

Just-in-time Sampling and Pre-filtering for Wearable Physiological Sensors: Going from Days to Weeks of Operation on a Single Charge

Nan Hua
Georgia Tech

Ashwin Lall
Georgia Tech

Justin Romberg
Georgia Tech

Jun (Jim) Xu
Georgia Tech

Mustafa al'Absi
Univ. of Minnesota

Emre Ertin
The Ohio State Univ.

Santosh Kumar
Univ. of Memphis

Shikhar Suri
Georgia Tech

ABSTRACT

Continuous monitoring of human physiology and behavior in natural environments via unobtrusively wearable wireless sensors is witnessing rapid adoption in both consumer healthcare and in scientific studies, since those portable and long-running devices can provide critical information for diagnosis and early prevention of disease, as well as invaluable data for scientific studies. Due to the requirement of continuous monitoring, these sensors, all operating on small wearable batteries, require frequent recharging. Lowering this recharging burden is essential for their widespread adoption. In this paper we explore mechanisms for significantly enhancing the lifetime of these wearable sensors at the cost of a small loss in their sensing accuracies. We propose two ideas that build upon our observation that collecting bursts of samples over short periods of time is sufficient to capture the most interesting and informative part of the signal. In the first part of this paper, we propose a general methodology for reconstructing bandlimited signals accurately from such short bursts of samples. While this reconstruction task is in nature an ill-conditioned problem, we show that the insertion of an analog “modulated pre-filter” hardware module before the ADC can almost surely alleviate this conditioning problem. In the second part of this paper, we describe *just-in-time sampling*, which by sampling in short bursts at the “right” times, can accurately track R-wave peaks in ECG signals. Using simulations on publicly available traces as well as self-collected data, we show the efficacy of this technique.

1. INTRODUCTION

This work was supported in part by NSF grants CNS-0905169, CNS-0910592, CNS-0910878, funded under the American Recovery and Reinvestment Act of 2009 (Public Law 111-5), NIH Grant U01DA023812 from National Institute for Drug Abuse (NIDA), and NSF grant CNS-0716423.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$10.00.

Remote health monitoring via wearable sensors is witnessing rapid adoption especially for elderly care and chronic diseases. For such patients, remote health monitoring makes diagnosis more accurate, improves the quality of care, and reduces the cost of care. For example, one study found that in-home monitoring and coaching after hospitalization for Congestive Heart Failures (CHF) reduced rehospitalizations for heart failure by 72%, and all cardiac-related hospitalizations by 63% [8]. Moreover, those devices could aid the diagnosis of diseases (e.g., heart diseases) that require long term monitoring and provide preventive care for susceptible population so that serious diseases can be detected at early stages. As the accuracy of remote monitoring devices improves, their cost decreases, and they become easier to use (accelerated by improvements in technology and special initiatives such as NIH’s Genes Environment Initiative (GEI) program), their adoption will accelerate.

Similarly, unobtrusively wearable sensors are revolutionizing scientific studies by providing continuous and objective measurements of daily behaviors such as exercise and dietary habits, and of personal exposures to psychosocial stress, addictive substances (e.g., smoking and drinking), and environmental pollutants (e.g., diesel exhaust), etc. [24]. These studies seek to investigate the causes, associated physiological responses, and public health consequences of prevalent human diseases such as chronic stress, addiction, obesity, panic disorders, etc., and to design and evaluate appropriate interventions for them. Use of wearable sensors makes it possible to collect objective data from the natural environment of subjects that have unprecedented ecological validity as compared to lab based studies, while not suffering from human biases and high compliance burden associated with collection of self-reports or ecological momentary assessments [14].

Although a variety of new wearable sensors have emerged that provide scientifically valid measurements from wearable form factors aided by the GEI initiative and other similar programs, they all face some major roadblocks to widespread adoption. Limited lifetime on small wearable batteries is one of them. It has been noted in experience reports of field deployment that the need for recharging adds to the participant burden and complicates the study logistics as the participants must be provided with a charger and instructed on how to take off the sensors and recharge them overnight [9]. Since compliance is known to be one of the biggest hindrances in behavioral studies, behavioral scientists usually place a high premium on devices that can run for the entire

length of the study without a need for recharging. Since several applications in both remote health monitoring and scientific studies involve detection of phenomena that can occur in an instant (e.g., stress, craving for a drug, panic attack, heart failure, etc.), it places a lower bound on the sampling frequency of sensors and an upper bound on the delay in inferring behavioral events. Given that behavioral studies or long term health monitoring may last for weeks, sometimes several months, it is a challenge to enable sensors to *last several months on wearable batteries* (usually of < 1100 mAh capacity), while still capturing a majority of behavioral events of interest.

To address the issue of power consumption, these sensing devices are equipped with very low-power hardware [4, 12] that maximizes their lifetime between recharges. Another common solution is to perform duty cycling [13, 17], wherein the device regularly cycles between on and off states. By remaining on for only a small proportion of the time, the device spends only a fraction of its normal power consumption. A significant drawback of regular duty cycling is that, for some signals of interest, this type of sampling may miss interesting events when it is powered down.

In this paper, we explore two novel techniques for enhancing the lifetime of these wearable sensors by an order of magnitude at the cost of a tolerable loss in the accuracy of measurements they produce. Both schemes perform on-off sampling of signals like in conventional duty cycling schemes to save power, that is, the ADC (Analog-to-Digital Converter) on the device alternates between the state of “on” for a short interval (say 100ms), wherein a burst of measurements (called samples) are gathered, and the state of “off” for a much larger interval (say 900ms). However, both circumvent the problem of significant information loss and signal distortion caused by the aforementioned conventional duty cycling scheme, in very different ways.

Our first technique, called *pre-filtering*, is a general methodology for reconstructing band-limited signals from “bursts” of samples collected during the aforementioned “on” intervals. However, extrapolating a band-limited signal from a burst of closely spaced samples is an inherently ill-conditioned problem. Our idea is to scramble the original signal with an analog modulated “pre-filter” that can operate without need for the ADC to be on. We show that if the signal is band-limited, it is possible to reconstruct the original signal from such bursts of samples with very little information loss and distortion.

Our second technique, called *just-in-time sampling*, is to turn on the ADC “just in time” to capture the most interesting or informative segments of the signal of interest and then to turn it off immediately after such segments are obtained to save power. Unlike in duty-cycling, here the ADC is turned on and off at “right times” as computed by our signal processing (detection, inference, and prediction) procedures. We show that the proposed simple mechanisms for predicting when to collect short bursts of samples are sufficient to guarantee small measurement error. In addition, just-in-time sampling allows us to turn off the entire sensing chain (not just the ADC) to save even more power.

