Smart Beamforming for Wireless Communications: A Novel Minimum Bit Error Rate Approach

S. Chen, N.N. Ahmad and L. Hanzo

Department of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ
United Kingdom
sqc@ecs.soton.ac.uk nna00r@ecs.soton.ac.uk lh@ecs.soton.ac.uk

Presented at 2nd IMA Conference on Mathematics in Communications
16-18 December 2002 Lancaster University

Motivations

- Spatial processing with adaptive antenna arrays has shown real promise for substantial capacity enhancement.
- Adaptive beamforming is capable of separating signals transmitted on the same carrier frequency but are separated in the spatial domain.
- Classical beamforming technique is based on minimizing the system mean square error.
- For a communication system, it the bit error rate, not the mean square error, that really matters.
- This motivates our derivation of a novel beamforming technique based directly on minimizing the system bit error rate.

System Model

- The system has M users (sources), and each transmits a binary phase shift keying (BPSK) signal on the same carrier frequency $\omega = 2\pi f$.
- The baseband signal of user i with signal power A_i^2 is

$$m_i(k) = A_i b_i(k), \quad b_i(k) \in \{\pm 1\}, \quad 1 \le i \le M$$

  Source 1 is the desired user and the rest are interfering users.

- The signals at the antenna array of L uniformly spaced elements are

$$x_l(k) = \sum_{i=1}^{M} m_i(k) \exp\left(j\omega t_l(\theta_i)\right) + n_l(k) = \bar{x}_l(k) + n_l(k), \ 1 \le l \le L$$

  $t_l(\theta_i)$: the relative time delay at element l for source i,
  θ_i: the direction of arrival for source i, and
  $n_l(k)$: a complex-valued white Gaussian noise with $E[|n_l(k)|^2] = 2\sigma_n^2$.

Matrix Form of System Model

- Define the steering vector for source i

$$\mathbf{s}_i = [\exp(j\omega t_1(\theta_i)) \cdots \exp(j\omega t_L(\theta_i))]^T$$

  the system matrix

$$\mathbf{P} = [A_1 \mathbf{s}_1 \cdots A_M \mathbf{s}_M]$$

  the bit vector

$$\mathbf{b}(k) = [b_1(k) \cdots b_M(k)]^T$$

  and the noise vector

$$\mathbf{n}(k) = [n_1(k) \cdots b_L(k)]^T$$

- Then, the array input vector $\mathbf{x}(k) = [x_1(k) \cdots x_L(k)]^T$ is expressed as

$$\mathbf{x}(k) = \bar{\mathbf{x}}(k) + \mathbf{n}(k) = \mathbf{P}\mathbf{b}(k) + \mathbf{n}(k)$$

Beamformer

- The beamformer output is

$$y(k) = \mathbf{w}^H \mathbf{x}(k) = \mathbf{w}^H \bar{\mathbf{x}}(k) + \mathbf{w}^H \mathbf{n}(k) = \bar{y}(k) + e(k)$$

where $\mathbf{w} = [w_1 \cdots w_L]^T$ is the complex-valued beamformer weight vector and $e(k)$ is Gaussian with zero mean and $E[|e(k)|^2] = 2\sigma_n^2 \mathbf{w}^H \mathbf{w}$.

- The estimate of the transmitted bit $b_1(k)$ is

$$\hat{b}_1(k) = \begin{cases} +1, & y_R(k) = \Re[y(k)] > 0, \\ -1, & y_R(k) = \Re[y(k)] \le 0, \end{cases}$$

- The classical MMSE beamforming solution is given by

$$\mathbf{w}_{\rm MMSE} = \left(\mathbf{P}\mathbf{P}^H + 2\sigma_n^2 \mathbf{I}_L\right)^{-1} \mathbf{p}_1$$

with $\mathbf{p}_1$ being the first column of $\mathbf{P}$

Signal States

- Denote the $N_b = 2^M$ possible sequences of $\mathbf{b}(k)$ as $\mathbf{b}_q$, $1 \le q \le N_b$. Let the first element of $\mathbf{b}_q$, corresponding to the desired user, be $b_{q,1}$.

