = \mathbb{E}_{\mathbf{a}}[\Lambda(f, \mathbf{a})] = \int \Lambda(f, \mathbf{a}) p(\mathbf{a}) d\mathbf{a}$$

$$\Rightarrow \mathrm{UCRB}(f) = \frac{1}{\mathbb{E}_{\mathbf{y}}\left[\left|\frac{\partial}{\partial f}\ln \Lambda_u(f)\right|^2\right]} = \frac{1}{\mathbb{E}_{\mathbf{y}}\left[\left|\frac{\partial}{\partial f}\ln \mathbb{E}_{\mathbf{a}}[\Lambda(f, \mathbf{a})]\right|^2\right]}$$

$\rightsquigarrow$ Often *untractable*
$\rightsquigarrow$ UCRB mainly analysed by Moeneclaey
$\rightsquigarrow$ Approximation at low SNR ($e^x = 1 + x + x^2/2$ if x small)

Conditional CRB

$a(n)$ are viewed as parameters of interest $\rightsquigarrow$ *deterministic*

Conditional CRB or Deterministic CRB

Conditional Likelihood is equal to Deterministic Likelihood

$$\Lambda_c(f) = \Lambda(f, \hat{\mathbf{a}}_f) \quad \text{where} \quad \frac{\partial \Lambda(f, \mathbf{a})}{\partial \mathbf{a}}_{|\hat{\mathbf{a}}_f} = 0$$

$$\Rightarrow \mathrm{CCRB}(f) = \frac{1}{\mathbb{E}_{\mathbf{y}}\left[\left|\frac{\partial}{\partial f}\ln \Lambda_c(f)\right|^2\right]}$$

Average CCRB or Asymptotic CCRB

$$< \mathrm{CCRB} >(f) = \frac{1}{\mathbb{E}_{\mathbf{y},\mathbf{a}}\left[\left|\frac{\partial}{\partial f}\ln \Lambda_c(f)\right|^2\right]}$$

$\rightsquigarrow$ CCRB not used although CML well spread
$\rightsquigarrow$ CCRB mainly analysed by Stoica and Vazquez

Modified CRB

$a(n)$ are viewed as *known* parameters

Modified CRB

$$\Rightarrow \mathrm{MCRB}(f) = \frac{1}{\mathbb{E}_{\mathbf{y},\mathbf{a}}\left[\left|\frac{\partial}{\partial f}\ln\Lambda(f,\mathbf{a})\right|^2\right]}$$

$\rightsquigarrow$ Closed-form expressions *tractable*
$\rightsquigarrow$ MCRB introduced by Mengali
$\rightsquigarrow$ MCRB very often used in COM/SP community

Gaussian CRB

$a(n)$ are viewed as *Gaussian process*

Gaussian CRB

Gaussian Likelihood

$$\Lambda_g(f) = \mathbb{E}_{\mathbf{a}}[\Lambda(f,\mathbf{a})]$$

where **a** is a Gaussian vector.

$$\Rightarrow \mathrm{GCRB}(f) = \frac{1}{\mathbb{E}_{\mathbf{y}}\left[\left|\frac{\partial}{\partial f}\ln\Lambda_g(f)\right|^2\right]}$$

$\rightsquigarrow$ Closed-form expressions *tractable*
$\rightsquigarrow$ Not valid for digital communications but this is still a bound for all the consistent estimates based on data sample covariance matrix
$\rightsquigarrow$ GCRB developed in SP community (Giannakis, Ghogho, Ciblat)

Bayesian CRB

f_0 is also viewed as stochastic variable with an a priori pdf $p(f)$

Let $\hat{\boldsymbol{\theta}}$ be an unbiased estimate of $\boldsymbol{\theta}_0 = [f_0, \mathbf{a}]$. Then

$$\mathrm{MSE}_{|\mathrm{Bayesian}} = \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^{\mathrm{T}}] \geq \mathbf{J}^{-1} = \mathrm{BCRB}$$

with

$$\mathbf{J} = \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\left[\frac{\partial \ln \Lambda(f, \mathbf{a})}{\partial \boldsymbol{\theta}} \frac{\partial \ln \Lambda(f, \mathbf{a})}{\partial \boldsymbol{\theta}}^{\mathrm{T}}\right]$$

Jensen's inquality

$$\mathbb{E}_{\boldsymbol{\theta}}[\mathrm{CRB}(\boldsymbol{\theta})] \geq \mathrm{BCRB}$$

Remarks

- If $\mathrm{CRB}(\boldsymbol{\theta})$ independent of $\boldsymbol{\theta}$ then $\mathrm{CRB} = \mathrm{BCRB}$
- No link in the literature between xCRB and BCRB
- If $\theta_0 = f_0$, then $\mathbb{E}_{\theta}[\mathrm{TCRB}(\theta)] \geq \mathrm{BCRB}$

Bayesian Algorithm

Deterministic approach :

- Optimal unbiased estimate does not always exist (except ML in asymptotic regime)

Stochastic/Bayesian approach :

- Optimal unbiased estimate always exists : the so-called MMSE estimator

$$\hat{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}}[\boldsymbol{\theta}] = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$$

Remarks

- *The MMSE is the mean of the a posteriori density*
- $p(\boldsymbol{\theta})$ must be differentiable
- SP community (Van Trees)

Link between xCRB (I)

All these bounds (except GCRB) lower-bound the mean square error !

Results

$$\mathrm{UCRB} \geq \mathrm{MCRB}$$

and

$$< \mathrm{CCRB} > \geq \mathrm{MCRB}$$

- *At high SNR :* $\mathrm{UCRB} = \mathrm{MCRB}$ (if the values of the parameters of nuisance belongs to a discrete set)
- *For large samples :* $\mathrm{CCRB} \overset{N\to\infty}{\longrightarrow} < \mathrm{CCRB} >$ (ergodism)
- *Under Gaussian assumption :* $\mathrm{UCRB} = \mathrm{GCRB}$

$\rightsquigarrow$ MCRB usually too optimistic
$\rightsquigarrow$ GCRB unable to take into acocunt high order information

Link between xCRB (II)

Application to blind synchronization

$a(n)$ belongs to a constellation and thus to a discrete set

$\rightsquigarrow$ At high SNR,

$$\mathrm{UCRB} = \mathrm{MCRB}$$

MCRB is of interest in digital communications

$\rightsquigarrow$ At low SNR,

$$\mathrm{UCRB} \gg \mathrm{MCRB}$$

Let M be the order of non-circularity (Moeneclaey IEEE COML 2001).

$$\mathrm{UCRB} = \mathcal{O}(1/\mathrm{SNR}^M) \quad \text{and} \quad \mathrm{MCRB} = \mathcal{O}(1/\mathrm{SNR})$$

$\rightsquigarrow$ GCRB likely useful for BPSK but not for other constellations

Example (I)

Harmonic retrieval where $a(n)$ is complex-valued white (discrete) process with $\mathbb{E}[|a(n)|^2] = 1$ and $\mathbb{E}[a(n)^2] = u$.

$$\mathrm{MCRB} = \frac{3\sigma^2}{2\pi^2 N^3} \quad \text{and} \quad \mathrm{GCRB} = \frac{3\left[(1-|u|^2) + 2\sigma^2 + \sigma^4\right]}{4\pi^2 |u|^2 N^3}$$

$$\mathrm{UCRB}_{|\text{low SNR}} = \frac{3\sigma^4}{4\pi^2 |u|^2 N^3}$$

and

$$\mathrm{UCRB}_{|\text{high SNR}} = \mathrm{MCRB} = \frac{3\sigma^2}{2\pi^2 N^3}$$

$\rightsquigarrow$ MCRB quite relevant BUT does not depend on non-circularity rate.
$\rightsquigarrow$ At low SNR, second-order noncircularity leads to GCRB=UCRB
$\rightsquigarrow$ If $|u| \neq 1$, floor error with GCRB not with MCRB and UCRB

Example (II)

We consider $u = 1$ (e.g. $a(n) \in \mathrm{BPSK}$)

$$\mathrm{MCRB} = \frac{3\sigma^2}{2\pi^2 N^3} \quad \text{and} \quad \mathrm{GCRB} = \frac{3\sigma^2}{2\pi^2 N^3} + \frac{3\sigma^4}{4\pi^2 N^3}$$

and

$$\mathrm{UCRB}_{|\text{high SNR}} = \frac{3\sigma^2}{2\pi^2 N^3} \quad \text{and} \quad \mathrm{UCRB}_{|\text{low SNR}} = \frac{3\sigma^4}{4\pi^2 N^3}$$

At high SNR

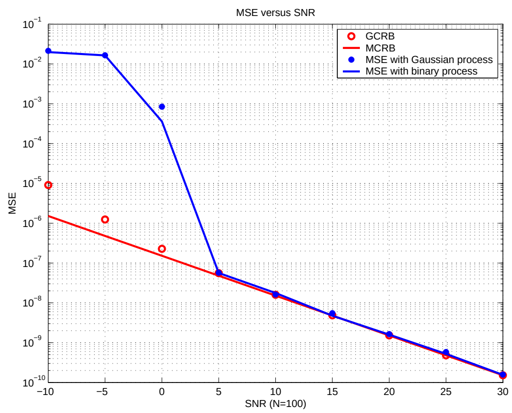
$$\mathrm{UCRB} = \mathrm{MCRB} = \mathrm{GCRB}$$

At low SNR

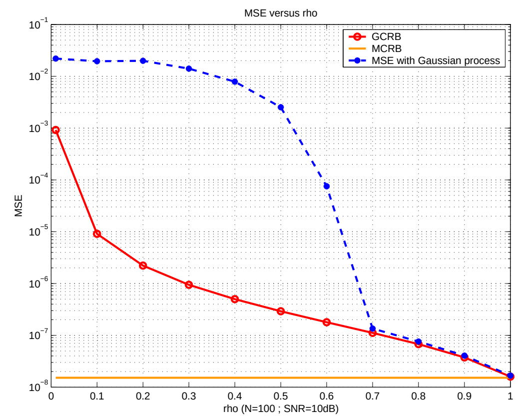
$$\mathrm{UCRB} = \mathrm{GCRB}$$

$\rightsquigarrow$ GCRB relevant for BPSK

Example (III)



MSE versus SNR



MSE versus $|u|$

⇝ For BPSK signal, we are lucky (GCRB≈MCRB) !

Asymptotic Gaussian CRB (I)

⇝ Several works for obtaining asymptotic (large sample) expressions for GCRB.

- Circular case : Ghogho 2001 (based on Whittle's theorem)
- Real-valued case : Ghogho 1999
- Non-circular case : Ciblat 2003 (large Toeplitz matrices)

	White	Colored
Circular	∞	$O(1/N)$
Real-valued	$O(1/N^3)$ No floor error Reached by Square Power	$O(1/N^3)$ No floor effect
Non-circular	$O(1/N^3)$ No floor error Reached by Square Power	$O(1/N^3)$ Floor effect

Asymptotic Gaussian CRB (II)

Second-order noncircular case : Ciblat (EURASIP SP 2005)

$$\mathrm{GCRB} \sim \frac{3}{4\pi^2 \xi N^3} \quad \text{with} \quad \xi = \int_0^1 \frac{c_a(e^{2i\pi f})\overline{c_a(e^{-2i\pi f})}}{\mathcal{X}(e^{2i\pi f})} df$$

$$\mathrm{MSE} \sim \frac{3\eta}{4\pi^2 N^3} \quad \text{with} \quad \eta = \frac{\int_0^1 |c_a(e^{2i\pi f})|^2 \mathcal{X}(e^{2i\pi f}) df}{\left(\int_0^1 |c_a(e^{2i\pi f})|^2 df\right)^2}$$

One can proven that (Cauchy-Schwartz inequality)

$$\mathrm{GCRB} = \mathrm{MSE} \text{ iff } a(n) \text{ white process}$$

Other types of bound

Remark

xCRB unable to predict and analyze the outliers effect

Solutions

Introducing other tighter lower bounds

- Deterministic approach
  - ⇝ Battacharyya bound
  - ⇝ Barankin bound
- Stochastic approach
  - ⇝ Ziv-Zakai bound
  - ⇝ Weiss-Weinstein bound

Battacharyya bound (I)

Review on CRB : consider the vector $\mathbf{z}$,

$$\mathbf{z} = \begin{bmatrix} \boldsymbol{\theta} - \boldsymbol{\theta}_0 \\ \frac{\partial \ln(p(\mathbf{y}|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \end{bmatrix}$$

By construction, $\mathbb{E}[\mathbf{z}\mathbf{z}^{\mathrm{T}}]$ is nonnegative matrix. This implies that

$$\begin{bmatrix} \mathrm{MSE} & 1 \\ 1 & \mathrm{FIM} \end{bmatrix} \geq 0$$

and

$$\mathrm{MSE} \geq \mathrm{FIM}^{-1} = \mathrm{CRB}$$

Battacharyya bound (II)

consider the vector $\mathbf{z}_N$,

$$\mathbf{z}_N = \begin{bmatrix} \boldsymbol{\theta} - \boldsymbol{\theta}_0 \\ \frac{\partial \ln(p(\mathbf{y}|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \\ \vdots \\ \frac{\partial^N \ln(p(\mathbf{y}|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^N} \end{bmatrix}$$

Once again $\mathbb{E}[\mathbf{z}_N\mathbf{z}_N^{\mathrm{T}}]$ is nonnegative matrix and this leads to

$$\mathrm{MSE} \geq \mathrm{BaB} = \mathrm{CRB} + \text{one positive term}$$

Barankin bound (I)

We consider "test-points" $\mathcal{E}_n = [\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}^{(n)} - \boldsymbol{\theta}_0]$.
Furthermore $\mathbf{B}_n = (B_{k,l})_{1 \leq k,l \leq n}$ is the following $n \times n$ matrix

$$B_{k,l} = \mathbb{E}_{\mathbf{y}}\left[\frac{p(\mathbf{y}|\boldsymbol{\theta}^{(k)})p(\mathbf{y}|\boldsymbol{\theta}^{(l)})}{p(\mathbf{y}|\boldsymbol{\theta}_0)^2}\right]$$

Definition

Barankin bound of order $n \rightsquigarrow \mathrm{BB}_n(\boldsymbol{\theta}_0) = \sup_{\mathcal{E}_n} \underbrace{\mathcal{E}_n(\mathbf{B}_n(\mathcal{E}_n) - \mathbf{1}_n\mathbf{1}_n^{\mathrm{T}})^{-1}\mathcal{E}_n^{\mathrm{T}}}_{S_n(\mathcal{E}_n)}$

with $\mathbf{1}_n = \mathrm{ones}(n, 1)$

$\rightsquigarrow$ *MSE of any unbiased estimator is greater than any* BB_n
$\rightsquigarrow$ *As* $n \rightarrow \infty$, BB_∞ *becomes even the tightest lower bound*

Barankin bound (II)

- BB_1 used (one test-point)
- Main task : closed-form expression for matrix **B**

Remark

$$\mathrm{CRB} = \lim_{\mathcal{E} \rightarrow 0} S_1(\mathcal{E})$$

$\rightsquigarrow$ CRB inspects the likelihood only around the true point
$\rightsquigarrow$ CRB and BaB unable to observe outliers

$$\mathrm{BB} = \sup_{\mathcal{E}} S_1(\mathcal{E})$$

$\rightsquigarrow$ BB scans all the research interval
$\rightsquigarrow$ BB takes into account outliers effect in lower bound.

- Pure harmonic retrieval : Knockaert in 1997
- Circular multiplicative noise : Messer in 1992 for DOA issue

Derivations

- Let $y(n) = ae^{2i\pi f_0 n} + b(n) \rightsquigarrow$ Information in *mean* of $y(n)$.
- Let $y(n) = a(n)e^{2i\pi f_0 n} + b(n) \rightsquigarrow$ Information in *variance* of $y(n)$.

Closed-form expression (Ciblat EURASIP SP 2005)

$$B_{k,l} = \begin{cases} \frac{1}{\sqrt{\det(\mathbf{Q}_{k,l})}} & \text{if } \mathbf{Q}_{k,l} > 0 \\ +\infty & \text{otherwise} \end{cases},$$

where

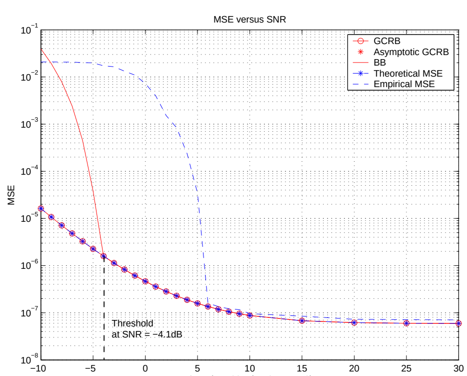
$$\mathbf{Q}_{k,l} = (\widetilde{\mathbf{R}}_{f(k)}^{-1} + \widetilde{\mathbf{R}}_{f(l)}^{-1})\widetilde{\mathbf{R}}_{f_0} - \mathbf{Id}_{2N}$$

and

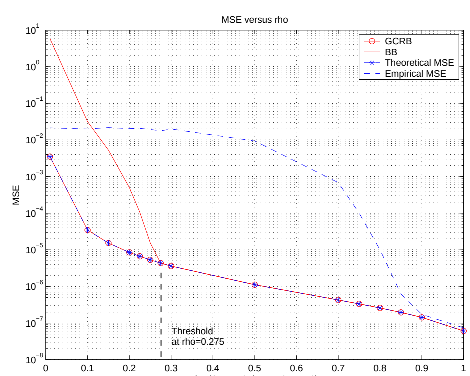
$$\widetilde{\mathbf{R}}_f = \begin{bmatrix} \frac{\mathbf{E}[\mathbf{y}_N \mathbf{y}_N^{\mathrm{H}}]}{\mathbf{E}[\mathbf{y}_N \mathbf{y}_N^{\mathrm{T}}]} & \frac{\mathbf{E}[\mathbf{y}_N \mathbf{y}_N^{\mathrm{T}}]}{\mathbf{E}[\mathbf{y}_N \mathbf{y}_N^{\mathrm{H}}]} \end{bmatrix}.$$

Numerical illustrations

$a(n)$ white Gaussian process with unit-variance and $\mathbb{E}[a(n)^2] = u$.



MSE versus SNR



MSE versus u

- Threshold analysis : $\mathrm{BB} = \max(\mathrm{GCRB}, S(1/4))$
- Important gap between BB and standard Square-Power estimate

Ziv-Zakai bound (I)

- Bayesian bound : random parameter
- Two classes :
  - Hölder inequality :
    - Bayesian Battacharyya
    - Bobrovsky-Zakai (1976)
    - Weiss-Weinstein bound (1985)
  - Kotelnikov inequality :
    - Ziv-Zakai (1969)
    - Bellini-Tartara (1975)

State-of-the-Art

Ziv-Zakai bound (ZZB) derivations

1. bearing estimation and additive noise (Bell IEEE IT 1997)
2. time-delay estimation (Weiss IEEE SP 1983)

Ziv-Zakai bound (II)

Definition

The mean square error (MSE) for φ_1 is bounded by

$$\mathrm{MSE} \geq \int_0^\infty h_1 \left(\max_{h_0} g(h_0, h_1) \right) dh_1 .$$

where

- $g(h_0, h_1) = \int \min(p(\varphi), p(\varphi + \mathbf{h})) P_e(\varphi, \varphi + \mathbf{h}) d\varphi$
- $\varphi = [\phi_0, f_0]$ and $\mathbf{h} = [h_0, h_1]$
- $p(.)$ is the *a priori* density function of φ
- $P_e(\varphi, \varphi + \mathbf{h})$ is the error probability when the optimal detector decides between the following two equally likely hypotheses

$$\left\{ \begin{array}{lcl} H_0 : \; y(n) & = & a(n) e^{2i\pi(\phi_0 + f_0 n)} + b(n) \\ H_1 : \; y(n) & = & a(n) e^{2i\pi((\phi_0 + h_0) + (f_0 + h_1) n)} + b(n) \end{array} \right.$$

$\rightsquigarrow$ *Detection theory with multiplicative noise*

Derivations

Result

$$\mathrm{MSE}_1 \geq \int_0^{1/2} (1/2 - h_1) h_1 (\max_{h_0}(1/2 - h_0) P_e(h_0, h_1)) dh_1$$

with

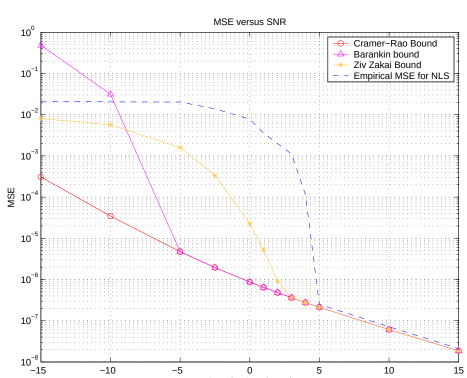
$$P_e(h_0, h_1) = \frac{(\theta_1/\theta_2)^{\alpha_1}}{\alpha_1} B(\alpha_1, \alpha_2) {}_2F_1(\alpha_1 + \alpha_2, \alpha_1, \alpha_1 + 1 - \theta_1/\theta_2)$$

where

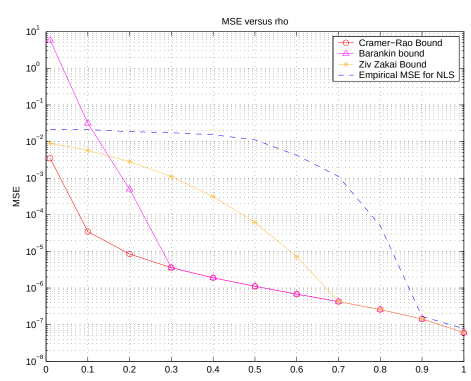
- $B(\alpha_1, \alpha_2) = \Gamma(\alpha_1 + \alpha_2)/\Gamma(\alpha_1)$ is called either the Euler's first integral or the Beta function
- ${}_2F_1(\alpha, \beta, \gamma; x)$ is the hyper-geometric function
- Closed-form expressions of θ_1, θ_2, α_1, α_2 depend on $\widetilde{\mathbf{R}}_\mathbf{h}$ and $\widetilde{\mathbf{R}}_0$

Numerical illustrations

$a(n)$ white Gaussian process with unit-variance and $\mathbb{E}[a(n)^2] = u$.



MSE versus SNR



MSE versus u

- Small gap between ZZB and standard Square-Power estimate

Bibliography

- *BCRB and CRB* : H.L. Van Trees,"Detection, Estimation, and Modulation Theory", Part 1, 1968.
- *xCRB* :
  - G. Vazquez, "Non-Data-Aided Digital Synchronization" in the book "Signal Processing Advances in Communications" edited by G.Giannakis et al., 2000.
  - M. Moeneclaey, "On the True and the Modified CRB for the estimation of a scalar paramter in the presence of nuisance parameter", IEEE TCOM, Nov. 1998.
- *Asymptotic analysis* :
  - P. Ciblat, "Asymptotic analysis of blind cyclic correlation based symbol rate estimation", IEEE IT, Jul. 2002.
  - M. Ghogho, "Frequency estimation in the presence of Doppler spread : performance analysis", IEEE SP, Apr. 2001.
- *ZZB* : J. Ziv, "Some lower bounds on Signal Processing, IEEE IT, May 1969.
- *Outliers effect* : D.C. Rife, "Single-tone parameter estimation from discrete-time observations", IEEE IT, Sep. 1974.

Channel estimation and Superresolution in UWB system

Philippe Ciblat

École Nationale Supérieure des Télécommunications, Paris, France

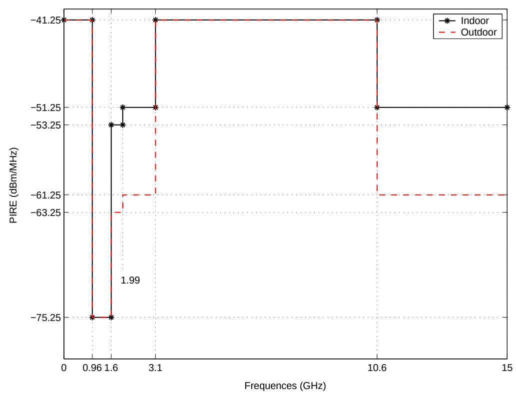
NEWCOM Autumn School :
Estimation Theory for wireless communications

Outline

1. UWB system
   - Impulse Radio
   - Multi-band
   - Channel Model
2. Channel estimation
   - Cramer-Rao Bound
   - Existing estimates
   - Comparison
3. Superresolution

Introduction

Digital communications system satisfies the following spectral mask :



Interest

- Spread spectrum technique
- Localization

Techniques

Approaches

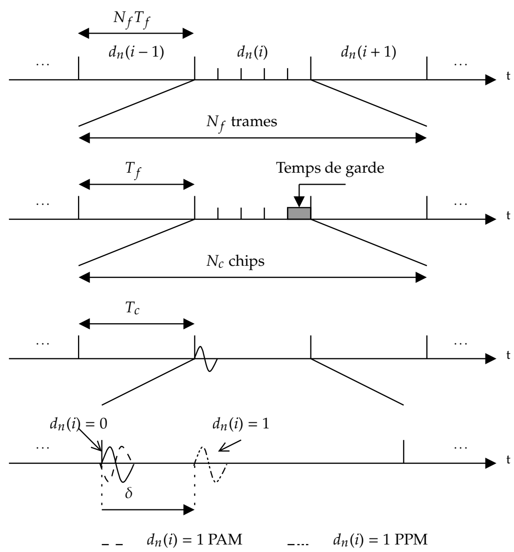
- Impulse Radio (IR)
- Multi-band (MB)

We hereafter focus on Impulse-Radio technique

- Pierce and Hopper 1952
- Winthington and Fullerton 1992
- Win and Scholtz 1993

IR-UWB transmit signal

- Time-Hopping (TH) IR-UWB signal associated with user n



Data stream

$$s(t) = \sum_{i=0}^{M-1} d_i b(t - iN_f T_f)$$

where

- M is the number of transmit symbols
- $\mathbf{d} = [d_0, \cdots, d_{M-1}]$ belongs to PAM
- N_f is the number of frame per symbol
- T_f is the duration of each frame

Superframe structure

The super frame composed by N_f frames is structured as follows

$$b(t) = \sum_{j=0}^{N_f-1} g(t - jT_f - \tilde{c}_j T_c)$$

where

- T_c is the chip duration
- N_c is the number of chips in one frame
- Time-hopping code in the j^{th} frame is given by $\tilde{c}_j \in \{0, \cdots, N_c - 1\}$
- $g(t)$ is the mono-cycle with the temporal support $[0, T_g)$

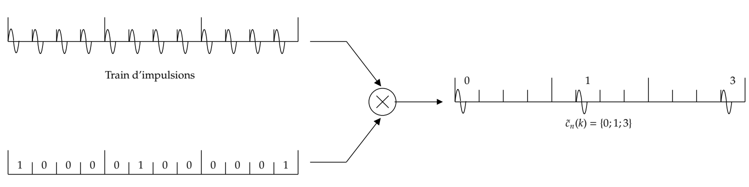
Developed code

For each frame j, let $\mathbf{c}_j = [c_j(0), \cdots, c_j(N_c - 1)]$ defined as follows

$$c_j(i) = \begin{cases} 1 & \text{if } i = \tilde{c}_j \\ 0 & \text{otherwise} \end{cases}.$$

Then $\mathbf{c} = [\mathbf{c}_0, \cdots, \mathbf{c}_{N_f-1}] = [c(0), \cdots, c(N_f N_c - 1)]$

$$s(t) = \sum_{i=0}^{M-1} d_i \sum_{j=0}^{N_f N_c-1} c(j) g(t - jT_c - iN_f T_f)$$



- Status of the chip (occupied/free) outside $g(t)$
- Le Martret & Giannakis 2002

Channel model

- Multi-path random channel
- Molish 2003

Impulse response

$$h(t) = \sum_{k=1}^{N_p} A_k \delta(t - \tau_k)$$

where

- A_k is the attenuation associated with the k^{th}-path
- τ_k is the delay associated with the k^{th}-path

Statistical channel model

- We focus on one cluster model

Statistical model

$$p(\tau_k|\tau_{k-1}) = \lambda e^{-\lambda(\tau_k - \tau_{k-1})}$$

$$A_k = \underbrace{(p_k.b_k)}_{a_k} e^{-\tau_k/\gamma}$$

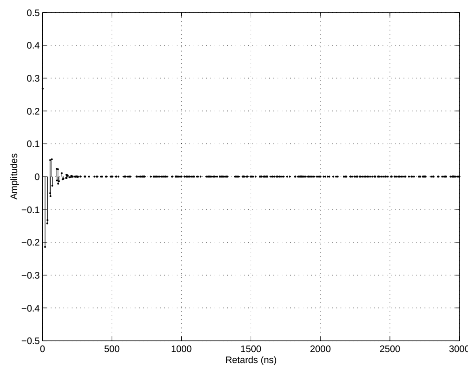
where

- a_k independent of τ_n^k
- p_k binary variable
- b_k log-normal variable

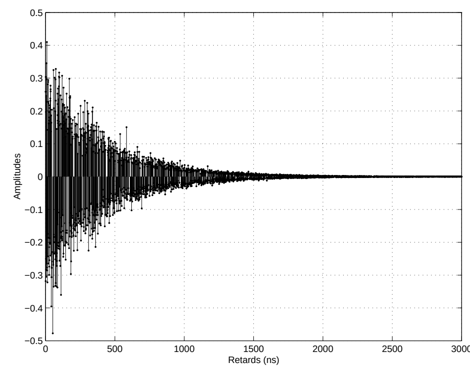
λ and γ are both deterministic parameters

Deterministic parameters

- λ is the path density
- γ is the RMS delay spread (i.e., length of impulse response)



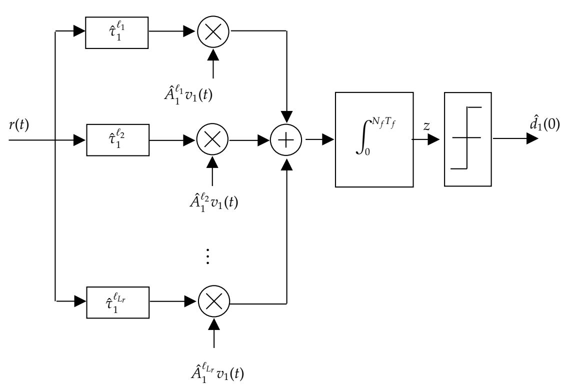
$\lambda = 0.1\mathrm{ns}^{-1}$ and $\gamma = 20\mathrm{ns}$



$\lambda = 1\mathrm{ns}^{-1}$ and $\gamma = 200\mathrm{ns}$

Receiver

- Rake receiver (for sake of simplicity)
- Correlation with the template $b(t) = \sum_{j=0}^{N_f N_c - 1} c_j g(t - jT_c)$ synchronized at each path



Path estimation is necessary

Fisher Information Matrix

$$J_{A_l,A_k} = \frac{2}{N_0} f_1^{(k,l)}, J_{A_l,\tau_k} = -\frac{2A_k}{N_0} f_2^{(l,k)}, J_{\tau_l,\tau_k} = \frac{2A_k A_l}{N_0} f_3^{(k,l)}$$

where

$$\begin{aligned} f_1^{(k,l)} &= \mathbb{E}_{\mathbf{d}}\left[\int s(t-\tau_k)s(t-\tau_l)dt\right] \\ f_2^{(k,l)} &= \mathbb{E}_{\mathbf{d}}\left[\int s(t-\tau_k)s'(t-\tau_l)dt\right] \\ f_3^{(k,l)} &= \mathbb{E}_{\mathbf{d}}\left[\int s'(t-\tau_k)s'(t-\tau_l)dt\right] \end{aligned}$$

with

- $s'(t) = ds(t)/dt$ and $\mathbb{E}_{\mathbf{d}}[\phi(\mathbf{d})] = \phi(\mathbf{d})$ if $\mathbf{d}$ is a known sequence

$\rightsquigarrow$ *CRB for DA scheme and MCRB for NDA scheme*

State-of-the-Art

1. Laurenti (September 2004) : one path
2. Huang (June 2004) : non-overlapping context (i.e., signal echoes are orthogonal)

$$f_m^{(k,l)} = 0 \quad \text{if} \quad k \neq l$$

3. Zhang (June 2004) : overlapping taken into account (but no closed-form expression for FIM)

Questions

- Non-overlapping assumption does not hold in realistic situation ?
- Closed-form expressions for $f_m^{(k,l)}$ even when $k \neq l$

Non-overlapping case

Straightforward derivations yield

$$\mathrm{CRB}_{\mathrm{DA}}(A_l) = \mathrm{MCRB}_{\mathrm{NDA}}(A_l) = \frac{N_0}{MN_f}\frac{E_3}{2(E_1E_3 - E_2^2)}$$

$$\mathrm{CRB}_{\mathrm{DA}}(\tau_l) = \mathrm{MCRB}_{\mathrm{NDA}}(\tau_l) = \frac{N_0}{MN_f}\frac{E_1}{2A_l^2(E_1E_3 - E_2^2)}$$

with $E_1 = \int g(t)^2 dt$, $E_2 = \int g(t)g'(t)dt$, and $E_3 = \int g'(t)^2 dt$

Remarks

- In DA scheme, performance does not depend on the training sequence
- Same expression in the context of single-path (when $N_p = 1$)

Overlapping case

Let

- $\Delta\tau_{k,l} = \tau_k - \tau_l = Q_{k,l}N_fT_f + q_{k,l}T_c + \varepsilon_{k,l}$ with the integer parts $Q_{k,l}$ and $q_{k,l}$, and the remainder $\varepsilon_{k,l}$

Main result

$$\begin{aligned} f_m^{(k,l)} &= M(\mathcal{C}(q)\mathcal{A}_m(\varepsilon) + \mathcal{C}(q+1)\mathcal{A}_m(\varepsilon - T_c) \\ &+ \mathcal{D}(q)\mathcal{B}_m(\varepsilon) + \mathcal{D}(q+1)\mathcal{B}_m(\varepsilon - T_c)) \end{aligned}$$

with

- $$\mathcal{C}(q) = \sum_{j=0}^{N_fN_c-q-1} c(j)c(j+q), \quad \mathcal{D}(q) = \sum_{j=0}^{q-1} c(j)c(j-q)$$
- $$\mathcal{A}_m(\varepsilon) = \frac{1}{M}\sum_{i=0}^{M-1}\mathbb{E}_{\mathbf{d}}[d_{-Q-1+i}d_i]r_m(\varepsilon), \ \mathcal{B}_m(\varepsilon) = \frac{1}{M}\sum_{i=0}^{M-1}\mathbb{E}_{\mathbf{d}}[d_{-Q+i}d_i]r_m(\varepsilon)$$
- $r_1(t) = g(t)\star g(-t)$, $r_2(t) = g'(t)\star g(-t)$, $r_3(t) = g'(t)\star g'(-t)$

Comments

- Code collisions plays an important role.
- The more $f_m^{k,l}$) (for $k \neq l$) is high, the more the CRB is high
- If $\varepsilon \in [T_g, T_c - T_g]$, there is no overlapping
- The more the path is dense, the more the CRB taking into account the overlapping is larger than the (simplified) CRB
- Deleuze & Ciblat & Le Martret (July 2004)

Average CRB (I)

$$\mathbb{E}_{\mathbf{x}}[\mathrm{CRB}(\mathbf{x})] = \mathbb{E}_{\mathbf{x}}[J(\mathbf{x})^{-1}] \geq (\mathbb{E}_{\mathbf{x}}[J(\mathbf{x})])^{-1}$$

Simplified expressions for $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ by averaging over

- *symbol sequence*
- *time-hopping code*

⇝ *In DA scheme, average CRB over all possible training sequences*

⇝ *In NDA scheme, MCRB is considered*

Average CRB (II)

- $\{d(i)\}_i$ i.i.d. symbols belonging to 2-PAM

Result

$$\mathbb{E}_{\mathbf{d}}[\mathcal{A}_m(\varepsilon)] = \delta_{Q,-1} r_m(\varepsilon), \quad \mathbb{E}_{\mathbf{d}}[\mathcal{B}_m(\varepsilon)] = \delta_{Q,0} r_m(\varepsilon)$$

- $\mathbf{c}_j$ is the realization of i.i.d. random vector whose each component admits the following distribution $p(c) = ((N_c - 1)\delta(c) + \delta(c-1))/N_c$.

Result

$$\left\{ \begin{array}{ll} \mathbb{E}_{\mathbf{c}}[\mathcal{C}(q)] = \frac{N_f N_c - q}{N_c^2} & \text{if } q \neq 0 \\ \mathbb{E}_{\mathbf{c}}[\mathcal{C}(0)] = N_f & \text{if } q = 0 \end{array} \right. , \quad \left\{ \begin{array}{ll} \mathbb{E}_{\mathbf{c}}[\mathcal{D}(q)] = \frac{q}{N_c^2} & \text{if } q \neq N_f N_c \\ \mathbb{E}_{\mathbf{c}}[\mathcal{D}(N_f N_c)] = N_f & \text{if } q = N_f N_c \end{array} \right.$$

Maximum Likelihood

- Lottici & Andrea & Mengali 2002
- No overlapping context
- Simulations done in a non-overlapping context
- ML carried out in DA and NDA schemes
  - DA scheme : derivations based on likelihood
  - NDA scheme : derivations based on true likehood at low SNR

Algorithm

$$J_{\mathrm{NDA}}(\tau) = \frac{1}{ME_b} \sum_{i=0}^{M-1} \frac{z_i(\tau, d_i = -1) + z_i(\tau, d_i = 1)}{2}$$

with $z_i(\tau, d_i) = d_i(r(t) \star b(-t)_{|t = iN_f T_f + \tau})$

- Localizations of peaks provide $\hat{\tau}$
- Magnitudes of peaks provide $\hat{A}$

Undersampling based method (I)

- Maravic & Vetterli 2003
- DA scheme
- Undersampling at period $T_s \gg T_p$ preceded by Anti-Aliasing Filter

Let $\tilde{r}(t)$ the noiseless receiver signal at the output of AAF

$$\tilde{R}(m) = \text{F.T.}(t \mapsto \tilde{r}(t))_{|f=mf_0} = \sum_{k=1}^{N_p} A_k \tilde{S}(m) e^{-2i\pi\tau_k m f_0}$$

then

$$\tilde{R}_s(m) = \tilde{R}(m)/\tilde{S}(m) = \sum_{k=1}^{N_p} A_k z_k^m$$

with $z_k = e^{-2i\pi\tau_k f_0}$

Undersampling based method (II)

$$\mathbf{R} = \begin{bmatrix} \tilde{R}_s(0) & \tilde{R}_s(1) & \cdots & \tilde{R}_s(N_p-1) \\ \tilde{R}_s(1) & \tilde{R}_s(2) & \cdots & \tilde{R}_s(N_p) \\ \vdots & \vdots & & \vdots \\ \tilde{R}_s(N_p-1) & \tilde{R}_s(N_p) & \cdots & \tilde{R}_s(2N_p-2) \end{bmatrix} \Leftrightarrow [\mathbf{R}]_{\ell,\ell'} = \sum_{k=1}^{N_p} A_k z_k^{\ell+\ell'}$$

Then

$$\mathbf{R} = V \Lambda V^{\mathrm{H}} \quad \text{with} \quad V = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ z_1^{N_p-1} & \cdots & z_{N_p}^{N_p-1} \end{bmatrix}$$

Undersampling based method (III)

Shift invariance

$$\overline{V} = \underline{V}\mathrm{diag}([z_1, \cdots, z_{N_p}])$$

where $\overline{V}$ and $\underline{V}$ denote the omition of the first and last row of V respectively

Then it exists a vector $\mathbf{x}_k$ such that

$$\overline{V}\mathbf{x}_k = z_k \underline{V}\mathbf{x}_k$$

$\rightsquigarrow$ *z_k is a generalized eigenvalue of $(\overline{V}, \underline{V})$*

Algorithm

For any k, z_k is the root of the polynomial

$$P(s) = \det(\overline{V} - s\underline{V})$$

This obviously provides $\hat{\tau}$ and $\hat{A}$

First-order cyclostationarity based method (I)

- Luo & Giannakis 2004
- Asymmetric PAM ($d_i \in \{-1, \theta\}$)
- ISI-less context ($\text{delay spread} < \text{guard-time}$)

$$r(t) = \sum_{i=0}^{M-1} d_i b_r(t - \tau_1 - iN_f T_f) \quad \text{with} \quad b_r(t) = \sum_{k=1}^{N_p} A_k b(t - \Delta\tau_{k,1})$$

If ISI-less, $\{b_r(t - \tau_1 - iN_f T_f)\}_i$ is a orthogonal set and thus $b_r(t)$ is a square-root Nyquist filter.

Problem

- *Optimal receiver is the matched filter $b_r(-t)$ shifted by τ_1*
- *Knowledge of $b_r(t)$ and τ_1 is needed*

First-order cyclostationarity based method (II)

$$\mathbb{E}[r(t)] = \frac{\theta - 1}{2} \sum_{i=0}^{M-1} b_r(t - \tau_1 - iN_f T_f)$$

The cyclostationary mean contains information about $b_r(t)$ and τ_1

Algorithm

If τ_1 is associated with the strongest path, then

$$\hat{\tau}_1 = \arg \max_{\tau \in [0, N_f T_f)} \left| \int_0^{2N_f T_f} \widehat{\mathbb{E}[r(t)]} b(t - \tau) dt \right|$$

and

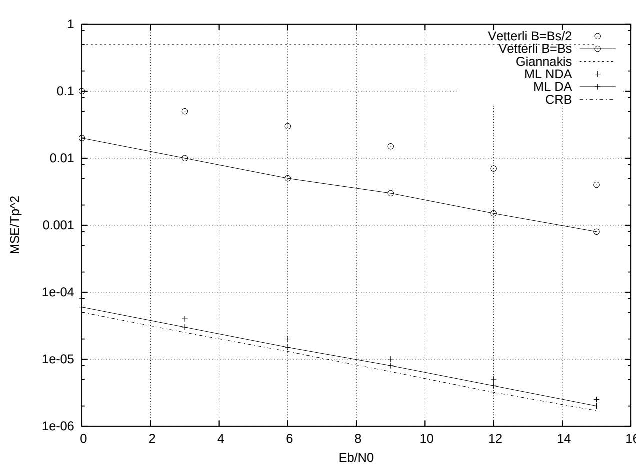
$$\hat{b}_r(t) = \frac{2}{\theta - 1} \mathbb{E}[\widehat{r(t + \hat{\tau}_1)}], \quad \text{for} \quad t \in [0, N_f T_f)$$

Non-overlapping case

Set-up

- $T_p = 1\text{ns}$, $T_c = 2T_p$, $N_c = 10$, and $N_f = 10$, $T_s = 200\text{ns}$, $M = 100$
- $\boldsymbol{\tau} = [5T_p, 10T_p, 15T_p]$ and $\mathbf{A} = [0.73, 0.67, 0.35]$

Such assumptions ensure the absence of overlapping

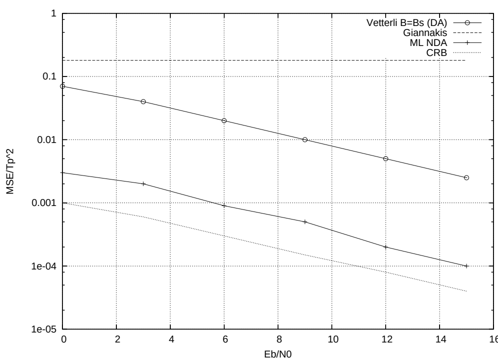


Overlapping case

Set-up

- $\boldsymbol{\tau} = \{kT_p/2\}_{k=1,\cdots,20}$
- $\mathbf{A}$ obeys a normalized exponential decreasing profile

Such assumptions ensure the presence of overlapping



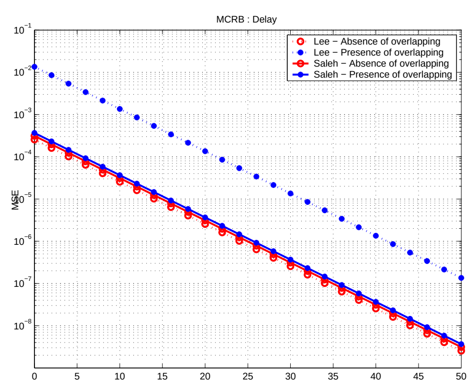
$\rightsquigarrow$ *ML non optimal in overlapping case*

Comparison

Question

Is there overlapping or not in realistic channel ?

