ADAPTATIVE SIGNAL SAMPLING AND SAMPLE QUANTIZATION FOR RESOURCE-CONSTRAINED STREAM PROCESSING

Deepak S. Turaga, Olivier Verscheure, Daby M. Sow and Lisa Amini
IBM T.J. Watson Research Center, 19 Skyline Drive, Hawthorne, NY, USA

Keywords: Remote Health Monitoring, ECG Compression, Low-complexity, Non-Uniform Sampling, Quantization.

Abstract: We propose a low-complexity encoding strategy for efficient compression of biomedical signals. At the heart of our approach is the combination of non-uniform signal sampling together with sample quantization to improve the source coding efficiency. We propose to jointly extract and quantize information (data samples) most relevant to the application processing the incoming data in the backend unit. The proposed joint sampling and quantization method maximizes a user-defined utility metric under system resource constraints such as maximum transmission rate or encoding computational complexity. We illustrate this optimization problem on electrocardiogram (ECG) signals, using the Percentage Root-mean-square Difference (PRD) metric as the utility function measuring the distortion between the original signal and its reconstructed (inverse quantization and linear interpolation) version. Experiments conducted on the MIT-BIH ECG corpus using the well-accepted *FAN* algorithm as the non-uniform sampling method show the effectiveness of our joint strategy: Same PRD as '*FAN* alone' at half the data rate for less than three times the (low) computational complexity of *FAN* alone.

1 INTRODUCTION

Remote Health Monitoring is an emerging technology allowing medical practitioners to extend their services to patients outside of traditional hospital settings. Common remote health monitoring systems are leveraging pervasive devices such as cellular phones to collect biomedical readings on patients and relay the data to servers while being non-intrusive and not restricting the mobility of patients (Mohomed et al., 2006). This usage of pervasive devices differ significantly from traditional client server usage models where the pervasive device acts as a client *receiving* data from a more powerful server. In the current model, the roles are reversed. Pervasive devices are used to *stream* data to back-end servers. Their resource scarceness creates interesting research challenges dictating the need for efficient, low complexity signal encoding schemes. This work proposes a generic method for streaming continuous signals under very strict resource constraints while minimizing the loss in valuable information the original signals carry.

While our method is applicable to a wide variety of signals, we describe it in the context of efficient, low complexity compression of electrocardiogram (ECG) signals. An ECG signal provides essential information to the cardiologist and is used for both monitoring and diagnostic purposes. An ECG monitoring device essentially measures the electrical impulses that stimulate the heart to contract. Between 125 and 500 sample points are collected every second, each coded on 8 or 12 bits (Nygaard et al., 2001). Thus, a single-lead uncompressed ECG signal requires between 1 kbps and 6 kbps of sustained wireless bandwidth. Any application based on wireless transmission of even moderate amounts of data must deal with the reality that usage of wireless spectrum will always incur some monetary cost. Efficient, low complexity compression is thus crucial to make remote health monitoring via low-end pervasive devices a reality.

The main goal of any compression technique is to achieve maximum data volume reduction while preserving the significant signal morphology features upon reconstruction (Jalaleddine et al., 1990). In ECG signal compression algorithms the goal is to achieve a minimum information rate, while retaining the relevant diagnostic information in the reconstructed signal. Compression techniques for ECG waveforms can be broadly classified into two main groups: direct time-domain techniques (Barr, 1988; Cox et al., 1968), and transform-domain techniques (Bradie, 1996; Hilton, 1997; Addison, 2005). Transform-based methods (e.g., wavelet-based) usually outperform time-domain techniques but require

S. Turaga D., Verscheure O., M. Sow D. and Amini L. (2008).
ADAPTATIVE SIGNAL SAMPLING AND SAMPLE QUANTIZATION FOR RESOURCE-CONSTRAINED STREAM PROCESSING.
In *Proceedings of the First International Conference on Bio-inspired Systems and Signal Processing*, pages 96-103
DOI: 10.5220/0001067400960103

a computational power beyond what a mainstream pervasive device can handle. Instead, well-accepted time-domain techniques, such as FAN (Barr, 1988) and AZTEC (Cox et al., 1968), rely on simple heuristics so as to non-uniformly sample the original waveform and retain only those data samples that contribute the most to the quality of the reconstructed (interpolated) signal.

Another well-known compression strategy is *quantization*. There are two types of quantization. Vector quantization, where the input symbols are gathered together in groups called vectors and processed to give the output, and scalar quantization, where each input symbol is treated separately in producing the output. Scalar quantization has a low computational complexity, is easy to implement and can achieve reasonably good compression performance if applied properly. There has been recent interests in the scientific community to design schemes performing jointly both quantization and uniform sampling in order to match the underlying system resource constraints (Derpich et al., 2006). Uniform sampling involves discarding samples of the data regularly to reduce the data rate. While uniform sampling can reduce the stream rate appropriately it does not guarantee the retention of all samples of interest (features), especially when the frequency characteristics of the signal are not well-behaved, which is clearly the case for ECG waveforms.