- Then, $\bar{\mathbf{x}}(k)$ only takes values from the signal state set defined as

$$\mathcal{X} \triangleq \{\bar{\mathbf{x}}_q = \mathbf{P}\mathbf{b}_q, \ 1 \le q \le N_b\}$$

- Therefore, $\bar{y}(k) \in \mathcal{Y} \triangleq \{\bar{y}_q = \mathbf{w}^H \bar{\mathbf{x}}_q, \ 1 \le q \le N_b\}$.

- Thus, $\bar{y}_R(k) = \Re[\bar{y}(k)]$ can only take values from the set

$$\mathcal{Y}_R \triangleq \{\bar{y}_{R,q} = \Re[\bar{y}_q], \ \ 1 \le q \le N_b\}$$

which can be divided into the two subsets conditioned on $b_1(k)$

$$\mathcal{Y}_R^{(\pm)} \triangleq \{\bar{y}_{R,q}^{(\pm)} \in \mathcal{Y}_R : \ b_1(k) = \pm 1\}$$

Bit Error Rate

- The conditional probability density function of $y_R(k)$ given $b_1(k) = +1$ is

$$p(y_R|+1) = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} \frac{1}{\sqrt{2\pi\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \exp\left(-\frac{\left(y_R - \bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}\right)$$

where $\bar{y}_{R,q}^{(+)} \in \mathcal{Y}_R^{(+)}$ and $N_{sb} = N_b/2$ is the number of the points in $\mathcal{Y}_R^{(+)}$.

- Thus the BER is given by

$$P_E(\mathbf{w}) = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} Q\left(g_{q,+}(\mathbf{w})\right)$$

where

$$Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^{\infty} \exp\left(-\frac{v^2}{2}\right) dv \quad \text{and} \quad g_{q,+}(\mathbf{w}) = \frac{\mathrm{sgn}(b_{q,1}) \bar{y}_{R,q}^{(+)}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}}$$

Minimum Bit Error Rate Beamformer

- The MBER beamforming solution is then defined as

$$\mathbf{w}_{\rm MBER} = \arg\min_{\mathbf{w}} P_E(\mathbf{w})$$

- There exists no closed-form solution, but with the gradient

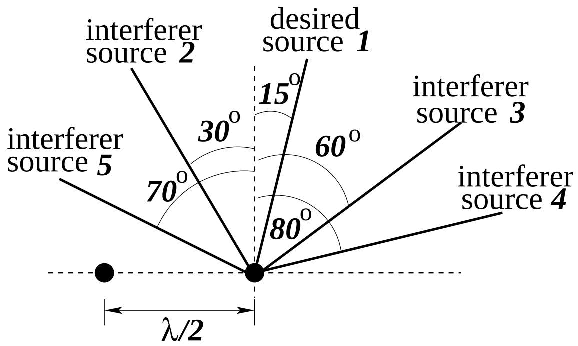
$$\nabla P_E(\mathbf{w}) = \frac{1}{2N_{sb}\sqrt{2\pi\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}\right) \mathrm{sgn}(b_{q,1}) \left(\frac{\bar{y}_{R,q}^{(+)} \mathbf{w}}{\mathbf{w}^H \mathbf{w}} - \bar{\mathbf{x}}_q^{(+)}\right)$$

a MBER solution can be obtained iteratively using a simplified conjugated gradient algorithm.

- BER is invariant to the size of $\mathbf{w}$. Thus, if $\mathbf{w}_{\rm MBER}$ is a MBER solution, $\alpha \mathbf{w}_{\rm MBER}$ is also a MBER solution for $\alpha > 0$.