We experimentally apply our just-in-time sampling algorithm to accurately track R-wave peaks (to determine beat-to-beat variations) in ECG signals, a measure used widely in both remote monitoring of health conditions related to heart and in scientific studies of stress, obesity, addiction, among

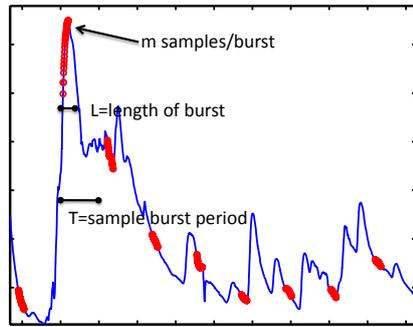


Figure 1: A bandlimited signal which is being sampled in bursts. It is clear that although the total number of samples is above twice the bandwidth, the samples do not give us information about salient features occurring between the bursts. Our ability to reconstruct this signal depends strongly on the time in between bursts and the length of the bursts and only very weakly on the number of samples per burst.

others [15]. Using simulations on publicly available traces as well as data collected from a mobile device, we show that we can track heartbeats with little error while keeping the entire sensing ADC on about 10-25% of the time.

The rest of this paper is organized as follows. In Section 2 we introduce the use of pre-filtering for recovering band-limited signals. Section 3 details the workings of our just-in-time sampling mechanism for intelligently power-cycling the sensor. Finally, we discuss related work in Section 4 before ending with our conclusions and suggested future work in Section 5.

2. MODULATED PRE-FILTERING

In this section, we consider the general problem of reconstructing a signal from samples taken in “bursts”. The scenario is illustrated in Figure 1. As we will discuss below, this problem is inherently ill-conditioned if the time between the burst is much greater than the inverse of the bandwidth of the signal. However, we will show how this type of acquisition can be *preconditioned* by performing some simple analog operations on the signal before it is sampled. We will treat the problem from a mathematical standpoint, focussing on the types of mathematical transformations that need to be applied to the signal before it is sampled, and leave aside for now the question of how to implement these components.

A signal $x(t)$ that is bandlimited to W is sampled K times every second. The samples are equally spaced, but are clustered into bursts of size m which occur over a time period L every T seconds (so $L \leq T$, the spacing between the samples is m/L , and the number of samples in each burst is $m = KT$). The question, then, is what the relationship between W , m , L , and T needs to be to accurately reconstruct the signal.

If the samples were equally spaced ($m = 1$, $K = 1/T$, $L = T$), then the celebrated Shannon-Nyquist theorem tells us that we can take $W \leq 1/(2T) = K/2$ — we can recover signals whose bandwidth is up to half the number of samples we are taking per second. The answer is somewhat murkier in the burst sampling case, but we can use some classical guidelines for a qualitative analysis. We can see from Figure 1 that if m is large, the samples within a burst

are highly correlated. This leads us to believe that after a point, increasing the number of samples in each burst does not really help in reconstructing the signal; in a sense, these excess samples are wasted. Indeed, consider for a moment that we observed the continuous signal over $[0, L]$ (or any other interval of length L). The space of signals we could observe are those that can be written as a W bandlimited signal that is then time-limited to $[0, L]$. The essential dimension of this space (when L is small) is $\approx 2WL$ (see [19] for a nice exposition). This means that no matter how many samples m we take, only about $2WL$ of them will be linearly independent. For $L \ll T$, this means that we will only be able to reconstruct signals with bandwidth $W \approx 1/T$, even though we are taking m/T samples per second. There is absolutely no benefit to taking the samples more quickly, all that matters is the spacing between the bursts.

We will show below that if the signal is modulated and then filtered in a certain way *before* it is sampled, then this sampling rate does make a difference; increasing m will give us a greater number of diverse observations. These operations, by spreading the signal out in frequency and then in time, effectively “precondition” the burst sampling operator by combining information about the original signal over long periods of times and at all frequencies into each sample.

To develop the ideas above more precisely, we will specialize the discussion to the following scenario which can be described neatly using the language of linear algebra. Suppose that $x(t)$ is a bandlimited periodic signal with period $T = 1$; we can write $x(t)$ as

$$x(t) = \sum_{\ell=0}^{b-1} \alpha_{\ell} e^{j2\pi\ell t},$$

where $\alpha_0, \dots, \alpha_{b-1}$ are the (non-zero) Fourier series coefficients of $x(t)$. Notice that we are taking $x(t)$ to be “half band”; it is complex valued, and can be reconstructed perfectly from b equally spaced samples (also complex valued) on $[0, 1]$. We do this only to simplify the notation in the exposition; the discussion translates easily to real signals with non-zero Fourier coefficients at negative frequencies.

We collect a single burst of samples of $x(t)$ spread out on $[0, L]$ at the beginning of the period $[0, 1]$. We observe

$$y_k = x(kL/m) = x(k/n), \quad k = 0, \dots, m-1,$$

where $n = m/L$. We will assume that things are arranged so that n is an integer; it can be interpreted as the *sample rate*, the number of samples per second we would be taking if the burst lasted the entire period. Combining the equations above, we can write

$$y_k = \frac{1}{\sqrt{n}} \sum_{\ell=0}^{b-1} \tilde{\alpha}_{\ell} e^{j2\pi\ell k/n}, \quad (1)$$

where we re-normalize the α_{ℓ} as $\tilde{\alpha}_{\ell} = \sqrt{n} \cdot \alpha_{\ell}$.

We can write this more compactly. Let F be the $n \times n$ normalized discrete Fourier transform (DFT) matrix

$$F_{\ell,k} = \frac{1}{\sqrt{n}} e^{-j2\pi k\ell/n}, \quad \ell, k = 0, \dots, n-1,$$

notice that $F^*F = I^*$. For convenience we will use the notation $x_d[k] = x(k/n)$ for $k = 0, \dots, n-1$ for the samples

Throughout the paper, we use A^ to denote the conjugate transpose of the matrix A . Here, since F is $n \times n$ and its rows are orthogonal, F^* is also the inverse of F .

of $x(t)$ taken uniformly over $[0, 1]$ with spacing $1/n$; we will assume that $n > b$, so the continuous-time signal $x(t)$ can be safely reconstructed from the n -vector x_d . Let R_m be the $m \times n$ “sampling matrix” such that $R_m v$ returns the first m entries of the n -vector v , and so $y = R_m x_d$. Using this notation, we can write $x_d = F^* R_b^* \tilde{\alpha}$; the operator R_b^* (the transpose of R_b) takes the b -vector $\tilde{\alpha}$ and zero-pads it to length n , while F^* synthesizes the signal and samples it on the dense grid. Then we can rewrite (1) as

$$y = R_m F^* R_b^* \tilde{\alpha}.$$

Below, we will use the notation $F_{mb}^* = R_m F^* R_b^*$. Notice that F_{mb}^* can be constructed by taking the Fourier matrix F^* , extracting the first b columns, and then extracting the first m rows.

The question, which was addressed qualitatively above, is whether we can stably recover $\tilde{\alpha}$, and hence $x(t)$, given the (possibly noisy) samples in y . The answer, unfortunately, is no in just about every meaningful situation. As we are not making any assumptions about the structure of $\tilde{\alpha}$, we will consider the least-squares estimate

$$\hat{\alpha} = (F_{bm} F_{mb}^*)^{-1} F_{bm} y,$$

where F_{bm} is the adjoint (conjugate transpose) of F_{mb}^* . The stability of this recovery is completely characterized by the singular values of F_{mb}^* . To have confidence in our estimate $\hat{\alpha}$, we would like F_{mb}^* to be *well-conditioned*. That is, we would like the spectrum (the set of singular values) of F_{mb}^* to consist of b non-zero values that are all about the same size.

