Real-time Recognition of Continuous-time Biomedical Signals Using the Diffusion Network

Yu-Su Hsu, Tang-Jung Chiu, and Hsin Chen

***Abstract*— Real-time recognition of multichannel, continuous-time physiological signals has been crucial for the development of implantable biomedical devices. This work investigates the feasibility of using the Diffusion Network, a stochastic recurrent neural network, to recognise continuous-time biomedical signals. In addition, a hardware-friendly approach for achieving real-time recognition is proposed and tested with both artificial and real biomedical data. Based on this approach, the Diffusion Network is demonstrated to exhibit great tolerance against noise and drifts in continuous-time signals being classified.**

I. Introduction

Many physiological signals (e.g. neural activities) are continuous-time in nature. Real-time recognition of multichannel, continuous-time physiological signals is thus one of the main challenges in the development of implantable biomedical devices, which aims to improve treatments to illness by delivering bio-feedbacks or by controlling prosthetic devices in real-time[1]. The difficulty of recognising continuous-time biomedical signals lies not only in the dimension of data, but also in the noisy and drifting nature of biomedical signals.

The Diffusion Network proposed by Movellan [2], [3] is a stochastic recurrent network whose stochastic dynamics can be trained to model the probability distributions of continuous-time sequences by the Monte-Carlo Expectation-Maximisation (EM) algorithm. As the stochasticity in many probability models has proved useful for generalising the natural variability in data [4][5], the Diffusion Network is potentially useful for modelling drifting and noisy physiological data. In addition, the analogy between the Diffusion Network and the Continuous Restricted Boltzmann Machine [4] suggests that the Diffusion Network is amenable to VLSI implementation[6]. It is thus of great interests to develop an intelligent embedded system capable of recognising multichannel, continuous-time biomedical signals in real time, based on the Diffusion Network.

However, the capability of the Diffusion Network in modelling biomedical data is seldom explored. Classification based on the likelihood under the model is not favourable for real-time recognition nor for hardware implementation. Therefore, this work examines the feasibility of using the Diffusion Network to model biomedical signals, and identifies a hardware-amenable method of recognising signals in real time with trained Diffusion Networks. The tolerance of the Diffusion Network against noise and drifts in data to be classified is also probed. Finally, the capability of the Diffusion Network is demonstrated with the real heartbeat data.

II. The Diffusion Network

The Diffusion Network [2][3] consists of continuous-time, continuous-valued, stochastic units(neurons) with full, recurrent connections, as illustrated in Fig.1a. The state of each stochastic unit is a random variable changing with time, i.e. the dynamics of the state is a stochastic process governed by stochastic differential equations. Let $x_i(t)$ represent the state of the unit i at time t, and w_{ij} the coupling strength from the unit i to the unit j. The stochastic dynamics of the Diffusion Network is governed by

$$dx_i(t) = \mu_i(t) \cdot dt + \sigma \cdot dB_i(t) \tag{1}$$

where $\mu_i(t)$ is a deterministic *drift* term, σ a constant, and $dB_i(t)$ the *Brownian* motion, a stochastic process whose increment $(dB_i(t+dt) - dB_i(t))$ is a Gaussian random variable with zero mean and variance dt [7]. The inclusion of the Brownian motion attributes to the stochasticity in the Diffusion Network, enriching greatly the representational capability of the Diffusion Network [3], as compared to deterministic recurrent neural networks [3]. The drift $\mu_i(t)$ in Eq.(2) is defined as

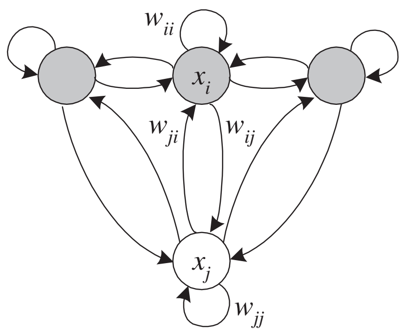
$$\mu_i(t) = \frac{1}{C_i}\left(\sum_j w_{ij} s_j(t) - x_i(t)/R_i\right) \tag{2}$$

where $C_i > 0$ and $R_i > 0$ are adaptable parameters called *input capacitance* and *transmembrane resistance* of the stochastic unit, and $s_i(t)$ represents the nonlinear transform of $x_i(t)$ defined as $s_i(t) = \varphi(x_i) = \tanh(a_i \cdot x_i)$, in which a_i is a constant controlling the slope of the nonlinear function.