Two statistical models :
Molish ($\lambda = 0.2\mathrm{ns}^{-1}$, $\gamma = 20\mathrm{ns}$) and Lee ($\lambda = 2\mathrm{ns}^{-1}$, $\gamma = 5\mathrm{ns}$)



$\rightsquigarrow$ *If path density is high, the non-overlapping model does not hold*

Definition

- *The superresolution is the smallest gap between two delays that we are able to distinguish from*
- The Cramer-Rao Bound $\mathrm{CRB}(\tau)$ is the smallest mean square error that we may reach when the value of the sought delay is τ

Superresolution definition

The superresolution $\tau_{\mathrm{res.}}$ satisfies the following equation

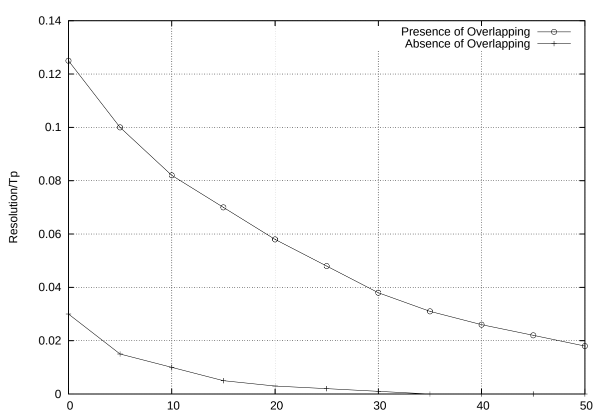
$$\tau_{\mathrm{res.}} = \sqrt{\mathrm{CRB}(\tau_{\mathrm{res.}})}$$

- When τ decreases, the overlapping increases
- To evaluate accurately the superresolution, we need the closed-form expression of $\mathrm{CRB}(\tau)$ in overlapping case

Superresolution versus SNR

Set-up

- $\boldsymbol{\tau} = [0 \tau]$, $\mathbf{A} = [1 0.5]$, and $M = 100$

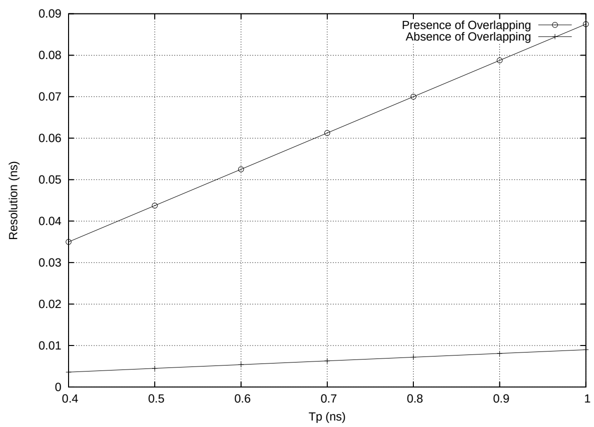


$\rightsquigarrow$ *Non-overlapping is too optimistic and does not make sense*

Superresolution versus T_p

Set-up

- $E_b/N_0 = 10\text{dB}$ and $M = 100$



$\rightsquigarrow$ *Resolution proportional to T_p*

Bibliography

- *CRB derivations* :
  - L. Huang et al., "Performance of ML channel estimator for UWB communications", IEEE COML, Jun. 2004.
  - J. Zhang et al., "CRB for time-delay estimation of UWB signals", ICC, Jun. 2004.
  - A.L. Deleuze, P. Ciblat et al., "CRB for channel parameters in UWB based system, IEEE SPAWC, Jul. 2004.
  - N. Laurenti et al., "On the performacne of TH-PPM and TH-PAM as transmission formats for UWB commmunications, IEEE VTC, Sept. 2004.
- *Estimator design* :
  - M.Z. Win and R.A. Scholtz, "On the energy capture of UWB signals in desne multipath environments", IEEE COML, Sept. 1998.
  - V. Lottici et al., "Channel estimation for UWB communications", IEEE TCOM, Dec. 2002.
  - I. Maravic, M. Vetterli et al., "High resolution acquisition methods for wideband communication systems", IEEE ICASSP, Apr. 2003.
  - X. Luo and G. Giannakis, "Blind timing and channel estimation for UWB multi-user ad hoc access", Asilomar Conference, Nov. 2004.
- *Comparison* : K. Bouchireb, "systèmes UWB : résolution de trajets pour la localisation dans les réseaux ad hoc", Master thesis (in French), Sept. 2005.

Estimation Theory for Wirelss Communication, 24-28 Oct 2005, PARIS.

Channel Estimation for Cyclic Prefixed Block Transmissions

M. Ghogho University of Leeds, UK

Aims

- To describe a few channel estimation techniques for cyclic-prefixed (CP) block transmissions, including OFDM and single-carrier (SC-)CP systems
- To address the issue of optimum training design and power allocation
- To introduce a new bandwidth efficient pilot assisted transmission technique

Outline

- Introduction
- Channel estimation for OFDM
  - OFDM signal model and preliminaries
  - Pilot-based channel estimation for OFDM
  - Blind channel estimation for OFDM
- Channel estimation for general CP systems
  - Affine precoding and MMSE channel estimation
  - Full rank orthogonal precoding
  - Rank-deficient orthogonal precoding
- Summary

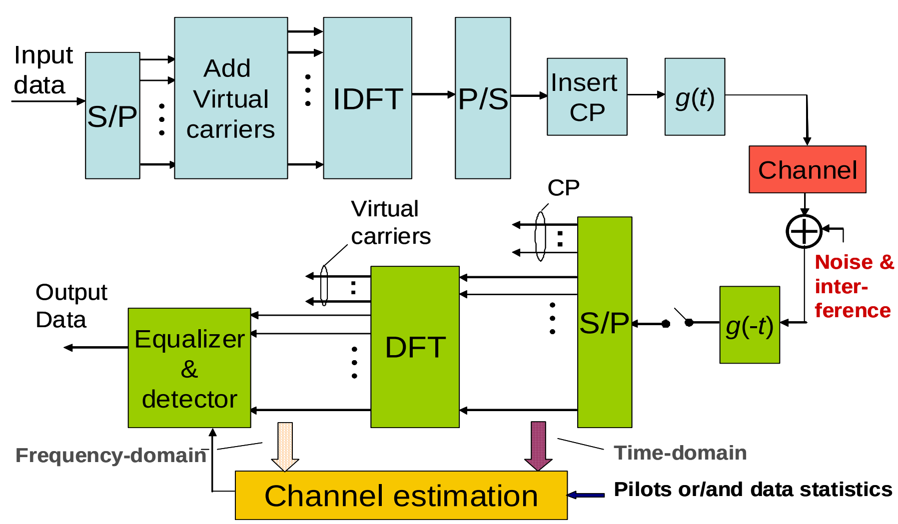
Introduction

- Why block transmissions?
  - existence of zero-forcing equalizer
  - block-by-block processing
- Why cyclic prefix?
  - FFT-based channel equalization
- Why channel estimation
  - required for coherent communication systems

Part 1: Channel Estimation for OFDM

OFDM signal model and preliminaries

- Block diagram



OFDM signal model and preliminaries (2)

- Frequency-domain (F-D) methods: either pilot-based or (semi-)blind
- Time-domain (T-D), generally (semi-)blind.

Assumptions:

- Channel impulse response (CIR) constant during each OFDM symbol

$$h(t) = \sum_{\ell=0}^{L} h_\ell \delta(t - \tau_\ell)$$

- $\tau_\ell = \ell T_s$, $T_s = T/N$ and T: duration of 1 OFDM block w/0 CP.
- $\boldsymbol{h} := [h_0 \cdots h_L]^T \sim \mathcal{CN}(0, \mathbf{R}_h)$, $\mathbf{R}_h = \text{diag}\{\sigma^2_{h_\ell}, \ell = 0 \cdots L\}$
- length of CP $= L$. Additive noise is Gaussian and white with variance σ_v^2.

OFDM signal model and preliminaries (3)

- Notations

- N: DFT size • N_a: # active carriers • N_p: # pilot carriers
- $\mathcal{A}$ ($\mathcal{P}$): set of active (pilot) carriers; $\mathcal{P} \subseteq \mathcal{A} \subseteq \{0, \cdots N-1\}$
- $\mathbf{F} = (1/\sqrt{N})\{\exp(-j2\pi nk/N)\}_{n,k=0}^{N-1}$ • $\mathbf{W} = (\sqrt{N})\mathbf{F}(:, 0 : L)$
- $\mathbf{T}_a$: active carriers selection matrix ($N \times N_a$)
- $\mathbf{T}_p$: pilot carriers selection matrix ($N \times N_p$)
- $\mathbf{T}_d$: data carriers selection matrix ($N \times N_d$) with $N_d = N_a - N_p$
- $\mathbf{W}_a = \mathbf{T}_a^T \mathbf{W}$ • $\mathbf{W}_\mathcal{P} = \mathbf{T}_p^T \mathbf{W}$ • $\mathbf{W}_\mathcal{D} = \mathbf{T}_d^T \mathbf{W}$
- σ_p^2 (resp σ_s^2) total power of pilot (resp. data) carriers; ;$\sigma_t^2 := \sigma_p^2 + \sigma_s^2$.
- $\mathbf{D}_{\boldsymbol{z}} = \text{diag}\{\boldsymbol{z}\}$.

OFDM signal model and preliminaries (4)

- VC insertion: $\mathbf{T}_{vc}$: N_a columns of a $N \times N$ identity matrix
- CP insertion: $\mathbf{T}_{cp} = \begin{bmatrix} \mathbf{0}_{L\times(N-L)}, & \mathbf{I}_L \\ \mathbf{I}_N & \end{bmatrix}$
- Transmitted block: $\boldsymbol{u}_{\text{cp}}(i) = \mathbf{T}_{cp}\mathbf{F}^{\mathcal{H}}\mathbf{T}_{sc}\boldsymbol{s}(i)$
- Input-output relationship $(N \geq N_a, P = L + N)$

$$x_{\text{cp}}(n) = \sum_{l=0}^{L} h(l)u_{\text{cp}}(n-l) + v_{\text{cp}}(n)$$

OFDM signal model and preliminaries (5)

- Received blocks

$$\boldsymbol{x}_{\text{cp}}(i) = [\mathbf{H}_1\boldsymbol{u}_{\text{cp}}(i) + \mathbf{H}_2\boldsymbol{u}_{\text{cp}}(i-1)] + \boldsymbol{v}(i)$$

- Discard CP to avoid IBI: $\mathbf{R}_{cp} := [\mathbf{0}_{N\times(P-N)}, \mathbf{I}_N] \rightarrow \mathbf{R}_{cp}\mathbf{H}_2 = \mathbf{0}$.
- Channel matrix: $\mathbf{H}_1$ Toeplitz $\Rightarrow$ $\mathbf{H}_c = \mathbf{R}_{cp}\mathbf{H}_1\mathbf{T}_{cp}$ circulant; so

$$\mathbf{F}\mathbf{H}_c\mathbf{F}^{\mathcal{H}} = \text{diag}(H_0 \cdots H_{N-1}) =: \mathbf{D}_H$$

  where $H_k = \sum_{\ell=0}^{L} h_\ell e^{-j2\pi\ell k/N}$
- Received blocks after CP removal

$$\boldsymbol{x}(i) \;=\; \mathbf{R}_{cp}\boldsymbol{x}_{\text{cp}}(i) = \mathbf{F}^{\mathcal{H}}\mathbf{D}_H\mathbf{T}_{sc}\boldsymbol{s}(i) + \boldsymbol{v}(i)$$

  and after FFT

$$\tilde{\boldsymbol{x}}(i) = \mathbf{D}_H\mathbf{T}_{sc}\boldsymbol{s}(i) + \tilde{\boldsymbol{v}}(i)$$

OFDM signal model and preliminaries (6)

- F-D received signal at the data carriers (dropping block index):

$$\tilde{x}_n = H_n s_n + \tilde{v}_n \qquad n \in \mathcal{D}$$

s_n: data symbol on nth carrier and $H_n = \sum_{\ell=0}^{L} h_\ell e^{-2j\pi \ell n/N}$.

- F-D signal at the pilot carriers, $\mathcal{P} = \{i_1, \cdots, i_{N_p}\} \subseteq \mathcal{A} = \mathcal{D} \cup \mathcal{P}$,

$$\tilde{x}_{i_m} = H_{i_m} c_m + \tilde{v}_{i_m}, \qquad m = 1, \cdots, N_p$$

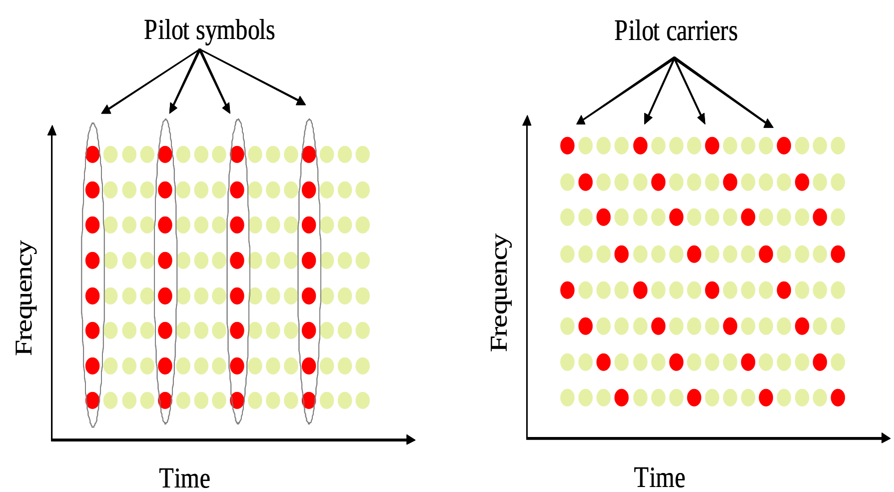
In vector form:

$$\tilde{\boldsymbol{x}}_{\mathcal{P}} = \mathbf{D}_{\boldsymbol{c}} \mathbf{W}_{\mathcal{P}} \boldsymbol{h} + \tilde{\boldsymbol{v}}_{\mathcal{P}}$$

$\boldsymbol{c} = [c_1, \cdots, c_{N_p}]^T$: known pilot symbols.

Pilot-Based Channel Estimation for OFDM

- Pilot placement



- Time-invariant or slowly varying channels - Time-varying channels

Pilot-Based Channel Estimation for OFDM (2)

- Minimum Mean Square Error (MMSE) Method

❑ MMSE CIR estimator

$$\hat{\boldsymbol{h}} = \left(\sigma^2 \mathbf{R}_h^{-1} + \mathbf{W}_{\mathcal{P}}^{\mathcal{H}} \mathbf{D}_{\boldsymbol{\rho}} \mathbf{W}_{\mathcal{P}}\right)^{-1} \mathbf{W}_{\mathcal{P}}^{\mathcal{H}} \mathbf{D}_{\boldsymbol{c}}^{\mathcal{H}} \tilde{\boldsymbol{x}}_{\mathcal{P}}$$

where $\mathbf{D}_{\boldsymbol{\rho}} = \mathrm{diag}\{|c_m|^2,\ m = 1 \cdots N_p\}$.

☞ The least square (LS) estimator is obtained by setting $\mathbf{R}_h^{-1} = 0$.

❑ Identifiability condition (since $c_m \neq 0$):

$$\mathrm{rank}\,(\mathbf{D}_{\boldsymbol{c}} \mathbf{W}_{\mathcal{P}}) = L + 1 \quad \Longleftrightarrow \quad N_p \geq L + 1$$

❑ MMSE estimate of H_n

$$\hat{H}_n = \boldsymbol{w}_n^{\mathcal{H}} \hat{\boldsymbol{h}}$$

where $\boldsymbol{w}_n^{\mathcal{H}} := \mathbf{W}(n, :)$

Pilot-Based Channel Estimation for OFDM (3)

- Performance of MMSE estimates

❑ MSEs of $\hat{\boldsymbol{h}}$ and the $\hat{H}_n$'s:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{h}}} := E\left\{(\hat{\boldsymbol{h}} - \boldsymbol{h})(\hat{\boldsymbol{h}} - \boldsymbol{h})^{\mathcal{H}}\right\} = \left(\mathbf{R}_h^{-1} + \frac{1}{\sigma_v^2} \mathbf{W}_{\mathcal{P}}^{\mathcal{H}} \mathbf{D}_{\boldsymbol{\rho}} \mathbf{W}_{\mathcal{P}}\right)^{-1}$$

$$\gamma_n := E\left\{|\hat{H}_n - H_n|^2\right\} = \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{\Sigma}_{\hat{\boldsymbol{h}}} \boldsymbol{w}_n$$

$$\bar{\gamma}_{\mathrm{mmse}} := \sum_{n \in \mathcal{D}} \gamma_n = \mathrm{Tr}\left\{\mathbf{W}_{\mathcal{D}} \boldsymbol{\Sigma}_{\hat{\boldsymbol{h}}} \mathbf{W}_{\mathcal{D}}^{\mathcal{H}}\right\}$$

☞ MSEs of LS estimates obtained by setting $\mathbf{R}_h^{-1} = 0$

Pilot-Based Channel Estimation for OFDM (4)

- Optimum pilot design for MMSE channel estimation

❑ Equalization carried out in F-D; so criterion based on γ_n. Minimizing the total (or average) mse:

$$\begin{aligned}\{\boldsymbol{\rho}^o, \mathcal{P}^o\} &= \arg\min_{\boldsymbol{\rho},\mathcal{P}} \bar{\gamma}_{\text{mmse}} \\ &= \arg\min_{\boldsymbol{\rho},\mathcal{P}} \text{Tr}\left\{\mathbf{W}_{\mathcal{D}}\left(\mathbf{R}_h^{-1} + \frac{1}{\sigma_v^2}\mathbf{W}_{\mathcal{P}}^{\mathcal{H}}\mathbf{D}_{\boldsymbol{\rho}}\mathbf{W}_{\mathcal{P}}\right)^{-1}\mathbf{W}_{\mathcal{D}}^{\mathcal{H}}\right\}\end{aligned}$$

under the constraints

$$\mathcal{P} \subseteq \mathcal{A}; \qquad \sum_{n=1}^{N_p} \rho_n = \sigma_p^2 \qquad (\mathbf{C1})$$

Pilot-Based Channel Estimation for OFDM (5)

- Optimum pilot design for MMSE channel estimation: no VC

❑ For any $(L \times L)$ positive-definite matrix, $\mathbf{B} = \{b_{k,\ell}\}_{k,\ell=0}^{L}$, we have

$$\text{Tr}\left\{\mathbf{B}^{-1}\right\} \geq \sum_{\ell=0}^{L} \frac{1}{b_{\ell,\ell}}$$

with equality iff $\mathbf{B}$ is diagonal.

❑ Since $\mathbf{R}_h$ is diagonal, $\bar{\gamma}_{\text{mmse}}$ is minimized if

$$\mathbf{W}_{\mathcal{D}}^{\mathcal{H}}\mathbf{W}_{\mathcal{D}} = N_d\mathbf{I} \ \text{ and } \ \mathbf{W}_{\mathcal{P}}^{\mathcal{H}}\mathbf{D}_{\boldsymbol{\rho}}\mathbf{W}_{\mathcal{P}} = \sigma_p^2\mathbf{I}$$

which is possible in the no-VC case

Pilot-Based Channel Estimation for OFDM (6)

- Optimum pilot design for MMSE channel estimation: no VC (cont.)

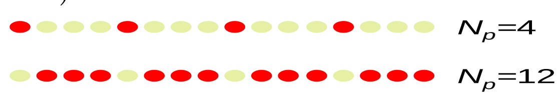
❑ An optimum design is

$$\boldsymbol{\rho}^o = \frac{\sigma_p^2}{N_p}\mathbf{1}^T$$

$$\mathcal{P}^o = \begin{cases} \mathcal{P}_1^o := \{t + iQ,\ i = 0, \cdots, N_p - 1\} & \text{if } Q := \frac{N}{N_p} \text{ integer} \\ \\ \mathcal{P}_2^o := \{0, \cdots, N-1\} - \mathcal{P}_1^o & \text{if } Q := \frac{N}{N-N_p} \text{ integer} \end{cases}$$

where t is arbitrary integer from $[0, Q-1)$.

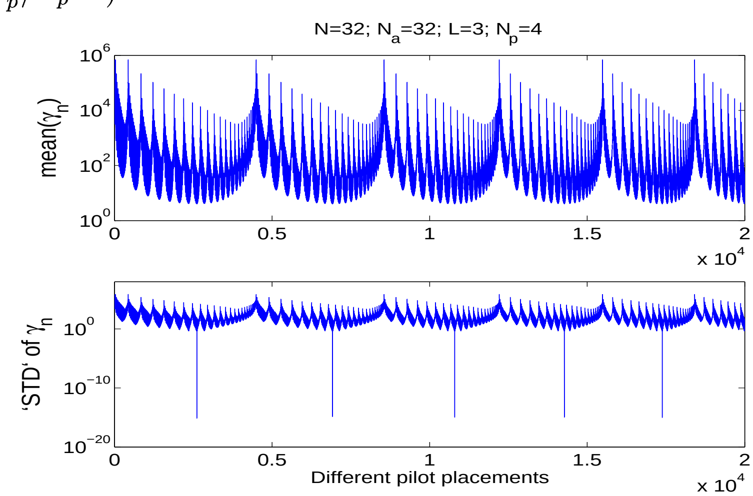
Example: $(N = 16)$



Pilot-Based Channel Estimation for OFDM (7)

- Optimum pilot design for MMSE channel estimation: no VC (cont.)

Illustration of the effect of pilot placement on estimation performance $(\boldsymbol{\rho} = \sigma_p^2/N_p\mathbf{1}^T)$



Pilot-Based Channel Estimation for OFDM (8)

- Optimum pilot design for MMSE channel estimation: no VC (cont.)

❑ The minimum of $\bar{\gamma}_{\text{mmse}}$ is (using $N_d = N - N_p$ since no VC)

$$\bar{\gamma}^o_{\text{mmse}} = (N - N_p)\gamma^o = (N - N_p)\sum_{\ell=0}^{L} \frac{\sigma_v^2\ \sigma_{h_\ell}^2}{\sigma_v^2 + \sigma_p^2\sigma_{h_\ell}^2}$$

☞ The MSE, $\bar{\gamma}_{\text{LS}}$, of LS estimate obtained using $\sigma^2_{h_\ell} = \infty$.

☞ Pilot design minimizing $\bar{\gamma}_{\text{mmse}}$ also minimizes the γ_n's individually, and with optimal design, all carriers experience the same channel estimation MSE, i.e. $\gamma^o_n = \gamma^o$.

☞ Minimizations of the MSE in the F-D and T-D are equivalent:

$$\arg\min_{\boldsymbol{\rho},\mathcal{P}} \bar{\gamma}_{\text{mmse}} = \arg\min_{\boldsymbol{\rho},\mathcal{P}} \text{Tr}\left\{\boldsymbol{\Sigma}_{\hat{\boldsymbol{h}}}\right\}$$

Pilot-Based Channel Estimation for OFDM (9)

- Optimum pilot design for MMSE channel estimation: no VC (cont.)

☞ For fixed σ_p^2 and with $\boldsymbol{\rho}^o$ and $\mathcal{P}^o$, γ^o is independent of N_p. But this is not exactly true if there is a mismatch between the assumed and the actual channel models, e.g. fractional path delays!

☞ Optimum pilot placement and power distribution design not unique, in general. However if $N_p = L + 1$, only equipowered and equispaced pilot carriers achieve minimum MSE.

☞ In the case of colored noise with unknown spectral density, use pilot carrier hopping, e.g. t in the above optimum design should vary across the blocks.

Pilot-Based Channel Estimation for OFDM (10)

- Under the above optimal placement and power distribution of the pilots, what are the optimal value of N_p, the optimal power allocation and the optimal data power distribution? We use a capacity-bound criterion

- Channel unknown at transmitter $\Rightarrow$ ideal training-based capacity maximized when $\sigma_s^2(n) := E\left\{|s_n|^2\right\} = \sigma_s^2/N_d$:

$$C_{\text{ideal}} = \frac{N_d}{N+L} E\left\{\log\left(1+\beta_{\text{ideal}}|g|^2\right)\right\} \qquad \text{(bits/symbol)}$$

where $g \sim \mathcal{CN}(0,1)$ and β_{ideal} is the ideal SNR ($\sigma_H^2 = \sum_\ell \sigma_{h_\ell}^2$)

$$\beta_{\text{ideal}} := \frac{\sigma_H^2 \sigma_s^2}{N_d \sigma_v^2}$$

Pilot-Based Channel Estimation for OFDM (11)

- Incorporating estimation error into signal model

- Treating estimation error as extra noise:

$$\tilde{x}_n = H_n s_n + \tilde{v}_n = \hat{H}_n s_n + \underbrace{e_n s_n}_{\text{extra noise}} + \tilde{v}_n$$

where $e_n = \hat{H}_n - H_n$ and $E\left\{|e_n s_n + \tilde{v}_n|^2\right\} = \gamma_n \sigma_s^2(n) + \sigma_v^2$.

- Orthogonality principle: $E\left\{\hat{H}_n e_n\right\} = 0$. Thus

$$E\left\{|\hat{H}_n|^2\right\} = \sigma_H^2 - \gamma_n < \sigma_H^2$$

☞ Equivalent to a *known channel* $\hat{H}_n$ system subjected to an additive noise $\tilde{v}_n' = e_n s_n + \tilde{v}_n$ which is *neither Gaussian nor independent* (though uncorrelated) of the data.

Pilot-Based Channel Estimation for OFDM (12)

- Effect of estimation on capacity

❑ Since noise $\tilde{v}'_n$ is *uncorrelated* from data, the capacity is lower bounded by that of a system subjected to Gaussian noise with same power as $\tilde{v}'_n$.

$$C > \underline{C} = \frac{1}{N+L} \sum_{n \in \mathcal{D}} E\left\{\log\left(1 + \beta(n)|g|^2\right)\right\}$$

where $\beta(n)$ effective SNR at nth carrier

$$\beta(n) := \frac{E\left\{|\hat{H}_n|^2\right\} E\left\{|s_n|^2\right\}}{E\left\{|\tilde{v}'_n|^2\right\}} = \frac{(\sigma_H^2 - \gamma_n)\sigma_s^2(n)}{\gamma_n \sigma_s^2(n) + \sigma_v^2}$$

Pilot-Based Channel Estimation for OFDM (13)

- Optimum data power distribution, no VC

❑ In this case, $N_d = N - N_p$ and with optimal design, $\gamma_n = \gamma^o, \forall n$. Hence, $\underline{C}$ maximized when $\sigma_s^2(n) = \sigma_s^2/N_d$:

❑ Maximum lower bound:

$$\underline{C} = \frac{N - N_p}{N+L} E\left\{\log\left(1 + \beta|g|^2\right)\right\}$$

where

$$\beta := \frac{(\sigma_H^2 - \gamma^o)\sigma_s^2}{\gamma^o \sigma_s^2 + (N - N_p)\sigma_v^2}$$

Pilot-Based Channel Estimation for OFDM (14)

- Optimal number of pilots: no VC

❏ Treating N_p as a continuous variable ν, it can be shown that

$$\frac{\partial \underline{C}}{\partial \nu} = \frac{1}{N+L} E\left\{ -\log\left(1+\beta|g|^2\right) + (N-\nu)\frac{\partial \beta}{\partial \nu}\frac{|g|^2}{1+\beta|g|^2} \right\} < 0$$

$\Rightarrow$ μ should be as small as possible, i.e.

$$N_p^o = L + 1$$

☞ $N_p = L+1$ also minimizes complexity at the receiver and maximizes bandwidth efficiency. However, $N_p = L+1$ might not be optimal in the case of channel modeling mismatch.

Pilot-Based Channel Estimation for OFDM (15)

- Optimum power allocation: no VC

❏ Let $\alpha = \sigma_s^2/\sigma_t^2$. Using $N_p = L+1$, $\mathcal{P}^o$, $\boldsymbol{\rho}^o$ we maximize $\underline{C}$

$$\alpha^o := \arg\max_{\alpha} \underline{C} = \arg\max_{\alpha} \beta$$

$\Rightarrow$ For the general case, solution can be found by polynomial rooting. Let β^o denote maximum value of β.

❏ Let $\xi = \sigma_H^2\sigma_t^2/(N-N_p)\sigma_v^2$, i.e. data SNR when $\sigma_s^2 = \sigma_t^2$.

❏ SNR losses due to channel estimation, estimation errors and both:

$$\frac{\xi}{\beta_{\text{ideal}}}(=\frac{1}{\alpha}), \qquad \frac{\beta_{\text{ideal}}}{\beta}, \qquad \frac{\xi}{\beta}$$

Pilot-Based Channel Estimation for OFDM (16)

- Optimum power allocation: no VC (cont.) High SNR regime:

❑ Approximations:

$$\sigma_H^2 - \gamma^o \approx \sigma_H^2, \quad \gamma^o \approx \frac{\sigma_v^2\,(L+1)}{\sigma_p^2}$$

$$\beta = (N - N_p)\xi \frac{\alpha(1-\alpha)}{(L+1)\alpha + (N-N_p)(1-\alpha)}$$

❑ Take $N_p = L+1$. For fixed pair (N, L), optimal value of α and β:

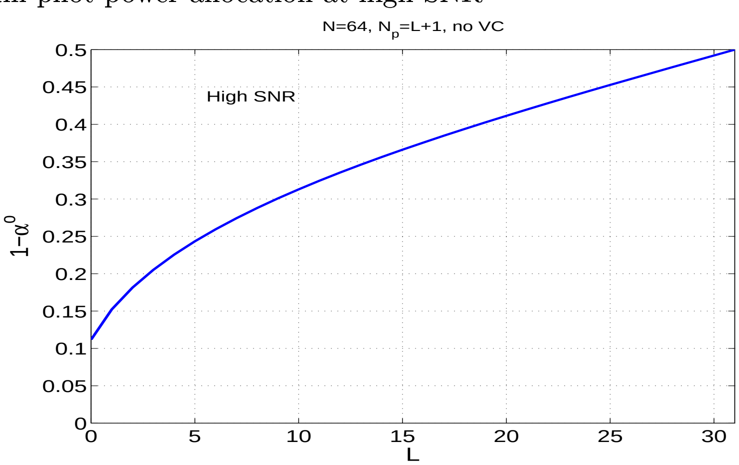
$$\alpha_\infty := \alpha^o|_{\text{high snr}} = \frac{1}{1+\sqrt{\frac{L+1}{N-L-1}}}; \quad \beta_\infty := \beta^o|_{\text{high snr}} = \xi\ \alpha_\infty^2$$

❑ For typical $N > 2(L+1)$, $\alpha \geq 0.5$ and maximum SNR losses (at high SNR) are resp 3dB, 3dB and 6dB. ☞ SNR loss decreases with N/L and $\to 0$ when $N >> L$.

Pilot-Based Channel Estimation for OFDM (17)

- Optimum power allocation: no VC (cont.)

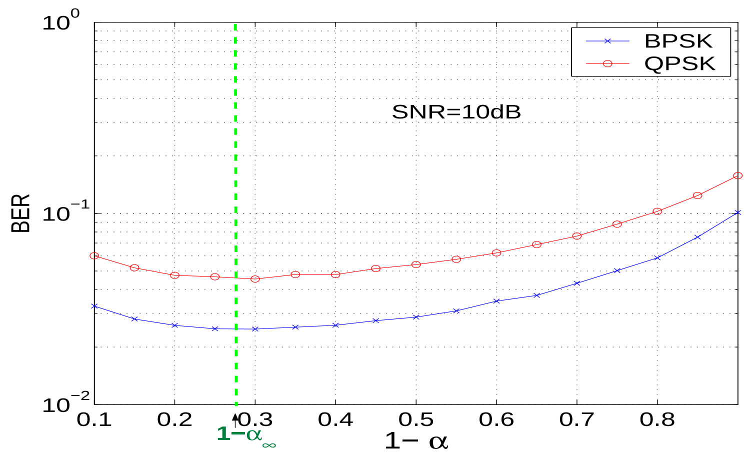
Optimum pilot power allocation at high SNR



Pilot-Based Channel Estimation for OFDM (18)

- Optimum power allocation: no VC (cont.)

BER performance: Rayleigh channel with exponential delay profile; $N = 64$ and $N_p = L + 1 = 8$.



Pilot-Based Channel Estimation for OFDM (19)

- Optimum power allocation: no VC (cont.)

❑ Rayleigh channels with equipowered taps, i.e. $\sigma_{h_\ell} = \sigma_h$:

$$\alpha_{\text{iid}} = \frac{1}{1+\sqrt{1-1/\xi}} \qquad \text{where} \quad \xi := \frac{N-N_p}{N-N_p-L-1}\left(1+\frac{L+1}{(N-N_p)\xi}\right)$$

➩ Max effective data SNR

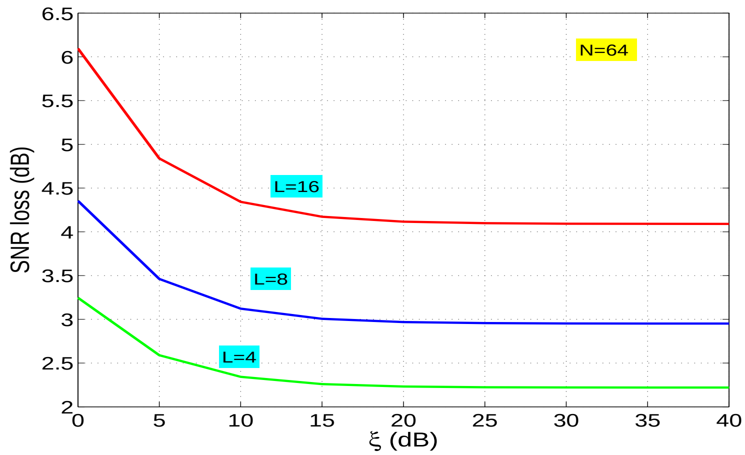
$$\beta_{\text{iid}} = \frac{\xi}{1+\frac{L+1}{(N-N_p)\xi}} \quad \alpha_{\text{iid}}^2$$

➩ Data SNR loss due estimation depends on both N/L and ξ.

Pilot-Based Channel Estimation for OFDM (20)

- Optimum power allocation: no VC (cont.)

SNR loss vs ξ, $N_p = L+1$.



Pilot-Based Channel Estimation for OFDM (21)

- Optimum pilot design for LS channel estimation: VC present

❏ Optimization wrt both $\mathcal{P}$ and $\boldsymbol{\rho}$ untractable in general.

❏ Complexity reduced if LS is used and $N_p = L+1$ (i.e. $\mathbf{W}_{\mathcal{P}}$ square).

❏ If total MSE, $\bar{\gamma}$, is used as criterion:

$$\{\boldsymbol{\rho}^o, \mathcal{P}^o\} = \arg\min_{\boldsymbol{\rho},\mathcal{P}} \bar{\gamma}_{\text{LS}} = \arg\min_{\boldsymbol{\rho},\mathcal{P}} \sum_{n=1}^{N_p} \frac{\psi_{n,n}}{\rho_n}$$

under $(\mathbf{C1})$ where $\boldsymbol{\Psi} := \mathbf{W}_{\mathcal{P}}^{-1^{\mathcal{H}}} \mathbf{W}_{\mathcal{D}}^{\mathcal{H}} \mathbf{W}_{\mathcal{D}} \mathbf{W}_{\mathcal{P}}^{-1}$.

❏ Minimizing wrt to $\boldsymbol{\rho}$ under $\sum \rho_n = \sigma_p^2$ gives

$$\rho_n^o = \sigma_p^2 \frac{\sqrt{\psi_{n,n}}}{\sum_{i=1}^{N_p} \sqrt{\psi_{i,i}}}, \ \forall n = 1, \cdots, N_p$$

Pilot-Based Channel Estimation for OFDM (22)

- Optimum pilot design: VC present (cont.)

❏ Optimization reduced to:

$$\mathcal{P}^o = \arg \min_{\mathcal{P} \subset \mathcal{A}} \left(\sum_{n=1}^{N_p} \sqrt{\psi_{n,n}} \right)^2$$

☞ Minimum total MSE of LS estimates:

$$\frac{\sigma_v^2}{\sigma_p^2} \left(\sum_{n=1}^{N_p} \sqrt{\psi_{n,n}} \right)^2$$

❏ Exhaustive search over all N_p-point subsets of $\mathcal{A}$.

Pilot-Based Channel Estimation for OFDM (23)

- Optimum pilot design: VC present (cont.)

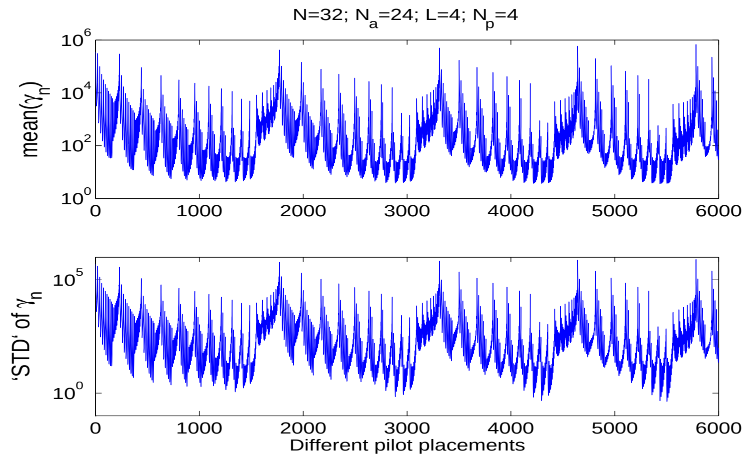
❏ Example: $N = 32,\ N_a = 24,\ N_p = L + 1 = 4$:

☞ Equispacing pilots in the active carrier region with one pilot placed near each edge of the VCs seems to be optimal.
☞ Pilot power ρ_n decreases when pilot close to VCs.

❏ Numerical examples show that setting $\boldsymbol{\rho}$ to be constant and optimizing wrt $\mathcal{P}$ lead to *almost* the same design

Pilot-Based Channel Estimation for OFDM (24)

- Optimum pilot design: VC present (cont.)



☞ $\mathcal{P}^o$ *almost* also minimizes 'STD' of γ_n. Perfect 'Fairness' in terms of estimation accuracy at different data carriers is impossible in general.

Pilot-Based Channel Estimation for OFDM (25)

- Optimum pilot design: VC present (cont.)

❑ The general problem is that of maximizing

$$\underline{C} = \frac{N}{N+L} \sum_{n \in \mathcal{D}} E \left\{ \log \left(1 + \frac{(\sigma_H^2 - \gamma_n)\sigma_s^2(n)}{\gamma_n \sigma_s^2(n) + \sigma_v^2} \right) |g|^2 \right\}$$

wrt $\mathcal{P}$, $\boldsymbol{\rho}$ σ_p^2 and the $\sigma_s^2(n)$'s for a constant σ_t^2; (orthogonality is valid only for MMSE estimator!)

❑ Maximization is untractable. A suboptimum solution is to use $\mathcal{P}$, $\boldsymbol{\rho}$ which minimize $\bar{\gamma}_{LS}$ and use the individual γ_n to maximize $\bar{C}$ wrt the $\sigma_s^2(n)$'s. ☞ Numerical examples show that no significant gain is obtained by accounting for the slight differences between the $gamma_n$'s.

Blind Channel Estimation for OFDM

- Two main classes of methods
  - methods exploiting the redundancy introduced by CP or/and virtual carriers: require large number of OFDM symbols.
  - methods exploiting the finite-alphabet (FA) property of the symbols: performance deteriorates with size of constellation.

  When the channel varies rapidly across the blocks, only the FA-based methods may be suitable.

Blind Channel Estimation for OFDM (2)

- FA-based blind channel estimation

- Assume $E\left\{s_n^M\right\} = \mu_M \neq 0$ and $E\left\{s_n^J\right\} = 0$ for $J < M$, e.g. $M = 2$ for BPSK and $M = 4$ for QPSK and QAM.
- Received ith block after CP removal and DFT (assume $N_a = N$):

$$\tilde{x}_n(i) = H_n s_n(i) + \tilde{w}_n(i), \qquad n = 0, \cdots, N-1$$

- Then

$$\tilde{y}_n(i) := [\tilde{x}_n(i)]^M = H_n^M s_n^M(i) + \xi_n(i)$$

  where $E\left\{\xi_n(i)\right\} = 0$. and

$$H_n^M = [1,\ e^{-j2\pi n/N},\ \cdots, e^{-j2\pi nM(L)/N}](\boldsymbol{h} *_M \boldsymbol{h}) =: \Omega(n,:)\boldsymbol{h}_M$$

- In vector form

$$[H_0^M, \cdots, H_{N-1}^M]^T =: \boldsymbol{H}_M = \boldsymbol{\Omega}\boldsymbol{h}_M$$

Blind Channel Estimation for OFDM (3)

- FA-based blind channel estimation (cont)

❑ Blind estimate of $\boldsymbol{H}_M$ and $\boldsymbol{h}_M$ using K blocks:

$$\begin{aligned}[\hat{\boldsymbol{H}}_M]_n &:= \widehat{H_n^M} = \frac{1}{\mu_M}\frac{1}{K}\sum_{i=1}^{K}\tilde{\boldsymbol{y}}(i) \\ \hat{\boldsymbol{h}}_M &= \boldsymbol{\Omega}^{\dagger}\hat{\boldsymbol{H}}_M = (1/N)\boldsymbol{\Omega}^{\mathcal{H}}\hat{\boldsymbol{H}}_M\end{aligned}$$

❑ Necessary condition: $N \geq ML+1$. For PSK, identifiability guaranteed even with one OFDM symbol.

❑ Blind estimate of $\boldsymbol{h}$:

$$\hat{\boldsymbol{h}} = \arg\min_{\boldsymbol{h}} \|\hat{\boldsymbol{h}}_M - \boldsymbol{h} *_M \boldsymbol{h}\|$$

Blind Channel Estimation for OFDM (4)

- FA-based blind channel estimation (cont)
- Minimum Distance Algorithm

❑ Estimate $\hat{H}_n$ using

$$\hat{H}_n = \lambda_n \left[\widehat{H_n^M}\right]^{1/M}$$

where $\lambda_n \in \{e^{j(2\pi/M)m}\}_{m=0}^{M-1}$ is the scalar ambiguity.

❑ Using exhaustive search over all M^N possible vectors $\boldsymbol{\lambda}$, and for each $\boldsymbol{\lambda}$, estimate time-domain vector $\hat{\boldsymbol{h}}$ and compute

$$\|\hat{\boldsymbol{h}}_M - \hat{\boldsymbol{h}} *_M \hat{\boldsymbol{h}}\|$$

❑ Final estimate of $\boldsymbol{h}$ is the minimizer of the above criterion.
☞ Reduced complexity because of discrete search. Other simpler algorithms exist.

Part 2: Channel Estimation for General CP Systems

Affine Precoding and MMSE Channel Estimation

- Assume
  - frequency-selective channel, constant over $K(\geq 1)$ blocks
- Received signal after CP removal

$$\begin{aligned} \boldsymbol{x}_i &= \mathbf{H}\boldsymbol{u}_i + \boldsymbol{v}_i \qquad i = 1 \cdots K \\ \boldsymbol{u}_i &= \boldsymbol{\Theta}_i \boldsymbol{s}_i + \boldsymbol{b}_i \end{aligned}$$

- $\boldsymbol{\Theta}_i$ $(N \times N)$ precoding matrix • $\boldsymbol{s}_i$: ith transmitted data block
- $\boldsymbol{b}_i$: ith pilot sequence • $\mathbf{H} = \mathrm{circ}([h_0 ... h_L 0 ... 0])$
- $\boldsymbol{v}_i$: AWGN, variance σ_v^2 • $\boldsymbol{s}_i$: independent of $\boldsymbol{v}_i$.

☞ Affine precoding includes TDM and superimposed training.

Affine Precoding and MMSE Channel Estimation (2)

Assume

- **(A1)** The non-zero elements of the $\boldsymbol{s}_i$'s are unknown, i.i.d zero-mean random variables drawn from a finite alphabet $\mathcal{M}$.
- Design criteria assume a fixed total pilot power in the frame

$$\sigma_b^2 = \frac{1}{K} \sum_{i=0}^{K-1} \sigma_b^2(i) \ ,$$

but the training power can vary from block to block.