We can examine the conditioning of F_{mb}^* by calculating the spectrum for a particular case. For concreteness, we take $n = 1000$, $b = 10$, $m = 10$ (in this case $L = m/n = 1/100$, so we are observing samples over just 1% of the fundamental interval). A plot of the log-spectrum is shown in Figure 2. From the plot, we see that F_{mb}^* is essentially rank-2, and even then there is a factor of ≈ 20 between the largest singular value and the second largest. This means we are acquiring at most 2 linearly independent samples, making it impossible to recover a vector with $b = 10$ entries. If we increase the sampling rate while keeping the length of the burst $L = m/n$ constant, the spectrums are almost identical. This indicates that these additional samples do not give us any new information about the signal — they are simply linear combinations of one or two principal components. (In this example, we have kept the ratio m/n constant. Increasing m without increasing n corresponds to taking samples over a longer period; increasing n without increasing m corresponds to taking samples at a faster rate, but making the burst over a shorter amount of time. In either case, the rank of F_{mb}^* will scale with m/n .)

The fundamental problem is that the signal is concentrated in frequency and the sample locations are concentrated in time. We will show below that if we spread the signal out in frequency and then convolve with a long and diverse pulse (which has the effect of diffusing local information about the signal across the entire interval) before we sample it, then the corresponding inverse problem is beautifully conditioned.

Our proposed sampling architecture is shown in Figure 3. The signal passes through a modulator, which mixes it against a pseudo-random binary waveform, then is passed through a filter with impulse response $h(t)$. Each of these compo-

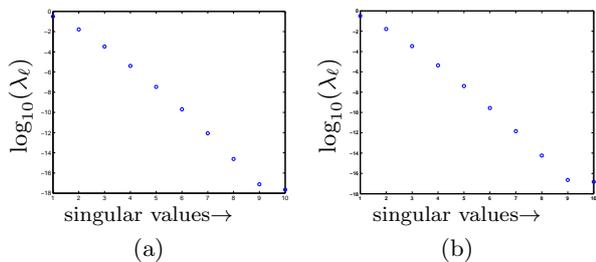


Figure 2: (a) Plot of the singular values (on a log scale) of F^*_{mb} for $b = 10, m = 10, n = 1000$. The $m \times b$ matrix is essentially rank-2. (b) Plot of the ten largest singular values for $b = 10, m = 100, n = 10,000$; the spectrum is nearly identical to (a) indicating that the conditioning of the burst-sampling process is independent of the number of samples m in the burst.

nents is analog, mapping a continuous time signal to another continuous time signal. However, because all of the signals involved are bandlimited and periodic, they can be represented with discrete sets of numbers (equally spaced samples or Fourier coefficients, for example). Much of the analysis below will model these analog components by how they map a discrete representation of the input to a discrete representation of the output; since the components are all linear, the action of each can be captured with a matrix. We will use parentheses (e.g. $x(t)$) to index a continuous-time signal and brackets (e.g. $x_d[k]$) to index a discrete representation.

The waveform $p(t)$ we mix x against in the modulator is periodic with the same period as $x(t)$ and is bandlimited to πn . We can write it as

$$p(t) = \theta(t) \star h_{lp}(t),$$

where $\theta(t)$ is constructed on $[0, 1]$ from independent and identically distributed random variables $\theta_0, \dots, \theta_{n-1}$ that take values $\pm\sqrt{n/m}$ with equal probability[†],

$$\theta(t) = \theta_i, \quad \text{for } i/n \leq t \leq (i+1)/n,$$

the convolution \star is circular, and $h_{lp}(t) = \sin(n\pi t)/\sin(\pi t)$ is the Dirichlet kernel. The samples of the output $u(t)$ of the modulator can be written in terms of the samples of the input as

$$u_d[k] := u(k/n) = \theta_k x_d[k], \quad \text{or } u_d = \Theta x_d,$$

where $\Theta = \text{diag}(\{\theta_0, \dots, \theta_{n-1}\})$.

As $x(t)$ (and also $u(t)$) is periodic, we will also take the subsequent filtering operation to be circular. We give the impulse response diversity by constructing it from a *random Fourier series*. We set

$$h(t) = \frac{1}{n} \sum_{\ell=0}^{n-1} \sigma_\ell e^{j2\pi\ell t},$$

where $\sigma_0, \dots, \sigma_{n-1}$ is another sequence of iid ± 1 random variables. The samples v_d of the output $v(t) = u(t) \star h(t)$ of the filter can then be written as

$$v_d = H u_d$$

[†]This choice of scaling factor, as we will see later, of $\sqrt{n/m} = \sqrt{1/L}$ makes the total energy of the samples roughly the same as the total energy of the signal $x(t)$.

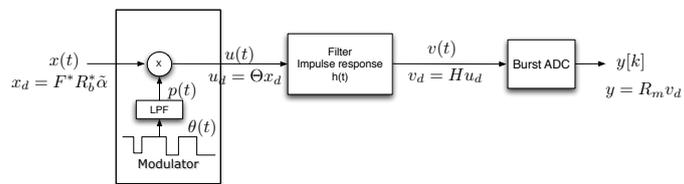


Figure 3: The prefiltering architecture. The bandlimited signal passes gets mixed against a pseudo-random waveform, passes through a filter with a broad frequency response, and then goes to the burst sampler. While it is impossible to recover $x(t)$ from burst samples, it can be very stably recovered from burst samples of the prefiltered signal $v(t)$.

where H is a circulant matrix with character $h[k] = h(k/n)$; we can decompose H as

$$H = F^* \Sigma F, \quad \Sigma = \text{diag}(\{\sigma_0, \dots, \sigma_{n-1}\}).$$

The samples that we observe can then be written as

$$y = R_m v_d = R_m F^* \Sigma F u_d = R_m F^* \Sigma F \Theta F^* R_b^* \tilde{\alpha}.$$

Our ability to recover $\tilde{\alpha}$ (and hence $x(t)$) from the burst of samples y using least-squares

$$\hat{\alpha} = (A^* A)^{-1} A^* y, \quad (2)$$

is now controlled by the singular values of the *random matrix* $A = R_m F^* \Sigma F \Theta F^* R_b^*$. A plot for $n = 1000, b = 10, m = 10$ and particular realization of Θ and Σ is shown in Figure 4(a) (the behavior does not vary too much over different realizations). We see that not only is $A^* A$ full rank, but it is also fairly well-conditioned; the ratio of the maximum singular value to the minimum singular value is about 17. Essentially, this means that we can recover, with reasonable stability, any $b = 10$ non-zero Fourier coefficients (and hence any b -bandlimited $x(t)$) from the burst of $m = 10$ samples. Each of the ten samples represents a linearly independent measurement of x .

