How Robust Is a Probabilistic Neural VLSI System Against Environmental Noise

C.C. Lu, C.C. Li, and H. Chen

The Dept. of Electrical Engineering,
The National Tsing-Hua University, Hsin-Chu, Taiwan 30013
`hchen@ee.nthu.edu.tw`

Abstract. Implementing probabilistic models in the Very-Large-Scale-Integration (VLSI) has been attractive to implantable biomedical devices for improving sensor fusion and power management. However, implantable devices are normally exposed to noisy environments which can introduce non-negligible computational errors and hinder optimal modelling on-chip. While the probabilistic model called the Continuous Restricted Boltzmann Machine (CRBM) has been shown useful and realised as a VLSI system with noise-induced stochastic behaviour, this paper investigates the suggestion that the stochastic behaviour in VLSI could enhance the tolerance against the interferences of environmental noise. The behavioural simulation of the CRBM system is used to examine the system's performance in the presence of environmental noise. Furthermore, the possibility of using environmental noise to induce stochasticity in VLSI for computation is investigated.

1 Introduction

In the development of implantable devices [1][2] and bioelectrical interfaces [3][4], exposing electronic systems to the noisy environment becomes inevitable. Although noisy data could be transmitted wirelessly out of implanted devices and processed by sophisticated algorithms, transmitting all raw data is power-consuming, and is unfavourable for long-term monitoring. Therefore, an intelligent embedded system which is robust against noise and able to extract useful information from high dimensional, noisy biomedical signals becomes essential. The Continuous Restricted Boltzmann Machine (CRBM) is a probabilistic model both useful in classifying biomedical data and amenable to the VLSI implementation [5]. The usefulness comes from the use of noise-induced stochasticity to represent natural variability in data. The VLSI implementation further explores the utility of noise-induced, continuous-valued stochastic behaviour in VLSI circuits [5]. This leads to the suggestion that stochastic behaviour in VLSI could be useful for discouraging environmental noise and computation errors. Therefore, based on the well-defined software-hardware mapping derived in [5] this paper use behavioural simulation to examines the maximum external noise that the CRBM system can tolerate to model both artificial and real biomedical (ECG) data. The possibility of using environmental noise to replace on-chip noise generators is also investigated.
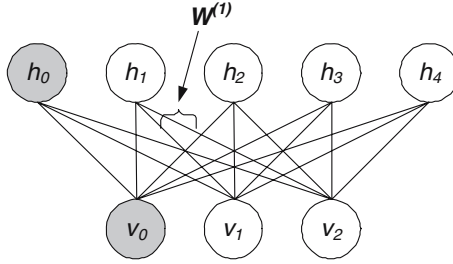
Fig. 1. The architecture of a CRBM model with two visible and four hidden neurons. v_0 and h_0 represent biasing unit with invariant outputs $v_0 = h_0 = 1$.

2   The CRBM Model

The CRBM model consists of one visible and one hidden layers of continuous-valued, stochastic neurons with inter-layer connections, as shown in Fig.1. Circle v_i represents visible neurons while circle h_j represents hidden neurons. The number of visible neurons corresponds to the dimensions of modeled data, while that of hidden neurons is chosen according to data complexity. Let s_i denotes the state of neuron v_i or h_j, and w_{ij} represents the bi-directional connection between v_i and h_j. The stochastic behaviour of a neuron s_i is described by [5]

$$s_i = \varphi_i \left(a_i \cdot \left(\Sigma_j w_{ij} \cdot s_j + N_i(\sigma, 0) \right) \right) \tag{1}$$

where $N_i(\sigma, 0)$ represents a zero-mean Gaussian noise with variance σ^2, and $\varphi_i(\cdot)$ a sigmoid function with asymptotes at ± 1 and slope controlled by a_i. As a *generative* model, the CRBM learns to "regenerate" training data distributions in its visible neurons. Testing data can then be categorised according to the responses of hidden neurons [5]. The training algorithm implemented in the CRBM system is defined by the following equation [6]

$$\triangle \lambda = \eta_\lambda \cdot \left(\langle s_i \cdot s_j \rangle_4 - \langle \hat{s}_i \cdot \hat{s}_j \rangle_4 \right) \tag{2}$$

where λ represent parameters w_{ij} or a_i, η_λ the updating rate, $\hat{s}_i$ and $\hat{s}_j$ the one-step Gibbs-sampled states. $\langle \cdot \rangle_4$ stands for taking the expectation over four training data. For parameter a_i, the training algorithm is the same with Eq.(2) but simply replace s_j and $\hat{s}_j$ by s_i and $\hat{s}_i$, respectively.

