Hungary-Serbia
IPA Cross-border Co-operation Programme

The project is co-financed by the European Union

CEFAIC
Center of Excellence For Advanced and Intelligent Control

CENTER OF EXELLENCE FOR ADVANCED AND INTELIGENT CONTROL

HUSRB/1203/221/020

Good neighbours
creating
common future

PARTNERS:

Faculty of Technical Sciences

University of Novi Sad

University of Szeged

Good neighbours creating common future

This document has been produced with the financial assistance of the European Union. The content of the document is the sole responsibility of Faculty of technical sciences, University of Novi Sad and can under no circumstances be regarded as reflecting the position of the European Union and/or the Managing Authority.

Hungary-Serbia
IPA Cross-border Co-operation Programme

Introduction to adaptive filters

Silvester Pletl

2014.01.10.

The project is co-financed by the European Union

Good neighbours creating common future

Introduction

Hungary-Serbia
IPA Cross-border Co-operation Programme

In this lecture we will consider an overview of adaptive signal processing:

- History
- Architectures
- Algorithms (we mainly consider the LMS algorithm)
- Applications

The project is co-financed by the European Union

Good neighbours creating common future

History

- Least squares - 19th Century mathematician **Gauss**.
- Least Squares is **widely used** off-line in practically every branch of science, engineering and business.
- Least mean squares first suggested for DSP in 1960 by Bernard **Widrow**.

The discrete mathematics of adaptive filtering, is fundamentally based on the least squares minimization theory of the celebrated 19th Century German mathematician Gauss.

Least squares is of course widely used in statistical analysis and virtually every branch of science and engineering.

For DSP however the problem of least squares minimization is applied to **real time data**. This presents the challenge of producing a real time implementation to operate on data arriving at high data rates (from 1kHz to 100MHz), and with loosely known statistics and properties.

Real Time DSP in the 1960's

Hungary-Serbia
IPA Cross-border Co-operation Programme

The first suggestion of adaptive DSP algorithms was in Widrow and Hoff's classic paper on the adaptive switching circuits and the least mean squares (LMS) algorithm in 1960 (IRE WESCON Conference). This paper stimulated great interest by providing a practical and potentially real time solution for least squares implementation.

Widrow (at Stanford University) **followed up this work** with two definitive and classic papers in the 1970s:

- B. Widrow et al. Adaptive Noise Cancellation: Principles and Applications. Proceedings of the IEEE, Vol. 63, pp. 1692-1716, 1975
- B. Widrow et al. Stationary and Non-stationary learning characteristics of the LMS adaptive filter. Proc. IEEE, Vol 64, pp.1151-1162, 1976.

And **nowadays**:

B. Widrow and E. Walach. *Adaptive Inverse Control.* New Jersey: Prentice-Hall, Inc., 2008

The project is co-financed by the European Union

Good neighbours creating common future

Modern Perspective

Hungary-Serbia
IPA Cross-border Co-operation Programme

- Since the 1970's there has been **considerable research** in adaptive signal processing algorithms and architectures.
- The advent of **powerful DSP** processors in the early 1980s has allowed many real time adaptive DSP systems to be developed.
- In the 1990s there has been a large growth in the **application** of real time adaptive DSP to solve many problems.
- **New algorithm** research continues for techniques such as the QR adaptive algorithm, least squares lattice, and more recently for neural networks (a class of non-linear adaptive Systems).
- Nowadays the majority of key innovations in adaptive signal processing research can be found in the **IEEE Trans on** Signal Processing
- FPGA

Theoretical prerequisites of Adaptive filters

Hungary-Serbia
IPA Cross-border Co-operation Programme

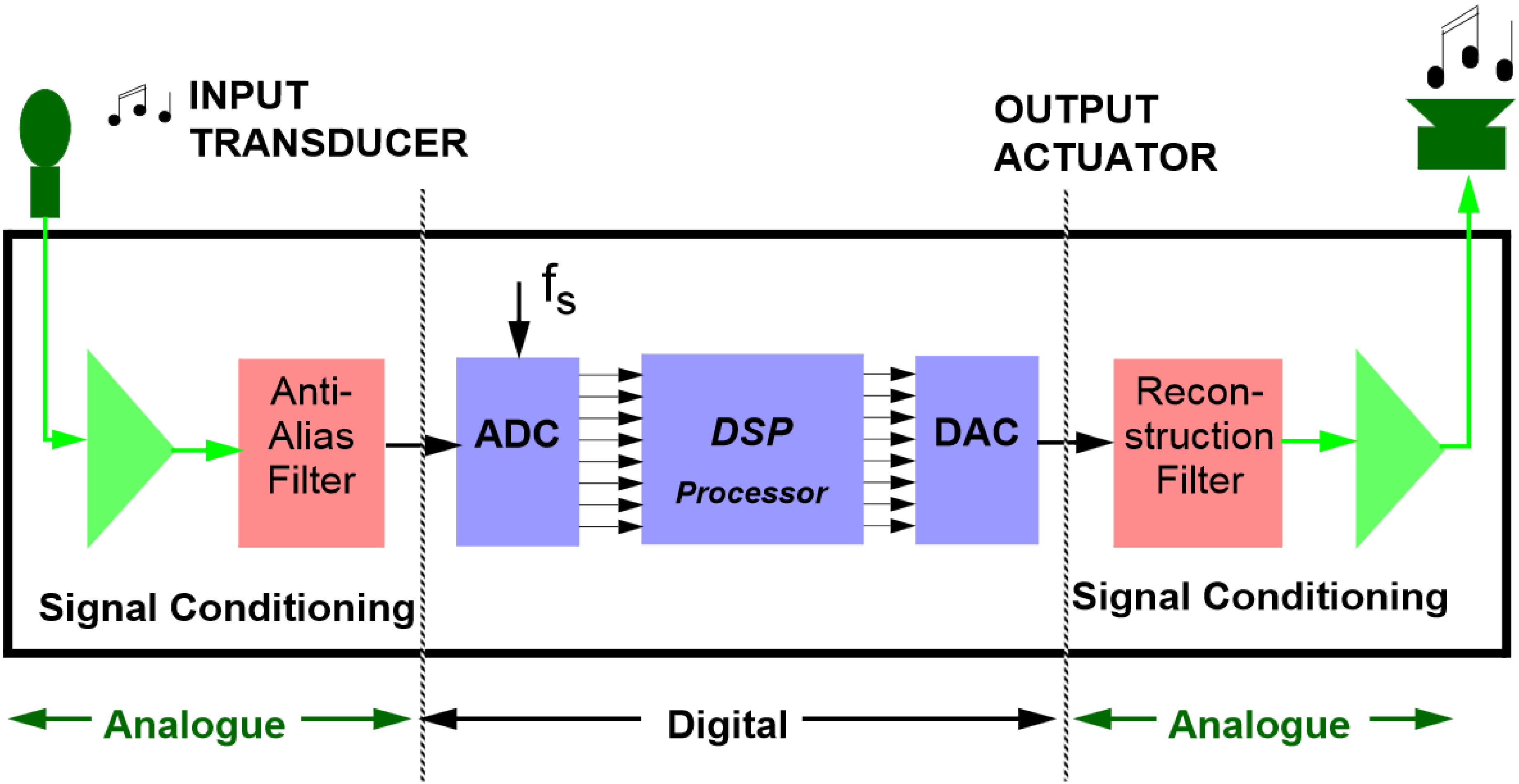
- **Digital Signal Processing Fundamentals**: Nyquist rate sampling, digital filters, Fourier transforms, analogue interfacing.
- **Statistical signal processing**: Correlation; Ergodicity; Means, variances; Stationarity; Wide sense stationarity; Frequency response / Power Spectrum.
- **Matrix algebra**: Addition, multiplication and matrix inverses; properties of the correlation/ covariance matrix; eigenvalues and eigenvectors; for QR matrix decomposition.