The conditioning of this system gets even better as we increase the sampling rate. As shown in Figure 4(b), with $m = 100, n = 10000$ the condition number is ~ 1.6 . This means that these added samples really are giving us diverse looks at this signal — if noise is added to the observations, taking these extra samples will decrease the variance of the final estimate in almost the same way that uniform oversampling does in the standard case.

Since the $m \times b$ matrix A has all of its singular values close to 1, we can model the samples that our architecture is taking as expansion coefficients in a *frame* which is nearly tight. This means that if there is noise added to the samples, the reconstruction (2) will give us the associated oversampling gains. That is, the total noise in the reconstruction will be smaller than the total noise in the measurements by a factor of about m/b on average.

Injecting this randomness into the measurement system is an effective preconditioner for the burst sampling process. The matrix A is random, as are its singular values. If we let f_ℓ be a column of the matrix $R_m F^*$ and let \tilde{f}_k^* be a row of $F^* R_b^*$, then we can write A as

$$A = \frac{1}{\sqrt{n}} \sum_{k,\ell} \sigma_\ell \theta_k W_{k,\ell} f_\ell \tilde{f}_k^*,$$

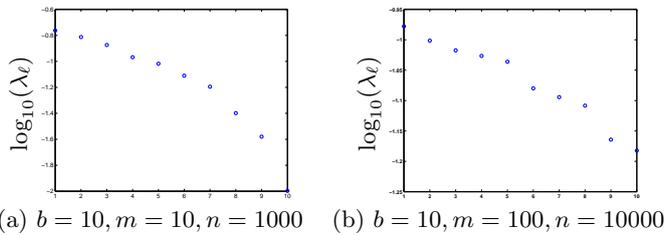


Figure 4: Log-spectrums of $A = R_m F^* \Sigma F \Theta F^* R_b^*$. In contrast to Figure 2, the matrix $A^T A$ is getting better and better conditioned as the number of samples in the burst increases.

where $W_{k,\ell} = e^{-j2\pi(\ell-1)(k-1)/n}$, and

$$A^* A = \frac{1}{n} \sum_{k,\ell,k',\ell'} \sigma_\ell \sigma_{\ell'} \theta_k \theta_{k'} W_{k,\ell} W_{k',\ell'}^* \langle f_{\ell'}, f_\ell \rangle \bar{f}_{k'} \bar{f}_k^*.$$

The expectation of $A^* A$ is

$$\begin{aligned} \mathbb{E}[A^* A] &= \frac{1}{n} \sum_{k,\ell,k',\ell'} \mathbb{E}[\sigma_\ell \sigma_{\ell'}] \mathbb{E}[\theta_k \theta_{k'}] W_{k,\ell} W_{k',\ell'}^* \langle f_{\ell'}, f_\ell \rangle \bar{f}_{k'} \bar{f}_k^* \\ &= \frac{1}{m} \sum_{k,\ell} \langle f_\ell, f_\ell \rangle \bar{f}_k \bar{f}_k^* = \sum_k \bar{f}_k \bar{f}_k^* = I, \end{aligned}$$

where the first equality comes from the fact that the θ_k and σ_ℓ are independent for all k, ℓ , the second equality comes from the fact that $\mathbb{E}[\sigma_\ell \sigma_{\ell'}] = 0$ and $\mathbb{E}[\theta_k \theta_{k'}] = 0$ unless $\ell = \ell'$ and $k = k'$, the third equality comes from the fact that $\langle f_\ell, f_\ell \rangle = m/n$ for all ℓ , and the last equality comes from the fact that $R_b F F^* R_b^* = I$.

In expectation, $A^* A$ is the identity, meaning that on average the reconstruction in (2) is very well behaved. Of course, we would like to be confident that A does not vary from its mean by too much. As m gets larger, we expect that the $A^* A$ will concentrate more and more around I (we see this numerically in Figure 4 as the conditioning gets better as m increases). We are currently working to make this precise by developing a tail bound on the random variable $\|I - A^* A\|$, where $\|\cdot\|$ is the standard matrix norm.

It is worth noting the neither the random modulator nor the random filter would be an effective preconditioner by themselves. If we were to simply burst sample the output of the modulator, we would observe the same samples as the standard burst sampler, only with their signs changed. This of course would not improve our ability to recover \tilde{a} from y , as the conditioning of ΘF_{mb}^* is exactly the same as that of F_{mb}^* . Similarly, if we were to simply pass the bandlimited signal through the broadband filter, the output will have the same bandlimit as the signal, and so the problem of recovering from the samples has not been made easier — the conditioning of $R_m F^* \Sigma R_b^*$ is exactly the same as that of F_{mb}^* .

Both the random modulator [20] and the random filter [18, 21] have been proposed before in the context of *compressive sensing* (CS). The high-level goal of CS is different than our goals here. In CS, we are interested in recovering a sparse signal from under-sampled measurements; the diversity that the random modulator or random filter introduce compensates for the small number of samples. Here, we are interested in a more classical signal model (bandlimited as opposed to sparse), and we use the diversity introduced by the two components working in concert to compensate for

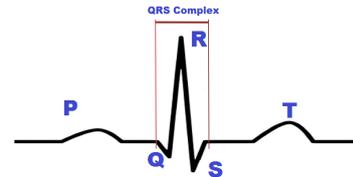


Figure 5: P-, Q-, R-, S-, T-Waves in One Beat

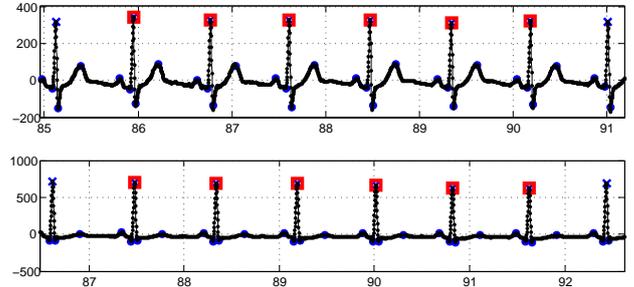


Figure 6: Examples of ECG signals

the fact that the samples are taken in bursts. In CS, the randomness helps us “invert” an underdetermined system, while here the randomness helps us precondition an overdetermined system. The idea prefiltering a signal before you sample also appears in the literature on sampling signals with finite rate of innovation [23]; our goal here is slightly different in that we are trying to overcome difficulties caused by taking the signals in bursts, rather than acquire signals which are concentrated in time.

3. JUST-IN-TIME SAMPLING

In this section we describe and evaluate our just-in-time sampling methodology. Whereas the pre-filtering described in the previous section requires some additional hardware support that we hope to build in the future, just-in-time sampling is an entirely software-driven solution. Moreover, just-in-time replaces regularly interspersed burst sampling with intelligent selection of sampling periods so as to capture only the information-rich parts of the signal. We present just-in-time sampling in the context of measuring features in cardiac signals, which we describe first.