This work investigates the benefit of jointly performing non-uniform sampling (e.g., FAN or AZTEC) and quantization operations in the context of remote health monitoring. The paper is organized as follows: Section 2 introduces some notations and describes, in generic terms, the concept of joint non-uniform sampling and quantization. This concept applied to signal compression is the subject of Section 3, while Section 4 formulates the problem specifically for ECG compression under resource constraints using FAN (Barr, 1988) as the non-uniform sampling technique. The problem is posed as an optimization problem. The optimization problem is solved in Section 5. Finally, our strategy is validated in Section 6. And, Section 7 gives concluding remarks.

2 SIGNAL COMPRESSION: NON UNIFORM SAMPLING AND QUANTIZATION

Let $x[k], 0 \leq k < N$ denote a discrete time signal represented with b_u bits per sample.

2.1 Non Uniform Sampling

Non uniform sampling of $x[k]$ extracts $N_{SOI} \leq N$ samples of interests (SOI) from $x[k]$. We denote such sampling by the operator $\mathcal{S} : x[k] \rightarrow x[k_i]$ where k_i corresponds to the location of the retained samples of interest. The operator $\mathcal{S}$ is often lossy, and only an approximation to the original signal $x_r[k]$ may be recovered by interpolating $x[k_i]$ appropriately. If, after sampling, we retain N_{SOI} out of N samples, the achieved compression ratio is $\frac{N_{SOI} b_u}{N b_u}$, corresponding to a rate $\frac{N_{SOI} b_u}{N}$ bits per sample. Additionally, in the compressed rate we also need to include the bits required to encode the locations of the retained samples, i.e. an additional b_{loc} bits per sample. The selectivity of the sampling operator $\mathcal{S}$ is controlled by a sampling sensitivity parameter ε, with low values of ε corresponding to low selectivity, i.e. most samples from $x[k]$ are retained. To explicitly indicate the dependence of $\mathcal{S}$ on ε, we present it as $\mathcal{S}_\varepsilon$.

2.2 Quantization

Quantization is another well known lossy technique used to reduce the signal rate, when applications can tolerate the resultant distortion. We denote the quantization operator as $\mathcal{Q} : x[k] \rightarrow \hat{x}[k]$ where $\hat{x}[k]$ uses $b_q < b_u$ bits per sample, thereby reducing the average data rate of the stream by a factor $\frac{b_u}{b_q}$.

Given a periodic signal such as ECG, with relatively stationary probability density function (under known context, i.e. physical activity, health state etc.) the quantizer sensitivity is controlled only by the number of desired reconstruction levels[1] $L = 2^{b_q}$. As before, to explicitly indicate the dependence of $\mathcal{Q}$ on L, we represent it as $\mathcal{Q}_L$.

2.3 Joint Non-uniform Sampling and Quantization

Quantization, when used in conjunction with non-uniform sampling can further reduces the rate of the stream. When quantization is applied prior to sampling we have the resultant signal $\mathcal{S}_\varepsilon(\mathcal{Q}_L(x[k]))$ and when the signal is sub-sampled before quantization, the resultant signal is $\mathcal{Q}_L(\mathcal{S}_\varepsilon(x[k]))$. Note that these operators are not commutative, and the two cases are likely to achieve different compression factors. The compression gain is multiplicative, i.e. the corresponding rate of signal $\mathcal{Q}_L(\mathcal{S}_\varepsilon(x[k]))$ is $\frac{N_{SOI} b_q}{N} + b_{loc}$ bits per sample.

[1]The optimal values of these reconstruction levels are known for a standard MSE quantizer

3 DESIGN OF JOINT NON-UNIFORM SAMPLING AND QUANTIZATION BASED COMPRESSION

We can exploit the multiplicative gain in compression achieved by joint sampling and quantization to design better signal compression schemes. However, different types of signals and applications can tolerate different levels of quantization noise and require different numbers of samples of interest. Hence the joint design of quantization and non-uniform sampling needs to be performed carefully. Consider the two different operator options $\mathcal{S}_\varepsilon(Q_L)$ and $Q_L(\mathcal{S}_\varepsilon)$, and let the corresponding rates be $\frac{N_{SOI}^{\mathcal{S}_\varepsilon(Q_L)} b_q^{\mathcal{S}_\varepsilon(Q_L)}}{N} + b_{loc}^{\mathcal{S}_\varepsilon(Q_L)}$ and $\frac{N_{SOI}^{Q_L(\mathcal{S}_\varepsilon)} b_q^{Q_L(\mathcal{S}_\varepsilon)}}{N} + b_{loc}^{Q_L(\mathcal{S}_\varepsilon)}$. In order to design a good compression scheme, we also need to formally define a distortion metric. Let $x_r[k]$ represent the reconstructed signal, after decompression, i.e. $x_r[k] = \mathcal{S}_\varepsilon^{-1}(Q_L^{-1}(Q_L(\mathcal{S}_\varepsilon(x[k]))))$ or $x_r[k] = Q_L^{-1}(\mathcal{S}_\varepsilon^{-1}(\mathcal{S}_\varepsilon(Q_L(x[k]))))$. Then the utility associated with the compression may be defined in terms of $x[k]$ and $x_r[k]$ as $\mathcal{U}(x[k], x_r[k])$. The goal of designing the right compression scheme is to maximize this utility under a rate constraint. If the desired rate constraint is b_{con} (in bits per sample), the optimal compression scheme may be designed by solving the following constrained optimizations:

