Analysis of Heart Rate Variability on a Massively Parallel Processor

Suchendra M. Bhandarkar[a] Sridhar Chirravuri[a] David Whitmire[b]
[a]Department of Computer Science
[b]Department of Biological and Agricultural Engineering
The University of Georgia
Athens, GA 30602-7404, USA

Abstract

The K_2 entropy, and the correlation dimension are commonly used measures of chaotic activity in complex dynamic systems. The design and implementation of a massively parallel algorithm on the MasPar MP-2 for the computation of K_2 entropy and the correlation dimension is presented. The algorithm is used to compute the K_2 entropy and correlation dimension of experimental heart rate data. Experimental results on canine subjects support the claim that K_2 entropy and the correlation dimension can be used as a measure of the heart rate variability and the level of chaos present in the nonlinear dynamics underlying the heart rate data. The parallel algorithm has potential for use as a tool for cardiac health assessment and diagnosis of certain cardiac ailments.

1 Introduction

Chaos theory is often used to provide an underlying model with which to formally study and characterize complex nonlinear dynamic systems. A *chaotic* system is essentially a deterministic system not a stochastic system[1]. Following the pioneering work by Lorenz[4], many dynamic systems which were earlier considered random, have been reclassified as chaotic.

An EKG measures the electrical activity of the heart and when plotted as a function of time, exhibits certain characteristic peaks which reflect the dominant electrocardiac phases. The R peak in the EKG is a dominant peak of cardiological significance[5]. The instantaneous heart rate is computed as the inverse of the time period between two successive R peaks and is observed to be a highly complex function of time. In most biological systems, the instantaneous heart rate is affected by several direct and indirect physiological and environmental factors. The term *heart rate variability* (HRV) denotes the fact that the instantaneous heart rate is a complex nonlinear function of time.

It is difficult to reconstruct a complex dynamic system such as the heart, which is believed to have several degrees of freedom, using only a relatively few time series data points from the EKG. Time series analysis techniques based on computation of correlation-dimension[3], Lyapunov exponents[1], Kolmogorov (i.e. K_2) entropy[3] and Approximate Entropy (ApEn)[6] been recently developed to characterize the nonlinear chaotic dynamics underlying complex physiological systems such as the heart[2]. A widely reported measure for chaotic time series analysis is the K_2 entropy[3].

The HRV is a measure of the nonlinear complexity of electrocardiac activity. Qualitative analysis of the HRV is often used by the cardiologist as an indicator of cardiovascular health[6]. The heart rate of healthy persons is known to have higher variability (i.e. higher degree of chaos) than that of cardiac patients. Due to its close association with cardiovascular health, it is important to be able to characterize and measure HRV from the EKG to enable *quantitative* analysis of the underlying electrocardiac activity.

1.1 K_2 Entropy

The K_2 entropy is an estimate of the lower bound for the metric Kolmogorov entropy computed from time series data[3]. The algorithm used to compute K_2 entropy can also be used to estimate a good lower bound for the correlation dimension of complex dynamic systems. The computation of K_2 entropy from time series data is described as follows:

Consider a time series with N points. Let x_i denote the *ith* point in the time series. The time series

Proceedings of the 1996 International Conference on Parallel Processing (ICPP '96)

data can be vectorized by constructing d-dimensional vectors where each vector consists of d consecutive points in the time series $(x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+d-1})$. In such a d-dimensional vectorization (or embedding) of the time series data, the correlation integral $C_d(\epsilon)$ is defined as:

$$C_d(\epsilon) = \lim_{N\to\infty} \left[\frac{1}{N^2} \left| \left\{ (n,m) : \left[\sum_{i=1}^{d} |x_{n+i} - x_{m+i}|^2 \right]^{1/2} < \epsilon \right\} \right| \right] \quad (1)$$

The K_2 entropy, is given by:

$$K_2 = \lim_{\epsilon\to 0} \lim_{d\to\infty} \left[\frac{1}{\tau} \ln \frac{C_d(\epsilon)}{C_{d+1}(\epsilon)} \right] \quad (2)$$

In our case, the Takens constant τ was assumed to be unity. The K_2 entropy is a *quantitative* measure of the complexity or degree of chaotic behavior of the dynamic system underlying the time series data. The K_2 entropy values are interpreted and classified as follows:
(i) K_2 entropy $= 0$ implies that the dynamic system is regular i.e. either *constant* or *periodic*,
(ii) K_2 entropy $>> 1$ (approaching ∞) implies that the dynamic system is *random*, and
(iii) $0 < K_2$ entropy $<< \infty$ implies that the dynamic system is *chaotic*.

For the purpose of verification of the algorithms for computation of $C_d(\epsilon)$ and the K_2 entropy, the Henon system was considered. The Henon system is a model chaotic system whose time series data is generated using the recurrence equations[1]:

$$\begin{aligned} X_{n+1} &= \alpha - X_n^2 + \beta Y_n \\ Y_{n+1} &= X_n \text{ where } \alpha = 1.4,\ \beta = 0.3 \end{aligned} \quad (3)$$

The Henon system is known to have a characteristic K_2 entropy of 0.325 and a correlation dimension of 1.22[3].