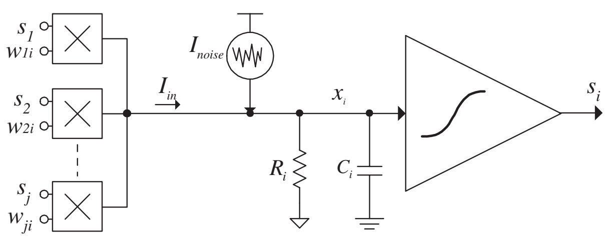
The Diffusion Network is a *generative* model able to model the probability distributions of sequences by the Monte-Carlo EM algorithm[3] With stochastic units divided into *visible* and *hidden units*, as shown in Fig.1a, the modelling refers to optimising parameter values, so that the trained Diffusion Network is able to "regenerate" the sequences as the stochastic dynamics of its visible neurons, and the regenerated sequences possess the same probability distribution as the data modelled. Therefore, the number of visible neurons is equal to the dimension of sequences to be modelled, while the number of hidden neurons is chosen to be the minimum number of hidden variables required

Yu-Su Hsu, Tang-Jung Chiu, and Hsin Chen are with the Institute of Electronics Engineering, The Tsing-Hua University, HsinChu, Taiwan (email: hchen@ee.nthu.edu.tw).

This work was supported by the National Science Council, Taiwan under the Grant Code NSC $96-2221-E-007-167-MY3$



(a)



(b)

Fig. 1. (a)The diagram of a Diffusion Network with one visible (white-coloured) and three hidden(grey-coloured) units. (b)The stochastic unit of the Diffusion Network in terms of electrical equivalent circuits.

for modelling high-order correlations among different dimensions of the sequences. The latter is normally identified through a trial process. As the Monte-Carlo EM algorithm maximises the likelihood of regenerating the modelled sequences, the trained Diffusion Network can classify whether an unknown sequence x_o^T belongs to the training dataset by calculating the likelihood of regenerating the sequence according to Eq.(4).

Consider a single run of a Diffusion Network with n visible and m hidden units in the time interval $[0, T]$. Given $x_i(0) = 0$ for all units, the dynamics of all units are sampled according to Eq.(1) and Eq.(2)[1]. Let λ denote the set of all parameters $w_{ij}, C_i, R_i,\ x_o : [0, T] \to \mathbb{R}^n$ an n-dimensional sequence representing the dynamics of visible units sampled during $[0, T]$, and $x_h : [0, T] \to \mathbb{R}^m$ an m-dimensional sequence representing the sampled dynamics of hidden units. The log-likelihood for the Diffusion Network with parameter λ to generate the joint dynamics (x_o, x_h) is given as [3]

$$\log p^{\lambda}(x_o, x_h) = \frac{1}{\sigma^2}\sum_{i=1}^{n}\int_0^T \mu_i(t)dx_i(t) - \frac{1}{2\sigma^2}\int_0^T \mu_i(t)^2 dt \quad (3)$$

To calculate the probability of generating a specific path x_o^T, the marginal probability $p^{\lambda}(x_o^T)$ can be estimated by *clamping* the dynamics of visible units to be x_o^T[3], sampling the corresponding dynamics of hidden units for l times, and marginalising over hidden samples as

$$\log \hat{p}^{\lambda}(x_o^T) = \sum_l \log p^{\lambda}(x_o, x_h^l) \quad (4)$$

where x_h^l represent the l-th Monte-Carlo sampled hidden dynamics. The estimated marginal likelihood in Eq.(4) is useful not only for sequence classification, but also for the indication of whether the EM algorithm has optimised the parameters, which is important for knowing when to stop a training process.

