# Enhanced Spectral Estimation Using FFT in Case of Data Loss

András Palkó, László Sujbert
Budapest University of Technology and Economics,
Department of Measurement and Information Systems,
Budapest, Hungary
Email: palkoandras96@gmail.com, sujbert@mit.bme.hu

***Abstract*—Spectral estimation plays a significant role in engineering practice. Along with the spreading of sensor networks, more and more data are transmitted through unreliable channels which may lead to lost data. The most common method of spectral estimation uses FFT, but this requires the whole record without any data loss. This paper presents a new FFT-based method for the problem which can be used for coherent sampling. Its efficiency and accuracy is demonstrated via theoretical analysis, simulation and measurement results.**

***Keywords—spectral estimation, data loss, FFT***

## I. Introduction

Data loss is usually caused by communication problems. For example, in sensor networks data are often transmitted via radio channel, which is known to be faulty if interference or noise occurs. Data loss can mean missing samples, invalid samples (e.g. ADC overdrive) or synchronization issues.

In engineering practice, spectral estimation plays an important role. If this is the measurement task and some of the samples are lost, then data loss becomes a serious problem. The spectral estimate of sampled signals can be calculated via DFT:

$$X(k) = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}nk} \, (n, k = 0, 1, \dots, N-1). \qquad (1)$$

The DFT can effectively be evaluated by the Fast Fourier Transform (FFT). In order to get the value of any point of the DFT, the whole record is needed, without data loss. An obvious solution is to wait for a complete record, but the number of samples which are needed can be the multiple of the DFT record size, which is unacceptable in most applications, where linear or exponential averaging is applied to reduce the measurement noise.

There are available methods which can be used, e.g. Lomb-Sclarge [5][6] or autoregressive analysis [7]. For our research, computationally effective and robust methods are preferred, two of them will be presented briefly. The first one is the extension [1] of the resonator-based observer (RBO) [2]. The second one [3] utilizes the FFT because of its particular efficiency in spectral estimation. This modifies the FFT blocks by zero padding them (here: replacing the samples with zeros) from the first lost sample.

A question arises why we don't replace only the lost samples with zeros. Replacing lost samples results in an additive noise, which can make difficult or impossible to find low magnitude spectral components. The aim of the methods is to reduce the power of this noise.

When data loss arises, time-domain interpolation (e.g., linear from nearest neighbors or Lagrange) is one of the first ideas to consider. However, interpolation methods cause a linear distortion in the spectrum (e.g., linear interpolation distorts the original spectrum with a sinc-squared function), in spite the additive noise of "replacement with zeros" method. Every method which uses surrounding samples gives a kind of memory. This makes the spectrum variable (even if it was constant) which we want to avoid.

In the paper a new method is presented which can be used effectively if the sampling is coherent. It provides the same accuracy as the RBO, but with much less computational complexity.

## II. Preliminaries

### A. Mathematical Description of Data Loss

#### 1) Indicator Function

Data loss can be modeled with an availability indicator function:

$$K_n = \begin{cases} 1 & \text{, if sample is available at n} \\ 0 & \text{, if sample is lost at n} \end{cases} \qquad (2)$$

Available samples will also be termed as processed samples. Using this we can define the data loss rate:

$$\gamma = \mathrm{P}\{K_n = 0\} \qquad (3)$$

where $\mathrm{P}\{\cdot\}$ is the probability operator. We can describe a signal with lost samples as

$$x_n = K_n x_{0,n}, \qquad (4)$$

where $x_{0,n}$ is the original signal (without data loss).

#### 2) Data Loss Models

There are different data loss models with different indicator functions. Random independent data loss is the simplest. It can be defined as

$$K_n = \begin{cases} 1 & \text{with } \mu = 1 - \gamma \text{ probability} \\ 0 & \text{with } \gamma \text{ probability} \end{cases} \text{ for } \forall n. \qquad (5)$$

In random block-based data loss, a block is formed from each $M$ samples. The same applies to the blocks as in the random, independent case:

$$\alpha_k = \begin{cases} 1 & \text{with } \mu = 1-\gamma \text{ probability} \\ 0 & \text{with } \gamma \text{ probability} \end{cases} \text{ for } \forall k$$
$$K_{kM+n} = \alpha_k \text{ for } \forall k\ \forall (n \in \{0, 1, \dots, M-1\}). \quad (6)$$

In Markov-chain based data loss, the indicator function is generated as the state of a Markov-chain, see Fig. 1.