$$\{Q_{opt}, \mathcal{S}_{opt}\} = argmax_{\{Q_L, \mathcal{S}_\varepsilon\}} [\mathcal{U}(x[k], x_r[k])]$$
$$\text{subject to } \frac{N_{SOI}^{Q_L(\mathcal{S}_\varepsilon)}(b_q^{Q_L(\mathcal{S}_\varepsilon)})}{N} + b_{loc}^{Q_L(\mathcal{S}_\varepsilon)} \leq b_{con} \quad (1)$$

and

$$\{\mathcal{S}_{opt}, Q_{opt}\} = argmax_{\{\mathcal{S}_\varepsilon, Q_L\}} [\mathcal{U}(x[k], x_r[k])]$$
$$\text{subject to } \frac{N_{SOI}^{\mathcal{S}_\varepsilon(Q_L)}(b_q^{\mathcal{S}_\varepsilon(Q_L)})}{N} + b_{loc}^{\mathcal{S}_\varepsilon(Q_L)} \leq b_{con} \quad (2)$$

As mentioned earlier, designing the quantizer Q_L requires determining the number of quantization levels L and designing the non-uniform sampling $\mathcal{S}_{epsilon}$ strategy requires determining the optimal value for ε for a given non uniform sampling scheme. We thus reduce the problem of finding Q_{opt} and $\mathcal{S}_{opt}$ to the identification of the values of L and ε that maximizes the utility. Consequently, since $b_q^{Q_L(\mathcal{S}_{epsilon})} = \lceil \log_2 L \rceil$,1 and 2 can be rewritten as:

$$\{\varepsilon_{opt}, L_{opt}\} = argmax_{\{L, \varepsilon\}} [\mathcal{U}(x[k], x_r[k])]$$
$$\text{subject to } \frac{N_{SOI}^{Q_L(\mathcal{S}_\varepsilon)}(\lceil \log_2 L \rceil)}{N} + b_{loc}^{Q_L(\mathcal{S}_\varepsilon)} \leq b_{con} \quad (3)$$

and

$$\{\varepsilon_{opt}, L_{opt}\} = argmax_{\{\varepsilon, L\}} [\mathcal{U}(x[k], x_r[k])]$$
$$\text{subject to } \frac{N_{SOI}^{\mathcal{S}_\varepsilon(Q_L)}(\lceil \log_2 L \rceil)}{N} + b_{loc}^{\mathcal{S}_\varepsilon(Q_L)} \leq r_{con} \quad (4)$$

If the order of the quantization and non-uniform sampling also needs to be determined, we may compare the optimal utilities in the two cases to determine the best order. Solving the joint optimization presented in equations 4 and 3 is non-trivial. This optimization is heavily dependent on the relationships between $\mathcal{U}$ and N_{SOI} and the pair (ε, L). For a generic sampling algorithm, for a signal with arbitrary characteristics, it is likely to be very difficult to determine the optimal solution without some form of computationally complex exhaustive search. In some cases, however, for sampling algorithms such as FAN, and for well-behaved signals such as ECG, we show that these relationships can be estimated experimentally, and modeled using simple parametric functions. This enables tractable, and low complexity algorithms to solve the optimization in real time. In the following sections, we present several parametric model based approaches to trade-off computational complexity for accuracy, while solving this optimization for the FAN algorithm with MSE quantization for the ECG signal.

4 ENCODING ECG SIGNALS FOR REMOTE HEALTH MONITORING

We illustrate our approach to jointly quantize and sample non uniformly waveforms by focusing on the representation of electrocardiogram (ECG) signals. This proposed technique implements adaptive sampling before quantization (i.e. $\mathcal{S}$ before Q).