Affine Precoding and MMSE Channel Estimation (3)

- Collecting K blocks:

$$\boldsymbol{x}_i = \mathbf{H}\boldsymbol{\Theta}_i \boldsymbol{s}_i + \mathbf{B}_i \boldsymbol{h} + \boldsymbol{v}_i, \quad i = 0, ..., K-1$$

  - $\mathbf{B}_i$: leading $(N \times L)$ of circ$(\boldsymbol{b}_i)$ • $\boldsymbol{h} = [h_0 \cdots h_L]^T$.

- MMSE channel estimate:

$$\hat{\boldsymbol{h}} = \frac{1}{\sigma_v^2} \left(\mathbf{R}_h^{-1} + \frac{1}{\sigma_v^2} \mathbf{B}^{\mathcal{H}} \mathbf{B} \right)^{-1} \mathbf{B}^{\mathcal{H}} \boldsymbol{x}.$$

  - $\boldsymbol{x} = [\boldsymbol{x}_1^T \cdots \boldsymbol{x}_K^T]^T$ • $\mathbf{B} = [\mathbf{B}_1^T \cdots \mathbf{B}_K^T]^T$

Affine Precoding and MMSE Channel Estimation(4)

- Identifiability condition:

$$\mathrm{rank}(\mathbf{B}) = L + 1 \qquad \text{(C2)}$$

- Frequency-domain counterpart:
  - let $\tilde{\boldsymbol{b}}_i$:= DFT of $\boldsymbol{b}_i$ and

$$\rho_n := \sum_{i=1}^{K} |\tilde{b}_i(n)|^2, \quad n = 0, ..., N-1$$

  - Let N_p: number of nonzero entries of $\boldsymbol{\rho} := [\rho_0 \cdots \rho_{N-1}]$

  ⇨ $\mathrm{rank}(\mathbf{B}) = \min(N_p, L+1)$

$$(C2) \qquad \Longleftrightarrow \qquad N_p \geq L+1$$

  i.e. combined training power across the blocks is non-zero at at least $L+1$ frequencies.

Affine Precoding and MMSE Channel Estimation(5)

- Orthogonal precoding

- Condition for decoupled channel estimation and data detection:

$$\begin{aligned} \tilde{b}_i^*(n)\ [\mathbf{F}^{\mathcal{H}} \boldsymbol{\Theta}_i \boldsymbol{s}_i]_n &= 0, \qquad \forall n, i \\ &\Updownarrow \\ \mathbf{T}_i \mathbf{F}^{\mathcal{H}} \boldsymbol{\Theta}_i \boldsymbol{s}_i &= 0, \qquad \forall i \qquad \text{(C3)} \end{aligned}$$

where $\mathbf{T}_i = \mathrm{diag}\{t_i(n), n = 0, \cdots, N-1\}$ with

$$t_i(n) = \begin{cases} 1 & \text{if } n \in \mathcal{P}_i \\ 0 & \text{otherwise} \end{cases}$$

Affine Precoding and MMSE Channel Estimation(6)

- Optimal training for orthogonal precoding

Result 1 Assume that $Q = N/N_p$ is an integer. Under (C3) and the constraint of fixed training power σ_b^2, the MSE of $\hat{\boldsymbol{h}}$ in orthogonal precoders is minimized when

$$\rho_n = \begin{cases} \frac{\sigma_b^2 N}{N_p} \sum_{\ell=0}^{N_p-1} \delta(n - \ell Q - m) & \text{if } Q := \frac{N}{N_p} \text{ integer} \\ \\ \frac{\sigma_b^2 N}{N_p} \sum_{\ell=0}^{N_p-1} [1 - \delta(n - \ell Q - m)] & \text{if } Q := \frac{N}{N - N_p} \text{ integer} \end{cases} \quad \text{(C4)}$$

- m : arbitrary integer from $[0, ..., Q-1]$

Affine Precoding and MMSE Channel Estimation(7)

- Result 1 implies that the pilot frequencies should be equispaced and that their *average* powers across the K blocks should be identical. Therefore, channel estimation performance is the same regardless of the distribution of the training power across the blocks.
- the minimum MSE of $\hat{\boldsymbol{h}}$ is independent of N_p, the number of pilot frequencies.
- Time-division multiplexing (TDM) is not an orthogonal precoding scheme. Condition (C3) implies that training should be superimposed onto the data in the time domain (but orthogonal in the frequency domain).
- The $K > 1$ scenario gives more flexibility for designing precoders. It is also useful if frequency hopping is desired.

Full-Rank Orthogonal Precoding

- Let $\mathcal{P}_i$: set of pilot frequencies during ith block
- **Result 2** Assume that $\mathbf{\Theta}_i$, $i = 1, ..., K-1$, are full rank, assumption (A1) holds and maximum possible data-rate is required. Then, the orthogonality condition (C3) is satisfied if and only if the nth entry of $\mathbf{\Lambda}_i \boldsymbol{s}_i$ is identically zero for $n \in \mathcal{P}_i$, where $\mathbf{\Lambda}_i$ is any permutation matrix, and the precoding matrix has the following form

$$\mathbf{\Theta}_i = \mathbf{F}^{\mathcal{H}} \left[\mathbf{T}_i \mathbf{W}_i \mathbf{T}_i + (\mathbf{I} - \mathbf{T}_i) \mathbf{A_i} \right] \mathbf{\Lambda}_i$$

where $\mathbf{W}_i$ and $\mathbf{A_i}$ are any $(N \times N)$ matrices such that $(\mathbf{T}_i \mathbf{W}_i \mathbf{T}_i + (\mathbf{I} - \mathbf{T}_i) \mathbf{A_i})$ is full-rank.

Full-Rank Orthogonal Precoding (2)

- $\mathbf{W}_i = \mathbf{A_i} = \mathbf{I} \rightarrow \mathbf{\Theta_i} = \mathbf{F}^{\mathcal{H}} \equiv$ OFDM with reserved pilot tones.
- Uncoded OFDM has poor performance because only diversity order one is possible through Rayleigh fading channels. This problem is overcome by employing either Galois field channel coding or LP-OFDM - LCP-OFDM.
- Here, we focus on SC-CP systems. Although such systems do not have full multipath diversity, their performance at realistic SNR values approaches that of maximum diversity systems. Further, maximum diversity at high SNR can be achieved if the constellations are first rotated prior to SC-CP modulation.
- Conventional SC-CP where $\mathbf{\Theta}_i = \mathbf{I}$ is not an orthogonal precoding scheme.

Full-Rank Orthogonal Precoding (3)

- Full-rank orthogonal single carrier (FROSC) precoding

❑ Let $\mathbf{T}_{\mathcal{D}_i}$ and $\mathbf{T}_{\mathcal{P}_i}$ be the data and pilot selection matrices, and $\bar{\mathbf{A}}_i$= non-zero $((N - N_{p_i}) \times N)$ submatrix of $(\mathbf{I} - \mathbf{T}_i)\mathbf{A_i}$

❑ FROSC is obtained by choosing $\mathbf{\Theta}$ to be the same as $\mathbf{I}$ except for the N_{p_i} pilot rows. This is achieved by

$$\mathbf{W}_i = \mathbf{I}, \quad \text{and} \quad \bar{\mathbf{A}} = (\mathbf{T}^T_{\mathcal{D}_i}\mathbf{F}^{\mathcal{H}}\mathbf{T}_{\mathcal{D}_i})^{-1}\mathbf{T}^{\mathcal{H}}_{\mathcal{D}_i}(\mathbf{I} - \mathbf{F}^{\mathcal{H}}\mathbf{T}_{\mathcal{P}_i}\mathbf{T}^T_{\mathcal{P}_i})$$

❑ Bandwidth efficiency of FROSC:

$$\zeta_{FROSC}(i) = \frac{N - N_{p_i}}{N + L}$$

Full-Rank Orthogonal Precoding (4)

- FROSC precoding (cont.)

❑ The Θ_i's are the same as $\mathbf{I}$ except for P_i rows are obtained using $\mathbf{A_i}$. An example of the structure of $\mathbf{\Theta}_i$ when $N = 8$ and $\mathcal{P}_i = \{0, 4\}$ is

$$\mathbf{\Theta}_i = \begin{pmatrix} \times & \times & \times & \times & \times & \times & \times & \times \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \times & \times & \times & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} ; \quad \boldsymbol{s}_i = \begin{pmatrix} 0 \\ \times \\ \times \\ \times \\ 0 \\ \times \\ \times \\ \times \end{pmatrix}$$

Full-Rank Orthogonal Precoding (5)

- FROSC precoding (cont.)

❑ Effectively, the precoding is redundant (or tall):

$$\boldsymbol{\Theta}_i \boldsymbol{s}_i = \bar{\boldsymbol{\Theta}}_i \bar{\boldsymbol{s}}_i \quad \text{with } \bar{\boldsymbol{\Theta}}_i := \boldsymbol{\Theta}_i \mathbf{T}_{\mathcal{D}_i}^T \text{ and } \bar{\boldsymbol{s}}_i = \mathbf{T}_{\mathcal{D}_i} \boldsymbol{s}_i$$

Previous example:

$$\bar{\boldsymbol{\Theta}}_i = \begin{pmatrix} \times & \times & \times & \times & \times & \times \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \times & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} ; \quad \bar{\boldsymbol{s}}_i = \begin{pmatrix} \times \\ \times \\ \times \\ \times \\ \times \\ \times \end{pmatrix}$$

Full-Rank Orthogonal Precoding (6)

- FROSC precoding: symbol detection

❑ Linear equalization: $\mathbf{H}$ is circulant $\Rightarrow$ equalization in the F-D

$$\widehat{\bar{\boldsymbol{s}}}_i = \lfloor \bar{\boldsymbol{\Theta}}_i^{\dagger} \mathbf{F}^{\mathcal{H}} (\mathbf{I} - \mathbf{T}_i) \mathbf{G} \mathbf{F} \boldsymbol{x}_i) \rfloor_{\mathcal{M}}$$

where $\mathbf{G} = \text{diag}\{g(k),\ k = 0, \cdots, N-1\}$ is the MMSE equalizer:

$$g(k) = \hat{H}_n^* / (|\hat{H}_n|^2 + \sigma_v^2)$$

❑ Ignoring the $n \in \mathcal{P}_i$ rows of $\bar{\boldsymbol{\Theta}}_i$, a simpler detection scheme is

$$\widehat{\bar{\boldsymbol{s}}}_i = \lfloor \mathbf{T}_{\mathcal{D}_i} \mathbf{F}^{\mathcal{H}} (\mathbf{I} - \mathbf{T}_i) \mathbf{G} \mathbf{F} \boldsymbol{x}_i) \rfloor_{\mathcal{M}}$$

Rank-Deficient Orthogonal Precoding

- Rank-deficient orthogonal single carrier (DROSC) precoding

❑ Full data-rate under (C3) requires ($\text{rank}(\boldsymbol{\Theta}_i) = N - P_i$)

$$[\mathbf{F}\boldsymbol{\Theta}_i]_n = 0, \quad n \in \mathcal{P}_i$$

❑ $\boldsymbol{s}_i$ cannot be recovered linearly. However, using the finite-alphabet property detection is still possible.

❑ DROSC is obtained by designing $\boldsymbol{\Theta}_i$ as

$$\begin{aligned}\boldsymbol{\Theta}_i^o &= \min_{\boldsymbol{\Theta}_i;\ \mathbf{F}_{\mathcal{P}_i}\boldsymbol{\Theta}_i=\mathbf{0}} \sum_{i=0}^{K-1} \|\boldsymbol{\Theta}_i - \mathbf{I}\|_2 \\ &\Downarrow \\ \boldsymbol{\Theta}_i^o &= \mathbf{F}^{\mathcal{H}}(\mathbf{I} - \mathbf{T}_i)\mathbf{F}\end{aligned}$$

Rank-Deficient Orthogonal Precoding (2)

❑ **Result 3** Assume $N/(L+1) = Q$ and $M = (L+1)/K$ are integers. A bandwidth efficient orthogonal precoding scheme is obtained as follows

- for $i = 0, ..., K-1$ chose $\mathcal{P}_i = \{nKQ + iQ, n = 0, ..., M-1\}$
- set $\boldsymbol{\Theta}_i = \mathbf{F}^{\mathcal{H}}(\mathbf{I} - \mathbf{T}_i)\mathbf{F}$
- add a training sequence according to condition (C3).

❑ Bandwidth efficiency of DROSC:

$$\zeta_{DROSC} = \frac{N}{N+L}$$

Rank-Deficient Orthogonal Precoding (3)

- Symbol detection

❑ Received signal $\boldsymbol{x}_i = \mathbf{H}\ [(\mathbf{I}-\mathbf{J})\boldsymbol{s}_i + \boldsymbol{b}_i] + \boldsymbol{v}_i$ with $\mathbf{J} = \mathbf{F}^{\mathcal{H}}\mathbf{T}_i\mathbf{F}$

❑ Remove training related term

$$
\begin{aligned}
\boldsymbol{z}_i &:= (\mathbf{I}-\mathbf{J})\ \boldsymbol{x}_i \\
&= (\mathbf{I}-\mathbf{J})\ \mathbf{H}\boldsymbol{x}_i + (\mathbf{I}-\mathbf{J})\ \boldsymbol{v}_i, \\
&= \mathbf{H}\ (\mathbf{I}-\mathbf{J})\boldsymbol{x}_i + \tilde{\boldsymbol{v}}_i \\
&= \mathbf{H}\ (\mathbf{I}-\mathbf{J})[(\mathbf{I}-\mathbf{J})\boldsymbol{s}_i + \boldsymbol{b}_i] + \tilde{\boldsymbol{v}}_i \\
&= \mathbf{H}\ (\mathbf{I}-\mathbf{J})\boldsymbol{s}_i + \tilde{\boldsymbol{v}}_i \qquad \text{since } (\mathbf{I}-\mathbf{J})^2 = \mathbf{I}-\mathbf{J}
\end{aligned}
$$

❑ MMSE equalizer: $\mathbf{G} = \mathrm{diag}\{|[\hat{H}_n|^2 + \tilde{\sigma}^2]^{-1}\hat{H}_n, n = 0, \cdots N-1\}$

$$\boldsymbol{u}_i = \mathbf{F}^{\mathcal{H}}\mathbf{G}\mathbf{F}\boldsymbol{z}_i$$

Rank-Deficient Orthogonal Precoding (4)

- Symbol detection, cont.

❑ Even if channel estimation is perfect and no noise, $\boldsymbol{u}_i \neq \boldsymbol{s}_i$:

$$\boldsymbol{u}_i = (\mathbf{I}-\mathbf{J})\boldsymbol{s}_i + \boldsymbol{\epsilon}_i \qquad (\boldsymbol{\epsilon}_{\mathrm{i}} : \text{ due to noise \& estimation errors})$$

❑ $\mathbf{I}-\mathbf{J}$: rank-deficient $\Rightarrow$ $\boldsymbol{s}_i$ cannot be recovered linearly

❑ Using finite alphabet property:

➩ Symbol vector detection $\leftarrow$ prohibitive

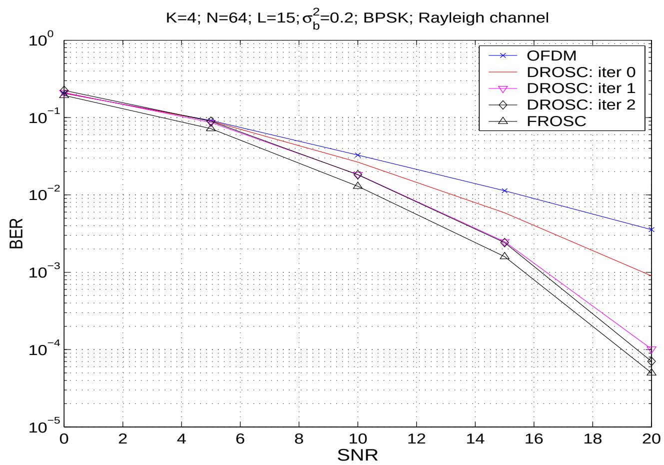
➩ Iterative symbol-by-symbol detection: (1-2 iterations suffice)

$$
\begin{aligned}
\hat{\boldsymbol{s}}_i^{(0)} &= \lfloor \boldsymbol{u}_i \rfloor \\
\hat{\boldsymbol{s}}_i^{(m)} &= \lfloor \boldsymbol{u}_i + \mathbf{J}\hat{\boldsymbol{s}}_i^{(m-1)} \rfloor
\end{aligned}
$$

Rank-Deficient Orthogonal Precoding (5)

- Simulation Results

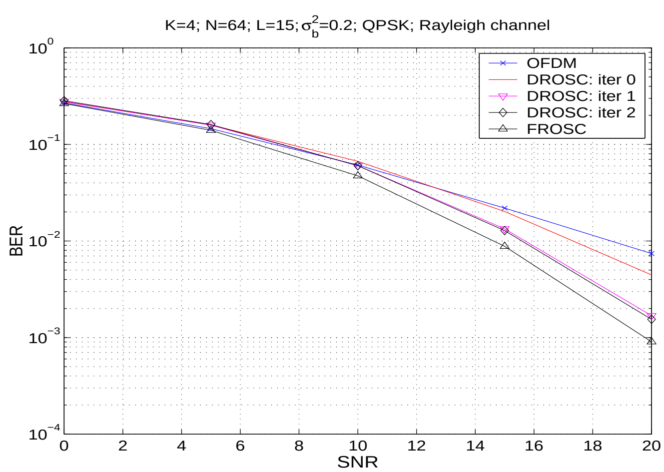
❑ BER vs SNR; $K = 4,\ N = 64,\ L = 15,\ \sigma_b^2 = 0.2$, BPSK.



Rank-Deficient Orthogonal Precoding (6)

- Simulation Results, cont.

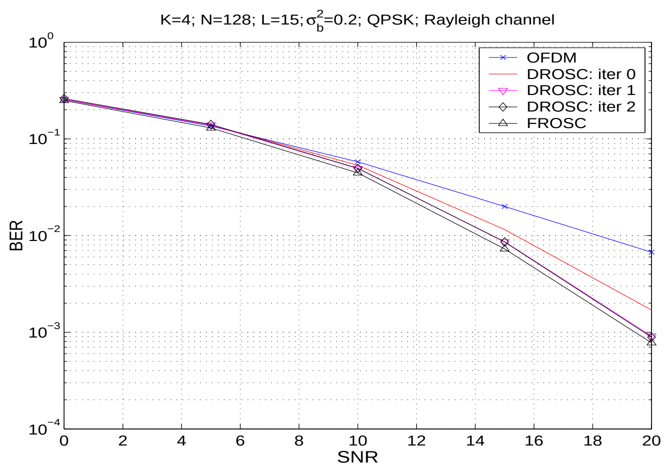
❑ BER vs SNR; $K = 4,\ N = 64,\ L = 15,\ \sigma_b^2 = 0.2$, QPSK.



Rank-Deficient Orthogonal Precoding (7)

- Simulation Results, cont.

❑ BER vs SNR; $K = 4, \; N = 128, \; L = 15, \; \sigma_b^2 = 0.2$, BPSK.



Summary

√ Pilot carrier design dramatically affects system performance

√ Blind techniques for OFDM may be more promising than for serial single-carrier systems

√ Affine precoding gives a general framework for block transmission schemes

√ OFDM or single-carrier CP systems? the saga continues...

References

[1] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electron.*, vol. 44, Aug. 1998.

[2] R. Heath and G.B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Processing*, vol. 47, March 1999.

[3] M. Medard, "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel," *IEEE Trans. Information Theory*, vol. 46, May 2000.

[4] S. Zhou and G. B. Giannakis, "Finite-alphabet based channel estimation for OFDM and related multicarrier systems," *IEEE Trans. Communications*, vol. 49, Aug. 2001.

[5] C. Li and S. Roy, "Subspace based blind channel estimation for OFDM by exploiting virtual carrier," *Proc. Globecom*, vol 1, San Antonio, TX, 2001.

[6] M. Morelli and Umberto Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Processing*, vol. 49, Dec. 2001.

[7] B. Muquet *et al*, "A subspace based blind and semi-blind channel identification method for OFDM systems," *IEEE Trans. Signal Processing*, vol. 50, July 2002.

[8] S. Adireddy, L. Tong and H. Viswanathan, "Optimal placement of training for frequency-slective block-fading channels," *IEEE Trans. Information Theory*, vol. 48, Aug 2002.

[9] How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Information Theory*, vol. 49, April 2003.

[10] S. Ohno, and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wirelesss OFDM," *IEEE Trans. Communications*, vol. 50, Dec 2002.

[11] Z. Liu *et al*, "Linear constellation precoding for OFDM with multipath diversity and coding gains," *IEEE Trans. Communications*, vol. 51, March 2003.

[12] Z. Wang and G. B. Giannakis, "Complex-field coding for OFDM over fading wireless channels," *IEEE Trans. Information Theory*, vol. 49, Mar. 2003.

[13] S. Ohno, and G. B. Giannakis, "Capacity optimizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Information Theory*, vol. 50, Sep 2004.

[14] G.T. Zhou *et al*, "First-order statistical method for channel estimation," *IEEE Signal Processing Lett.*, 10, Mar 2003.

[15] A. G. Orozco-Lugo *et al*, "Channel estimation using implicit training," *IEEE Trans. Signal Processing*, vol. 52, Jan 2004.

[16] M. Ghogho, *et al*, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training", *IEEE Signal Processing Lett.*, 12, Mar 2005.

[17] M. Ghogho *et al*, "SISO and MIMO Channel Estimation and Symbol Detection using Data-Dependent Superimposed Training," *IEEE ICASSP'05*, Philadelphia, March 2005.

[18] M. Ghogho and A. Swami, "Optimal Training for Affine-Precoded and Cyclic-Prefixed Block Transmissions," *Proc SSP'05*, Bordeaux, France, July 2005.

[19] M. Ghogho and A. Swami, "Training Design for Multipath Channel and Frequency-Offset Estimation in MIMO Systems," to be published in *IEEE Trans. Signal Processing*, vol. .

Estimation Theory for Wirelss Communication, 24-28 Oct 2005, PARIS.

Carrier Frequency-Offset for OFDM and Related Multicarrier Systems

M. Ghogho University of Leeds, UK

Aims and General Outline

Aims:

- To present data-aided and (semi-)blind CFO estimation algorithms for OFDM
- To give a unified framework for several existing algorithms

General outline

- Motivation and context
- Null-subcarrier-based CFO estimation
- Blind CFO estimation exploiting data properties

Motivation and Context

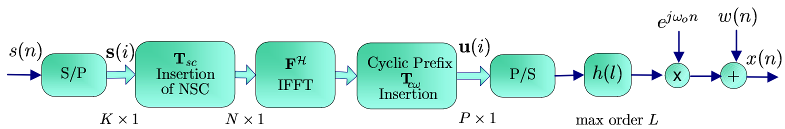
- High data rates (up to 54 Mbps) with Coded-OFDM
  - ⋆ IEEE802.11a, HIPERLAN/2, MMAC; DAB, DVB
- OFDM turns frequency-selective to flat fading channels
  - ⋆ Timing-Offset (TO) as a pure-delay channel
- Low-complexity equalization and easy decoding
  - ⋆ convolutional coded OFDM (across subcarriers)
- Challenges
  - Non-constant modulus $\Rightarrow$ large peak-to-average power ratio
  - Sensitivity to Carrier Frequency-Offset (CFO)
    - Inter-Carrier Interference (ICI)
    - At $E_s/N_0 = 19$dB: CFO/subcarrier spacing $= 1.26\%$
      $\Longrightarrow$ SNR degradation 10 dB

Part 1: Null-Subcarrier-based CFO Estimation

Outline

- Signal model
- Deterministic ML estimator
- Identifiability issues
- CRB and optimal placement of null subcarriers
- Performance analysis
- Repetitive Slot-Based CFO Estimation
- Comparisons
- Summary

Signal Model



- NSC insertion: $\mathbf{T}_{sc}$: K cols of a $N \times N$ permutation matrix
- CP insertion: $\mathbf{T}_{cp} = \begin{bmatrix} \mathbf{0}_{L\times(N-L)}, & \mathbf{I}_L \\ \multicolumn{2}{c}{\mathbf{I}_N} \end{bmatrix}$
- Transmitted block: $\boldsymbol{u}_{\text{cp}}(i) = \mathbf{T}_{cp}\mathbf{F}_N^{\mathcal{H}}\mathbf{T}_{sc}\boldsymbol{s}(i)$
- Input-output relationship $(N \geq K, P = L + N)$

$$x_{\text{cp}}(n) = e^{j\omega_o n} \sum_{l=0}^{L} h(l) u_{\text{cp}}(n-l) + w_{\text{cp}}(n)$$

Goal: Estimate CFO ω_o based only on knowledge of $\mathbf{T}_{sc}$ without channel state information

Signal Model (2)

- Received blocks

$$\boldsymbol{x}_{\text{cp}}(i) = e^{j\omega_o iP}\mathbf{D}_P(\omega_o)\left[\mathbf{H}_1\boldsymbol{u}(i) + \mathbf{H}_2\boldsymbol{u}(i-1)\right] + \boldsymbol{w}(i)$$

where $\mathbf{D}_P(\omega_o) = \text{diag}(e^{jk\omega_o}, k = 0, ..., P-1)$

- Discard CP to avoid IBI: using $\mathbf{R}_{cp} := [\mathbf{0}_{N\times(P-N)}, \mathbf{I}_N]$:

$$\mathbf{R}_{cp}\mathbf{H}_2 = \mathbf{0}, \quad \mathbf{R}_{cp}\mathbf{D}_P(\omega_o) = \mathbf{D}_N(\omega_o)\mathbf{R}_{cp}, \qquad \mathbf{R}_{cp}\mathbf{D}(\omega_o)\mathbf{H}_2 = \mathbf{0}$$

- Channel matrix: $\mathbf{H}_1$ Toeplitz $\Rightarrow$ $\mathbf{H}_c = \mathbf{R}_{cp}\mathbf{H}_1\mathbf{T}_{cp}$ circulant; so

$$\mathbf{F}_N\mathbf{H}_c\mathbf{F}_N^{\mathcal{H}} == \text{diag}(H_0 \cdots H_{N-1}) =: \mathbf{D}_H$$

where $H_k = \sum_{\ell=0}^{L} h_\ell \exp(-j2\pi\ell k/N)$

Signal Model (3)

- Received blocks after CP removal

$$\boldsymbol{x}(i) = \mathbf{R}_{cp}\boldsymbol{x}_{\text{cp}}(i) = e^{j\omega_o iP}\mathbf{D}_N(\omega_o)\mathbf{F}_N^{\mathcal{H}}\mathbf{D}_H\mathbf{T}_{sc}\boldsymbol{s}(i) + \boldsymbol{w}(i)$$

- Perform FFT:

$$\begin{aligned}\tilde{\boldsymbol{x}}(i) &= \mathbf{F}_N\boldsymbol{x}(i)\\ &= e^{j\omega_o iP}\underbrace{[\mathbf{F}_N\mathbf{D}_N(\omega_o)\mathbf{F}_N^{\mathcal{H}}]}_{\text{diagonal?}}\mathbf{D}_H\mathbf{T}_{sc}\boldsymbol{s}(i) + \tilde{\boldsymbol{w}}(i)\\ &= \mathbf{D}_H\mathbf{T}_{sc}\boldsymbol{s}(i) + \tilde{\boldsymbol{w}}(i) \qquad \text{iff } \omega_o = 0\end{aligned}$$

- $\hookrightarrow$ CFO causes ICI; degrades BER

Signal Model (4)

- After discarding CP, but before FFT (dropping block index)

$$x(k) = \sum_{n\in\mathcal{A}} H_n s_n e^{j2\pi k(n+\nu_o)/N} + w(k) \quad k = 0, ..., N-1$$

• $\nu_o = N\frac{\omega_o}{2\pi}$ is unknown CFO ; $-N/2 < \nu_o \leq N/2$ s_n unknown data symbols
• $\mathcal{A} \subset \mathcal{N} = \{-N/2+1, ..., N/2\}$: active sub-carriers
$\mathcal{Z} = \mathcal{N} - \mathcal{A}$: set of NSC's

-

$$\begin{aligned}a(k) &= \sum_{n\in\mathcal{A}} H_n s_n e^{j2\pi kn/N}\\ x(k) &= a(k)\exp(j2\pi k\xi_o/N) + w(k)\end{aligned}$$

• Estimate CFO in additive + multiplicative noise

Deterministic ML Estimator

- Treat $\alpha_n := H_n s_n$ as non-random unknowns
- Receiver knows NSC set

$$\boldsymbol{x} = \mathbf{D}(\nu_o)\ \Phi_{\mathcal{A}}\ \boldsymbol{\alpha}\ +\ \boldsymbol{w}$$

$$\begin{aligned}
\mathbf{D}(\nu_o) &= \mathrm{diag}\{1, e^{j2\pi\nu_o/N}, ..., e^{j2\pi(N-1)\nu_o/N}\} \\
\Phi_{\mathcal{A}} &= \mathbf{F}_N^{\mathcal{H}} \mathbf{T}_{sc} \\
\boldsymbol{\alpha} &= [\alpha_{n_1}\ \ \dots\ \ \alpha_{n_{N_a}}]^T; \qquad n_\ell \in \mathcal{A},\ \ell = 1, ..., N_a
\end{aligned}$$

Deterministic ML Estimator (2)

- Gaussian Problem. Concentrate LLF wrt α_n's:

$$\begin{aligned}
\hat{\nu}_o &= \arg\max_{\nu} \sum_{\tau} r(\tau) \psi_{\mathcal{A}}^*(\tau) e^{-j2\pi\tau\nu/N} \\
r(\tau) &= \sum_{k=0}^{N-1-\tau} y^*(k) y(k+\tau) = r^*(-\tau) \\
\psi_{\mathcal{A}}(\tau) &= \frac{1}{N_a} \sum_{n \in \mathcal{A}} e^{j2\pi n\tau/N}
\end{aligned}$$

- Peak-pick windowed correlogram; window dictated by $\mathcal{A}$.
- $N_a = N \Rightarrow \psi_{\mathcal{A}}(\tau) = \delta(\tau) \Rightarrow$ CFO is *not* identifiable
  $\hookrightarrow$ Need NSC's

Deterministic ML Estimator (3)

- Interpretation of DML

❏ MLE maximizes $J_A(\nu)$ or minimizes $J_z(\nu)$

$$\hat{\nu}_o = \arg\max J_a(\nu) = \arg\min J_z(\nu)$$

where

$$J_a(\nu) = \sum_{n\in\mathcal{A}} |X(\nu+n)|^2 \qquad J_z(\nu) = \sum_{n\in\mathcal{Z}} |X(\nu+n)|^2$$

with $X(f)$=DTFT of $\boldsymbol{x}$

➩ Peak-pick (null-pick) sum of shifted periodograms

➩ $\hat{\nu}$: frequency shift that minimizes total energy at NSC's

Identifiability Issues

- Identifiability study assumes noiseless case

❏ Identifiability is guaranteed iff

$$|\mathbf{D}(\nu_o)\mathbf{\Phi}_{\mathcal{A}}\alpha - \mathbf{D}(\nu)\mathbf{\Phi}_{\mathcal{A}}\alpha\|_2 \neq 0 \;\; \forall\, \nu \neq \nu_o$$

❏ Equivalently $J(\nu) < J(\nu_o)$ where

$$J(\nu) = \boldsymbol{\alpha}^{\mathcal{H}}\, \mathbf{G}_{\mathcal{A}}(\nu-\nu_o)\; \boldsymbol{\alpha}$$

with

$$\mathbf{G}_{\mathcal{A}}(\epsilon) = \mathbf{T}_{sc}^{\mathcal{H}}\mathbf{F}\mathbf{D}^{\mathcal{H}}(\epsilon)\mathbf{F}^{\mathcal{H}}\mathbf{T}_{sc}$$

➩ $J(\nu_o) = |\boldsymbol{\alpha}|^2$.

➩ Channel zeros $\alpha_n = 0$: it suffices to have $N_a \geq L+1$

Identifiability Issues (2)

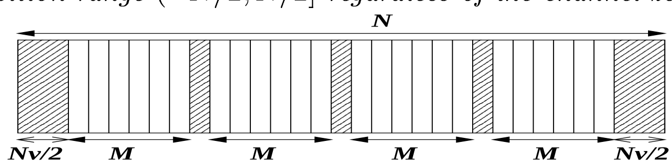
- Ambiguity due to number and location of NSC's
  - Global maxima of $J(\nu)$ at $\nu = \nu_o + m$; unique global at m=0?
  - For $\nu = \nu_o + m$, $\mathbf{G}_{\mathcal{A}}$ is diagonal of ones and zeros
  - $J(m + \nu_o) = \sum_{n_\ell \in \mathcal{A}} |\alpha_{n_\ell} g_{n_\ell}(m)|^2$
  - If for some $m \neq 0$, $g_{n_i}(m) \neq 0$ whenever $\alpha_{n_i} \neq 0$: $\hookrightarrow$ Identifiability lost
  - Identifiability is restored in $(-M/2, M/2]$ by choosing $\mathcal{A}$ st. $\forall m \in [1, M/2]$, $g_{n_i}(m) = 0$ for at least $L+1$ values of i, $n_i \in \mathcal{A}$. (because channel has a maximum of L zeros)

Identifiability Issues (3)

- Let $P(m) := \{n_p : n_p \neq n_k + m, n_p, n_k \in \mathcal{A}\}$. Need $P(m) \geq L + 1$, for $0 < |m| \leq M/2$
- For consecutive NSC, $P(m) = \min(m, N_z, N_a)$. With $m = 1 \rightarrow L = 0 \rightarrow$ VSC-based estimator is viable only for AWGN channel.
- If $M \geq 2$, need $\min(N_a, N_z) > L$.
- For equi-spaced NSC's, CFO is uniquely identifiable in $(-N/2N_z, N/2N_z)$, if $L < N_z < N - L$.
- For equi-spaced active sub-carriers, CFO is uniquely identifiable in $(-N/2N_a, N/2N_a)$, if $L < N_a < N - L$.
- For NSC with distinct spacing, CFO is uniquely identifiable in $[-N/2, N/2)$ iff $L + 1 < N_z < N - L$.

Identifiability Issues (4)

- *If the number of consecutive NSC $N_v > L$, the number of equispaced NSC $N_n > L$ and the spacing between the equispaced NSC is $M > L$, then the CFO is uniquely identifiable in the entire acquisition range $(-N/2, N/2]$ regardless of the channel zeros.*



- Tradeoffs between acquisition range, performance, maximum tolerable delay spread.
- Identifiability conditions are relaxed if multiple blocks used and null-subcarrier hopping is performed.

CRB and Optimal Placement of Null Subcarriers

- Conditional CRB (CCRB)

- CCRB treats $\alpha_n = H_n s_n$ as non-random unknowns

$$CCRB_{\mathcal{A}}(\nu_o) = \frac{\sigma^2}{8\pi^2 N}\left[\boldsymbol{\alpha}^{\mathcal{H}}\Phi_{\mathcal{A}}^{\mathcal{H}}\mathbf{Q}\left(\mathbf{I} - \frac{N_a}{N}\Psi_{\mathcal{A}}\right)\mathbf{Q}\Phi_{\mathcal{A}}\boldsymbol{\alpha}\right]^{-1}$$

$$\mathbf{Q} = N^{-3/2}\mathrm{diag}\{0, ..., N-1\}$$
$$\Phi_{\mathcal{A}} = \mathbf{F}^{\mathcal{H}}\mathbf{T}_{sc}, \quad \Psi_{\mathcal{A}} = \Phi_{\mathcal{A}}\Phi_{\mathcal{A}}^{\mathcal{H}}$$

- If no NSC i.e. $N_a = N \longrightarrow CCRB(\nu_o) = \infty$.

- CCRB is channel-dependent.

CRB and Optimal Placement of Null Subcarriers (2)

- Modified CRB (MCRB)

❑ Rayleigh fading $\mathbf{R}_h = E\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^{\mathcal{H}}\}$.

❑ $\alpha_n := H_n s_n$; $\mathbf{S} = \text{diag}\{s_n, n \in \mathcal{A}\}$; $\quad \mathbf{R}_\alpha = \mathbf{S}\mathbf{R}_h\mathbf{S}^H$

❑ Channel-independent CRB:

$$MCRB_{\mathcal{A}}(\nu_o) = \frac{1/(8\pi^2 N)}{\text{Tr}\left\{\mathbf{R}^{-1}\mathbf{QRQ} - \mathbf{Q}^2\right\}}$$

where

$$\mathbf{R} \quad = \quad \Phi_{\mathcal{A}}\mathbf{R}_\alpha\Phi_{\mathcal{A}}^{\mathcal{H}} + \sigma^2\mathbf{I}$$

❑ Blind case: reasonable to assume $\mathbf{R}_\alpha$ diagonal

CRB and Optimal Placement of Null Subcarriers (3)

❑ $\rightarrow$ MCRB is a function of $\mathcal{A}$: # and placement of NSC's:

$$MCRB_{\mathcal{A}}(\nu_o) = \frac{1/(8\pi^2 N\eta)}{\frac{N}{N_a}\text{Tr}\left\{\mathbf{Q}^2\right\} - \text{Tr}\left\{\Psi_{\mathcal{A}}\mathbf{Q}\Psi_{\mathcal{A}}\mathbf{Q}\right\}}$$

- $\eta = N_a\gamma^2/(N_a + N\gamma)$ is channel-independent
- $\gamma = E|H_n|^2/\sigma^2$ is the average SNR

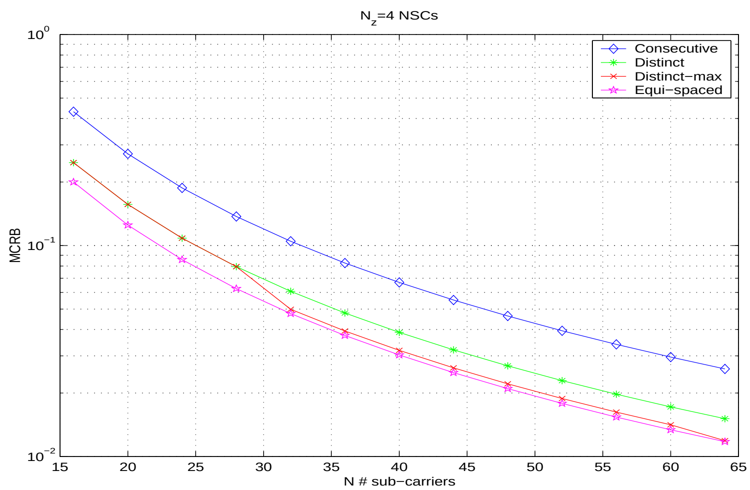
❑ *The optimal (in the sense of minimum MCRB) placement of a fixed number of active sub-carriers, N_a, is given by*

$$\mathcal{A}^* = \arg\min_{\mathcal{A}} \sum_{k,\ell=0}^{N-1} k\ell|\psi_{\mathcal{A}}(k,\ell)|^2$$

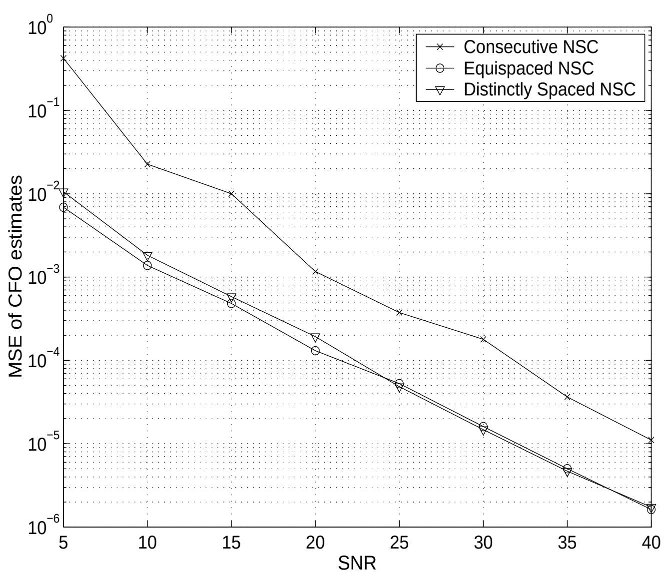
For $N_a \leq N/2$: equispace active sub-carriers
For $N_a \geq N/2$: equispace null sub-carriers
Average performance improves with # NSC's $N_z = N - N_a$

CRB and Optimal Placement of Null Subcarriers (4)

MCRB for different NSC placements; $N_z = 4$; one block



Performance Analysis



1 OFDM block

N = 64

$N_a = 54$

$N_z = 10$

$\nu_o \in$ [-2,2)

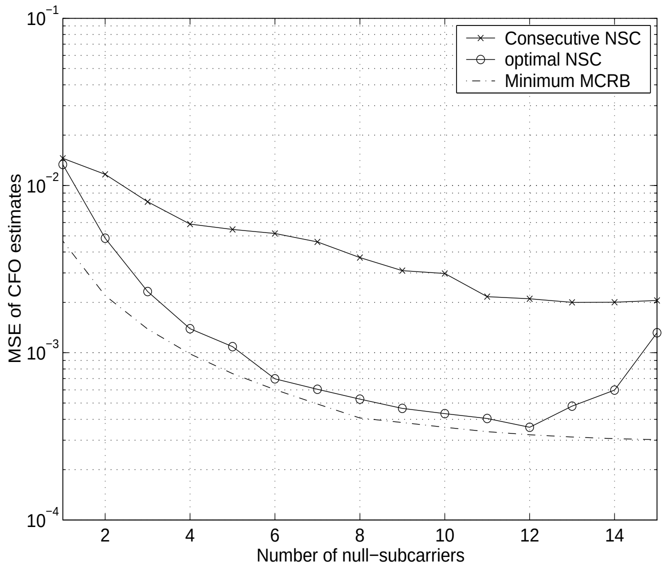
L = 8

$E\left\{|h_l|^2\right\} = e^{0.2l}$

Distinct:

1,2,4,7,...,56

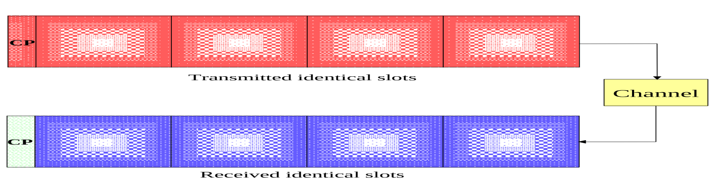
Performance Analysis (2)



1 OFDM block
N = 16
L = 4
$E\left\{|h_l|^2\right\} = e^{0.2l}$
$\nu_o \in$ [-2,2)
SNR = 15 dB
QPSK

Repetitive Slot-Based CFO Estimation

Motivation: CFO acquisition not requiring channel estimation



☞ J identical slots obtained by nulling all carriers not multiples of J

• $\boldsymbol{u} := \mathbf{F}^H \boldsymbol{s}$ made of J identical slots $(N = JQ) \rightarrow u(k) = u(k + \ell Q)$, $k = 0 \ldots Q - 1;\ \ell = 0 \ldots J - 1$

$$\hookrightarrow \quad x(k + \ell Q) = z(k)\ e^{j2\pi\nu\ell/J} + w(k + \ell Q)$$

$$z(k) = e^{j2\pi\nu k/N} H_c(k, :)\boldsymbol{u}$$

Repetitive Slot-Based CFO Estimation (2)

- We ignore the dependence between $\boldsymbol{z}$ and ν. Nonlinear Least Squares Estimator (NLLS):

$$\{\hat{\nu}_{REP}, \hat{\boldsymbol{z}}\} = \min_{\nu, \boldsymbol{z}} \sum_{\ell=0}^{J-1} \sum_{k=0}^{Q-1} \left| x(k+\ell Q) - z(k) e^{j2\pi\nu\ell/J} \right|^2$$

$$\hookrightarrow \qquad \hat{\nu}_{REP} = \arg\max_{\nu} \sum_{k=0}^{Q-1} \xi_\nu(k)$$

$$\xi_\nu(k) = \frac{1}{J} \left| \sum_{\ell=0}^{J-1} e^{-j2\pi\ell\nu/J} x(k+\ell Q) \right|^2$$

Acquisition range increases with J: $-\frac{J}{2} \leq \hat{\nu}_{REP} < \frac{J}{2}$

Repetitive Slot-Based CFO Estimation (3)

- NLS estimator can be rewritten as

$$\hat{\nu}_{REP} = \arg\max_{\nu} \sum_{m=1}^{J-1} \mathrm{Re}\left[r(mQ) e^{-j2\pi m\nu/J} \right]$$

$$r(\tau) = \sum_{k=0}^{M-\tau-1} x^*(k) x(k+\tau)$$

  - if $J = 2$, $\rightarrow$ closed-form solution (Schmidl/Moose algorithms)

$$\hat{\nu}_{REP} = \frac{1}{\pi} \arg\{r(N/2)\}$$

  - if $J > 2$, $\rightarrow$ no closed-form solution...