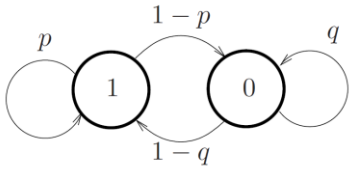


Fig. 1. Two-state Markov-chain. State 1: $K_n = 1$, state 0: $K_n = 0$

### B. *Spectral Estimation Using the Resonator-Based Observer [2]*

The resonator-based observer was designed to follow the state variables of the conceptual signal model [2] which generates signals according to their Fourier-series. This way, the observed state variables can be the Fourier coefficients or their rotating versions.

This structure has been modified to be able to handle data loss [1]. The main idea is to modify the conceptual signal model to generate signals with lost samples, then design a state observer for this system. The conceptual signal model can be described as:

$$\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n$$
$$y_n = K_n \mathbf{c}\mathbf{x}_n \quad (7)$$
$$\mathbf{A} = \langle z_i \rangle \quad \mathbf{c} = [1, 1, \dots, 1] \quad z_i = \mathrm{e}^{\mathrm{j}2\pi f_i}$$

where $\langle \cdot \rangle$ is the diagonal matrix formation operator, $\mathbf{x}_n$ is the state vector in the time step $n$, $y_n$ is the output signal and $f_i$ is the relative frequency of the $i$th component. Fourier-coefficients can be extracted from the state vector as:

$$\langle z_i^{-n} \rangle \mathbf{x}_n = X_i \quad (8)$$

where $X_i$ is the column vector of the Fourier-coefficients. The equation of the observer is the following:

$$\hat{\mathbf{x}}_{n+1} = \mathbf{A}\hat{\mathbf{x}}_n + \mathbf{g}K_n(y_n - \hat{y}_n) = \mathbf{A}\hat{\mathbf{x}}_n + \mathbf{g}K_n e_n$$
$$\hat{y}_n = \mathbf{c}\hat{\mathbf{x}}_n \quad (9)$$

where $\hat{\mathbf{x}}_n$ is the estimated state vector, $\mathbf{g}$ is the feedback vector, $\hat{y}$ is the estimated signal and $e_n$ is the estimation error. Fig. 2. shows the RBO for signals with lost samples.

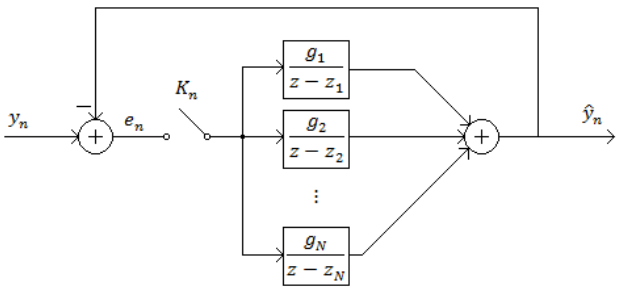


Fig. 2. Resonator-based observer for signals with lost samples.

It is worth noting that the estimation error is multiplied by the indicator function which can be interpreted in two ways. First, if the sample is lost, measurement update isn't performed. Second, at the lost samples the structure acts as its estimate was accurate.

RBO can be used for coherent and incoherent sampling if the frequencies of the signal components are known. The characteristic polynomial can be set arbitrarily with the feedback vector $\mathbf{g}$, which implies, for example, exponential averaging can be done without extra computation. The structure can be applied in real-time and offers fairly precise spectral estimation even at high data loss rate. The main drawback is the complexity: RBO is a quadratic algorithm, while the complexity of FFT-based algorithms is linearithmic ($O(N \log N)$).

### C. *Spectral Estimation Using FFT with Zero Padding [3]*

The procedure of spectral estimation using FFT with zero padding for a record is the following:

1. $x_n$ and $K_n$ $(n = 0, 1, \dots, N-1)$ input FFT record and indicator function are given. $N$ is the size of FFT.
2. $L = \min\{N;\ n, \text{where } K_n = 0\}$ is the position of the first lost sample in the block.
3. If $L \leq N_{min}$, the block is discarded. ($N_{min} \geq \frac{N}{4}$ is recommended.) Else, a new indicator function is generated:
$$K'_n = \begin{cases} 1 & \text{, if } n \leq L \\ 0 & \text{, if } n > L \end{cases} \quad (10)$$
4. The signal is multiplied by $\frac{N}{L}$ and the new indicator function is applied:
$$y_n = \frac{N}{L} x_n K'_n \quad (11)$$
5. DFT of $y_n$ is computed, with a window function for $L$ samples, the result is the spectral estimate of the record.