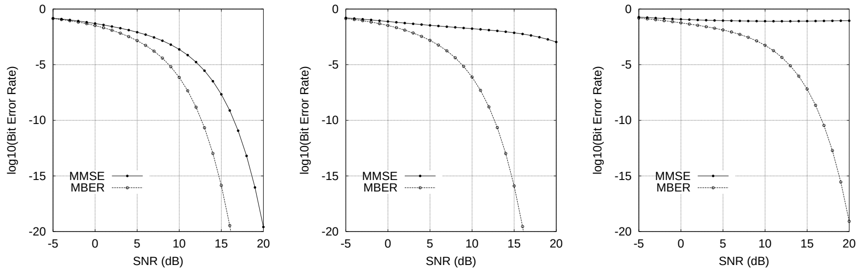
Example

Locations of the desired source and the interfering sources with respect to the two-element linear array with $\lambda/2$ element spacing, λ being the wavelength.



Definitions: SNR= $A_1^2/2\sigma_n^2$, SIR$_i = A_1^2/A_i^2$ for $i = 2, \cdots, M$.

Bit Error Rate Comparison



(a) (b) (c)

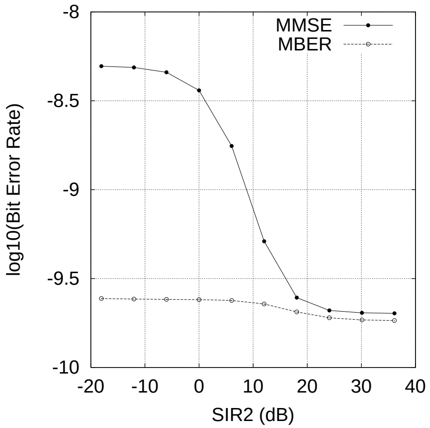
(a): SIR$_i = 0$ dB, $i = 2, 3, 4, 5$;

(b): SIR$_2 = -6$ dB and SIR$_i = 0$ dB, $i = 3, 4, 5$;

(c): SIR$_i = -6$ dB, $i = 2, 3, 4, 5$;

Near-Far Effect

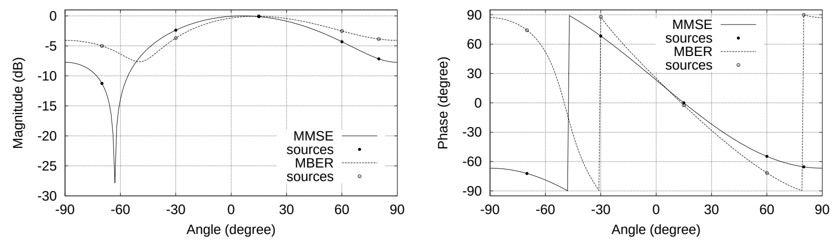
The near-far effect to bit error rate performance. SNR= 10 dB, SIR$_i$ = 24 dB for $i = 3, 4, 5$, varying SIR$_2$.



- The MBER solution appears to be robust to the near-far effect.

Beam Pattern Comparison

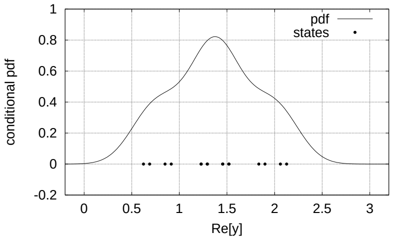
SNR= 10 dB, SIR$_i = 0$ dB, $i = 2, 3, 4, 5$.



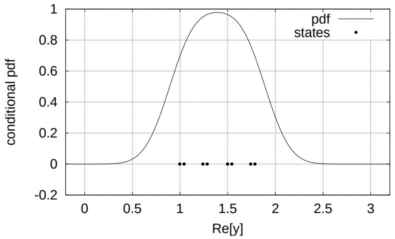
- Let $F(\theta)$ be the normalized DFT of the beamformer weight vector.
- Traditionally, the magnitude of $F(\theta)$ is used to judge the performance of a beamformer.
- Magnitude response along can be misleading, as in this case.
- At the four angles for the four interfering sources, the phase responses of the MBER solution are much closer to $\pm\frac{\pi}{2} \Rightarrow$ a much better response of $y_R(k) = \Re[y(k)]$.

Probability Density Function Comparison

Conditional probability density function of beamformer given $b_1(k) = +1$ and subset $\mathcal{Y}_R^{(+)}$. SNR$=10$ dB, SIR$_i = 0$ dB, $i = 2, 3, 4, 5$.