III. Hardware amenability of the Diffusion Network

Hardware implementation of the Diffusion Network is important for running continuous-valued, continuous-time stochastic dynamics of multiple units in parallel and in real-time. Exploiting the diffusion process inherent in analogue circuits, Fig.1b shows the translation of the stochastic differential equation in Eq.(1) into an equivalent-circuit model for the stochastic unit. The diagram indicates that the Diffusion Network can be implemented simply by integrating analogue multipliers, capacitors, resistors, noise generators, and sigmoid circuits. Furthermore, [4] has proved that the Diffusion Network is simply different from the Continuous Restricted Boltzmann Machine(CRBM) by the inclusion of parameters R_i and C_i. As the CRBM has been demonstrated in Very-Large-Sale-Integration (VLSI) implementation[6], the analogy between the Diffusion Network and the CRBM indicates that the Diffusion Network can be realised in VLSI simply by adding one resistor and one capacitor into the CRBM system in VLSI, whose stochastic units contain all the components expect for R_i and C_i in Fig.1. Therefore, the VLSI implementation of the Diffusion Network is undoubtedly feasible and potentially useful as a intelligent system for many implantable devices.

From another point of view, an analogy also exists between the Diffusion Network and the Cellular Neural Network [8], [9], a deterministic recurrent network with localised connections. The Diffusion Network differs from the Cellular Neural Network mainly by the inclusion of the Browian motion. The successful development of cellular neural network in both applications and VLSI implementation [10], [11], [12] suggests not only the hardware amenability of the Diffusion Network but also the rich computational power of the Diffusion Network in VLSI.

IV. Real-time Recognition Based on the Deterministic Dynamics of Hidden Units

Although calculating the log-likelihood of an unknown sequence is one mathematically-plausible way of classification, the VLSI implementation of Eq.(4) requires complicated circuits and discourages real-time classification. As the dynamics of hidden units must depend on and correlate to those of visible units owing to full connection among units, we investigate the possibility of classifying data according to the deterministic dynamics of hidden units when the dynamics of visible neurons are *clamped* to sequences to be classified.

[1]Discrete-time approximation to the stochastic differential equations was adopted in the numerical simulation in a computer

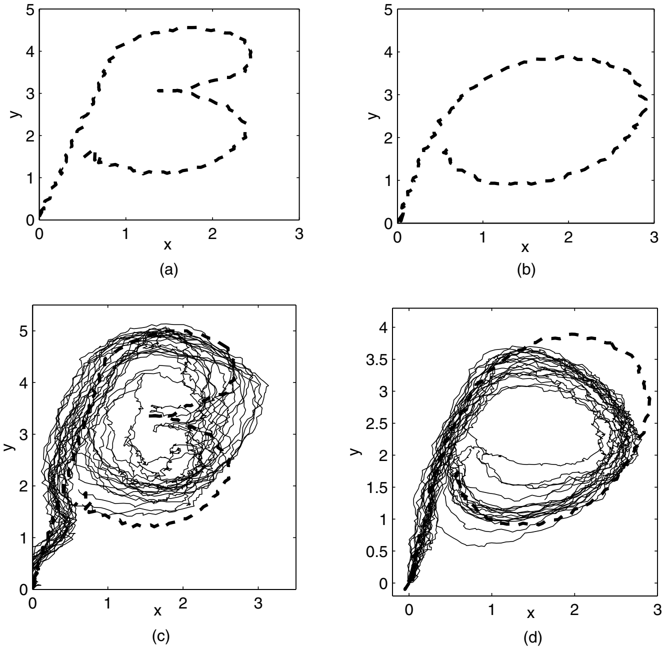


Fig. 2. (The handwritten character (a)β and (b)ρ. The 20 sequences of visible dynamics regenerated by the Diffusion Network trained on (c)β and (d)ρ. The dashed lines are training data in (a) and (b).