The project is co-financed by the European Union

Good neighbours creating common future

A Generic DSP System

Hungary-Serbia
IPA Cross-border Co-operation Programme



The project is co-financed by the
European Union

Good neighbours creating common future

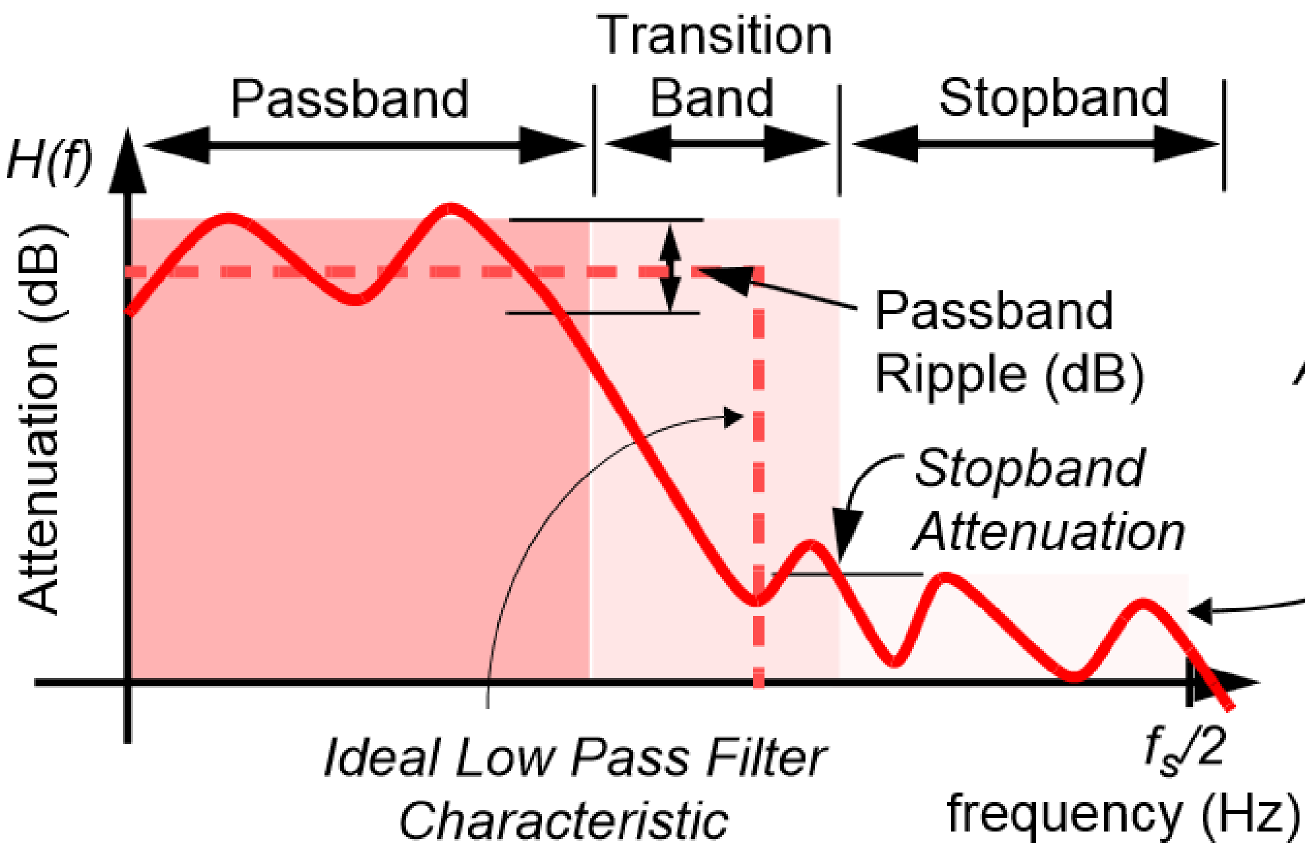
Fixed digital filters

One of the most common input/output DSP systems is a digital filter.

Fixed digital filters can be designed with a **wide variety of techniques**.

Most digital filter design software takes input of frequency response via “graphical” parameters. The **user inputs desirable parameters** to specify the acceptable tolerances from the ideal filter.

Fixed digital filters



Digital Filters

Hungary-Serbia
IPA Cross-border Co-operation Programme

Infinite Impulse Response (**IIR**) filters which are linear filters with feedback.

Finite Impulse Response (**FIR**) filter performs a linear combination:

$$y_k = \sum_{n=0}^{N-1} w_n x(k-n)$$

The difference equation for a simple 5 weight FIR filter is:

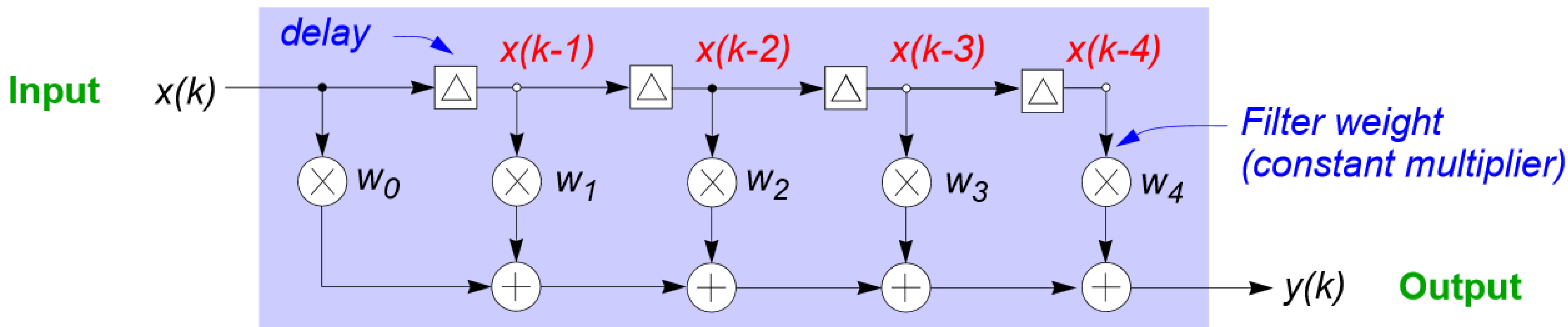
$$y(k) = x(k)w_0 + x(k-1)w_1 + x(k-2)w_2 + x(k-3)w_3 + x(k-4)w_4$$

The project is co-financed by the European Union

Good neighbours creating common future

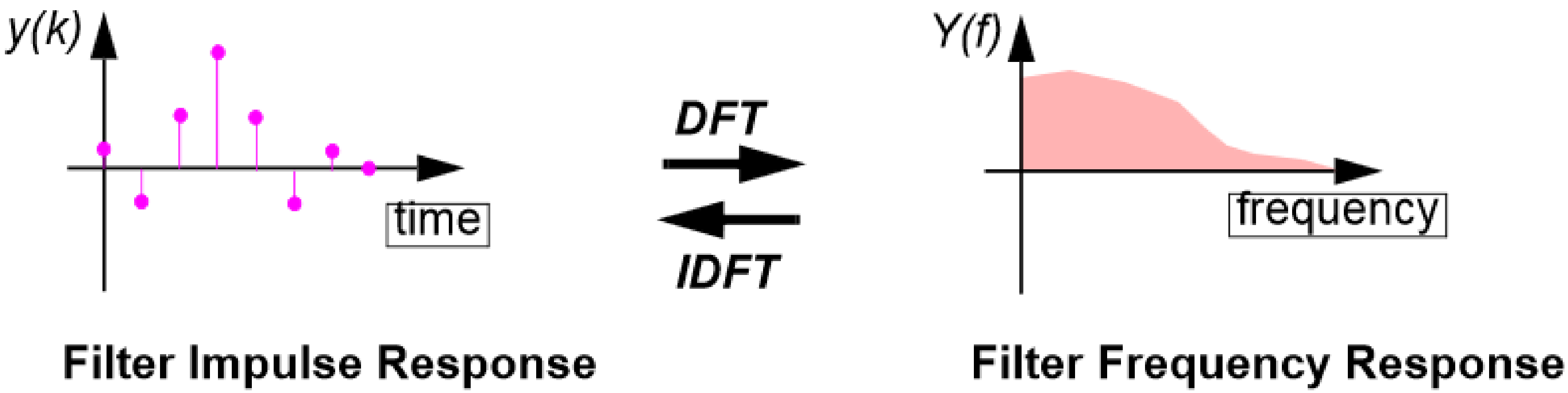
Signal flow graph of 5 weight FIR filter

$$y(k) = x(k)w_0 + x(k-1)w_1 + x(k-2)w_2 + x(k-3)w_3 + x(k-4)w_4$$



Response of a digital filter

The frequency response of a digital filter is found by taking the discrete Fourier Transform (DFT) of the impulse response



The project is co-financed by the European Union

Adaptive Digital Filters

Hungary-Serbia
IPA Cross-border Co-operation Programme

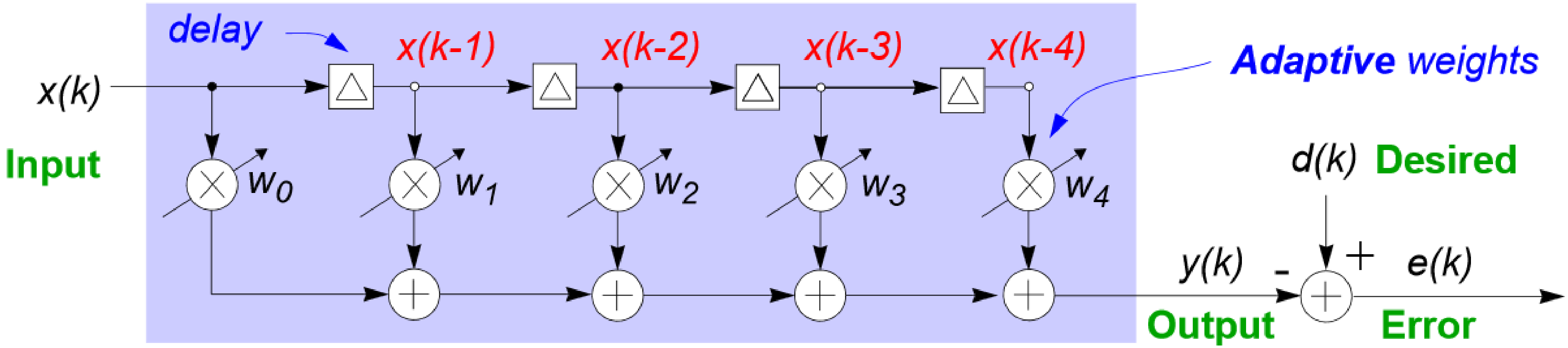
- Adaptive digital filters are self learning filters, whereby an FIR (or IIR) is designed based on the characteristics of input signals. No other frequency response information or specification information is available.
- There are a large number of applications suitable for the implementation of adaptive digital filters.