(a) MMSE

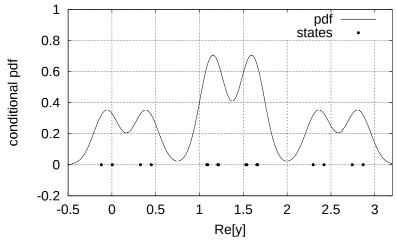


(b) MBER

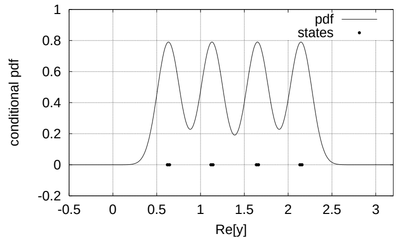
- The beamformer weight vector is normalized to a unit length, so that the BER is mainly determined by the minimum distance of the subset $\mathcal{Y}_R^{(+)}$ to the decision threshold $y_R = 0$.
- This minimum distance is much larger for the MBER beamformer.

Probability Density Function Comparison

Conditional probability density function of beamformer given $b_1(k) = +1$ and subset $\mathcal{Y}_R^{(+)}$. SNR$=15$ dB, SIR$_i = -6$ dB, $i = 2, 3, 4, 5$.



(a) MMSE



(b) MBER

- The beamformer weight vector is normalized to a unit length.
- Note that $\mathcal{Y}_R^{(+)}$ and $\mathcal{Y}_R^{(-)}$ are no longer linearly separable for the MMSE beamformer $\Rightarrow$ a high BER floor.

Block-Data Adaptive MBER Algorithm

- Given a block of K training samples $\{\mathbf{x}(k), b_1(k)\}$, a Parzen window estimate of the beamformer p.d.f. is

$$\hat{p}(y_R) = \frac{1}{K\sqrt{2\pi\rho_n^2 \mathbf{w}^H\mathbf{w}}} \sum_{k=1}^{K} \exp\left(-\frac{(y_R - y_R(k))^2}{2\rho_n^2 \mathbf{w}^H \mathbf{w}}\right)$$

where the kernel width ρ_n is related to the noise standard deviation σ_n.

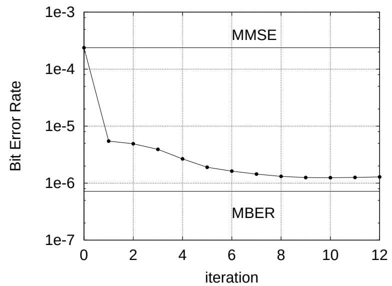
- From this estimated p.d.f., the estimated BER is given by:

$$\hat{P}_E(\mathbf{w}) = \frac{1}{K}\sum_{k=1}^{K} Q\left(\hat{g}_k(\mathbf{w})\right) \quad \text{with} \quad \hat{g}_k(\mathbf{w}) = \frac{\mathrm{sgn}(b_1(k))y_R(k)}{\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}}$$

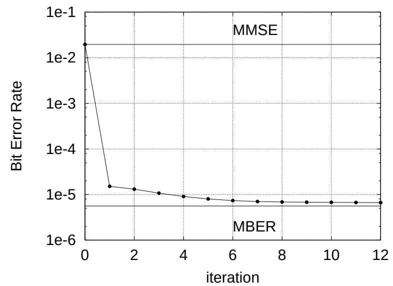
- Upon substituting $\nabla P_E(\mathbf{w})$ by $\nabla \hat{P}_E(\mathbf{w})$ in the conjugate gradient updating mechanism, a block-data based adaptive algorithm is obtained.

Convergence of Block Adaptive Algorithm

Convergence rate of the block-data based adaptive MBER algorithm for a block size of $K = 200$. The initial weight vector is set to $\mathbf{w}_{\mathrm{MMSE}}$.