The modelled distribution of a trained CRBM is obtained by initializing visible neurons with random values, and then Gibbs sampling hidden and visible neurons alternatively for multiple steps. The N-th step samples of visible neurons are called the *N-step reconstruction*, and it approximates the modelled distribution when N is large. The similarity between *N-step reconstruction* and training data indicates how well training data is modelled.

3   The Robustness Against Environmental Noise

Following a brief introduction to the architecture of the CRBM system, this section investigates the influences of environmental noise on the performance of the CRBM system.

3.1   The CRBM System

The prototype CRBM system containing two visible and four hidden neurons has been demonstrated able to reconstruct two-dimensional artificial data with noise-induced stochastic behaviour in VLSI [5]. Fig.2 shows the modular diagram of the CRBM systems excluding its learning circuits. The CRBM neurons mainly comprise of multipliers to calculate the products $(w_{ij} \cdot s_j)$ in Eq.(1), and sigmoid circuits with $\{a_i\}$ controlling
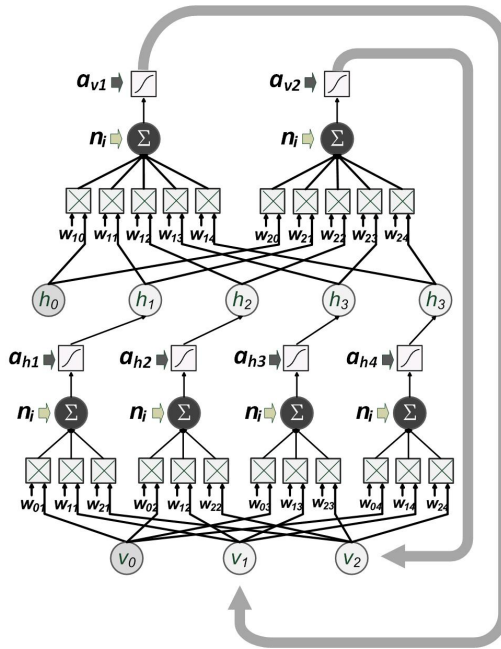

Fig. 2. The modular diagram of the CRBM system with two visible and four hidden neurons

Table 1. The mapping of parameter values between software simulation and hardware implementation

	Matlab	VLSI(V)
s_i	[-1.0, 1.0]	[1.5, 3.5]
w_{ij}	[-2.5, 2.5]	[0.0, 5.0]
a_i	[0.5, 9.0]	[1.0, 3.0]

the slope of φ_i. On the other hand, multi-channel, uncorrelated noise $\{n_i\}$ are injected into the neurons to make the outputs, $\{v_i\}$ and $\{h_i\}$, probabilistic. Parameters $\{w_{ij}\}$ and $\{a_i\}$ are stored as voltages across capacitors, and are adaptable by on-chip learning circuits. The learning circuits can also refresh $\{w_{ij}\}$ and $\{a_i\}$ to specific values after training. Table 1 summarises the mapping for all parameters between software simulation and VLSI implementation, which has been proved useful for simulating the effects of non-ideal training offsets on the performance of the CRBM system [5].

In an implantable device containing digital-signal-processing circuits, multi-purpose sensors, and wireless transceivers, a VLSI system unavoidably suffers from various environmental noise including substrate noise, sensory noise, and electromagnetic interferences. As these interferences mainly affect the precision of voltage signals in VLSI, voltage-represented s_i, w_{ij}, and a_i in the CRBM system are expected to experience serious effects, while the influence on current-mode learning circuits are assumed to be negligible. Therefore, the influence of environmental noise on the CRBM system was simulated by replacing s_i, w_{ij}, and a_i in Eq.(1) and Eq.(2) by the following equations

$$w' = w + n_w$$
$$a' = a + n_a \tag{3}$$
$$s' = s + n_s$$

where n_s, n_w, and n_a represent zero-mean, uncorrelated noise with either Gaussian or Uniform distributions.