The project is co-financed by the European Union

Good neighbours creating common future

Signal flow graph of 5 weight Adaptive FIR filter

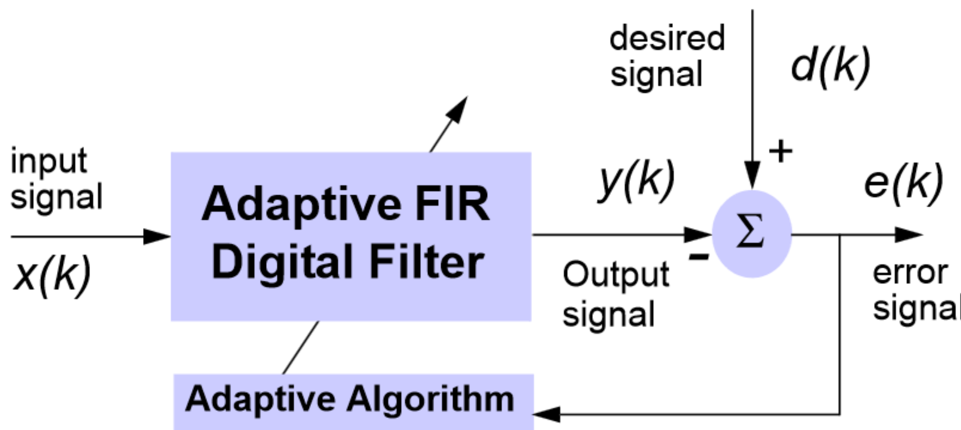
An adaptive digital filter is often represented by a signal flow graph with adaptive nature of weights shown:



An adaptive digital filter will therefore “adapt” to its environment. **The environment will be defined by the input signals x(k) and d(k)** to the adaptive digital filter.

General Closed Loop Adaptive filtering

"The aim is to adapt the digital filter such that the input signal x(k) is filtered to produce y(k) which when subtracted from desired signal d(k), will minimize the power of the error signal e(k)."



$$e(k) = d(k) - y(k)$$

$$y(k) = \mathrm{Filter}(x(k))$$

Adaptive Algorithms

The aim of the adaptive algorithm is to minimize the error signal power over a period of time.

This can be approached in two ways:

- Minimization of the mean squared error signal:

$$E[e^2(k)] = \frac{1}{M_2 - M_1} \sum_{n=M_1}^{M_2-1} e^2(k) \quad \text{for large } (M_2 - M_1)$$

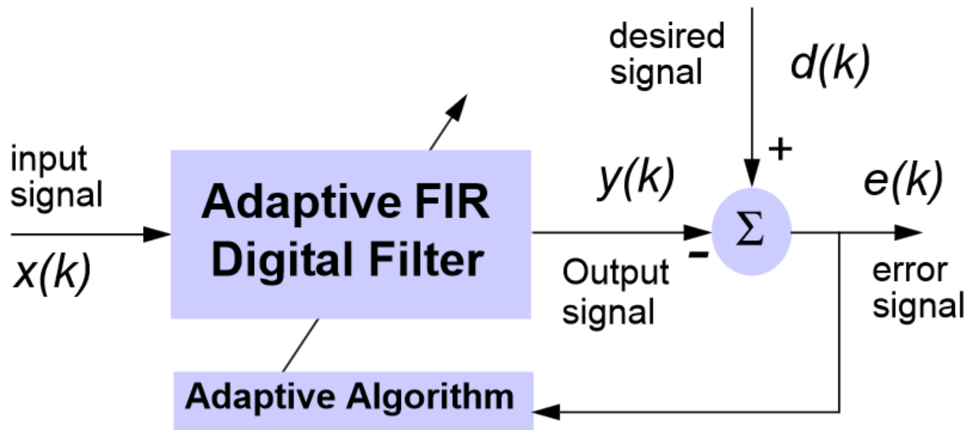
- Minimization of the total sum of error squares:

$$v_k = \sum_{s=0}^{k} e^2(s)$$

Minimizing the Mean Squared Error

Hungary-Serbia
IPA Cross-border Co-operation Programme

If the statistics of $x(k)$ and $d(k)$ are wide sense stationary and ergodic then we can choose to minimize the mean squared error signal:

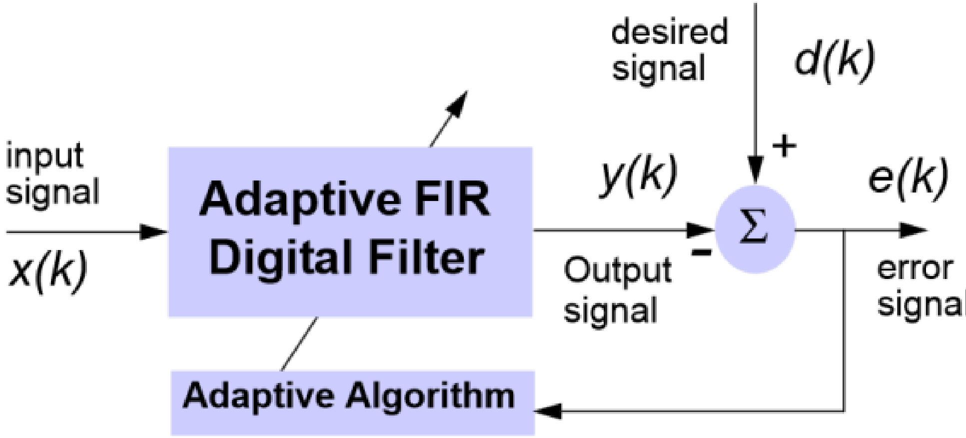


$$e(k) = d(k) - y(k)$$

$$y(k) = \text{Filter}(x(k))$$

The project is co-financed by the European Union

Good neighbours creating common future

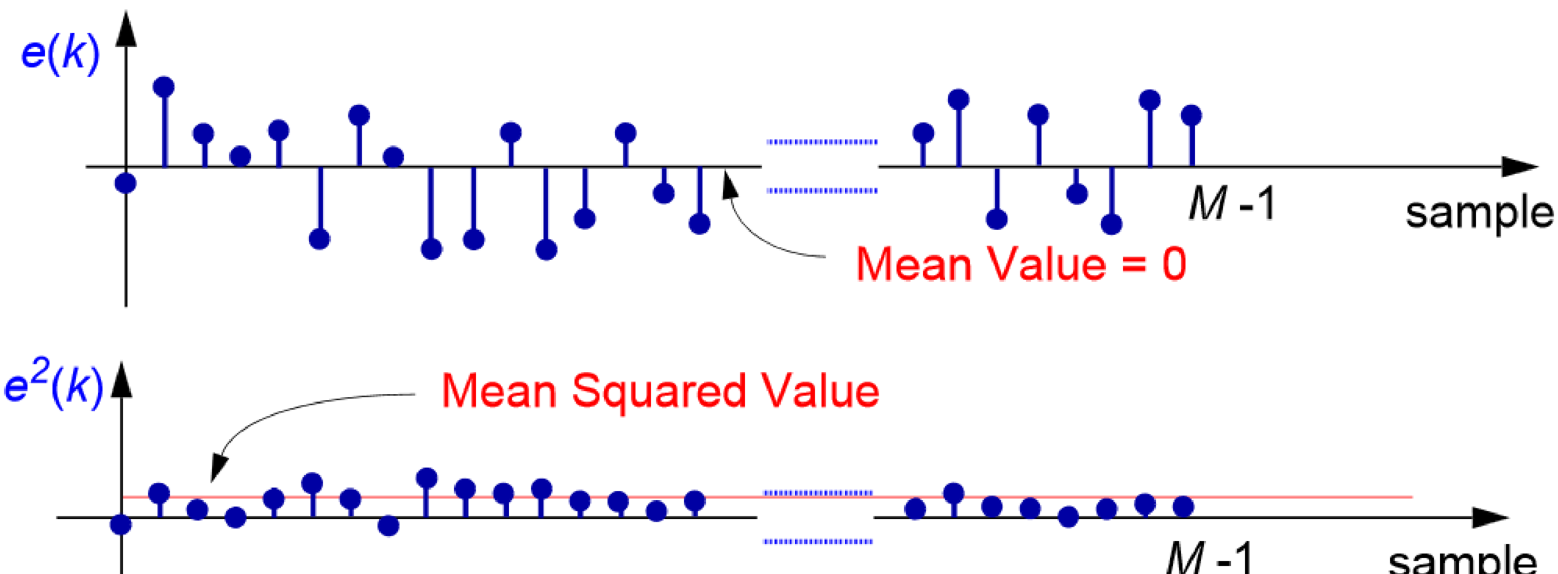


$$e(k) = d(k) - y(k)$$

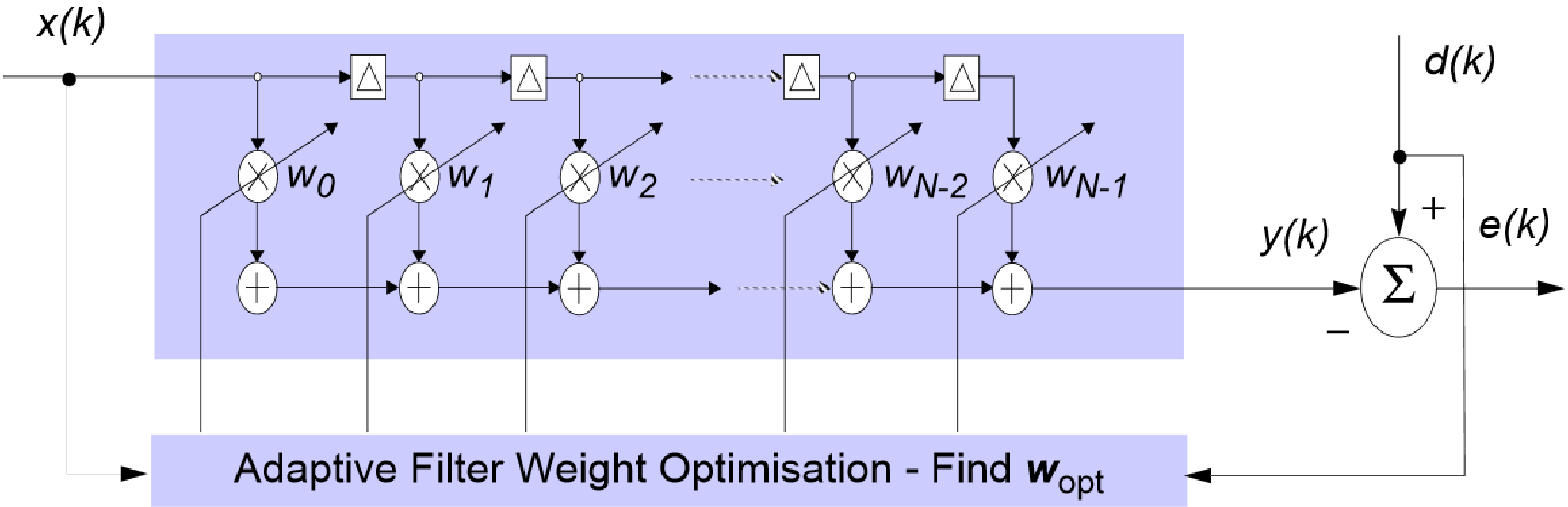
$$y(k) = \mathrm{Filter}(x(k))$$

$$\mathrm{MSE} = E[e^2(k)] = \frac{1}{M}\sum_{n=0}^{M-1} e^2(n), \quad \text{for large } M$$

The mean squared error (MSE) is in fact a measure of the signal power. Mean and mean squared value are assumed constant for wide sense stationary signals.



Adaptive Algorithm for FIR Filter



$$y(k) = \sum_{n=0}^{N-1} w_n x(k-n) = \mathbf{w}^T \mathbf{x}(k)$$

$$e(k) = d(k) - y(k) \qquad \mathbf{w}_{opt} = \text{function of}(x(k), d(k))$$

Mean Squared Error

Hungary-Serbia
IPA Cross-border Co-operation Programme

Consider a (trivial) **one weight filter** case.

Consider the MSE equation defining the so called MSE performance surface.

$$\zeta = E[e^2(k)] = E\left[\left(d(k) - wx(k)\right)^2\right]$$

$$\zeta = E\left[\left(d(k)\right)^2\right] - 2wE[d(k)x(k)] + w^2 E\left[\left(x(k)\right)^2\right]$$

$$R = E\left[\left(x(k)\right)^2\right] \qquad p = E[d(k)x(k)]$$

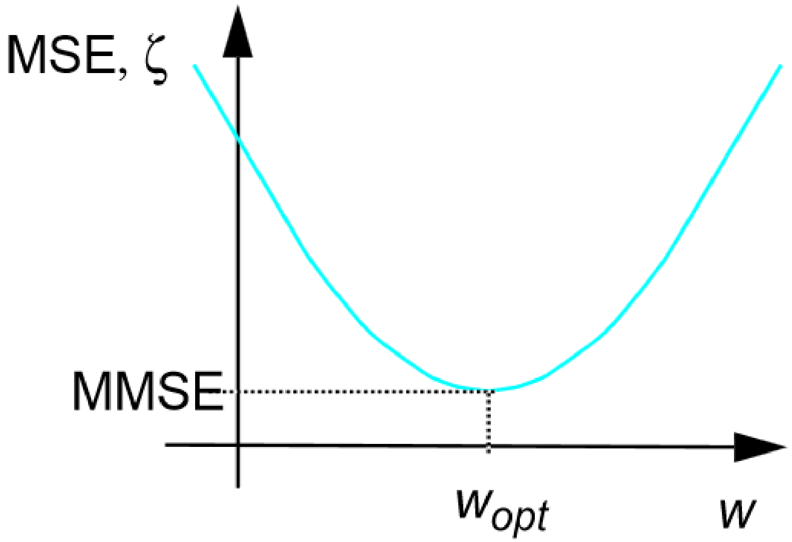
$$\zeta = E\left[d_k^2\right] - 2pw + w^2 r$$

The project is co-financed by the European Union

Good neighbours creating common future

Minimum Mean Squared Error Solution

$E\left[d_k^2\right]$ a constant, and w, r, and p are all scalars. Hence the performance surface is a **parabola** (upfacing). Plotting this performance surface gives:

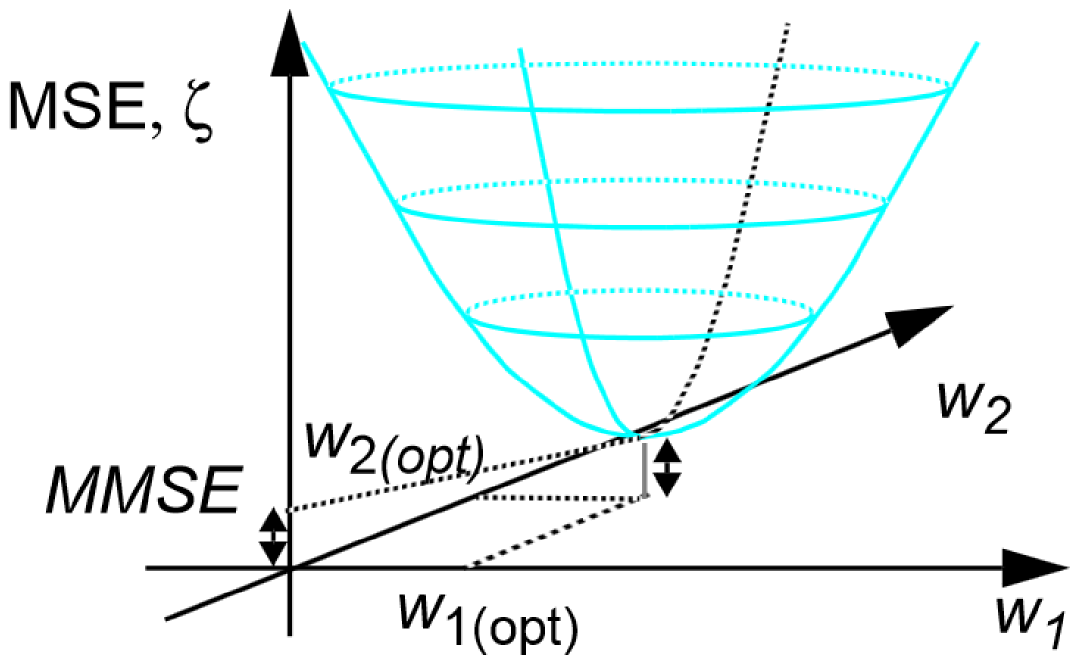


$$\zeta = E\left[d_k^2\right] - 2pw + w^2 r$$

The m $\dfrac{d\zeta}{dw} = 2rw - 2p = 0 \Rightarrow w_{opt} = r^{-1}p$ nt = 0

Minimum Mean Squared Error Solution

If the filter has two weights the performance surface is a **paraboloid** in 3 dimensions:



Minimum Mean Squared Error Solution

If the filter has more than three weights then we cannot draw the performance surface in three dimensions, however, mathematically there is only one minimum point which occurs when the gradient vector (with respect to w) is zero. A performance surface with more than three dimensions is often called a **hyperparaboloid**.