3.1 Cardiac Signal Processing

Tracking the ECG signal of a patient in the field is useful for numerous cardiac conditions such as arrhythmia as

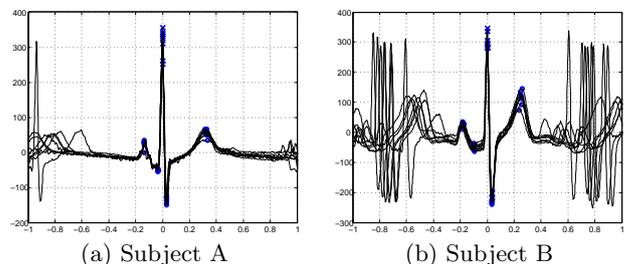


Figure 7: Eight superimposed beats for two subjects

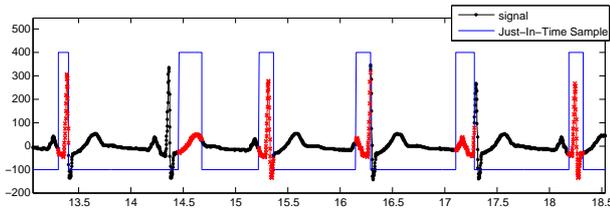


Figure 8: Illustration of just-in-time sampling

well as an important source of information for non-cardiac ailments such as diabetes, renal failure, and alcoholism [2]. Moreover, ECG signal is proven to be useful in scientific studies of psychological or behavior problems such as stress, obesity, addiction, among others [15]. In this paper, our primary interest is in the measurement of heart rate variability (HRV). Although HRV might not be enough for the diagnosis of cardiovascular disease which might need clinical-level equipment with strong batteries, long-term continuous measurement of HRV with affordable cost and easy portability could be very helpful for a wide range of applications such as the applications mentioned above. There is a large body of related works that have been done for processing and analyzing the ECG signal, which we will summarize in Section 4.

Here we give a very brief introduction to the nature of the ECG signal and introduce some medical terminology. (Please refer to [2] for a more detailed treatment of the topic.) The ECG signal captures depolarization in the cardiac muscles as they expand and contract while pumping blood through the body. In Figure 5, we present an illustration of a single beat, which is comprised of multiple features that are called the P-, Q-, R-, S-, and T-waves. Multiple waves taken together are called complexes, such as the PQR complex. The principal wave is the R-wave and is the most distinctive feature in the signal. Hence, we define the period of the heart beat (medically termed the R-R interval) as the length of time between the peaks of consecutive R-waves.

In Figure 6, we provided two examples of ECG signals. To visualize variations in the beat signals, we have eight beats of each person centered and overlapped in each subfigure of Figure 7. From these two figures we make the following observations.

1. Different people may have different beat “profiles”. However, for the same person, the shape of each beat does not change often. The eight beats superimposed in Figure 7 were randomly sampled from a period spanning an hour.
2. For different beats from the same person, the heights of the P-, Q-, R-, S-, and T-waves might be very different. However, the relative locations of these waves to one another are very consistent.

3.2 Just-in-time Overview

The problem that we address here is the measurement of heart rate variability (HRV) based on ECG signals.

Ideally, if we had the omniscient ability to predict the position of each R-wave peak, we could turn the ADC on only around the peak. Sampling a short burst centered around the peak would significantly reduce the energy consumption in the measurement device. However, because of the high variation in the ECG signal, it is a very hard problem to ex-

actly predict the next peak position. Several factors, such as baseline drift, measurement noise, not to mention the lack of information from unsampled periods make this a daunting task. The goal of just-in-time sampling is to make this prediction as accurate as possible, and to infer the location of the peak from the waveforms in the rest of the signal should the prediction procedure miss it.

In the following we review characteristics of modern ECG sensors and highlight the features that are matched to the just-in-time sampling strategy proposed in this paper. ECG sensors record the electrical activity on the body surface. A significant source of power consumption for an ECG sensor is the analog frontend. As an example a low power ECG monitor [12] uses $600\mu A$ for the analog frontend and $150\mu A$ of the ADC unit for a total of $750\mu A$. Commercial ECG ASICs can further reduce the power consumption to under $1mW$ ($300\mu A$) per channel combining analog frontend and digital A/D sections in a single optimized chip[1]. While modern ADCs have fast submillisecond wake-up times, the standard strategy of turning off the ECG sensor between the ECG samples is not feasible since the ECG analog frontend have long transients due to the time-constants of the various filters and amplifiers. Just-in-time sampling strategy relies on turning off the frontend *and* the ADC for inactive periods of fraction of a second providing sizable power savings at a nominal increase in the computation load. The required power consumption for computation is small compared to the realized savings from turning off the ECG sensor for long time windows. For a concrete example, TI MSP430’s energy consumption is only $165\mu A/MIPS$, and the next generation of chips can perform powerful computations at even lower power levels. Moreover, these chips consume very little power in stand-by (RTC) mode ($0.7\mu A$ for TI MSP430). This means that the cost of continuously performing one million instructions per second would be equivalent to keeping the ECG sensor on $\frac{165}{750} \approx 22\%$ of the time. As long as we constrain our computation costs to (say) $0.1MIPS$, the cost of the computation for the prediction is negligible in comparison with the savings in power draw from the ECG sensor. In other words, using a little energy for intelligent processing gives us large savings in the energy cost of the ECG sensor.

Novel ASIC designs that integrate Analog frontend and Successive Approximation ADCs into a single chip (e.g. [25]) recently proposed by academia might also result in dramatic power savings for EEG/ECG acquisition sensors. However, ASIC is a feasible alternative only when they are manufactured in bulk, considering the large financial investment and time needed to develop a commercially viable ASIC. In contrast, our scheme could leverage the chips now available on the market. Moreover, our algorithm can be easily integrated into the ECG ASIC through combination of analog peak detection and digital logic for active event detection to provide further power savings for ASIC design.

Our just-in-time sampling framework is comprised of two components: feature identification and prediction. The prediction component decides when to sample next, while the feature identification component deciphers the information sampled, makes an estimate of heart beat location based on it, and decides when to turn off the ADC. If the feature identification module cannot extract any useful feature to make the decision, the ADC is kept on until there is enough information to make a decision. Some instances of just-in-

time execution are illustrated in Figure 8. We see that while most R-waves are completely covered by our samples, some are only partially covered (e.g., second from the right), and other are completely missed (e.g., second from the left).

Of the abovementioned tasks, it would seem that the identification of R-waves is the easiest and should have well-established methods to solve it. However, even though there is a great deal of recent literature [3, 7, 10] on this topic, all this works attempts to robustly count/detect R-waves from a continuously sampled signal. The design target of our algorithm is fundamentally different; we need to identify the beat as accurately as possible from a small windowed sample. If the QRS complex is covered in whole or in part, we need to robustly identify it. If the QRS complex hasn't been sampled or identified, either because it has already passed or hasn't arrived yet, which is possible since the prediction is inherently difficult, we still try to identify the information contained in the area outside QRS complex, such as in the P- or T-wave (cf. Figure 5).

Here we summarize the key ideas of just-in-time sampling:

1. If we can intelligently turn on the ADC only around the peak of each beat, we would be able to save a great amount of energy in sampling.
2. Since the shapes of the beats are very similar, even if the samples do not cover the peak of the beat, as long as they cover some area inside the P-T region which contains some information to identify a peak, we are still able to estimate the location of the R-wave.
3. Since the shapes of beats are stable for the same person, we can build a profile for each subject to make our algorithm both efficient and robust.