3.2   Modelling Artificial Data in the Presence of Environmental Noise

To illustrate the characteristics of the CRBM, as well as to identify a quantitative index for how well the CRBM models a dataset, the CRBM with two visible and four hidden neurons was first trained to model the artificial data in Fig.3(a) in the absence of environmental noise. The training data contains one elliptic and one circular clusters of 1000 Gaussian-distributed data points. With $\sigma = 0.2$, $\eta_w = 0.02$, $\eta_a = 0.2$, and after 15,000 training epochs, the CRBM regenerated the 20-step reconstruction of 1000 data points as shown in Fig.3(b), indicating that the CRBM has modelled data. While visual comparison between Fig.3(a) and (b) can hardly tell how well the data is modelled, the following index is employed to measure the similarity quantitatively.

Let $P^T(\mathbf{v})$ and $P^M(\mathbf{v})$ represent the probability distribution of training data and that modelled by the CRBM, respectively. The *Kulback-Leibler (KL) Divergence* defined as Eq.(4) [7] measures the difference between $P^T(\mathbf{v})$ and $P^M(\mathbf{v})$.

$$G = \Sigma_{\mathbf{v}} P^T(\mathbf{v}) log \frac{P^T(\mathbf{v})}{P^M(\mathbf{v})} \tag{4}$$

where $\mathbf{v}$ denotes the subset of visible space, and G equals zero when $P^T(\mathbf{v}) = P^M(\mathbf{v})$. As explicit equations for describing the modelled distribution, $P^M(\mathbf{v})$, are normally intractable, $P^T(\mathbf{v})$ and $P^M(\mathbf{v})$ were statically-estimated by dividing the two-dimensional space into 10x10 square grids, counting the number of data points in each grid, and
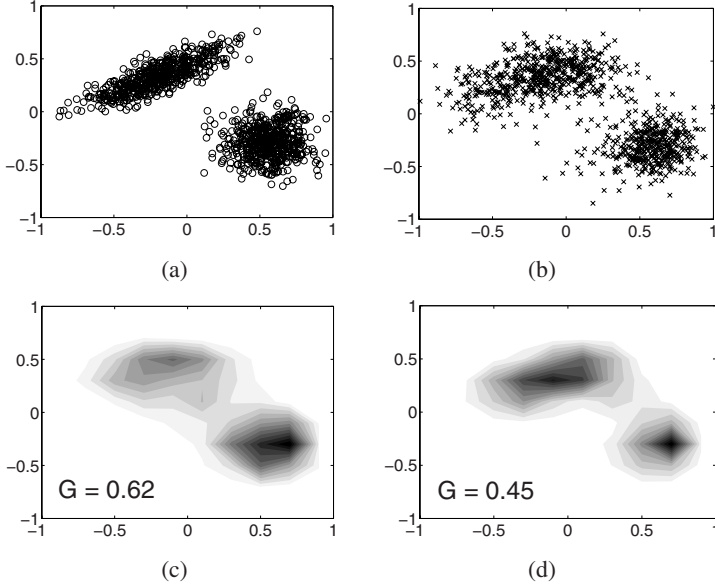
Fig. 3. (a)Artificial training data. (b)The 20-step reconstruction with 1000 points generated by the CRBM trained after 15,000 epochs. (c)The statistical density after 12,500 epochs. (d)The statistical density corresponding to (b).

normalising the counts with respect to the total number of data points. Fig.3(c)-(d) shows the statistical density of the 20-step reconstructions generated by the CRBM after 12,500 and 15,000 training epochs, respectively. The G values calculated according to Eq.(4) are also shown at the bottom-left corners. This result indicates the probability distribution of training data is modelled with negligible errors as G is smaller than 0.45(Fig.3(d)).

Based on the criterion of $G \leq 0.45$, the maximum environmental noise, both Gaussian- and uniformly-distributed, the CRBM can tolerate to model the artificial data are identified and summarised in Table 2. The maximum tolerable noise levels are expressed in terms of voltages based on the mapping in Table 1, and obviously, the tolerable levels are much greater than the noise levels existing in contemporary VLSI technologies. The first three rows in Table 2 show the tolerance identified when only one type of parameters experiences noise. Parameter $\{a_i\}$ has slightly smaller tolerance than $\{w_{ij}\}$ and $\{s_i\}$ because the mapping for $\{a_i\}$ has a largest ratio between the numerical and voltage ranges. This leads $\{a_i\}$ to experience largest numerical errors in the existence of the same noise levels. The forth row in Table 2 shows the tolerance when noise exists in all parameters, the more realistic case. The tolerance is not seriously degraded, indicating that the training algorithm of the CRBM system can compensate for noise-induced errors among parameters and maintain a satisfactory tolerance. Finally, Table 2 reveals that the tolerance against Gaussian-distributed noise is much better than