The project is co-financed by the European Union

Solution for N weights

Hungary-Serbia
IPA Cross-border Co-operation Programme

When:

$$y(k) = \sum_{n=0}^{N-1} w_n x(k-n) = \mathbf{w}^T \mathbf{x}(k) \qquad e(k) = d(k) - y(k)$$

Where:

$$\mathbf{w} = [w_0 \;\; w_1 \;\; w_2 \;\; \ldots\ldots \;\; w_{N-2} \;\; w_{N-1}]^T$$

$$\mathbf{x}(k) = [x(k) \;\; x(k-1) \;\; x(k-2) \;\; \ldots\ldots \;\; x(k-N+2) \;\; x(k-N+1)]^T$$

Consider the squared error:

$$e^2(k) = (d(k) - \mathbf{w}^T \mathbf{x}(k))^2$$

$$= d^2(k) + \mathbf{w}^T [\mathbf{x}(k) \mathbf{x}^T(k)] \mathbf{w} - 2 d(k) \mathbf{w}^T \mathbf{x}(k)$$

The project is co-financed by the European Union

Good neighbours creating common future

Mean Squared Error

Taking expected (or mean) values (and dropping "(k)" for notational convenience):

$$E[e^2(k)] = E[d^2] + \mathbf{w}^T E[\mathbf{x}\mathbf{x}^T]\mathbf{w} - 2\mathbf{w}^T E[d\mathbf{x}]$$

Writing in terms of the correlation matrix, **R** and the cross correlation vector, **p**, gives:

$$E[e^2(k)] = E[d^2(k)] + \mathbf{w}^T \mathbf{R} \mathbf{w} - 2\mathbf{w}^T \mathbf{p}$$

$$\mathbf{R} = E[\mathbf{x}\mathbf{x}^T] \qquad \mathbf{p} = E[d_k \mathbf{x}_k]$$

Mean Squared Error

Correlation Matrix: Assuming that x(k) and d(k) are wide sense stationary ergodic processes (i.e. mean and variance are constant) the **correlation matrix** for a 3 weight adaptive FIR filter

exar

$$\boldsymbol{R} = E[\boldsymbol{x}\boldsymbol{x}^T] = E\begin{bmatrix} x_k \\ x_{k-1} \\ x_{k-2} \end{bmatrix}\begin{bmatrix} x_k & x_{k-1} & x_{k-2} \end{bmatrix}$$

$$= E\begin{bmatrix} (x_k^2) & (x_k x_{k-1}) & (x_k x_{k-2}) \\ (x_{k-1} x_k) & (x_{k-1}^2) & (x_{k-1} x_{k-2}) \\ (x_{k-2} x_k) & (x_{k-2} x_{k-1}) & (x_{k-2}^2) \end{bmatrix}$$

The project is co-financed by the European Union

The **cross correlation vector**, **p**, for a 3 weight adaptive filter:

$$\boldsymbol{p} = E[d_k \boldsymbol{x}_k] = E\begin{bmatrix} d_k x_k \\ d_k x_{k-1} \\ d_k x_{k-2} \end{bmatrix} = \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix}$$

Minimum Mean Squared Error Solution

Hungary-Serbia
IPA Cross-border Co-operation Programme

Consider the MSE equation defining the so called MSE performance surface,

$$\zeta = E[e^2(k)] :$$

$$E[e^2(k)] = E[d^2(k)] + \boldsymbol{w}^T\boldsymbol{R}\boldsymbol{w} - 2\boldsymbol{w}^T\boldsymbol{p}$$

This equation is quadratic in the vector **w**. Hence there is **only one minimum value** of ζ, denoted MMSE (minimum mean square error) and which occurs at, $\mathbf{w}_{\mathbf{opt}}$.

Minimum Mean Squared Error Solution

Hungary-Serbia
IPA Cross-border Co-operation Programme

The MMSE is found from setting the (partial derivative) gradient vector ∇, to zero:

$$\nabla = \frac{\partial \zeta}{\partial \boldsymbol{w}} = 2\boldsymbol{R}\boldsymbol{w} - 2\boldsymbol{p} = \boldsymbol{0}$$

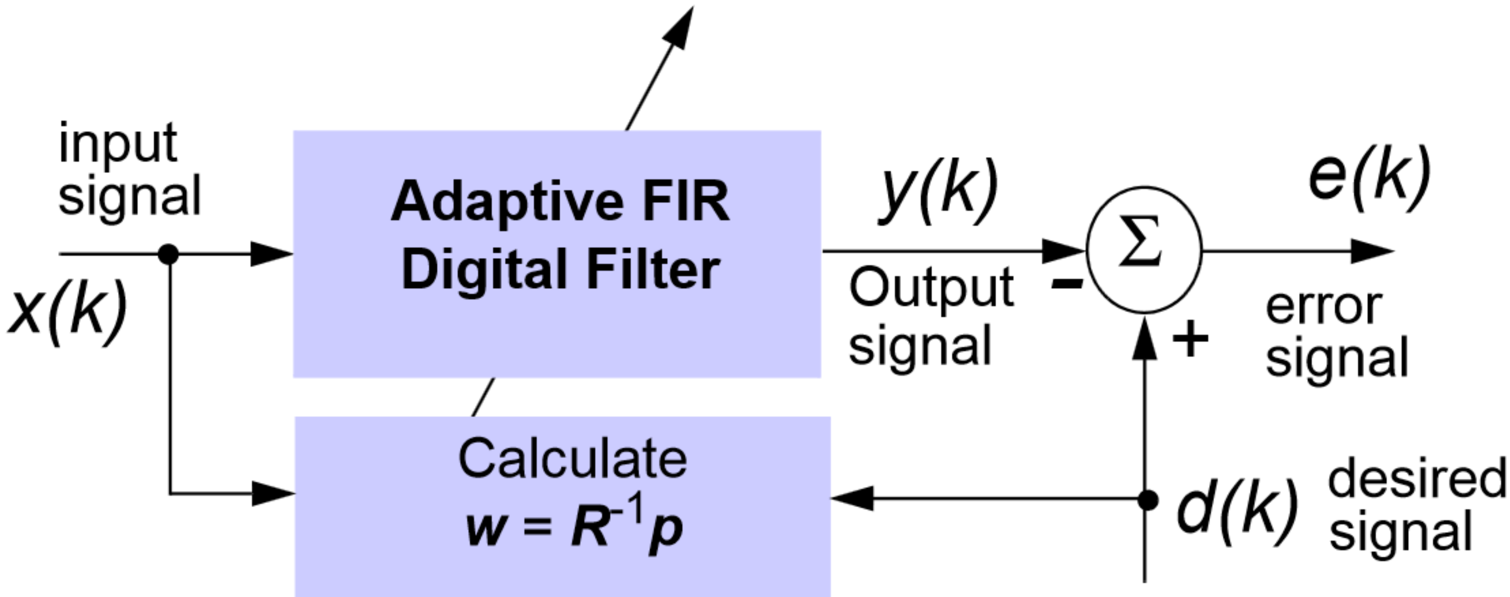
$$\Rightarrow \boldsymbol{w}_{opt} = \boldsymbol{R}^{-1}\boldsymbol{p}$$

This solution is termed the Wiener-Hopf solution (and is the optimum solution for the mean squared error minimization).

The project is co-financed by the European Union

Good neighbours creating common future

The Wiener-Hopf Solution

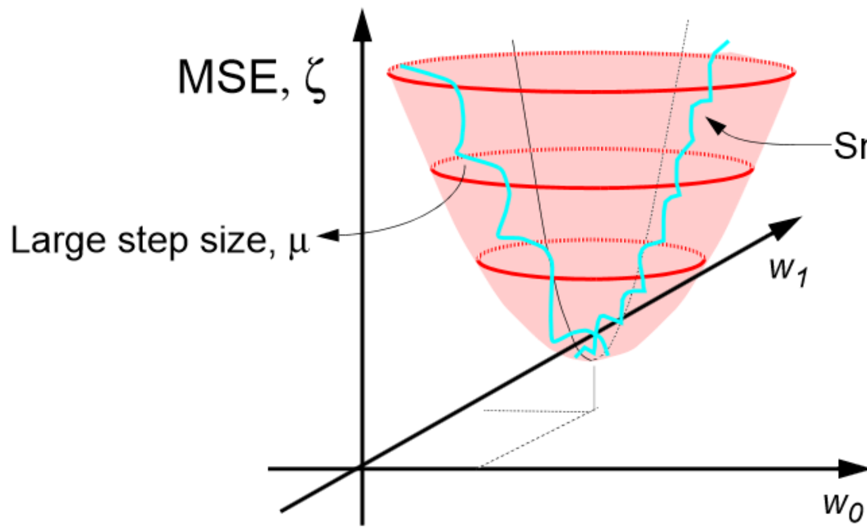


The Wiener-Hopf is **NOT** however **a useful real time algorithm** due to the heavy computation required, and if the statistics of x(k) and d(k) change then the $\mathbf{w}_{\mathbf{opt}}$ vector must be recalculated again.