Repetitive Slot-Based CFO Estimation (4)

- Relationship between DML and NLS estimators

❑ Repetition of identical slots: VSC absent

➩ $\mathcal{K} = \{mJ,\ m = 0, ..., M/J - 1\}$ and

$$\psi_{\mathcal{K}}(\tau) = \frac{K}{M}\delta(\tau - mQ) \qquad m = 0, \pm 1, \pm 2, ...$$

➩ The repetitive slot-based and NSC-based are identical:

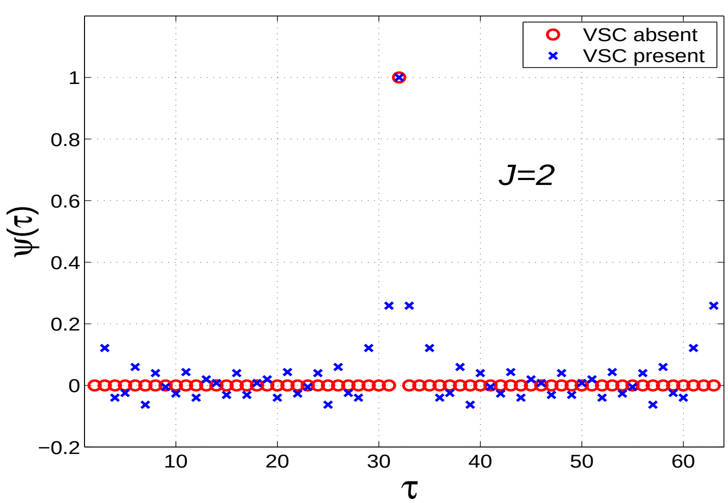
$$\hat{\nu}_{REP} \equiv \hat{\nu}_{NSC}$$

if no VSC (consecutive NSC dictated by system design)

Repetitive Slot-Based CFO Estimation (5)

- Relationship between DML and NLS estimators (cont.)

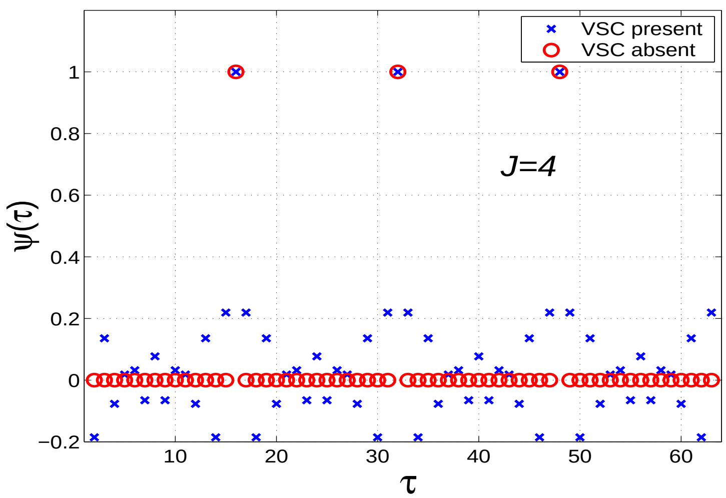
Plot of $\psi(\tau)$; $N = 64$;15 VSCs



Repetitive Slot-Based CFO Estimation (6)

- Relationship between DML and NLS estimators: J=4

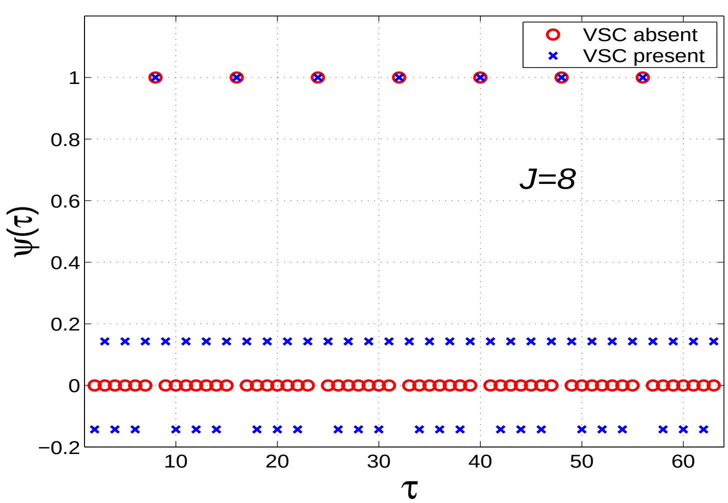
Plot of $\psi(\tau)$; N=64; 15 VSCs



Repetitive Slot-Based CFO Estimation (7)

- Relationship between DML and NLS estimators: J=8

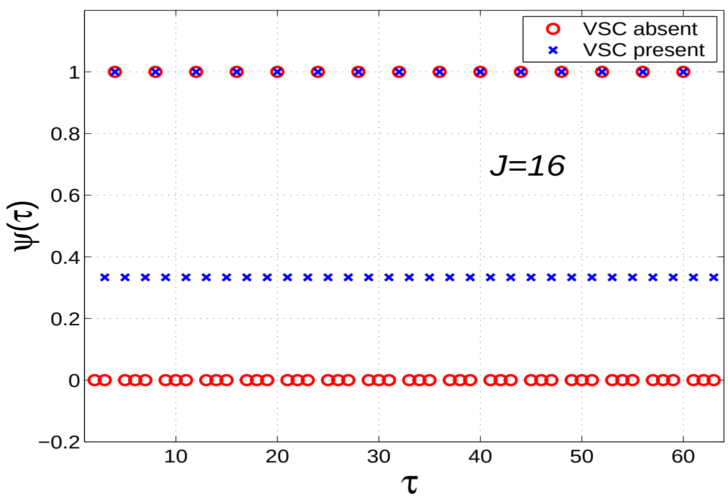
Plot of $\psi(\tau)$; N=64; 15 VSCs



Repetitive Slot-Based CFO Estimation (8)

- Relationship between DML and NLS estimators: J=16

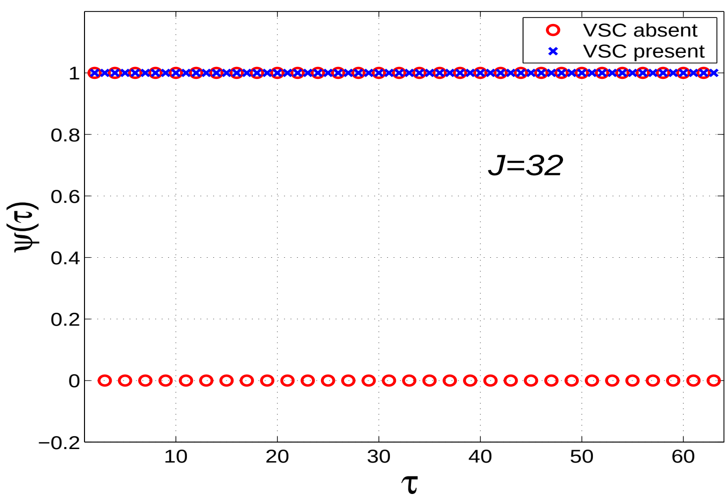
Plot of $\psi(\tau)$; N=64; 15 VSCs



Repetitive Slot-Based CFO Estimation (9)

- Relationship between DML and NLS estimators: J=32

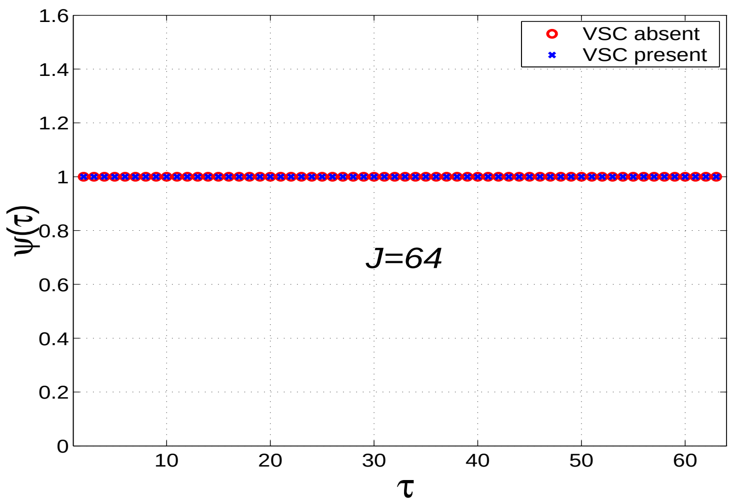
Plot of $\psi(\tau)$; N=64; 15 VSCs



Repetitive Slot-Based CFO Estimation (10)

- Relationship between DML and NLS estimators: J=64

Plot of $\psi(\tau)$; N=64; 15 VSCs



Repetitive Slot-Based CFO Estimation (11)

- Relationship between DML and NLS estimators (cont.)

  - Repetition of identical slots: VSC present

    - Most of the correlation coefficients contribute to the ML estimator
    - $\hat{\nu}_{REP}$ consists of using only the $(J-1)$ highest correlation coefficients, and is therefore an approximate ML estimator.
    - DML is computationally more demanding than NLS.
    - If $J = 2$, NLS is obtained in closed-form. If $J > 2$, no closed-form expression. Approximations given by the following algorithms.

Repetitive Slot-Based CFO Estimation (12)

- The 'BLUE' estimator: optimal combining of the correlations' phases.

❏ To avoid phase wrapping, the algorithm is based on

$$\varphi(m) = [\arg\{r(mQ)\} - \arg\{r((m-1)Q)\}]_{2\pi}$$

❏ Deriving the average (over Rayleigh channel) statistics of the $\varphi(m)$'s, the BLUE estimator is

$$\check{\nu}_{REP} = \frac{J}{2\pi}\sum_{m=1}^{p} w(m)\varphi(m)$$

p: design parameter (optimum value=$J/2$) and

$$w(m) = 3\frac{(J-m)(J-m+1) - p(J-p)}{p(4p^2 - 6pJ + 3J^2 - 1)}$$

☞ The amplitude of the correlations not exploited in BLUE...

Repetitive Slot-Based CFO Estimation (13)

- Approximate NLLS (ANNLS) estimator

❏ Rewrite the NLS criterion

$$\sum_{m=1}^{J-1} |r(mQ)| \cos(\phi_m - 2\pi m\nu/J)$$

ϕ_m: unwrapped phase of $r(mQ)$

❏ Small error approx. $\sin(\phi_m - j2\pi m\nu/J) \approx (\phi_m - j2\pi m\nu/J) \rightarrow$ ANLS estimator:

$$\tilde{\nu}_{REP} = \frac{J}{2\pi}\frac{\sum_{m=1}^{J-1} m|r(mQ)|\phi_m}{\sum_{m=1}^{J-1} m^2|r(mQ)|}$$

Repetitive Slot-Based CFO Estimation (14)

- Optimum number of identical slots (cont.)

❑ The repetitive-slot structure-based Conditional CRB:

$$\text{CCRB}(\nu) = \frac{3}{2\pi^2 N(1-1/J^2)\ SNR}\ \frac{1}{\gamma_H}$$

where we assumed no VSC and $|s_m| = 1, \forall m$ and where

$$\gamma_H = \sum_{m=0}^{N/J-1} \frac{|H_{nJ}|^2}{\sigma_H^2};$$

frequency diversity decreases with J

❑ Averaged CCRB:

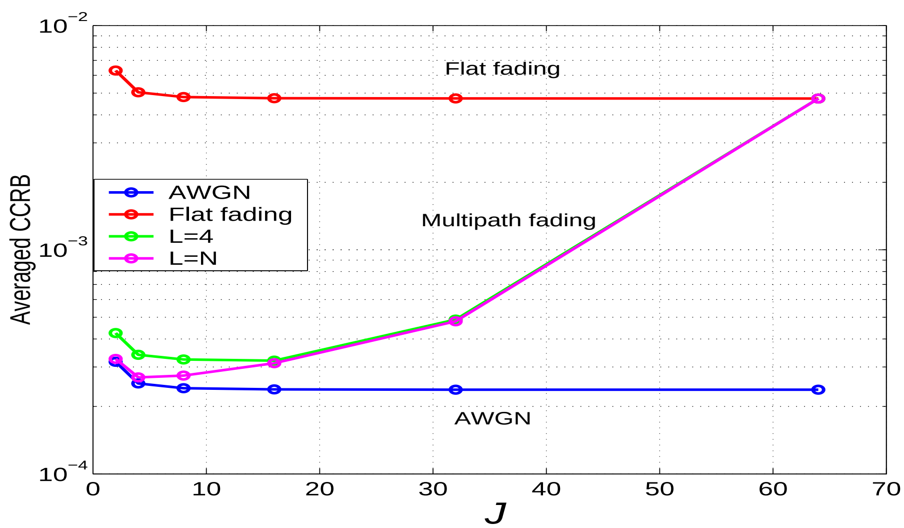
$$\text{ACCRB}(\nu) = \frac{3}{2\pi^2 N(1-1/J^2)\ SNR} E\left\{\frac{1}{\gamma_H}\right\}$$

→ no closed-form expression

→ Monte-Carlo simulations

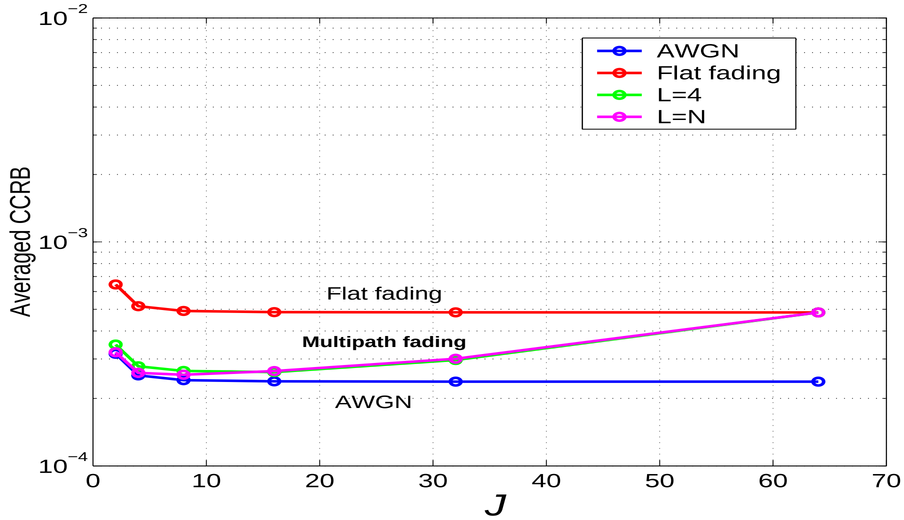
Repetitive Slot-Based CFO Estimation (15)

- Optimum number of identical slots (cont.) Rayleigh channel



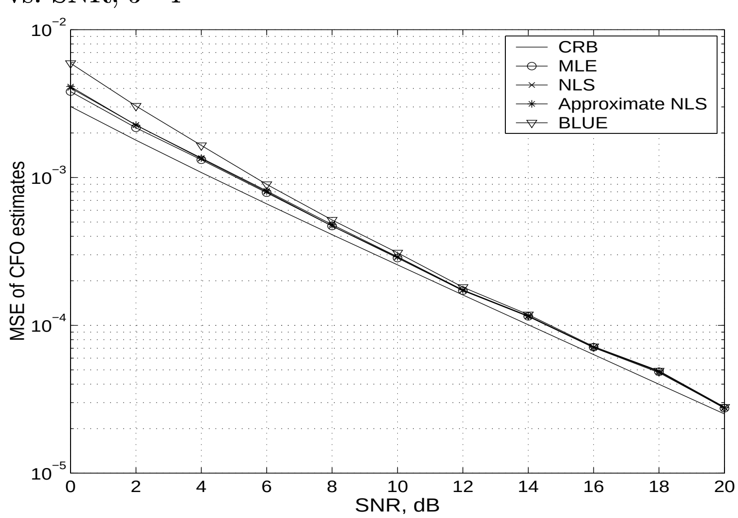
Repetitive Slot-Based CFO Estimation (16)

- Optimum number of identical slots (cont.) Ricean channel $\kappa = 4$



Comparisons

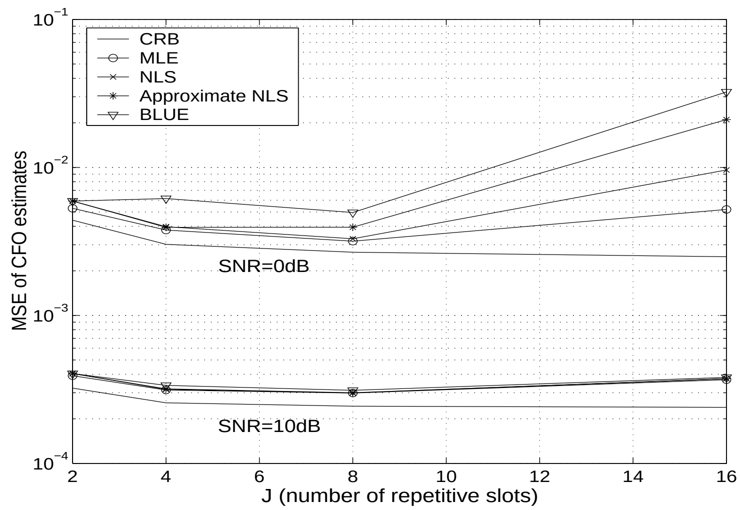
- MSE vs. SNR, J=4



$N = 64,\ N_a = 49,\ L = 15$, CFO $\in [-2, 2]$, $E\left\{|h_\ell|^2\right\} = e^{-0.2\ell}$, QPSK

Comparisons (2)

- MSE vs. # Repeated slots, J



Summary

- A computationally efficient algorithm
- Analytical performance analysis and CRB
- Relationship between the repetitive slot-based and the NSC-based MLE
  - Equivalent in the absence of VSC's
  - NSC is better if VSC's present

Part 2: Blind CFO estimation

Outline

- Constant-modulus algorithm
- Finite-alphabet algorithm
- Comparative study

Constant-Modulus Algorithm

- Assuming $|s_n| = 1,\ \forall n$, wlog

$$\hookrightarrow \quad H_n s_n = |H_n| e^{j\theta_n}; \qquad \theta_n = \angle H_n s_n$$

$$\hookrightarrow \quad x(k) = e^{j2\pi k\nu_o/N} \sum_{n\in\mathcal{A}} |H_n| e^{j\theta_n} e^{j2\pi kn/N} + w(k), \quad k = 0, ..., N-1$$

- The $|H_n|$'s are parameterized by only $(L+1)$ coefficients, the h_ℓ's
- The $H_n s_n$'s are parameterized by only $(N_a + L + 1)$ coefficients instead of $2N_a$, $\qquad (N_a = card(\mathcal{A}))$
- $w(k)$ is assumed AWGN

Constant-Modulus Algorithm (2)

- Deterministic Max-Likelihood
  - Treat $\{|H_n|\}, \{\theta_n\}$ as non-random unknowns
  - DML criterion

$$J(\nu, |\mathbf{H}|, \boldsymbol{\theta}) = \sum_{k=0}^{N-1} \left| x(k) - e^{j2\pi k\nu/N} \sum_{n\in\mathcal{A}} |H_n| e^{j\theta_n} e^{j2\pi kn/N} \right|^2$$

- can be rewritten as

$$J(\nu, |\mathbf{H}|, \boldsymbol{\theta}) = \sum_{k=0}^{N-1} |x(k)|^2 + \sum_{n\in\mathcal{A}} |H_n|^2 - 2N\mathrm{Re}\left[\sum_{n\in\mathcal{A}} |H_n| X(n+\nu) e^{-j\theta_n}\right]$$

- $X(f)$: DTFT of $\{x(k)\}$ at frequency f/N

$$X(f) = \sum_{k=0}^{N-1} x(k) e^{-j2\pi kf/N}$$

Constant-Modulus Algorithm (3)

- Deterministic Max-Likelihood, cont.
  - Setting $\partial J/\partial\theta_n = 0$,

$$\widehat{\theta}_n = \arg\{X(n+\nu)\}$$

    - If $|H_n| = 0$, θ_n becomes non-identifiable
    - $N_a > L$ ensures that $H_n \not\equiv 0$, $\forall n \in \mathcal{A}$
  - DML of $\{H_n\}$ and ν_o obtained by minimizing

$$\begin{aligned} J(\nu, |\mathbf{H}|) &= J_{VSC}(\nu) + J_A(\nu, |\mathbf{H}|) \\ J_{VSC}(\nu) &= \sum_{n\in\mathcal{Z}} |X(n+\nu)|^2 \qquad \text{due to VSC} \\ J_A(\nu, |\mathbf{H}|) &= \sum_{n\in\mathcal{A}} \left(|X(n+\nu)| - |H_n|\right)^2 \qquad \text{due to CM} \end{aligned}$$

Constant-Modulus Algorithm (4)

- Non-Dispersive Channel

❑ $H_n = h_0,\ \forall n \in \mathcal{A}$. Criterion becomes

$$\begin{aligned} J(\nu, |\mathbf{H}|) &= \sum_{n\in\mathcal{Z}} |X(n+\nu)|^2 + \sum_{n\in\mathcal{A}} \left(|X(n+\nu)| - |h_0|\right)^2 \\ &= \sum_{n=0}^{N-1} |X(n+\nu)|^2 + N_a|h_0|^2 - 2|h_0| \sum_{n\in\mathcal{A}} |X(n+\nu)| \end{aligned}$$

❑ DML of CFO:

$$\hat{\nu}_o = \arg\max_{\nu} \sum_{n\in\mathcal{A}} |X(n+\nu)|$$

☞ VSC-based estimator is equivalently obtained by maximizing the L_2-norm

$$\arg\min_{\nu} J_{VSC}(\nu) = \arg\max_{\nu} \sum_{n\in\mathcal{A}} |X(n+\nu)|^2$$

Constant-Modulus Algorithm (5)

- Dispersive Channel

❑ $J_{VSC}(\nu)$ is not a function of $|\mathbf{H}|$

❑ $J_A(\nu, |\mathbf{H}|)$ should be minimized wrt $|\mathbf{H}|$ under the constraint:

$$|H_n|^2 = \sum_{l,p=0}^{L} h_l h_p^* e^{-j2\pi(l-p)n/N}$$

❑ we modify $J_A(\nu, |\mathbf{H}|)$ into

$$J'_A(\nu, |\mathbf{H}|) = \sum_{n\in\mathcal{A}} \left(|X(n+\nu)|^2 - |H_n|^2\right)^2$$

Constant-Modulus Algorithm (6)

- Dispersive Channel, cont.

  - $|H_n|^2$ can be re-parameterized as

$$|H_n|^2 = \boldsymbol{c}_n^T \boldsymbol{\lambda}, \qquad n \in \mathcal{A}$$

$$\begin{aligned}
\boldsymbol{c}_n &= [1,\ \sqrt{2}\cos(2\pi n/N), \cdots, \sqrt{2}\cos(2\pi nL/N), \\
&\quad \sqrt{2}\sin(2\pi n/N), \cdots, \sqrt{2}\sin(2\pi nL/N)]^T \\
\boldsymbol{\lambda} &= [g_0, \sqrt{2}\mathrm{Re}\,[g_1], \cdots, \sqrt{2}\mathrm{Re}\,[g_L], \sqrt{2}\mathrm{Im}\,[g_1], \cdots, \sqrt{2}\mathrm{Im}\,[g_L]]^T \\
g_i &= \sum_{l=0}^{L-i} h_l^* h_{l+i}
\end{aligned}$$

Constant-Modulus Algorithm (7)

- Dispersive Channel, cont.

  - $\boldsymbol{\lambda}$ estimate:

$$\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} J'_A(\nu, |\mathbf{H}|) = \mathbf{C}_2^\dagger \sum_{n\in\mathcal{A}} |X(n+\nu)|^2 \boldsymbol{c}_n \ ,$$

$$\mathbf{C}_2 := \sum_{m\in\mathcal{A}} \boldsymbol{c}_m \boldsymbol{c}_m^T \ .$$

  - CFO estimate: obtained by minimizing $J(\nu) = J_{VSC}(\nu) + J_{CM}(\nu)$

$$J_{VSC}(\nu) = \sum_{n\in\mathcal{Z}} |X(n+\nu)|^2; \quad J_{CM}(\nu) = \sum_{n\in\mathcal{A}} \left(|X(n+\nu)| - \sqrt{Y(n;\nu)}\right)^2$$

$$Y(n;\nu) = \boldsymbol{c}_n^T \mathbf{C}_2^\dagger \sum_{n\in\mathcal{A}} |X(n+\nu)|^2 \boldsymbol{c}_n$$

Constant-Modulus Algorithm (8)

- Dispersive Channel, cont.

❏ The proposed VSC&CM estimate:

$$\hat{\nu}_o = \arg\min_{\nu} \sum_{n\in\mathcal{A}} \left(Y(n;\nu) - 2|X(n+\nu)|\sqrt{Y(n;\nu})\right)$$

$$Y(n;\nu) = \boldsymbol{c}_n^T \mathbf{C}_2^{\dagger} \sum_{n\in\mathcal{A}} |X(n+\nu)|^2 \boldsymbol{c}_n$$

$$\mathbf{C}_2 := \sum_{m\in\mathcal{A}} \boldsymbol{c}_m \boldsymbol{c}_m^T \qquad (\text{pre}-\text{computatble})$$

$$X(f) = \frac{1}{N}\sum_{k=0}^{N-1} x(k) e^{-j2\pi k f/N}$$

Constant-Modulus Algorithm (9)

- Extension to Multiple Blocks: Time-Invariant Channel

❏ Signal model for M blocks: (CFO and fading assumed constant across the set of blocks)

$$x_m(k) = e^{j2\pi k\nu_o/N} \sum_{n\in\mathcal{A}} H_n s_{m,n} e^{j2\pi kn/N} + w_m(k), \qquad m = 1,..,M$$

❏ VSC&CM CFO estimate:

$$\hat{\nu}_o = \arg\min_{\nu} \sum_{n\in\mathcal{A}} \left[Z(n;\nu) - 2\left(\frac{1}{M}\sum_{m=1}^{M} |X_m(n+\nu)|\right)\sqrt{Z(n;\nu)}\right]$$

$$Z(n;\nu) = \boldsymbol{c}_n^T \mathbf{C}_2^{\dagger} \sum_{n\in\mathcal{A}} \left(\frac{1}{M}\sum_{m=1}^{M} |X_m(n+\nu)|^2\right) \boldsymbol{c}_n$$

Constant-Modulus Algorithm (10)

- Extension to Multiple Blocks: Time-varying Channel

❑ Signal model for M blocks:

$$x_m(k) = e^{j2\pi k\nu_o/N} \sum_{n\in\mathcal{A}} H_{m,n} s_{m,n} e^{j2\pi kn/N} + w_m(k)$$

❑ VSC&CM CFO estimate:

$$\begin{aligned} \hat{\nu}_o &= \arg\min_{\nu} \sum_{m=1}^{M} J_m(\nu) \\ J_m(\nu) &= \sum_{n\in\mathcal{A}} \left(Y_m(n;\nu) - 2|X_m(n+\nu)| \sqrt{Y_m(n;\nu}\right) \end{aligned}$$

Finite-Alphabet Algorithm

❑ PSK constellations of size M satisfy:

$$s_n^M = 1$$

→ In the noiseless case

$$[X(n+\nu_o)]^M = H_n^M = \left[\sum_{l=0}^{L} h_l e^{-j2\pi ln/N}\right]^M = \sum_{l=0}^{ML} v_l e^{-j2\pi ln/N} = \boldsymbol{\gamma}_n^H \mathbf{v}$$

- $\boldsymbol{\gamma}_n = [1, e^{j2\pi n/N}, ..., e^{j2\pi MLn/N}]^T$; $\mathbf{v}$: $(ML+1)\times 1$

Finite-Alphabet Algorithm (2)

❑ Proposed criterion:

$$\begin{aligned} J(\nu) &= wJ_{VSC}(\nu) + (1-w)\bar{J}_{FA}(\nu, \mathbf{v}) \\ \bar{J}_{FA}(\nu, \mathbf{v}) &= \sum_{n \in \mathcal{A}} \left| [X(n+\nu]^M - \boldsymbol{\gamma}_n^H \mathbf{v} \right|^2 \end{aligned}$$

• If $ML + 1 < N_a$, $\boldsymbol{u}$ can be estimated as:

$$\hat{\mathbf{v}} = \boldsymbol{\Gamma}^{\dagger} \sum_{n \in \mathcal{A}} [X(n+\nu)]^M \boldsymbol{\gamma}_n \ ,$$

$$\boldsymbol{\Gamma} := \sum_{n \in \mathcal{A}} \boldsymbol{\gamma}_n \boldsymbol{\gamma}_n^H \ .$$

Finite-Alphabet Algorithm (3)

❑ The finite alphabet-based criterion becomes

$$J_{FA}(\nu) = \sum_{n \in \mathcal{A}} \left| [X(n+\nu]^M - Z(n;\nu) \right|^2$$

• $Z(n;\nu) = \boldsymbol{\gamma}_n^H \boldsymbol{\Gamma}^{\dagger} \sum_{n \in \mathcal{A}} [X(n+\nu)]^M \boldsymbol{\gamma}_n$

↪ Proposed VSC&FA-based estimator:

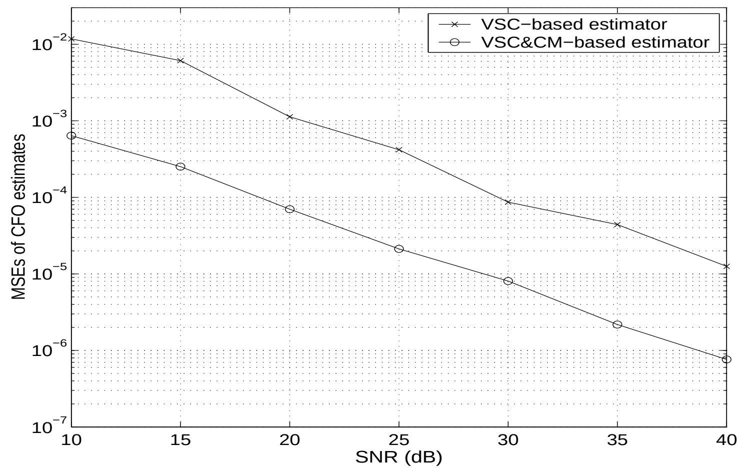
$$\hat{\nu}_o = \arg\min_{\nu} \ [\ w\ J_{VSC}(\nu)\ +\ (1-w)\ J_{FA}(\nu)\ ]$$

w: weight parameter to be adjusted. If no VSC, $w = 0$.

Comparative Study

- VSC vs CM: performance vs SNR.

MSE of CFO estimators vs. SNR; $L = 6$

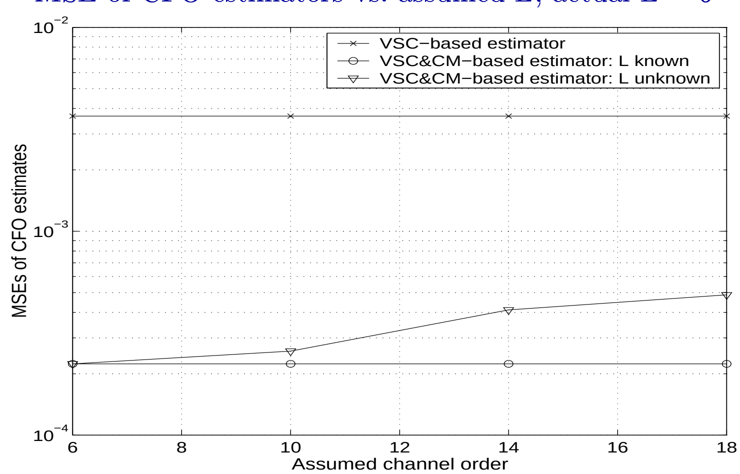


$N = 64,\ N_a = 49$, CFO in $[-2, 2]$ and $E\left\{|h_\ell|^2\right\} = e^{-0.2\ell}$; 8PSK.

Comparative Study (2)

- VSC vs CM: unknown channel order.

MSE of CFO estimators vs. assumed L; actual $L = 6$

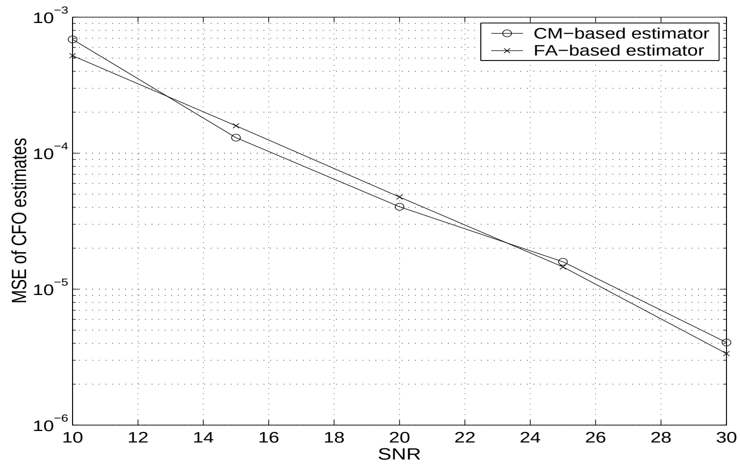


$N = 64,\ N_a = 49$, CFO in $[-2, 2]$ and $E\left\{|h_\ell|^2\right\} = e^{-0.2\ell}$; 8PSK.

Comparative Study (3)

- CM versus FA: BPSK case

MSE of CFO estimators vs. SNR; actual $L = 6$



$N = N_a = 64$, CFO in $[-2, 2]$ and $E\left\{|h_\ell|^2\right\} = e^{-0.2\ell}$

Summary

- CMA greatly outperforms VSC-based estimators
- CMA works even when the system is fully loaded
- CMA outperforms FA for M-PSK with $M > 2$
- Performance of CM close to data-aided algorithms
- Complexity is however greater than VSC and data-aided algorithms.

References

[1] M. ghogho and A. Swami, "Carrier Frequency Synchronization for OFDM Systems," in Signal Processing for Mobile Communications Handbook, Ibnkahla Edt., CRC, 2004.

[2] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Communications*, vol. 45, pp. 1613-1621, Dec. 1997.

[3] T. Pollet, M van Bladel and M. Moeneclaey "BER sensitivity of OFDM systems to CFO and Wiener phase noise", *IEEE Trans. Communications*, vol. 43, March 95.

[4] T. Pollet and M. Peeters, "Synchronization with DMT modulation", *IEEE Commun Mag*, 37(4), pp. 80-86, April 1999.

[5] L. Wei and S. Schegel, "Synchronization requirement for multi-user OFDM on satellite mobile and two path Rayleigh fading channel", *IEEE Trans. Communications*, vol. 43, pp. 887-895, Feb 1995.

[6] N. Lashkarian and S. Kiaei, "Class of cyclic-based estimators for frequency-offset estimation of OFDM systems", *IEEE Trans. Communications*, vol. 48, pp. 2139-2149, Dec 2000.

[7] X. Ma, and G. B. Giannakis, "Exploiting the Cyclic-Prefix for Blind Frequency-Offset Estimation in OFDM," *Proc. of 35th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, November 4-7, 2001.

[8] P.H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction", *IEEE Trans. Communications*, vol. 42, pp. 2908-2914, Oct 1994.

[9] M. Luise and Reggianini, "Carrier frequency offset acquisition and tracking for OFDM systems", *IEEE Trans. Communications*, vol. 44, Mar 1996.

[10] H. Liu and U. Tureli, "A high efficiency carrier estimator for OFDM communications," *IEEE Communications Lett.*, vol. 2, pp. 104-106, Apr. 1998.

[11] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications," *IEEE Communications Lett.*, vol. 3, pp. 75-77, Mar. 1999.

[12] H. Bölcskei, "Blind high-resolution uplink synchronization of OFDM-based multiple access schemes", *Proc. IEEE SPAWC'99*, pp. 166-169, Annapolis, MD, 1999.

[13] H.-K. Song, Y.-H. You, J.-H. Paik, and Y.-S. Cho, "Frequency-offset synchronization and channel estimation for OFDM-based transmission", *IEEE Communications Lett.*, vol. 4, pp. 95-97, Mar 2000.

[14] M. Morelli, A.N.D'Andrea, and U. Mengali, "Frequency ambiguity resolution in OFDM systems", *IEEE Communications Lett.*, vol. 4, pp. 134-136, April 2000

[15] U. Tureli, H. Liu, and M. Zoltowski, "OFDM blind carrier offset estimation: ESPRIT", *IEEE Trans. Communications*, vol. 48, Sept 2000.

[16] M. Morelli, A.N.D'Andrea, and U. Mengali, "Feedback frequency synchronization for OFDM applications", *IEEE Communications Lett.*, vol. 5, pp. 28-30, Jan 2001.

[17] J. Li, G. Liu and G.B. Giannakis, "Carrier frequency offset estimation for OFDM-based WLAN's" *IEEE Signal Processing Lett.*, 8, pp. 80-82, March 2001.

[18] M. Ghogho, A. Swami and G. B. Giannakis, "Opimized null-subcarrier selection for CFO estimation in OFDM over frequency-selective fading channels," *GLOBECOM*'2001, San Antonio, USA, Nov. 2001.

[19] X. Ma, C. Tepedelenlioglu, G. B. Giannakis, and S. Barbarossa, "Non-Data Aided Carrier-Offset Estimators for OFDM with Null Sub-Carriers: Identifiability, Algorithms, and Performance," *IEEE Journal on Selected Areas in Communications*, vol 19, pp. 2504-2515, Dec. 2001.

[20] S. Barbarossa, M. Pompili, and G. B. Giannakis, "Channel-independent synchronization of OFD multiple access systems", *IEEE Journal on Selected Areas in Communications*, vol 20, pp. 474-486, Feb 2002.

[21] M. Ghogho and A. Swami, "A blind frequency offset synchronization for OFDM transmitting constant-modulus symbols", *IEEE Communications Lett.*, vol. 5, August 2002.

[22] M. Ghogho and A. Swami, "Blind frequency-offset Estimation for OFDM and Multicarrier Systems," *2nd IEEE International Symposium on Signal Processing and Information Technology, ISSPIT2002*, Marrakech, Morocco. Dec. 2002

[23] M. Ghogho, "On blind frequency-offset synchronization for OFDM communications," *XI European Signal Processing Conference, EUSIPCO'2002*, Toulouse, France, Sep. 3-6, 2002.

[24] M. Ghogho and A. Swami, "Semi-blind frequency offset synchronization for OFDM," *IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP'02)*, Orlando, Florida, USA, May 2002.

[25] J.-J. van de Beek, M. Sandell, and P.O. Borjesson, "ML estimation of time and frequency offset in OFDM systems", *IEEE Trans. Signal Processing*, vol. 45, pp. 1800-1805, July 1997.

[26] A.J. Coulson, "Maximum likelihood synchronization for OFDM using a pilot symbol: algorithms", *IEEE Journal on Selected Areas in Communications*, vol 19, pp. 2486-2494, Dec 2001.

Soft information aided parameter estimation

L. Vandendorpe

Thanks to C. Herzet, V. Ramon, A. Dejonghe, X. Wautelet, M. Moeneclaey (UGent)

UCL Communications and Remote Sensing Lab, Louvain-la-Neuve, Belgium

1

Outline

- Introduction/motivation
- The EM algorithm
- Coding and the MAP algorithm
- Synchronization of coded systems with the EM algorithm
- Illustration of performance
- CSI estimation for coded MIMO transmission
- Illustration and performance
- Cramer-Rao bound with coded/prior information

2

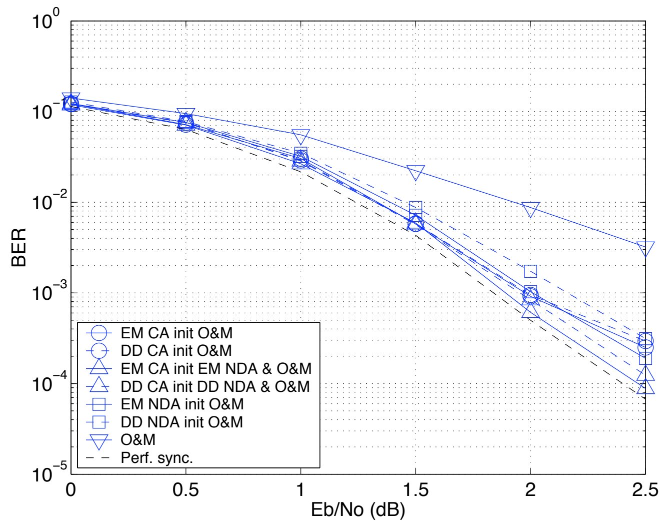
Outline

- **Introduction/motivation**
- The EM algorithm
- Coding and the MAP algorithm
- Synchronization of coded systems with the EM algorithm
- Illustration of performance
- CSI estimation for coded MIMO transmission
- Illustration and performance
- Cramer-Rao bound with coded/prior information

Motivation

- Synchronization or parameter estimation required even if not primary goal (data)
- Synchronization/CSI required at the RX; CSI also of interest for TX
- Recent advances in coding (error correcting codes): operation point at (very) low SNRs; powerful with perfect sync.
- **Can we still reliably estimate parameters at low SNRs ?**
  - Increase of number of pilot symbols decreases spectral efficiency
  - Problem for short block transmission; use the information carried by the whole block
  - Turbo receivers (for instance) produce soft information
  - **How to use this soft information for sync/CSI estimation ?**
- The EM algorithm is a nice framework to derive soft-data aided estimation algorithms; adaptations are desirable however

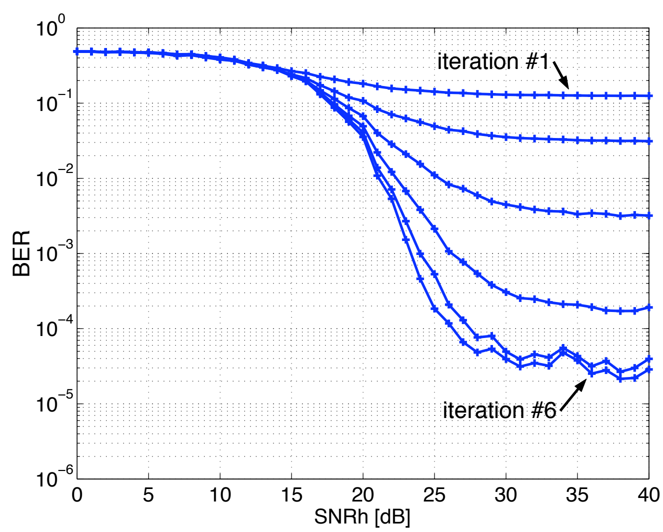
Illustration: impact of timing estimation



- Turbo code performance for various timing synchronizers

5

Illustration: impact of CSI



- Turbo equalizer for BICM over Porat channel

6

Parameter estimation

- Assume data symbols a_k, observation vector $\mathbf{r}$, parameter vector θ
- Ultimate goal (min SER): detection/decoding given by

$$\begin{aligned}\hat{a}_k &= \arg\max_{\tilde{a}_k} \; p(\tilde{a}_k \,|\, \mathbf{r}) \\ &= \arg\max_{\tilde{a}_k} \int_\theta p(\tilde{a}_k \,|\, \mathbf{r}, \theta)\, p(\theta |\, \mathbf{r}) \mathrm{d}\theta \qquad (1)\end{aligned}$$

- Suboptimal approach:

$$\begin{aligned}\hat{a}_k &= \arg\max_{\tilde{a}_k} \int_\theta p(\tilde{a}_k \,|\, \mathbf{r}, \theta)\; p(\theta |\, \mathbf{r}) \mathrm{d}\theta \qquad (2) \\ &\simeq \arg\max_{\tilde{a}_k} \; p(\tilde{a}_k \,|\, \mathbf{r}, \theta = \arg\max_{\tilde{\theta}} \; p(\tilde{\theta} \,|\, \mathbf{r})) \qquad (3)\end{aligned}$$

Maximum likelihood parameter estimation

- Assume no prior information about parameters (uniform distribution)
- About the estimates:

$$\begin{aligned}\hat{\theta} &= \arg\max_{\tilde{\theta}} \; p(\mathbf{r} \,|\, \tilde{\theta}) \qquad (4) \\ &= \arg\max_{\tilde{\theta}} \sum_{\mathbf{a}} p(\mathbf{r} \,|\, \mathbf{a}, \tilde{\theta})\; p(\mathbf{a}) \qquad (5)\end{aligned}$$

- Function of the information we have about the transmitted sequence

ML parameter estimation: DA mode

- Assume one uses pilots only
- We transmit a sequence of pilot symbols $\mathbf{a}_{\text{pilot}}$

$$\hat{\theta} = \arg\max_{\tilde{\theta}} \; p(\mathbf{r}_{\text{pilot}} \,|\, \mathbf{a}_{\text{pilot}}, \tilde{\theta}) \tag{6}$$

- Easy to compute
- Only exploits part of the available information

ML parameter estimation: NDA mode

- All transmitted sequences assumed equiprobable

$$\begin{aligned}\hat{\theta} &= \arg\max_{\tilde{\theta}} \sum_{\mathbf{a}} p(\mathbf{r} \,|\, \mathbf{a}, \tilde{\theta}) \; p(\mathbf{a}) && (7)\\ &= \arg\max_{\tilde{\theta}} \sum_{\mathbf{a}} p(\mathbf{r} \,|\, \mathbf{a}, \tilde{\theta}) \; \left(\frac{1}{|\mathcal{A}|}\right)^{N} && (8)\\ & && (9)\end{aligned}$$

- Untractable problem

ML parameter estimation: NDA mode

- All transmitted sequences assumed equiprobable

$$\hat{\theta} = \arg\max_{\tilde{\theta}} \sum_{\mathbf{a}} p(\mathbf{r} \mid \mathbf{a}, \tilde{\theta})\, p(\mathbf{a}) \quad (10)$$

$$= \arg\max_{\tilde{\theta}} \sum_{\mathbf{a}} \underbrace{p(\mathbf{r} \mid \mathbf{a}, \tilde{\theta})}_{\text{low SNR approx.}} \left(\frac{1}{|\mathcal{A}|}\right)^{N} \quad (11)$$

(12)

- Viterbi-Viterbi (phase), Oerder-Meyr (timing)

ML parameter estimation: Code aided mode

- Only existing codewords have non-zero probability:

$$\hat{\theta} = \arg\max_{\tilde{\theta}} \sum_{\mathbf{a}} p(\mathbf{r} \mid \mathbf{a}, \tilde{\theta})\, p(\mathbf{a}) \quad (13)$$

$$= \arg\max_{\tilde{\theta}} \sum_{\mathbf{a} \in \mathcal{B}} p(\mathbf{r} \mid \mathbf{a}, \tilde{\theta})\, p(\mathbf{a}) \quad (14)$$

- with $\mathcal{B} \subset \mathcal{A}^N$
- Untractable problem

Previous work (non exhaustive !)