Table 2. The maximum gaussian- and uniformly-distributed noise tolerable by the CRBM system during modelling artificial data

	Gaussian-distributed noise	Uniformly-distributed noise
n_w	[-0.27V, 0.27V]	[-0.18V, 0.18V]
n_a	[-0.16V, 0.16V]	[-0.08V, 0.08V]
n_s	[-0.18V, 0.18V]	[-0.06V, 0.06V]
n_w,n_a,n_s	[-0.13V, 0.13V]	[-0.04V, 0.04V]
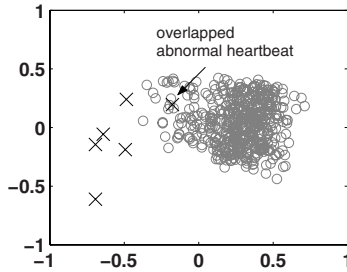
Fig. 4. The projection of 500 ECG training data to its first two principle components. The projection of the five abnormal ECGs in the dataset are denoted by black crosses.

that against uniformly-distributed noise. This is attributed to the fact that the training data and the noise incorporated in the CRBM neurons are Gaussian-distributed. From another point of view, if the distribution of environmental noise is known, training the CRBM system with noise inputs $\{n_i\}$ modified to have the known distribution can enhance the tolerance against a specific type of noise.

3.3 Modelling Biomedical Data in the Presence of Environmental Noise

The tolerable environmental noise for modelling high-dimensional, real-world data was examined in the context of recognising electrocardiograms (ECG), extracted from the MIT-BIH database as in [6]. The training dataset contains 500 heartbeats with only 5 abnormal heartbeats. The testing dataset contains 1700 heartbeats with 27 abnormal heartbeats. Each heartbeat is sampled as a 65-dimemsional datum, and Fig.4 shows the projection of the training dataset onto its first two principle components. Although the dimension reduction makes the quantitative index G remain applicable, pilot simulation showed that modelling training data satisfactorily does not guarantee the detection of abnormal heartbeats with 100% accuracy. This is because the distributions of normal and abnormal heartbeats overlap with each other, as shown in Fig.4. Therefore, detecting abnormal heartbeats with 100% accuracy was used as a stricter criterion for identifying the tolerable noise during modelling ECG data.
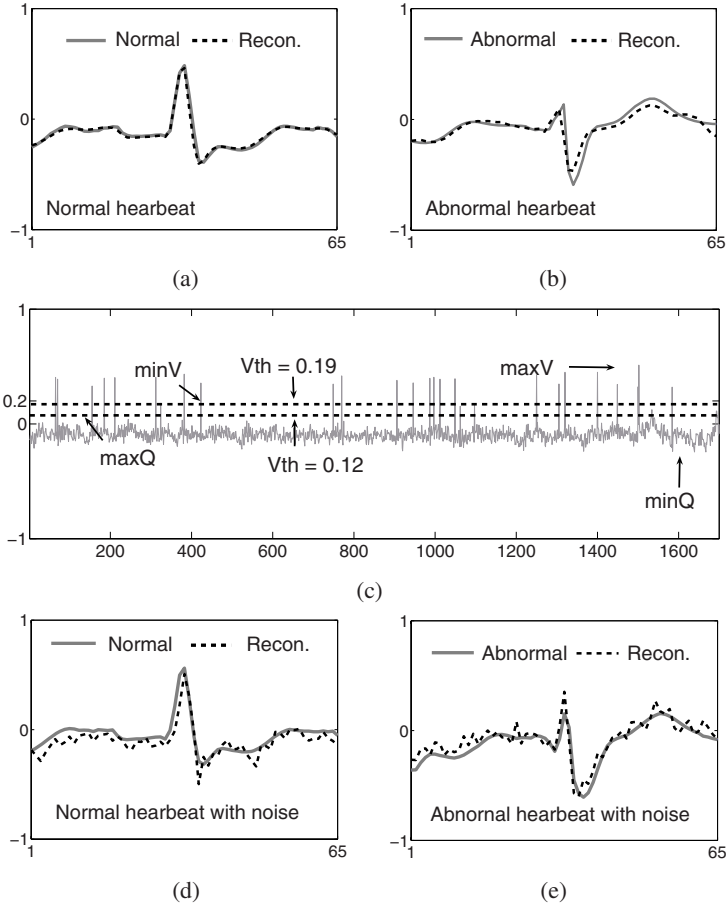
Fig. 5. (a) (b) Heartbeat signals sampled from training data (solid) and reconstruction generated by the trained CRBM (dashed) without noise. (c) Corresponding response of hidden neuron to 1700 testing data. (d)(e) Heartbeat signals and reconstruction generated by the trained CRBM with uniformly-distributed external noise in the range [-0.01, 0.01].