3.3 Algorithm

As outlined above, our algorithm has a prediction and a feature identification component. For prediction, we adopt a simple solution: we estimate the size of the current R-R interval to be the minimum of the previous four R-R intervals. This is reasonable most of the time since there is usually very little variation in R-R intervals from beat to beat. Unfortunately, this is not reliable all the time (e.g., due to sudden increases in the HRV at the onset of strenuous activity, measurement error, and normal variations). In such cases, we have to rely on the feature identification component to get us back on track, which encapsulates much of the complexity of the algorithm.

The algorithm handles feature identification as follows. The device keeps the ADC on until the sampled signal shows features that are strong enough for the algorithm to guess the offset of the sampled signal. As long as some identifiable feature appears in the signal, the algorithm will estimate the location of the next peak and use this to compute the delay before taking the next set of samples. Hence, the problem comprises of two related parts: how to select robust and easily-computable features and how to decide when to start sampling again.

Another challenging aspect for our algorithm is that it cannot use absolute values of signals to detect peaks. The signal collected may be influenced by base-line drift or may have some low-frequency components. Traditional peak-detection algorithms [7] use a low-pass filter, such as the moving average filter, to filter out the low-frequency components. However, since we are using on-off sampling, low-pass filter are not effective for such short periods. Hence, all our

algorithms are designed to be robust against low-frequency noise and to capture the features of the transitions of the signal rather than to perform absolute thresholding. We next describe how we achieve these goals.

3.3.1 Threshold-Based Sections (TBS)

For convenience of processing, we first split the signal sampled into a series of rising and falling sections. The simplest way to do this would be to simply cut sections into monotonically increasing/decreasing regions. That is, suppose t_0 is the starting point and t_1, t_2, \dots are local maximum/minimum points, then each segment starting from point t_j to point t_{j+1} is considered a section. However, this may introduce too many small sections due to small fluctuations in the signal. To overcome this problem, we define a concept called Threshold-Based Sections (TBS). These threshold-based sections are still delimited by the local maximum/minimum points. However, they disregard small sections whose range of signal values is less than a threshold V . The TBS can be computed in a single pass over the data.

In the Figure 9(a), we present an example to illustrate this concept of TBS. In the figure, we can see some local maximum/minimum points skipped by the cuttings with threshold 20, but retained in the cuttings with threshold 5. There are also some small fluctuations that are skipped by both thresholds. Applying TBS to the ECG signal results in section shown in Figure 9(b). We can see that the peaks of P-, Q-, R-, S-, and T-waves are all border points of some TBS.

For segmented TBS, we next define some interesting features of each section, namely the range of the signal values in the TBS and the maximum rising/falling speeds over some window width w . Suppose the signal values in a section from s to t are $x(s), \dots, x(t)$. We define the following for each TBS:

$$\begin{aligned} \text{value range} &= \max_{k \in [s, t]} x(k) - \min_{k \in [s, t]} x(k) \\ \text{maximum rising speed} &= \max_{k \in [s, t-w]} \frac{x(k+w) - x(k)}{w} \\ \text{maximum falling speed} &= \min_{k \in [s, t-w]} \frac{x(k+w) - x(k)}{w} \end{aligned}$$

3.3.2 The first feature: VRMRS

With the help of the definitions above, we can now define the following feature of the latest section $[s, t]$ to identify the QRS complex, which we called the **VRMRS feature**:

1. If value range $\geq Th_{R-range}$ and maximum rising speed $\geq Th_{R-speed}$, then the section is marked as the rising side of an R-wave.
2. If value range $\geq Th_{R-range}$ and maximum falling speed $\leq -Th_{R-speed}$, then the section is marked as the falling side of an R wave.

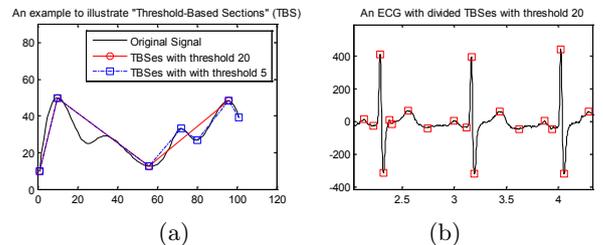


Figure 9: Illustration of the concept of TBS

| Current Section | is <i>not</i> the first sampled section | is the first sampled section |
|---|--|--|
| is in the <i>rising</i> side of R wave | The start of section would be Q Peak. Stop sampling. [Fig(a)] | Continue sampling till the changepoint (next section begins), which would be R peak [Fig(b)] |
| is in the <i>falling</i> side of R wave | The start of section would be R Peak. Stop sampling. [Fig(d)] | Continue sampling till the changepoint (next section begins), which would be S peak [Fig(c)] |

Table 1: Actions after observing the QRS detection feature

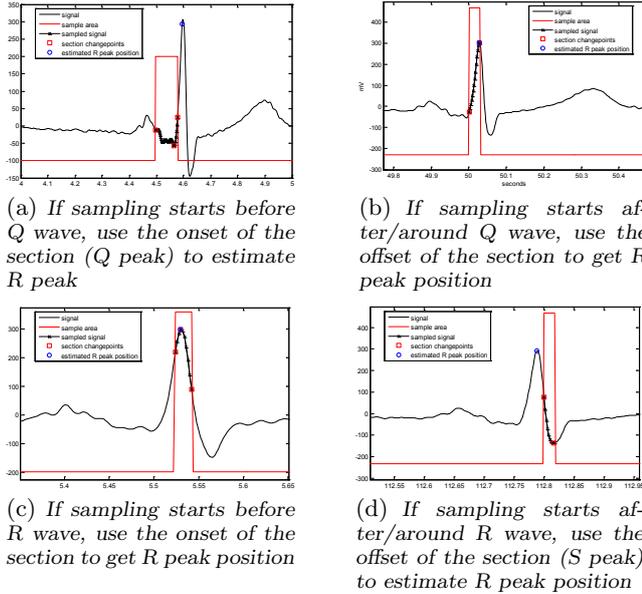


Figure 10: Use VRMRS feature to get/estimate position of R wave

For the setting of $Th_{R-range}$, we found experimentally that $V_R/3$ works well, where the V_R is the profiled height of R. Though $Th_{R-speed}$ could also be determined by profiling, we simply use $10\mu V/ms$ in all evaluations since the rapidity of the change is itself a very strong feature.

If our sensor controller identifies that the currently sampled section has one of the features described above, it estimates the location of the R-wave peak accordingly. More precisely, there are four possible situations listed in Table 1 and exemplified in Figure 10. In short, the crux of the idea is to use the location of the Q-wave or S-wave peak to estimate the location of the R-wave peak whenever we miss it.

In spite of its simplicity, this VRMRS feature works well on signals of different people. However, its weakness is that it is only useful when some portion of the QRS complex is observed. Since it is possible for the sampled signal to be outside of the QRS complex we develop the following method to recover from this.