Fig. 3. shows the procedure graphically.

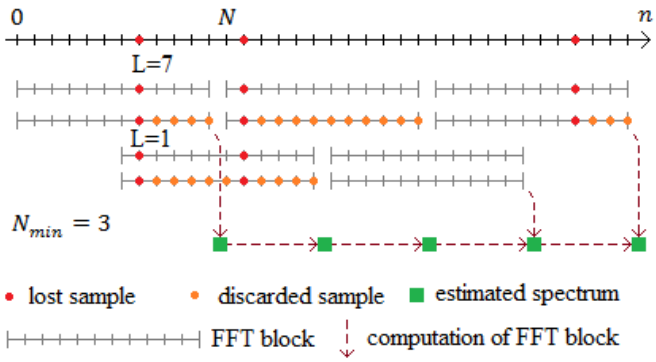


Fig. 3. Spectral estimation with zero padding FFT method

Here overlapping FFT-blocks are used. The aim of this is to make estimate converge faster. Based on [4], the maximal recommended overlap ratio is 75%.

The final spectral estimate is computed from the FFTs of the blocks with an averaging procedure. Averaging can be done both with magnitude spectra and complex spectra. Magnitude spectra can be averaged both for coherently and

incoherently sampled signals, but complex spectra can only be averaged in the case of coherent sampling.

The main benefit of this method is its linearithmic complexity, which makes it easy to apply in real-time measurements. It can be used effectively for searching dominant components even with high data loss rates. If we don't need to know the phase information, this method can be used for both coherently and incoherently sampled signals.

## III. Proposed Method: Replacement FFT

### A. Description of the Method

The idea is to use FFT for spectral estimation, and to try to achieve the same behavior at lost samples as RBO has. This means that at positions with lost samples, the method needs to behave as its estimate was accurate. This can be done by computing a replacement value for each lost sample via IDFT. The procedure is the following:

1. Wait for the first $N$ samples (FFT block), substitute lost samples with zeros and compute the DFT of the block, the result is $FFT_1$.
2. Wait for the next FFT block.
3. For the positions of lost samples, compute the replacement value with IDFT from $FFT_1$.
4. Compute the DFT of the new block, the result is $FFT_2$.
5. After applying the appropriate phase shift on $FFT_2$, compute the new spectral estimate from $FFT_1$ and $FFT_2$ via exponential averaging, and store it in $FFT_1$.
6. If the measurement is not over, continue from the second step.

Fig. 4. shows the method graphically.

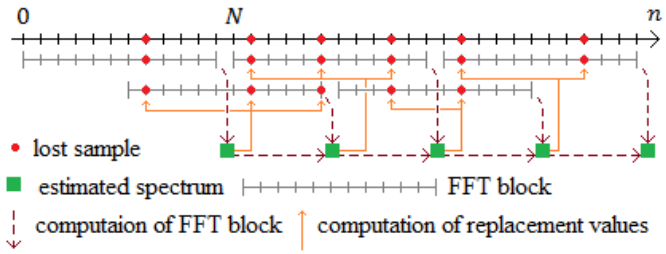


Fig. 4. Spectral estimation with replacement FFT method

It should be noted that in the first FFT block the lost samples are substituted with zeros, thus a noise is added to the initial estimate. It depends on the actual application if it is a problem or not, because exponential averaging reduces this noise over time. If it is a problem, we can wait for the first block without lost samples.

With the phase shift on $FFT_2$ the phase of the fundamental harmonic (in the DFT base functions) is made equal in $FFT_1$ and $FFT_2$. Similar phase shift is necessary at the computation of the replacement values. These can be done only for coherently sampled signals. That's why this method is not applicable for incoherently sampled signals. The phase shifts can be done automatically by implementing the method with a circular buffer.

### B. Computation of the Replacement Values

The computation of a single replacement value needs $O(N)$ operations using IDFT. In an $N$ samples long block there are on average $\gamma N$ lost samples, so the replacement values can be computed with $O(\gamma N^2)$ operations. If the data loss rate isn't small enough, these operations make the method complexity quadratic. In this case, IFFT can be used to compute a whole replacement block and use only the necessary positions of it. Of course, this solution needs more memory and at low data loss rates it is slower than individual computation.