4.1 Brief Background on ECG Signals

A typical electrocardiogram monitoring device generates large volumes of digital data. Depending on the intended application, the sampling rate may range from 125 to 1000 Hz, with each data sample digitized to a 8-16 bit value. This translates to a minimum data rate of 15 KB per minute. Transmitting this signal over a low-bandwidth channel, especially when aggregating data from multiple sensors, requires compression. The data needs to also be recorded over long periods, often as much as 24 hours, and doctors may wish to build a database of ECG recordings for their patients. Minimizing the storage resources also requires data compression.

4.2 Adaptive Sampling

FAN (Barr, 1988) is a standard sampling technique for ECG signal compression and was reported in 1964 by (Gardenhire, 1964). It extracts samples of interest by approximating the signal using a piecewise linear representation, and discards all but the terminal points along these line segments. More precisely, the FAN algorithm replaces the signal with straight line segments such that none of the original points lies further from the line segment than some predetermined maximum deviation threshold τ. Figure 1 visually describes the algorithm. The first point $x[k_0]$ is accepted as non-redundant (permanent sample). Two slopes $\{L_1, U_1\}$ are drawn between $x[k_0]$ and $\{x[k_1]-\tau, x[k_1]+\tau\}$. The third sample point $x[k_2)]$ falls within the area bounded by the two slopes. Thus new slopes $\{L_2, U_2\}$ are calculated between $x[k_0]$ and $x[k_2]\pm\tau$ respectively. Then the two pairs of slopes are compared and the most restrictive are retained: $U_2 = \min(U_2, U_1)$ and $L_2 = \max(L_1, L_2)$. Since sample $x[k_1]$ lies inside the range it is thus discarded; while $x[k_2]$ is accepted as a permanent sample and the procedure above is repeated, comparing future sample values to the most restrictive lines. During signal reconstruction, the discarded samples are linearly interpolated from their neighboring retained samples.

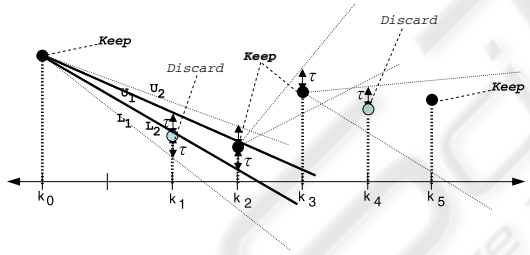

Figure 1: FAN algorithm for non-uniform sampling.

The deviation threshold τ determines the quality of the approximation with large τ leading to more samples being discarded, and coarser signal approximation. In our setting, this threshold τ maps directly to the sampling sensitivity ε, and we use the two interchangeably. The FAN algorithm has been used widely for ECG signal compression as it is extremely computationally lightweight ($O(N)$ for N samples), and performs reasonably well in practice, in terms of retaining samples and features of interest. However, for small target bit-rates (under 2 bits per sample), the FAN algorithm often underperforms computationally more complex ($O(N^2)$) algorithms such as Cardinality Constrained Shortest Path (CCSP). In this bit-rate range, we wish to improve the performance of FAN by combining it with quantization. Combination with a simple quantization can retain the low-complexity nature of FAN, while improving its compression quality.

4.3 Joint FAN Sampling and Quantization

The reconstruction quality of compressed ECG signals is often captured using the percentage root-mean-square difference (PRD) between the original signal and its reconstructed (inverse quantization and linear interpolation) version. The reconstructed signal $x_r[k]$ is determined from the sampled and quantized signal by inverse quantization and linear interpolation.

Hence, the utility function is defined as:

$$\mathcal{U}(x[k], x_r[k]) = -100 * \sqrt{\frac{\sum_{j=1}^{N}(x[j]-x_r[j])^2}{\sum_{j=1}^{N}x[j]^2}} \quad (5)$$

Finally the joint sampling and quantization problem, given a rate constraint b_{con} (in bits per sample), may be written as the following optimization:

$$\{\varepsilon_{opt}, L_{opt}\} = argmax_{\{\varepsilon,L\}}\left[\mathcal{U}(x[k], x_r[k])\right]$$
$$\text{subject to } \frac{N_{SOI}^{s_\varepsilon(Q_L)} b_q^{s_\varepsilon(Q_L)}}{N} + b_{loc}^{s_\varepsilon(Q_L)} \leq b_{con} \quad (6)$$

The search complexity for a naive implementation of the solution to this problem is $O(|\Omega_\tau| \times |\Omega_L|)$ where Ω_ε is the set of possible values for ε, Ω_L is the set of possible values for L and $|\bullet|$ is the cardinality operator. This is a constant factor that multiplies the complexity of the FAN algorithm (thereby linearly increasing the complexity). However, this is a worst case metric as it assumes no apriori knowledge of the underlying ECG signal. Due to the periodic nature of the ECG signal, the designed answer is likely to change slowly with time (across consecutive windows of N samples each), and hence we can distribute this complexity over several windows. This may be done by either solving the optimization once every Z windows, thereby reducing the overhead complexity to $O(\frac{|\Omega_\varepsilon|\times|\Omega_L|}{Z})$ or by reducing the space of possible search values, i.e. the number of elements in each set (allowing only for small variations in the previously designed values).