Fig.5(a)(b) shows heartbeat signals reconstructed by a CRBM system trained without noise. Fig.5(c) shows the response of hidden neuron h_2 to 1700 testing data $\{\mathbf{d}\}$, calculated according to Eq.(5). The abnormal heartbeats can be detected with 100% accuracy by setting any threshold between minV and maxQ.

$$h_2 = \varphi_2(a_2 \cdot (\mathbf{w}^{(2)} \cdot \mathbf{d})) \tag{5}$$

With uniformly-distributed noise ranging between -0.01V and 0.01V, the trained CRBM system was able to reconstruct both normal and abnormal ECG signal satisfactorily, as shown in Fig.5(d)(e). Comparison between Fig.5(a)(b) and Fig.5(d)(e) indicates that the influence of environmental noise injection introduce extra fluctuations in the waveform.

Table 3. The maximum gaussian- and uniformly-distributed noise tolerable by the CRBM system to detect abnormal ECGs reliably

	Gaussian-distributed noise	Uniformly-distributed noise
n_w	[-0.39V, 0.39V]	[-0.29V, 0.29V]
n_a	[-0.15V, 0.15V]	[-0.11V, 0.11V]
n_s	[-0.09V, 0.09V]	[-0.05V, 0.05V]
n_w,n_a,n_s	[-0.03V, 0.03V]	[-0.03V, 0.03V]

Table 3 summarises the maximum environmental noise the CRBM can tolerate to model and to detect abnormal ECGs with 100% accuracy. The tolerable noise levels in Table 3 are mostly smaller than those in Table 2 because the ECG data require more sophisticated modelling. On the contrary, parameter $\{w_{ij}\}$ has comparable tolerance in both modelling tasks because the presence of n_w simply distorts the weight vectors and results in fluctuated reconstructions like those in Fig.5(d)(e), while these effects do not impede the CRBM from modelling the distinguishable features between abnormal and normal ECGs. This also explains the relatively smaller tolerance against n_s, i.e. the noise in $\{s_i\}$, as n_s does distort the features of ECGs, making it difficult to identify any distinguishable feature between normal and abnormal ECGs.

4   Noise-Enhanced Robustness in the CRBM System

By reducing the standard deviation of the noise injected into CRBM neurons to $\sigma = 0.1$ in Eq.(1), the maximum environmental noise the CRBM can tolerate to model the artificial data in Fig.3(a) is summarised in Table 4. Comparison between Table 2 and Table 4 indicates that reducing $\sigma = 0.1$ helps to enhance the robustness against environmental noise, as part of environmental noise is incorporated to compensate for the reduction of the "internal noise" n_i, which is essential for inducing stochasticity for modelling the variability of training data. Therefore, as shown by Table 4, the robustness against environment noise, especially for uniformly-distributed noise, is improved significantly. The worst tolerable level is still greater than 110mV, corresponding to a signal-to-noise ratio less than 20 for a CRBM system. The advantage of incorporating noise-induced stochasticity in VLSI is clearly demonstrated.

Furthermore, it is interesting to investigate whether the internal noise n_i could be completely replaced by environmental noise to induce stochasticity for computation. By substituting $s_j + n_s$ for s_j and setting $N_i(\sigma, 0) = 0$ in Eq.(1), the term $\sum w_{ij} \cdot (s_j + n_s)$ becomes a random variable n'_i with mean value $\sum w_{ij} \cdot s_j$ and a variance given as

$$var(n'_i) = \Sigma_j w_{ij}^2 \cdot var(n_s) \qquad (6)$$

If n'_i has the same variance as n_i, the stochasticity induced by n_s should have the same level as that induced by n_i in Fig.2. The CRBM trained on the artificial data was found

Table 4. The tolerable interference of external gaussian- and uniformity-distributed noise with reduced variations of on-chip noise generator to model artificial data

	Gaussian-distributed noise	Uniformy-distributed noise
n_w	[-0.46V, 0.46V]	[-0.46V, 0.46V]
n_a	[-0.33V, 0.33V]	[-0.14V, 0.14V]
n_s	[-0.28V, 0.28V]	[-0.18V, 0.18V]
n_w, n_a, n_s	[-0.16V, 0.16V]	[-0.11V, 0.11V]