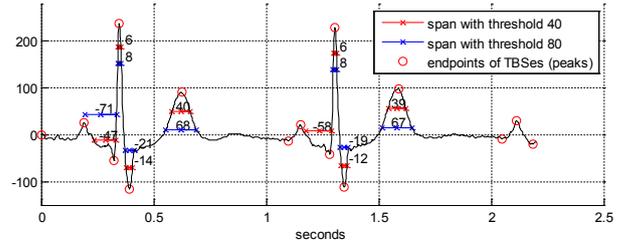


Figure 11: An illustration of peak-span

3.3.3 Mining information from the “weak” signal regions

As we have discussed above, even if we miss the QRS complex, we should still be able to estimate the location of the R-wave peak if we could identify the P- or T-wave. However, P- and T-waves are not as “strong” as the QRS complex and differ significantly from one person to another. In this section, we design another simple feature to aid in detection even in “weak” signal regions, and then discuss how we can further improve the result at the cost of additional computation.

The feature we propose is called the “peak-span”, defined as follows. Suppose i is the point connecting two TBS complexes, and hence should be a local maximum or minimum point, then the span of the peak i with threshold V' is defined by:

$$SP_{V'}(i) = \left[\min_{(k>i) \wedge (|x(k)-x(i)| \geq V')} k - \max_{(k<i) \wedge (|x(k)-x(i)| \geq V')} k \right].$$

For convenience, we let the span be negative when the peak point is a concave point.

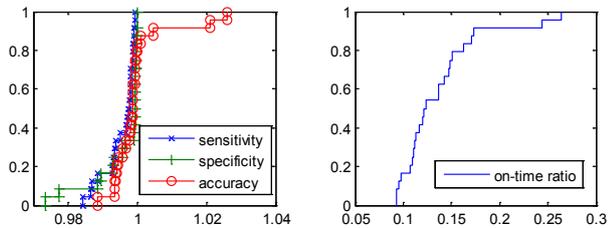
The calculation of the peak-span can easily be added to the algorithm for computing threshold-based sections without much increase in computation. Moreover, during the single-pass segmentation of the TBS complexes with threshold V , we can get the span of the peak with multiple thresholds V'_1, V'_2, \dots at the same time.

One examples of peak-span is presented in Figure 11.

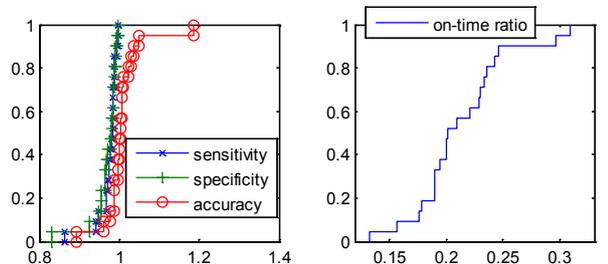
Since the shapes of either P,Q,R,S waves are very different for each people, the peak-span are very different for different people in the same way. Hence, the only way to use it is to build a profile for each person. We do this using a type of nearest-neighbor clustering algorithm; the details are omitted here in interest of space.

3.3.4 Matched Filter

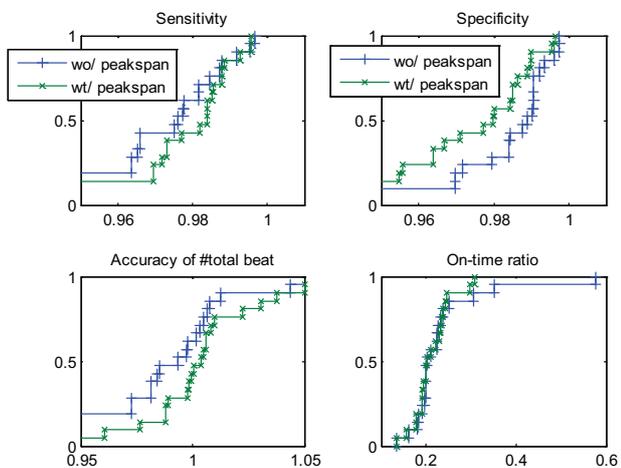
The final step of our algorithm, after finding the features described above, is to use a matched filter [22] to check the sampled region against previously recorded profiles. The reason why matched filter works here is that, in a short range of time such as 20ms, the sampled signal can be regarded as a linear transformation of the corresponding part of the profile signal. The features above help us identify a handful of profiles signals to be matched against, allowing us to pick the closest fit. Since this matching process is only needed when the features are found (as opposed to every point in the sampled region), it is well within the computation budget of 0.1 MIPS.



(a) The cumulative distribution function (y-axis) of the results of Apnea Data



(b) The cumulative distribution function (y-axis) of the results of Fieldstream's Data



(c) The cumulative distribution function (y-axis) of the results of Fieldstream's Data (Comparing two methods)

Figure 12: Statistics of Results

3.4 Evaluation

In this section, we evaluate the performance of our algorithms on two real data-sources. The first data source is the Apnea-ECG Database[‡]. We use the first 24 traces, each of which is about three hours long. Our second source is data that we collected in the field, which we refer to as FieldStream data. Our data is comprised of 24 traces, also three hours in length. For some traces from the Apnea Database, the first 10-20 minutes were flat-lined, and hence we manually had to remove them. For both the Apnea and FieldStream data, we first upsampled the data to 500Hz since they were recorded at 50–100Hz to save storage space. We did not perform any pre-processing of the data other than those mentioned above, such as low- or high-pass filters to remove power-line or base-line drift.

We used the following statistical measures to evaluate the

results:

- **Sensitivity:** the total number of identified correct beats, divided by the total number of real beats.
- **Specificity:** the total number of identified correct beats, divided by the total number of identified beats.
- **Accuracy:** the total number of beats identified divided by the total number of real beats.
- **On-Time ratio:** the total on-time of the ADC, divided by the total time.

We generated ground truth to compare our results using code by Rupert Ortner[§], which implements the method proposed in [3]. This is not a perfect solution as it has some errors such as mistakenly marking some thin S- or T-waves as R-waves and hence counting more beats, or missing short R-waves. These errors were found using manual inspection. When we changed the upsampled frequency from 500Hz to 100Hz, we found a 0.05–1% change in the number of beats detected by the program, which means the “ground truth” we use has at least 0.05–1% error itself.

The cumulative distribution function of the statistics listed above are shown in Figure 12. In Figure 12(a) and (b), we present the result for the Apnea and FieldStream data separately. The results are achieved after using all methods discussed previously, including the two features and matched filter. For the Apnea data, we achieved specificity and sensitivity above 96% in most cases with on-time ratio less than 20%. We also observed that the results for the FieldStream traces were slightly worse than the results for the Apnea data. This is expected, since data collected in field understandably has more noise compared with that collected in laboratory settings.

In Figure 12(c), we present the result on Fieldstream data with two combinations of methods, only using the first feature (the VRMRS feature described in Section 3.3.2) and using all methods including “peak-span”. We observe that the peak-span method, which tries to detect the “weak” signals outside the QRS complex, improves the sensitivity while slightly lowering the specificity. The reason for this is that the simpler method reacts quickly but perhaps not very accurately. The accuracy after using the peak-span method is very close to 1 in most cases. Since even a perfect algorithm may have up to 0.05–1% error in the total number of beats, this result is fairly promising. In our experiments, the ADC-on-time ratio was within 10–25% almost all the time.