Additional improvement in performance may be obtained by actually designing the complete quantizer (including the design of the optimal resonstruction levels) dynamically. This however comes at a cost of increased complexity. In the worst case, a standard k-means based implementation of quantizer design has complexity $O(N^L)$. Of course, this cost may also be distributed across several windows (due to the nature

of the ECG signal) to reduce the computational complexity. The design of optimal low-cost quantizers in conjunction with the sampling is an interesting direction of future research.

5 MODEL BASED SEARCH STRATEGY

A model based search strategy is enabled by the reasonably stationary characteristics of the ECG signal, and the somewhat predictable behavior of the FAN algorithm. Specifically, we observe, that in a particular operating region (defined by the rate constraint, e.g. number of bits per sample, and the corresponding quality metric, i.e. PRD) we may develop simple parametric models that capture the effect of L and ε on the utility (PRD) and the rate (bits per sample). As an example, we run the FAN algorithm several times on a real ECG signal with different values of ε and plot the resulting number N_{SOI} of samples retained, and the corresponding distortion *PRD* in Figure 2.

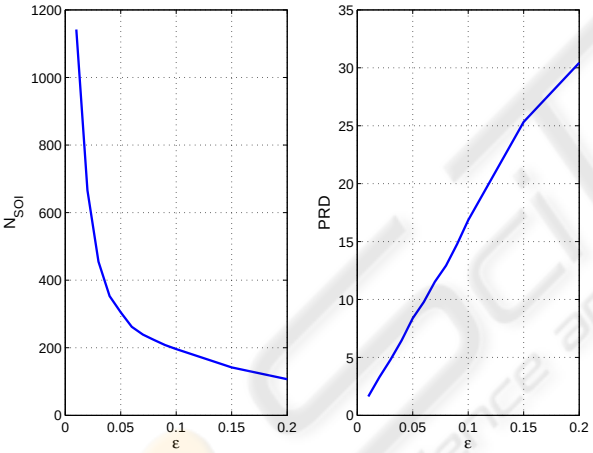

Figure 2: N_{SOI} and *PRD* as functions of ε.

As is clear, N_{SOI} has an almost exponentially decaying relationship with ε, while the *PRD* has a near-linear relationship with ε, and we can capture these relationships very simply as follows

$$N_{SOI}(\varepsilon) = \mu e^{-\nu\varepsilon} \tag{7}$$

and

$$PRD(\varepsilon) = \alpha + \beta\varepsilon \tag{8}$$

where μ, ν, α, and β are the model parameters. If we now combine this sampling with quantization using L levels, we can derive the resulting bit-rate for the compressed signal (in bits per sample) as

$$b_q^{Q_L(s_\varepsilon)} = \frac{N_{SOI} log_2(L)}{N} \tag{9}$$

Using Equation 7, we may rewrite this as

$$b_q^{Q_L(s_\varepsilon)} = \frac{\mu e^{-\nu\varepsilon} log_2(L)}{N}. \tag{10}$$

In order to build a similar model for the *PRD* as a joint function of ε and L, we plot the resulting PRD (after FAN followed by quantization) in Figure 3. Clearly,

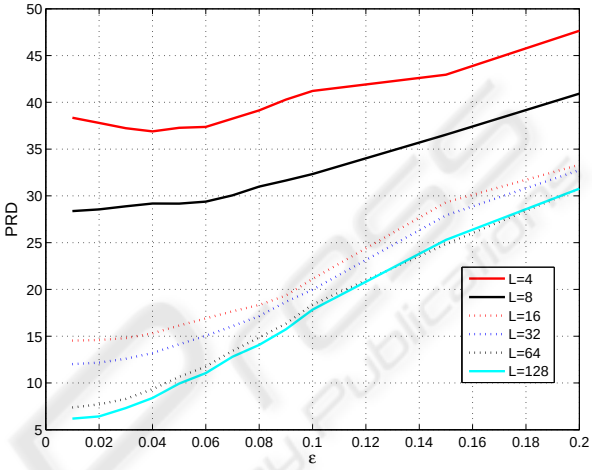

Figure 3: *PRD* as function of ε and L.

the slope and intercept of the line relating *PRD* to ε change with L. After further investigation, we find that this relationship may be captured as

$$\alpha(L) = \gamma e^{-\rho L} \tag{11}$$

and

$$\beta(L) = \eta \times log(\xi L_Q). \tag{12}$$

Combining these equations, we may rewrite the model for *PRD* after joint sampling followed by quantization as

$$PRD = \gamma e^{-\rho L_Q} + \eta\varepsilon log(\xi L_Q) \tag{13}$$

These models for *PRD* and $b_q^{Q_L(s_\varepsilon)}$ are validated for a real ECG signal in Figure 4 and Figure 5. While the models tend to underestimate the real values (especially for small ε), the shapes of the curves remain similar allowing for a search strategy using this model.