(a)

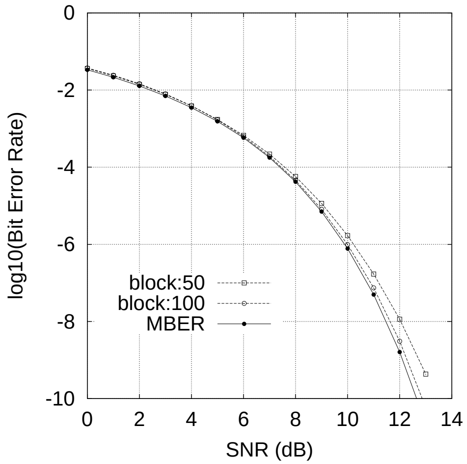


(b)

(a): SNR$=10$ dB, SIR$_i = 0$ dB for $i = 2, 3, 4, 5$, adaptive gain $\mu = 1.0$ and $\rho_n^2 = 6\sigma_n^2 = 0.3$. (b): SNR$=10$ dB, SIR$_3 =$SIR$_4 = 0$ dB, SIR$_2 =$SIR$_5 = -6$ dB, adaptive gain $\mu = 0.5$ and $\rho_n^2 = 2\sigma_n^2 = 0.1$.

Effect of Block Size

Effect of block size on the performance of the block-data based adaptive MBER algorithm for $\text{SIR}_2 = -6$ dB and $\text{SIR}_i = 0$ dB, $i = 3, 4, 5$.



Least Bit Error Rate Algorithm

- Consider a single-sample p.d.f. estimate of the beamformer output

$$\tilde{p}(y_R, k) = \frac{1}{\sqrt{2\pi}\rho_n} \exp\left(-\frac{(y_R - y_R(k))^2}{2\rho_n^2} \right)$$

- This leads to a single-sample BER estimate $\tilde{P}_E(\mathbf{w}, k)$.
- Using the instantaneous stochastic gradient

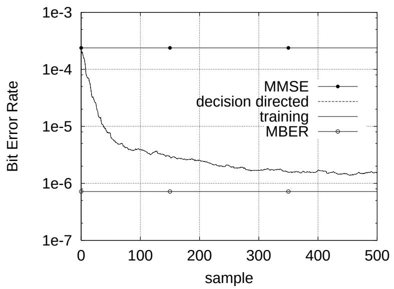
$$\nabla \tilde{P}_E(\mathbf{w}, k) = -\frac{\text{sgn}(b_1(k))}{2\sqrt{2\pi}\rho_n} \exp\left(-\frac{y_R^2(k)}{2\rho_n^2} \right) \mathbf{x}(k)$$

- leads to the LBER algorithm

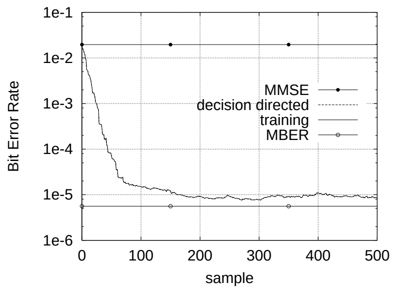
$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\text{sgn}(b_1(k))}{2\sqrt{2\pi}\rho_n} \exp\left(-\frac{y_R^2(k)}{2\rho_n^2} \right) \mathbf{x}(k)$$

Learning Curves of LBER Algorithm

Learning curves of the LBER algorithm averaged over 20 runs, the initial weight vector is set to $\mathbf{w}_{\text{MMSE}}$, solid curve is for training and dashed curve for decision-directed adaptation with $\hat{b}_1(k)$ substituting $b_1(k)$ (two curves are indistinguishable).



(a)



(b)

(a): SNR= 10 dB, $\text{SIR}_i = 0$ dB for $i = 2, 3, 4, 5$, $\mu = 0.03$ and $\rho_n^2 = 8\sigma_n^2 = 0.4$.
(b): SNR= 10 dB, $\text{SIR}_3 = \text{SIR}_4 = 0$ dB, $\text{SIR}_2 = \text{SIR}_5 = -6$ dB, $\mu = 0.02$ and $\rho_n^2 = 4\sigma_n^2 = 0.2$.