- Basically two different paths are followed:
  - Parameter estimation can be embedded in the SISO module (”augmented trellis”) [Colavolpe(2000)][Anastasopoulos,Chugg (2001)][Mielczarek(2002)]
  - Iterative detection/parameter estimation, coined turbo sync/parameter estimation
    - Carrier phase estimation in turbo coded systems: [Lottici, Luise (2002)]; [Burr (2002)]; [Oh,Cheun (2001)]; [Morlet (2000)]; [Langlais (2000)].
    - Timing recovery: [Mielczarek, Svensson (2002)]; [Li Zhang, Burr (2002)]
    - Channel estimation: [Kobayashi-Boutros-Caire (2001)], [Guenach2000], [Kaleh-Vallet (1994)]
  - The methods proposed for turbo-sync are rather ”ad-hoc”
  - The EM framework provides a more structured approach

Outline

- Introduction/motivation
- **The EM algorithm**
- Coding and the MAP algorithm
- Synchronization of coded systems with the EM algorithm
- Illustration of performance
- CSI estimation for coded MIMO transmission
- Illustration and performance
- Cramer-Rao bound with coded/prior information

EM algorithm (1/3)

- Expectation-Maximization
- Seminal paper of [Dempster, Laird, Rubin, 1977]
- Can be used for the ML estimate or also the MAP estimate (Bayes framework, accounting for prior distribution)
- Example: assume observed data r and set of parameters to be estimated b
- The ML estimate of b is obtained as

$$\hat{b} = \arg\max_b p_r(r|b) \tag{15}$$

EM algorithm (2/3)

- Assume that instead of the *incomplete data* r one has access to the *complete data* z from which r may be obtained by a many-to-one mapping $r = H(z)$
- Definition of the complete data non unique; idea: $p_z(z|b)$ more easily obtained
- EM algorithms proceeds as follows
  - E-step (expectation): compute $\mathcal{Q}[b, \hat{b}^i] = \mathrm{E}[\ln p_z(z|b)|r, \hat{b}^i]$
  - M-step (maximization): solve $\hat{b}^{i+1} = \arg\max_b \mathcal{Q}[b, \hat{b}^i]$

EM algorithm (3/3)

- Idea: $\ln p_z(z|b)$ is not available; it is therefore a random variable and one maximizes its expectation given the observation r and the most recent value of the estimate $\hat{b}^i$
- Converges under mild conditions
- Can produce a local maximum
- Likelihood never decreases

Parameter estimation in the presence of nuisance (1/3)

- Let the complete data $\mathbf{r}$ denote a random vector obtained by expanding the received modulated-signal $r(t)$ onto a suitable basis and let $\mathbf{b}$ indicate a deterministic vector of parameters (sync parameters) to be estimated
- $\mathbf{r}$ also depends on a random discrete-valued nuisance parameter vector $\mathbf{a}$ independent of $\mathbf{b}$ and with a priori probability density function $p(\mathbf{a})$ (the data)
- Find the ML estimate $\hat{\mathbf{b}}$ of $\mathbf{b}$: $\hat{\mathbf{b}} = \arg\max_{\tilde{\mathbf{b}}}\{\ln p(\mathbf{r}|\tilde{\mathbf{b}})\}$, where

$$p(\mathbf{r}|\tilde{\mathbf{b}}) = \int_{\mathbf{a}} p(\mathbf{r}|\mathbf{a}, \tilde{\mathbf{b}})\, p(\mathbf{a})\, d\mathbf{a} \tag{16}$$

Parameter estimation in the presence of nuisance (2/3)

- Set $\mathbf{r}$ as the *incomplete* data set and $\mathbf{z} \triangleq [\mathbf{r}^T, \mathbf{a}^T]^T$ as the *complete* data set
- EM algorithm :

$$\mathcal{Q}(\tilde{\mathbf{b}}, \hat{\mathbf{b}}^{(n-1)}) = \int_{\mathbf{z}} p(\mathbf{z}|\mathbf{r}, \hat{\mathbf{b}}^{(\mathbf{n-1})}) \ \ln p(\mathbf{z}|\tilde{\mathbf{b}}) \ d\mathbf{z} \quad (17)$$

$$\hat{\mathbf{b}}^{(n)} = \arg\max_{\tilde{\mathbf{b}}} \{\mathcal{Q}(\tilde{\mathbf{b}}, \hat{\mathbf{b}}^{(n-1)})\} \quad (18)$$

Parameter estimation in the presence of nuisance (3/3)

- Using now the Bayes rule and taking into account the independence of $\mathbf{a}$ and $\mathbf{b}$ we may write

$$p(\mathbf{z}|\tilde{\mathbf{b}}) = p(\mathbf{r}, \mathbf{a}|\tilde{\mathbf{b}}) = p(\mathbf{r}|\mathbf{a}, \tilde{\mathbf{b}})\, p(\mathbf{a}|\tilde{\mathbf{b}}) = p(\mathbf{r}|\mathbf{a}, \tilde{\mathbf{b}})\, p(\mathbf{a}).$$

- It comes

$$\begin{aligned} \mathcal{Q}(\tilde{\mathbf{b}}, \hat{\mathbf{b}}^{(n-1)}) &= \int_{\mathbf{a}} p(\mathbf{a}|\mathbf{r}, \hat{\mathbf{b}}^{(n-1)}) \ \ln p(\mathbf{r}|\mathbf{a}, \tilde{\mathbf{b}}) \ d\mathbf{a} \\ &+ \int_{\mathbf{a}} p(\mathbf{a}|\mathbf{r}, \hat{\mathbf{b}}^{(n-1)}) \ \ln p(\mathbf{a}) \ d\mathbf{a}. \end{aligned} \quad (19)$$

- Finally, with the independence assumption

$$\mathcal{Q}(\tilde{\mathbf{b}}, \hat{\mathbf{b}}^{(n-1)}) = \int_{\mathbf{a}} p(\mathbf{a}|\mathbf{r}, \hat{\mathbf{b}}^{(\mathbf{n-1})}) \ \ln p(\mathbf{r}|\mathbf{a}, \tilde{\mathbf{b}}) \ d\mathbf{a}. \quad (20)$$

Parameter estimation in the presence of nuisance: comments

- Knowledge of a posteriori sequence (symbol) probabilities required

$$p(\mathbf{a}|\mathbf{r}, \hat{\mathbf{b}}^{(\mathbf{n-1})}) \tag{21}$$

- Should take into account the code information if any
- For convolutional code: can be computed exactly
- For turbo code or any iterative device, should be delivered after "a number" of iterations
- How do we get marginal a posteriori probabilities ?

Outline

- Introduction/motivation
- The EM algorithm
- **Coding and the MAP algorithm**
- Synchronization of coded systems with the EM algorithm
- Illustration of performance
- CSI estimation for coded MIMO transmission
- Illustration and performance
- Cramer-Rao bound with coded/prior information

How to improve coding (1/2) ?

- Classical codes:
  - ▷ block codes (BCH, Reed-Solomon,...)
  - ▷ convolutional codes (NSC, RSC)

  ⇒ Efficiency is increased by increasing the length of the codewords (block codes) or the code memory (convolutional codes).

  ⇒ Exponentially increasing complexity of the associated Maximum Likelihood (ML) decoding.

- Concatenated codes
  - ▷ Outer block code and inner convolutional code separated by an interleaver.
  - ▷ Separate decoding of the codes.

How to improve coding (2/2) ?

- Turbo-codes and iterative decoding (1995):

  ▷ Combination of several simple codes (constituent codes) in order to form a powerful global code.

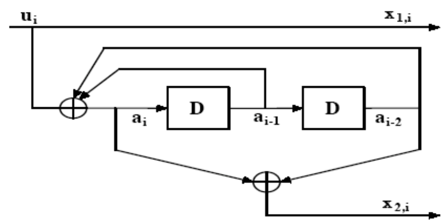
  ⇒ Attractive ML performances for the global code.

  ▷ Iterative decoding technique which allows the separate decoding of the constituent codes.

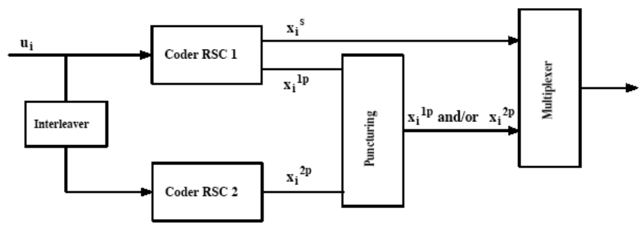
  ⇒ Performances close to those of the untractable ML decoding of the global code.

Classical turbo coding (1)

- rate-1/2 RSC code:



- Coding scheme:



Classical turbo coding (2)

- Parallel concatenation of 2 identical rate-1/2 RSC constituent codes.

- Pseudo-random interleaver: random permutation of the input sequence $\mathbf{u}$.

  $\Rightarrow$ The two constituent encoders are coding the same information sequence $\mathbf{u}$ but in a different order.

- For each input binary information symbol u_i, we keep:

  $\triangleright$ the systematic output $x_i^s = u_i$ of the first RSC encoder.

  $\triangleright$ the coded outputs x_i^{1p} and x_i^{2p} of the two RSC encoders.

Classical turbo coding (3)

- The outputs are multiplexed to form the sequence:

$$\{\ldots, u_i, x_i^{1p}, x_i^{2p}, u_{i+1}, x_{i+1}^{1p}, x_{i+1}^{2p}, u_{i+2}, x_{i+2}^{1p}, x_{i+2}^{2p}, \ldots,\}$$

$\Rightarrow$ code rate $r = 1/3$.

- The code rate may be increased through puncturing.

$\Rightarrow$ Classically the code rate is increased to 1/2 as follows:

$$\{\ldots, u_i, x_i^{1p}, u_{i+1}, x_{i+1}^{2p}, u_{i+2}, x_{i+2}^{1p}, u_{i+3}, x_{i+3}^{2p}, \ldots,\}$$

- In practice, only the trellis of the first constituent code is terminated with negligible impact on the performances of the global turbo-code.

Decoding complexity

- Maximum Likelihood decoding of the global turbo-code ?
  - ▷ $\mathcal{O}(2^N)$ complexity!

    N = information sequence length.
  - ▷ Totally untractable!

$\Rightarrow$ Suboptimal iterative decoding technique (turbo-decoding).

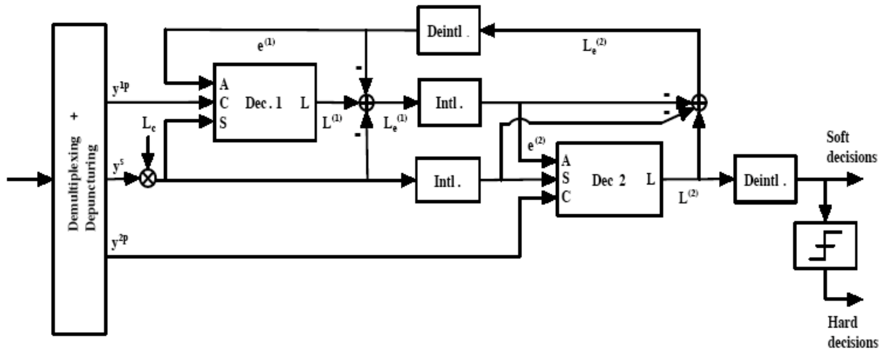
- ▷ $\mathcal{O}(\mathrm{n}(2^K + 2^K))$ complexity!

  K = constraint length of the constituent codes.

  n = number of iterations
- ▷ Performances (after convergence) close to those of ML decoding.

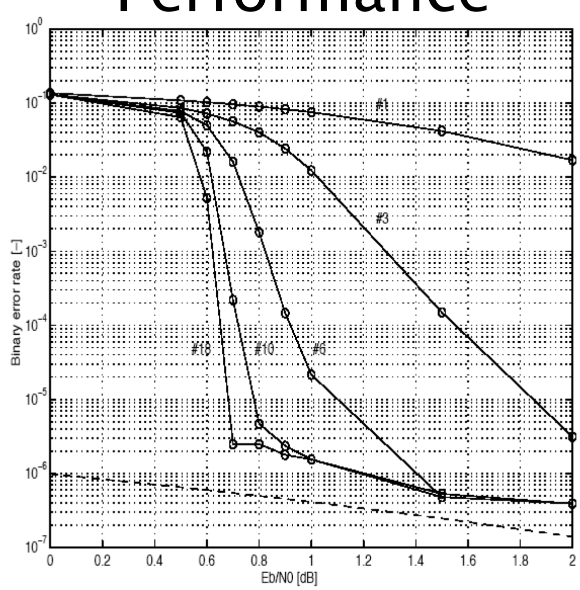
Iterative decoding

- Iterative decoding scheme:



- Soft information exchange between two soft-in/soft-out decoders.
- Progressive improvement in the reliability of the decisions.

Performance



Possible schemes

- Concatenation method:
  - ▷ Parallel concatenation of two or more constituent codes.
  - ▷ Serial concatenation of two or more constituent codes.
  - ▷ Hybrid concatenation of two or more constituent codes.

- Constituent codes:
  - ▷ rate-r convolutional codes (NSC or RSC).
  - ▷ rate-r block codes.

- In all cases:
  - ▷ Attractive asymptotic ML performances.
  - ▷ Iterative decoding.

Soft decisions and soft-in/soft-out (SISO) decoding

Soft decisions (1)

- Hard decision:

  A discrete symbol from the input constellation is associated with each received sample at the demodulator.

- Soft decision:

  A continuous value is kept at the demodulator.

$\Rightarrow$ Reliability measure associated with the symbol.

$\Rightarrow$ Allows the full exploitation of the available information.

Soft decisions (2)

- Soft decision vs. hard decision: 2dB Gain!
- Soft decision in the binary case: *Log-Likelihood Ratio* (LLR).
- LLR of a discrete binary random variable $\mathbf{U}$:

$$L_U(u) = \ln\left(\frac{P_U(u=1)}{P_U(u=0)}\right)$$

Absolute value $\Rightarrow$ Reliability of the decision.

Sign $\Rightarrow$ hard decision.

$$\hat{u} = \begin{cases} 1 & \text{if } L_U(u) \geq 0 \\ 0 & \text{if } L_U(u) < 0 \end{cases}$$

Soft output of a channel (1)

- Information symbol $u \in \{0, 1\}$ BPSK mapped to symbol $b \in \{+1, -1\}$.
- Memoryless channel associating the input symbol $b \in \{+1, -1\}$ with the received sample y.
- The LLR of symbol u given the reception of symbol y is:

$$L(u|y) = \ln\left(\frac{P(u=1|y)}{P(u=0|y)}\right) = \ln\left(\frac{P(b=+1|y)}{P(b=-1|y)}\right)$$

Using the Bayes rule:

$$\begin{aligned} L(u|y) &= \ln\left(\frac{P(y|u=1)}{P(y|u=0)}\right) + \ln\left(\frac{P(u=1)}{P(u=0)}\right) \\ &= \ln\left(\frac{P(y|b=+1)}{P(y|b=-1)}\right) + \ln\left(\frac{P(u=1)}{P(u=0)}\right) \\ &= L_c y + L_a(u) \end{aligned}$$

Soft output of a channel (2)

- Two terms in $L(u|y)$:
  - ▷ $L_c y$ is called *soft output of the channel*.
    Soft information associated with u, brought by the reception of y.
  - ▷ $L_a(u)$ corresponds to the information available *a priori* at the receiver about u, independently of the reception of y.
- In the case of an AWGN channel, with noise variance σ^2:

$$\ln\left(\frac{P(y|b=+1)}{P(y|b=-1)}\right) = \ln\left(\frac{\exp(-\frac{1}{2\sigma^2}(b-1)^2}{\exp(-\frac{1}{2\sigma^2}(b+1)^2}\right) = \frac{2}{\sigma^2}y$$

$\Rightarrow$ The reliability value of the channel is given by $L_c = \frac{2}{\sigma^2}$.

SISO decoder (1)

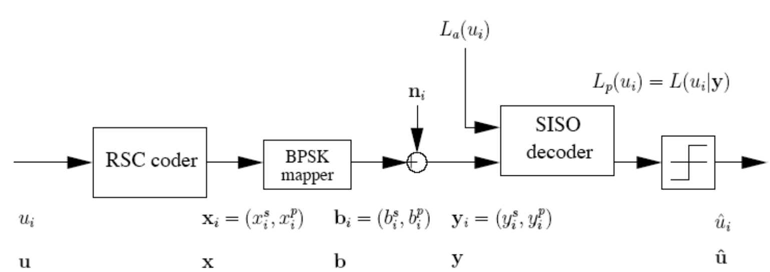
- Decoder working with soft values at its inputs and outputs

$\Rightarrow$ *Soft-In/Soft-out* (SISO) decoder.

- Particular case here: rate-1/2 systematic code (straightforward generalization).

SISO decoder (2)

- Coder input: binary information symbols u_i $(i = 1, \ldots, N)$
- Coder output: coded symbols x_i^s, x_i^p.
- Coder output sequence: $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ with $\mathbf{x}_i = (x_i^s, x_i^p)$.
- BPSK mapping $\Rightarrow$ sequence $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_N)$ with $\mathbf{b}_i = (b_i^s, b_i^p)$ and $b_i^s = 2x_i^s - 1$, $b_i^p = 2x_i^p - 1$.
- Channel $\Rightarrow$ output sequence $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$ with $\mathbf{y}_i = (y_i^s, y_i^p)$.



SISO decoder (3)

- Inputs of the SISO decoder:
  - ▷ Sequence $\mathbf{y}$ of the received symbols.

    Equivalently: sequences $\mathbf{y}^s = (y_1^s, \ldots, y_N^s)$ and $\mathbf{y}^p = (y_1^p, \ldots, y_N^p)$.

    Equivalently: sequences of soft channel values $L_c\mathbf{y}^s$ and $L_c\mathbf{y}^p$.
  - ▷ Sequence $\mathbf{L}_a$ of a priori information about the information symbols $\{u_i\}$ $(i = 1, \ldots, N)$:

$$L_a(u_i) = \ln\left(\frac{P(u_i = 1)}{P(u_i = 0)}\right)$$

- Output of the SISO decoder:
  - ▷ LLR of the a posteriori probabilities of the information symbols:

$$L_p(u_i) = \ln\left(\frac{P(u_i = 1|\mathbf{y})}{P(u_i = 0|\mathbf{y})}\right) = \ln\left(\frac{P(u_i = 1|\mathbf{y}^s, \mathbf{y}^p)}{P(u_i = 0|\mathbf{y}^s, \mathbf{y}^p)}\right)$$

SISO decoder (4)

- A SISO decoder is implemented with algorithms able to estimate the symbol a posteriori probabilities.
- From SISO decoder output, decoded symbols obtained via hard decision:

$$\hat{u}_i = \begin{cases} 1 & \text{if } L_p(u_i) \geq 0 \\ 0 & \text{if } L_p(u_i) < 0 \end{cases}$$

- SISO decoder + hard decision ⇒ symbol-by-symbol MAP decoding:

$$\hat{u}_i = \arg\max_u P(u|\mathbf{y})$$

- Fundamental property (SYSTEMATIC CODE):

$$L_p(u_i) = (L_c y_i^s) + L_a(u_i) + L_e(u_i)$$

⇒ The *a posteriori* LLR $L_p(u_i)$ can be split into three terms.

SISO decoder (5)

⇒ The *a posteriori* LLR $L_p(u_i)$ can be split into three terms:

▷ $L_c y_i^s$: information about symbol $x_i^s = u_i$ through direct (noisy) observation at the output of the channel.

▷ $L_a(u_i)$: a priori information about the information symbol u_i.

▷ $L_e(u_i)$: *extrinsic information* about the information symbol u_i.

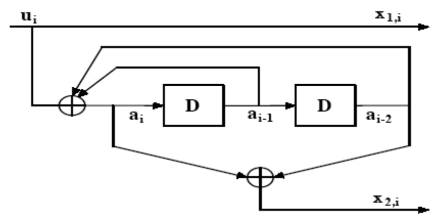
⇒ Supply of soft information brought by the decoding process.

⇒ Depends on y_m^s $(m = 1, \ldots, N; m \neq i)$, y_m^p $(m = 1, \ldots, N)$, $L_a(u_m)$ $(m = 1, \ldots, N; m \neq i)$.

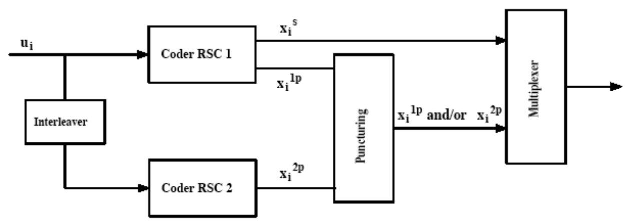
Iterative decoding

Classical turbo coding scheme

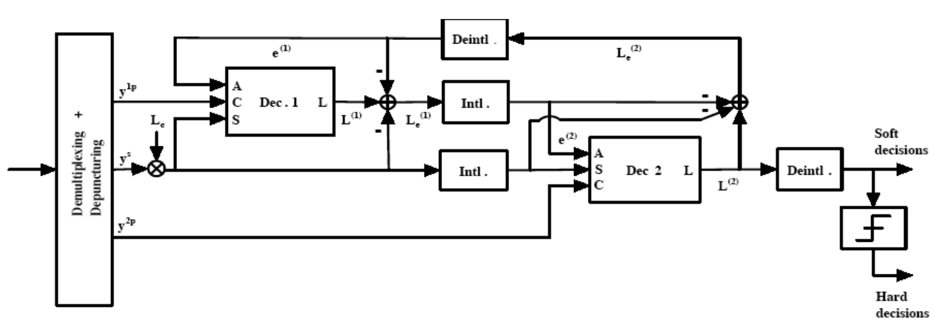
- rate-1/2 RSC code:



- Coding scheme:



Iterative decoding (1)



- Demultiplexing $\Rightarrow$ sequence $\mathbf{y}^s$ (systematic output of CC1), sequences $\mathbf{y}^{1p}$ and $\mathbf{y}^{2p}$ (coded outputs of CC1 and CC2).
- If puncturing: missing values are replaced by 0.

Iterative decoding (2)

- Decoding scheme based on the association of 2 SISO decoders corresponding to the 2 constituent codes of the turbo-code.
- These SISO decoders collaborate through an extrinsic information exchange.
- Iterative processing leads to progressive increase in the reliability of the decisions.
- Performances close (after convergence) to those of the untractable ML decoding of the turbo-code.

Iterative decoding (3)

- The first decoder ensures the decoding of the first constituent code based on the received sequences $\mathbf{y}^s$, $\mathbf{y}^{1p}$ and on the a priori information sequence $\mathbf{L}_a^{(1)}$ about the transmitted symbols.
- At the first iteration: no a priori information $\Rightarrow L_a^{(1)}(u_i) = 0 \quad \forall i$.
- It outputs a sequence $\mathbf{L}_p^{(1)}$ of a posteriori LLRs $L_p^{(1)}(u_i)$:

$$L_p^{(1)}(u_i) = \ln\left(\frac{P(u_i = 1|\mathbf{y}^s, \mathbf{y}^{1p})}{P(u_i = 0|\mathbf{y}^s, \mathbf{y}^{1p})}\right)$$

- The extrinsic component $\mathbf{L}_e^{(1)}$ is then extracted from the output $\mathbf{L}_p^{(1)}$:

$$L_e^{(1)}(u_i) = L_p^{(1)}(u_i) - L_c y_i^s - L_a^{(1)}(u_i)$$

Iterative decoding (4)

- The second decoder ensures the decoding of the second constituent code based on the received sequences $\mathbf{y}^s$ (interleaved), $\mathbf{y}^{2p}$ and on the a priori information sequence $\mathbf{L}_a^{(2)}$ about the transmitted symbols.

- $\mathbf{L}_a^{(2)}$ is obtained by interleaving of the extrinsic information sequence $\mathbf{L}_e^{(1)}$ produced by decoder 1.

- The second decoder outputs a sequence $\mathbf{L}_p^{(2)}$ of a posteriori LLRs $L_p^{(2)}(u_j)$:

$$L_p^{(2)}(u_j) = \ln\left(\frac{P(u_j = 1|\mathbf{y}^s, \mathbf{y}^{2p})}{P(u_j = 0|\mathbf{y}^s, \mathbf{y}^{2p})}\right)$$

- Again, the extrinsic component $\mathbf{L}_e^{(2)}$ is extracted from the output $\mathbf{L}_p^{(2)}$:

$$L_e^{(2)}(u_j) = L_p^{(2)}(u_j) - L_c y_j^s - L_a^{(2)}(u_j)$$

Iterative decoding (5)

- A second iteration may now begin:

  The sequence $\mathbf{L}_e^{(2)}$ of extrinsic information produced by decoder 2 becomes (after deinterleaving) the sequence $\mathbf{L}_a^{(1)}$ of a priori information for the decoder 1.

- The fundamental principle is that the extrinsic information provided by one of the decoders becomes the a priori information for the other.

$\Rightarrow$ Improved quality of the decoding for each of the SISO decoders.

- Through iterations: progressive increase in the reliability of the decisions.

Iterative decoding (6)

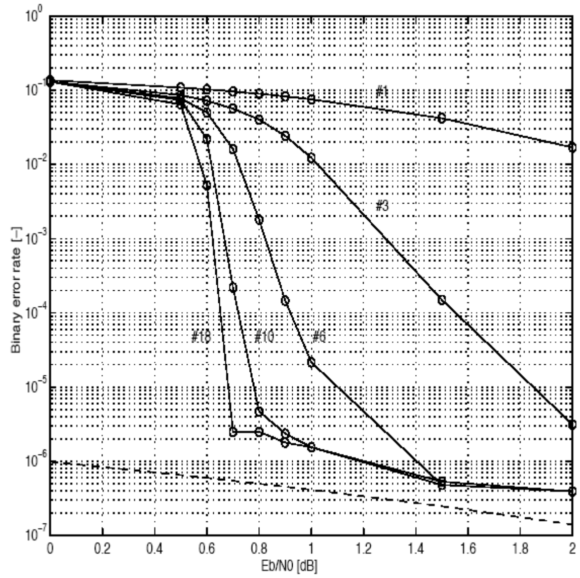
- At the last iteration, the best estimation available about the transmitted symbols is given by the deinterleaved a posteriori output of the second decoder.
- The final hard decision is:

$$\hat{u}_i = \begin{cases} 1 & \text{if } L_p^{(2)}(u_i) \geq 0 \\ 0 & \text{if } L_p^{(2)}(u_i) < 0 \end{cases}$$

Iterative decoding (7)

- This scheme will perform efficiently if the two SISO decoders are decorrelated information sources one for each other.
- This decorrelation is possible thanks to the interleaver.
- This is also the reason why only the extrinsic part of the a posteriori LLRs at the output of the SISO decoders is used during the exchange process.

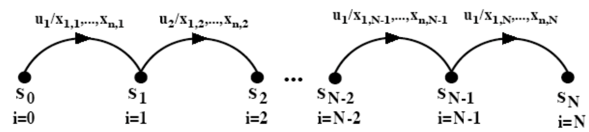
Performance



Symbol by symbol algorithm

Markov process

- Markov process:
  - ▷ State s_i in finite set $\mathcal{S}$ a each time i ($i = 0, \ldots, N$).
  - ▷ Input: sequence $\mathbf{u}$, output: sequence $\mathbf{x}$.
  - ▷ Particluar case: 1 input symbol, n output symbols:



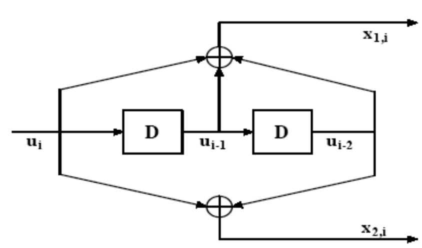
- At time i, transition between states $s_{i-1} = s'$ and $s_i = s$ caused by symbol u_i ($i = 1, \ldots, N$) generates symbols $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,n})$ of sequence $\mathbf{x}$.

- Fundamental property:

$$P(s_i|s_{i-1}, \ldots, s_0) = P(s_i|s_{i-1})$$

Convolutional code (1)

- Convolutional code = Markov process



- State = content of the shift-registers.

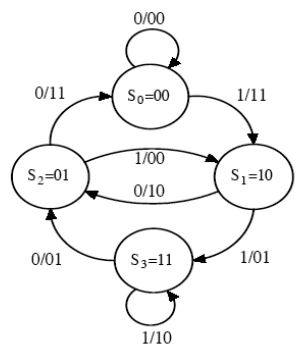
⇒ In the case of an NSC code:

$$s_i = (u_i, \ldots, u_{i-M+1})$$

- Memory $M \Rightarrow 2^M$ possible states S_j ($j = 0, \ldots, 2^M - 1$).

Convolutional code (2)

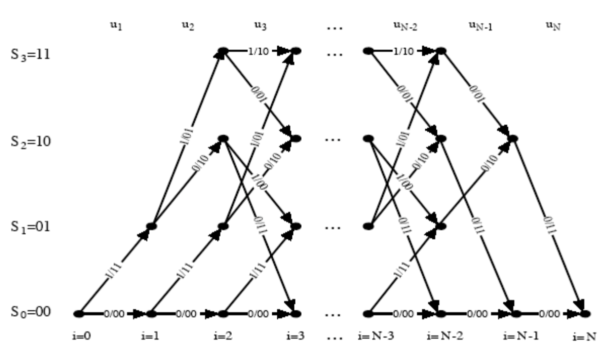
- State diagram representation of a convolutional code:



- Encoding of a sequence ⇒ path through the state diagram.

Convolutional code (3)

- Trellis representation of a convolutional code:



- Encoding of a sequence ⇒ path through the trellis diagram.

Transmission scheme (1)

- Rate $r = 1/n$ convolutional encoder.
- Memory M encoder $\Rightarrow 2^M$ possible states in set $\mathcal{S}$.
- Coder state at timestep i: s_i.
- At timestep i, transition (s', s) between states $s_{i-1} = s'$ and $s_i = s$.

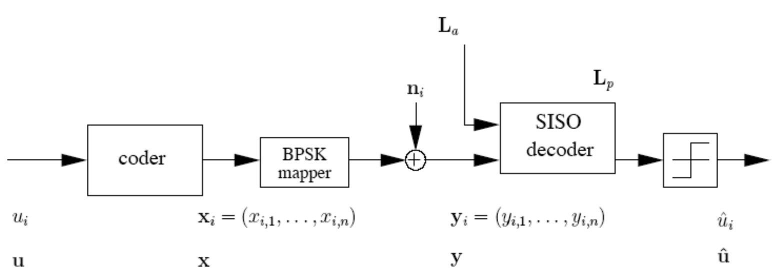
- Input: binary information symbols u_i $(i = 1, \ldots, N)$
- Output: coded symbols $x_{i,1}, \ldots, x_{i,n}$.
- Output sequence: $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ with $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,n})$.

- BPSK mapping $\Rightarrow$ sequence $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_N)$
  with $\mathbf{b}_i = (b_{i,1}, \ldots, b_{i,n})$ and $b_{i,j} = 2x_{i,j} - 1$.

- Channel $\Rightarrow$ output sequence $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$ with $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,n})$.

Transmission scheme (2)

- Transmission scheme:



SISO decoder

- Input of the SISO decoder:
  - ▷ Received sequence $\mathbf{y}$.
  - ▷ A priori LLR sequence $\mathbf{L}_a$ with entries $L_a(u_i) = \ln \frac{P(u_i=1)}{P(u_i=0)}$.
- Output of the SISO decoder:
  - ▷ A posteriori LLR sequence $\mathbf{L}_p$ with entries $L_p(u_i) = \ln \frac{P(u_i=1|\mathbf{y})}{P(u_i=0|\mathbf{y})}$.
- Data:
  - ▷ initial state s_0 and final state s_N.
  - ▷ Code trellis.
  - ▷ Noise variance σ^2.

BCJR algorithm (1)

- Symbol-by-symbol a posteriori probability (APP) evaluation

$\Leftrightarrow$ Minimization of the symbol error rate $\Rightarrow$ optimal!

- BCJR algorithm (1974):

  Evaluation of the a posteriori probabilities of the states and transitions of a Markov source observed through a discrete-time memoryless channel.

BCJR algorithm (2)

- The BCJR algorithm provides the a posteriori states and transitions probabilities:

$$P(s_i = s|\mathbf{y}) \text{ or } P(s_i = s, \mathbf{y})$$

and:

$$P(s_{i-1} = s', s_i = s|\mathbf{y}) \text{ or } P(s_{i-1} = s', s_i = s, \mathbf{y})$$

on the basis of:

$\Rightarrow$ the received sequence: $\mathbf{y}$.

$\Rightarrow$ the channel type $\rightarrow p(\mathbf{y}_i|s_{i-1} = s', s_i = s)$.

$\Rightarrow$ the transitions a priori probabilities: $p(s_i = s|s_{i-1} = s')$.

- Slight modification necessary to obtain a SISO decoder.

$\Rightarrow$ "MAP" algorithm.

MAP algorithm (1)

- Slight modification of the BCJR algorithm $\Rightarrow$ "MAP" algorithm.

- The goal of the MAP algorithm is to provide an APP LLR (soft output):

$$L_p(u_i) = \ln\left(\frac{P(u_i = 1|\mathbf{y})}{P(u_i = 0|\mathbf{y})}\right)$$

based on the received sequence $\mathbf{y}$ and the a priori information sequence $\mathbf{L}_a$.

$\Rightarrow$ Optimal algorithm for the implementation of a SISO decoder.

- Combined with hard detection, it realizes MAP decoding:

$$\hat{u}_i = \begin{cases} 1 & \text{if } L_p(u_i) \geq 0 \\ 0 & \text{if } L_p(u_i) < 0 \end{cases}$$

Equivalent to:

$$\hat{u}_i = \arg\max_u P(u|\mathbf{y})$$

MAP algorithm (2)

- The MAP algorithm provides the a posteriori LLR:

$$L_p(u_i) = \ln\left(\frac{P(u_i = 1|\mathbf{y})}{P(u_i = 0|\mathbf{y})}\right)$$

- As the knowledge of $s_{i-1} = s'$ and $s_i = s$ determines u_i, we have

$$L_p(u_i) = \ln\left(\frac{\sum_{\mathcal{S}^+} p(s_{i-1} = s', s_i = s|\mathbf{y})}{\sum_{\mathcal{S}^-} p(s_{i-1} = s', s_i = s|\mathbf{y})}\right)$$

where S_+ (resp. S_-) is the set of transitions $(s_{i-1} = s', s_i = s)$ caused by a symbol $u_i = 1$ (resp. $u_i = 0$).

- This can be simplified as:

$$L_p(u_i) = \ln\left(\frac{\sum_{\mathcal{S}^+} p(s_{i-1} = s', s_i = s, \mathbf{y})}{\sum_{\mathcal{S}^-} p(s_{i-1} = s', s_i = s, \mathbf{y})}\right)$$

MAP algorithm (3)

- The probability $p(s_{i-1} = s', s_i = s, \mathbf{y})$ is computed as (BCJR algorithm):

$$\begin{aligned}
&p(s_{i-1} = s', s_i = s, \mathbf{y}) \\
&= p(s_{i-1} = s', \mathbf{y}_{j<i}) p(\mathbf{y}_{j\geq i}, s_i = s|s_{i-1} = s', \mathbf{y}_{j<i}) \\
&= p(s_{i-1} = s', \mathbf{y}_{j<i}) p(\mathbf{y}_{j\geq i}, s_i = s|s_{i-1} = s') \\
&= p(s_{i-1} = s', \mathbf{y}_{j<i}) p(\mathbf{y}_i, \mathbf{y}_{j>i}, s_i = s|s_{i-1} = s') \\
&= p(s_{i-1} = s', \mathbf{y}_{j<i}) \frac{p(\mathbf{y}_i, \mathbf{y}_{j>i}, s_i = s, s_{i-1} = s')}{p(s_{i-1} = s')} \\
&= p(s_{i-1} = s', \mathbf{y}_{j<i}) \frac{p(\mathbf{y}_i, s_i = s, s_{i-1} = s')}{p(s_{i-1} = s')} p(\mathbf{y}_{j>i}|\mathbf{y}_i, s_i = s, s_{i-1} = s') \\
&= p(s_{i-1} = s', \mathbf{y}_{j<i}) p(\mathbf{y}_i, s_i = s|s_{i-1} = s') p(\mathbf{y}_{j>i}|s_i = s) \\
&= p(s_{i-1} = s', \mathbf{y}_{j<i}) p(\mathbf{y}_i|s_{i-1} = s', s_i = s) P(s_i = s|s_{i-1} = s') p(\mathbf{y}_{j>i}|s_i = s)
\end{aligned}$$

NB: if $s_i = s$ is known, events after time i do not depend on $\mathbf{y}_{j<i+1}$.

MAP algorithm (4)

- Defining:
  - $\triangleright\ \alpha_{i-1}(s') = p(s_{i-1} = s', \mathbf{y_{j<i}})$,
  - $\triangleright\ \beta_i(s) = p(\mathbf{y_{j>i}}|s_i = s)$,
  - $\triangleright\ \gamma_i(s', s) = p(\mathbf{y}_i, s_i = s|s_{i-1} = s')$

    $= p(\mathbf{y_i}|s_{i-1} = s', s_i = s)p(s_i = s|s_{i-1} = s')$,

We have:

$$p(s_{i-1} = s', s_i = s, \mathbf{y}) = \alpha_{i-1}(s').\gamma_i(s', s).\beta_i(s)$$

MAP algorithm (5)

- Parameters α are computed as follows:

$$\begin{aligned}
\alpha_i(s) &= p(s_i = s, \mathbf{y}_{j<i+1}) \\
&= \sum_{s' \in \mathcal{S}} p(s_{i-1} = s', s_i = s, \mathbf{y}_{j<i+1}) \\
&= \sum_{s' \in \mathcal{S}} p(s_{i-1} = s', s_i = s, \mathbf{y}_{j<i}, \mathbf{y}_i) \\
&= \sum_{s' \in \mathcal{S}} p(s_{i-1} = s', \mathbf{y}_{j<i}) p(s_i = s, \mathbf{y}_i|s_{i-1} = s', \mathbf{y}_{j<i}) \\
&= \sum_{s' \in \mathcal{S}} p(s_{i-1} = s', \mathbf{y}_{j<i}) p(s_i = s, \mathbf{y}_i|s_{i-1} = s') \\
&= \sum_{s' \in \mathcal{S}} \alpha_{i-1}(s').\gamma_i(s', s)
\end{aligned}$$

MAP algorithm (6)

- Parameters α are obtained via a forward recursion:

$$\alpha_i(s) = \sum_{s' \in \mathcal{S}} \alpha_{i-1}(s').\gamma_i(s', s)$$

for $(i = 0, \ldots, N-1)$ and $\forall s \in \mathcal{S}$.

- The initial conditions are:

$$\alpha_0(s_0) = 1 \text{ and } \alpha_0(s \neq s_0) = 0$$

$\Leftrightarrow$ The initial state is known to be s_0.

MAP algorithm (7)

- Parameters β are computed as follows:

$$\begin{aligned}
\beta_{i-1}(s') &= p(\mathbf{y}_{j>i-1}|s_{i-1} = s') \\
&= \sum_{s \in \mathcal{S}} p(s_i = s, \mathbf{y}_{j>i-1}|s_{i-1} = s') \\
&= \sum_{s \in \mathcal{S}} p(s_i = s, \mathbf{y}_{j>i}, \mathbf{y}_i|s_{i-1} = s') \\
&= \sum_{s \in \mathcal{S}} \frac{p(s_i = s, \mathbf{y}_{j>i}, \mathbf{y}_i, s_{i-1} = s')}{p(s_{i-1} = s')} \\
&= \sum_{s \in \mathcal{S}} p(\mathbf{y}_{j>i}|s_i = s, \mathbf{y}_i, s_{i-1} = s') \frac{p(s_i = s, \mathbf{y}_i, s_{i-1} = s')}{p(s_{i-1} = s')} \\
&= \sum_{s \in \mathcal{S}} p(\mathbf{y}_{j>i}|s_i = s) p(s_i = s, \mathbf{y}_i|s_{i-1} = s') \\
&= \sum_{s \in \mathcal{S}} \beta_i(s).\gamma_i(s', s)
\end{aligned}$$

MAP algorithm (8)

- Parameters β are obtained via a backward recursion:

$$\beta_{i-1}(s') = \sum_{s \in \mathcal{S}} \beta_i(s).\gamma_i(s', s)$$

for $(i = 2, \ldots, N+1)$ and $\forall s' \in \mathcal{S}$.

- If trellis termination, the initial conditions are:

$$\beta_N(s_N) = 1 \text{ and } \beta_N(s \neq s_N) = 0$$

$\Leftrightarrow$ The final state is known to be s_N.

- If no trellis termination, the initial conditions are:

$$\beta_N(s) = \frac{1}{\#\mathcal{S}} \ \forall s \in \mathcal{S}$$

$\Leftrightarrow$ The final state is unknown.