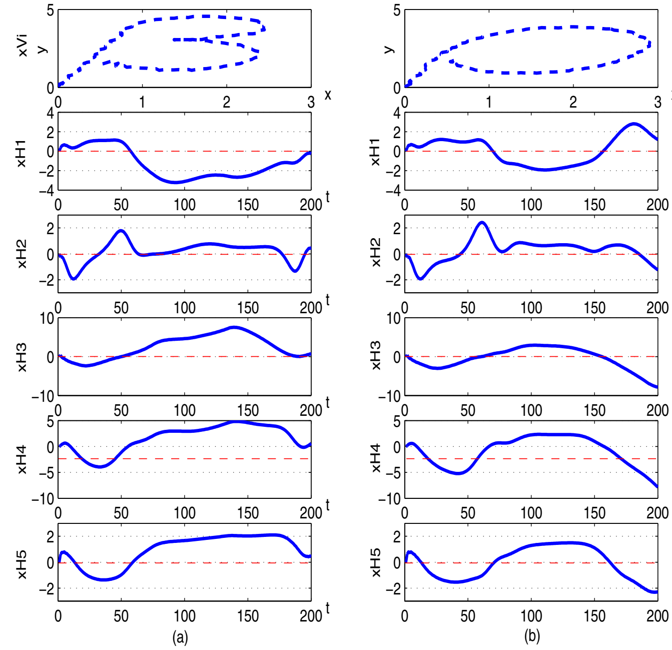


Fig. 3. The deterministic dynamics of hidden units of the Diffusion Network trained on handwritten ρ when the dynamics of the visible units are clamped to be (a)β and (b)ρ

The feasibility of this classification method is examined by training the Diffusion Network to recognise hand-written characters β and ρ, as shown in Fig.2a and b, respectively. Both characters are sampled as discrete-time, two-dimensional sequences with 200 sampled points across time, $t \in [0, 200]$. The characters are also carefully written to have the same initial and end points, as well as to differ only by the extra inward bending for β. This makes the classification task more difficult and therefore suitable for probing the ability of the Diffusion Network to distinguish between the two sequences by the proposed method

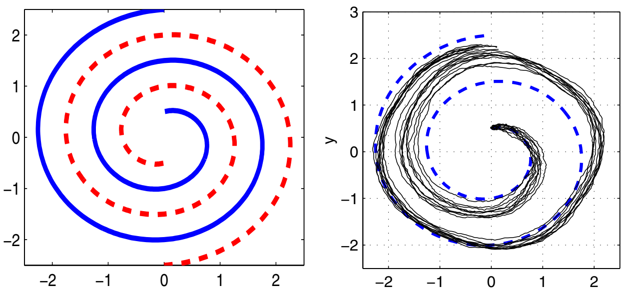
A Diffusion Network with two visible and five hidden units was trained to model one of the hand-written characters. With $\sigma = 0.2$, $R_i = 1$, and $a_i = 1$, parameters w_{ij} and C_i were trained for 100 epochs, at which the likelihood calculated according to Eq.(4) no longer has significant increment with further parameter updates. Let $x_{Vi}(t)$ and $x_{Hi}(t)$ denote the dynamics of visible unit i and hidden unit j, respectively, sampled according to Eq.(1). Given $x_{Vi}(0) = x_{Hi}(0) = 0$, the 20 sequences sampled from the visible units of the Diffusion Network trained on β and ρ are shown in Fig.2c and Fig.2d, respectively. The overlap between the training and regenerated sequences indicate clearly that the Diffusion Network had modelled both characters satisfactorily.

Let x_β represent the sequence of handwritten β. With the dynamics of visible units *clamped* to the handwritten β, i.e. setting $x_{Vi} = x_\beta$ for $t \in [0, 200]$, the deterministic dynamics of hidden units derived according to Eq.(5) are shown in Fig.3a. Contrarily, the deterministic dynamics of hidden units sampled when visible units are clamped to the handwritten ρ are shown in Fig.3b.