It can be easily suspected that there is a data loss rate, where the two procedures have the same computational requirement, this is called critical data loss rate ($\gamma_{crit}$). Below it, IDFT, above it, IFFT needs less operations.

An individual replacement value can be computed with $4N$ real operations using IDFT, for the whole block we need $4\gamma N^2$ steps. Assuming the usage of radix-2 IFFT, the replacement block can be computed with $\frac{N}{2}\log_2 N$ complex multiplications and $N\log_2 N$ complex additions, which means $5N\log_2 N$ real operations. We also need to check every position if there was data loss ($N$ operations) and replace the lost samples with the computed values ($\gamma N$ operations). In total, IFFT-based replacement has $5N\log_2 N + N + \gamma N$ steps.

At the critical data loss rate, the two procedures have the same number of operations, from which we obtain

$$\gamma_{crit} = \frac{5\log_2 N + 1}{4N - 1}. \tag{12}$$

Considering that this is only an estimate (e.g. different operations need different number of machine cycles, SIMD instruction execution, etc.), we can rewrite (12) as

$$\gamma_{crit} \approx \frac{5\log_2 N}{4N}. \tag{13}$$

The evaluation of (13) for different $N$ values can be found in Table 1.

TABLE I. Critical Data Loss Rates

| $N$ | $\gamma_{crit}$ | $N\gamma_{crit}$ | $N$ | $\gamma_{crit}$ | $N\gamma_{crit}$ |
|---|---|---|---|---|---|
| 16 | 31,250% | 5 | 4096 | 0,366% | 15 |
| 64 | 11,719% | 7,5 | 16384 | 0,107% | 17,5 |
| 256 | 3,906% | 10 | 65536 | 0,031% | 20 |
| 1024 | 1,221% | 12,5 | 262144 | 0,009% | 22,5 |

The first and fourth column show the size of the FFT, the second and fifth ones show the critical data loss rate and the third and last ones show the critical number of lost samples in a block. Based on this, if the data loss rate and the FFT size are known in advance, we can decide which replacement procedure is faster. If data loss rate varies within a large interval which contains $\gamma_{crit}$ and speed is critical, it should be taken into consideration to count the number of lost samples in each block and use the appropriate method. If this means too much overhead, IFFT-based replacement should be used.

## IV. Simulations and Measurement Results

The proposed method was examined and compared with RBO and zero padding FFT via simulations and

measurements. Some results are presented to demonstrate the features of the proposed method.

### A. Simulation Results

Simulation parameters: $N$=256 FFT size, $N_{min}$=$N$/4 minimal valid block size for zero padding FFT, 75% overlap ratio, exponential averaging for all three methods with the same time constant (1000), square wave with 1/64 relative frequency input signal with additive white noise (SNR=20 dB), $L$=50*$N$ simulation time, random independent data loss with 0.1% data loss rate.

Fig. 5. shows the settling of the spectral estimation. The error was formed as the Euclidean ($L_2$) norm of the difference of the original and the estimated magnitude spectra. It must be noted that the original spectrum was calculated without windowing for RBO and replacement FFT but with Hanning window for zero padding FFT. The reason of this asymmetry is that in zero padding FFT windowing should be used, but it's problematic to use a window function with RBO. Replacement FFT behaves similarly to RBO and provides accurate estimate only with coherent sampling, that's why it doesn't need windowing. The Euclidean norm of the noise magnitude spectrum (Noise FFT) is displayed for comparison. In the bottom, the data availability ($K$) is shown: high level means available, low level means lost samples.

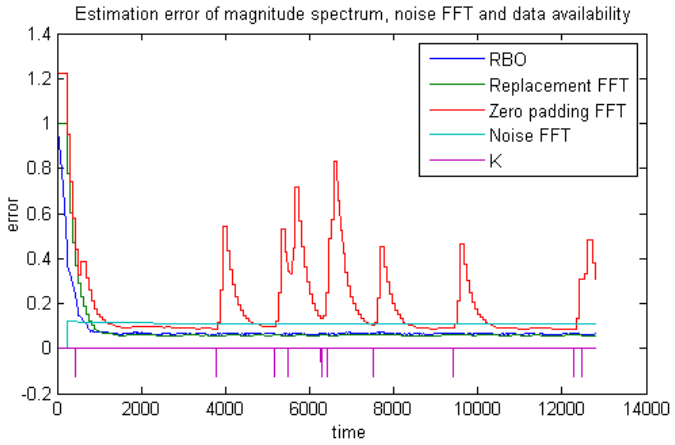


Fig. 5. Simulation results of coherently sampled square wave with 0.1% data loss rate

It can be seen that when there are complete blocks, all the three methods give fairly accurate results. The precision of the replacement FFT is the same as that of the RBO. Data loss leads to a peak in the error of zero padding FFT, but the estimates of replacement FFT and RBO are unaffected.