5.1 Three Times FAN Strategy

In order to compress the ECG signal under a rate constraint, we first partition it into fixed size windows (each with W samples[2]). Then per window, we run

[2]Note that since we process the window independently, we have $N = W$

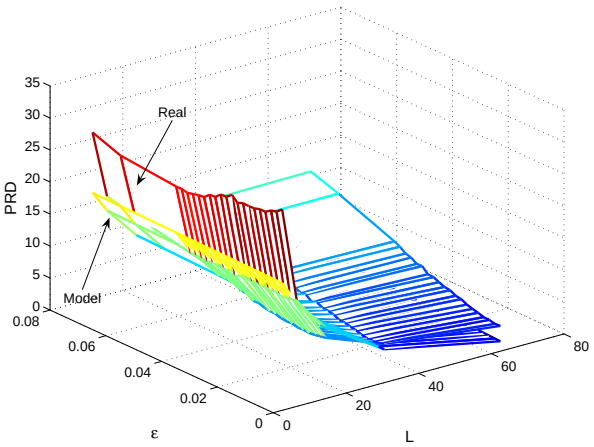

Figure 4: *PRD*: Real value versus model prediction.

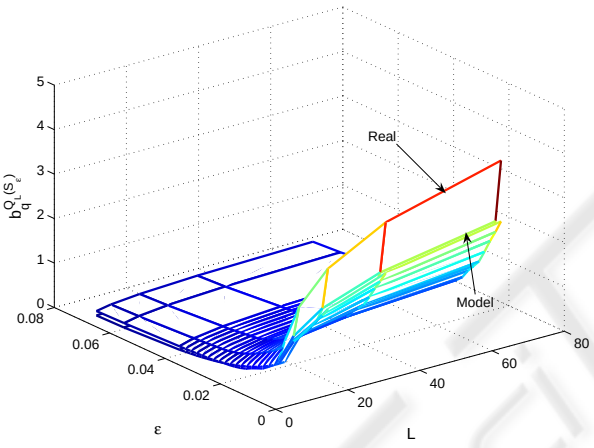

Figure 5: b_q: Real value versus model prediction.

the FAN algorithm for two different values of ε (a high value and a low value), to determine two values for N_{SOI} followed by quantization with two different numbers of levels L (a high value and a low value) to determine four values of *PRD*. The four values of *PRD* provide us with four equations to solve for parameters γ, ρ, η and ξ. Similarly, the two values of N_{SOI} provide us with two equations to solve for parameters μ and ν. Once we determine the model for a given window, it is straightforward to determine the optimal parameter settings for ε and L under any specified rate constraint. Once we determine the optimal parameter settings, we then need to run FAN once with the selected ε_{opt} followed by quantization with L_{opt} levels. Hence, per window we run the FAN algorithm three times.

5.2 Two Times FAN Strategy

We exploit the near stationarity of the ECG signal characteristics to reduce the complexity of the Three Times FAN strategy. Specifically, while for the first window we employ the same approach (with two times FAN followed by two times quantization) for every subsequent window we run the FAN algorithm for only one additional value of ε followed by quantization with two different values of L. This provides us with two values of *PRD* and one value of N_{SOI}. In order to compute the model parameters, we then combine this with two values of *PRD* and one value of N_{SOI} computed from the previous window. We alternate between recomputing the *PRD* and N_{SOI} for the high ε, and the *PRD* and N_{SOI} for the low ε (correspondingly reusing these for the low ε and high ε, respectively, from the previous window), for every successive window. Note that it is possible to easily extend this approach to recompute the model parameters only once every Z windows, to further reduce complexity. We examine some of the tradeoffs between complexity and accuracy by comparing the performance of these algorithms, and using that to identify trends for other extensions.

6 EXPERIMENTAL RESULTS

We evaluate the performance of these algorithms on ECG signals from the MIT-BIH database. Specifically, we use a subset of this database, consisting of 10 different ECG signals, of duration 8000 samples each. For these signals we evaluate the different strategies described in Table 1. We compare FANEX, FANQEX, FANQMS-3 and FANQMS-2 in terms of their distortion (PRD)-rate (bits per sample) curves, and also in terms of their computational complexity. We present results for different processing window sizes (W) to identify the general performance trend variations. Each window consists of W samples of the signal, and is analyzed and processed independently by the different algorithms, specifically in terms of computing the optimal parameters ε and L, and using the FAN algorithm with these parameters. We limit the search space for the exhaustive search strategies by considering a finite small set of possible values that ε and L can take. For our experiments we have $\varepsilon \in \{0.002, 0.004, \cdots, 0.04, 0.05, 0.06\}$ and $L \in \{4, 8, 16, 32, 64\}$. We first consider a processing window of size $W = 1000$ samples, and present the distortion-rate (D-R) curve averaged across these signals (across the 8 windows per signal) for the four different algorithms in Figure 6.