$$dx_i(t) = \mu_i(t) \cdot dt \tag{5}$$

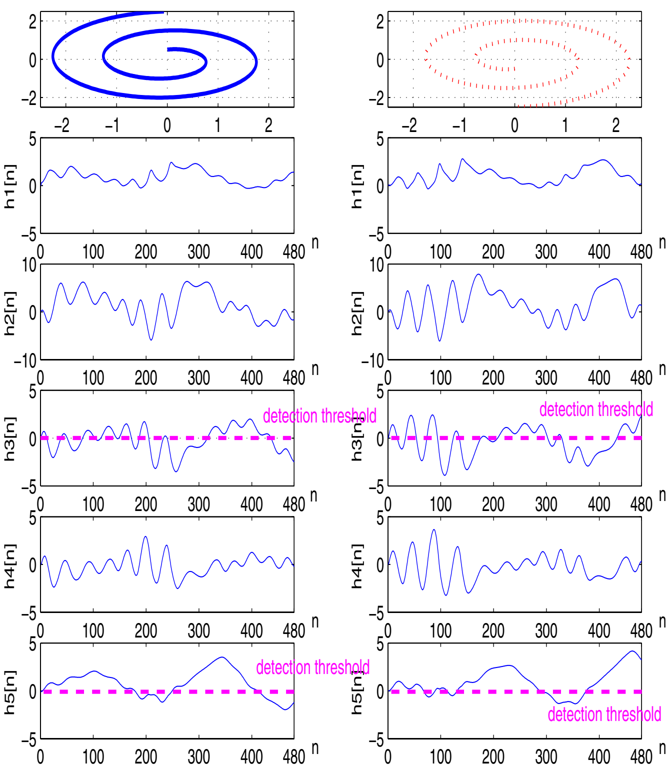
Eq.(5) differs from Eq.(1) by the exclusion of the Brownian motion, transforming the Diffusion Network into a deterministic recurrent network. Interestingly, Fig.3 shows that the deterministic dynamics of hidden units in response to β and ρ have significantly distintive features, especially for the final value sampled at $t = 200$. This allows us to distinguish between x_ρ and x_β simply by comparing the final values of hidden dynamics to the threshold values illustrated in Fig.3. Repeating the experiment gives similar results consistently, and the easily-distinguishable hidden dynamics are resulted from the from fact that x_β has a valley in the x-dimension during $t \in [110, 150]$, while x_ρ has a peak during the same period of time. The significantly-different dynamics are then amplified by the strong coupling between x_{V1} and x_{H4}, for example, leading x_{H4} to have very different terminal values for x_ρ and x_β(Fig.3).

This promising result suggests that the proposed classification is not only simple but reliable, as compared to the calculation of likelihood. In addition, the proposed method is more hardware-friendly, requiring simply a comparator for threshold detection. In a hardware-implemented Diffusion Network, the hidden dynamics can be obtained in real-time as visible dynamics are clamped to the sequences to be classified. This proposed method therefore facilitates real-time recognition greatly, which is crucial for applications like neural prostheses, requiring real-time recognition of multichannel neural signals for the control of prosthetic devices[1].



(a) (b)

Fig. 4. The two-dimensional spiral curves to be classified (b)The 20 regenerated visible sequences(solid lines) sampled from the Diffusion Network trained on one of the spiral curve(dashed line).



(a) (b)

Fig. 5. The dynamics of hidden units of the DN trained on the first spiral curve when visible units are clamped to (a)the first (b)the second spiral curves

V. Tolerance Against Noise and Offset

Biomedical signals are normally noisy and drifting. It is thus important to probe the tolerance of the proposed classification method against noise and offsets. The classification of spiral curves in Fig.4a is one of the most difficult artificial tasks in pattern classification[13], as the classification relies on capturing the correlation between the two dimensions of the spiral series, and then drawing a third nonlinear spiral curve to separate the two curves. To examine the Diffusion Network's tolerance against noise and offsets, the Diffusion Network was trained to model one of the spiral series, and then used to classify "distorted" spiral series obtained

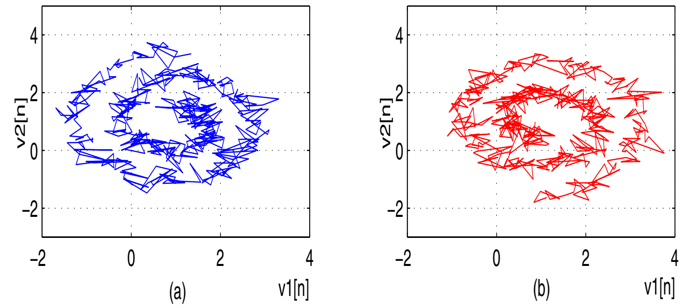


Fig. 6. The distorted spiral curves obtained by adding offsets and noise to the two types of spiral curves in Fig.4a.