4. RELATED WORK

There is a large body of research on detecting QRS complexes in ECG signals [3, 7, 10]. [3] proposes a frequency-domain method which decomposes the signal by several filter banks and makes decisions based on a heuristic combination of the results from each sub-band. We use this method to generate the ground truth for our evaluation. [7, 10] make use of time-domain heuristics, although they also need some frequency-domain pre-processing. The features they use includes high amplitude, steep edges, and sharp peaks. In [7], the expected location of the next beat is also taken into account since there is small probability of seeing two beats very close to one another.

[‡]<http://www.physionet.org>

[§]Downloaded from <http://www.koders.com/matlab/>

While all these methods are designed for real-time implementation, our goals are very different. Instead of having the entire continuous signal, we assume that we have short bursts of samples with which to work. As a result, the features used by earlier work cannot be directly used here. However, we still try to draw from previous work. For example, our value range feature is similar to detecting high amplitudes and our maximum rising/falling speeds are similar to the threshold of slopes in [7].

There has also been work done on the detection and elimination of artifacts in ECG data. [6] models the effect of missed and spurious R-waves and demonstrates that every beat is important to heart period variability analysis. [5] proposes methods to identify artifacts. We leave the incorporation of these methods into our framework as future work.

Finally, there has been a slew of work on designing prototype hardware for portable and low-power ECG monitors [11, 17, 16, 13]. However, all this work focuses on the hardware aspect of the problem and typically employs simple duty cycling to capture the signal. As a contrast, our algorithms could intelligently turn off the ADCs (and also the entire sensing chain) and save the majority of power. To the best of our knowledge, ours is the first work to propose intelligent cycling for capturing the target signal in a power-efficient manner.

5. CONCLUSIONS

In this paper we explore different mechanisms for sampling short bursts of physiological signals with the view of extending the battery life of wearable sensing devices. We present two different approaches that are both capable of large energy savings by turning off the ADC on the device without a significant loss in measurement information. In the case of pre-filtering, we are able to reconstruct a complete band-limited signal with the help of specialized analog hardware. Just-in-time sampling, on the other hand, requires standard hardware and can be implemented on current devices. However, just-in-time sampling is not as generic as pre-filtering and could only capture some interesting part of the signal, rather than reconstruction of the complete signal. Currently, it is only tailored for detecting the R waves of ECG signals and hence measuring heart rate variability. We show, via simulation on data sets collected from two different sources, that just-in-time sampling is extremely effective in capturing the R waves, while reducing the power consumption by an order of magnitude.

Future work includes capturing locations of every n th heartbeat (for small n) and extrapolating the other positions from these. This could conceivably cut down the number of samples needed by an order of magnitude and result in further extension of the battery life. We also hope to find other physiological signals (e.g., respiration rate) that may benefit from our proposed methodology.

6. REFERENCES

- [1] TI's ADS1298 chip. <http://focus.ti.com/docs/prod/folders/print/ads1298.html>.
- [2] U. Acharya, J. Suri, J. Span, and S. Krishnan, editors. *Advances in Cardiac Signal Processing*. Springer, 2007.
- [3] V. Afonso, W. Tompkins, T. Nguyen, and S. Luo. Ecg beat detection using filter banks. *Biomedical Engineering, IEEE Transactions on*, 46(2):192–202, feb. 1999.
- [4] A. Y. Benbasat and J. A. Paradiso. A framework for the automated generation of power-efficient classifiers for embedded sensor nodes. In *Sensys*, 2007.
- [5] G. Berntson, K. Quigley, J. Jang, and S. Boysen. An approach to artifact identification: Application to heart period data. *Psychophysiology*, 27(5):586–598, 1990.
- [6] G. Berntson and J. Stowell. ECG artifacts and heart period variability: Don't miss a beat! *Psychophysiology*, 35(01):127–132, 2001.
- [7] I. Christov. Real time electrocardiogram qrs detection using combined adaptive threshold. *BioMedical Engineering OnLine*, 3(1):28, 2004.
- [8] M. Coye, A. Haselkorn, and S. DeMello. Remote patient management: technology-enabled innovation and evolving business models for chronic disease care. *Health Affairs*, 28(1):126, 2009.
- [9] A. Dias and et. al. Measuring Physical Activity With Sensors: A Qualitative Study. *Studies in health technology and informatics*, 2009.
- [10] I. Dotsinsky and T. Stoyanov. Ventricular beat detection in single channel electrocardiograms. *BioMedical Engineering OnLine*, 3(1):3, 2004.
- [11] R. Fensli, E. Gunnarson, and T. Gunderson. A wearable ecg-recording system for continuous arrhythmia monitoring in a wireless tele-home-care situation. In *IEEE CBMS*, 2005.
- [12] R. Fonseca, P. Dutta, P. Levis, and I. Stoica. Quanto: Tracking energy in networked embedded systems. In *OSDI*, 2007.
- [13] T. R. Fulford-Jones, G.-Y. Wei, and M. Welsh. A portable, low-power, wireless two-lead ekg system. In *IEEE EMBS*, 2004.
- [14] J. Houtveen and E. de Geus. Noninvasive Psychophysiological Ambulatory Recordings. *European Psychologist*, 14(2):132–141, 2009.
- [15] S. Kreibitz. *Autonomic Nervous System Activity in Emotion: A Review*. *Biological Psychology*, 2010.
- [16] D. Lucani, G. Cataldo, J. Cruz, G. Villegas, and S. Wong. A portable ecg monitoring device with bluetooth and holter capabilities for telemedicine applications. In *IEEE EMBS*, 2006.
- [17] C. Park, P. H. Chou, Y. Bai, R. Matthews, and A. Hibbs. An ultra-wearable, wireless, low power ecg monitoring system. In *BioCAS*, 2006.
- [18] J. Romberg. Compressive sensing by random convolution. *SIAM J. Imaging Sci.*, 2(4):1098–1128, 2009.
- [19] D. Slepian. On bandwidth. *Proceedings of the IEEE*, 64(3):292–300, March 1976.
- [20] J. A. Tropp, J. N. Laska, M. F. Duarte, J. Romberg, and R. G. Baraniuk. Beyond Nyquist: efficient sampling of sparse bandlimited signals. *IEEE Trans. Inform. Theory*, 56(1), January 2010.
- [21] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk. Random filters for compressive sampling and reconstruction. In *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, volume 3, pages III–872–875, Toulouse, France, May 2006.
- [22] S. V. Vaseghi. *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons.
- [23] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Signal Proc.*, 50:1417–1428, June 2002.
- [24] F. Wilhelm and P. Grossman. Emotions beyond the laboratory: Theoretical fundamentals, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, 2010.
- [25] R. Yazicioglu, P. Merken, R. Puers, and C. Van Hoof. A 200 μ v eight-channel acquisition asic for ambulatory ecg systems. In *International Solid-State Circuits Conference (ISSCC)*, 2008.