In Figure 6 we observe that the schemes with joint

Table 1: Algorithms Considered.

Name	Uses Quantization	ε only or (ε, L) search strategy
FANEX	No	Exhaustive
FANQEX	Yes	Exhaustive
FANQMS-3	Yes	Model Based - Three Times FAN
FANQMS-2	Yes	Model Based - Two Times FAN

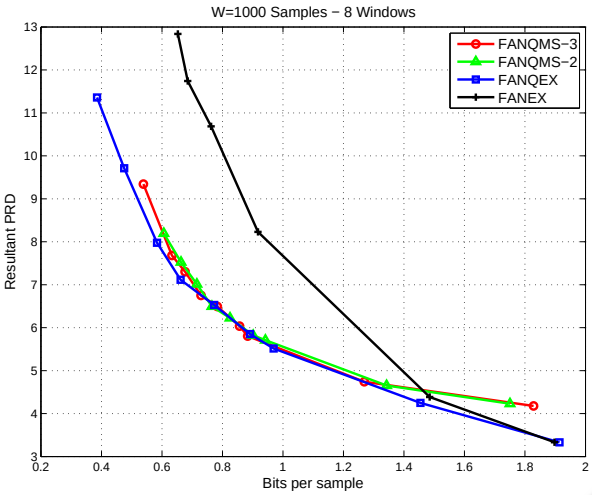

Figure 6: D-R Curves: $W = 1000$.

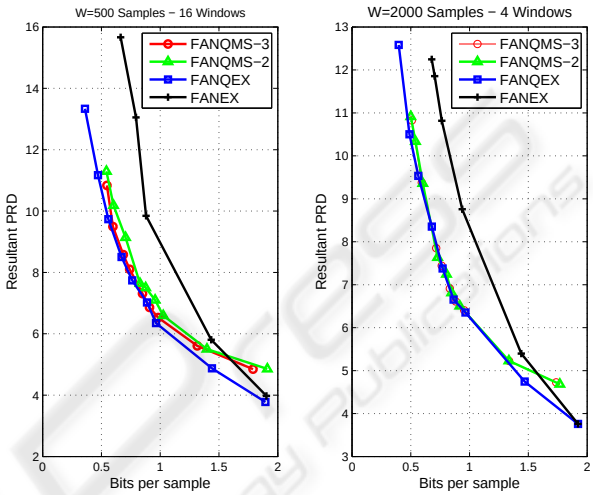

Figure 7: D-R Curves: $W = 500$(left), $W = 2000$(right) .

quantization and FAN significantly outperform the FAN only scheme, a compression factor of 2 for the same PRD. This makes the performance of the FAN algorithm comparable to the state-of-the art compression algorithms (with significantly higher complexity). Furthermore, we find that the model based searches FANQMS-3 and FANQMS-2 have performance very close to that achieved by the exhaustive search for target bit-rates less than 2 bits per sample. As the target bit-rate starts to approach 2 bits per sample, the model based search strategies underperform the FANEX strategy, as the models are inaccurate[3] for this range. However, note that for this higher bit-rate range, the performance of the FAN algorithm by itself is comparable to the best ECG compression algorithms presented, thereby limiting any gains obtained by additionally quantizing the signal. We also repeat these experiments for a smaller window ($W = 500$) and a larger window ($W = 2000$, and the results are presented in Figure 7.

From Figure 7, the same performance trend is observed as in Figure 6 for the four algorithms, however it is clear that the performance of FANQEX, FANQMS-2, and FANQMS-3 are closer to each other for larger W. This may be explained by the fact that a larger window size allows the model based algorithms to fit better parameterized curves, improving the performance of the model based search schemes. This is also evident from the fact that on average, the PRD for the same target bit rate decreases with increasing W. We also compare the computational complexity of these algorithms in terms of the amount of CPU time consumed per window. These CPU times are labeled t_{FANEX}, t_{FANQEX}, $t_{FANQMS-2}$ and $t_{FANQMS-3}$ respectively. We also label the time taken to run the FAN algorithm on one window as t. Instead of presenting absolute numbers, we present relative ratios of the complexity of these algorithms to hide the dependency on the underlying computer architecture, operating system etc. These complexity ratios for the different algorithms are presented in Table 2.