(a)   (b)
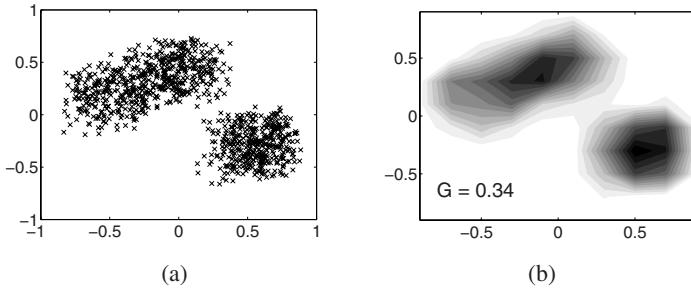
Fig. 6. (a) The 20-step reconstruction with 1000 points to model artificial in Fig.3(a) generated by the CRBM trained after 15,000 epochs. (b) The statistical density corresponding to (a).

to have an average value between 0.8 and 1 for $\Sigma_j w_{ij}^2$. According to Eq.6, a uniformly-distributed noise n_s ranging between [-0.3V, 0.3V] will yield an equivalent noise n_i' with a variance of 0.04. Fig.6 shows the results of training the CRBM to model the artificial data in Fig.3(a) without n_i, but with uniformly-distributed noise n_s in the range [-0.3V, 0.3V]. Fig.6(a) and (b) depicts the 20-step reconstruction and its statistical density, respectively. The corresponding G value is 0.34, revealing that the trained CRBM has used the stochasticity induced by environmental noise in $\{s_i\}$ to model data optimally. This supports the suggestion in [8] that intrinsic noise of MOSFETs could be used rather than suppressed to achieve robust computation, based on algorithms like the CRBM. This is especially important when the VLSI technology moves towards the deep-sub-micron era.

5   Conclusion

The behavioural simulation of the CRBM system demonstrates that the CRBM system has satisfactory robustness against environmental noise, confirming the potential of using the CRBM system as an intelligent system in implantable devices. As the promising performance mainly comes from the incorporation of noise-induced stochasticity, the robustness can be further enhanced if the distribution of environmental noise is known and incorporated during training, or by reducing the internal noise of the CRBM system.

It is also demonstrated that environmental noise can be used to induce the stochasticity essential for the CRBM system to model data optimally. In other words, the robustness of the CRBM system can be optimised by training the CRBM to model data with stochasticity induced by environmental noise the CRBM system is exposed to. All these concepts will be further examined by hardware testing with the CRBM system.

References

1. Tong, B.T., Johannessen, E.A., Lei, W., Astaras, A., Ahmadian, M., Murray, A.F., Cooper, J.M., Beaumont, S.P., Flynn, B.W., Cumming, D.R.S.: Toward a miniature wireless integrated multisensor microsystem for industrial and biomedical applications. Sensors Journal 2(6), 628–635 (2002)
2. Johannessen, E.A., Lei, W., Li, C., Tong, B.T., Ahmadian, M., Astaras, A., Reid, S.W.J., Yam, P.S., Murray, A.F., Flynn, B.W.A., Beaumont, S.P.A., Cumming, D.R.S., Cooper, J.M.A.: Implementation of multichannel sensors for remote biomedical measurements in a microsystems format. IEEE Transactions on Biomedical Engineering 51(3), 525–535 (2004)
3. Mingui, S., Mickle, M., Wei, L., Qiang, L., Sclabassi, R.J.: Data communication between brain implants and computer. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11(2), 189–192 (2003)
4. Nicolelis, M.A.L.: Actions from thoughts. Nature 409, 403–407 (2001)
5. Hsin, C., Fleury, P.C.D., Murray, A.F.: Continuous-valued probabilistic behavior in a VLSI generative model. IEEE Transactions on Neural Networks 17(3), 755–770 (2006)
6. Chen, H., Murray, A.F.: Continuous restricted Boltzmann machine with an implementable training algorithm. IEE Proceedings-Vision Image and Signal Processing 150(3), 153–158 (2003)
7. Hinton, G.E., Sejnowski, T.J.: Learning and Relearning in Boltzmann Machine. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, pp. 283–317. MIT, Cambridge (1986)
8. Hamid, N.H., Murray, A.F., Laurenson, D., Roy, S., Binjie, C.: Probabilistic computing with future deep sub-micrometer devices: a modelling approach. In: IEEE International Symposium on Circuits and Systems, pp. 2510–2513 (2005)