TABLE I

The accuracy in classifying two types of spiral curves distorted by various levels of offsets and noise

$\delta_x = \delta_y$	$\sigma_x = \sigma_y$				
	0.2	0.4	0.6	0.8	1.0
0.4	100%	100%	100%	100%	100%
0.5	100%	100%	100%	100%	100%
0.6	100%	100%	100%	100%	100%
0.7	100%	100%	100%	100%	99%
0.8	100%	100%	97%	97%	94%
0.9	78%	71%	77%	74%	79%
1.0	51%	53%	62%	63%	67%

according to Eq.(6).

Let x_{S1} and x_{S2} : $[0, 480] \rightarrow \mathbb{R}^2$ represent the spiral series ending at (0,3) and (0,-3) in Fig.4, respectively. A Diffusion Network with two visible and five hidden neurons was trained to model x_{S1} with $\sigma = 0.2$, $R_i = 1$, and $a_i = 1$. After 80 training epochs, the Diffusion Network regenerated its visible dynamics as shown in Fig.4b, indicating that the spiral series has been modelled satisfactorily. With visible units of the trained Diffusion Network clamped to x_{S1} and to x_{S2}, the deterministic dynamics of hidden units are shown in Fig.5. As shown in Fig.5, the two different spiral curves can be easily distinguished by comparing the final values of the hidden dynamics x_{H3} or x_{H5} to the thresholds indicated by the dashed lines.

Let $x_{Six}(t)$ and $x_{Siy}(t)$ represent the coordinate of either x_{S1} or x_{S2} at time t. To probe the tolerance against noise and offsets, the trained Diffusion Network was employed to classify a set of "distorted" spiral curves $x_T = (x_{Tx}, x_{Ty}) : [0, 480] \rightarrow \mathbb{R}^2$ generated according to

$$x_{Tx}(t) = x_{Six}(t) + \delta_x + \sigma_x \cdot N(0, 1) \quad (6)$$

$$x_{Ty}(t) = x_{Siy}(t) + \delta_y + \sigma_y \cdot N(0, 1) \quad (7)$$

where δ_x and δ_y denote the offsets added to the x and y dimensions, respectively, $N(0, 1)$ denotes a zero-mean Gaussian noise with unit variance, and σ_x and σ_y scale the noise variance. Fig.6 illustrates two types of distorted spiral curves with $\delta_x = \delta_y = 1$ and $\sigma_x = \sigma_y = 1$. The distorted spiral curves x_T were classified according to the final value

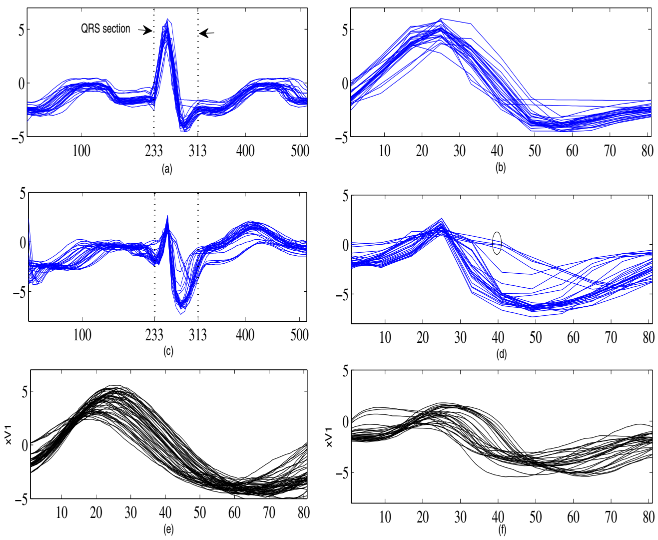


Fig. 7. (a)The normal ECGs and (b)its QRS section (corresponding to t=[233,313] in (a)) (c)The abnormal ECGs and (d) its QRS section. The circled ECG traces in (d) has less significant drop than others. (e)(f):The visible dynamics regenerated by the Diffusion Network after 100 epochs of training on (e)normal and (f)abnormal ECGs.