### B. Measurement Results

Measurements were conducted with a Sharc ADSP-21364 Ez-kit Lite DSP board. A noise generator has been used to independently control the data availability.

Fig. 6. shows the measurement results of the processing of a square wave sampled in a special way: the sampling is incoherent for the first harmonic, but coherent for the third harmonic. That's why the third harmonic and its higher harmonics are measured correctly with all the methods. In this measurement, FFT methods were compared with 0.1% data loss rate and $N$=4096.

The other components are measured incorrectly with the replacement FFT even with applying a window function, because the problem arises from averaging complex spectra. Examining the results of zero padding FFT, it can be stated that windowing should be used with the method.

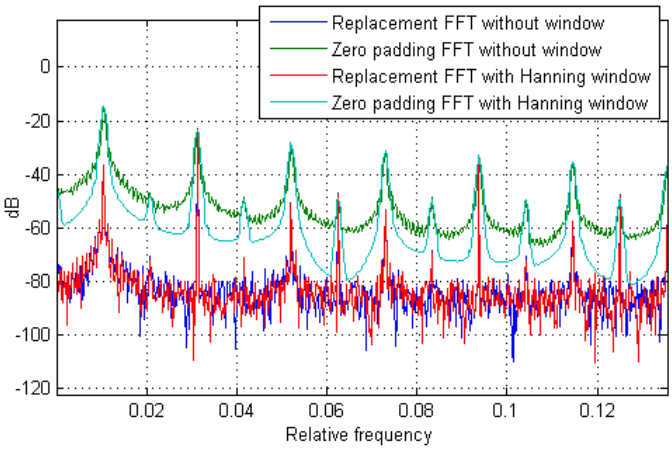


Fig. 6. Measurement results of incoherently sampled square wave with 0.1% data loss (zoomed)

Other measurements have shown that RBO and replacement FFT yield the same results.

## V. Conclusion

In this paper, a new method of spectral estimation in the case of data loss was introduced and examined. This method calculates a replacement value for each lost sample from the latest estimate. The replacement FFT can be used effectively for coherently sampled signals, even at high data loss rates and provides distortion-free spectral estimate. This method has linearithmic complexity which makes it beneficial for real-time applications.

## References


[1] Gy Orosz, L. Sujbert, G. Péceli, „Analysis of Resonator-Based Harmonic Estimation in Case of Data Loss", IEEE Transactions on Instrumentation and Measurement 62:(2) p. 510-518., 2013

[2] G .Péceli, „A Common Structure for Recursive Discrete Transforms", IEEE Transactions on Circuits and Systems 33:(10) p. 1035-1036., 1986

[3] L. Sujbert, Gy. Orosz, „FFT-based Spectrum Analysis in the Case of Data Loss," IEEE Transaction on Instrumentation and Measurement,vol. 65, no. 5, pp. 968–976, May 2016.

[4] F. Harris, "On the use of Windows for Harmonic Analysis with the Discrete Fourier Transform", In Proc. IEEE, vol. 66, no. 1, pp. 51-83, Jan. 1978.

[5] N. R. Lomb, „Least-Squares Frequency Analysis of Unequally paced Data", Astrophysics And Space Science 39 (1976), pp. 447-462

[6] J. D. Sclarge, „Studies in Astronomical Time Series Analysis. III. Fourier Transforms, Autocorrelation Functions, and Cross-Correlation Functions of Unevenly Spaced Data", Theoretical Studies Branch, Space Science Division, NASA-Ames Research Center, vol. 343 pp.874-887, No.2, 1989

[7] P. M. T. Broersen, S. de Waele, R. Bos, „Application of Autoregressive Spectral Analysis to Missing Data Problems", IEEE Transactions On Instrumentation And Measurement, vol. 53., no. 4, August 2004