Gradient Techniques

An iterative equation to find the MMSE can be performed by “jumping” down the inside of the performance surface in the direction of steepest gradient -∇(k).

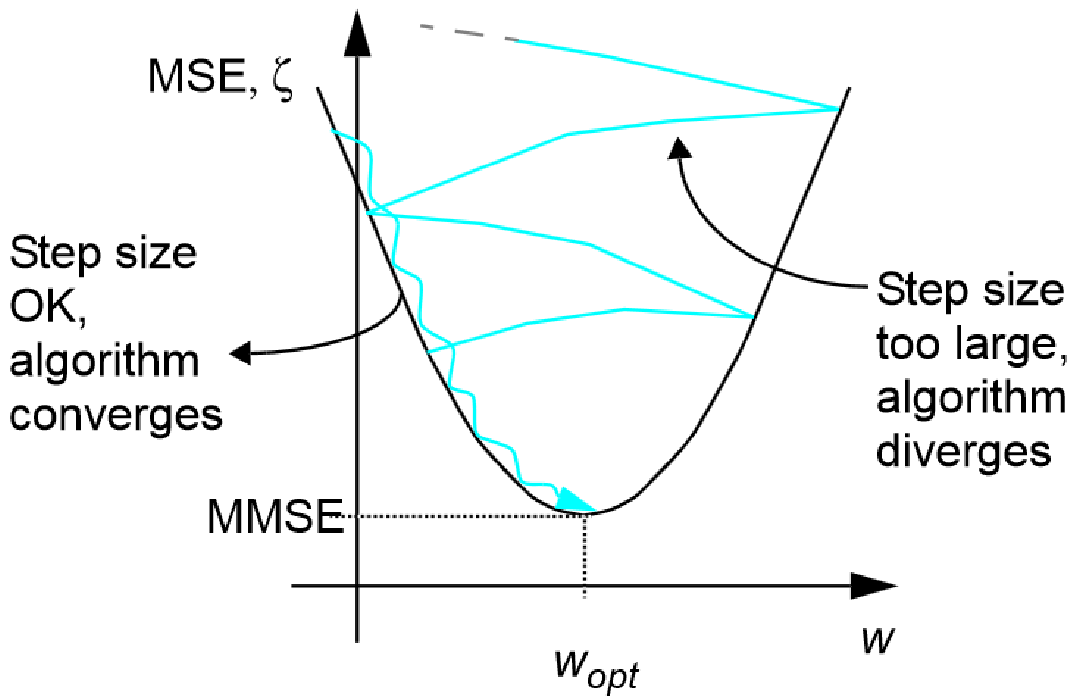
$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(-\nabla(k)) \quad :$$



$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

Step size

The step size, μ, controls the speed of adaption and also the stability of the (feedback) algorithm. If μ is too large then the algorithm will climb the inside of the parabola and hence be unstable (diverge). For example in a one weight case:



The LMS Algorithm

Hungary-Serbia
IPA Cross-border Co-operation Programme

Widrow suggested that instead of calculating the derivative of the mean squared error, use instead, the instantaneous squared error,
$e^2(k) = (d(k) - \boldsymbol{w}^T\boldsymbol{x}(k))^2$

Calculating gradient estimate , gives:

$$\hat{\nabla}(k) = \frac{\partial}{\partial \boldsymbol{w}(k)} e^2(k) = 2e(k)\left(\frac{\partial}{\partial \boldsymbol{w}(k)} e(k)\right) = -2e(k)\boldsymbol{x}(k)$$

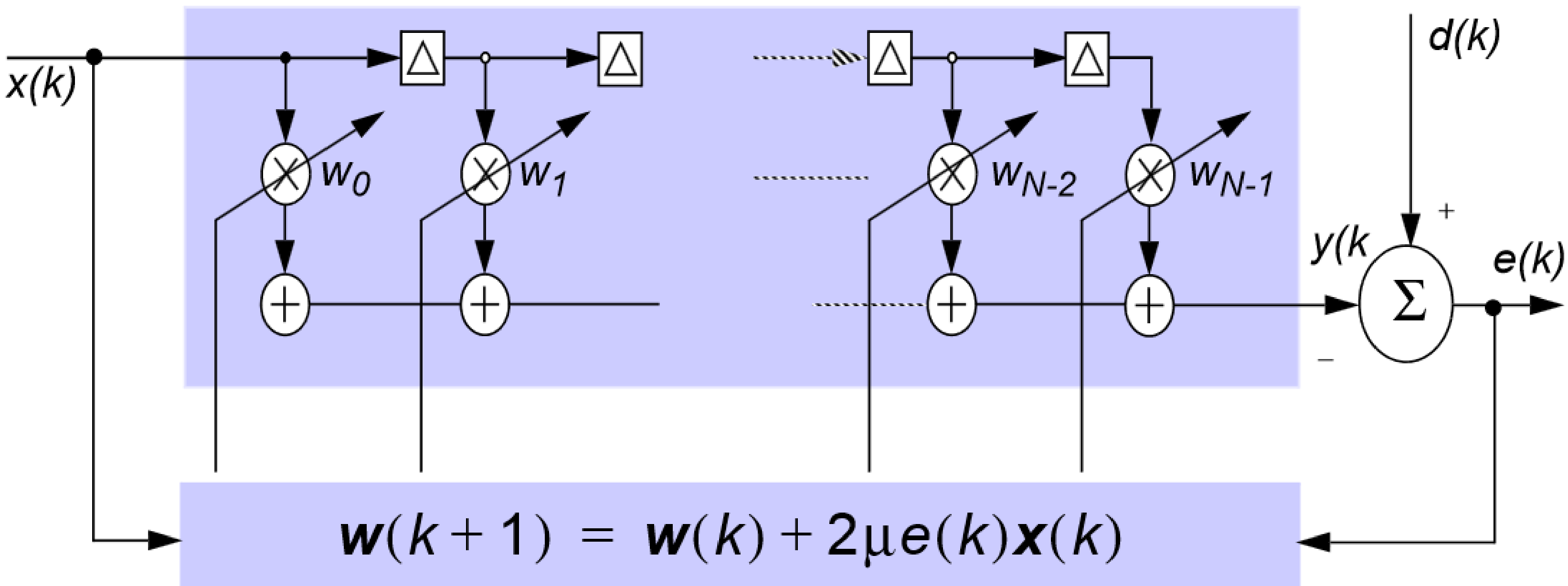
The LMS Algorithm

The LMS iterative weight update algorithm is:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu\left(-\frac{\partial}{\partial \boldsymbol{w}(k)}(e^2(k))\right)$$

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + 2\mu e(k)\boldsymbol{x}(k)$$

The LMS Algorithm



The LMS Algorithm

Hungary-Serbia
IPA Cross-border Co-operation Programme

The FIR filter requires N MACs (multiply-accumulates).

The LMS update requires N MACs.

2N MACs to implement each LMS algorithm iteration. Hence $2Nf_s$ MACs per second (where f_s is the application sampling frequency.)

The project is co-financed by the European Union

Good neighbours creating common future

LMS Stability

The stability of the LMS is dependent on the magnitude of the step size parameter, μ.

For convergence we require that:

$$0 < \mu < \frac{1}{\lambda_{max}}$$

Where λ_{max} is the largest eigenvalue of the **R** matrix.