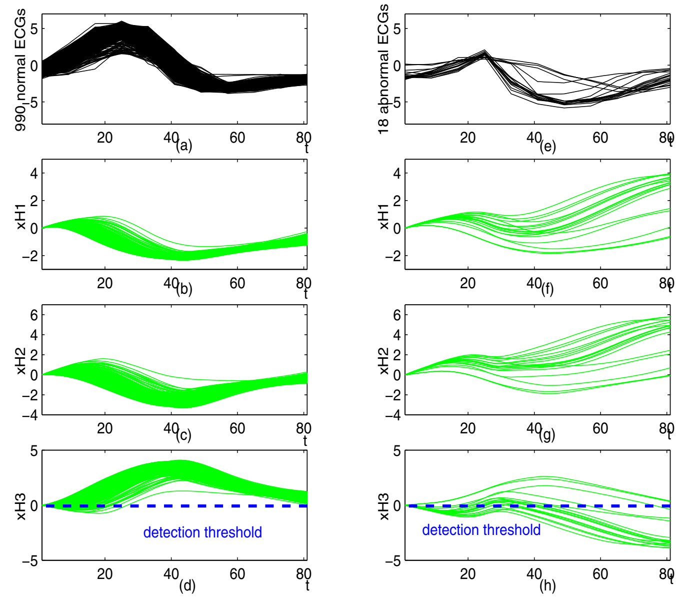


Fig. 8. The deterministic dynamics of the three hidden units of the trained Diffusion Network sampled when visible units are clamped to normal ECGs((a)-(d)), and abnormal ECGs((e)-(h))

of x_{H3} when the dynamics of visible units are clamped to x_T. If $x_{H3}[480] > 0$, x_T is classified as x_{S1}, while if $x_{H3}[480] \leq 0$, x_T is classified as x_{S2}.

Table I summarises the accuracy in classifying the two types of spiral curves distorted with various levels of offsets and noise. The results indicate that the Diffusion Network with the proposed classification method exhibits remarkable tolerance against noise and offsets. With offsets smaller than 0.6, the Diffusion Network is able to classify the spiral curves with 100% accuracy, regardless of the levels of noise added into the spiral curves. With noise variance smaller than 0.6, the Diffusion Network is also able to tolerate an offset up to 0.7 while maintaining an accuracy of 100%. The tolerance can be further improved if noise and offsets are added to only one dimension. Moreover, Table.I reveals that the Diffusion Network tolerate noise better than offsets, owing to that offsets impact deterministic dynamics much more than zero-mean noise.

The promising performance above supports the suggestion in [3] that the Diffusion Network is able to use its stochasticity to generalise variability in data of the same type. As deterministic hidden dynamics are obtained by "turning off" the noise, the proposed classification method would effectively lead the Diffusion Network to ignore the variability in data and thus to classify data reliably.

VI. Recognition of Real Heartbeat Data

Electrocardiograms (ECGs) extracted from the MIT-BIH database were further used to examine the Diffusion Network's ability to classify real biomedical signals. Fig.7a-d shows the extracted normal and abnormal ECGs (The latter are called *ventricular ectopic beats*). All ECGs are sampled as discrete-time, one-dimensional sequences with $t \in [0, 513]$, and are aligned by centering their most positive peaks at $t = 257$. As the abnormal ECGs shown in Fig.7 are mainly diagnosed according to the QRS section of the ECGs in clinics[14], we simply trained the Diffusion Network to learn the QRS sections of normal and abnormal ECGs in Fig.7b and Fig.7d, respectively.

A Diffusion Network with one visible and three hidden units was trained to model ten normal ECGs sampled from the dataset in Fig.7b. The training aimed to regenerates the one-dimensional ECG traces as the dynamics of the visible unit, while the minimum number of hidden units required for modelling the ECGs was determined by experimental trials. With $\sigma = 0.2$, $R_i = 1$, $a_i = 1$ and after 100 training epochs, the Diffusion Network regenerated visible dynamics as shown in Fig.7e. The similarity between Fig.7e and Fig.7b indicates that the QRS section of normal ECGs has been modelled satisfactorily. Similar result is obtained in the experiment of modelling the QRS section of abnormal ECGs, as shown in Fig.7f.