Figure 1 shows the plot of $C_d(\epsilon)$ as a function of ϵ on a log-log scale for various values of the embedding dimension d for the Henon system. This plot is referred to as the C_d plot and the curves in the plot are referred to as the d-curves. With reference to Figure 1, the region where all the d-curves merge into a single horizontal line is called the *saturation* region. The *depopulation* region in the C_d plot is characterized by the abrupt termination of the d-curves for small values of ϵ which is particularly noticeable in the case of d-curves at higher d values. The *scaling* region in

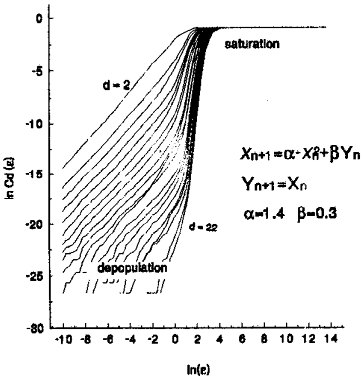

Figure 1: Correlation integral curves for the Henon system

the correlation integral plot is defined as the one between the saturation and depopulation regions where all the d-curves have a constant slope and are parallel to each other. The K_2 entropy is computed as the asymptotic distance between successive d-curves in the scaling region in the limit as d approaches infinity. The correlation-dimension is computed as the asymptotic slope of the d-curves in the scaling region in the limit as d approaches infinity.

The computational complexity of the serial algorithm for the computation of K_2 entropy stems primarily from the computation of C_d (equation (1)) where one is required to compute the Euclidean distance between $N \cdot (N-1)/2$ vector pairs in $\mathcal{R}^d$. Thus the computational complexity of the serial algorithm is $O(N^2)$, which for large values of N limits the usefulness of the algorithm for time-critical and real-life applications.

2 K_2 Entropy Computation on the MasPar MP-2

The MasPar MP-2 is a massively parallel SIMD computer with the processing elements (PE's) interconnected in a 2-D toroidal mesh topology. A data parallel algorithm on the MasPar MP-2 is coded in the MPL programming language. The *plural* construct in MPL allows the replication of a variable in the local memory of each PE. Data parallelism stems from operations carried out on plural variables since a plural

Proceedings of the 1996 International Conference on Parallel Processing (ICPP '96)
0190-3918/96 $10.00 © 1996 IEEE

Table 1: Run time statistics for computation of C_d on various systems

Run time (in minutes)				
	Number of Data Points (N)			
Systems	2K	4K	8K	16K
SUN Sparc-2	126	480	-	-
Pentium P5-60	311	960	-	-
Cray YMP C-90	12	40	156	612
MasPar MP-2	3.2	6.5	14.9	40

variable can be operated upon by several PE's concurrently. The MasPar MP-2 offers nearest-neighbor communication along the eight principal directions on the 2-D toroidal mesh via the *xnet* MPL primitive and global communication via a global router network.

The algorithm for computation of $C_d(\epsilon)$ was parallelized on the MasPar MP-2 and the resulting data parallel algorithm is depicted in Figure 2. With reference to Figure 2, the set of instructions contained in the *for* loop with index d counts the number of vector pairs in a d-dimensional embedding of the time series data with inter-vector distances in $\mathcal{R}^d \leq \epsilon$. The algorithm then computes the distances between *nproc* vector pairs concurrently where *nproc* is the number of PE's. For a given value of d, the *for* loops indexed by i and j enable one to repeat the *for* loop indexed by n to count all vector pairs with inter-vector distances in $\mathcal{R}^d \leq \epsilon$. Thus, the *for* loops indexed by i and j assist in breaking up the counting process into a number of smaller *for* loops of size $\leq$ *nproc*. The upper bounds for the variables i, j, and *repeat_count* can be determined at compile time. Thus, the data parallel algorithm in Figure 2 is scalable over different sizes of the PE array *nproc* and the data set N.

The run time statistics of the algorithm for computation of C_d on various computing platforms, are tabulated in Table 1. The data parallel algorithm was implemented on a 2048 processor MasPar MP-2 at the University of Georgia with a value of $nproc =$ 2048 for each of the data sets. The implementation on the Cray YMP C-90 was optimized using several of the vectorization constructs provided by the compiler as well as by ensuring that a healthy memory/cache hit ratio was consistently maintained during data access. The data structures were designed to be small enough to fit into the main memory (RAM) of the Cray YMP C-90 (1 GByte), the SUN Sparc-2 workstation (32 MBytes) and the Pentium P5-60 workstation (32 MBytes) without the need for swapping. Our results indicate that the execution times for the

```
int i, d, k, n, N, FinalD, repeat_count;
float eps,increment=6000.0,eps_final=120000.0;
plural float X, distance;
for (i=0; i < 200; i++)  CompareVec[i] = 0.0;
 for (d = 1; d < FinalD; d++)
 {
  repeat_count = (int) N/nproc ;
  if ((N % nproc) > 0)
     repeat_count = repeat_count + 1;
  for (j = 0; j < (repeat_count); j++)
  {
   for (n = (j * nproc), distance = 0;
   (n < ((j+1)*nproc)) && (n < (N-(2*d-1))); n++)
   {
    for (i = j; i < (repeat_count); i++)
    {
     if (((iproc + i*nproc) >= n) &&
          ((iproc + i*nproc + d) < N-1))
     {
      for (k = 0, distance = 0; k <= (d-1); k++)
      {
       distance = distance +
        (X[n+k] - X[1+iproc+i*nproc+k])*
            (X[n+k] - X[1+iproc+i*nproc+k]);
      }
     }
     distance = sqrt(distance);
     for (eps = 0; eps <= eps_final;
                     eps = eps + increment)
     {
      if ( distance <= eps )
      {
       .......
      }
       .......
     }
     .......
    }
   }
  }
 }
```