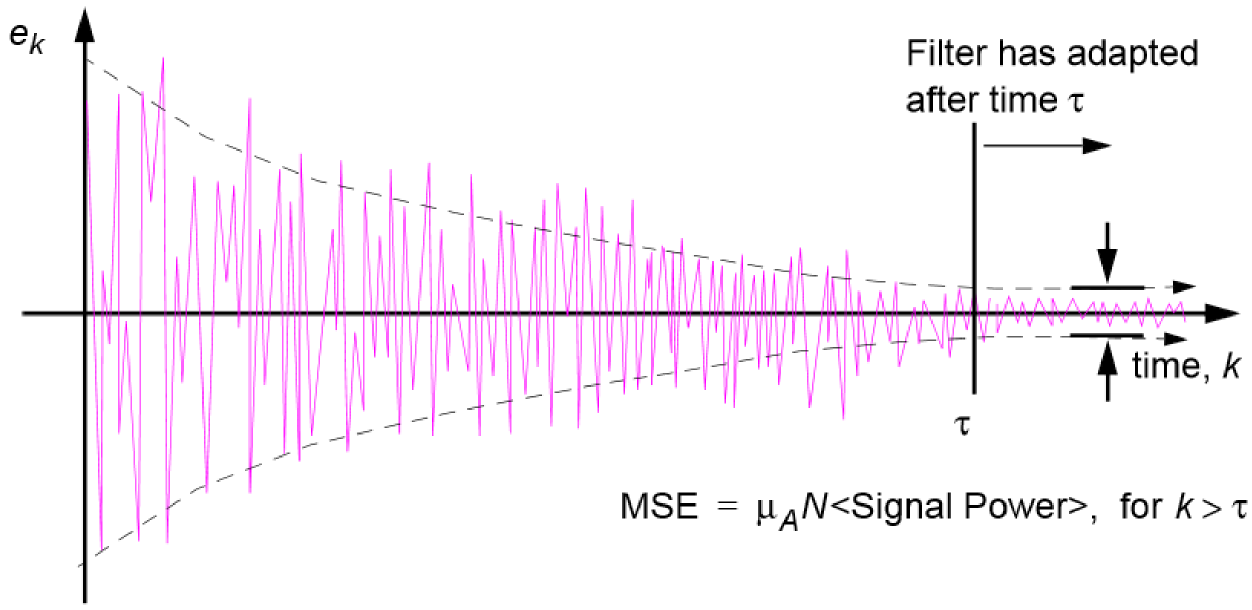
The previously derived bound is not convenient to calculate, and hence not particularly useful for practical purposes. However using the linear algebraic result that:

$$\mathrm{trace}[\boldsymbol{R}] = \sum_{n=0}^{N-1} \lambda_n \qquad \lambda_{max} \leq \mathrm{trace}[\boldsymbol{R}]$$

The project is co-financed by the European Union

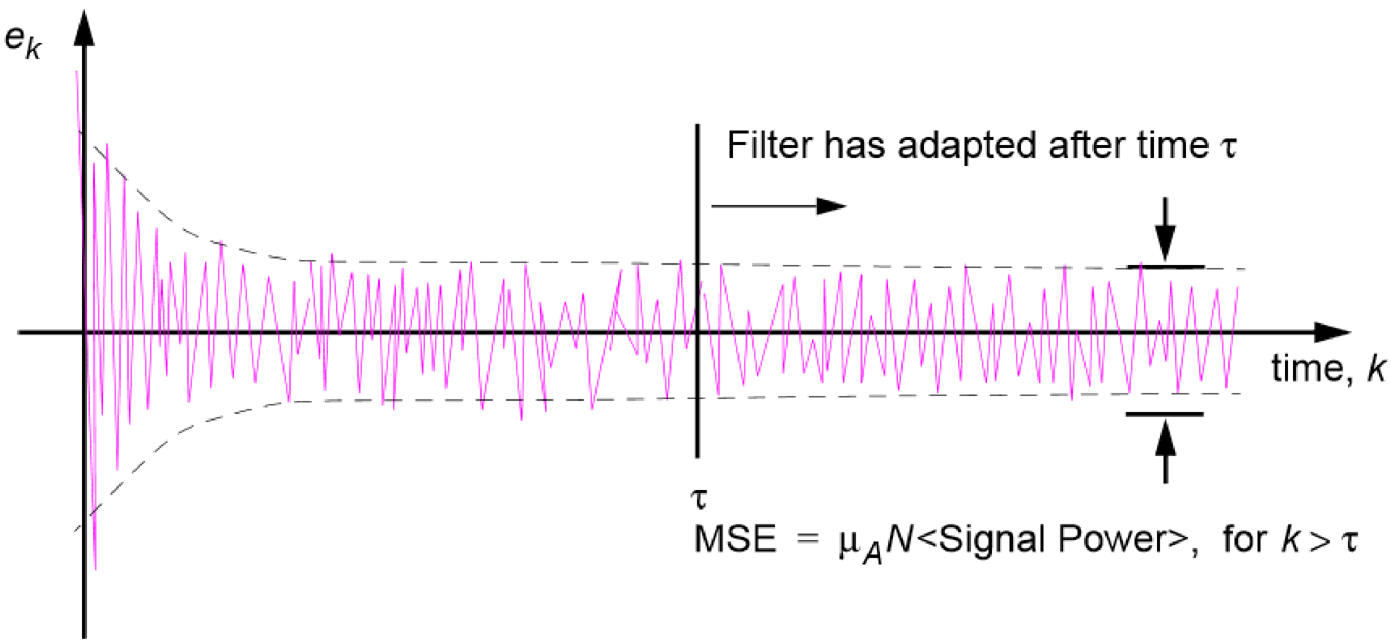
LMS Convergence - Small Step Size

For a small step size, μ_A, the LMS algorithm will converge slowly, and with a small misadjustment error:



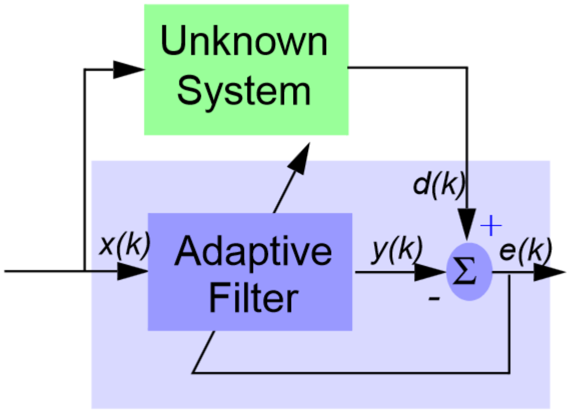
LMS Convergence - Large Step Size

For a large step size, μ_B, the LMS algorithm will converge quickly, and with a large misadjustment error:

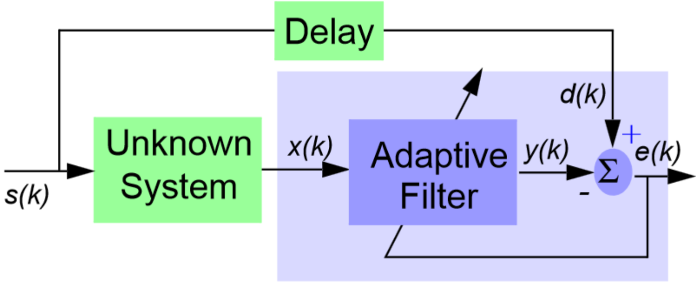


Application Examples

System Identification:



Inverse System Identification:

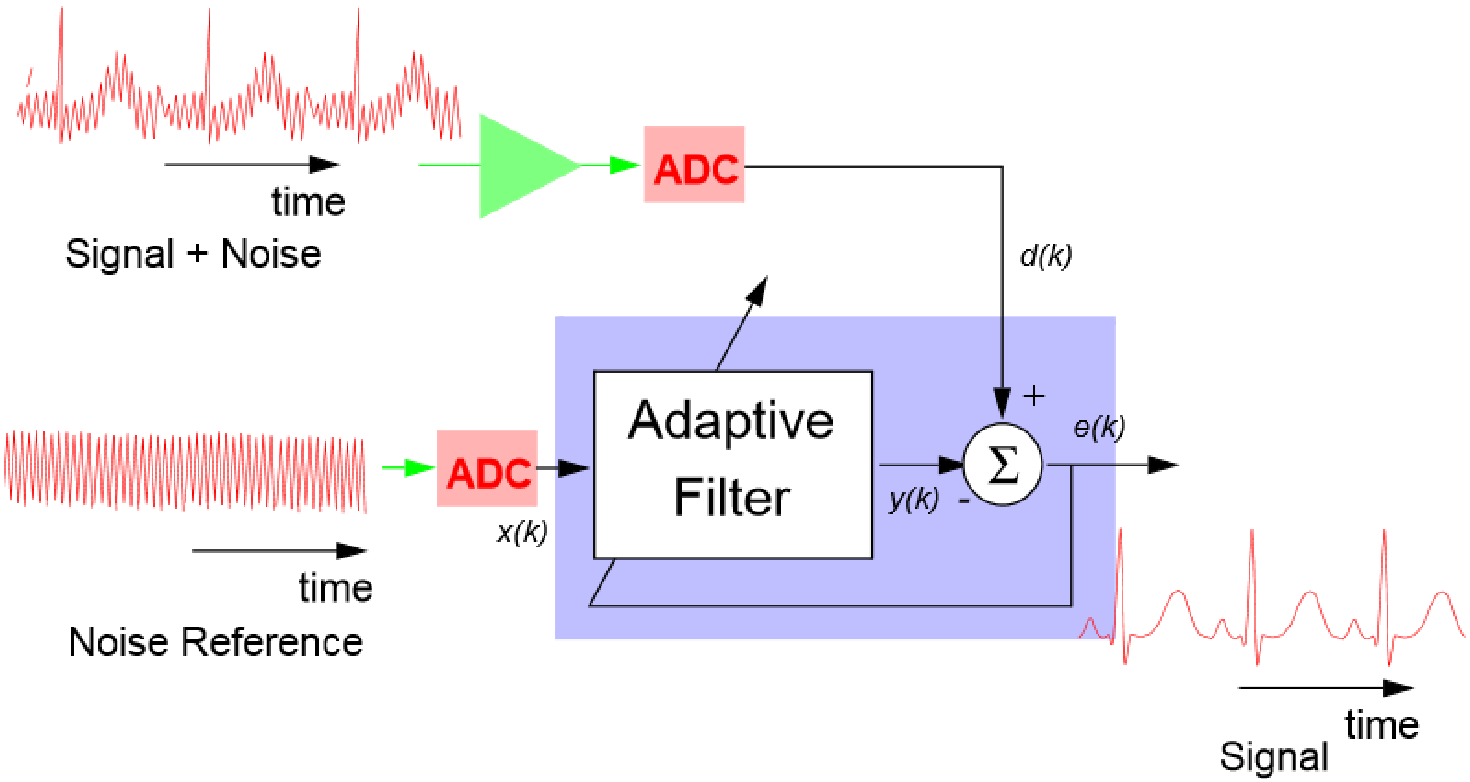


Noise Cancellation:

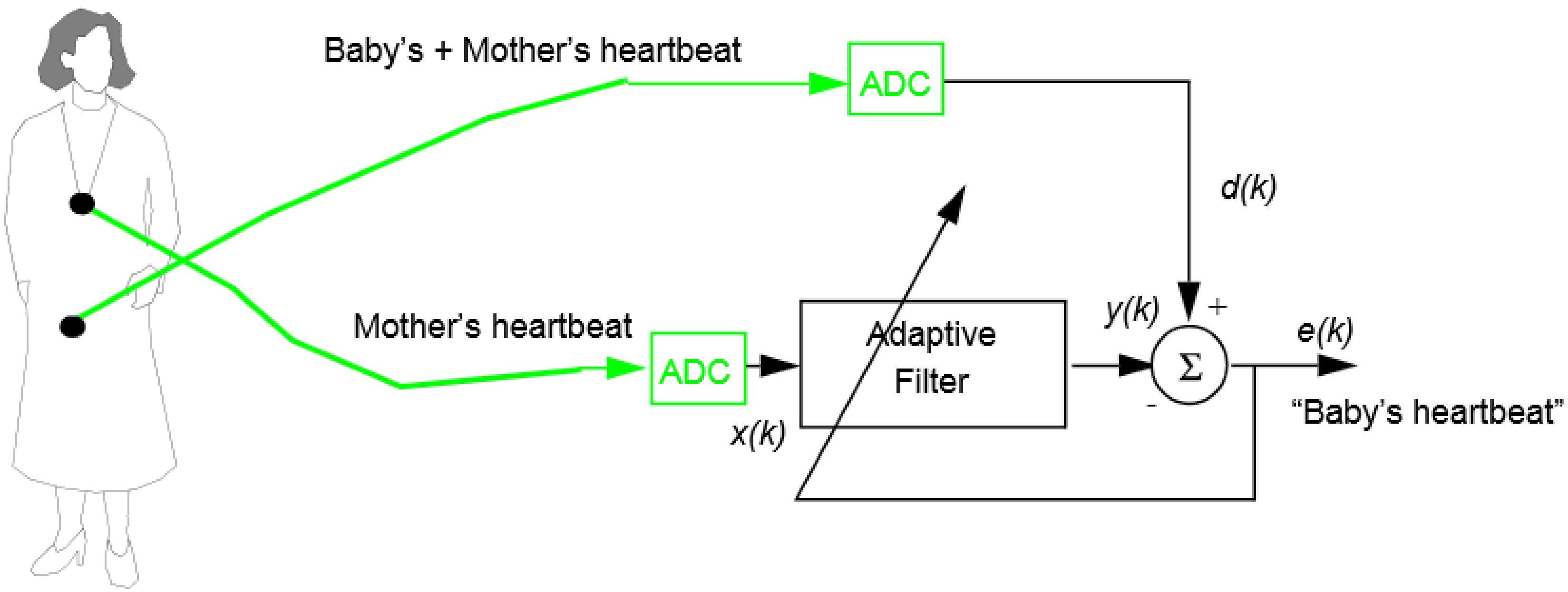
Mains Hum Noise Suppression

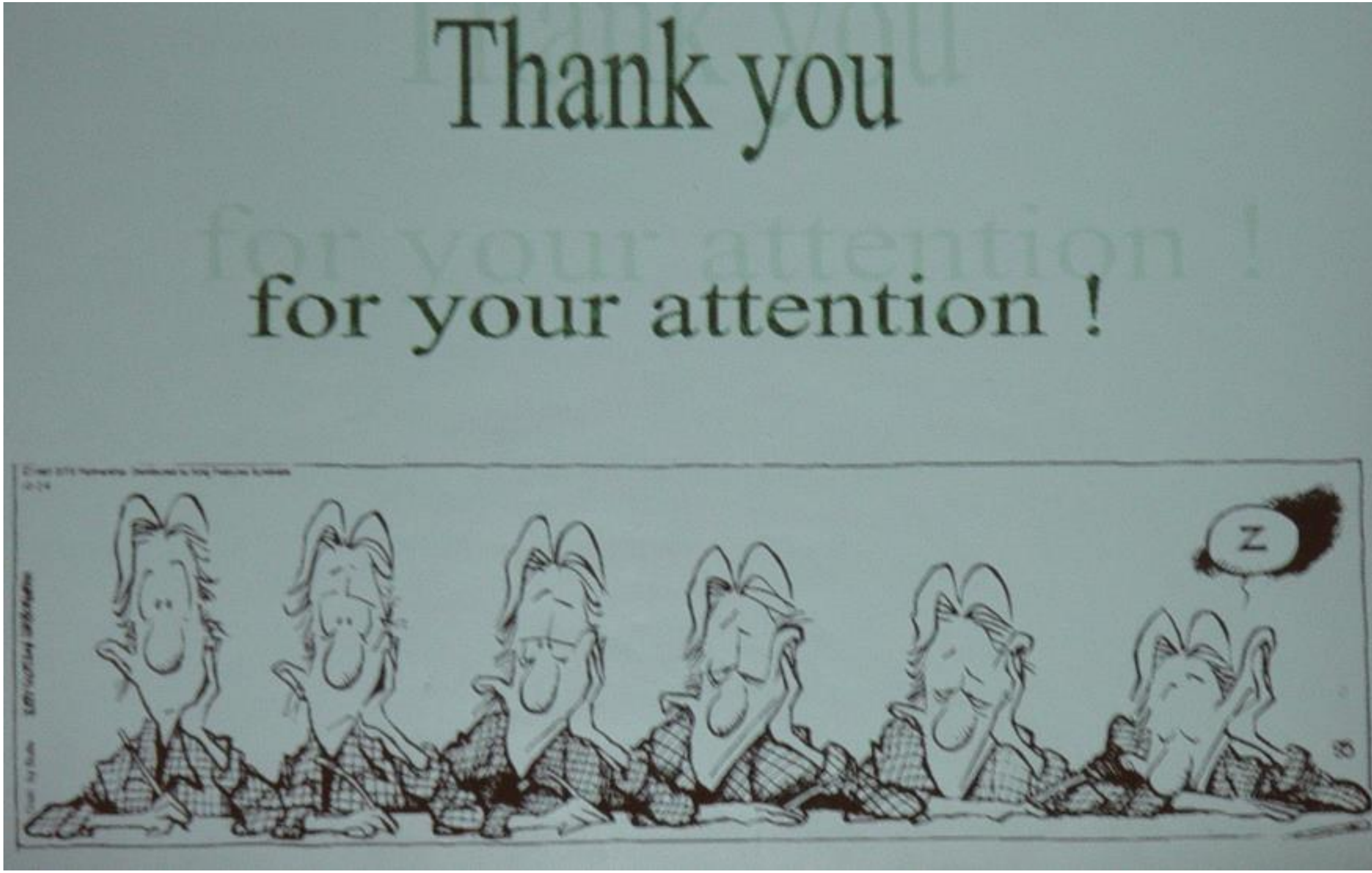
Hungary-Serbia
IPA Cross-border Co-operation Programme

Using a 50 Hz noise reference, electrical mains hum can be removed from the ECG (electrocardiograph, heartbeat signal).



The ECG main's hum noise canceller is a classic example first presented by Widrow et al. and frequently quoted in many texts and papers for example purposes.





- We woluld also like to thank for the Xilinx University Program

The project is co-financed by the European Union

Good neighbours creating common future