The Diffusion Network trained on normal ECGs were then used to generate deterministic dynamics of hidden units, in response to the clamping of visible units to 990 normal and 18 abnormal ECGs, sampled from a 30-minutes long ECG recording provided by the MIT-BIH database. This task examines the Diffusion Network's ability to detect abnormal ECGs nearly in real time during the ECG recording of a patient. Fig.8 shows the deterministic dynamics sampled when visible units were clamped to the QRS section of normal ECGs and abnormal ECGs. Even if the QRS section of both types have similar initial and end values, Fig.8 reveals that the hidden units, especially for x_{H3}, in the trained Diffusion Network has significantly distinctive dynamics. Therefore, abnormal ECGs can be easily detected by comparing $x_{H3}[81]$ to the threshold indicated as the dashed line in Fig.8. If $x_{H3}[81] > 0$, the trace is classified as normal ECGs, and if $x_{H3}[81] < 0$ the trace is classified as abnormal ECGs. Based on the classification method, only two abnormal ECGs are misclassified, as shown by Fig.8h. The accuracy is thus proved to be 99.8% (calculated as $1006/1008$). The slight inaccuracy mainly come from the fact that the S segment of several abnormal ECGs do not have significantly lower value than that of normal ECG, as shown in Fig.7d.

VII. Conclusion

This paper investigates the feasibility of applying the Diffusion Network, a stochastic recurrent network, to the recognition of continuous-time, continuous-valued biomedical signals. A new classification method is also proposed to facilitate real-time recognition especially for hardware implementation. Simulation results demonstrate that the Diffusion Network can classify both artificial and heartbeat data reliably with considerable tolerance against noise and offsets, demonstrating the advantage of using stochasticity to generalise variability in data. As the dynamics of the Diffusion Network are governed by stochastic differential equations which are shown hardware-amenable, it is important to look into the VLSI implementation of the Diffusion Network, making the Diffusion Network useful for real biomedical applications.

VIII. Acknowledgement

The authors would like to thanks to the National Science Council in Taiwan for funding this research (Grant code : NSC 96-2221-E-007 -167 -MY3).

References

[1] M. Lebedev and M. Nicolelis, "Brain-machine interfaces: past, present and future," *TRENDS in Neuroscience*, vol. 29, no. 9, pp. 536–546, 2006.

[2] J. R. Movellan, "A learning theorem for networks at detailed stochastic equilibrium," *Neural Computation*, vol. 10, no. 5, pp. 1157–1178, 1998.

[3] J. R. Movellan, P. Mineiro, and R. J. Williams, "A Monte-Carlo EM approach for partially observable diffusion processes: Theory and applications to neural networks," *Neural Computation*, vol. 14, no. 7, pp. 1507–1544, 2002.

[4] H. Chen and A. F. Murray, "A continuous restricted boltzmann machine with an implementable training algorithm," *IEE Proceedings of Vision, Image and Signal Processing*, vol. 150, no. 3, pp. 153–158, 2003.

[5] D. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. *, pp. 109–118, 1990.

[6] H. Chen and A. F. Murray, "Continuous-valued probabilistic behaviour in a vlsi generative model," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 755–770, 2006.

[7] E. Wong and B. Hajek, *Stochastic Process in Engineering Systems*, 2nd ed. Springer-Verlag New York Inc., 1971.

[8] L. Chua and T. Roska, "The cnn paradigm," *IEEE Transactions on Circuits and Systems-I: Foundamental Theory and Applications*, vol. 40, no. 3, pp. 147–156, 1993.

[9] L. Chua and L. Yang, "Cellular neural networks: Theory," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 10, pp. 1257–1272, 1988.

[10] L. Chou and L. Yang, "Cellular neural networks: Applications," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 10, pp. 1273–1290, 1988.

[11] T. Roska and L. Chua, "The cnn universal machine: An analogic array computer," *IEEE Transactions on Circuits and Systems Ii-Analog and Digital Signal Processing*, vol. 40, no. 3, pp. 163–173, 1993.

[12] L. Chua, T. Roska, T. Kozek, and A. Zarandy, "Cnn universal chips crank up the computing power," *IEEE Circuits and Devices Magazine*, vol. 12, no. 4, pp. 18–28, 1996.

[13] S. Singh, "Quantifying structural time varying changes in helical data," *Technical Reort*, *.

[14] P. Gomis, K. Suarez, G. Passariello, I. Mendoza, P. Caminal, and P. Lander, "Abnormal intra-qrs signals and late potentials in the high-resolution ecg associated with chagasic myocarditis," *IEEE Computers in Cardiology*, vol. 8, no. 11, pp. 633–636, 1996.