October 27, 2005 Newcom Automn School 47

MAP algorithm (9)

- $\gamma_i(s', s)$ associated with a transition between states $s_{i-1} = s'$ and $s_i = s$:

$$\begin{aligned}\gamma_i(s', s) &= p(\mathbf{y}_i, s_i = s | s_{i-1} = s') \\ &= p(\mathbf{y_i} | s_{i-1} = s', s_i = s).P(s_i = s | s_{i-1} = s')\end{aligned}$$

In terms of symbols:

$$\gamma_i(s', s) = p(\mathbf{y_i} | u_i, s_{i-1} = s').P(u_i)$$

$\triangleright$ $p(\mathbf{y_i} | u_i, s_{i-1} = s')$ is evaluated on the basis of the received symbol and the channel type.

$\triangleright$ $P(u_i)$ is evaluated on the basis of the a priori information $L_a(u_i)$.

- $\gamma_i(s', s)$ = metric associated with the transition $(s_{i-1} = s', s_i = s)$.
The same as in MAP sequence estimation and SOVA.

October 27, 2005 Newcom Automn School 48

MAP algorithm: summary

- The MAP algorithm computes the a posteriori LLR $L_p(u_i)$ of the information bits u_i (for $i = 1, \ldots, N$):

$$L_p(u_i) = \ln\left(\frac{\sum_{\mathcal{S}^+} \alpha_{i-1}(s').\gamma_i(s',s).\beta_i(s)}{\sum_{\mathcal{S}^-} \alpha_{i-1}(s').\gamma_i(s',s).\beta_i(s)}\right) \qquad (i = 1, \ldots, N)$$

- $\alpha \Rightarrow$ forward recursion with appropriate initial condition:

$$\alpha_i(s) = \sum_{s' \in \mathcal{S}} \alpha_{i-1}(s').\gamma_i(s',s) \qquad (i = 0, \ldots, N-1; \forall s \in \mathcal{S})$$

- $\beta \Rightarrow$ backward recursion with appropriate initial condition:

$$\beta_{i-1}(s') = \sum_{s \in \mathcal{S}} \beta_i(s).\gamma_i(s',s) \qquad (i = 2, \ldots, N+1; \forall s' \in \mathcal{S})$$

- $\gamma \Rightarrow$ calculated based on the received symbols and the a priori information:

$$\overline{\gamma}_i(s',s) = p(\mathbf{y_i}|s_{i-1} = s', s_i = s).P(s_i = s|s_{i-1} = s') \qquad \forall i; \forall (s',s) \in \text{trellis}$$

MAP algorithm: log MAP

- The MAP algorithm has numerical problems.

$\Rightarrow$ Implementation in the logarithmic domain:

- Define $\overline{\alpha}_i(s) = \ln(\alpha_i(s))$, $\overline{\beta}_i(s) = \ln(\beta_i(s))$ and $\overline{\gamma}_i(s',s) = \ln(\gamma_i(s',s))$.

- The a posteriori LLR becomes:

$$\begin{aligned} L_p(u_i) &= \ln\left(\frac{\sum_{\mathcal{S}^+} \exp(\overline{\alpha}_{i-1}(s')) \,.\, \exp(\overline{\gamma}_i(s',s)) \,.\, \exp(\overline{\beta}_i(s))}{\sum_{\mathcal{S}^-} \exp(\overline{\alpha}_{i-1}(s')) \,.\, \exp(\overline{\gamma}_i(s',s)) \,.\, \exp(\overline{\beta}_i(s))}\right) \\ &= \ln\left(\sum_{\mathcal{S}^+} \exp\left(\overline{\alpha}_{i-1}(s') + \overline{\gamma}_i(s',s) + \overline{\beta}_i(s)\right)\right) \\ &- \ln\left(\sum_{\mathcal{S}^-} \exp\left(\overline{\alpha}_{i-1}(s') + \overline{\gamma}_i(s',s) + \overline{\beta}_i(s)\right)\right) \end{aligned}$$

MAP algorithm: max log MAP

- Using the approximation:

$$\ln(\exp(x) + \exp(y) + \exp(z)) \approx \max(x, y, z)$$

we have

$$\begin{aligned} L_p(u_i) &\approx \max_{\mathcal{S}^+}(\overline{\alpha}_{i-1}(s') + \overline{\gamma}_i(s', s) + \overline{\beta}_i(s)) \\ &- \max_{\mathcal{S}^-}(\overline{\alpha}_{i-1}(s') + \overline{\gamma}_i(s', s) + \overline{\beta}_i(s)) \end{aligned}$$

▷ The forward recursion for parameters $\overline{\alpha}_i(s)$ becomes:

$$\overline{\alpha}_i(s) = \max_{s' \in \mathcal{S}}(\overline{\alpha}_{i-1}(s') + \overline{\gamma}_i(s', s))$$

with initial conditions :

$$\overline{\alpha}_0(s_0) = 0 \text{ and } \overline{\alpha}_0(s \neq s_0) = -\infty$$

MAP algorithm: max log MAP

▷ The backward recursion for parameters $\overline{\beta}_{i-1}(s')$ becomes:

$$\overline{\beta}_{i-1}(s) = \max_{s \in \mathcal{S}}(\overline{\beta}_i(s) + \overline{\gamma}_i(s', s))$$

with initial conditions:

$$\overline{\beta}_N(s_N) = 0 \text{ and } \overline{\beta}_N(s \neq s_N) = -\infty \quad \text{if trellis termination}$$

or initial conditions:

$$\overline{\beta}_N(s) = \ln\left(\frac{1}{\#\mathcal{S}}\right) \; \forall s \in \mathcal{S} \qquad \text{if no trellis termination}$$

▷ Parameter $\overline{\gamma}_i(s', s)$:

$$\overline{\gamma}_i(s', s) = \ln(p(\mathbf{y_i}|s_{i-1} = s', s_i = s)) + \ln(P(s_i = s|s_{i-1} = s'))$$

⇒ Metric calculated for each transition between states $s_{i-1} = s'$ and $s_i = s$ on the basis of the received symbol and the a priori information.

MAP algorithm:log MAP

- An optimal implementation in the logarithmic domain is possible.

$\Rightarrow$ Instead of approximation, use exact expression:

$$\begin{aligned}\ln(\exp(x)+\exp(y)) &= \max(x,y)+\ln(1+\exp(-|x-y|)) \\ &= \overset{*}{\max}(x,y)\end{aligned}$$

If more than two entries:

$$\begin{aligned}\ln(\exp(x)+\exp(y)+\exp(z)) &= \overset{*}{\max}(x,y,z) \\ &= \overset{*}{\max}(\overset{*}{\max}(x,y),z)\end{aligned}$$

$\Rightarrow$ Generalized maximum function.

MAP algorithm:log MAP

$\Rightarrow$ LOG-MAP algorithm.

- Proceeds exactly as the MAX-LOG-MAP algorithms

  if we replace every max function with a $\max^*$ function:

$$\begin{aligned}L_p(u_i) &= \overset{*}{\max_{\mathcal{S}^+}}(\overline{\alpha}_{i-1}(s')+\overline{\gamma}_i(s',s)+\overline{\beta}_i(s)) \\ &- \overset{*}{\max_{\mathcal{S}^-}}(\overline{\alpha}_{i-1}(s')+\overline{\gamma}_i(s',s)+\overline{\beta}_i(s))\end{aligned}$$

- Optimal algorithm!
- Numerical problems solved.
- 2 instances of a generalized VA.
- Complexity $\mathcal{O}(2^K)$ where K is the code constraint length.

Summary of algorithms

- Optimal algorithm: MAP.
  - ▷ Consider all paths in the trellis at each step.

    Divide them into 2 sets at step i.
- Optimal algorithm in the log. domain: LOG-MAP.
  - ▷ Consider all paths in the trellis at each step.

    Divide them into 2 sets at step i.
- Suboptimal algorithm in the log. domain: MAX-LOG-MAP.
  - ▷ Consider 2 paths per step:

    The best with bit 0 and the best with bit 1 at step i

Metric computation (1)

- Particular case: rate-1/2 RSC code.

  Notations already defined.
- Transition metric $\overline{\gamma}_i(s', s) = \ln(\gamma_i(s', s))$

  suited for LOG-MAP, MAX-LOG-MAP and SOVA algorithms.
- Metric $\overline{\gamma}_i(s', s)$ for a transition between states $s_{i-1} = s'$ and $s_i = s$:

$$\begin{aligned} \overline{\gamma}_i(s_{i-1} = s', s_i = s) &= \ln\left(\mathbf{p}(\mathbf{y_i}|s_{i-1} = s', s_i = s)\right) \\ &+ \ln\left(P(s_i = s|s_{i-1} = s')\right) \end{aligned}$$

  or equivalently, in terms of symbols:

$$\begin{aligned} \overline{\gamma}_i(s', s) &= \ln\left(p(\mathbf{y_i}|u_i, s_{i-1} = s')\right) + \ln(P(u_i)) \\ &= \ln(p(\mathbf{y_i}|u_i, u_{i-1}, \ldots, u_{i-M})) + \ln(P(u_i)) \\ &= \ln(p(\mathbf{y_i}|x_i^s, x_i^p)) + \ln(P(u_i)) = \ln(p(\mathbf{y_i}|b_i^s, b_i^p)) + \ln(P(u_i)) \end{aligned}$$

Metric computation (2)

▷ The first term $\ln\left(p(\mathbf{y}_i|b_i^s, b_i^p)\right)$ depends on the received symbols.

Considering an AWGN channel with noise variance σ^2:

$$p(\mathbf{y}_i|b_i^s, b_i^p) = \frac{1}{\sigma^2 2\pi} \exp\left(-\frac{(y_i^s - b_i^s)^2 + (y_i^p - b_i^p)^2}{2\sigma^2}\right)$$

or, in the logarithmic domain:

$$\ln(p(\mathbf{y}_i|b_i^s, b_i^p)) = -\ln(\sigma^2 2\pi) - \frac{(y_i^s - b_i^s)^2 + (y_i^p - b_i^p)^2}{2\sigma^2}$$

which may be developed as:

$$\ln(p(\mathbf{y}_i|b_i^s, b_i^p)) = -\ln(\sigma^2 2\pi) + \frac{y_i^s b_i^s + y_i^p b_i^p}{\sigma^2} - \frac{(y_i^s)^2 + (b_i^s)^2 + (y_i^p)^2 + (b_i^p)^2}{2\sigma^2}$$

Metric computation (3)

▷ The second term $\ln\left(P(u_i)\right)$ is calculated on the basis of the a priori information:

$$L_a(u_i) \approx \ln\left(\frac{P(u_i = 1)}{P(u_i = 0)}\right)$$

We may write:

$$P(u_i) = \begin{cases} \frac{\exp(L_a(u_i))}{1+\exp(L_a(u_i))} & \text{if } u_i = 1 \\ \frac{1}{1+\exp(L_a(u_i))} & \text{if } u_i = 0 \end{cases}$$

or, in the logarithmic domain:

$$\ln\left(P(u_i)\right) = L_a(u_i)u_i - \ln\left(1 + \exp(L_a(u_i))\right)$$

Metric computation (4)

- Combining those two terms, we obtain:

$$\begin{aligned}\overline{\gamma}_i(s',s) = & - \ln(\sigma^2 2\pi) + \frac{y_i^s b_i^s + y_i^p b_i^p}{\sigma^2} - \frac{(y_i^s)^2 + (b_i^s)^2 + (y_i^p)^2 + (b_i^p)^2}{2\sigma^2} \\ & + L_a(u_i)u_i - \ln\left(1 + \exp(L_a(u_i))\right)\end{aligned}$$

- Suppressing the terms common to all hypotheses (terms which do not depend on u_i, b_i^s or b_i^p):

$$\overline{\gamma}_i(s',s) = \frac{y_i^s b_i^s + y_i^p b_i^p}{\sigma^2} + L_a(u_i)u_i$$

- Remembering that $L_c = \frac{2}{\sigma^2}$ for an AWGN channel:

$$\overline{\gamma}_i(s',s) = \frac{1}{2}(L_c y_i^s)b_i^s + \frac{1}{2}(L_c y_i^p)b_i^p + L_a(u_i)u_i$$

Metric computation (5)

- Noting that $b_i^s = 2x_i^s - 1$ and $b_i^p = 2x_i^p - 1$:

$$\overline{\gamma}_i(s',s) = (L_c y_i^s)x_i^s + (L_c y_i^p)x_i^p + L_a(u_i)u_i$$

- Remembering that $x_i^s = u_i$:

$$\overline{\gamma}_i(s',s) = (L_c y_i^s + L_a(u_i))u_i + (L_c y_i^p)x_i^p$$

$\Rightarrow$ For each transition $(s_{i-1} = s', s_i = s)$ in the trellis (characterized by u_i, $x_i^s = u_i$ and x_i^p), we can compute the metric on the basis of the a priori information $L_a(u_i)$ and the soft outputs of the channel $L_c y_i^s$ and $L_c y_i^p$.

Metric : fundamental property

- SISO decoder fundamental property for a rate-1/2 RSC code:

$$L_p(u_i) = L_c y_i^s + L_a(u_i) + L_e(u_i)$$

- Expression of the transition metric:

$$\begin{aligned} \gamma_i(s', s) &= \exp\left(\overline{\gamma}_i(s', s)\right) \\ &= \exp((L_c y_i^s + L_a(u_i))u_i + L_c y_i^p x_i^p) \end{aligned}$$

can be written as:

$$\gamma_i(s', s) = \exp((L_c y_i^s + L_a(u_i))u_i)\gamma_i^e(s', s)$$

with:

$$\gamma_i^e(s', s) = \exp\left(L_c y_i^p x_i^p\right)$$

Metric : fundamental property

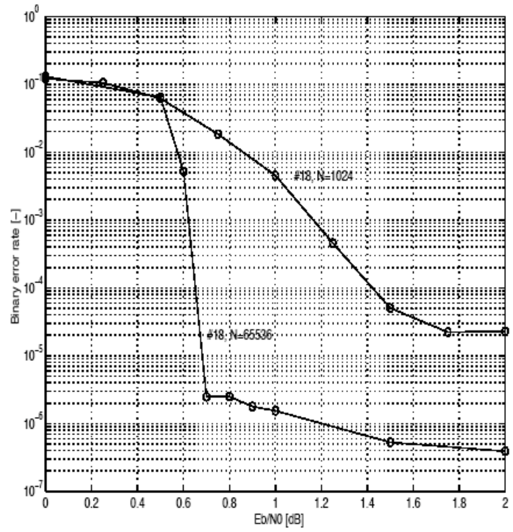
- According to the MAP algorithm:

$$\begin{aligned} L_p(u_i) &= \ln\left(\frac{P(u_i = 1|\mathbf{y})}{P(u_i = 0|\mathbf{y})}\right) \\ &= \ln\left(\frac{\sum_{\mathcal{S}^+} \alpha_{i-1}(s').\gamma_i(s', s).\beta_i(s)}{\sum_{\mathcal{S}^-} \alpha_{i-1}(s').\gamma_i(s', s).\beta_i(s)}\right) \\ &= \ln\left(\frac{\sum_{\mathcal{S}^+} \alpha_{i-1}(s').\exp((L_c y_i^s + L_a(u_i))u_i)\gamma_i^e(s', s).\beta_i(s)}{\sum_{\mathcal{S}^-} \alpha_{i-1}(s').\exp((L_c y_i^s + L_a(u_i))u_i)\gamma_i^e(s', s).\beta_i(s)}\right) \end{aligned}$$

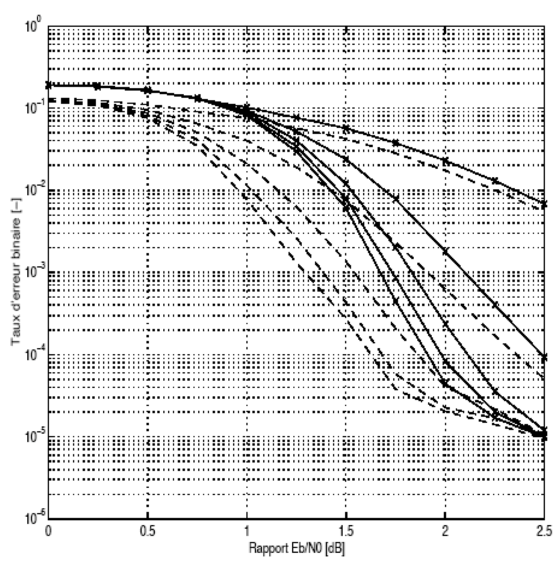
- Factors $\exp((L_c y_i^s + L_a(u_i))u_i)$ identical for all transitions in $\mathcal{S}^+$ and $\mathcal{S}^- \Rightarrow$

$$\begin{aligned} L_p(u_i) &= \ln\left(\frac{\exp((L_c y_i^s + L_a(u_i)).1)\sum_{\mathcal{S}^+} \alpha_{i-1}(s').\gamma_i^e(s', s).\beta_i(s)}{\exp((L_c y_i^s + L_a(u_i)).0)\sum_{\mathcal{S}^-} \alpha_{i-1}(s').\gamma_i^e(s', s).\beta_i(s)}\right) \\ &= L_c y_i^s + L_a(u_i) + \ln\left(\frac{\sum_{\mathcal{S}^+} \alpha_{i-1}(s').\gamma_i^e(s', s).\beta_i(s)}{\sum_{\mathcal{S}^-} \alpha_{i-1}(s').\gamma_i^e(s', s).\beta_i(s)}\right) \\ &= L_c y_i^s + L_a(u_i) + L_e(u_i) \end{aligned}$$

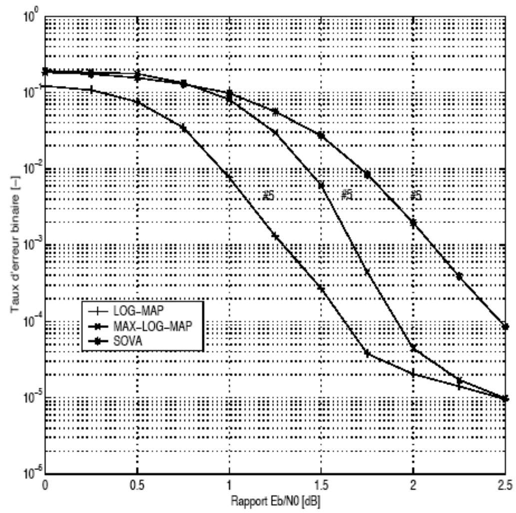
Impact of interleaver



Log MAP vs MAX LOG MAP



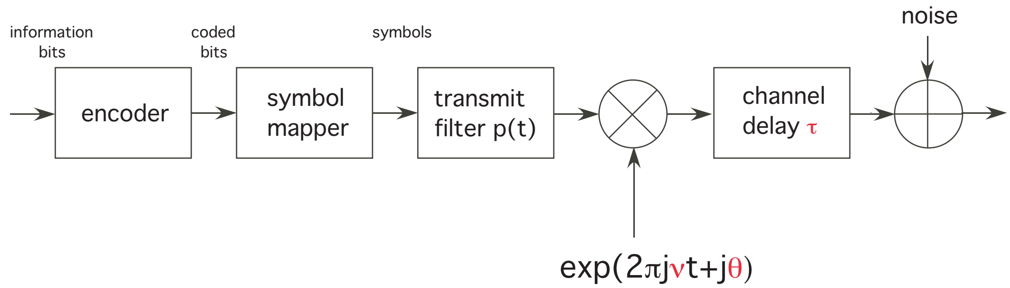
Comparison



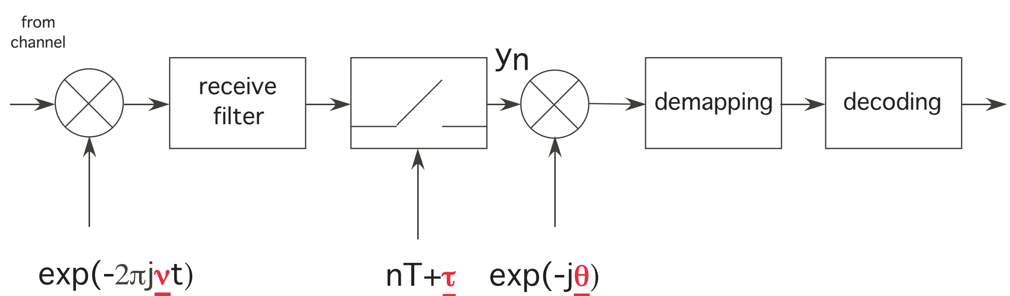
Outline

- Introduction/motivation
- The EM algorithm
- Coding and the MAP algorithm
- **Synchronization of coded systems with the EM algorithm**
- Illustration of performance
- CSI estimation for coded MIMO transmission
- Illustration and performance
- Cramer-Rao bound with coded/prior information

Transmitter setup



Receiver setup



Observation model

- Received signal

$$r(t) = \mathbf{A} \sum_{k=0}^{K-1} a_k \, p(t - kT - \tau) \; e^{j(2\pi\nu t + \theta)} + w(t), \tag{22}$$

- A: amplitude; τ: timing; (ν, θ) carrier frequency and phase offset
- $w(t)$ AWGN
- a_k data symbols

EM algorithm

- It comes

$$\begin{aligned} \ln p(\mathbf{r}|\mathbf{a}, \tilde{\mathbf{b}}) &= -2\tilde{A} \, \mathrm{Re}\{\sum_{k=0}^{K-1} a_k^* \, y_k(\tilde{\nu}, \tilde{\tau}) \, e^{-j\tilde{\theta}}\} \\ &+ \tilde{A}^2 \sum_{k=0}^{K-1} |a_k|^2, \end{aligned} \tag{23}$$

where

$$y_k(\tilde{\nu}, \tilde{\tau}) \triangleq \int_{-\infty}^{+\infty} r(t) \, e^{-j(2\pi\tilde{\nu}t)} \, p(t - kT - \tilde{\tau}) \, dt. \tag{24}$$

Posterior averages

- Expectation step:

$$\begin{aligned} \mathcal{Q}(\tilde{\mathbf{b}}, \hat{\mathbf{b}}^{(n-1)}) &= -2\tilde{A}\ \mathrm{Re}\{\sum_{k=0}^{K-1} \eta_k^*(\mathbf{r}, \hat{\mathbf{b}}^{(n-1)})\, y_k(\tilde{\nu}, \tilde{\tau})\, e^{-j\tilde{\theta}}\} \\ &+ \tilde{A}^2 \sum_{k=0}^{K-1} \rho_k(\mathbf{r}, \hat{\mathbf{b}}^{(n-1)}). \end{aligned} \tag{25}$$

- With following posterior values

$$\begin{aligned} \eta_k(\mathbf{r}, \hat{\mathbf{b}}^{(n-1)}) &\triangleq \int_{\mathbf{a}\in\mathcal{A}^K} a(k)\ p(\mathbf{a}|\mathbf{r}, \hat{\mathbf{b}}^{(n-1)})\, d\mathbf{a} \\ \rho_k(\mathbf{r}, \hat{\mathbf{b}}^{(n-1)}) &\triangleq \int_{\mathbf{a}\in\mathcal{A}^K} |a(k)|^2\, p(\mathbf{a}|\mathbf{r}, \hat{\mathbf{b}}^{(n-1)})\, d\mathbf{a} \end{aligned}$$

- Note: depend on symbol marginal posterior probabilities !

28

EM estimates

- Maximization step leads to partially decoupled solutions [ICC2003]

$$[\hat{\nu}^{(n)}, \hat{\tau}^{(n)}] = \arg\max_{\tilde{\nu},\tilde{\tau}}\{|\sum_{k=0}^{K-1} \eta_{\mathbf{k}}^*(\mathbf{r}, \hat{\mathbf{b}}^{(\mathbf{n-1})})\, y_k(\tilde{\nu}, \tilde{\tau})|\} \tag{26}$$

$$\hat{\theta}^{(n)} = \arg\{\sum_{k=0}^{K-1} \eta_{\mathbf{k}}^*(\mathbf{r}, \hat{\mathbf{b}}^{(\mathbf{n-1})})\, y_k(\hat{\nu}^{(n)}, \hat{\tau}^{(n)})\} \tag{27}$$

$$\hat{A}^{(n)} = \frac{|\sum_{k=0}^{K-1} \eta_{\mathbf{k}}^*(\mathbf{r}, \hat{\mathbf{b}}^{(\mathbf{n-1})})\, y_k(\hat{\nu}^{(n)}, \hat{\tau}^{(n)})|}{\sum_{k=0}^{K-1} \rho_{\mathbf{k}}(\mathbf{r}, \hat{\mathbf{b}}^{(\mathbf{n-1})})}. \tag{28}$$

29

Comparison with pilot aided solution

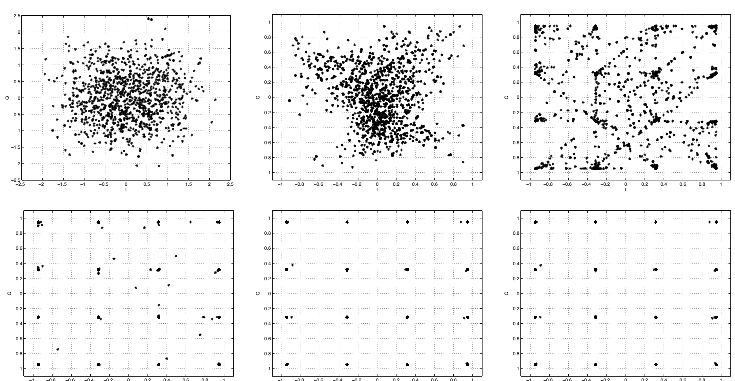
- If pilots had been used

$$[\hat{\nu}, \hat{\tau}] = \arg\max_{\tilde{\nu},\tilde{\tau}} \{ | \sum_{k=0}^{K-1} \mathbf{a}_k^* \, y_k(\tilde{\nu}, \tilde{\tau}) | \} \quad (29)$$

$$\hat{\theta} = \arg\{ \sum_{k=0}^{K-1} \mathbf{a}_k^* \, y_k(\hat{\nu}, \hat{\tau}) \} \quad (30)$$

$$\hat{A}^{(n)} = \frac{| \sum_{k=0}^{K-1} \mathbf{a}_k^* \, y_k(\hat{\nu}, \hat{\tau}) |}{\sum_{k=0}^{K-1} |\mathbf{a}_k|^2}. \quad (31)$$

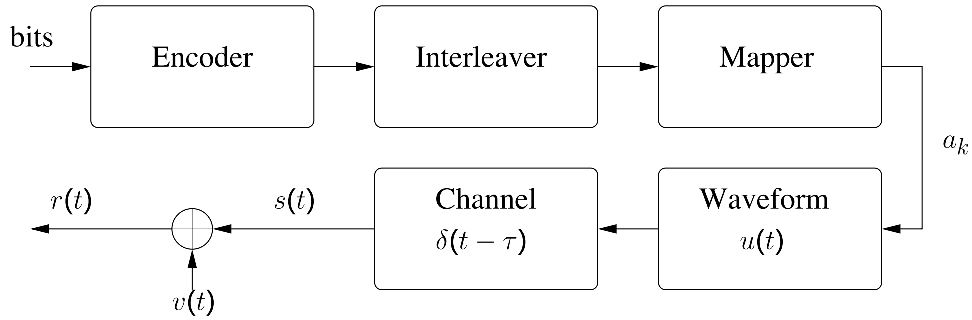
Posterior mean values



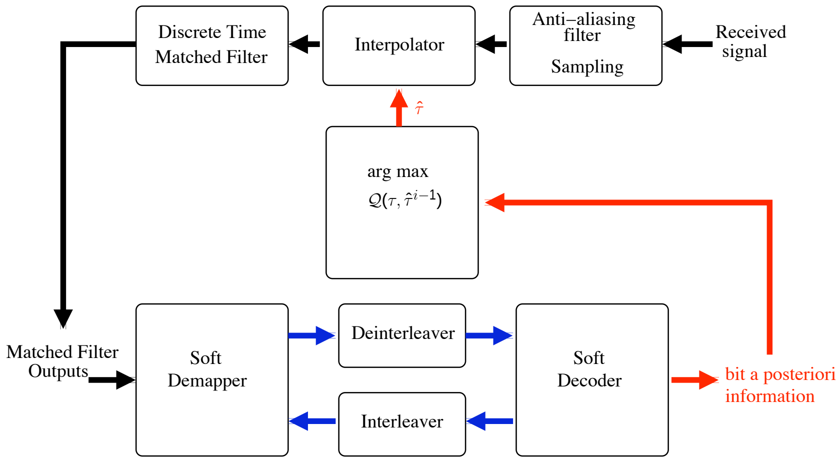
Discussion

- Solution only requires marginal symbol a posteriori probabilities
- Delivered by trellis based MAP module implemented by means of BCJR algorithm (when code or *supercode* not too complex)
- Also available in a turbo receiver after *sufficient* number of iterations

BICM transmitter



BICM iterative demapper/decoder with timing estimation



Discussion

- A turbo receiver is supposed to deliver bit posterior probabilities after an infinite number of iterations
- Approximation: use these bit APPs obtained after one or several iterations to build symbol APPS
- Use them in the EM algorithm

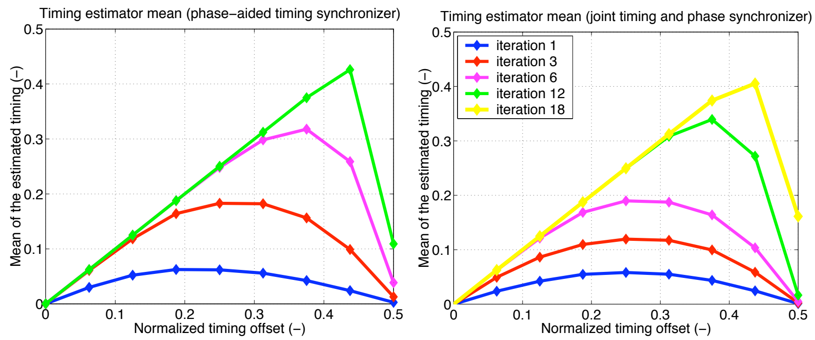
Outline

- Introduction/motivation
- The EM algorithm
- Coding and the MAP algorithm
- Synchronization of coded systems with the EM algorithm
- **Illustration of performance**
- CSI estimation for coded MIMO transmission
- Illustration and performance
- Cramer-Rao bound with coded/prior information

Setup

- 16-QAM, ”medium unconditioned bit-wise mutual information” mapping, convolutional code, length= 3, code rate= 1/2
- Timing only or joint phase/timing estimation
- Startup : $\hat{\tau}^{(0)} = 0$ or $\hat{\tau}^{(0)} = 0, \hat{\theta}^{(0)} = 0$ ($\theta = 15$ degrees)
- $E_b/N_0 = 4$dB
- One turbo iteration per EM iteration (no reset of extrinsic information)

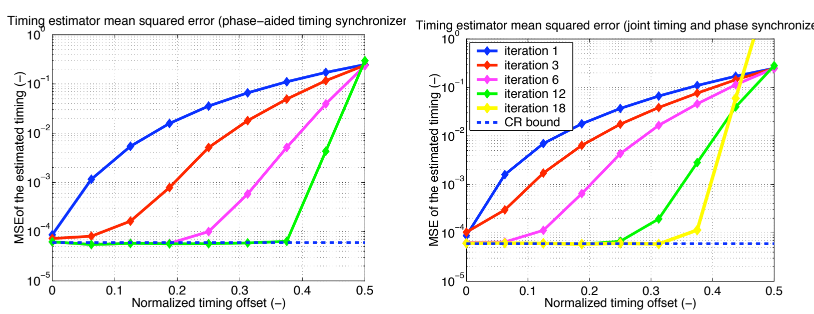
Results: mean



38

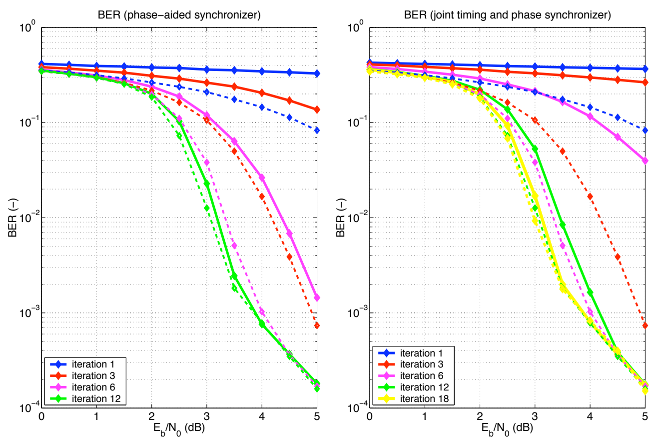
Symbol timing or joint phase/timing estimation

Results: MSE



39

Results: BER ($\tau/T = 0.25$)



Steepest descent implementation

- No closed form solution for the symbol timing
- Steepest descent leads to

$$\hat{\epsilon}^{(n)} \triangleq \hat{\tau}^{(n+1)} - \hat{\tau}^{(n)} = \beta \sum_k |\eta_k^{(n)}| \times \mathrm{Re}\{e^{-j\arg(\eta_k^{(n)})}\dot{y}(kT + \hat{\tau}^{(n)})\} \quad (32)$$

- Proposal: design a best linear unbiased estimator [SPAWC2003]

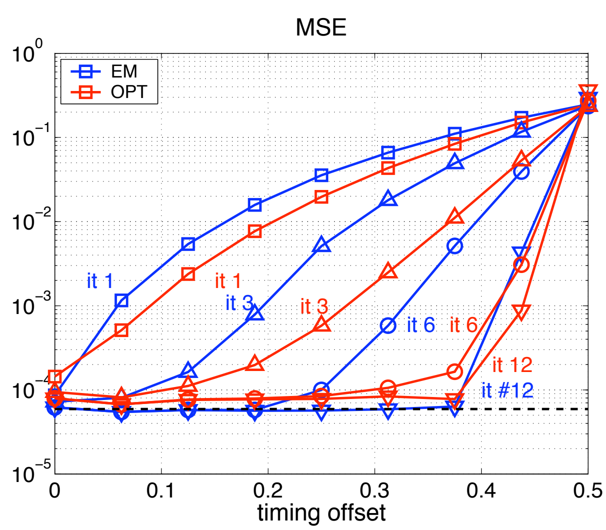
BLUE estimator

- BLUE estimator (with some simplification) leads to

$$\hat{\epsilon}^{(n)} = \beta' \sum_k \frac{\mathsf{E}[h_I(k)]}{\sigma^2_{w_I(k)} + \sigma^2_{e_I(k)}} \times \mathrm{Re}\{ e^{-j\arg(\eta_k^{(n)})} (\dot{y}(kT + \hat{\tau}^{(n)}) - \sum_{\mathbf{k}'} \eta_{\mathbf{k}'}^{(\mathbf{n})} \dot{\mathbf{x}}_{\mathbf{k}-\mathbf{k}'})\}$$
$$+ \beta' \sum_{\mathbf{k}} \frac{\mathsf{E}[\mathbf{h_Q(k)}]}{\sigma^2_{\mathbf{w_Q(k)}} + \sigma^2_{\mathbf{e_Q(k)}}} \times \mathrm{Im}\{ \mathbf{e}^{-\mathbf{j}\arg(\eta_{\mathbf{k}}^{(\mathbf{n})})} (\dot{\mathbf{y}}(\mathbf{kT} + \hat{\tau}^{(\mathbf{n})}) - \sum_{\mathbf{k}'} \eta_{\mathbf{k}'}^{(\mathbf{n})} \dot{\mathbf{x}}_{\mathbf{k}-\mathbf{k}'})\}$$

- Idea: not only projection in phase with $\eta_k^{(n)}$ contains useful information but also that in quadrature (**red** term).
- Also: perform soft interference cancellation of self noise (**blue** term)

Results with improved design



Acquisition

- Does not solve acquisition
- Conventional methods with ambiguity resolution can be used to initialize the EM estimates.
- Or run the EM with different initial values [Wymeersch2004] . Can work without pilots at low SNRs ((M)CRB reached at 1dB).
- Solves convergence towards local minimum.

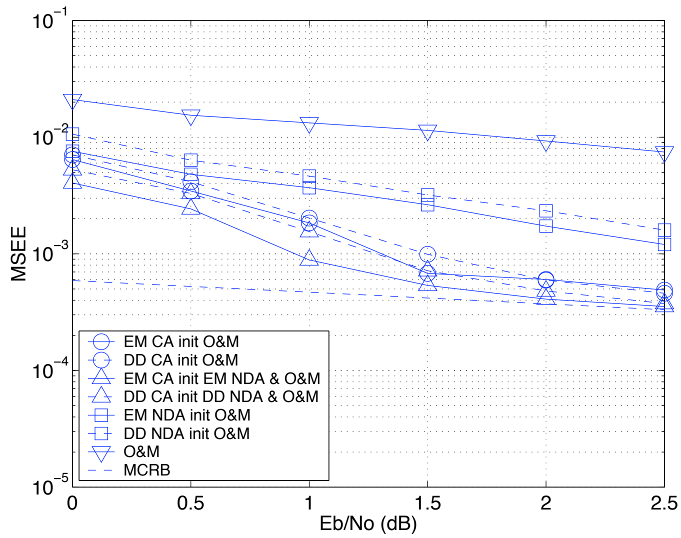
44

Turbo coded system

- 512 BPSK symbols
- Timing changed randomly at each new frame
- MSE and BER with different initial values for the EM

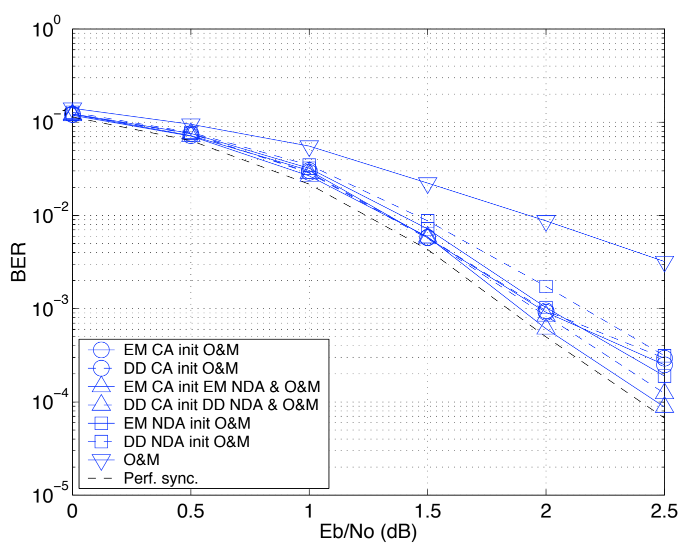
45

MSE results



46

BER results



47

Conclusion

- Soft data aided synchronization works
- Cramér Rao bound can be reached
- Initial value has large impact

48

Outline

- Introduction/motivation
- The EM algorithm
- Coding and the MAP algorithm
- Synchronization of coded systems with the EM algorithm
- Illustration of performance
- **CSI estimation for coded MIMO transmission**
- Illustration and performance
- Cramer-Rao bound with coded/prior information

49

FS MIMO scheme

- FS MIMO channels with n_t transmit and n_r receive antennas
- Observation model for polyphase component m and RX antenna j

$$\underline{r}_m^{(j)} = \underline{\underline{A}}\,\underline{h}_m^{(j)} + \underline{n}_m^{(j)}. \tag{33}$$

- Objective: estimate the $\underline{h}_m^{(j)}$; the symbols $a_i(m)$ are nuisance parameters
- Estimation of noise variance can be handled as well

EM algorithm

- Follow path similar to soft data aided synchronization [Wautelet2003]

$$\ln p(\mathcal{R}|\underline{\underline{A}}, \tilde{\mathcal{B}}) = -\frac{1}{\tilde{\sigma}_n^2} \sum_{j=1}^{n_R} \sum_{m=0}^{M_s-1} (\underline{r}_m^{(j)} - \underline{\underline{A}}\,\underline{\tilde{h}}_m^{(j)})^H (\underline{r}_m^{(j)} - \underline{\underline{A}}\,\underline{\tilde{h}}_m^{(j)}) \tag{34}$$

- Channel estimation at step (n)

$$\underline{\hat{h}}_{m,\mathrm{EM}}^{(j)(n)} = E[\underline{\underline{A}}^H \underline{\underline{A}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}]^{-1} E[\underline{\underline{A}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}]^H \underline{r}_m^{(j)} \tag{35}$$

- Noise-variance estimation

$$\begin{aligned}\hat{\sigma}_{n,\mathrm{EM}}^{2\,(n)} = \frac{1}{n_R M_s L_r} \sum_{j=1}^{n_R} \sum_{m=0}^{M_s-1} \Big[& \underline{r}_m^{(j)\,H} \underline{r}_m^{(j)} + \underline{\hat{h}}_{m,\mathrm{EM}}^{(j)\;H} E[\underline{\underline{A}}^H \underline{\underline{A}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}]\, \underline{\hat{h}}_{m,\mathrm{EM}}^{(j)} \\ & -2\mathrm{Re}\left\{ \underline{r}_m^{(j)\,H} E[\underline{\underline{A}}|\mathcal{R}, \hat{\mathcal{B}}^{(n-1)}]\, \underline{\hat{h}}_{m,\mathrm{EM}}^{(j)} \right\} \Big].\end{aligned}$$

Comparison with pilot aided solution

- Pilot aided solution for the channel

$$\underline{\hat{h}}^{(j)}_{m,\mathrm{DA}} = \left(\underline{\underline{A_p}}^H \underline{\underline{A_p}}\right)^{-1} \underline{\underline{A_p}}^H \underline{r_p}^{(j)}_m. \tag{36}$$

- For the noise-variance (biased):

$$\hat{\sigma}^2_{n,\mathrm{DA}} = \frac{1}{n_R M_s L_r} \sum_{j=1}^{n_R} \sum_{m=0}^{M_s-1} (\underline{r_p}^{(j)}_m - \underline{\underline{A_p}}\, \underline{\hat{h}}^{(i,j)}_{m,\mathrm{DA}})^H (\underline{r_p}^{(j)}_m - \underline{\underline{A_p}}\, \underline{\hat{h}}^{(i,j)}_{m,\mathrm{DA}}). \tag{37}$$

- Biased can be removed
- EM Channel estimation at step (n)

$$\underline{\hat{h}}^{(j)(n)}_{m,\mathrm{EM}} = E[\underline{\underline{A}}^H \underline{\underline{A}} | \mathcal{R}, \hat{\mathcal{B}}^{(n-1)}]^{-1}\, E[\underline{\underline{A}} | \mathcal{R}, \hat{\mathcal{B}}^{(n-1)}]^H\, \underline{r}^{(j)}_m \tag{38}$$

- Posterior averages of products also needed

Problems

- Posterior average of product not delivered by e. g. turbo receivers
- Solution for the channel estimate delivered at each EM iteration is **biased**
  - Degrades the BER
  - Pointed out by [Kobayashi et al.,2001]; ad-hoc solutions proposed
- Solution for the noise variance estimate delivered at each EM iteration is also **biased**: bias can be partly removed.