It is evident from Table 2 that FANQMS-3 has 29 (FANQMS-2 has 45 times) lower complexity than FANQEX and 4 times (FANQMS-2 has 7 times) lower complexity than FANEX. Further, as expected, FANQMS-2 has lower complexity than FANQMS-3. This observation holds across the two different window sizes considered. Furthermore, FANQMS-3 has 4 times the complexity of FAN, while FANQMS-2 has 3 times the complexity of running FAN one time. This implies that the search for the optimal ε and L has the complexity 1.5 times that of the FAN algorithm. Note that by reusing the model parameters

[3] Assumptions on linear, exponential and log-linear relationships are violated in this range.

Table 2: Complexity Comparison.

W	$\frac{t_{FANQEX}}{t_{FANQMS-3}}$	$\frac{t_{FANEX}}{t_{FANQMS-3}}$	$\frac{t_{FANQMS-3}}{t}$	$\frac{t_{FANQEX}}{t_{FANQMS-2}}$	$\frac{t_{FANEX}}{t_{FANQMS-2}}$	$\frac{t_{FANQMS-2}}{t}$
500	29.18	4.50	4.18	47.75	7.35	2.80
1000	29.06	4.49	4.21	46.07	7.12	2.64

across more windows (updating model infrequently), this overhead can also be significantly reduced. This is also indicated by the comparing the rows of Table 2, as the complexity gains for FANQMS-2 increase as W increases from 500 to 1000 (the ratio $\frac{t_{FANQMS-2}}{t}$ decreases from 2.80 to 2.64).

7 CONCLUSIONS

We present a low-complexity joint non-uniform sampling and quantization based strategy for signal compression. Specifically, we combine the FAN algorithm with a minimum mean-squared error quantization strategy to compress ECG signals. We first formulate the joint design of non-uniform sampling and quantization for compression, as a constrained optimization problem in terms of maximizing the relevant distortion metric given the desired compression rate. The solution of this optimization yields the optimal sampling sensitivity, and the number of levels to be used by the quantizer. In general, and for arbitrary signals, it may not be possible to solve this optimization efficiently. However, for ECG signals, we show that we can develop simple parametric models to capture the impact of the FAN algorithm and quantization on the resulting distortion (PRD) and rate, especially in very low bit-rate operating regions. Using these models we can efficiently determine the optimal FAN selectivity parameter ε and quantization levels L to minimize the PRD for a given rate constraint. We design two model based algorithms, one that re-estimates model parameters for every window (W samples), and another that updates model parameters every alternate window. We show that with these strategies, we can achieve up to 2 times the compression rate of FAN (for the same PRD) with a complexity less than 3 times that of FAN alone. We also show that the performance of these algorithms approaches (within 10% in rate when $\varepsilon < 1.8$) an exhaustive search based strategy for different signals, and window sizes. Given the low complexity of FAN our algorithms still remain significantly lower complexity than state-of-the-art transform based compression schemes, while achieving comparable performance. Directions for future research include design of the optimal search strategy to re-estimate model parameters (how often, optimal window size etc.), theoretical analysis of the signal frequency and statistical properties as well as algorithm complexity for rate-distortion-complexity optimal joint sampling and quantization, and application of these ideas for other multi-dimensional medical signals.

REFERENCES

Addison, P. S. (2005). Wavelet transforms and the ecg: a review. In *Physiological Measurement*, volume 26.

Barr, R. C. (1988). Adaptive sampling of cardiac waveforms. In *Journal of Electrocardiology*, volume 21, pages 57–60.

Bradie, B. (1996). Wavelet packet-based compression of single lead ecg. In *IEEE Transactions on Biomedical Engineering*, volume 43, pages 493–501.

Cox, J., Fozzard, H., Nolle, F. M., and Oliver, G. (1968). Aztec: A preprocessing system for real-time ecg rhythm analysis. In *IEEE Transactions on Biomedical Engineering*, volume 15, pages 128–129.

Derpich, M., Quevedo, D., Goodwi, G., and Feue, A. (2006). Quantization and sampling of not necessarily band-limited signals. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3.

Gardenhire, L. W. (1964). Redundancy reduction the key to adaptive telemetry. In *National Telemetering Conference*, pages 1–16.

Hilton, M. L. (1997). Wavelet and wavelet packet compression of electrocardiograms. In *IEEE Transactions on Biomedical Engineering*, volume 44, pages 394–402.

Jalaleddine, S., Hutchens, C., Strattan, R., and Coberly, W. (1990). Ecg data compression techniques-a unified approach. In *IEEE Transactions on Biomedical Engineering*, volume 37, pages 329–343.

Mohomed, I., Ebling, M. R., Jerome, W., and Misra, A. (2006). Harmoni: Motivation for a health-oriented adaptive remote monitoring middleware. In *UbiHealth'06, Fourth International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications*, Irvine, CA, USA.

Nygaard, R., Melnikov, G., and Katsaggelos, A. (2001). A rate distortion optimal ecg coding algorithm. In *IEEE Transactions on Biomedical Engineering*, volume 48, pages 28–40.