Figure 2: Data parallel algorithm for computation of C_d

Proceedings of the 1996 International Conference on Parallel Processing (ICPP '96)
0190-3918/96 $10.00

Table 2: Comparison of results for the model Henon system. **K_2**: K_2 entropy, **CD**: Correlation Dimension

	K_2	**CD**
Grassberger and Procaccia	0.325	1.22
MasPar MP-2 Implementation	0.343	1.26

serial implementation scale as $O(N^2)$, where N is the number of data points in the time series. The massively parallel implementation on the MasPar MP-2 was observed to scale as $O(N^2/P)$ where P is the total number of PE's participating in the computation (i.e. *nproc*).

We have implemented a serial exhaustive search algorithm for determination of the optimal scaling region in the C_d plot where all the d-curves are optimally parallel to each other. The objective function for the exhaustive search algorithm is formulated so as to have a minimum in the scaling region. The global minimum on the objective function surface yields the angle of the slope and y-intercept of a line that is optimally orthogonal to all the d-curves and whose intersection points with the d-curves lie entirely within the scaling region. These intersection points constitute the *optimal scaling region* and are used to determine the values of the K_2 entropy and correlation-dimension. The values thus computed agree well with the values reported by Grassberger and Procaccia[3] (Table 2).

3 Experimental Results

The heart rate data of a cohort of ethanol induced canine subjects was obtained from the Department of Anesthesiology, Medical College of Georgia, Augusta, Georgia. The heart rate data was analyzed on the MasPar MP-2. The average run time of the data parallel algorithm for the computation of the C_d plot from a data set of 1800 heart rate points was less than three minutes. The observed speedup on a data size of 1800 on 2048 PE's was approximately 1000. The serial algorithm for determination of the optimal scaling region in the C_d plot was used to compute the asymptotic values of the K_2 entropy and the correlation dimension. The serial algorithm ran in less than 30 seconds on the DEC 5000 workstation which serves as the front end host for the MasPar MP-2 system.

It was our experimental observation that the heart rate data of a typical subject in this cohort under non-intoxicated conditions exhibited a K_2 entropy value in the range [0.8, 1.0] and a correlation dimension value in the range [2.4, 2.6]. With increasing blood ethanol concentration, the K_2 entropy values dropped to the range [0.66, 0.68] and the correlation dimension values to the range [1.15, 1.45]. It has been empirically shown that ethanol intoxication results in reduction in the HRV. Our experiments have shown that the values of the K_2 entropy and correlation dimension computed from experimental heart rate data are a decreasing function of the blood ethanol concentration. Our experimental results support our hypothesis that the values of the K_2 entropy and correlation-dimension provide a quantitative measure of the HRV.

4 Conclusions and Future Research

The present data parallel implementation on the MasPar MP-2 satisfies current cardiological requirements for off-line analysis and also provides scope for improved precision with data sets of larger size (Table 1). On-line analysis of EKG data would require turn around times in the millisecond range which is 3–4 orders of magnitude smaller than the turn around time of our current massively parallel implementation. We intend to address this issue in our future research.

References

[1] G.L. Baker, and J.P. Gollub, *Chaotic Dynamics, An Introduction*, Cambridge University Press, New York, NY, 1990.

[2] L. Glass, and D. Kaplan, "Time-series analysis of complex dynamics in physiology and medicine", *Medical Progress Through Technology*, 19(3), 1993, pp. 115-128.

[3] P. Grassberger, and I. Procaccia, "Estimation of Kolmogorov entropy from a chaotic signal", *Phy. Rev. A*, 28(4), 1983, pp. 2591-2593.

[4] E. Lorenz, "Deterministic non-periodic flow", *J. Atmosph. Sci.*, 20(2), 1963, pp. 131-141.

[5] R.E. Phillips, and M.K. Feeney, *The Cardiac Rhythms*, W.B. Saunders, Philadelphia, PA, 1973.

[6] S.M. Pincus, and R.R. Viscerello, "Approximate entropy: A regularity measure for fetal heart-rate analysis", *Obstetrics & Gynecology*, 79(2), 1992, pp. 249-255.

Proceedings of the 1996 International Conference on Parallel Processing (ICPP '96)
0190-3918/96 $10.00