Proposed solution: BLUE design

- Target Best Linear Unbiased Estimator assuming a priori information for the symbols
- Estimation at step (n):

$$\underline{\hat{h}}_{m,\mathrm{UEM}}^{(j)} = (E[\underline{\underline{A}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}]^H E[\underline{\underline{A}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}])^{-1}\, E[\underline{\underline{A}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}]^H\, \underline{r}_m^{(j)}. \tag{39}$$

Other possibility: ECM

- Expectation Conditional Maximization
- Update one value at a time; take the most recent value for others
- Avoid matrix inversion

$$\hat{h}_{l,m,\mathrm{ECM}}^{(i,j)(n)} = \frac{\{E[\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}]^H\, \underline{r}_m^{(j)}\}_{Li+l} - \{E[\underline{\underline{S}}^H\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}]\, \underline{\tilde{h}}_{l,m}^{(i,j)(n)}\}_{Li+l}}{\{E[\underline{\underline{S}}^H\underline{\underline{S}}|\mathcal{R},\hat{\mathcal{B}}^{(n-1)}]\}_{Li+l,Li+l}}, \tag{40}$$

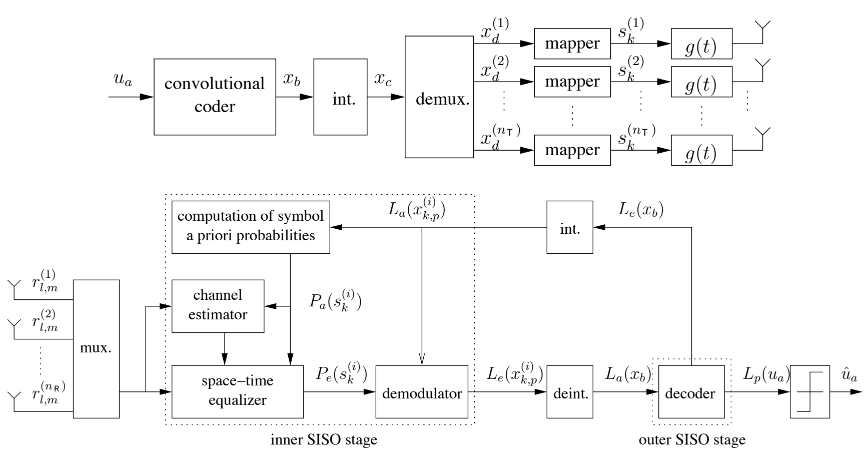
- This solution is also biased and the bias can be removed

Outline

- Introduction/motivation
- The EM algorithm
- Coding and the MAP algorithm
- Synchronization of coded systems with the EM algorithm
- Illustration of performance
- CSI estimation for coded MIMO transmission
- **Illustration and performance**
- Cramer-Rao bound with coded/prior information

56

ST BICM Transmitter and receiver



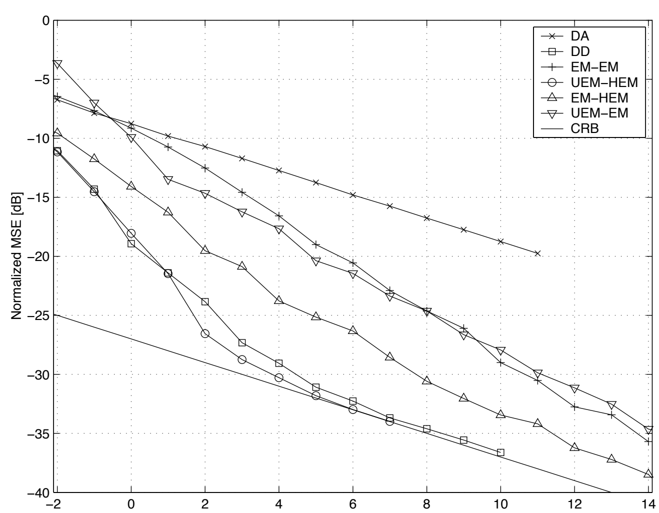
57

Simulation parameters

- Space time BICM [Tonello 2000]
- Random interleaver, 8-PSK, Gray Mapping
- r = 0.5 convolutional encoder, generator polynomials (23,35) (octal)
- frame: 2000 information bits (1336 symboles)
- Flat Rayleigh fading channel 4×4; 4×5 pilot symbols (orthogonal)
- FS GSM Typical Urban 4×4; 4×55 pilot symbols
- Iterative space equalization/demodulation (MMSE filter based) and decoding (BCJR) [Wautelet 2004]
- 6 iterations
- Noise variance estimated in a way similar to CSI

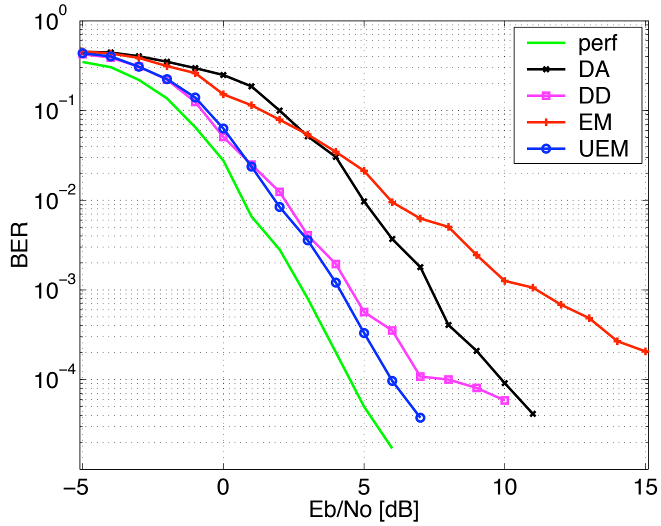
58

Results for Flat $4 * 4$ MIMO@10 it



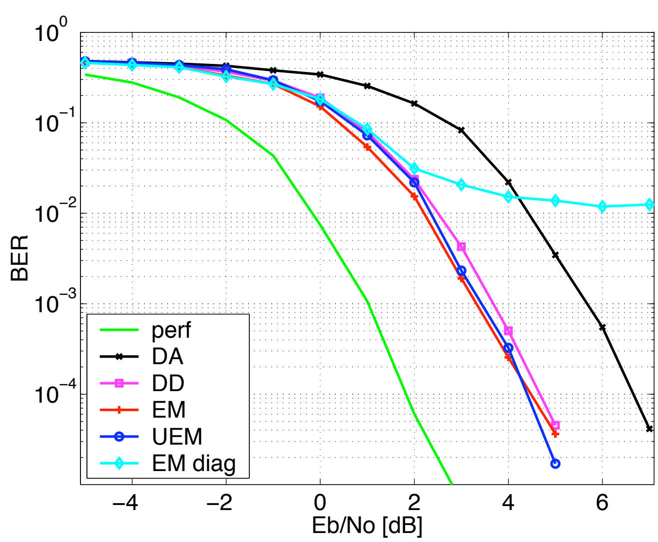
59

Results for Flat $4 * 4$ MIMO



60

Results for FS $4 * 4$ MIMO

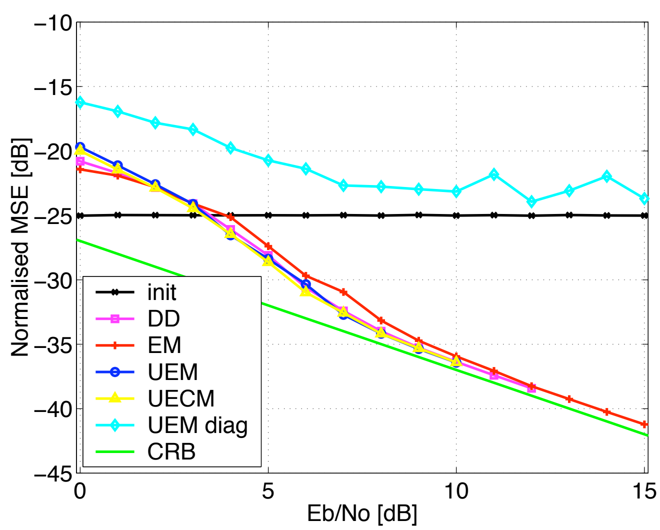


61

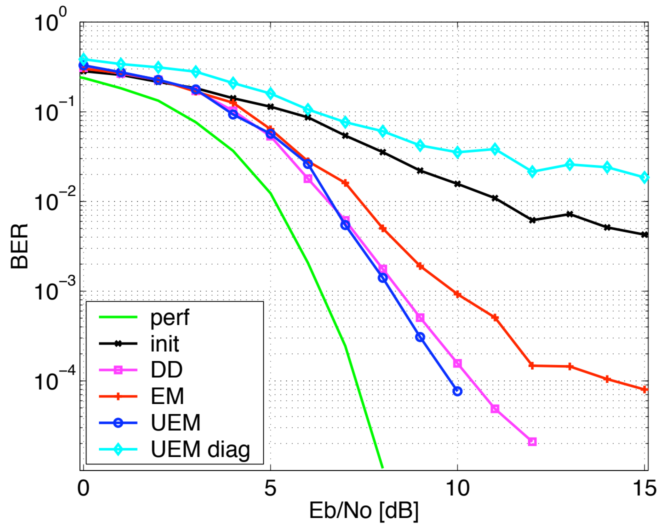
Simulation setup

- Space time BICM, 16-QAM, Gray Mapping
- r = 0.5 convolutional encoder, generator polynomials $[7_8, 5_8]$
- frame: 1001 information symbols
- Initialization with CSI corrupted by noise: normalized MSE of $-25dB$
- FS Hiperlan 2/B channel 2×2

Results for Hiperlan II $2 * 2$ channel



Results for Hiperlan II $2 * 2$ channel



64

Global conclusions

- EM : nice framework for the use of soft information in a synchronization/parameter estimation context
- Improvements have to be introduced wrt pure EM design

65

Outline

- Introduction/motivation
- The EM algorithm
- Coding and the MAP algorithm
- Synchronization of coded systems with the EM algorithm
- Illustration of performance
- CSI estimation for coded MIMO transmission
- Illustration and performance
- **Cramer-Rao bound with coded/prior information**

Cramer-Rao bound

- Channel with n_T inputs and n_R outputs; bursts of $n_T L_s$ complex symbols $s_k^{(i)}$ are sent
- Model:

$$r_k^{(j)} = \sum_{i=1}^{n_T} \sum_{l=0}^{L-1} h_l^{(i,j)} \, s_{k-l}^{(i)} + n_k^{(j)}, \tag{41}$$

- Let

$$\underline{h}_R = [\Re\{\underline{h}\}^T \, \Im\{\underline{h}\}^T]^T. \tag{42}$$

- We have

$$E_{\underline{r}|\underline{h}}[(\hat{\underline{h}}_R - \underline{h}_R)(\hat{\underline{h}}_R - \underline{h}_R)^T] \geq \text{CRB}(\underline{h}_R). \tag{43}$$

$$\text{CRB}(\underline{h}_R) = \underline{\underline{J}}^{-1}(\underline{h}_R). \tag{44}$$

- Fisher Information Matrix

$$\{\underline{\underline{J}}(\underline{h}_R)\}_{l,k} = E_{\underline{r}|\underline{h}_R}\left[\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_l} \frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_k}\right]_{|\tilde{\underline{h}}_R = \underline{h}_R}, \tag{45}$$

Cramer-Rao bound

- With nuisance (data) parameters:

$$p(\underline{r}|\underline{\tilde{h}}_R) = \int p(\underline{r}|\underline{\tilde{h}}_R, \underline{s})\, p(\underline{s})\mathrm{d}\underline{s} \tag{46}$$

- We have

$$\frac{\partial \ln p(\underline{r}|\underline{\tilde{h}}_R)}{\partial\{\underline{\tilde{h}}_R\}_l} = \frac{1}{p(\underline{r}|\underline{\tilde{h}}_R)} \frac{\partial p(\underline{r}|\underline{\tilde{h}}_R)}{\partial\{\underline{\tilde{h}}_R\}_l} \tag{47}$$

- So use the substitution

$$\frac{\partial p(\underline{r}|\underline{\tilde{h}}_R)}{\partial\{\underline{\tilde{h}}_R\}_l} = p(\underline{r}|\underline{\tilde{h}}_R) \frac{\partial \ln p(\underline{r}|\underline{\tilde{h}}_R)}{\partial\{\underline{\tilde{h}}_R\}_l} \tag{48}$$

68

CRAMER-RAO BOUND WITH CODED/PRIOR INFORMATION

Cramer-Rao bound

- With nuisance (data) parameters:

$$\begin{aligned}
\frac{\partial \ln p(\underline{r}|\underline{\tilde{h}}_R)}{\partial\{\underline{\tilde{h}}_R\}_l} &= \frac{1}{p(\underline{r}|\underline{\tilde{h}}_R)} \frac{\partial p(\underline{r}|\underline{\tilde{h}}_R)}{\partial\{\underline{\tilde{h}}_R\}_l} && (49)\\
&= \frac{1}{p(\underline{r}|\underline{\tilde{h}}_R)} \frac{\partial \int p(\underline{r}|\underline{\tilde{h}}_R, \underline{s})\, p(\underline{s})\mathrm{d}\underline{s}}{\partial\{\underline{\tilde{h}}_R\}_l} && (50)\\
&= \frac{1}{p(\underline{r}|\underline{\tilde{h}}_R)} \int p(\underline{s}) \frac{\partial p(\underline{r}|\underline{\tilde{h}}_R, \underline{s})\, \mathrm{d}\underline{s}}{\partial\{\underline{\tilde{h}}_R\}_l} && (51)\\
&= \int \frac{p(\underline{s})p(\underline{r}|\underline{\tilde{h}}_R, \underline{s})}{p(\underline{r}|\underline{\tilde{h}}_R)} \frac{\partial \ln p(\underline{r}|\underline{\tilde{h}}_R, \underline{s})}{\partial\{\underline{\tilde{h}}_R\}_l} \mathrm{d}\underline{s} && (52)\\
&= \int p(\underline{s}|\underline{\tilde{h}}_R, \underline{r}) \frac{\partial \ln p(\underline{r}|\underline{\tilde{h}}_R, \underline{s})}{\partial\{\underline{\tilde{h}}_R\}_l} \mathrm{d}\underline{s} && (53)
\end{aligned}$$

69

Cramer-Rao bound

- The effect of the prior distribution of nuisance parameters $\underline{s}$ is captured through the posterior probability
- This posterior probability $p(\underline{s}|\underline{\tilde{h}}_R, \underline{r})$ is exactly what is delivered by an $\underline{\tilde{h}}_R$-aided MAP receiver
- Basic formula for CRB computation over coded system
- Assumes exact posterior probabilities are delivered: true MAP (turbo ?)

Cramer-Rao bound

- About the partial derivatives

$$\frac{\partial \ln p(\underline{r}|\underline{\tilde{h}}, \underline{s})}{\partial \Re\{\tilde{h}_l^{(i,j)}\}}_{|\underline{\tilde{h}}=\underline{h}} = \frac{2}{\sigma_n^2} \sum_{k=1}^{L_s} \Re\Big\{ s_{k-l}^{(i)*} r_k^{(j)} - \sum_{i'=1}^{n_T} \sum_{l'=0}^{L-1} h_{l'}^{(i',j)} s_{k-l'}^{(i')} s_{k-l}^{(i)*} \Big\}$$

$$\frac{\partial \ln p(\underline{r}|\underline{\tilde{h}}, \underline{s})}{\partial \Im\{\tilde{h}_l^{(i,j)}\}}_{|\underline{\tilde{h}}=\underline{h}} = \frac{2}{\sigma_n^2} \sum_{k=1}^{L_s} \Im\Big\{ s_{k-l}^{(i)*} r_k^{(j)} - \sum_{i'=1}^{n_T} \sum_{l'=0}^{L-1} h_{l'}^{(i',j)} s_{k-l'}^{(i')} s_{k-l}^{(i)*} \Big\}.$$

- Using $\eta_k^{(i)} = E_{\underline{s}|\underline{r},\underline{h}_R}[s_k^{(i)}]$ and $\rho_{k,k'}^{(i,i')} = E_{\underline{s}|\underline{r},\underline{h}_R}[s_k^{(i)}\, s_{k'}^{(i')*}]$, we finally have

$$\frac{\partial \ln p(\underline{r}|\underline{\tilde{h}})}{\partial \Re\{\tilde{h}_l^{(i,j)}\}}_{|\underline{\tilde{h}}=\underline{h}} = \frac{2}{\sigma_n^2} \sum_{k=1}^{L_s} \Re\Big\{ \eta_{k-l}^{(i)*} r_k^{(j)} - \sum_{i'=1}^{n_T} \sum_{l'=0}^{L-1} h_{l'}^{(i',j)} \rho_{k-l',k-l}^{(i',i)} \Big\}$$

$$\frac{\partial \ln p(\underline{r}|\underline{\tilde{h}})}{\partial \Im\{\tilde{h}_l^{(i,j)}\}}_{|\underline{\tilde{h}}=\underline{h}} = \frac{2}{\sigma_n^2} \sum_{k=1}^{L_s} \Im\Big\{ \eta_{k-l}^{(i)*} r_k^{(j)} - \sum_{i'=1}^{n_T} \sum_{l'=0}^{L-1} h_{l'}^{(i',j)} \rho_{k-l',k-l}^{(i',i)} \Big\}.$$

Cramer-Rao bound for given mutual information

- Instead of setting $p(\underline{s})$ for each sequence or symbol, one can instead assume a pdf for the symbol probability
- Usually LLR are gaussian distributed
- One can set the mutual information (MI) between $p(\underline{s})$ and the sequence
- Amounts to fixing the LLR distribution : MI=0 $\leftrightarrow$ NDA; MI=1 $\leftrightarrow$ DA
- For a given MI, one has a lower bound given on the CRB given by

$$E_{\underline{r}|\underline{h},\mathrm{MI}}[(\underline{\hat{h}}_R - \underline{h}_R)(\underline{\hat{h}}_R - \underline{h}_R)^T] \geq E_{p(\underline{s})|\mathrm{MI}}[\underline{\underline{J}}^{-1}(\underline{h}_R)], \tag{54}$$

- With Jensen's inequality for matrices:

$$\begin{aligned} E_{\underline{r}|\underline{h},\mathrm{MI}}[(\underline{\hat{h}}_R - \underline{h}_R)(\underline{\hat{h}}_R - \underline{h}_R)^T] &\geq (E_{p(\underline{s})|\mathrm{MI}}[\underline{\underline{J}}(\underline{h})])^{-1} & (55)\\ &= \mathrm{CRB}_{\mathrm{MI}} & (56) \end{aligned}$$

Cramer-Rao bound for random channel

- For an *estimate unbiased on average* : $E_{\underline{h}_R,\underline{r}}[\underline{\hat{h}}_R] = m_{\underline{h}_R}$,
- Lower bound given by

$$E_{\underline{h},\underline{r}}[(\underline{\hat{h}}_R - \underline{h}_R)(\underline{\hat{h}}_R - \underline{h}_R)^T] \geq \mathrm{CRB}_{\mathrm{Rand}}. \tag{57}$$

- with

$$\mathrm{CRB}_{\mathrm{Rand}} = (E_{\underline{h}_R}[\underline{\underline{J_2}}(\underline{h}_R)])^{-1}, \tag{58}$$

- and $\underline{\underline{J_2}}(\underline{h}_R)$ is a matrix whose elements are

$$\{\underline{\underline{J_2}}(\underline{h}_R)\}_{l,k} = E_{\underline{r}|\underline{h}_R}\left[\frac{\partial \ln p(\underline{r},\underline{\tilde{h}}_R)}{\partial\{\underline{\tilde{h}}_R\}_l}\frac{\partial \ln p(\underline{r},\underline{\tilde{h}}_R)}{\partial\{\underline{\tilde{h}}_R\}_k}\right]_{|\underline{\tilde{h}}_R=\underline{h}_R}. \tag{59}$$

- Valid for estimators knowing the prior channel distribution or the joint pdf $p(\underline{r},\underline{\tilde{h}}_R)$

Cramer-Rao bound for random channel

- For a *conditionally unbiased estimator* : $E_{\underline{r}|\underline{h}_R}[\hat{\underline{h}}_R] = \underline{h}_R$.

$$E_{\underline{r},\underline{h}_R}[(\hat{\underline{h}}_R - \underline{h}_R)(\hat{\underline{h}}_R - \underline{h}_R)^T] \geq \text{CRB}_{\text{CU}}, \quad (60)$$

$$\text{CRB}_{\text{CU}} = E_{\underline{h}_R}[\underline{\underline{J}}^{-1}(\underline{h}_R)]. \quad (61)$$

- J of the "usual" CRB (see 44)
- Averaging over r and channel NOT simultaneous (inversion in between)
- With Jensen's inequality for matrices:

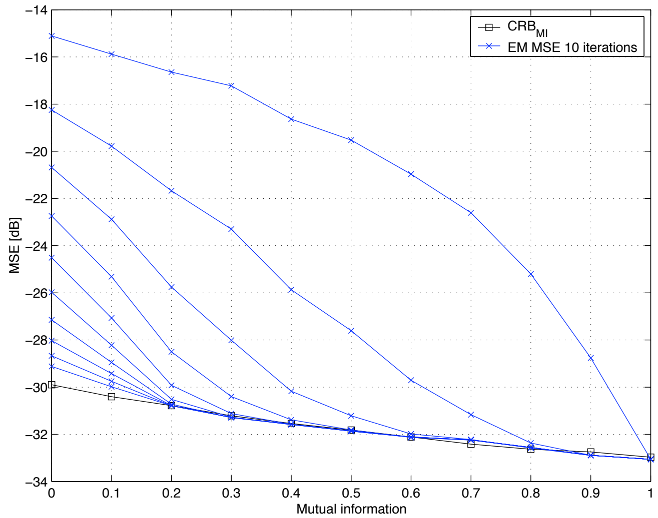
$$\text{CRB}_{\text{CU2}} = (E_{\underline{h}_R}[\underline{\underline{J}}(\underline{h}_R)])^{-1} \quad (62)$$

$$\text{CRB}_{\text{CU2}} \leq \text{CRB}_{\text{CU}}. \quad (63)$$

Results

- Burst sent over Porat channel
- MAP equalizer, no coding, BPSK
- E_s/N_0 = 0 dB
- CRB decreases with increasing MI (means closer to DA mode)
- Result also for EM estimation: achieves the CRB after 10 iterations

Results for Porat channel



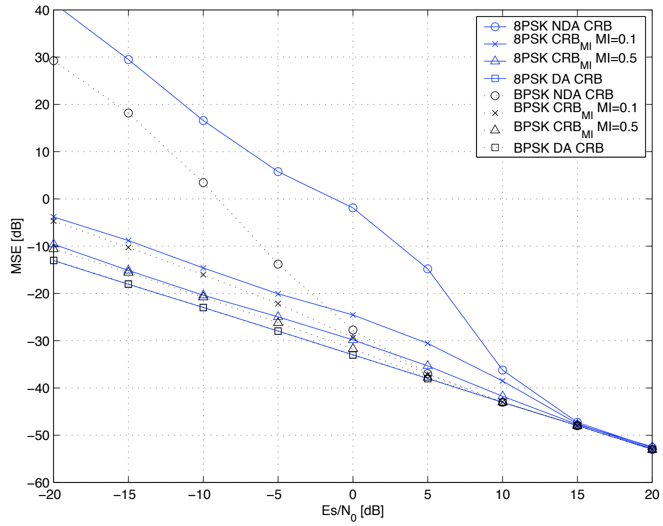
76

Results for di□erent constellations

- SISO Proakis B channel
- All bounds converge to the DA CRB at high E_s/N_0
- For MI=0.1, smaller constellation better: less uncertainty about symbols for low E_s/N_0
- For large MI information brought by constellation less crucial
- All same DA CRB

77

Results for Proakis B



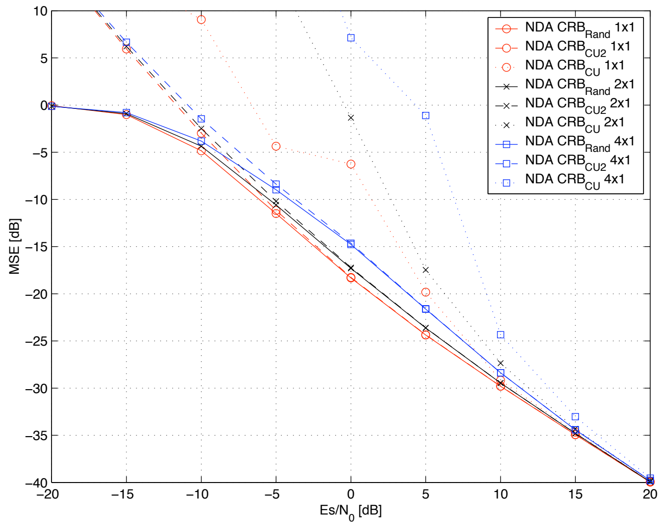
78

Results for random channels

- Flat Rayleigh fading
- 1,2 or 4 TX antennas
- All NDA: MI=0
- Benefitial knowledge of channel distribution for low E_s/N_0
- Degradation with increasing number of antennas: less information about data (more interference)

79

Results for MISO flat Rayleigh



80

Thank you !

81

Parameter estimation for the Alamouti scheme: impact of diversity on "estimability"

L. Vandendorpe (UCL)

Thanks to J. Louveaux

1

Outline

- Introduction/motivation
- Alamouti scheme
- CRB and nuisance parameters
- Results for Alamouti

2

Outline

- Introduction/motivation
- Alamouti scheme
- CRB and nuisance parameters
- Results for Alamouti

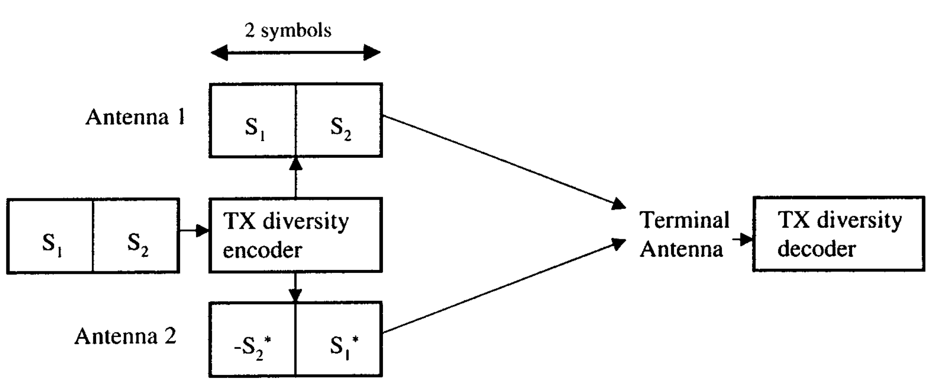
Motivation

- Alamouti benefits from order 2 diversity
- Effect known for detection: slope of BER curve changes accordingly
- What about sensitivity to synchronisation errors ?
- Does diversity impact the sensitivity and the CRB ?

Outline

- Introduction/motivation
- Alamouti scheme
- CRB and nuisance parameters
- Results for Alamouti

Transmitter



Model

- Transmitted signal (baseband)

$$x_0(t) = \sum_{n=0}^{N-1} [s_0(n)u(t-2nT) - s_1^*(n)u(t-2nT-T)] \quad (1)$$
$$x_1(t) = \sum_{n=0}^{N-1} [s_1(n)u(t-2nT) + s_0^*(n)u(t-2nT-T)] \quad (2)$$

- Received signal

$$\begin{aligned} r(t) &= h_0 \sum_{n=0}^{N-1} [s_0(n)u(t-2nT-\tau) - s_1^*(n)u(t-2nT-T-\tau)] \\ &+ h_1 \sum_{n=0}^{N-1} [s_1(n)u(t-2nT-\tau) + s_0^*(n)u(t-2nT-T-\tau)] \\ &+ n(t) \quad (3) \end{aligned}$$

Question

- h_0, h_1 are both complex circular gaussian (Rayleigh fading)
- What is the impact on the "estimability" of τ
- To be compared with a non diversity situation

Transmitter

- Transmitted signal

$$x(t) = \sum_{n=0}^{N-1} s(n)u(t - nT - \tau) \quad (4)$$

- Received signal

$$r(t) = h \sum_{n=0}^{N-1} s(n)u(t - nT - \tau) + n(t) \quad (5)$$

- with

$$s(n) = s_r(n) + js_i(n) \quad (6)$$

Likelihood function

- Assuming h

$$p[r;\tau|h_r,h_i] = C\ \exp[\int_{-\infty}^{\infty} -|r(t) - h \sum_{n=0}^{N-1} s(n)u(t - nT - \tau)|^2/2N_0] \quad (7)$$

- After expansion/simplification

$$p'[r;\tau|h_r,h_i] = C\ \exp[h_r\, A_r/N_0 + h_i\, A_i/N_0]\ \exp[-(h_r^2 + h_i^2)B/2N_0] \quad (8)$$

$$A_r = \sum_n [s_r(n)y_r(n) + s_i(n)y_i(n)] \quad (9)$$

$$A_i = \sum_n [s_r(n)y_i(n) - s_i(n)y_r(n)] \quad (10)$$

$$B = \sum_n |s(n)|^2 \quad (11)$$

$$y(n) = y_r(n) + jy_i(n) = \int_{-\infty}^{\infty} r(t)u^*(t - nT - \tau)\, \mathrm{d}t \quad (12)$$

Outline

- Introduction/motivation
- Alamouti scheme
- CRB and nuisance parameters
- Results for Alamouti

Cramér Rao bound

- The CRB: (for any unbiased estimator):

$$\sigma_{\hat{\tau}}^2 \geq \frac{1}{-\mathsf{E}\left[\frac{\partial^2 \ln p[r;\tau]}{\partial \tau^2}\right]} \tag{13}$$

- where $\mathsf{E}[.]$ means is expectation wrt to $p[r;\tau]$
- How to handle h, or a nuisance parameter ?
- 4 possible cases

CRB for case 1: joint estimation

- If nothing is known about h, should be estimated together with τ
- Compute the Fisher information matrix J with ($\underline{\theta}^T = [\tau, h_r, h_i]$)

$$J_{i,j} = \mathsf{E}\left[\frac{\partial \ln p[r;\underline{\theta}]}{\partial \theta_i}\frac{\partial \ln p[r;\underline{\theta}]}{\partial \theta_j}\right] = -\mathsf{E}\left[\frac{\partial^2 \ln p[r;\underline{\theta}]}{\partial \theta_i \partial \theta_j}\right] \tag{14}$$

$$\sigma^2_{\hat{\theta}_i} \geq \left[J^{-1}\right]_{ii} \tag{15}$$

- where $\mathsf{E}[.]$ means is expectation wrt to $p[r;\underline{\theta}]$
- Not interesting here: we want the effect of the distribution of h

CRB for case 2: nuisance parameter

- h has to be "removed" in the likelihood function
- Situation comparable with the symbols
- Called "nuisance parameters"
- "Proper" handling of nuisance
- Averaging over h

$$\begin{aligned} p[r;\tau] &= \int_{h_r} \mathrm{d}h_r \int_{h_i} \mathrm{d}h_i \, T_{h_r,h_i}(h_r,h_i)\, p[r;\tau|h_r,h_i] && (16)\\ &= C' \exp[\alpha^2 \sum_n |s^*(n)y(n)|^2] && (17)\end{aligned}$$

$$\alpha^2 = \frac{1}{2N_0^2}\left[\frac{1}{\sigma_h^2} + \frac{\sum_n |s(n)|^2}{N_0}\right]^{-1} \tag{18}$$

CRB for case 2: nuisance parameter (cont'd)

- This corresponds to a "non-h-aided solution"; for any estimator that does not use the knowledge (estimation) of h
- The CRB: (for any unbiased estimator):

$$\sigma^2_{\hat{\tau}} \geq \frac{1}{-\mathsf{E}\left[\frac{\partial^2 \ln p[r;\tau]}{\partial \tau^2}\right]} \tag{19}$$

- where $\mathsf{E}[.]$ means is expectation wrt to $p[r;\tau]$

CRB for case 3: h aided solution

- Assume h is known and compute the h-aided CRB for τ

$$\sigma^2_{h,\hat{\tau}} \geq \frac{1}{-\mathsf{E}\left[\frac{\partial^2 \ln p[r;\tau,h_r,h_i]}{\partial \tau^2}\right]} \tag{20}$$

- Then compute the average of this CRB over the statistics of h

$$\sigma^2_{MCB,\hat{\tau}} = \int_{h_r} \mathrm{d}h_r \int_{h_i} \mathrm{d}h_i \, T_{h_r,h_i}(h_r,h_i) \frac{1}{-\mathsf{E}\left[\frac{\partial^2 \ln p[r;\tau,h_r,h_i]}{\partial \tau^2}\right]} \tag{21}$$

CRB for case 4: bound modified wrt h

- Compute $p[r;\tau,h_r,h_i]$
- Compute

$$\sigma^2_{m,\hat{\tau}} \geq \frac{1}{-\mathsf{E}_{r,h_r,h_i}\left[\dfrac{\partial^2 \ln p[r;\tau,h_r,h_i]}{\partial \tau^2}\right]} \tag{22}$$

- where $\mathsf{E}_{r,h_r,h_i}[.]$ means expectation wrt to both r and h

Outline

- Introduction/motivation
- Alamouti scheme
- CRB and nuisance parameters
- Results for Alamouti

Discussion

- Cases 2 and 4: same solution for Alamouti or non Alamouti !
- If normalization such that identical number of symbols, and total emitted power
- Value for MCRB:

$$\left(\frac{E_s}{N_0}\right)^{-1} \frac{1}{N_{na} W_s^2} \tag{23}$$

$$\bar{E}_s = 2\sigma_h^2 \sigma_s^2 \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega\, |U(\omega)|^2 \tag{24}$$

$$W_s^2 = \frac{\int_{-\infty}^{\infty} \mathrm{d}\omega\, \omega^2\, |U(\omega)|^2}{\int_{-\infty}^{\infty} \mathrm{d}\omega\, |U(\omega)|^2} \tag{25}$$

Cramér Rao bounds

Discussion

- Apparently: no benefit from diversity when non h aided solution
- Is this logical ? Yes
- One should remember that the detector providing diversity IS h aided
- A non-h aided detector would maximize (see above)

$$p[r;\tau] = C' \exp[\alpha^2 \sum_n |s^*(n)y(n)|^2] \tag{26}$$

- Something similar for non h aided Alamouti detection
- So the diversity in detection is measured by considering the h aided detector and then average the $BER(h)$ over the statistics of h
- One should "mimic" this for estimation

h-aided Alamouti detector

- Detection structure:

$$\hat{s}_0(n') = h_0^* \int_{-\infty}^{\infty} r(t)u(t-2n'T)\mathrm{d}t + h_1^* \left[\int_{-\infty}^{\infty} r(t)u(t-2n'T-T)\mathrm{d}t\right]^* \quad (27)$$

$$\hat{s}_1(n') = h_1^* \int_{-\infty}^{\infty} r(t)u(t-2n'T)\mathrm{d}t - h_0 \left[\int_{-\infty}^{\infty} r(t)u(t-2n'T-T)\mathrm{d}t\right]^* \quad (28)$$

- Structure of decision variables

$$\hat{s}_0(n') = \left[|h_0|^2 + |h_1|^2\right] s_0(n') + h_0^*\nu_0(n) + h_1\nu_1^*(n) \quad (29)$$

$$\hat{s}_1(n') = \left[|h_0|^2 + |h_1|^2\right] s_1(n') + h_1^*\nu_0(n) - h_0\nu_1^*(n) \quad (30)$$

Impact of diversity on error bound

- For Q-QAM modulation, symbol error bounded by :

$$P_b < 2\,(1 - \frac{1}{\sqrt{Q}})\,\exp^{-\frac{3\,SNR}{2(Q-1)}} \quad (31)$$

- Averaging over the SNR distribution normalized such that the average received energy is constant, it comes for Alamouti

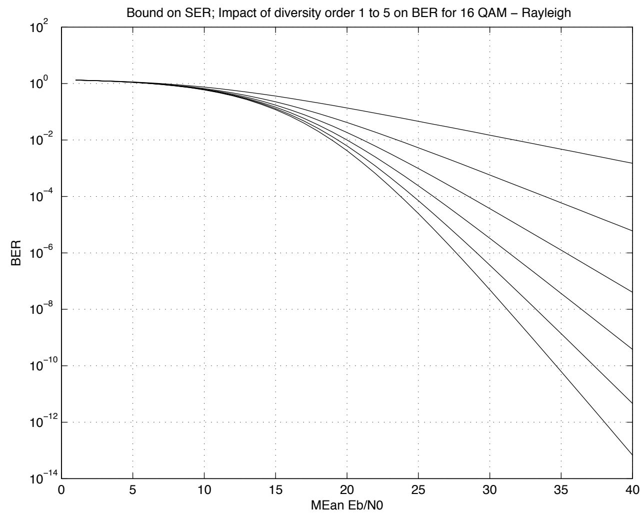
$$\bar{P}_b < 2\,(1 - \frac{1}{\sqrt{Q}})\left[\frac{0.75}{Q-1}\frac{\bar{E}_s}{N_0} + 1\right]^{-2} \quad (32)$$

- For non Alamouti

$$\bar{P}_b < 2\,(1 - \frac{1}{\sqrt{Q}})\left[\frac{1.5}{Q-1}\frac{\bar{E}_s}{N_0} + 1\right]^{-1} \quad (33)$$

- where $\bar{E}_s$ is the average received energy per branch in the non-Alamouti case
- slope of the SER determined by diversity order: this is how diversity materializes !

Illustration for $Q = 16$-QAM and Rayleigh channels



Case 3 Non Alamouti

- Bound for given h_0:

$$\left(\frac{E_s}{N_0}\right)^{-1} \frac{1}{N_{na} W_s^2 \, |h_0|^2 / 2\sigma_h^2} \tag{34}$$

- $|h_0|^2$ is χ^2 with 2 degrees of freedom
- for $u = |h_0|^2 / 2\sigma_h^2$,

$$T(u) = \exp^{-u} \text{ and } \int_0^\infty u^{-1} \exp^{-u} \mathrm{d}u = \infty \tag{35}$$

- Average of h-aided bound is infinite

Case 3 Alamouti

- Bound for given h_0, h_1:

$$\left(\frac{E_s}{N_0}\right)^{-1} \frac{4}{N_{na}\, W_s^2\, (|h_0|^2 + |h_1|^2)/\sigma_h^2} \tag{36}$$

- $|h_0|^2 + |h_1|^2$ is χ^2 with 4 degrees of freedom
- for $u = (|h_0|^2 + |h_1|^2)/\sigma_h^2$,

$$T(u) = 0.25\, u\, \exp^{-u/2} \text{ and } \int_0^\infty u^{-1}\, 0.25\, u\, \exp^{-u/2} \mathrm{d}u = 0.5 \tag{37}$$

- Average of h-aided bound is finite and given by

$$\left(\frac{E_s}{N_0}\right)^{-1} \frac{2}{N_{na}\, W_s^2} \tag{38}$$

Thank you !

Mobile Localisation

Karim Abed-Meraim

ENST-PARIS, TSI department TSI
abed@tsi.enst.fr

Outline

- Generalities.
- Mobile localisation using time of arrival.
- Mobile localisation using angle of arrival.
- Conclusion.

Generalities

Introduction

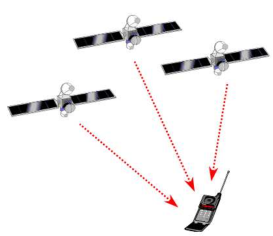
- **Objective:** Find the mobile position (x, y) in a cellular network.
- **Interest:**
  - Localisation services: Emergency, hotels, close restaurants, ...
  - Trafic Localisation, navigation, ...
- **Possible approaches:**
  - Use of GPS (satellite) system.
  - Terrestrial base station (BS) based localization: (Focus on the mobile localization in UMTS-FDD).
  - Hybrid solutions (GPS + BS).

Introduction: Some history...

- GPS is the first localization system (operational since 1991). Developped by US army mainly for military applications and navigation aid.

- New requirement by the FCC (federal communications commission) for all mobile operators to provide a localisation service for emergencies (911 service):
  - *Phase 1*: Localization with a precision ≤ 125m in 67% of the cases.
  - *Phase 2*: Localization with a precision ≤ 300m in 99% of the cases.

GPS



- Advantage : high precision.
- Drawbacks :
  - Requires the visibility by at least 3 satellites.
  - Generation of new mobiles : extra cost for the mobile operators.
  - Heavy initialization system.

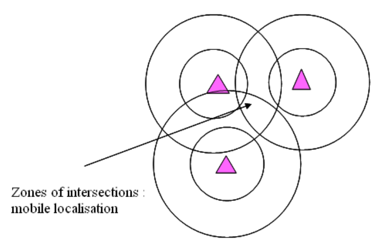
Localisation techniques

1. **Distance measures ⇒ at least 3 base stations (BSs)** :
   - Power measure : exists in the standard.
   - Time of arrival (TOA) : Synchronisation of the BSs.

2. **Angle of arrival (AOA) ⇒ at least 2 BSs** :
   - Installation of multi-sensor antennae : up-link.

3. **Angle-distance measure ⇒ 1 BS** :
   - AOA + distance measure (in the near field case).

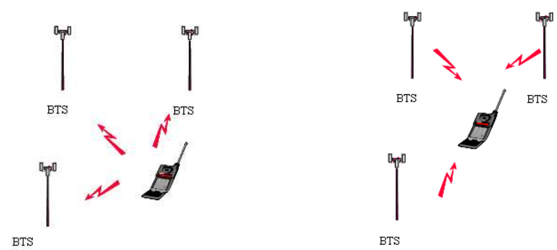
Timing Advance (TA)



- TA: Proportional to the propagation time between the BS and the mobile.
- Quantification with 6 bits of the TA ⇒ precision error of about 500m!!!
- Triangulation with the TA and al least 3 BS.

TOA / OTD (1)

Localisation via time delays



- Necessitates the synchronisation of the BSs.
- Necessitates the use of at least 3 BSs.

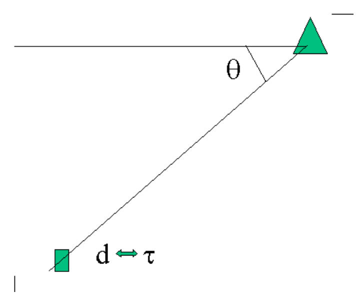
TOA / OTD (2)

- Time Of Arrival
  - Installation of heavy and expensive equipments at the BSs.
  - sensitive to multi-paths.
- Observed Time Difference
  - Certain improvement over the previous technique (signals are synchronized in the down-link).
  - Drawbacks:
    * Generation of new mobiles.
    * sensitive to multi-paths.

Power measures

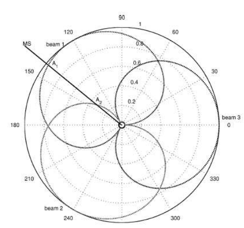
- $P_r = P_e(\frac{\lambda}{4\pi d})^{\alpha}$.
- Advantages:
  - Exists already in the standard.
  - Triangulation possible with more than 3 BSs. de base.
- Drawbacks:
  - Very sensitive to the received power model (tough modelization problem!!).

Angles and delays



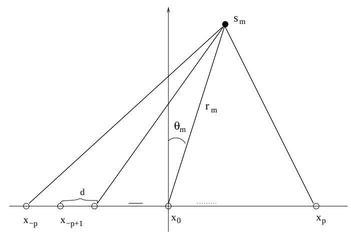
- Avantages :
  - Localisation is possible with only 1 BS (if synchronization).
- Drawbacks :
  - Requires multiple receivers (antenna array) at the BS.
  - Even more sensitive to multipaths effect.

Powers and delays



- Avantages :
  - Localisation is possible with the existing BSs that use ‘sectorial’ sensors.
- Drawbacks :
  - Very sensitive to the received model power.
  - Requires time synchronization of the BS with the mobiles.

Angles and ranges (Near Field)



- Avantages:
  - Localisation is possible with only 1 BS (without synchronization).
- Drawbacks:
  - Applicable only in the near-field!!

Differences between GSM & UMTS

- Advantages in favor of UMTS:
  - Better time resolution due to the oversampling w.r.t symbol duration.
  - Frequency re-use factor equal to 1: Mobile seen by neighboring cells.
- Advantages en faveur du GSM :
  - Relatively reduced multi-paths effect.

Preliminary results for the GSM

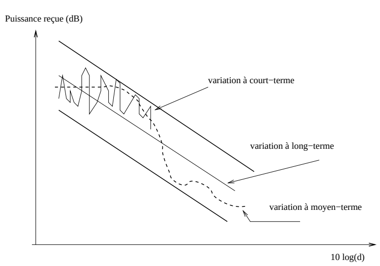
Power measure	140 meters (experiment realised in Paris)
Timing Advance	550 meters
OTA/TOA	110 meters
GPS	5 to 10 meters
Angle of arrival	$\approx$ 100 meters

Limiting factors in UMTS-FDD

- **Estimation accuracy:** An error of one chip period $T_c \Rightarrow$ an error of 73m.
- **Hearing problem (particular to l'UMTS-FDD):** communication between the mobile and the far-located BSs.
  - First considered solutions:
    - ∗ Down-link: use of ldle periods.
    - ∗ Up-link: $\nearrow$ mobile power.
      $\Longrightarrow$ Reduces the system capacity and the mobile autonomy.
- **Non-line of sight (NLOS) problem:**
  - Considered solutions: Use redundant measures and perform selection.

Mobile Localization in UMTS-FDD Using OTD (Down-link)

Signal Model



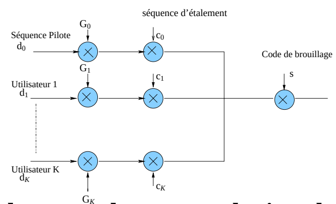
Tap	Retard relatif (ns)	Puissance relative σ_i (dB)	Spectr de Doppler $S(f)$
1	100	-3.2	CLASS
2	200	-5	CLASS
3	500	-4.5	CLASS
4	600	-3.6	CLASS
5	850	-3.9	CLASS
6	900	0.0	CLASS
7	1050	-3.0	CLASS
8	1350	-1.2	CLASS
9	1450	-5.0	CLASS
10	1500	-3.5	CLASS

$$h(t,\tau) = \sum_{i=1}^{M} h_i(t)\delta(t-\tau_i(t))$$

Channel model in a macro-cellular environment proposed by CODIT

The down link

- **Why chosing the down-link:**
  - A high power pilot existing during all the transmission period.
  - Transmitted signals are synchronized.
- **Signal transmitted by the BS:**



- **Propagation channel assumed constant during the slot period l:**

$$h^l(t) = \sum_{r=1}^{R} \beta_{r,l} g(t - \tau_r)$$

Estimation of TOAs

- **Principle (RAKE estimator):**
  - Estimation of $\hat{h}_l(k)$: Correlation between the l-th slot received signal and the shifted version of the pilot signal.
  - TOAs Estimation: Averaging over L slots.

$$\hat{h}(k) = \frac{1}{L} \sum_{l=1}^{L} |\hat{h}_l(k)|$$

- **Estimation accuracy:** $T_c/2$
  Refining the accuracy:
  - By oversampling.
  - By using high resolution methods.
- **Floor effect:** RAKE estimator is not robust against interferences.

Hearing problem

- **Objective:** Improve the robustness of channel estimate against interferences especially for far-located BSs.
- **Difficulty:** The mobile does not know the other user's signatures.
- **Proposed solutions:**
  - Projection of the channel estimate onto the principal subspace of its covariance matrix $\mathbf{\Gamma}$ (RAKE-SP).

    $$\mathbf{h}_l = \mathbf{U}\mathbf{g}_l$$

    where $\mathbf{U}$ *represents the matrix of principal eigenvectors of* $\mathbf{\Gamma}$.
  - Remove (substract) the pilot signal of the serving BS to estimate the channels of far-located BSs.

    $$\tilde{x}_l(i) = x_l(i) - \hat{p}_l^1(i)$$

High resolution (MUSIC) algorithm

- : Estimation of the channel covariance matrix

$$\hat{\mathbf{\Gamma}} = \frac{1}{J}\sum_{j=1}^{J} \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^H \longrightarrow_{J\to\infty} \mathbf{A}(\tau)\mathbf{G}\mathbf{A}(\tau)^H + \sigma_0 \mathbf{R}_0$$

- Estimation of the generalized eigenvectors of $\hat{\mathbf{\Gamma}}$:

$$\hat{\mathbf{\Lambda}}\mathbf{e}_i = \lambda_i \mathbf{R}_0 \mathbf{e}_i$$

- Delay estimation by minimising:

$$v(\tau) = \frac{\mathbf{r}_\tau \mathbf{r}_\tau^H}{\mathbf{r}_\tau \mathbf{E}\mathbf{E}^H \mathbf{r}_\tau^H}$$

where $\mathbf{E}$ represents the matrix of noise eigenvectors of $\mathbf{\Lambda}$ and $\mathbf{r}_\tau$ is the pilot signal autocorrelation vector evaluated for a time lag τ.

Discussion

- MUSIC allows a better estimation of the time delay (see simulation results).
- However, MUSIC is relatively expensive $\Rightarrow$ especially for the down-link (limited mobile power).
- One should reduce its complexity (size of vector $\mathbf{h}$) by using a windowing around the first peak of the RAKE $\Rightarrow$ Two step procedure where MUSIC represents the 'refinement' step.

Triangulation with more than 3 BSs

- Relation between the TOAs and the mobile position (x, y) :

$$\hat{t}_i = \frac{\sqrt{(x-x_i)^2 + (y-y_i)^2}}{c} + t_0 + w_i$$

  t_0= temps de référence et w_i= bruit d'estimation.

- System resolution:
  - Solving the system in the least squares sence (non-linear equations).
  - Explicit solution (after linearization):

$$\begin{pmatrix} c^2(t_2^2 - t_1^2) \\ \vdots \\ c^2(t_I^2 - t_1^2) \end{pmatrix} = -2 \begin{pmatrix} x_{2,1} & y_{2,1} & c(t_2 - t_1) \\ \vdots & \vdots & \vdots \\ x_{I,1} & y_{I,1} & c(t_I - t_1) \end{pmatrix} \begin{pmatrix} x \\ y \\ t_0 \end{pmatrix} + \begin{pmatrix} K_2 - K_1 \\ \vdots \\ K_C - K_1 \end{pmatrix}$$

Triangulation: Chan's method

- If the number of BS is 3:
  - One solves w.r.t. r_1 :

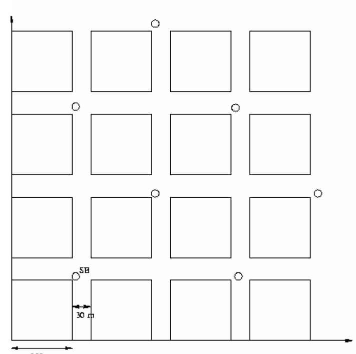
$$\begin{pmatrix} x \\ y \end{pmatrix} = - \begin{pmatrix} x_{2,1} & y_{2,1} \\ x_{3,1} & y_{3,1} \end{pmatrix} \left[\begin{pmatrix} r_{2,1} \\ r_{3,1} \end{pmatrix} r_1 + \frac{1}{2} \begin{pmatrix} r_{2,1}^2 - K_2 + K_1 \\ r_{3,1}^2 - K_3 + K_1 \end{pmatrix} \right]$$

  - Then, we solve a second order polynomial equation in r_1 :

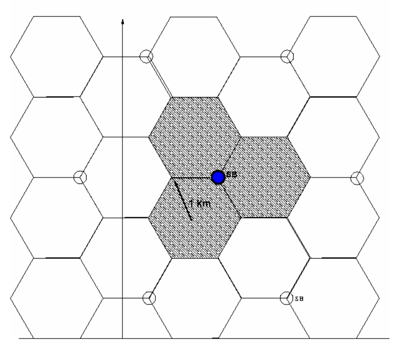
$$r_1^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} 2x_1 & 2y_1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_1 & y_1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

- Among the 2 possible solutions, one choses the one withing the area covered by the serving BS.

MICRO & MACRO cells)



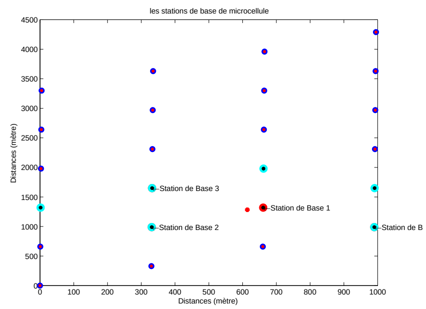
(g) Micro-cell (Manhattan)



(h) Macro-cell

Simulation

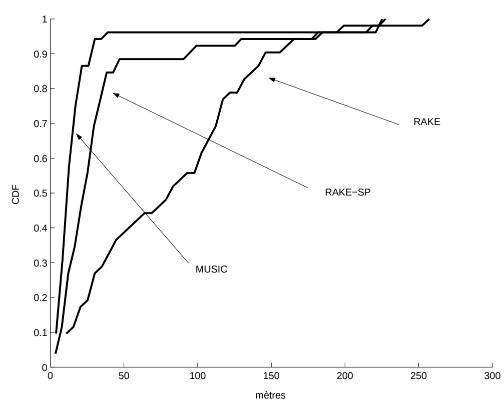
- Simulation in a micro-cell environment (Manhattan).
- Three paths per channel, triangulation with 4 BSs.
- Additif noise representing 10% of the total received power of the furthest BS.
- Loose power control (the ratio between the maximal and minimal powers is ≤ 10).



Micro-cell environment

RAKE-SP

- Comparison of the performance obtained by MUSIC, RAKE-SP and RAKE.



Random position of the mobile, 30 users, $L = 240$ slots.

Dealing with NLOS

- **Proposed solution:** Selection of the 3 'most coherent' TOA measures (we assume mobile hearing by more than 3 BSs).
- **Coherence criterion:**
  - Coherence of the estimated position $M_{i,j,l}(t_i, t_j, t_l)$ (using BSs i, j and l) with TOA t_k assuming a time reference t_0 known:

$$\xi^k_{i,j,l}(t_0) = \| \sqrt{(x_{i,j,l} - x_k)^2 + (y_{i,j,l} - y_k)^2} - c(t_k - t_0) \|^2$$

    Minimisation of $\xi^k_{i,j,l}$ over all possible choices of i, j, l

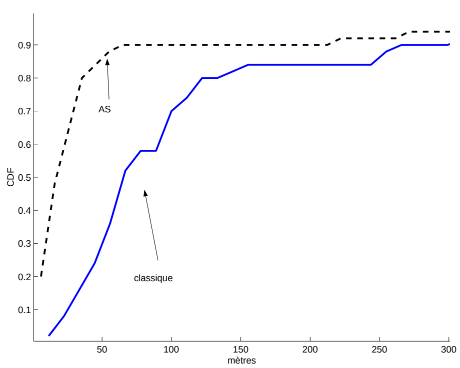
$$\hat{i}, \hat{j}, \hat{l} = \arg \min_{i,j,l,k} \xi^k_{i,j,l}(t_0)$$

  - The time reference t_0 being unknown (one minimizes numerically):

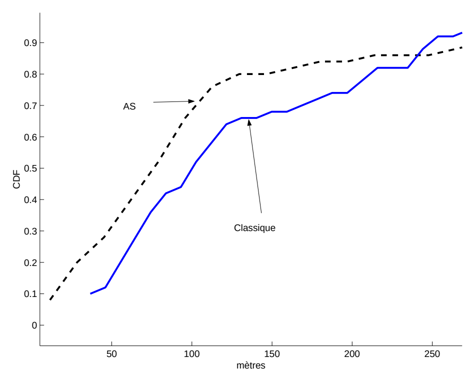
$$\hat{i}, \hat{j}, \hat{l}, \hat{t_0} = \arg \min_{i,j,l,k,t_0} \xi^k_{i,j,l}(t_0)$$

Algorithm of selection (AS)

random position of the mobile at each run, $L = 120$ slots, 8 BSs, K=15



(i) TOAs estimated by MUSIC, NLOS on the 2nd BS



(j) TOAs estimated by RAKE-SP, NLOS on the 2nd BS

Mobile Localisation Using

Angle of Arrival (Up-Link)

Estimation of the AOA

- Requires at least two sensors $\Rightarrow$ Applicable in the uplink.
- Possible with existing BSs but poor estimation accuracy.
- Estimation using ‘smart antennae’ $\Rightarrow$ array processing for source localization.

Array Processing: Basic Concepts

Objectives

- Signal processing extracts information from measured signals.
- Array signal processing uses a group of sensors:
  - Signal enhancement / noise reduction.
    - Coherence adding.
    - Spatial filtering.
  - Source / channel characterizations :
    - number of sources.
    - location 'direction finding'.
    - waveforms 'information from the sources'.

Applications

- Wireless communications.
- Interference mitigation.
- Radar / Sonar.
- Biomedical.
- Speech.
- Seismic.
- ....

Coherent adding

- Let us have an array of M sensors $(m = 1, \cdots, M)$:

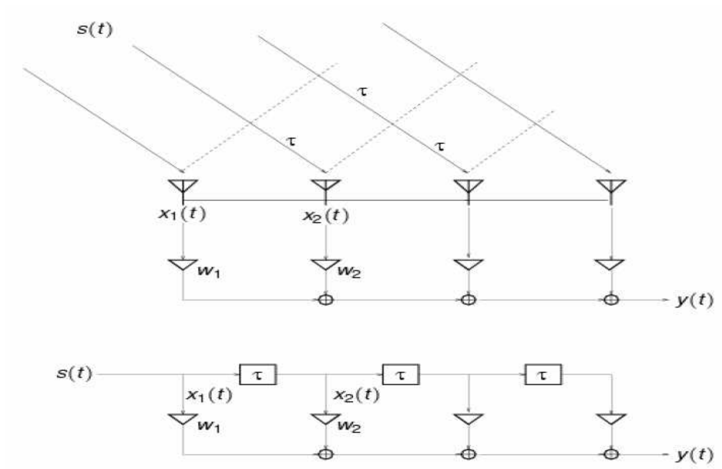
$$x_m(t) = s(t) + n_m(t), \qquad \text{noise variance } \sigma^2$$

- If the noise on the antennas is uncorrelated, then

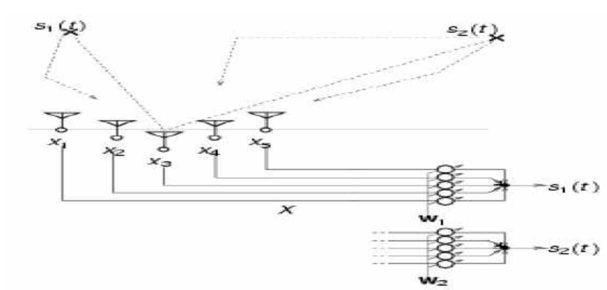
$$y(t) = \frac{1}{M}\sum_{m=1}^{M} x_m(t) = s(t) + \frac{1}{M}\sum_{m=1}^{M} n_m(t), \qquad \text{noise variance } \frac{1}{M}\sigma^2$$

Hence the noise power is reduced by a factor M.

Spatial filtering



Spatial filtering



- Cancelling out interferers : Source separation
  - Classical beamforming requires known ’look directions’, or a reference signal.
  - Blind beamforming : no a priori direction information. Relies on structural properties.

Data model

Baseband signal

- An antenna receives a real valued bandpass signal with center frequency f_c,

$$z(t) = \Re\{s(t)e^{j2\pi f_c t}\} = x(t)\cos(2\pi f_c t) - y(t)\sin(2\pi f_c t)$$

- The baseband signal is

$$s(t) = x(t) + jy(t)$$

  It is the complex envelope of $z(t)$

- $s(t)$ is recovered from $z(t)$ by demodulation : multiplying the received signal with $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ followed by low pass filtering.

Data model

Small delays of narrow band signals

- Recall $z(t) = \Re\{s(t)e^{j2\pi f_c t}\}$. We investigate the effect of small delays of $z(t)$ on the baseband signal $s(t)$

$$z_\tau(t) \triangleq z(t-\tau) = \Re\{s(t-\tau)e^{-j2\pi f_c \tau}e^{j2\pi f_c t}\}$$

- The complex envelope of the delayed signal is

$$s_\tau(t) = s(t-\tau)e^{-j2\pi f_c \tau}$$

Data model

Small delays of narrow band signals

- Let W be the bandwidth of $s(t)$. If $e^{-j2\pi f\tau} \approx 1$ for all frequencies $|f| \leq \frac{W}{2}$, then

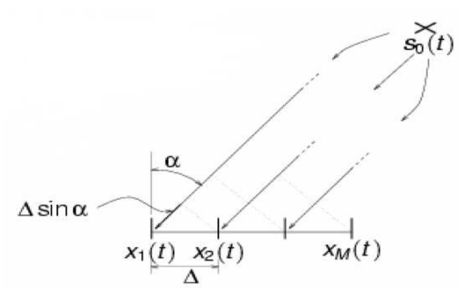
$$s(t-\tau) = \int_{-\frac{W}{2}}^{\frac{W}{2}} S(f) e^{j2\pi f(t-\tau)} df \approx \int_{-\frac{W}{2}}^{\frac{W}{2}} S(f) e^{j2\pi f t} df = s(t)$$

For narrowband signals, time delays shorter than the inverse bandwidth amount to phase shifts of the complex envelope.

Data model

Antenna array response

$$\tau = \frac{L}{c} = \frac{d\sin(\alpha)}{c} = \frac{\lambda\Delta\sin(\alpha)}{c} = \frac{\Delta\sin(\alpha)}{f_c}$$



- Far field assumption : planar waves.
- $a(\alpha)$: Antenna gain pattern.

Data model

Antenna array response

- Let $s(t)$ be the baseband signal at the first antenna : $x_1(t) = a(\alpha)s(t)$
- The signal received by x_2 at a distance of Δ wavelengths experiences an addition delay τ.
- If τ is small compared to the inverse bandwidth of $s(t)$, then

$$s_\tau(t) = s(t)e^{-j2\pi\Delta\sin(\alpha)}$$

- Collect the received signals into a vector $\mathbf{x}(t)$:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} e^{-j2\pi\Delta_1\sin(\alpha)} \\ \vdots \\ e^{-j2\pi\Delta_M\sin(\alpha)} \end{bmatrix} a(\alpha)s(t) = \mathbf{a}(\alpha)s(t)$$

$\mathbf{a}(\alpha)$ is the array response vector. For uniform linear array $\Delta_k = (k-1)\Delta$.

Data model

Array manifold

$$\mathbf{x}(t) = \mathbf{a}(\alpha)s(t)$$

- The array manifold :

$$\mathbf{\Omega} = \{\mathbf{a}(\alpha) : -\pi \leq \alpha \leq \pi\}$$

- The knowledge of $\mathbf{\Omega}$ allows direction finding (i.e. determine α from $\mathbf{x}$).

Spatial Localisation

- Find the number and positions of the sources.
- Sweep all space directions using beamforming
  - Matched filter $\Rightarrow$ Bartelett's method.
  - MVDR $\Rightarrow$ Capon's method.
- Exploit the data model & covariance matrix structure
  - MUSIC (subspace) algorithm
  - ESPRIT algorithm.

Bartlett's method

- Estimate the covariance and sweep all angles

$$\varphi(\theta) = E(|y(t)|^2) = \mathbf{w}^H \mathbf{R} \mathbf{w}$$

- Sum-delay (matched filter) beamforming

$$\mathbf{w} = \frac{\mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{a}(\theta)} \Rightarrow \varphi(\theta) = \frac{\mathbf{a}(\theta)^H \mathbf{R} \mathbf{a}(\theta)}{(\mathbf{a}(\theta)^H \mathbf{a}(\theta))^2}$$

- For a uniform linear array (ULA)

$$\varphi(\theta) = \frac{1}{N^2} \mathbf{a}(\theta)^H \mathbf{R} \mathbf{a}(\theta)$$

Computation using Fourier transform

- Development of the quadratic transform

$$\varphi(\theta) = \mathbf{a}(\theta)^H \mathbf{R} \mathbf{a}(\theta) = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \alpha_n^* R_{nm} \alpha_m$$

- For ULA

$$\alpha_n = (e^{-j2\pi\nu_\theta})^n \Rightarrow \varphi(\theta) = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} R_{nm} (e^{-j2\pi\nu_\theta})^{n-m}$$

- Fourier transform

$$\varphi(\theta) = \sum_{q=-N+1}^{N-1} (e^{-j2\pi\nu_\theta})^q \sum_{n=\max(0,q)}^{N-1+\min(0,q)} R_{n,n-q}$$

Capon's method (MVDR)

- Sweep all angle positions with the MVDR spatial filter

$$\mathbf{w} = \frac{\mathbf{R}^{-1}\mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{R}^{-1} \mathbf{a}(\theta)}$$

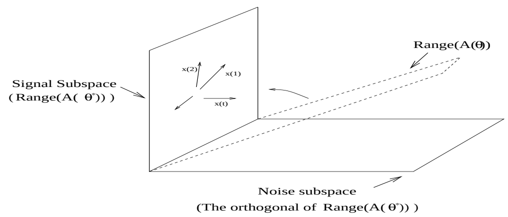
- The localisation function becomes

$$\varphi(\theta) = \frac{1}{\mathbf{a}(\theta)^H \mathbf{R}^{-1} \mathbf{a}(\theta)}$$

car $\varphi(\theta) = \mathbf{w}^H \mathbf{R} \mathbf{w} = \frac{\mathbf{a}(\theta)\mathbf{R}^{-1}}{\mathbf{a}(\theta)^H \mathbf{R}^{-1} \mathbf{a}(\theta)} \mathbf{R} \frac{\mathbf{R}^{-1}\mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{R}^{-1} \mathbf{a}(\theta)}$

- Can be computed using Fourier transform but with $\mathbf{R}^{-1}$ instead of $\mathbf{R}$.

MUSIC (subspace) method



Principle: Assume the following model: $\mathbf{x}(n) = \mathbf{A}(\theta)\mathbf{s}(n)$ with

$$\text{Range}(\mathbf{A}(\theta)) = \text{Range}(\mathbf{A}(\theta')) \Longleftrightarrow \theta = \theta'$$

Thus, θ can be estimated as:

$$\hat{\theta} = \arg\min_{\theta} \text{d}(\text{Range}\{\mathbf{x}(n)\},\ \text{Range}(\mathbf{A}(\theta)))$$

MUSIC

- Estimate the signal (resp. noise) subspace as the principal (resp. minor) eigen-subspace of the data covariance matrix $\mathbf{R}_x$:

$$\mathbf{R}_x = \sum_n \mathbf{x}(n)\mathbf{x}^H(n) = [\mathbf{E}_s\ \mathbf{E}_n] \begin{bmatrix} \mathbf{\Lambda}_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}_s^H \\ \mathbf{E}_n^H \end{bmatrix}$$

where $\quad \text{Range}(\mathbf{E}_s) = \text{Range}(A(\theta)) \perp \text{Range}(\mathbf{E}_n).$

- Orthogonal relation still valid if additive white noise.

MUSIC

- The source angle locations are estimated by minimizing:

$$\min_{\theta} \mathbf{a}(\theta)^H \mathbf{E}_n \mathbf{E}_n^H \mathbf{a}(\theta)$$

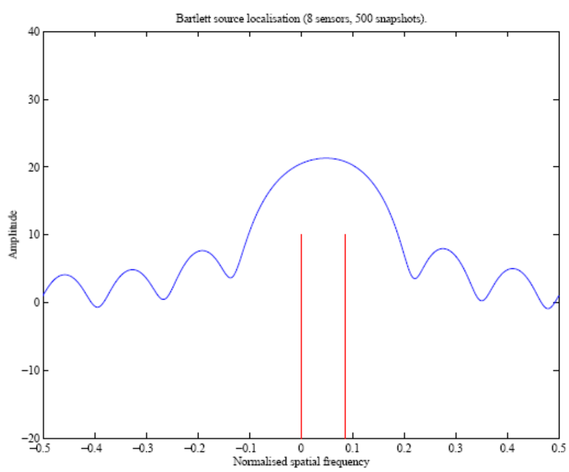
- Or equivalently by maximizing the MUSIC localisation function

$$\varphi(\theta) = \frac{1}{\mathbf{a}(\theta)^H \mathbf{E}_n \mathbf{E}_n^H \mathbf{a}(\theta)}$$

The P sources locations correspond to the P maximas of the above function.

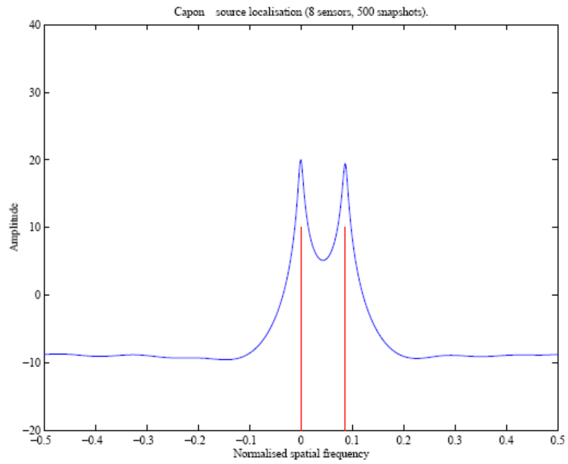
Example (1)

- Bartlett's method (SNR = 20dB)



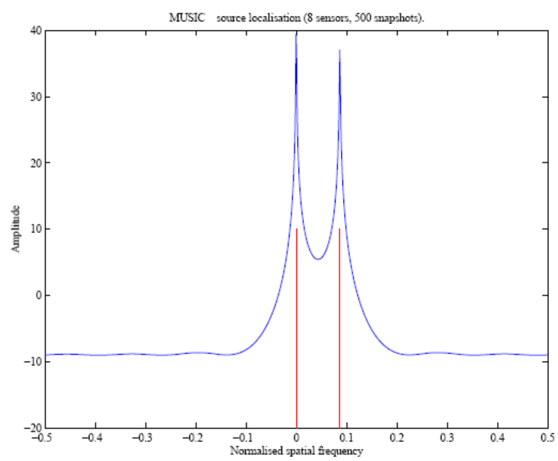
Example (2)

- Capon's method (SNR = 20dB)



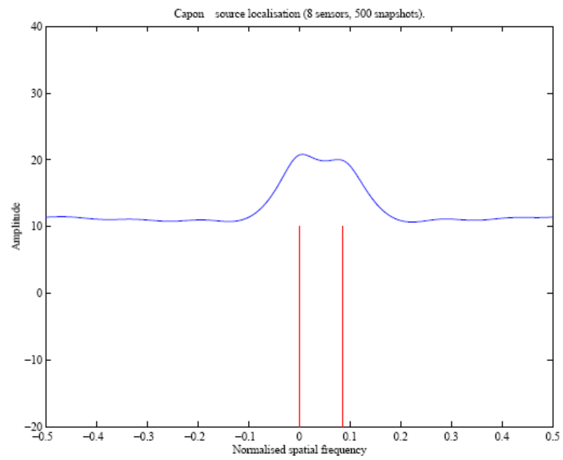
Example (3)

- MUSIC method (SNR = 20dB)



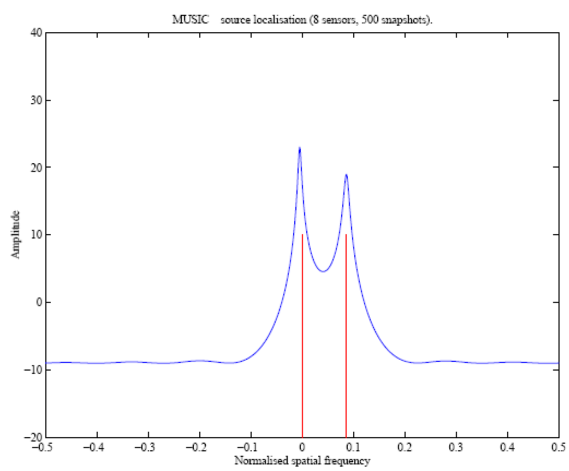
Example (4)

- Capon's method (SNR = 0dB)



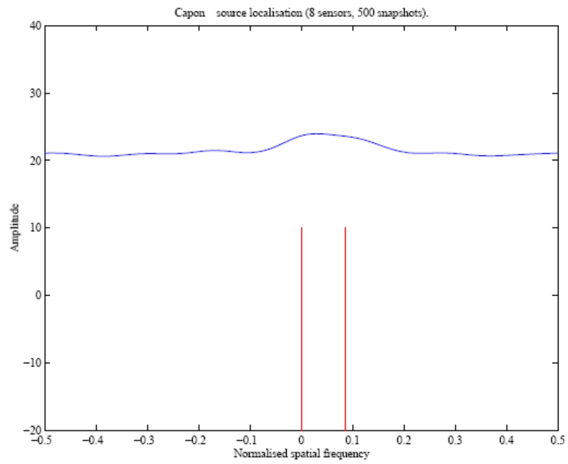
Example (5)

- MUSIC method (SNR = 0dB)



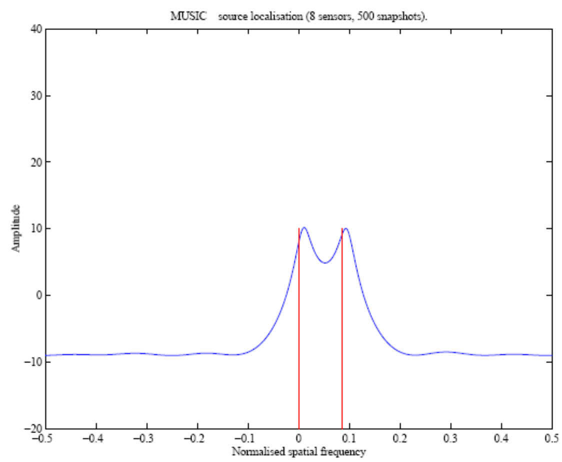
Example (6)

- Capon's method (SNR = -10dB)



Example (7)

- MUSIC method (SNR = -10dB)



ESPRIT Method

- Consider a ULA
- Structure of the directional vector

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 \\ e^{-j2\pi f \frac{d}{C} \sin\theta} \\ \vdots \\ e^{-j2\pi f (N-1) \frac{d}{C} \sin\theta} \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-j2\pi \nu_\theta} \\ \vdots \\ (e^{-j2\pi\nu_\theta})^{N-1} \end{bmatrix}$$

- By removing the first or the last entry of this vector, one obtains two linearly dependent subvectors of $\mathbf{a}(\theta)$.

Rotational invariance

- Oo the directional vector

$$\mathbf{a}(\theta) = \begin{bmatrix} \mathbf{a}_1(\theta) \\ \text{row } N \end{bmatrix} = \begin{bmatrix} \text{row } 1 \\ \mathbf{a}_2(\theta) \end{bmatrix} \Rightarrow \mathbf{a}_2(\theta) = \mathbf{a}_1(\theta) e^{-j2\pi\nu_\theta}$$

- On matrix $\mathbf{A}$

$$\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_P)]$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \text{row } N \end{bmatrix} = \begin{bmatrix} \text{row } 1 \\ \mathbf{A}_2 \end{bmatrix} \Rightarrow \mathbf{A}_2(\theta) = \mathbf{A}_1 \Phi$$

$$\Phi = \text{diag}(e^{-j2\pi\nu_{\theta_p}})$$

- Matrix Φ provides directly the desired angles.

ESPRIT method

- The same transform on the eigenvectors of the signal subspace leads to

$$\mathbf{U}_1 = \mathbf{A}_1\mathbf{T}, \;\; \mathbf{U}_2 = \mathbf{A}_2\mathbf{T}$$

$$\mathbf{U}_2 = \mathbf{A}_1\Phi\mathbf{T} = \mathbf{U}_1\mathbf{T}^{-1}\Phi\mathbf{T}$$

- Il suffit de trouver Ψ tel que $\mathbf{U}_2 = \mathbf{U}_1\Psi$ By least squares estimation:

$$\Psi = (\mathbf{U}_1^H\mathbf{U}_1)^{-1}\mathbf{U}_1^H\mathbf{U}_2$$

- Φ and Ψ have the same eigenvalues

$$\mathrm{Eig}(\Psi) = \mathrm{diag}(e^{j2\pi\nu_{\theta_p}})$$

'Generalized' ESPRIT method

- ESPRIT can be used, not only for ULA but for any array containing 2 sub-arrays such that the 2nd is the translated version of the first one. Hence, for a source located at θ

$$\begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1(\theta) \\ \mathbf{a}_2(\theta) \end{bmatrix} s(t) \;\text{ with }\; \mathbf{a}_2(\theta) = \mathbf{a}_1(\theta)e^{-j2\pi\nu_\theta}$$

- Space shift: plays the role of the inter-sensors distance in ULA.

$$\begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \mathbf{s}(t) \Rightarrow \mathbf{A}_1 = \mathbf{A}_2\Phi$$

ESPRIT method: algorithm

- Eigen-decomposition of the covariance algorithm (noiselesss case)

$$\mathbf{R} = \mathbf{E}\left(\begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix}^H \right) = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix} \Lambda \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix}^H$$

- Rotational invariance property for $\mathbf{A}$ and $\mathbf{U}_s$

$$\begin{aligned} \mathbf{U}_1 &= \mathbf{A}_1\mathbf{T} \\ \mathbf{U}_2 &= \mathbf{A}_2\mathbf{T} \end{aligned}$$

$$\exists\Psi \text{ such that } \mathbf{U}_2 = \mathbf{U}_1\Psi$$

- Matrix Φ is the matrix of eigenvalues of Ψ

$$\text{Eig}(\Psi) = \text{diag}(e^{j2\pi\nu_{\theta_p}})$$

Other localization methods: ML

- In the AWGN and deterministic inputs case, the likelihood function can be expressed as:

$$L(\theta, s(t), \sigma^2) = \prod_{t=1}^{T} (\pi\sigma^2)^{-N} e^{-\frac{\|x(t)-As(t)\|^2}{\sigma^2}}$$

- Let Π_A be the orthogonal projection matrix on Range$(\mathbf{A})$

$$\Pi_A = \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H \Rightarrow \mathbf{A}\hat{\mathbf{s}}(t) = \Pi_A\mathbf{x}(t)$$

$$\overline{\Pi}_A = \mathbf{I} - \Pi_A \Rightarrow \theta = \arg\min_{\theta} \text{Tr}(\overline{\Pi}_A\hat{\mathbf{R}})$$

Other methods: Weighted subspace fitting

- Exploits the relation between $\mathbf{U}_s$ and $\mathbf{A}$

$$\exists \mathbf{T} \text{ such that } \mathbf{U}_s = \mathbf{AT}$$

- Minimise the LS distance

$$\{\hat{\theta}, \hat{\mathbf{T}}\} = \arg\min_{\theta, \mathbf{T}} \|\hat{\mathbf{U}}_s - \mathbf{AT}\|_W^2$$

- Solving in $\mathbf{T}$ first followed by an estimation of θ

$$\hat{\mathbf{T}} = \mathbf{A}^{\#}\mathbf{U}_s \text{ with } \mathbf{A}^{\#} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$$

$$\text{then } \hat{\theta} = \arg\min_{\theta} \mathrm{Tr}(\overline{\Pi}_A \mathbf{U}_s \mathbf{W} \mathbf{U}_s)$$

Asymptotic optimal weighting $\mathbf{W} = (\Lambda_s - \sigma^2\mathbf{I})^2\Lambda_s^{-1}$

Discussion

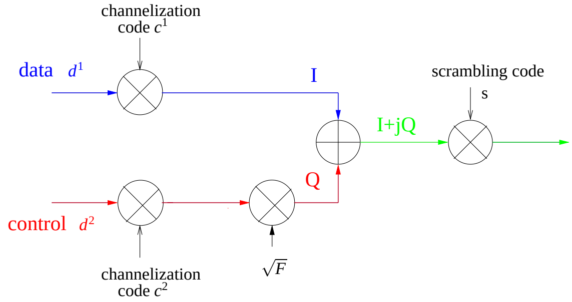
- Many existing localizaton methods.
- Compromise between resolution (MUSIC, ESPRIT, ..) and robustness and computational complexity (Beamforming).

- Many existing extentions:
  - Joint estimation of angles and delays (JADE algorithm)
  - Generalisation to wide-band sources,
  - Tracking and adaptive processing, ...

Application to Mobile Localisation in UMTS-FDD

Up-Link

- Transmitted signal by k-th user:



- Propagation channel

$$\mathbf{h}_k(t) = \sum_{i=1}^{R_k} \mathbf{a}(\theta_{k,i})\beta_{k,i}\mathbf{g}(t - \tau_{k,i})$$

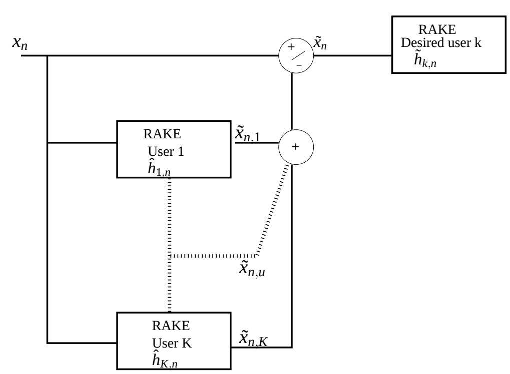
Joint AOA and TOA estimation

- **Raison:** Correspondance between the AOAs and TOAs of the multi-paths ⇒ for joint angle-delay localization. Also, the direct path is chosen as the one associated with the smallest TOA.

- **State of the art**
  - Maximum likelihood approach [Wax & al. 1997].
  - Subspace methods: Time Space Time-MUSIC [Wax & al. 2001].
  - ESPRIT-like methods [Vanderveen & al. 1998].
- **Proposed method:**
  - Delay estimation using the channel FT matrix by ESPRIT.
  - Estimation of only the desired angle (i.e. the one corresponding to the smallest delay).

Hearing problem

- AOA-TOA estimation algorithms require a first channel estimation:
  - RAKE-type estimator: non-robust to near-far effect (interferences).
  - RAKE-estimator with interference cancellation (PIC):



Localization with antenna array:

- Channel model: $\mathbf{h}(t) = \begin{bmatrix} h_1(t) \\ \vdots \\ h_N(t) \end{bmatrix} = \sum_{i=1}^{d} \mathbf{a}(\theta_i)\beta_i \mathbf{g}(t - \tau_i)$
- For a uniform circular array : $\mathbf{a}(\theta_i) = \begin{bmatrix} e^{j\xi\cos(\theta_i - \gamma_1)} \\ \vdots \\ e^{j\xi\cos(\theta_i - \gamma_N)} \end{bmatrix}$
- Channel matrix :

$$\begin{aligned} \mathbf{H} &\triangleq \left[\mathbf{h}(0)\,\mathbf{h}(\frac{T}{P})\cdots\mathbf{h}(LT - \frac{T}{P})\right] \\ &= [\mathbf{a}(\theta_1)\cdots\mathbf{a}(\theta_d)] \begin{bmatrix} \beta_1 & & 0 \\ & \ddots & \\ 0 & & \beta_d \end{bmatrix} \begin{bmatrix} \mathbf{g}_{\tau_1} \\ \vdots \\ \mathbf{g}_{\tau_d} \end{bmatrix} \\ &= \mathbf{A}(\theta)\mathbf{B}\mathbf{G}(\tau) \end{aligned}$$

Delays Estimation

- The FT of $\mathbf{H}$ transforms $\mathbf{G}(\tau)$ (up to a diagonal matrix) into:

$$\mathbf{V}(\tau) = \begin{bmatrix} 1 & \chi_1 & \chi_1^2 & \cdots & \chi_1^{LP-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \chi_d & \chi_d^2 & \cdots & \chi_d^{LP-1} \end{bmatrix}$$

  where $\chi_i = e^{\frac{-j2\pi\tau_i}{L}}$, $1 \le i \le d$.
- Matrix $\mathbf{H}_F$ has the rotational invariance property that allows for the estimation of τ_i using ESPRIT algorithm.

Angle estimation

- Once the delays are estimated we estimate the angle of the LOS path according to:
  - Inversion of the delays matrix:

$$\mathbf{H}' = \mathbf{H}\mathbf{G}(\tau)^{-1}$$

  - Selection of the first column $\mathbf{h}_1$ de $\mathbf{H}'$ and estimation of the AOA of the first path by maximizing :

$$\|\mathbf{a}(\theta)^H \mathbf{h}_1\|$$

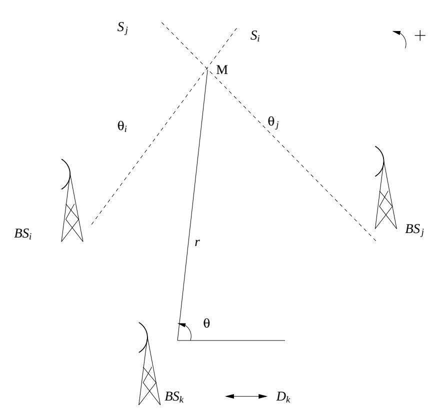
Proposed method for the NLOS

- **Idea:** Selection of the 2 'most reliable' measures: Coherence criterion of 2 given AOAs.
- **Coherence measure:**
  - If we know the distribution of the mobile position D_k w.r.t. a BS k:

$$P(\theta_i, \theta_j / D_k) = D_k(M)$$

    $\Rightarrow$ We would select the pair (θ_i, θ_j) that maximizes $P(\theta_i, \theta_j / D_k)$.
  - To give equal opprtunity to all BSs, we chose:

$$\hat{i}, \hat{j} = \arg \max_{i,j,k} P(\theta_i, \theta_j / D_k)$$

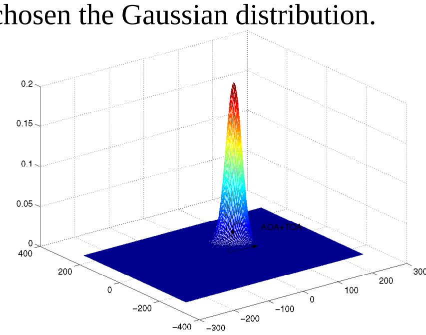


‘A priori’ mobile position distribution

Many possible distributions: We have chosen the Gaussian distribution.

- σ_r et σ_θ are ad-hoc.
- $\mu_\theta = \theta_k$.
- $\mu_r = d_k$: t_k is linked to d_k via the relation:

$$d_k = c(t_k - t_{0_k})$$



$$D(r,\theta) = \frac{1}{2\pi\sigma_r\sigma_\theta} e^{-\left(\frac{r-\mu_r}{\sqrt{2}\sigma_r}\right)^2} e^{-\left(\frac{\theta-\mu_\theta}{\sqrt{2}\sigma_\theta}\right)^2}$$

Synchronisation constraint

- **Problem:** Necessitates between the mobile and the BSs. Too constraining!!

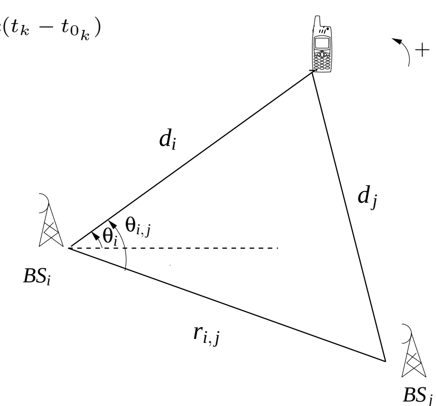
$$d_k = c(t_k - t_{0_k})$$

- **Alternative solution:**
  - Use a similar technique to the Timing Advance in GSM.
  - Estimate the time references by minimizing:

$$\hat{t}_{0_1}, \ldots, \hat{t}_{0_I} =$$

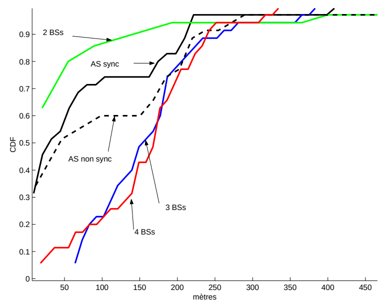
$$\arg\min \sum_{i=1}^{I}\sum_{j=1}^{I} \|d_j(t_{0_j}) - d_j(t_{0_i})\|^2$$

where $d_{i,j}$ is given by: $\sqrt{r_{i,j}^2 + d_i^2 - 2\cos(\theta_{i,j}) r_{i,j} d_i}$

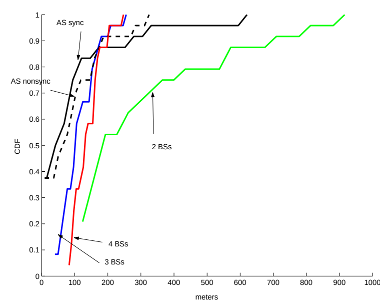


Localization results

- Comparison with standard triangulation techniques.



(k) NLOS sur les BS 3 et 4



(l) NLOS sur les BS 2 et 4

Random mobile position at each run, $L = 80$ slots, 4 BSs, $K = 20$.

Concluding Remarks

Conclusion

- **Main difficulties (Hearing + NLOS)**: No fully satisfactory solution (i.e. still an open problem). We have presented certain solutions using, when possible, partial interference cancellation and selection of the 'best' AOA/TOA estimates. Other solutions exist, e.g.
  - Using 'a priori' learning of the dependence of the channel impulse response on the mobile position (too expensive and requires regular up-dating),
  - Using a 'super calculator' which captures both the transmitted and received signals to extract the desired information,
  - Using Idle periods: reduces significantly the system capacity.

Conclusion

- **Estimation accuracy**: The best estimates are computationally demanding and the power in the downlink is 'restricted'. Good 'intermediate' solutions especially in adaptive schemes.
- **Tracking**: Many tracking algorithms exist using subspace tracking, Kalman filtering, particular filtering, gradient techniques, etc. Tracking might improve the estimation accuracy (at least for slowly moving mobiles) due to memory effect.
- **Hybrid solution**: Use both GPS and terrestrial BS signals for mobile location.